POPULATION AND COMPARATIVE GENOMICS OF PLANT PATHOGENIC BACTERIA

EDITED BY: Jeffrey Jones, Erica M. Goss, Jonathan Michael Jacobs, Ralf Koebnik, Neha Potnis, Sujan Timilsina and Veronica Roman-reyna

PUBLISHED IN: Frontiers in Microbiology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-892-9
DOI 10.3389/978-2-88976-892-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

POPULATION AND COMPARATIVE GENOMICS OF PLANT PATHOGENIC BACTERIA

Topic Editors:
Jeffrey Jones, University of Florida, United States
Erica M. Goss, University of Florida, United States
Jonathan Michael Jacobs, The Ohio State University, United States
Ralf Koebnik, Plant Health Institute, France
Neha Potnis, Auburn University, United States
Sujan Timilsina, University of Florida, United States
Veronica Roman-reyna, The Ohio State University, United States

Citation: Jones, J., Goss, E. M., Jacobs, J. M., Koebnik, R., Potnis, N., Timilsina, S., Roman-reyna, V., eds. (2022). Population and Comparative Genomics of Plant Pathogenic Bacteria. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-892-9

Table of Contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 31 August 2022
DOI 10.3389/fmicb.2022.1012034

OPEN ACCESS

EDITED AND REVIEWED BY
Trevor Carlos Charles,
University of Waterloo, Canada

*CORRESPONDENCE
Sujan Timilsina
sujan.timilsina@ufl.edu
Erica M. Goss
emgoss@ufl.edu
Ralf Koebnik
ralf.koebnik@ird.fr
Neha Potnis
nzp0024@auburn.edu
Jeffrey B. Jones
jbjones@ufl.edu

SPECIALTY SECTION
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

RECEIVED 05 August 2022
ACCEPTED 17 August 2022
PUBLISHED 31 August 2022

CITATION
Timilsina S, Goss EM, Koebnik R, Potnis N and Jones JB (2022) Editorial: Population and comparative genomics of plant pathogenic bacteria. *Front. Microbiol.* 13:1012034. doi: 10.3389/fmicb.2022.1012034

Editorial: Population and comparative genomics of plant pathogenic bacteria

Sujan Timilsina[1]*, Erica M. Goss[1,2]*, Ralf Koebnik[3]*, Neha Potnis[4]* and Jeffrey B. Jones[1]*

[1]Department of Plant Pathology, University of Florida, Gainesville, FL, United States, [2]Emerging Pathogens Institute, University of Florida, Gainesville, FL, United States, [3]Plant Health Institute of Montpellier, University of Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France, [4]Department of Entomology and Plant Pathology, Auburn University, Auburn, AL, United States

KEYWORDS

taxonomy, evolution, ecology, effector, host specificity

Editorial on the Research Topic
Population and comparative genomics of plant pathogenic bacteria

Plant pathogenic bacteria are among the major challenges in crop production, amplified by their continuous and rapid evolution and resulting in the emergence of virulent, competitive, and elusive plant pathogens. To further understand these bacteria, their interactions with host plants, and their responses to disease management, comprehensive methods are necessary. In recent years, phytobacteriology research has used population and comparative studies to quantify the diversity of plant pathogenic bacteria and their virulence mechanisms, including the effectors secreted from different protein secretion systems, phytohormones, or toxins that mediate their interactions with the hosts and/or competing microbes in the phyllosphere. Studies comparing strains over time, within and among geographic regions, and across phylogenetic trees has provided insights into scales of variation and changes in pathogens over time and in response to selection pressures. In this series of articles, population and comparative genomics were used to investigate plant pathogens representing diverse bacterial genera including Gram-negative bacteria composed of facultative anaerobes (i.e., *Dickeya, Pantoea,* and *Pectobacterium*) as well as obligate aerobes (i.e., *Pseudomonas*, *Xanthomonas*, and *Xylella*) as well as the Gram-positive genus, *Clavibacter*.

Genomics approaches clarify taxonomy and provide resources for comparative studies

Since ANI has been introduced as a means to reevaluate taxonomic assignments, genomics approaches have been used for comparative studies at different taxonomic levels. In this Focus Issue, reference pathotype strains of all described *Xanthomonas translucens* pathovars were sequenced using PacBio long-read technology and complete

genome sequences were assembled, which gave robust proof for three major clades in this species, each with distinct host preferences (Goettelmann et al.). Moreover, type III effector (T3E) profiling revealed 21 TAL effector classes and 29 additional type III effectors, some of them displaying clade or strain specificity. Further investigation of these genes could help to identify genes that are critically involved in pathogenicity and/or host adaptation, setting the grounds for the development of new resistant cultivars.

Expansion of whole genome sequencing of species within the same genus has offered insighst into host specialization and into phylogenetic relationships within and between species within genera. Prior to the identification of the *Xylella* species, *X. taiwanensis*, characterization was restricted to different subspecies of *Xylella fastidiosa*, but lacked analyses at the genus level due to the lack of genome sequences for multiple species (Weng et al.). As a result of having the *X. taiwanensis* genome available as an outgroup in phylogenetic comparisons of *X. fastidiosa* subspecies, there was strong support for *X. fastidiosa* subspecies *pauca* being the basal lineage of *X. fastidiosa* and for *Xylella* being derived from the paraphyletic genus *Xanthomonas*.

Genome sequencing is also revealing previously unrecognized diversity, as is illustrated by the characterization of *Xanthomonas* causing bacterial leaf canker on water spinach (*Ipomoea aquatic*) by Hu et al. a survey of the disease in Dongguan City, China, where water spinach is grown in small plots, found the *Xanthomonas* causal agent along with *Pantoea*, which had a synergistic effect on disease. Whole genome sequencing using short and long-read technologies of a representative *Xanthomonas* strain showed >97% ANI to *X. perforans*, which is a classic pathogen of tomato, but was more recently also identified on pepper and eucalyptus. The ANI below 98% indicates that this is an atypical *X. perforans*. The catalog of effectors in this strain will advance efforts to understand the relationships between effector repertoires and host range in *Xanthomonas*.

Soft rot pectobacteria are devastating plant pathogens with a global distribution and a broad host range. To close the shortage for genome data of *Pectobacterium aroidearum*, the causal agent of soft rot on the popular houseplant *Syngonium podophyllum*. Xu et al. present the complete genome sequence of strain L6. About 10% of the predicted genes were potentially related to pathogenesis using the Virulence Factors of Pathogenic Bacteria database, including genes related to toxins, plant cell-wall degrading enzymes, and bacterial secretion systems. This study provides novel information for the discovery of potential pathogenicity factors and the development of more effective strategies against this pathogen.

Historically, less work has been done on Gram-positive plant pathogens. Among them, *Clavibacter michiganensis*, a member of the Actinobacteria, is a causal agent of bacterial canker of tomato. Oh et al. provide the complete genome sequence of the type strain of *C. michiganensis*. When compared with other strains, it turned out that the chromosomal DNA sequences were almost identical, whereas its plasmids were found to carry distinct gene content among *C. michiganensis* strains. The genome information of the type strain LMG 7333^T will help in understanding the genetic diversity of *Clavibacter*'s plasmid complements and how they may relate to virulence on host plants.

Using genomics to investigate variation and evolution of type III effector repertoires

Traditionally, T3Es have been thought to be important host-range determinants. Such effectors have been characterized as avirulence factors that define race structure in plant pathogens. Understanding the race structure in pathogen population and factors associated with race specificity are of importance when assessing efficacy and durability of host resistance. Rosenthal et al. used a comparative genomics approach on genome sequences of different races of *Xanthomonas hortorum* pv. *vitians* and identified two T3Es, XopAQ and XopAF2, as potential mediators of gene-for-gene interactions between race 1 and 3 strains and wild lettuce. The close examination of the neighborhood regions of these effectors indicated proximity of *xopAF2* to prophage sequences, thus, likelihood of gain or loss of this effector in the pathogen population *via* phage-mediated transfer, in response to host selection pressure.

Shah et al. analyzed T3Es of *X. translucens* strains that cause bacterial leaf streak in small grain cereals, which include three pathovars: *cerealis*, *translucens*, and *undulosa*. Four effectors, XopAJ, XopAL1, XopE3, and XopM were found in all analyzed *X. translucens* pv. *translucens* strains, but not in members of the other two pathovars, *cerealis* and *undulosa*. Whether these effectors act as host specificity factors, as hypothesized by the authors, remains to be experimentally proven. Interesting differences were also reported in the TAL effectors, which activate the expression of host genes, including two novel dipeptides (called RVDs for Repeat Variable Diresidues) that contact the bound DNA and provide sequence specificity to these transcription activators (Shah et al.).

The *Pseudomonas syringae* species complex represents a classic system to address convergence of diverse strains belonging to distantly related phylogroups onto a common host. Ruinelli et al. conducted association analyses on genomes of *P. syringae* strains associated with diseases on *Prunus* as well as those from other hosts and from non-agricultural environments. They identified horizontal gene transfer (HGT) as a predominant evolutionary force that has mediated gain/loss of genes such as the T3E, *hopAY*, among strains belonging to diverse phylogroups in a relatively short evolutionary timescale. Such independent acquisition events of genes important for successful pathogen colonization, but also loss of specific genes

or their inactivation, have shaped what we define as host ranges in plant pathogenic bacteria.

Heritable changes in bacteria, including effectors, often occur during conjugation or horizontal gene transfer of mobile genetic elements. Integrative Conjugative Elements (ICEs) are among the mobile replicons that combine the characteristics of plasmids with an ability to carry genes. ICEs are also known to carry cargo regions that can drastically change bacterial phenotypes, including their interactions with hosts. Baltrus et al. demonstrated that a T3E ICE from *P. syringae* pv. *maculicola*—PmaICE-DQ—can transfer to other strains *via* conjugation, resulting in new phenotypes of the recipient. Moreover, another ICE carrying multiple effector genes—PmaICE-AOAB—was found to be adjacent to the PmaICE-DQ. The study demonstrated the transfer of ICEs regions to naive strains. The presence of effectors within ICEs enables quick loss and gain of effectors, thus providing additional flexibility to the pathogen under selection pressure.

Whole genome sequencing can capture shifts in effector content in pathogens over time. Bernal et al. screened hundreds of *X. perforans* strains from tomato fields in the Midwestern USA over a period of 4 years, representing multiple varieties, seed producers, and growers. They found that the *X. perforans* population shifted from race T3 to race T4, mostly caused by a SNP in *avrXv3* that inactivated this effector gene by an early stop codon. While host selection is a likely explanation for the loss of effector function, few tomato varieties with resistance to *X. perforans* T3 made up <7% of the acreage in this geographic region. Analysis of whole genome sequences showed that most strains were nearly identical to each other. The authors discuss the possible mechanisms of persistence of this clonal population from year to year, which is critical knowledge needed to manage bacterial spot of tomato.

Beyond type III effectors

A recent paradigm to explain genetic basis of host adaptation or specificity has suggested involvement of multiple genetic determinants (not limited to T3Es), each contributing small or large effects toward adaptation of pathogen to a specific host or tissue type. Liyanapathiranage et al. conducted large-scale comparative and evolutionary analyses to test the hypothesis that specific gain or loss events of multiple type VI secretion systems are not random, but occurred as independent events during the adaptation of *Xanthomonas* species to specific hosts. They observed that specific cases of loss or gain events happened in certain clades. Whether these secretion systems and their effectors contribute to an adaptation to vascular or non-vascular lifestyles or provide preference toward dicots or monocots remains to be addressed experimentally.

Exhaustive genomics resources stimulate comparative studies aiming at cataloging known and discovering new traits, thus better understanding how they are linked to the ecology of the pathogen. Bacterial toxin-antitoxin (TA) systems, consisting of two or more adjacent genes, are one of these traits of interest, which are implicated in genome maintenance, antibiotics persistence, phage defense, and virulence. Kandel et al. used bioinformatics tools to screen the genomes of hundreds of *P. syringae* strains representing the genetic and lifestyle diversity of the *P. syringae* species complex for TA systems. They show that *P. syringae* strains encode on average 15 TA systems per genome, which belong to 26 different families and are thought to target diverse cellular functions. Further functional characterization of the predicted TA systems could reveal how these widely prevalent gene modules may affect *P. syringae* ecology, virulence, and disease management practices.

Another approach to identify previously unknown virulence factors is using genome-wide association studies (GWAS). Agarwal et al. applied a GWAS approach on *Pantoea ananatis* that causes center rot of onion. This bacterium does not utilize the type II or type III secretion system, but possesses a biosynthetic gene cluster—HiVir/PASVIL—for production of a phosphonate secondary metabolite and also another gene cluster—*alt*—that confers tolerance to thiosulfinates. The pangenome association analyses indicated that although HGT events, including transfer of PASVIL cluster, may have contributed toward diversification and niche adaptation of *P. ananatis*, other factors (e.g. tyrosine kinase, N-acetylmuramoyl-L-alanine amidase, and HTH-type transcriptional regulator) are likely also involved in the pathogenicity of *P. ananatis* on onion.

High-throughput sequencing offers incredible opportunities for identifying genes involved in various microbial processes including fitness, virulence, and antibiotic resistance. Transposon mutagenesis was useful in identifying essential genes for growth on media as well as in planta. Random barcode transposon-site sequencing (RB-TnSeq) was used in comparing three *Dickeya* strains from two species to identify genes involved in fitness as measured by bacterial growth in potato tubers and growth in minimal and complex media (Helmann et al.). Notably, many of the metabolic traits that were required for growth in minimal medium were also important for efficient growth in potato tubers, such as amino acid, carbohydrate, and nucleotide biosynthesis. Moreover, growth of all three *Dickeya* species required the pectate degradation gene, *kduD*, whereas disruption in three putative DNA-binding proteins had different effects at the strain level.

The last two decades have witnessed adoption of comparative and population genomics methods being integrated into addressing several questions related to pathogen biology, epidemiology as well as translational research to guide disease management practices based on changes in pathogen population structure. The articles published in this Research Focus show such translational aspects, informing meaningful hypotheses that demand further functional

genomics studies. Thanks to the plant pathologists who have taken initiative to sequence several hundreds, if not thousands, of genomes of their favorite bacterial genus, we have rich databases with genotypic and phenotypic data. The articles published in this issue also address this important aspect of the need to explore already available data in genome databases, rather than focusing efforts on sequencing more isolates.

Author contributions

ST, EG, RK, NP, and JJ contributed to the manuscript preparation and finalizing. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 May 2021
doi: 10.3389/fmicb.2021.684092

Complete Genome Sequence of *Xylella taiwanensis* and Comparative Analysis of Virulence Gene Content With *Xylella fastidiosa*

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] and 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

OPEN ACCESS

Edited by:
Jeffrey Jones,
University of Florida, United States

Reviewed by:
Prem P. Kandel,
Pennsylvania State University (PSU), United States
Lindsey Burbank,
Agricultural Research Service, United States Department of Agriculture, United States

***Correspondence:**
Chi-Wei Tsai
chiwei@ntu.edu.tw
Chih-Horng Kuo
chk@gate.sinica.edu.tw

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 22 March 2021
Accepted: 27 April 2021
Published: 21 May 2021

Citation:
Weng L-W, Lin Y-C, Su C-C, Huang C-T, Cho S-T, Chen A-P, Chou S-J, Tsai C-W and Kuo C-H (2021) Complete Genome Sequence of Xylella taiwanensis and Comparative Analysis of Virulence Gene Content With Xylella fastidiosa. Front. Microbiol. 12:684092. doi: 10.3389/fmicb.2021.684092

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 was 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.

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

INTRODUCTION

The gammaproteobacterium *Xylella fastidiosa* (*Xf*) is an insect-vectored plant pathogen that resides in plant xylem and is fastidious (Wells et al., 1987). To date, at least 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* (Denancé et al., 2019; Potnis et al., 2019; Vanhove et al., 2019). 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 are 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), that draft assembly was produced with only ∼20-fold coverage of Roche/454 GS-FLX reads and is highly fragmented (i.e., 85 contigs; N50 = 121 kb). Moreover, 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 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, United States). 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, United States) 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, including 27 from *Xf*, were used for comparative analysis (**Table 1**). Our taxon sampling mainly focused on the strains that could represent the known *Xylella* diversity (Vanhove et al., 2019). Two other Xanthomonadaceae genera were also included. 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.

Chromosomal level comparisons of nucleotide sequences were conducted using fastANI v1.1 (Jain et al., 2018). Homologous gene clusters were identified based on protein sequences 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 genomes by homolog 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

TABLE 1 | List of the genome sequences analyzed.

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*	ATCC 35879	United States: Florida	GCA_011801475.1	Complete	2,607,257	2,189	133	2,322
Xylella fastidiosa subsp. *fastidiosa*	Bakersfield-1	United States: Bakersfield, California	GCA_009664125.2	Complete	2,575,627	2,198	70	2,268
Xylella fastidiosa subsp. *fastidiosa*	GB514	United States: 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	United States: California	GCA_000019765.1	Complete	2,573,987	2,235	118	2,353
Xylella fastidiosa subsp. *fastidiosa*	Temecula1	United States: California	GCA_000007245.1	Complete	2,521,148	2,107	64	2,171
Xylella fastidiosa subsp. *morus*	MUL0034	United States: California	GCA_000698825.1	Complete	2,666,577	2,266	151	2,417
Xylella fastidiosa subsp. *morus*	Mul-MD	United States: Maryland	GCA_000567985.1	101 contigs	2,520,555	2,121	155	2,276
Xylella fastidiosa subsp. *multiplex*	AlmaEM3	United States: 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	United States: California	GCA_000166835.1	32 contigs	2,622,328	2,272	123	2,395
Xylella fastidiosa subsp. *multiplex*	Griffin-1	United States: Georgia	GCA_000466025.1	84 contigs	2,387,314	1,911	303	2,214
Xylella fastidiosa subsp. *multiplex*	M12	United States: California	GCA_000019325.1	Complete	2,475,130	2,041	116	2,157
Xylella fastidiosa subsp. *multiplex*	sycamore Sy-VA	United States: 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, São Paulo	GCA_001456195.1	Complete	2,748,594	2,273	179	2,452
Xylella fastidiosa subsp. *pauca*	9a5c	Brazil: Macaubal, São 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, São Paulo	GCA_001456315.1	Complete	2,877,548	2,456	185	2,641
Xylella fastidiosa subsp. *pauca*	J1a12	Brazil: Jales, São 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, São Paulo	GCA_001456275.1	Complete	2,732,490	2,274	178	2,452

(Continued)

TABLE 1 | Continued

Species	Strain	Location	Accession	Assembly	Genome size (bp)	CDS (Intact)	CDS (Pseudo)	CDS (All)
Xylella fastidiosa subsp. *sandyi*	Ann-1	United States: 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	United States: 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

genes, and 2,176 intact coding sequences (CDSs). 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; **Table 1**). Among all 40 representative Xanthomonadaceae genomes, the genome sizes and the numbers of intact CDSs have a correlation coefficient of 0.989 ($p < 2.2e^{-16}$). Compared to those *Xf* representatives with similar genome sizes (i.e., ∼2.73–2.88 Mb), such as those five *Xfp* strains from Brazil or the *Xfs* strain from the United States, the *Xt* PLS229^T genome has fewer intact CDSs (i.e., 2,273–2,456 vs. 2,173) and fewer pseudogenes (i.e., 153–242 vs. 132). It is unclear if these observations were caused by annotation artifacts or have any biological meaning.

Molecular Phylogeny and Genome Divergence

A total of 779 single-copy protein-coding genes were found to be shared by the 40 Xanthomonadaceae genomes compared (**Table 1**). Based on the concatenated alignment of the protein sequences derived from these genes, a robust maximum likelihood phylogeny was inferred (**Figure 1**). 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* (**Figure 1**) and has experienced genome reduction since their divergence (**Table 1**).

When the genetic divergence was measured by overall nucleotide sequence conservation using fastANI (Jain et al., 2018), comparisons within each of the five *Xf* subspecies found that 88.8–99.8% of the chromosomal segments are shared and those segments have 98.5–100% average nucleotide identity (ANI) (**Figure 2**). For between-subspecies comparisons, 86.6–94.8% of the chromosomal 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 chromosomal 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 nucleotide sequence identity drops 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% protein 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*

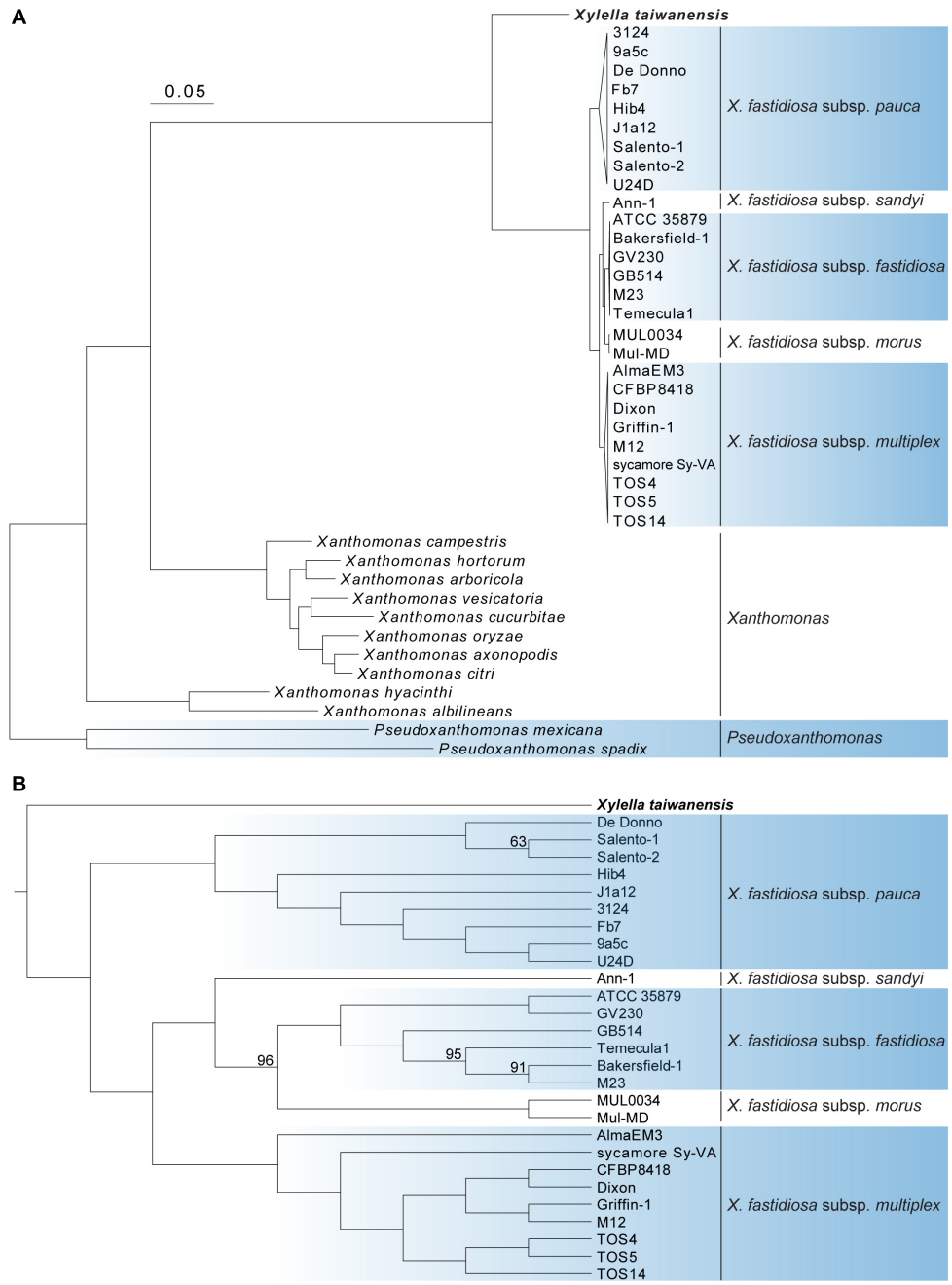


FIGURE 1 | Molecular phylogeny of *Xylella* and related species in the family Xanthomonadaceae. The maximum likelihood phylogeny was based on 779 shared single-copy genes, the concatenated alignment contains 252,319 aligned amino acid sites. The genus *Pseudoxanthomonas* was included as the outgroup.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

comparison (88.6%), and higher than the *Pseudoxanthomonas mexicana-Pseudoxanthomonas spadix* comparison (75.4%).

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

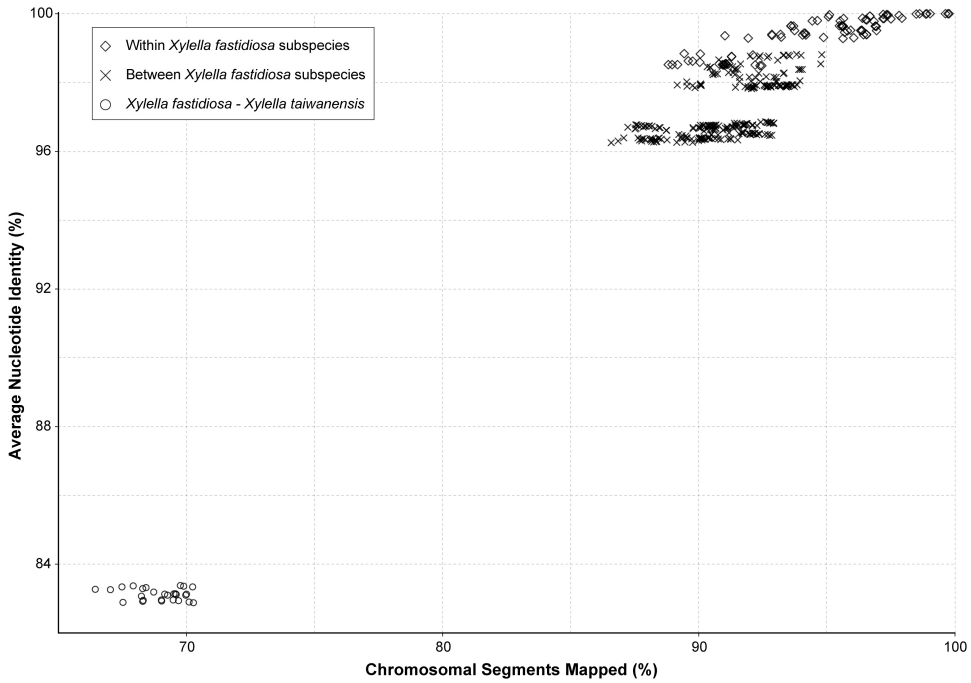


FIGURE 2 | Genome similarity among the representative *Xylella* strains. The pairwise comparisons were classified into three categories: (1) within the same *X. fastidiosa* subspecies, (2) between different *X. fastidiosa* subspecies, and (3) between *X. fastidiosa* and *X. taiwanensis*.

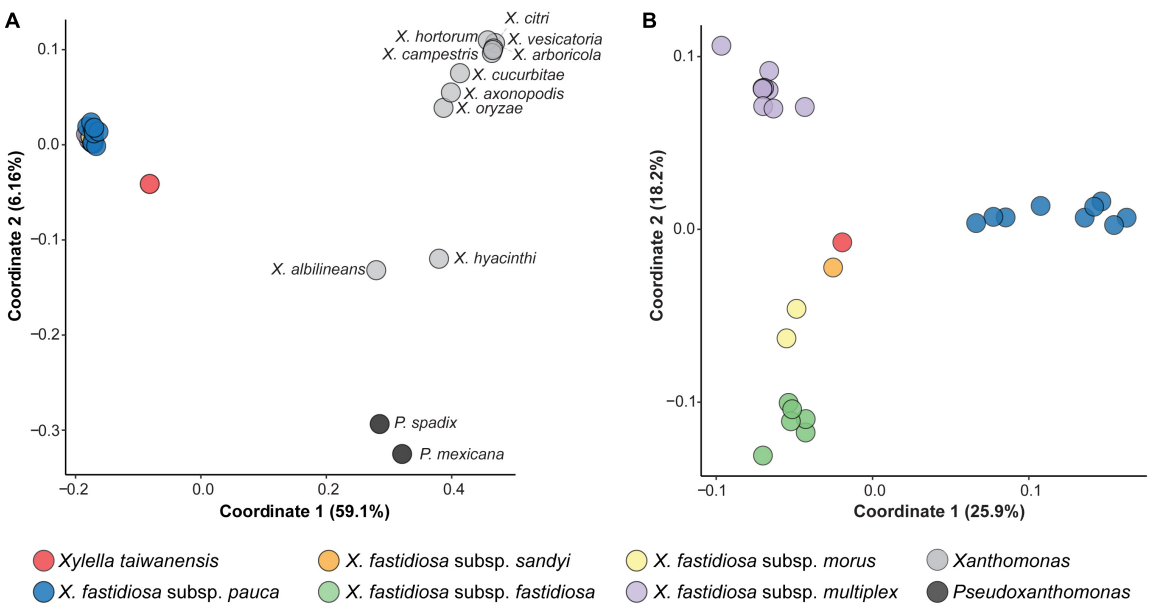


FIGURE 3 | Principal coordinates analysis of gene content dissimilarity. The % variance explained by each coordinate was provided in parentheses. **(A)** Based on the 11,455 homologous gene clusters found among all 40 Xanthomonadaceae genomes analyzed in this work. **(B)** Based on the 5,395 homologous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

the phylogenetic clades inferred based on sequence divergence of the 779 core genes (**Figure 1A**). All 27 *Xf* genomes form a tight cluster (**Figure 3A**) despite their differences in the number of intact CDSs (range: 1,911–2,456; av. ± std. dev.: 2,156 ± 144; **Table 1**). In contrast, although the *Xt* genome has 2,176 intact CDSs, which is close to the average observed among those 27 *Xf* representatives, it does not fall into the *Xf* cluster (**Figure 3A**). This result indicates that the

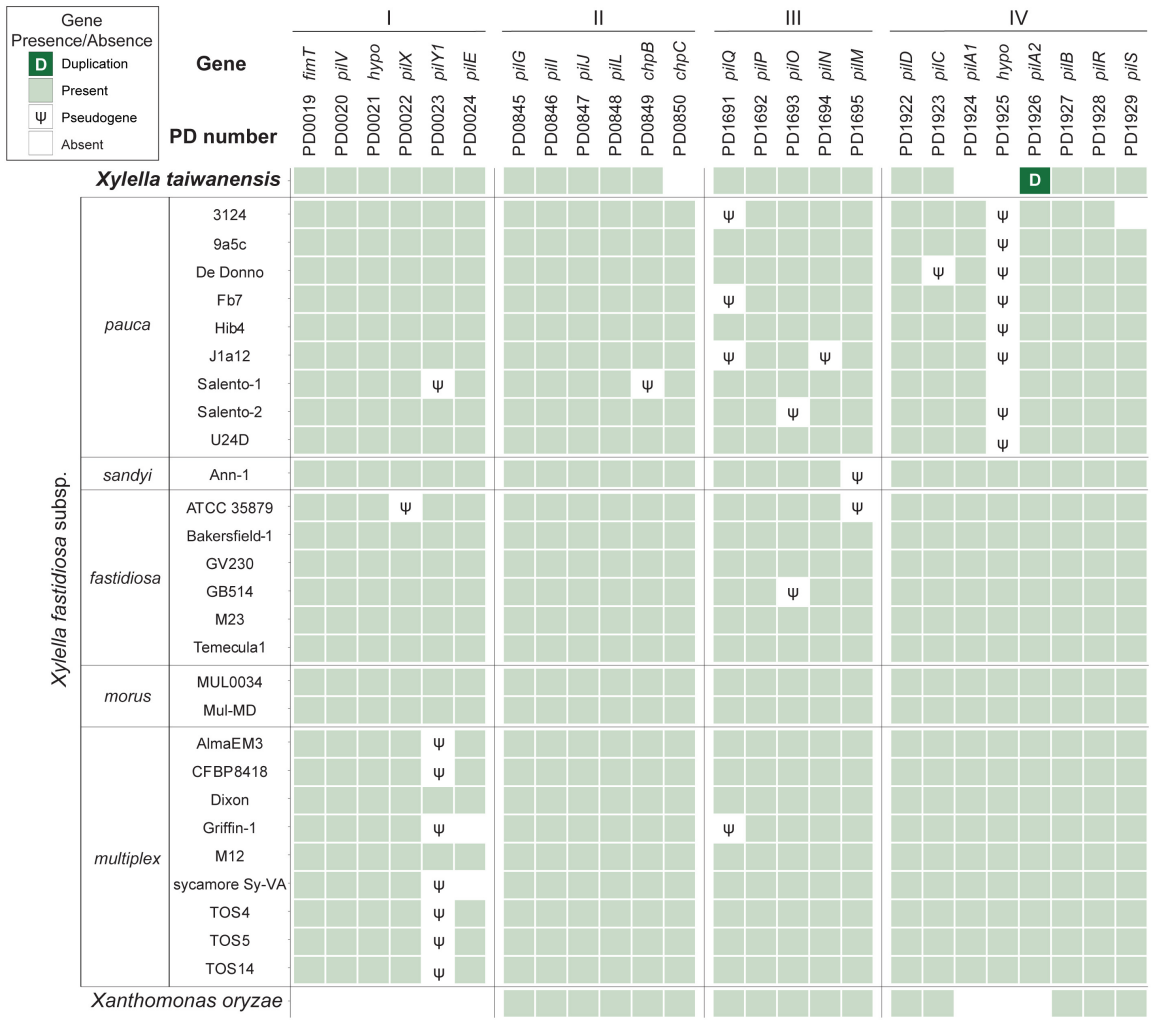


FIGURE 4 | Distribution of the type IV pili genes found among representative *Xylella* genomes. *Xanthomonas oryzae* was included as the outgroup. The genes were identified by the PD numbers based on the annotation of Temecula1 genome; those with adjacent PD numbers are located in neighboring regions on the chromosome. A total of four gene clusters (labeled as I–IV) located in syntenic regions were found to be conserved. Gene names were provided when available; "*hypo*" indicates those annotated as encoding hypothetical proteins. Patterns of gene presence and absence were illustrated in the format of a heatmap. For gene absence, those with identifiable pseudogenes were labeled accordingly. One case of tandem gene duplication was observed for the PD1926 (*pilA2*) homolog in the *Xylella taiwanensis* genome.

gene content divergence between these two *Xylella* species is much higher than the divergence among *Xf* subspecies. For the within-*Xylella* comparison based on 5,395 homologous gene clusters, the grouping patterns are consistent with the taxonomic assignments and *Xt* is most similar to *Xfs* (**Figure 3B**). It is interesting that *Xt* and *Xfs* are similar in having a narrow host range (i.e., *Xt* is restricted to pear and *Xfs* is mostly known for oleander infections), while other *Xf* subspecies can infect a wide range of hosts (Baldi and La Porta, 2017; European Food Safety Authority, 2018; Rapicavoli et al., 2018). However, it is also important to note that the host range information may be limited by sampling and experimental efforts. As more research results become available, this information may be updated. For example, *Xfs*-related strains have been reported to infect coffee (Jacques et al., 2016) and whether *Xt* can infect a wider range of plants remains to be investigated.

Virulence Genes and Pathogenicity Factors

Based on the current knowledge of putative virulence genes and pathogenicity factors identified in *Xf*, type IV pili (T4P) are important for twitching motility and movements within infected plants (Meng et al., 2005; Li et al., 2007; Burdman et al., 2011; Cursino et al., 2011; Hao et al., 2017). The *Xylella* T4P genes are organized into four major gene clusters with 25 homologs and highly conserved among the 28 representative genomes examined (**Figure 4**). Among these four T4P gene clusters, cluster II that corresponds to the *pil-chp* operon (Cursino et al., 2011;

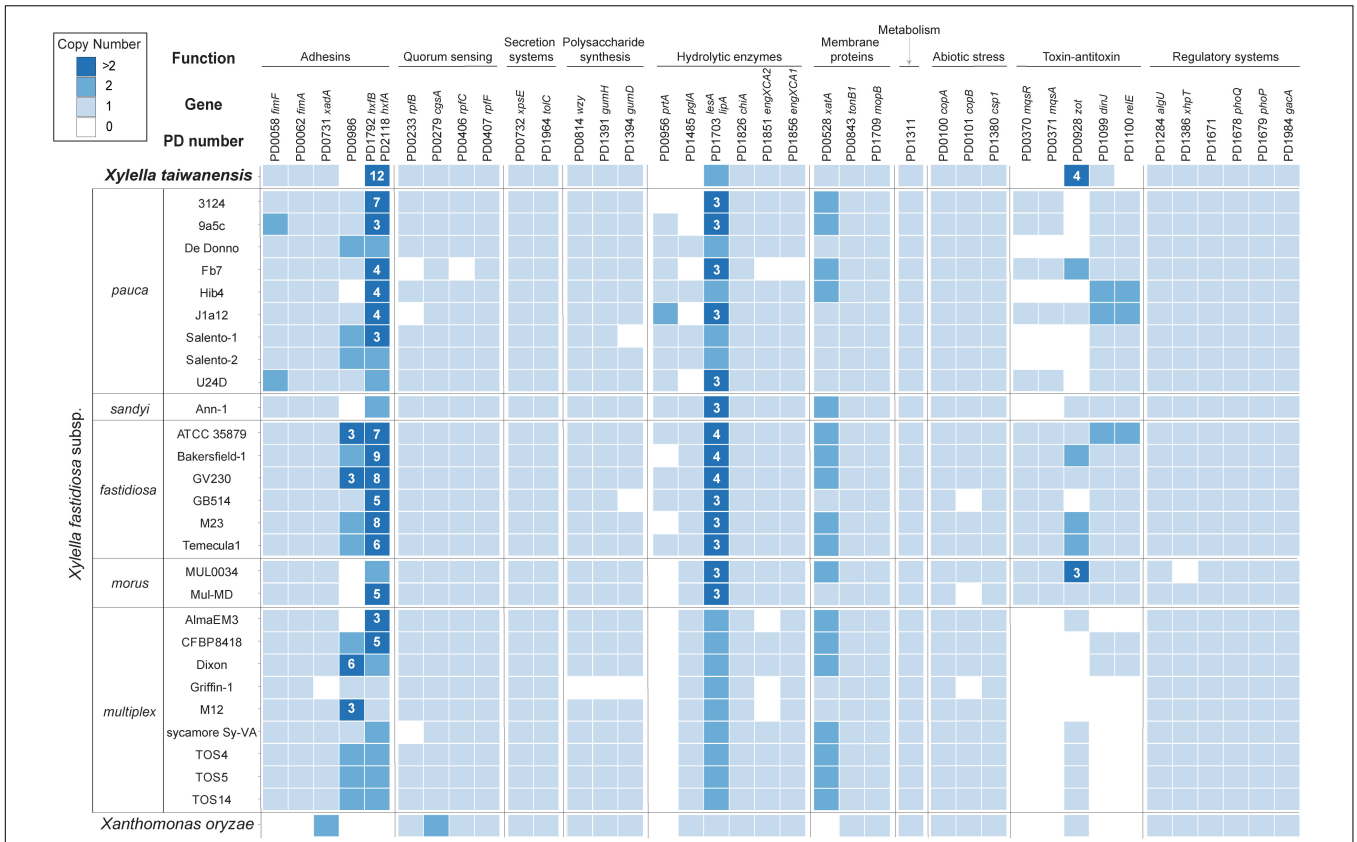


FIGURE 5 | Distribution of putative virulence genes and pathogenicity factors among representative *Xylella* genomes. *Xanthomonas oryzae* was included as the outgroup. The homologous gene clusters were identified by the PD numbers based on the annotation of Temecula1 genome; gene names were provided when available. Gene copy numbers were illustrated in the format of a heatmap; values higher than two were 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.

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

In addition to the T4P genes, many other *Xylella* pathogenicity factors have been identified and these additional virulence genes may be classified into 10 major functional categories (Lee et al., 2014; Merfa et al., 2016; Chen et al., 2017; Rapicavoli et al., 2018; Ge et al., 2021). Among these categories, secretion systems and metabolism are the most conserved ones with no variation in gene copy number across all *Xylella* representatives (**Figure 5**). Additionally, those genes involved in regulatory systems are also highly conserved. 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 38 homologous gene clusters and six are absent in the *Xt* genome (**Figure 5**). These include the genes that encode a putative adhesin (PD0986, hemagglutinin-like protein), two hydrolytic enzymes (PD0956, serine protease; PD1485, polygalacturonase), one pair of toxin-antitoxin (PD0370, motility quorum sensing regulator MqsR ribonuclease; PD0371, MqsA antitoxin), and another separate toxin (PD1100, endoribonuclease). Notably, the *mqsR-mqsA* toxin-antitoxin system genes (Lee et al., 2014; Merfa et al., 2016) are differentially distributed among those *Xf* subspecies. Homologs of these two genes are entirely conserved in all *Xff* and *Xf* subspecies *morus* strains, present in five out of the nine *Xfp* strains, and completely absent in *Xfs* and *Xf* subspecies *multiplex*.

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. Similarly, PD0370 is another antivirulence factor that reduces the virulence of *Xfp* against citrus when overexpressed (Merfa et al., 2016). 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).

CONCLUSION

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 species 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 (**Figures 4** and **5**). 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 (Burdman et al., 2011; Rapicavoli et al., 2018). However, it is important to note that the current knowledge of *Xylella* virulence genes is mostly derived from those strains that are relatively easy to culture and transform, such that only limited diversity has been investigated in molecular genetics studies. Moreover, infection experiments for the investigation of gene functions were limited to a small number of plant species. For further improvements, experimental studies that examine more diverse *Xylella* lineages and plant hosts, as well as the combined effects of multiple virulence genes will be critical.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, CP053627, //www.ncbi.nlm.nih.gov/, SRR11805344, https://www.ncbi.nlm.nih.gov/, SRR11805345.

AUTHOR CONTRIBUTIONS

C-WT and C-HK: conceptualization, funding acquisition, project administration, and supervision. L-WW, Y-CL, and C-TH: investigation, validation, and visualization. C-CS, S-TC, A-PC, S-JC, and C-HK: methodology. C-CS, C-WT, and C-HK: resources. L-WW and C-HK: writing—original draft. L-WW, Y-CL, C-CS, C-TH, S-TC, A-PC, S-JC, C-WT, and C-HK: writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

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

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).

REFERENCES

Anderson, I., Teshima, H., Nolan, M., Lapidus, A., Tice, H., Rio, T. G. D., et al. (2013). Genome sequence of *Frateuria aurantia* type strain (Kondô 67T), a xanthomonade isolated from *Lilium auratium* Lindl. *Stand. Genomic. Sci.* 9, 83–92. doi: 10.4056/sigs.4338002

Baldi, P., and La Porta, N. (2017). *Xylella fastidiosa*: host range and advance in molecular identification techniques. *Front. Plant Sci.* 8:944.

Burbank, L. P., and Stenger, D. C. (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

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

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

Chen, H., Kandel, P. P., Cruz, L. F., Cobine, P. A., and De La Fuente, L. (2017). The major outer membrane protein MopB is required for twitching

movement and affects biofilm formation and virulence in two *Xylella fastidiosa strains*. *Mol. Plant Microbe Interact.* 30, 896–905. doi: 10.1094/mpmi-07-17-0161-r

Cho, S.-T., Zwolińska, A., Huang, W., Wouters, R. H. M., Mugford, S. T., Hogenhout, S. A., 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:e000760-20.

Cursino, L., Galvani, C. D., Athinuwat, D., Zaini, P. A., Li, Y., De La Fuente, L., et al. (2011). Identification of an operon, Pil-Chp, that controls twitching motility and virulence in *Xylella fastidiosa*. *Mol. Plant Microbe Interact.* 24, 1198–1206. doi: 10.1094/mpmi-10-10-0252

Denancé, N., Briand, M., Gaborieau, R., Gaillard, S., and Jacques, M.-A. (2019). Identification of genetic relationships and subspecies signatures in *Xylella fastidiosa*. *BMC Genomics* 20:239.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

European Food, and Safety Authority. (2018). Update of the *Xylella* spp. host plant database. *EFSA J.* 16:e05408.

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

Ge, Q., Cobine, P., and De La Fuente, L. (2021). The influence of copper homeostasis genes copA and copB on *Xylella fastidiosa* virulence is affected by sap copper concentration. *Phytopathology* doi: 10.1094/PHYTO-12-20-0531-R Online ahead of print.

Gouran, H., Gillespie, H., Nascimento, R., Chakraborty, S., Zaini, P. A., Jacobson, A., et al. (2016). The secreted protease PrtA controls cell growth, biofilm formation and pathogenicity in *Xylella fastidiosa*. *Sci. Rep.* 6:31098.

Guilhabert, M. R., and Kirkpatrick, B. C. (2005). Identification of *Xylella fastidiosa* antivirulence genes: hemagglutinin adhesins contribute to X. fastidiosa biofilm maturation and colonization and attenuate virulence. *Mol. Plant Microbe Interact.* 18, 856–868. doi: 10.1094/mpmi-18-0856

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Hao, L., Athinuwat, D., Johnson, K., Cursino, L., Burr, T. J., and Mowery, P. (2017). *Xylella fastidiosa* pil-chp operon is involved in regulating key structural genes of both type I and IV pili. *Vitis* 56, 55–62.

Jacques, M.-A., Denancé, N., Legendre, B., Morel, E., Briand, M., Mississipi, S., et al. (2016). New coffee plant-infecting *Xylella fastidiosa* variants derived via homologous recombination. *Appl. Environ. Microbiol.* 82, 1556–1568. doi: 10.1128/aem.03299-15

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.

Kandel, P. P., Chen, H., and De La Fuente, L. (2018). A short protocol for gene knockout and complementation in *Xylella fastidiosa* shows that one of the type IV pilin paralogs (PD1926) is needed for twitching while another (PD1924) affects pilus number and location. *Appl. Environ. Microbiol.* 84:e01167-18.

Lee, M. W., Tan, C. C., Rogers, E. E., and Stenger, D. C. (2014). Toxin-antitoxin systems *mqsR/ygiT* and *dinJ/relE* of *Xylella fastidiosa*. *Physiol. Mol. Plant Pathol.* 87, 59–68. doi: 10.1016/j.pmpp.2014.07.001

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Li, Y., Hao, G., Galvani, C. D., Meng, Y., De La Fuente, L., Hoch, H. C., et al. (2007). Type I and type IV pili of *Xylella fastidiosa* affect twitching motility, biofilm formation and cell–cell aggregation. *Microbiology* 153, 719–726. doi: 10.1099/mic.0.2006/002311-0

Lo, W.-S., Chen, L.-L., Chung, W.-C., Gasparich, G. E., and Kuo, C.-H. (2013). Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genom.* 14:22. doi: 10.1186/1471-2164-14-22

Lo, W.-S., Gasparich, G. E., and Kuo, C.-H. (2018). Convergent evolution among ruminant-pathogenic *Mycoplasma* involved extensive gene content changes. *Genome Biol. Evol.* 10, 2130–2139. doi: 10.1093/gbe/evy172

Meng, Y., Li, Y., Galvani, C. D., Hao, G., Turner, J. N., Burr, T. J., et al. (2005). Upstream migration of *Xylella fastidiosa* via pilus-driven twitching motility. *J. Bacteriol.* 187, 5560–5567. doi: 10.1128/jb.187.16.5560-5567.2005

Merfa, M. V., Niza, B., Takita, M. A., and De Souza, A. A. (2016). The MqsRA toxin-antitoxin system from *Xylella fastidiosa* plays a key role in bacterial fitness, pathogenicity, and persister cell formation. *Front. Microbiol.* 7:904.

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633

Parkinson, N., Cowie, C., Heeney, J., and Stead, D. (2009). Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int. J. Syst. Evol. Microbiol.* 59, 264–274. doi: 10.1099/ijs.0.65825-0

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., et al. (2009). The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics* 10:616. doi: 10.1186/1471-2164-10-616

Potnis, N., Kandel, P. P., Merfa, M. V., Retchless, A. C., Parker, J. K., Stenger, D. C., et al. (2019). Patterns of inter- and intrasubspecific homologous recombination inform eco-evolutionary dynamics of *Xylella fastidiosa*. *ISME J.* 13, 2319–2333. doi: 10.1038/s41396-019-0423-y

Rapicavoli, J., Ingel, B., Blanco-Ulate, B., Cantu, D., and Roper, C. (2018). *Xylella fastidiosa*: an examination of a re-emerging plant pathogen. *Mol. Plant Pathol.* 19, 786–800. doi: 10.1111/mpp.12585

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotech.* 29, 24–26. doi: 10.1038/nbt.1754

Rodriguez-R, L. M., Grajales, A., Arrieta-Ortiz, M. L., Salazar, C., Restrepo, S., and Bernal, A. (2012). Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiol.* 12:43. doi: 10.1186/1471-2180-12-43

Roper, M. C., Greve, L. C., Warren, J. G., Labavitch, J. M., and Kirkpatrick, B. C. (2007). *Xylella fastidiosa* requires polygalacturonase for colonization and pathogenicity in *Vitis vinifera* grapevines. *Mol. Plant Microbe Interact.* 20, 411–419. doi: 10.1094/mpmi-20-4-0411

Su, C.-C., Chang, C.-J., Yang, W.-J., Hsu, S.-T., Tzeng, K.-C., Jan, F.-J., et al. (2012). Specific characters of 16S rRNA gene and 16S–23S rRNA internal transcribed spacer sequences of *Xylella fastidiosa* pear leaf scorch strains. *Eur. J. Plant Pathol.* 132, 203–216. doi: 10.1007/s10658-011-9863-6

Su, C.-C., Deng, W.-L., Jan, F.-J., Chang, C.-J., Huang, H., and Chen, J. (2014). Draft genome sequence of *Xylella fastidiosa* pear leaf scorch strain in Taiwan. *Genome Announc.* 2:e00166-14.

Su, C.-C., Deng, W.-L., Jan, F.-J., Chang, C.-J., Huang, H., Shih, H.-T., et al. (2016). *Xylella taiwanensis* sp. nov., causing pear leaf scorch disease. *Int. J. Syst. Evol. Microbiol.* 66, 4766–4771. doi: 10.1099/ijsem.0.001426

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

Vanhove, M., Retchless, A. C., Sicard, A., Rieux, A., Coletta-Filho, H. D., De La Fuente, L., et al. (2019). Genomic diversity and recombination among *Xylella fastidiosa* subspecies. *Appl. Environ. Microbiol.* 85:e02972-18.

Wells, J. M., Raju, B. C., Hung, H.-Y., Weisburg, W. G., Mandelco-Paul, L., and Brenner, D. J. (1987). *Xylella fastidiosa* gen. nov., sp. nov: gram-negative, xylem-limited, fastidious plant bacteria related to Xanthomonas spp. *Int. J. Syst. Bacteriol.* 37, 136–143. doi: 10.1099/00207713-37-2-136

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:e1005595. doi: 10.1371/journal.pcbi.1005595

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.

Zhang, S., Chakrabarty, P. K., Fleites, L. A., Rayside, P. A., Hopkins, D. L., and Gabriel, D. W. (2015). Three new Pierce's disease pathogenicity effectors identified using *Xylella fastidiosa* biocontrol strain EB92-1. *PLoS One* 10:e0133796. doi: 10.1371/journal.pone.0133796

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fmicb.2021.674952

Genomics-Enabled Novel Insight Into the Pathovar-Specific Population Structure of the Bacterial Leaf Streak Pathogen *Xanthomonas translucens* in Small Grain Cereals

Syed Mashab Ali Shah[1†], Moein Khojasteh[1,2,3†], Qi Wang[1], S. Mohsen Taghavi[2], Zhengyin Xu[1], Pejman Khodaygan[4], Lifang Zou[1], Sedighe Mohammadikhah[4], Gongyou Chen[1] and Ebrahim Osdaghi[3*]*

[1] School of Agriculture and Biology/State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University, Shanghai, China, [2] Department of Plant Protection, School of Agriculture, Shiraz University, Shiraz, Iran, [3] Department of Plant Protection, University of Tehran, Karaj, Iran, [4] Department of Plant Protection, Faculty of Agriculture, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran

The Gram-negative bacterium *Xanthomonas translucens* infects a wide range of gramineous plants with a notable impact on small grain cereals. However, genomics-informed intra-species population structure and virulence repertories of the pathogen have rarely been investigated. In this study, the complete genome sequences of seven *X. translucens* strains representing an entire set of genetic diversity of two pathovars *X. translucens* pv. *undulosa* and *X. translucens* pv. *translucens* is provided and compared with those of seven publicly available complete genomes of the pathogen. Organization of the 25 type III secretion system genes in all the 14 *X. translucens* strains was exactly the same, while TAL effector genes localized singly or in clusters across four loci in *X. translucens* pv. *translucens* and five to six loci in *X. translucens* pv. *undulosa*. Beside two previously unreported endogenous plasmids in *X. translucens* pv. *undulosa*, and variations in repeat variable diresidue (RVD) of the 14 strains, *tal1a* of *X. translucens* pv. *translucens* strain XtKm8 encode the new RVDs HE and YI which have not previously been reported in xanthomonads. Further, a number of truncated *tal* genes were predicted among the 14 genomes lacking conserved *Bam*HI site at N-terminus and *Sph*I site at C-terminus. Our data have doubled the number of complete genomes of *X. translucens* clarifying the population structure and genomics of the pathogen to pave the way in the small grain cereals industry for disease resistance breeding in the 21st century's agriculture.

Keywords: bacterial leaf streak, *Xanthomonas translucens,* nanopore sequencing, genome structure, small grain cereals

OPEN ACCESS

Edited by:
Sujan Timilsina,
University of Florida, United States

Reviewed by:
Veronica Roman-reyna,
The Ohio State University, United States
Sanzhen Liu,
Kansas State University, United States

****Correspondence:***
Gongyou Chen
gyouchen@sjtu.edu.cn
Ebrahim Osdaghi
eosdaghi@ut.ac.ir

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: *02 March 2021*
Accepted: *27 April 2021*
Published: *28 May 2021*

Citation:
Shah SMA, Khojasteh M, Wang Q, Taghavi SM, Xu Z, Khodaygan P, Zou L, Mohammadikhah S, Chen G and Osdaghi E (2021) Genomics-Enabled Novel Insight Into the Pathovar-Specific Population Structure of the Bacterial Leaf Streak Pathogen Xanthomonas translucens in Small Grain Cereals. Front. Microbiol. 12:674952. doi: 10.3389/fmicb.2021.674952

INTRODUCTION

The Gram-negative phytopathogenic members of the genus *Xanthomonas* cause devastating diseases on hundreds of agricultural crops i.e., the members of *Poaceae* family, e.g., barley, maize, oat, rice, rye, sugarcane, triticale, and wheat (Jacques et al., 2016; Sapkota et al., 2020). Diseases caused by xanthomonads on gramineous crops comprise leaf streak, black chaff, blight, or wilt symptoms (Denancé et al., 2016). Bacterial leaf streak of small grain cereals caused by different pathovars of *X. translucens* is one of the economically important diseases of wheat and barley worldwide. The disease occurs in many countries across the globe with a particular importance in regions characterized by high precipitations. *Xanthomonas translucens* pv. *undulosa*, *X. translucens* pv. *translucens* and *X. translucens* pv. *cerealis* are the only member of nine pathovars within the species that are reported to have wide geographic distribution causing economic yield losses on wheat and barley (Sapkota et al., 2020). Hence, these pathovars are included in the A2 (high risk) list of quarantine pathogens by the European and Mediterranean Plant Protection Organization, and are under strict quarantine control and zero tolerance in several countries (EPPO, 1998; Sapkota et al., 2020). All the three pathovars possess broad host range where *X. translucens* pv. *cerealis* infects rye, oat, bromus and wheat; *X. translucens* pv. *translucens* infects barley, oat, rye and harding grass; and *X. translucens* pv. *undulosa* infects triticale, oat, rye, bromus, barley and wheat (Khojasteh et al., 2019). The remaining seven pathovars of the species are associated with different grasses having narrow host range with lower economic impact (Sapkota et al., 2020).

The *X. translucens* pathogens are seed-borne and can impact grain quantity and quality by reducing number of kernels per spike, and grain weight (Shane et al., 1987). Yield losses due to the bacterial leaf streak mainly depend on resistance/susceptibility of cultivars grown, environmental conditions and availability of primary inoculum (CABI, 2020; Sapkota et al., 2020). Despite the pivotal economic importance of the bacterial leaf streak in cereals industry, different pathovars of the pathogen have rarely been subjected to phylogenomics and comparative genomics in order to clarify molecular characteristics and virulence repertories of each of the above mentioned pathovars. Similar to the bacterial spot of solanaceous vegetables caused by four genetically distinct xanthomonads (Potnis et al., 2015; Osdaghi et al., 2016, 2017) determination of the exact yield lose corresponding to each of the small grains-pathogenic xanthomonads on wheat and barley is practically neither possible nor reliable. Unless otherwise exactly determined, the economic lose attributed to the bacterial leaf streak disease is usually considered as a whole. Recently, Khojasteh et al. (2019) have investigated the phylogenetic relationships and phylogeography of all the available *X. translucens* strains and proposed the Middle Eastern countries, i.e., Fertile Crescent as the center of diversity of the pathogen.

Plant pathogenic xanthomonads translocate a cocktail of different effector proteins into host plant cells referred to as type-III effectors (T3Es) using the type three secretion system (T3SS). The T3Es are further categorized into TALEs (transcription activator-like effectors) and non-TALEs which also known as Xops (*Xanthomonas* outer proteins). It has been shown that the bacterial leaf streak pathogen isolated from barley and wheat crops possess high genetic diversity in terms of TALEs (Khojasteh et al., 2020). During the past decade, high throughput complete genome sequencing technologies have provided substantial progresses in the understanding of molecular mechanisms underlying plant colonization, pathogenicity and survival of the bacterial leaf streak pathogens (Sapkota et al., 2020). Genome-informed investigation of *X. translucens* species complex provides valuable information on the virulence repertories, pathogenicity mechanisms, and host adaptation of the bacterial leaf streak pathogens (Peng et al., 2016). Until 2021, complete genome resources of seven *X. translucens* strains, i.e., *X. translucens* pv. *cerealis* strains CFBP 2541 and NXtc01 (Pesce et al., 2015; Shah et al., 2019), *X. translucens* pv. *translucens* strain DSM 18974^{T} (Jaenicke et al., 2016), and *X. translucens* pv. *undulosa* strains Xtu 4699, ICMP 11055, P3 and LW16 (Peng et al., 2016, 2019; Falahi Charkhabi et al., 2017) were publicly available. Using these genome resources, it has been shown that TALEs have pivotal contribution to the virulence and adaptation of the bacterial leaf streak pathogens facilitating host plant colonization, fitness and proliferation within the host plant tissues. So far, three TALEs, i.e., ICMP 11055 Tal2 and Tal4b, and NXtc01_Tal1 have been reported to have effective contribution to the virulence of the bacterial leaf streak pathogens (Falahi Charkhabi et al., 2017; Shah et al., 2019). Recently, a TALE named Xtu 4699_Tal8 has functionally been characterized, inducing expression of wheat gene TaNCED located on short arm of chromosome 5B to promote disease susceptibility (Peng et al., 2019). Despite the abundance of data on geographic distribution and genetic diversity of the bacterial leaf streak agents (Curland et al., 2018; Khojasteh et al., 2020), complete genome sequence-based population structure and genomic repertories of the pathogens mostly remain uninvestigated.

In this study, in order to provide a comprehensive insight into the population structure, genomic content and pathogenicity determinants of *X. translucens*, we have selected seven *X. translucens*, i.e., four X. *translucens* pv. *undulosa* and three *X. translucens* pv. *translucens* strains among a collection of 57 strains isolated in Iran during the past couple of decades (Khojasteh et al., 2019). The representative strains were selected based on their host range and pathovar status, multilocus sequence analysis and typing (MLSA/MLST) scheme as well as the Southern blot-based TALE diversity as described previously (Khojasteh et al., 2020). The strains were sequenced using Oxford Nanopore PromethION long-read direct DNA sequencing platform. To evaluate the genomic variations among and between *X. translucens* pathovars we applied a comparative genomic workflow, taking into account the complete genome sequences of a set of seven reference strains. Complete genome sequencing revealed that the *X. translucens* pv. *undulosa* strains XtKm15 and XtLr8 each harbors two plasmids which have not previously been reported in any strain of this pathovar. We analyzed all the genomes for novel genes that might be important for pathogenicity, particularly TAL and non-TAL effectors with homologs in other *Xanthomonas* strains. These

data provided an important insight into the *X. translucens*-gramineous crops pathosystem and pave the way for future development of resistant cultivars.

MATERIALS AND METHODS

Bacterial Strains, Growth Conditions and Genomic DNA Extraction

The bacterial strains used in this study are listed in **Table 1**. A set of seven *X. translucens* strains, i.e., four *X. translucens pv. undulosa* strains: XtKm12, XtKm15, XtLr8, and XtFa1 and three *X. translucens* pv. *translucens* strains: XtKm8, XtKm9, and XtKm34 were selected from a collection of 57 strains isolated from wheat and barley in Iran during 2008 to 2017 (Khojasteh et al., 2019). The criteria used for selection of the strains were their host range and pathovar status, MLSA-based genetic diversity and TALE repertories of the strains (Khojasteh et al., 2020). The bacterial strains were streaked onto nutrient agar (NA) medium or nutrient broth medium (NB: NA without agar) when required and incubated at 28°C. All the strains were stored at −80°C in nutrient broth (NB) medium amended with 50% sterile glycerol. The genomic DNA of the bacterial strains was extracted from a 24 h culture in NB medium using the Hipure bacterial DNA extraction kit (Magen, Guangzhou, Guangdong, China) as recommended by the manufacturer. The quality and quantity of the DNAs were spectrophotometrically evaluated and adjusted to 1500 ng/μL using the NanoDrop ND-100 (NanoDrop Technologies, Waltham, MA, United States) and then confirmed by 1.0% agarose gel electrophoresis.

Complete Genome Sequencing, Assembly, and Annotation

Genomic DNA of all *X. translucens* strains were sequenced using long-read Nanopore sequencing technology plus NovaSeq 6000 sequencing (OE Biotech Co. Ltd., Shanghai, China). Sequencing libraries were prepared and added to PromethION flow cells and transferred into the Oxford Nanopore sequencer for real-time single molecule sequencing. The NovaSeq 6000 short reads were also produced for assembly and polished by using Racon v1.4.3 (Vaser et al., 2017) and Pilon v1.22. For demultiplexing ONT-guppy v4.0.11 was used for promethion sequencing. After demultiplexing, the obtained reads were assembled using Flye v2.5 (Kolmogorov et al., 2019) and Canu v1.7 (Koren et al., 2017) command-line services via default parameters. Nanopore data was first mapped to the assembled genomes with Minimap2-2.9 (Li, 2016) and then corrected three times with Racon v1.4.3. The Nanopore raw data of XtKm9, XtKm34, XtFa1, XtKm12, XtKm15 and XtLr8 were assembled using Flye with default parameters, while Canu (with default parameters) was used for XtKm8. Then all the assembled genomes were corrected three times via Minimap2+racon. Finally, all the genomes were polished three times with Pilon (v1.22) (Walker et al., 2014) using the high-quality short reads (generated by Trimmomatic-0.36; Bolger et al., 2014). Subsequently, genome annotation was performed using the GeneMarkS^{+} v4.6 suite implemented in the NCBI

TABLE 1 | Genomic characteristics of *Xanthomonas translucens* strains investigated in this study. The first seven strains were sequenced in this study while the next seven reference strains were retrieved from the NCBI GenBank.

	Strains sequenced in this study							Reference sequences obtained from the NCBI GenBank						
	XtFa1	**XtLr8**	**XtKm12**	**XtKm15**	**XtKm8**	**XtKm9**	**XtKm34**	**NXtc01**	**ICMP 11055**	**CFBP 2541**	**DSM 18974^{T}**	**LW16**	**P3**	**Xtu 4699**
Host	Wheat	Wheat	Wheat	Ryegrass	Barley	Barley	Barley	Wheat	Wheat	Bromegrass	Barley	Wheat	Wheat	Wheat
Region	Fars	Lorestan	Kerman	Kerman	Kerman	Kerman	Kerman	Xinjiang	Kerman	–	Minnesota	North Dakota	North Dakota	Kansas
Country	Iran	Iran	Iran	Iran	Iran	Iran	Iran	China	Iran	United States	United States	United States	United States	United States
Year	2016	2016	2015	2015	2014	2015	2015	2016	1983	1941	1933	2009	2009	1999
Genome length (bp)	4,605,208	4,563,212	4,581,137	4,560,646	4,792,950	4,689,955	4,680,513	4,622,298	4,761,583	4,518,140	4,715,357	4,746,074	4,618,583	4,561,137
G + C content (%)	68.06	68.04	68.01	68.04	67.79	67.87	67.79	67.23	67.8	67.34	67.7	67.8	68.1	68.1
Protein-coding genes	3,699	3,731	3,654	3,728	3,873	3,763	3,768	3,733	3,835	3609	3804	3752	3723	3636
RNA genes	63	63	63	63	63	64	63	64	63	60	64	64	63	63
Pseudo-genes	218	211	205	211	208	200	185	256	224	294	190	234	215	205
CRISPR arrays	1	1	1	1	1	1	1	1	1	3	1	1	1	ND
TAL effector genes	7	8	7	8	8	5	7	2	7	2	8	8	8	8
Non-TAL T3E genes	30	29	31	29	34	36	33	35	27	31	36	29	30	27
Genome coverage (×)	478	490	313	475	445	484	385	80	118	1926	NA	90	90	60
Acc. no.	CP063996	CP063993	CP064000	CP063997	CP064004	CP064003	CP064001	CP038228	CP009750	NZ_CM003052	LT604072	CP043540	CP043500	CP008714

Prokaryotic Genome Annotation Pipeline with default settings (Borodovsky and Lomsadze, 2014). The assembled genomes were masked via RepeatMasker v4.0.7 (Tarailo-Graovac and Chen, 2009). For functional classification, the putative genes were annotated against five databases, i.e., KEGG, NR, COG, Swiss-Prot and GO with default parameters as described previously (Chen et al., 2020). Further, tRNA and rRNA genes were predicted using tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997) and rRNAmmer v1.2 (Lagesen et al., 2007), respectively, while sRNAs were predicted using BLAST against the Rfam database (Griffiths-Jones et al., 2003). CRISPR sequences were predicted using PILER-CR v1.06 (Buchfink et al., 2015) and CRT1.2-CLI (Eddy, 2009). Prophages were also predicted using PhiSpy v2.3 (Akhter et al., 2012). The circular genome maps were generated using Circos to show annotation information.

Phylogenomics, Comparative Genomics, and Pan-Genome Analysis

In order to determine the precise phylogenetic position of the strains sequenced in this study, all the publicly available genome sequences assigned as *X. translucens* (up to November 2020) were retrieved from the NCBI GenBank database and included in the phylogenetic analyses. Average nucleotide identity (ANI) was calculated among all *X. translucens* genome sequences using both one-vs.-one and all-vs.-all strategies via different algorithms, i.e., JSpeciesWS, ANI calculator, and OrthoANIu as detailed previously (Osdaghi et al., 2020). ANI calculator estimates all-vs.-all distances in a collection of genomes and builds a similarity clustering (Rodriguez-R and Konstantinidis, 2016). The OrthoANIu algorithm is an improved iteration of the original OrthoANI algorithm which uses USEARCH instead of BLAST (Yoon et al., 2017). JSpecies Web Server (JSpeciesWS) online service measures the ANI based on BLAST + (ANIb) and MUMmer (ANIm), as well as correlation indexes of tetra-nucleotide signatures (Richter et al., 2016). ANI-based Neighbor-Joining distance clustering plot was constructed using the ANI calculator online service for all the *X. translucens* strains. Further, to determine the gene pool of seven Iranian *X. translucens* strains and to compare these strains to the publicly available complete genomes of reference strains, the complete genome data of seven *X. translucens* strains, i.e., *X. translucens* pv. *cerealis* strains CFBP 2541 and NXtc01 (Pesce et al., 2015; Shah et al., 2019), *X. translucens* pv. *translucens* strain DSM 18974^T (Jaenicke et al., 2016), and *X. translucens* pv. *undulosa* strains Xtu 4699, ICMP 11055, P3 and LW16 (Peng et al., 2016, 2019; Falahi Charkhabi et al., 2017) were retrieved from the NCBI GenBank database and included in all the subsequent analyses. We carried out a pan-/core-genome analysis and functional assignment to the COGs categories by Roary 3.8.0 using the procedure described by Page et al. (2015). In brief, FASTA files of all the 14 *X. translucens* strains were transformed to GFF3 format using Prokka (Seemann, 2014) to create a nucleotide alignment using Roary 3.8.0 (with a 95% BLASTp percentage identity cut-off) to cluster the genes into core and accessory genomes. Phandango (Hadfield et al., 2018) and R packages including seqinR (Charif and Lobry, 2007) as well as tidyverse (Wickham et al., 2019) were applied to visualize the resulted output graphs. IslandViewer 4 was used for the identification and visualization of genomic islands (Bertelli et al., 2017) while pairwise genome collinearity alignment and visualization of the seven strains sequenced in this study was performed using BRIG 0.95 (Alikhan et al., 2011). The complete genome sequences of the strains DSM 18974^T and Xtu 4699 were used as reference genomes for *X. translucens* pv. *translucens* and *X. translucens* pv. *undulosa* strains, respectively. Further, Mauve software was used to illustrate locally collinear blocks among the genomes obtained in this study and those of the reference strains in the two pathovars (Darling et al., 2010). Genome-wide comparisons and visualization of orthologous clusters were performed using the online service OrthoVenn (Wang et al., 2015).

Type III Secretion System Repertory

Database searches were performed for the prediction of genes encoding different secretion systems including type III secretion system (T3SS) and type III effectors (T3Es) in *X. translucens* genomes using KEGG orthologies (KO) by implementing KofamKOALA (Aramaki et al., 2020) and BLASTn/BLASTp as described previously (Peng et al., 2016; Pesce et al., 2017). In brief, the sequences of 63 T3Es were retrieved from http://xanthomonas.org/ and two T3Es, i.e., XopE4 and XopE5 were obtained from EuroXanth DokuWiki. A dataset including the information of all 65 T3Es are shown in the **Supplementary Dataset 1**. For further confirmation, all annotated effector sequences were searched through NCBI GenBank, https://www.uniprot.org/, http://xanthomonas.org/ and EuroXanth DokuWiki, and all the genome sequences were analyzed one-vs.-one using BLAST or PSI-BLAST. Amino acid sequences (BLASTp) were used in all the analyses while nucleotide sequences (BLASTn) were also implemented in the investigations for further confirmation. For BLASTP, *e*-value = 1e-5 with a 50% query coverage and 35% sequence similarity were considered as cut-off criteria as recommended previously (Pearson, 2013; Wichmann et al., 2013). Ortho MCL v. 2.0 was used to generate groups of orthologous proteins with default parameters.

TALE Repertory of *X. translucens*

The genome sequences of 14 *X. translucens* strains were used for TALE prediction and TALE-based phylogenetic analysis. DisTAL v1.1 was used to align and phylogenetically classify TALEs based on their repeat arrangement (Pérez-Quintero et al., 2015). For the analysis of TALEs repeat variable diresidue (RVDs), we used AnnoTALE v1.2 that contained 516 *TALE* genes from 33 *Xanthomonas* strains (up to December 2020). First we analyzed all the *X. translucens* genomes and merged the TALEs RVDs output file into publicly available 516 TALEs RVDs. Then, the TALEs were grouped into different classes on the basis of RVDs that indicates their possible functional and evolutionary relationships (Grau et al., 2016; Erkes et al., 2017). TALE-CRR (central repeat region)-based tree was generated using DisTAL v1.1 with default parameters (Pérez-Quintero et al., 2015), while the resulting tree was visualized using FigTree v1.4.4 (Rambaut and Drummond, 2012). Furthermore, TALE repertory of the seven strains sequenced in study which have previously been

investigated by Southern blotting of *Bam*HI-digested genomic DNAs (Khojasteh et al., 2020) was confirmed using complete genome sequence data.

Data Availability

The dataset produced in this complete genome sequencing project is available at the NCBI GenBank/DDBJ/EMBL database under the accession numbers CP063993, CP063994, CP063995, CP063996, CP063997, CP063998, CP063999, CP064000, CP064001, CP064003, and CP064004 as detailed in **Table 1**.

RESULTS

The Genome of *X. translucens*

Complete genome sequencing of *X. translucens* strains were performed using Oxford Nanopore PromethION platform. The assembled sequences of all seven strains consisted of a single circular chromosome. Interestingly, in two of the seven *X. translucens* strains sequenced in this study, i.e., XtKm15 and XtLr8 two plasmids were found. The plasmids of the strain XtKm15 were designated as XtKm15_P1 (41,956 bp) and XtKm15_P2 (45,639 bp), while the two plasmids in XtLr8 were designated as XtLr8_P1 (45,351 bp) and XtLr8_P2 (40,770 bp) as detailed in **Table 1**. General features of the *X. translucens* sequences obtained in this study as well as the seven reference complete genome sequences retrieved from the GenBank are comparatively presented in **Table 1**. Genome size in the strains sequenced in this study ranged from 4,560,646 bp in XtKm15 to 4,792,950 bp in XtKm8, while the GC% content of the strains was from 67.79% in XtKm8 and XtKm34 to 68.06% in XtFa1. The number of protein coding genes ranged between 3,654 in XtKm12 to 3,873 in XtKm8. **Supplementary Figure 1** shows the circular diagram and genome features of the seven chromosomes and four plasmids resulted from this genome sequencing project. BLASTn-based investigation showed non-significant (<2% query coverage) similarities between the genome sequences of XtKm15 and XtLr8 and their accompanying plasmids (data not shown). However, the plasmid XtKm15_P1 had 92% query coverage and 99.5% sequence identity with XtLr8_P2 while XtKm15_P2 had 86% query coverage and 99.5% sequence identity with XtLr8_P1.

Phylogeny of *X. translucens*

Neighbor-joining distance clustering plot constructed using the genome sequences of 43 publicly available *X. translucens* whole genome sequences via ANI Calculator online service with all-vs.-all strategy revealed high genetic diversity among *X. translucens* complex species. Members of the three pathovars, i.e., pv. cerealis, pv. translucens, and pv. undulosa which are commonly referred to as translucens group were clustered in a monophyletic clade including all the 14 completed genome sequences investigated in this study (**Figure 1**). ANI values among the members of each of the translucens and graminis groups were higher than those observed between the members of the two groups (**Supplementary Figure 2**). However, the two clades *X. translucens* pv. *undulosa* and *X. translucens* pv. *translucens* were phylogenetically closer to one another while *X. translucens* pv. *cerealis* strains were clustered in a distinct clade. The three *X. translucens* pv. *translucens* strains sequenced in this study, i.e., XtKm8, XtKm9, and XtKm34 were clustered in two subclades where XtKm8 and XtKm34 were close to the reference strain DSM 18974^{T} while XtKm9 was clustered with the other three strains of this pathovar. Surprisingly, three strains, i.e., SLV-2, SIMT-07, and BLSB3 that are labeled as *X. translucens* pv. *translucens* in the NCBI GenBank database were clustered within the *X. translucens* pv. *undulosa* strains indicating a mislabeling in the identification of these strains. The four *X. translucens* pv. *undulosa* strains sequenced in this study, i.e., XtFa1, XtLr8, XtKm12, and XtKm15 were scattered through different subclades indicating that these strains are a proper set of representatives within the pathovar which were selected on the basis of MLSA results. As for the graminis group of *X. translucens*, all the strains designated as *X. translucens* pv. *graminis* were clustered in a monophyletic clade as well as the *X. translucens* pv. *arrhenatheri* strains. However, the five strains designated as *X. translucens* pv. *poae* were divided into two phylogenetically distinct clades where the two strains Utah5-P1 and LMG 728 were placed close to the *X. translucens* pv. *arrhenatheri* clade while the three strains ATCC 33804, CNC2-P4, and B99 were clustered in a distinct clade apart from the other members of graminis group as shown in **Figure 1**. Furthermore, the strains 569 and F5 have not been included in either group of *X. translucens* where the ANI values between these two strains and the other members of *X. translucens* were consistently below 93%, suggesting that the strains 569 and F5 could not be considered as *X. translucens*. Tacking together, a formal and comprehensive taxonomic study is warranted to address the two last taxonomic issues within *X. translucens* members and further refine the classification of this complex species.

Comparative Genomics

The core-genome of 14 *X. translucens* strains investigated in this study consisted of 2,175 genes appeared in >99% of the strains with >95% sequence similarity, while no soft core genes (presented in 95% to 99% of the strains) was detected. The number of shell genes presenting in 15% to 95% of the 14 strains was 2,384, while the number of cloud genes that found in 0% to 15% of the strains was 3,022. The pan-genome (total genes) of the 14 *X. translucens* strains was determined as 7,581 (**Figure 2A**). Distribution of the clusters of orthologous groups (COG) affiliated to biological functions is shown in **Figure 2B**. The highest proportion of unique genes in a certain COG was found in the orthologous groups assigned to replication, recombination and repair (L), and general function (R) making 65% and 48% of the total genes, respectively; followed by transcription (K) cluster. The least number of unique genes was found in translation, ribosomal structure and biogenesis (J) and energy production and conversion (C) orthologous groups, while no unique gene was found in the cell motility (N) COG. The highest number of core genes was predicted in the general function cluster (R) while the COGs assigned to defense mechanisms (V) showed the least number of core genes. The number of accessory genes with unknown function (S), cell wall/membrane/envelope biogenesis (M) and

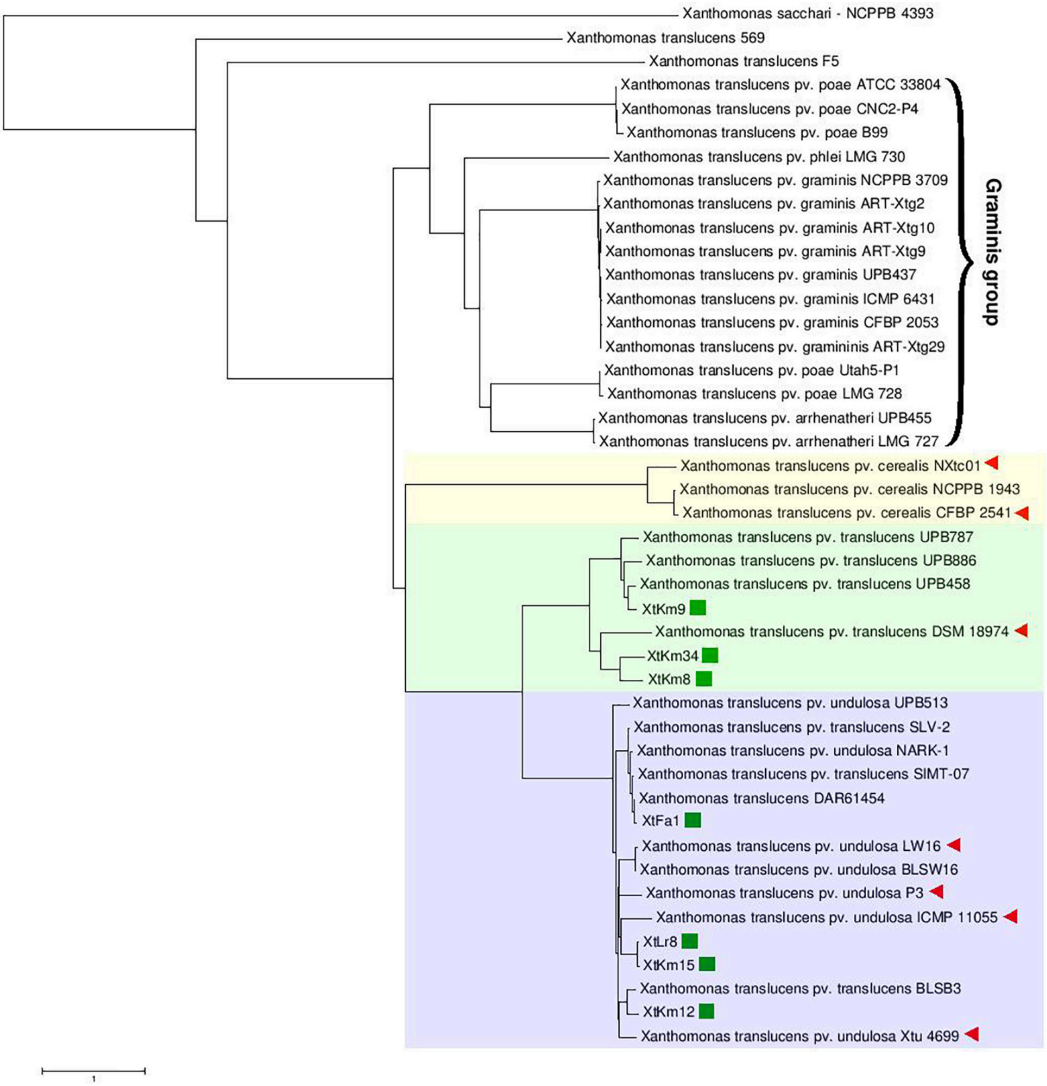


FIGURE 1 | Average Nucleotide Identity (ANI)-based Neighbor-Joining distance clustering plot of all available *Xanthomonas translucens* strains constructed using the ANI calculator online service. Members of *X. translucens* were clustered in two phylogenetically distinct clades, i.e., graminis and translucens groups. Among the strains sequenced in this study, XtKm8, XtKm9, and XtKm34 clustered within the *X. translucens* pv. *translucens* clade while the strains XtFa1, XtLr8, XtKm12, and XtKm15 clustered in the *X. translucens* pv. *undulosa* clade. Three strains, i.e., SLV-2, SIMT-07, and BLSB3 that labeled as *X. translucens* pv. *translucens* in the NCBI GenBank database were clustered within the *X. translucens* pv. *undulosa* strains indicating a mislabeling in the identification of these strains. Based on the ANI values, the strains 569 and F5 could not be included within *X. translucens*, where the ANI values between these two strains and the other members of the species were consistently below 93%.

intracellular trafficking, secretion, and vesicular transport (U) were proportionally similar to those of the core genes and unique genes in their respective COG as shown in **Figure 2B**. Translation, ribosomal structure and biogenesis (J) COG along with the coenzyme transport and metabolism (H) had the least number of accessory genes.

The core- vs. pan-genome plot generated with BPGA 1.3v software revealed an open pan-genome for the 14 *X. translucens* strains investigated in this study. **Figure 2C** represent the power-fit curve resulted from the equation [$f(x) = a.x\hat{}b$], where the exponent $b > 0$ indicates that the genome is open (Bosi et al., 2015). Therefore, the number of dispensable or accessory genes increases with the increase in the number of genomes indicating that the pan-genome of *X. translucens* has not yet been closed (**Figure 2C**). The "$b = 0.143$" indicates an open pan-genome for the 14 *X. translucens* but may be closed soon. In the other word, the unique gene pool should be increased by addition of newly sequenced *X. translucens* genomes. The pan-genome expansion analysis is biased and limited by the number of strains and pathovars used in the analysis. The number of singletons (genes unique to a single strain) in each of the 14 *X. translucens* strains is presented in **Figure 2D**. Considering the entire dataset, the highest number of singletons was found in the strain NXtc01 (369 singletons) followed by CFBP 2541

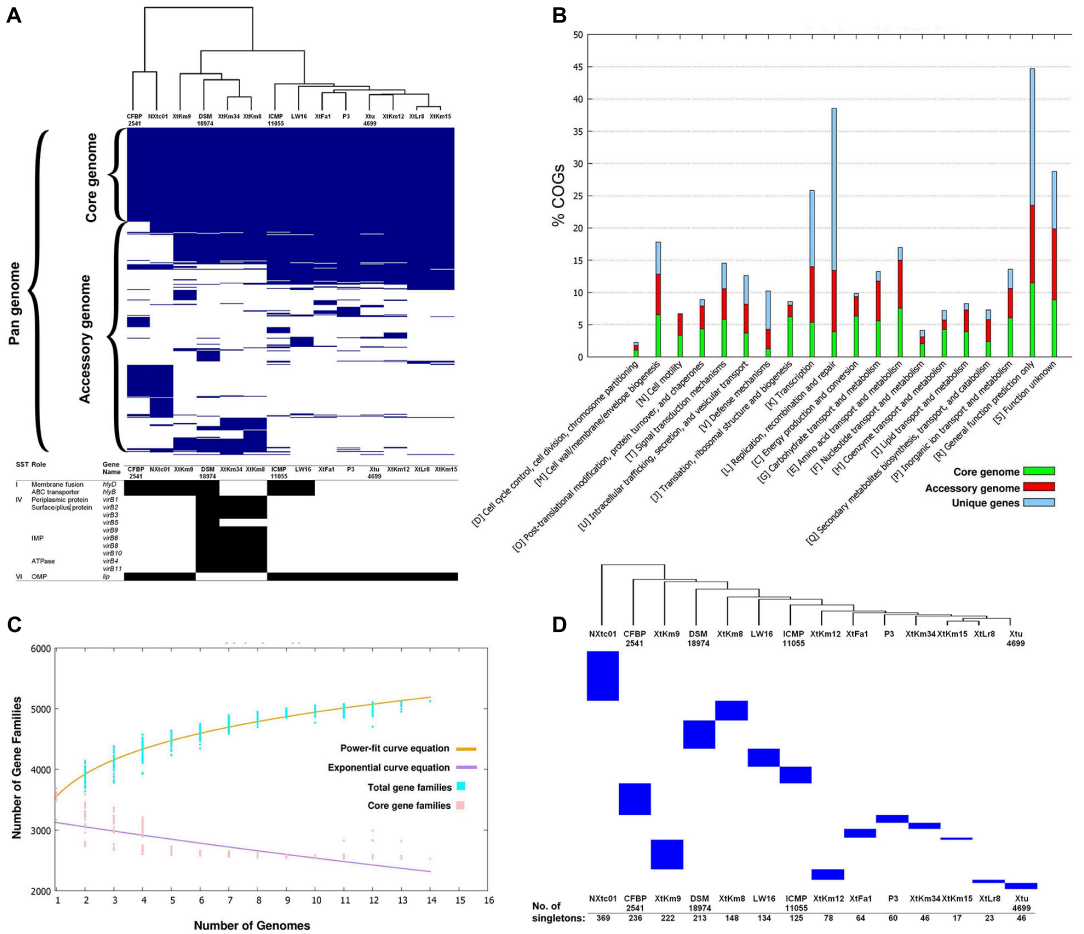


FIGURE 2 | Core- vs. pan-genome analyses of 14 *Xanthomonas translucens* strains using different criteria. **(A)** Ratio and size plot of core- and pan-genome generated using Roary 3.8.0 and visualized with Phandango. Each column corresponds to one strain while each row represents an orthologous gene family. The black and white table at the bottom of core- vs. pan-genome heat map created using KofamKOALA where black and white squares indicate presence and absence of a certain gene, respectively. **(B)** COG distribution of core, accessory and unique genes in the 14 *X. translucens* strains obtained using BPGA v. 1.3. **(C)** Total number of distinct gene families referring to the pan-genome of *X. translucens* strains illustrated by BPGA v. 1.3. **(D)** The number of singletons in each of the 14 *X. translucens* strains obtained using Roary 3.8.0.

(236 singletons) both belonging to *X. translucens* pv. *cerealis*. As for the strains sequenced in this study, the strain XtKm9 had the highest number of singletons (222) followed by XtKm8 (148 singletons) both belonging to *X. translucens* pv. *translucens*. The least number of singletons was found in XtKm15 (17 singletons) followed by XtLr8 (23 singletons) both belonging to *X. translucens* pv. *undulosa*, and the *X. translucens* pv. *translucens* strain XtKm34 (46 singletons).

Distribution of genomic islands - part of a genome that has evidence of having horizontal origins - in the sequences of seven strains obtained in this study and their respective reference genomes is shown in **Supplementary Figure 3**. The largest genomic islands were detected in the strain ICMP 11055 while most of the strains possessed a unique pattern of island distribution in their genome. In order to provide a comparative scheme for the genomes obtained in this study with their respective reference genomes, BRIG 0.95 was used to whole genome-based comparative genomics using DSM 18974^{T} and Xtu 4699 as reference genomes for *X. translucens* pv. *translucens* and *X. translucens* pv. *undulosa*, respectively (**Figures 3A,B**). Three *X. translucens* pv. *translucens* strains sequenced in this study lacked the genomics islands detected in DSM 18974^{T} (**Figure 3A**). For instance, a large island between the positions of 2,200 kbp and 2,500 kbp in DSM 18974^{T} was lacking in all the three *X. translucens* pv. *translucens* strains sequenced in this study. As for the *X. translucens* pv. *undulosa* strains, a large fraction of a genomic island in the position of 3,100–3,300 kbp was lacking in the four strains sequenced in this study. The islands between 3,900 kbp and 4,100 kbp in the genome of Xtu 4699 were also lacking in the query strains (**Figure 3B**). One-vs.-all collinearity test was performed among the genomes obtained in this study and their respective reference genomes using Mauve, where the organization of locally collinear blocks (LCBs) determined genome rearrangements and segmentation (**Figures 4A,B**). In the *X. translucens* pv. *translucens* genomes, the order of LCBs in the strain XtKm34 was almost entirely

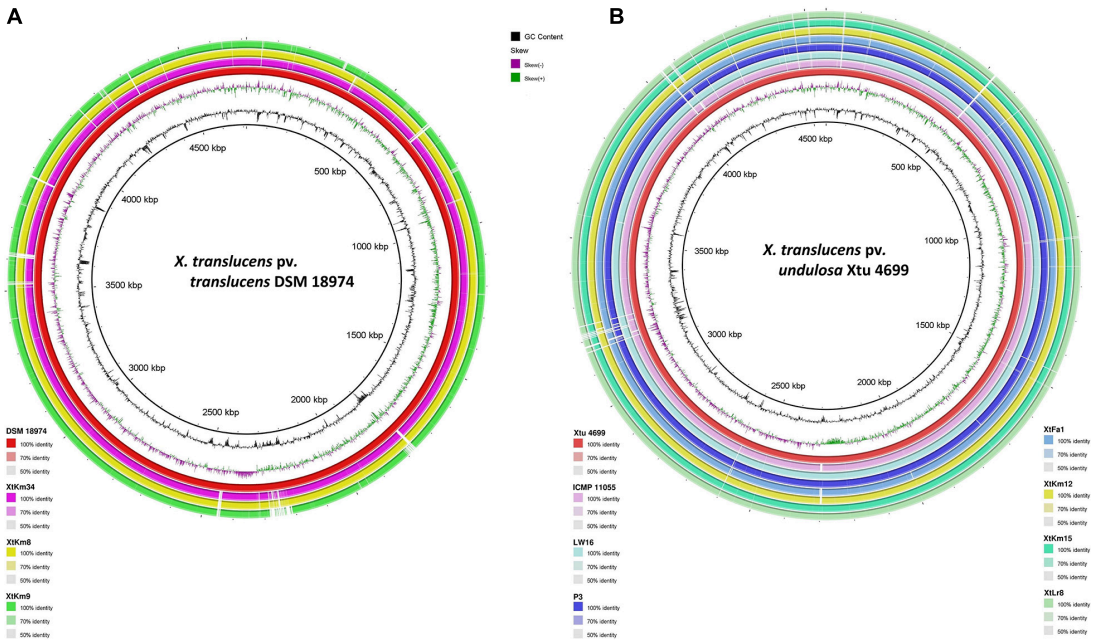


FIGURE 3 | Pairwise genome collinearity alignment of the seven *Xanthomonas translucens* strains sequenced in this study performed using BRIG 0.95. The genomes of DSM 18974^T **(A)** and Xtu 4699 **(B)** were used as reference genomes for *X. translucens* pv. *translucens* and *X. translucens* pv. *undulosa* strains, respectively.

in congruence with the LCBs in the reference genome DSM 18974^T except for a reversion in a 100 kbp (nucleotides 400–500 kbp) fragment in XtKm34. The strains XtKm8 and XtKm9 showed more variations in their LCB arrangement where almost 40% of the XtKm9 genome experienced a reversion as shown in **Figure 4A**. As for *X. translucens* pv. *undulosa*, the strains XtKm12, XtKm15, and XtLr8 had the LCB arrangement similar to the reference strain Xtu 4699. However, the genome of XtFa1 had multiple translocations, inversions and rearrangements (**Figure 4B**). Each LCB is a homologous region of sequence shared by the reference genome and the genomes under study with no rearrangements of homologous sequence. Hence, from the functional point of view it would be more probable for a LCB to have similar biological feature in their corresponding genomes.

Orthologous gene clusters were determined using OrthoVenn online service through four-vs.-four and five-vs.-five designations of the strains as shown in **Figure 5**. Three *X. translucens* pv. *translucens* strains sequenced in this study shared 3,296 proteins with the reference genome DSM 18974^T, while the strains XtKm8, XtKm9, and XtKm34 each had one, eighteen, and one unique proteins in their genomes (**Figure 5A**). As for the *X. translucens* pv. *undulosa*, the four strains sequenced in this study showed 3,375 shared proteins with the reference strain Xtu 4699. The strains XtFa1, XtLr8, XtKm12 and XtKm15 had two, zero, three and one unique proteins, respectively (**Figure 5B**). The protein contents of the four plasmids identified in the strains XtLr8 (i.e., XtLr8_P1 and XtLr8_P2) and XtKm15 (i.e., XtKm15_P1 and XtKm15_P2) were evaluated against the previously reported plasmid Xtc-CFBP 2541-G1-Mol002 in *X. translucens* pv. *cerealis* CFBP 2541. Surprisingly, only one protein sequence was found to be shared among the four plasmids identified in this study with no unique protein in each of the four plasmids. The two plasmids XtLr8_P1 and XtKm15_P2 had 42 shared proteins hypothecating their similar origin. The plasmids XtLr8_P2 and XtKm15_P1 had also 47 shared proteins. The reference plasmid Xtc-CFBP 2541-G1-Mol002 showed 13 unique proteins suggesting its genomic distinction from the four plasmids identified in this study (**Figure 5C**). We have also performed a BLASTn search using the genomes of the four plasmids identified in this study to determine the closest plasmids to them in the GenBank (**Supplementary Figures 4A–D**). For each plasmid, the top four plasmids with the highest sequence similarity were selected for an OrthoVenn-based orthologous gene clusters determination. The plasmid XtKm15_P1 shared 30 proteins with the *X. albilineans* plasmids GPE PC73 and pXaFJ1, *X. hortorum* pv. *pelargonii* plasmid CFBP2533_p47 and a *Xanthomonas* sp. CPBF 424 plasmid 2 (**Supplementary Figure 4A**); while the plasmid XtKm15_P2 had 28 shared proteins with the *Aminobacter* sp. plasmid pBAM1, *Cupriavidus necator* plasmid pENH91, *Delftia acidovorans* plasmid pNB8c and *Diaphorobacter* sp. plasmid pDCNB as shown in **Supplementary Figure 4B**. As for the plasmid XtLr8_P1, 37 shared proteins were detected among the *Yersinia pestis* plasmid pIP1203, *X. vesicatoria* plasmid pLM159.2, as well as the two plasmids pAKD1 and pSN1104-59 from uncultured bacteria (**Supplementary Figure 4C**). The plasmid XtLr8_P2 shared 21 proteins with the *X. hortorum* pv. *gardneri* plasmid pICMP7383.2, *X. hortorum*

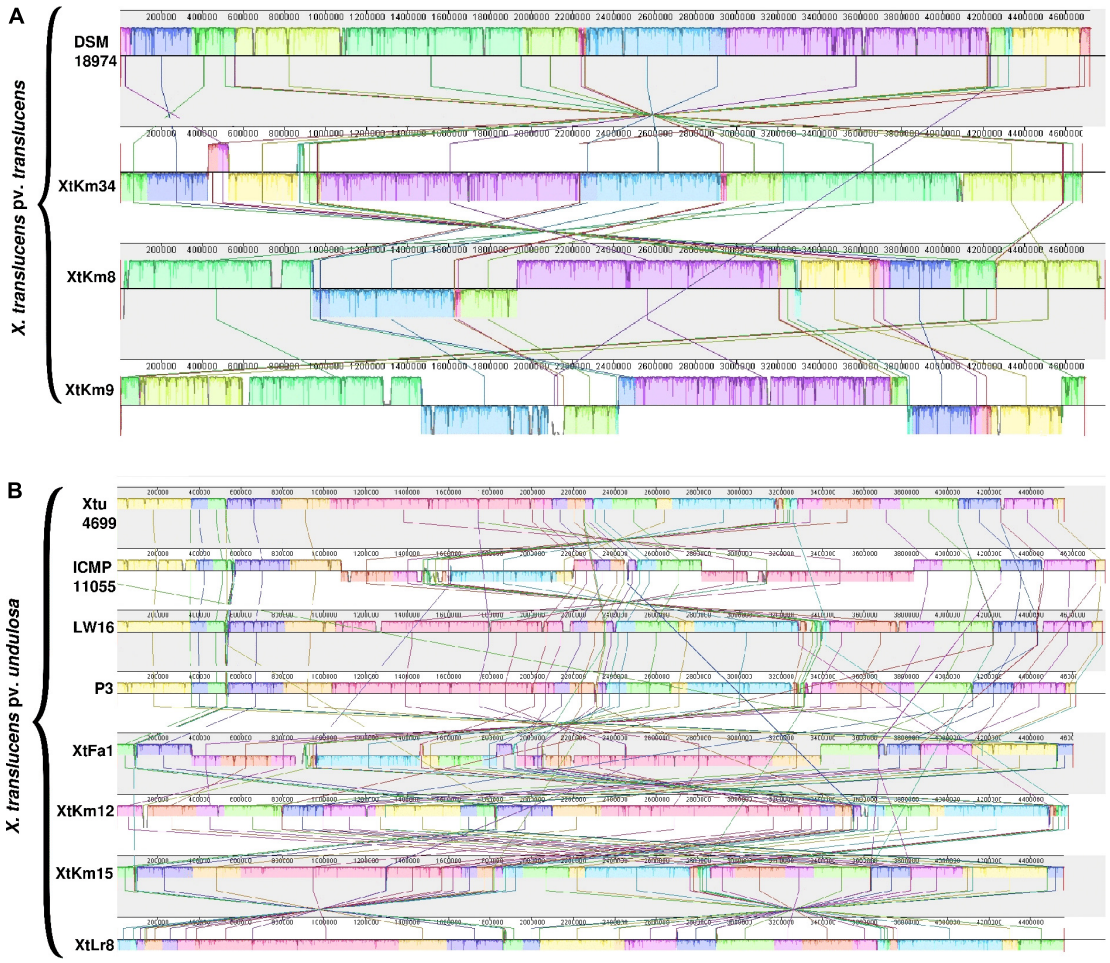


FIGURE 4 | Pairwise alignment among the chromosomal DNA of *Xanthomonas translucens* strains sequenced in this study and the strains DSM 18974^T and Xtu 4699 as reference genomes for **(A)** *X. translucens* pv. *translucens* and **(B)** *X. translucens* pv. *undulosa* strains, respectively using MAUVE software. Colors show conserved and highly related genomic regions (locally collinear blocks). Blocks shifted below the center line indicate segments that align in the reverse orientation as inversions relative to the respective reference strain. Each contiguously colored region is a locally collinear block, which is a region without rearrangement of homologous backbone sequence. Lines between two genomes trace each orthologous locally collinear block.

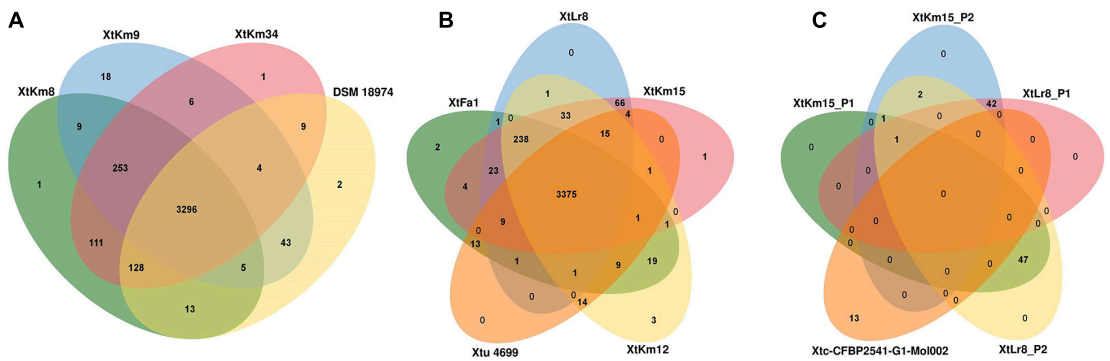


FIGURE 5 | Venn diagrams constructed using the OrthoVenn online service showing the shared gene families (orthologous clusters) among different sets of strains. The numbers indicate the number of shared gene families among certain set of genomes. **(A)** *X. translucens* pv. *translucens* strains, **(B)** *X. translucens* pv. *undulosa* strains, and **(C)** plasmids XtKm15_P1 and XtKm15_P2 associated with the strain XtKm15 as well as XtLr8_P1 and XtLr8_P2 associated with the strain XtKm8. The genomes of DSM 18974^T and Xtu 4699 were used as reference genomes for *X. translucens* pv. *translucens* and *X. translucens* pv. *undulosa* strains, respectively while the plasmid Xtc-CFBP 2541-G1-Mol002 associated with *X. translucens* pv. *cerealis* strain CFBP 2541 was used as reference for the four plasmids.

pv. *pelargonii* plasmid CFBP2533_p47, *X. hortorum* plasmid pB07007 and a *Xanthomonas* sp. CPBF 424 plasmid 2 (**Supplementary Figure 4D**).

Type III Secretion System of *X. translucens*

The *X. translucens* genomes were evaluated for potential variations in their secretion systems and the corresponding genes. Significant variations were found among the 14 *X. translucens* dataset in their membrane fusion (*hlyD*) and ABC transporter (*hlyB*) genes where these genes were lacking in the *X. translucens* pv. *translucens* strains XtKm8 and XtKm34 but present in XtKm9 and the reference strain DSM 18974^T. None of the *X. translucens* pv. *undulosa* strains sequenced in this study had these genes. The strain XtKm9 was more similar to the *X. translucens* pv. *cerealis* strains than to the other *X. translucens* pv. *translucens* strains in the evaluated features. Except for the *hlyD* and *hlyB* genes, all the *X. translucens* pv. *undulosa* strains were similar to each other in the secretion system repertories as shown in **Figure 2A**. To assess the pathogenicity repertories of *X. translucens* we compared the T3SS features among the 14 dataset. Due to the pivotal role of T3SS in delivering virulence associated effector proteins into host cells any defect in the T3SS will be leading to an attenuated virulence or a complete loss of bacterial pathogenicity. The T3SS in all *X. translucens* strains is encoded by 25 genes from *hpaH* to *hpaD* (>23 kb). The structural organization of the 25 genes in all *X. translucens* was almost identical (**Supplementary Figure 5A**). We identified six *hpa*, eight *hrp* and eleven *hrc* genes which are conserved in all *X. translucens* strains. The T3SS regulatory genes, *hrpG* and *hrpX,* were positioned inside the *hrp* gene cluster in all *X. translucens*, different from other Xanthomonas species (**Supplementary Figure 5A**). The structure and arrangement of T3SS in the 14 *X. translucens* dataset were compared against a set of five plant pathogenic xanthomonads representing members from clade I and II of the genus (**Supplementary Figure 5B**). In *X. translucens* an unknown ORF gene was found between the *hpaT* and *hrcC* loci while the *hpaB* gene was located between the *hrpE* and *hrpG* genes which were different from the other xanthomonads.

Non-TAL Effectors (Xops) of *X. translucens*

In silico analyses revealed that the 14 *X. translucens* strains investigated in this study encode a set of 29–36 non-TAL effectors (Xop) as shown in **Table 2**. As for the strains sequenced in this study, the number of Xops was 30 in XtFa1, 29 in XtLr8, 31 in XtKm12, 29 in XtKm15, 34 in XtKm8, 36 in XtKm9, and 33 in XtKm34. Furthermore, 29 and 30 Xops were predicted in the genome sequences of the strains LW16 and P3, respectively (**Table 2**). Among the Xops predicted, 17 Xops, i.e., AvrBs2, XopAA, XopAF1, XopAM, XopAP, XopAV, XopAZ, XopC2, XopF, XopG, XopK, XopN, XopP, XopQ, XopV and XopX as well as XopR (possessing a frameshift mutation) were conserved among all the 14 *X. translucens* dataset. Our results revealed eight core Xops, i.e., AvrBs2, XopF, XopK, XopN, XopP, XopQ, XopX, and XopR in the 14 *X. translucens* dataset while in the previous studies the two effectors XopL and XopZ have also been predicted as the core T3Es. Although two copies of XopL have previously been reported in Xtu 4699 and ICMP 11055 strains we did not find this T3E in our analyses, while three copies of XopL were found in CFBP 2541 instead of previously reported four (Peng et al., 2016; Falahi Charkhabi et al., 2017; Shah et al., 2019). All the 14 *X. translucens* dataset contained multiple copies of AvrBs2 (n = 2), XopAF1 (n = 1–2), XopAZ (n = 2), XopF1 (n = 2), XopL (n = 0–4), XopP (n = 1–3) and XopX (n = 3). Several inconsistencies were observed among the results obtained in this study and those reported in the literatures regarding the presence/absence and the copy number of T3Es in *X. translucens* (Peng et al., 2016; Falahi Charkhabi et al., 2017; Shah et al., 2019). A number of Xops were detected in the sequences of reference *X. translucens* strains in our analyses which have not previously been reported in their respective strains. For instance, in the strains Xtu 4699 and ICMP 11055 XopAV (n = 1) and XopAZ (n = 2), in the strain NXtc01 XopAV (n = 2) and XopZ (two copies instead of one), in CFBP 2541 AvrXccA1 (n = 1), XopAV (n = 2) and XopAZ (n = 2) and in DSM 18974^T XopAJ (n = 1), XopAL1 (n = 1), XopAV (n = 2), XopAZ (n = 2), XopE5 (n = 1) and XopM (n = 1) were different in our analyses from those reported in the literatures (**Table 2**).

On the other hand, a number of Xops have previously been reported in certain reference strains but have not been detected in our analyses. For instance, AvrBs1, XopAD, XopE1, XopE2 and XopZ1 in CFBP 2541; XopE1 and XopE2 in NXtc01; XopAH and XopE2 in DSM 18974^T and XopB, XopE1, XopE5, XopG1 and XopL in Xtu 4699 and ICMP 11055 were absent in our analyses. Furthermore, in Xtu 4699, two copies of XopAF1 have previously been reported while only a single copy was found in our analyses. A single copy of XopAF1 was found in all *X. translucens* pv. *undulosa* and *X. translucens* pv. *cerealis* strains except for the strain XtFa1 which had two copies, while all *X. translucens* pv. *translucens* strains carried two copies of XopAF1. AvrXccA1 was detected only in *X. translucens* pv. *cerealis* strains while XopAD was predicted in all 14 *X. translucens* dataset except for CFBP 2541. Further, XopAH has not been detected in any of seven strains sequenced in this study. XopAJ, XopAL1 XopE3 and XopM were found in all *X. translucens* pv. *translucens* strains but not in *X. translucens* pv. *undulosa* and *X. translucens* pv. *cerealis* strains (**Table 2**), indicating their host specificity nature. Considering the seven reference strains, inconsistencies between the data reported in the literature and those obtained in this study could be due to the continuous up-gradation in the T3E databases.

TALE Diversity in *X. translucens*

Complete genome sequence-based investigations have determined the TALE repertories of the seven strains sequenced in this study where the number of TALEs in the strains XtKm8 = 8, XtKm9 = 5, XtKm34 = 7, XtKm12 = 7, XtKm15 = 8, XtFa1 = 7 and XtLr8 = 8 (**Supplementary Table 1**). In most of the TALE genes, conserved *Bam*HI and/or *Sph*I sites at either N- and/or C-terminus were missing. TALEs of XtKm8 encoded proteins with 7 to 22 RVDs, while TALEs of XtKm9 encoded 12 to 17 RVDs, and TALEs of XtKm34 encoded 7 to 19 RVDs.

TABLE 2 | Predicted type III effectors (T3Es) in the seven *Xanthomonas translucens* strains sequenced in this study as well as the seven reference strains obtained from the NCBI GenBank.

	Xtu								Xtt				Xtc	
T3Es	**XT4699**	**ICMP11055**	**P3**	**LW16**	**XtLr8**	**XtKm15**	**XtKm12**	**XtFa1**	**DSM 18974^T**	**XtKm8**	**XtKm34**	**XtKm9**	**NXtc01**	**CFBP 2541**
AvrBs1	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
AvrBs2	++	++	++	++	++	++	++	++	++	++	++	++	++	++
AvrXccA1	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	+	+
TALEs	8	7	8	8	8	8	7	7	8	8	7	5	2	2
XopAA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
XopAD	+	+	+	+	+	+	+	+	+	+	+	+	+	ND
XopAF1	+	+	+	+	+	+	+	++	++	++	++	++	+	+
XopAH	+	+	+	+	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
XopAJ	ND	ND	ND	ND	ND	ND	ND	ND	+	+	+	+	ND	ND
XopAK	+	+	+	+	+	+	+	+	+	+	+	+	ND	ND
XopAL1	ND	ND	ND	ND	ND	ND	ND	ND	+	+	+	+	ND	ND
XopAM	+	+	+	+	+	+	+	+	+	+	+	+	+	+
XopAP	+	+	+	+	+	+	+	+	+	+	+	+	+	+
XopAV/AY	+	+	+	ND	ND	ND	++	ND	+	ND	ND	+	++	++
XopAZ	++	++	++	++	++	++	++	++	++	++	++	++	++	++
XopB	ND	ND	+	+	+	+	+	+	+	+	+	+	+	+
XopC2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
XopE1	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
XopE2	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
XopE3	ND	ND	ND	ND	ND	ND	ND	ND	+	+	+	+	ND	ND
XopE4	ND	ND	ND	ND	+	+	+	+	ND	ND	ND	ND	ND	ND
XopE5	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	+	+
XopF1	++	++	++	++	++	++	++	++	++	++	++	++	++	++
XopG1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
XopJ5	ND	ND	ND	ND	+	+	+	+	+	+	+	+	+	ND
XopK	+	+	+	+	+	+	+	+	+	+	+	+	+	+
XopL	ND	ND	++	++	+	+	+	+	++	++	++	++	++++	+++
XopM	ND	ND	ND	ND	ND	ND	ND	ND	+	+	+	+	ND	ND
XopN	+	+	+	+	+	+	+	+	+	+	+	+	+	+
XopP	++	++	++	++	++	++	++	++	+++	++	+	+++	+++	+++
XopQ	+	+	+	+	+	+	+	+	+	+	+	+	+	+
XopR	F	F	F	F	F	F	F	F	F	F	F	F	F	F
XopV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
XopX	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
XopY	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
XopZ1	+	+	+	+	+	+	+	+	+	+	+	+	+	ND

Hypothetical T3Es were predicted using BLASTn and BLASTp. The annotated T3E sequences were obtained from the NCBI GenBank, https://www.uniprot.org/, http://xanthomonas.org/ and EuroXanth DokuWiki. The number of T3Es varied between 29 and 36 while eight genes were identified as core T3Es (highlighted in gray). The number of plusses (+) indicates the copy number of T3Es detected in each genome; ND, not detected; F, the gene harbors a frameshift mutation. Green highlight: T3Es have been reported previously but not found in our analyses. Red highlight: T3Es have not been reported previously but found in our analyses. Blue highlight: T3Es have been reported previously in full length but found to be frameshifted in our analyses. Orange highlight: two copies of T3Es have been reported previously but we found a single copy. Yellow highlight: T3Es have been reported previously in four copies but we found in three copies.

TALEs of the strains XtKm15, XtLr8, and XtKm12 encoded 14–18, 12–18, and 15–18 RVDs respectively, while the TALEs of XtFa1 encoded 13 to 18 RVDs. A schematic map of *tal* genes in all 14 *X. translucens* strains is illustrated in **Figure 6A** where all chromosomes are linearized starting from DNA gyrase subunit B (*gyrB*) gene. The *tal* genes were distributed among four loci in *X. translucens* pv. *translucens*, while in five to six loci in *X. translucens* pv. *undulosa* and in two loci in *X. translucens* pv. *cerealis*. In the *X. translucens* pv. *translucens* strains DSM 18974^T and XtKm8 two loci consisted of a single gene, one locus consisted of two *tal* genes and the other locus consisted of four, all oriented in the same direction. However, in the strains XtKm9 and XtKm34 three loci consisted of a single gene, one locus in XtKm34 consisted of four and one locus in XtKm9 consisted of two genes. In *X. translucens* pv. *cerealis* both loci comprised of a single *tal* gene entirely conserved between the two strains, while

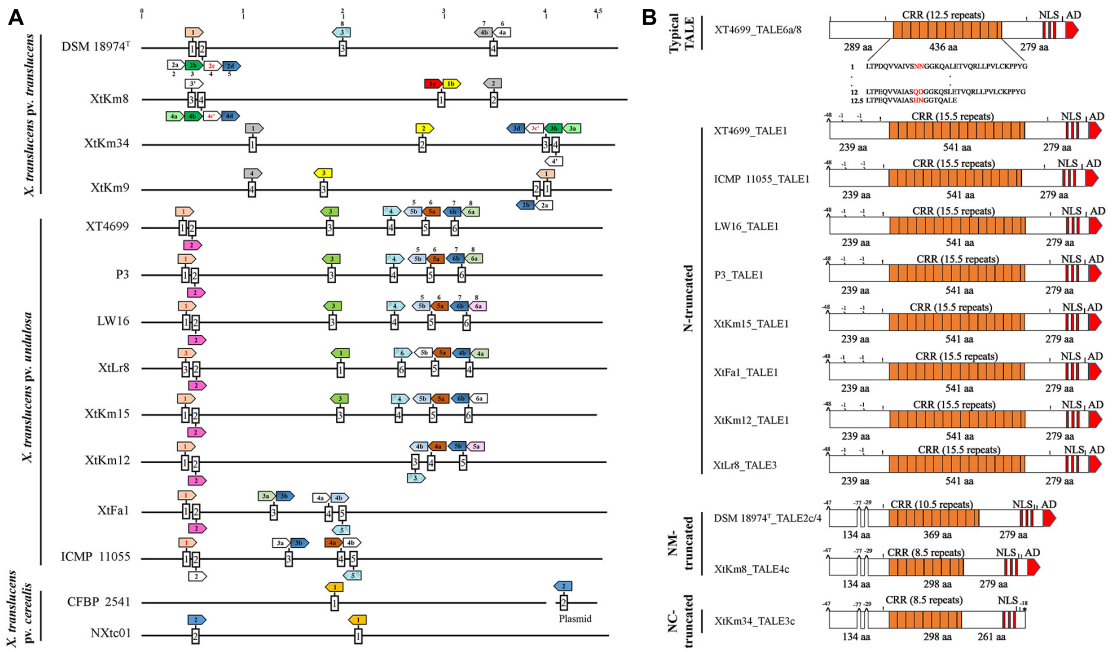


FIGURE 6 | TALE (transcription activator-like effector) repertories of the seven *Xanthomonas translucens* strains sequenced in this study in comparison to the seven reference sequences obtained from the NCBI GenBank. **(A)** A linearized plot of the chromosomes of 14 genomes starting from DNA gyrase subunit B (*gyrB*) gene to indicate the distribution and localization of TALE genes in each strain. Each box represents a TALE locus where TALEs that had identity at the 12th and 13th codons (Zero to three RVDs difference) are indicated by colored arrows while pseudogenes are indicated by red fonts inside the arrows. Names of the previously characterized TALEs are indicated above the corresponding arrow. TALEs having RVDs ≤ 9 are indicated by an apostrophe. The remaining TALEs (≥four RVD differences), unique to each strain are not colored. Scale at top of chromosomes in MB. **(B)** Schematic maps of truncated tal genes where Xtu 4699_TALE6a/8 was used as a reference gene at top. All other TALEs have short N-terminus while XtKm34_TALE3c have deletion of amino acids (aa) both at N- and C-termini. The number of RVDs in each TALE is specified. Symbols shown above the TALE horizontal line (|) *Bam*HI site (!) *Sph*I site, ˆ deletion of 29 to 77 aa at start and middle of N-termini, ′ means deletion of single aa and * denote missing activation domain at the said position comparing with reference TALE. Length of CRR, N- and C-termini are displayed in aa.

in the strain CFBP 2541 one of the *tal* genes was located on an endogenous plasmid. In *X. translucens* pv. *undulosa* strains Xtu 4699, P3, LW16, XtLr8 and XtKm15, four loci possessed single *tal* gene whereas in the strains XtKm12, XtFa1 and ICMP 11055 three loci had single genes. The remaining two loci in all the *X. translucens* pv. *undulosa* strains had two *tal* genes oriented in the same direction (**Figure 6A**). The number of *tal* genes in the seven *X. translucens* strains sequenced in this study was also confirmed using southern blot technique as detailed previously (Khojasteh et al., 2020).

In addition, a number of truncated *tal* genes were predicted among the 14 *X. translucens* dataset. For instance, the TALEs XtKm8_tal4c, XtLr8_tal3, DSM 18974^T_tal2c and tal1 in the strains XtFa1, XtKm12, XtKm15, Xtu 4699, ICMP 11055, LW16 and P3 were distinct from the other TALE genes having shortened N-terminus while XtKm34_tal3c have shortened both N- and C-termini (**Figure 6B**). Xtu 4699 contained eight *tal* genes on six loci where TALE6a/8 which is known as virulent factor contributing to bacterial leaf streak development in wheat was used as a reference gene to determine pseudogenes (**Figure 6B**). Some *tal* genes including Xtu 4699_tal1, ICMP 11055_tal1, LW16_tal1, P3_tal1, XtKm15_tal1, XtFa1_tal1, XtKm12_tal1 and XtLr8_tal3 has 48 amino acid (aa) deletion at the start of N- terminus and also lack classically conserved *Bam*HI site. Single aa deletions were also found in all these *tal* genes at the mid of N-terminus. Some other *tal* genes, e.g., DSM 18974^T _tal2c/4, XtKm8_tal4c and XtKm34_tal3c had 48 aa deletions at start and 29 to 77 aa at middle of N-terminus. XtKm34_tal3c carried a premature stop codon, probably encoding a protein with a C-terminal truncation of 18 aa, leading to deletion of activation domain (**Figure 6B**). All truncated *tal* genes lack classically conserved *Bam*HI site at N-terminus and *Sph*I site at C-terminus of some genes.

TALEs of all 14 *X. translucens* strains composed of both 34 and 35 amino acid repeat types except for the last repeat of each TALE. In each repeat, 12th and 13th amino acids (termed RVD) comprised of some unusual RVDs, i.e., KG, QD, YK, YD, NF, GI, KI and Y*. These RVDs have rarely been found in xanthomonads. In the strain XtKm8, two unique RVDs HE and YI were identified in the TALE XtKm8_TALE1a that have not previously been reported. Among the 14 *X. translucens* strains TALEs were grouped into 16 classes, each class consisting of perfectly or nearly conserved TALEs (up to three variations in RVDs) with the exception of the classes 2, 8, 9, 12, and 16. Only one TALE in class 12 (XtFa1_TALE4a) and class 8 (XtKm9_TALE2a), two TALEs in class 2 (XtKm9_TALE1 and DSM 18974^T _TALE1), three in class 9 (ICMP 11055_TALE2, XtKm8_TALE4a and XtKm34_TALE3a) and almost all of class

16 were found distinct from the remaining TALEs in the class (**Supplementary Table 1**).

Four *X. translucens* TALEs have previously been reported to have contribution to virulence and host susceptibility on wheat plants. All these four TALEs were detected in our dataset and were grouped into separate classes where NXtc01_TALE1 was in class 6, ICMP 11055_TALE2 was in class 9, Xtu 4699_TALE6a/8 was in class 11, and ICMP 11055_TALE4b was in class 16. The Xtu 4699_TALE6a/8 encodes the major virulence determinant for Xtu 4699 whose function is to promote disease susceptibility by targeting host gene TaNCED-5BS that encodes 9-cis-epoxycarotenoid dioxygenase (Peng et al., 2019). The class 11 that contains this TALE comprised of perfectly identical four TALEs of *X. translucens* pv. *undulosa* including P3_TALE6a/8, Xtu 4699_TALE6a/8, XtFa1_TALE3a, and XtLr8_TALE4a, suggesting their similar functioning in the host plant (**Supplementary Table 1**). Unexpected similarity in the CRR domain of TALEs between the *X. translucens* pv. *translucens* strains XtKm8, XtKm9 and XtKm34 and those of the *X. translucens* pv. *undulosa* strains XtKm12, XtKm15, XtFa1 and XtLr8 led us to strictly explore their relationship with other *X. translucens* strains (**Figures 7A,B**). Based on the RVDs-CRR structure of all available complete genomes, *X. translucens* TALEs were classified into five clades (I-V) as shown in **Figure 7**. Each clade comprised of two or more TALE classes while TALEs of the class 16 that had different TALE-RVDs distributed throughout all clades. In addition, virulence-associated TALEs, i.e., Xtu 4699_tal6a/8, ICMP 11055_tal2, ICMP 11055_tal4b, ICMP 11055 and NXtc01_tal1, were classified in clades II, III and V, respectively. TALEs possessing similar RVDs (TALEs-classes see **Supplementary Table 1**) grouped on a same node (shown with colored nodes in **Figures 7A,B**) disseminated either in a single clade (i.e., clade-III) or multiple classes in a single clade, while different TALEs (i.e., class 16) have not been clustered in a single clade suggesting that identical CRRs may encode similar RVDs.

DISCUSSION

In this study, we have provided the complete genome sequences of seven highly virulent *X. translucens* strains chosen among a set of 57 strains isolated from wheat, barley and ryegrass across the cereal growing areas in Iran. The overall genome structure, phylogenetic position, core *hrp* cluster, non-TALE T3Es and TALE contents of the seven strains were compared with all available complete genomes of *X. translucens* in the NCBI GenBank, i.e., two *X. translucens* pv. *cerealis* strains CFBP 2541 and NXtc01 isolated in the United States and China, respectively, the *X. translucens* pv. *translucens* strain DSM 18974^{T} isolated in the United States and four *X. translucens* pv. *undulosa* strains ICMP 11055, Xtu 4699, P3 and LW16 isolated in Iran and the United States. Despite the temporal and geographic distinctions among the origin of the strains, all the genomes were highly syntenous and their non-TALE T3Es as well as TALE repertoires were highly conserved particularly at pathovar level. The data generated in this study, has doubled the number of publicly available complete genome resources of the bacterial leaf streak pathogen providing a wider framework for the population structure of *X. translucens*.

In the previous study, Khojasteh et al. (2019) have demonstrated high genetic diversity of the bacterial leaf streak pathogens in Iran which is in congruence with the fact

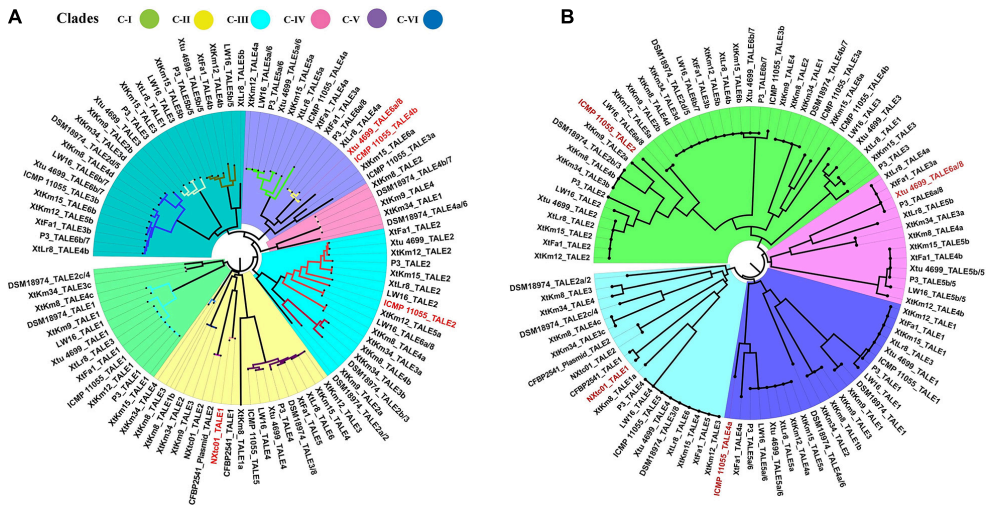


FIGURE 7 | Phylogenetic tree based on TALEs-CRR **(A)** and RVDs **(B)** in the seven *Xanthomonas translucens* strains sequenced in this study as well as the seven reference sequences obtained from the NCBI GenBank. TALEs were designated numerically according to the locus order (**Figure 6A**). Former TALE names are pointed out in the tree with slash, while red font indicates virulence associated TALEs. **(A)** DisTAL was used to construct a Neighbor-joining tree for a set of 93 TALE sequences from 14 *X. translucens* strains. TALEs were classified into six major clades as block shaded in different colors. TALE classes (**Supplementary Table 1**) are represented with colored nodes with the exception of class 16 distributed among all clades. **(B)** FuncTAL was used to build a Neighbor-joining tree for a set of 93 TALE sequences based on TALE-RVDs. TALEs were classified into four major groups (16 classes; **Supplementary Table 1**) as block shaded in different colors.

that the center of origin of cultivated wheat is determined in the Fertile Crescent overlapping with Iranian Plateau in Karacadağ Mountains in southeast Turkey (Heun et al., 1997; Brandolini et al., 2016). On the other hand, it has been speculated that the new world population of the bacterial leaf streak pathogens has been originated from the Iranian Plateau as indicated by phylogeographic analyses (Khojasteh et al., 2019). Except for the strains ICMP 11055 and NXtc01 which were isolated in Iran and China, respectively, the *X. translucens* strains having available complete genome resources were originated from the United States, narrowing our understanding of the population structure and genomic features of the species. The complete genome sequences provided in this study include a set of taxonomically diverse representatives of the species all isolated from the old world. Comparative genomics and phylogenomics among the entire set of 14 strains provide a comprehensive insight into the global population of *X. translucens*.

The members of *Xanthomonas* encode a typical Hrp-T3SS comprising six hpa (hrp associated), eleven hrc (hrp conserved) and eight hrp genes (Timilsina et al., 2020). Functional analyses have proven the pivotal importance of *X. translucens* T3SS for pathogenicity, induction of HR and delivery of T3Es. However, functional variations were reported for different set of T3SS genes. For instance, mutant of *hrcC* of wheat pathogens NXtc01 and Xtu 4699, and *hrcT* of barley pathogens UPB886, UPB787R and *X. translucens* pv. *hordei* resulted complete loss of the symptom development on host and HR on non-host plants compared to the wild type strain (Peng et al., 2016; Pesce et al., 2017; Shah et al., 2019). In contrast, the *hrcE*, *hrpG* and *hrcR* mutants of grass pathogen Xtg-Xtg29 cannot eliminate disease symptoms completely and colonization is also not effected (Wichmann et al., 2013). Comparison of core *hrp* cluster revealed similar organization in all seven *X. translucens* strains sequenced in this study which was in congruence with their aggressiveness and virulence features described previously (Khojasteh et al., 2020). This study also reveals the non-TALEs/Xop effectors repertoire in *X. translucens* strains in comparison to the other reference genomes of plant pathogenic *Xanthomonas* spp. (Peng et al., 2016; Shah et al., 2019). However, with respect to the previous studies we have noted variations in the copy number, frame-shift mutation and presence or absence of individual Xops (**Table 2**). Alignment and comparison of genomic data indicate that a core set of T3SEs identified previously is present in the sequenced *X. translucens* genomes but surprisingly two core effectors XopL in Xtu 4699 and ICMP 11055, and XopZ in CFBP 2541 which have previously been reported to present in these strains were not found in our analyses (**Table 2**). These inconsistencies in the T3SEs might be due to up-gradation of T3Es database implemented in the analyses. The unique T3E repertoire in different pathovars of *X. translucens* and within individual strains might reflect host specificity of the strains to various small grain cereals or specific genotypes of a host (Jacques et al., 2016).

So far, none of the T3Es of *X. translucens* have been functionally characterized nor tested for their contribution to colonization and virulence of the pathogen. TALEs that act like transcription factor inside host nucleus are important virulence factors facilitating the proliferation of the pathogens with the ability to directly bind to the promoter region of the target genes. Comparison of TALEs of the 14 *X. translucens* genomes revealed divergent subfamilies (**Figure 7**). All *X. translucens* pv. *translucens* strains, i.e., DSM 18974^{T}, XtKm8, XtKm9, and XtKm34 share two TALE genes tal4b = tal2 = tal1 = tal4 and tal2d = tal4d = tal3d = tal2b, while the first three strains (DSM 18974^{T}, XtKm8 and XtKm9) and last three strains (XtKm8, XtKm9, and XtKm34) share additional one tal gene in each including tal2b = tal4b = tal3b and tal1b = tal2 = tal3. Similarly, one tal of all *X. translucens* (tal2d = tal4d = tal3d = tal2b), and two additional TALEs of DSM 18974^{T} and one of XtKm9 (tal1 = tal1 and tal3/8) are also found in all *X. translucens* pv. *undulosa* strains. Other common TALEs in *X. translucens* pv. *undulosa* strains Xtu 4699, P3, LW16, XtLr8, XtKm15, XtKm12, XtFa1 and ICMP 11055 include tal2 of all strains except ICMP 11055, tal5a and tal4a of all strains excluding XtFa1, tal5b with the exception of XtLr8 and tal4b except for ICMP 11055 and tal3 of some strains (**Figure 6A**). Some other TALEs were also found common in two or more strains (**Figures 6**, **7**). These identical *tal* genes were acquired prior to pathovars divergence, implying their important role in the pathogenesis of the bacterial leaf streak in small grain cereals. Other *tal* genes which are unique in each strain suggest independent acquisition in the lineage.

A family of TALE variants categorized into two forms, i.e., iTALEs (interfering TALEs) and truncTALEs (truncated TALEs) has been introduced in the previous studies (Ji et al., 2016, 2020; Read et al., 2016). The iTALEs lack C-terminal transcription activation domains due to the introduction of premature stop codon in the coding sequence of the genes, whereas truncTALEs have large deletion of the coding sequence at the 3' end of the genes (Ji et al., 2020). TALEs comparison of all *X. translucens* strains exhibit one iTALE, i.e., XtKm34_TALE3c possessing truncation of 18 aa at the C-terminal domain due to premature stop codon (**Figure 6B**), None of the other TALEs had deletion at C-terminus but all TALE variants harbored two conserved internal deletions from 1–77 aa at the N-terminus. Surprisingly, all truncaTALEs retained their CRR, NLS and AD but the variants suffer a large 47–48 aa deletions in the N-terminal region that removes a part of the type III secretion signal. Furthermore, all the truncTALEs also lack classically conserved *Bam*HI site of the N-terminus and found improper distribution of C-terminus *Sph*I site in DSM 18974^{T}_TALE2c/4, XtKm8_TALE4c and XtKm34_TALE3c (**Figure 6B**).

In conclusion, results of the present study revealed a greater diversity in the virulence determinants and pathogenicity repertories among the worldwide population of *X. translucens* than the one that had been described before. More specifically, based on the comparative genomics of the 14 strains we have noted that the strains isolated in Iran are similar to the new world strains in T3E arrangement and non-TAL effectors. However, significant variations were observed in the TALE repertories of the strains. On the other hand, our results suggest that the genomic contents of the bacterial leaf streak pathogens should be further investigated using a pool of strains from all the known hosts of the pathogen, including gramineous weeds

in the center of origin of the host crop. Previous studies revealed that presence of plasmids have major impact on metabolic functions and host adaptation (Niu et al., 2015). The presence of plasmids in *Xanthomonas* significantly enhances the tolerance to the stresses of heavy metal ions. Our data could be helpful to further elucidate the biological significance of these plasmids and the adaptive evolution of *X. translucens* pv. *undulosa*. If done, these evaluations will pave the way of searching for new sources of resistance among the wild population of wheat species and will help to find new breeding strategies to develop resistant cultivars. Only future studies based on population genetics, comparative genomics, and pathogenicity assays of a wider collection of strains isolated from different hosts and geographical regions can shed more light on these areas.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, CP063993, CP063994, CP063995, CP063996, CP063997, CP063998, CP063999, CP064000, CP064001, CP064003, and CP064004.

AUTHOR CONTRIBUTIONS

GC and EO conceived and designed the study with assistance from SS and MK. SS and MK carried out the experiments with assistance from PK, LZ, and SM. SS and MK analyzed and interpreted the data with assistance from ST, QW, and ZX. EO prepared the article with assistance from MK, SS, and GC. All the authors revised the final version of the manuscript, while GC and EO acted as the corresponding authors. All authors contributed to the article and approved the submitted version.

FUNDING

This study was financially supported by the National Natural Science Foundation of China (Grant No. 31830072), the National Key Research and Development Program of China (Grant No. 2016YFD0100601), the National Transgenic Major Program (Grant No. 2016ZX08001-002), Chinese Scholarship Council (Grant No. 2017GXZ018098), and Shiraz University (Iran). MK benefited from a grant provided by the Iranian Ministry of Science and Technology for a 6-month sabbatical stay at Shanghai Jiao Tong University in China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.674952/full#supplementary-material

Supplementary Figure 1 | Circular map of the chromosomal DNA and plasmids of *Xanthomonas translucens* strains sequenced in this study. The figures were generated using Circos software v0.69. The four genomes in the top row are *X. translucens* pv. *undulosa*, the three genomes in the middle row are *X. translucens* pv. *translucens* and the bottom row shows four plasmids in the strains XtLr8 and XtKm15. In each circular DNA map, from the outermost to inner the circles show nucleotide numbering (circle 1), COG annotation of forward (circle 2) and reverse (circle 3) strands, non-coding RNAs (rRNA red, tRNA blue, sRNA green; circle 4), GC content (circle 5) and GC skew (circle 6).

Supplementary Figure 2 | Average nucleotide identity (ANI) values among the entire set of *Xanthomonas translucens* including the strains sequenced in this study.

Supplementary Figure 3 | Visualization of genomic islands in the chromosomal DNA of the seven *Xanthomonas translucens* strains sequenced in this study as well as the reference strains obtained from the NCBI GenBank using IslandViewer online service.

Supplementary Figure 4 | Venn diagrams showing the shared gene families (orthologous clusters) among different sets of plasmids constructed using the OrthoVenn online service. Each diagram indicates the number of shared gene families between one of the four plasmids identified in this study, i.e., XtKm15_P1, XtKm15_P2, XtLr8_P1 and XtLr8_P2 and the four closest plasmids for each of these plasmids in the NCBI GenBank determined by BLASTn.

Supplementary Figure 5 | Comparison of the type three secretion system (T3SS) gene clusters in plant pathogenic xanthomonads. **(A)** Organization of the 25 T3SS genes from *hpaH* to *XopF* (about 25 kb) in the 14 *X. translucens* strains evaluated in this study was exactly the same. **(B)** Comparison of the arrangement of T3SS in the 14 *X. translucens* dataset against a set of five plant pathogenic xanthomonads showed that the *X. translucens* strains harbor *hrp* cluster flanked by two *hrp* regulatory genes *hrpX* and *hrpG* which is present outside the *hrp* cluster in all the other evaluated xanthomonads. Furthermore, *hrcL* and *hrcD* were found unique to *X. translucens* which were replaced by *hrpB5* and *hrpD5* in other xanthomonads, respectively.

Supplementary Table 1 | Comparison of repeat variable diresidue (RVDs) in the seven *Xanthomonas translucens* strains sequenced in this study as well as the seven reference strains obtained from the NCBI GenBank. TALEs of all 14 *X. translucens* strains were classified into 16 classes where virulence-associated TALEs, i.e., Xtu 4699_TALE6a/8, ICMP 11055_TALE2 and TALE4b and NXtc01_TALE1 are highlighted in yellow, while former TALE names are indicated with slash. Unusual RVDs are shown in red font. Unique RVDs are underlined.

Supplementary Dataset 1 | The sequences of all 65 type III effectors (T3Es) retrieved from public databases and include in this study.

REFERENCES

Akhter, S., Aziz, R. K., and Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity-and composition-based strategies. *Nucleic Acids Res.* 40:e126. doi: 10.1093/nar/gks406

Alikhan, N. F., Petty, N. K., Zakour, N. L. B., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2020). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi: 10.1093/bioinformatics/btz859

Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing Group, Lau, B. Y., Hoad, G., et al. (2017). IslandViewer 4: expanded

prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

Borodovsky, M., and Lomsadze, A. (2014). Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. *Curr. Protoc. Microbiol.* 32, 1E.7.1–1E.7.17.

Bosi, E., Fani, R., and Fondi, M. (2015). "Defining orthologs and pangenome size metrics," in *Bacterial Pangenomics*, eds Alessio M., Marco G., and Marco F. (New York, NY: Humana Press), 191–202. doi: 10.1007/978-1-4939-1720-4_13

Brandolini, A., Volante, A., and Heun, M. (2016). Geographic differentiation of domesticated einkorn wheat and possible Neolithic migration routes. *Heredity* 117, 135–141. doi: 10.1038/hdy.2016.32

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

CABI (2020). *CABI Invasive Species Compendium: Datasheet Report for Xanthomonas translucens pv. translucens: Bacterial Leaf Streak of Barley*. Wallingford: CAB International.

Charif, D., and Lobry, J. R. (2007). "SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis," in *Structural Approaches to Sequence Evolution*, eds Bastolla U., Porto M., Roman H.E., Vendruscolo M (Berlin: Springer), 207–232. doi: 10.1007/978-3-540-35306-5_10

Chen, G., Khojasteh, M., Taheri-Dehkordi, A., Taghavi, S. M., Rahimi, T., and Osdaghi, E. (2020). Complete genome sequencing provides novel insight into the virulence repertories and phylogenetic position of dry beans pathogen *Curtobacterium flaccumfaciens* pv. *flaccumfaciens*. *Phytopathology* 111, 268–280. doi: 10.1094/phyto-06-20-0243-r

Curland, R. D., Gao, L., Bull, C. T., Vinatzer, B., Dill-Macky, R., Von Eck, L., et al. (2018). Genetic diversity and virulence of wheat and barley strains of *Xanthomonas translucens* from the Upper Midwestern United States. *Phytopathology* 108, 443–453. doi: 10.1094/phyto-08-17-0271-r

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147

Denancé, N., Lahaye, T., and Noël, L. D. (2016). Genomics and effectomics of the crop killer Xanthomonas. *Front. Plant Sci.* 7:71. doi: 10.3389/fpls.2016.00071

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23, 205–211.

EPPO (1998). *Xanthomonas translucens. [Distribution map]*. Wallingford: CAB International.

Erkes, A., Reschke, M., Boch, J., and Grau, J. (2017). Evolution of transcription activator-like effectors in *Xanthomonas oryzae*. *Genome Biol. Evol.* 9, 1599–1615. doi: 10.1093/gbe/evx108

Falahi Charkhabi, N., Booher, N. J., Peng, Z., Wang, L., Rahimian, H., Shams-Bakhsh, M., et al. (2017). Complete genome sequencing and targeted mutagenesis reveal virulence contributions of Tal2 and Tal4b of *Xanthomonas translucens* pv. *undulosa* ICMP11055 in bacterial leaf streak of wheat. *Front. Microbiol.* 8:1488. doi: 10.3389/fmicb.2017.01488

Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R. D., Wilson, G. G., et al. (2016). AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. *Sci. Rep.* 6:21077.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Res.* 31, 439–441. doi: 10.1093/nar/gkg006

Hadfield, J., Croucher, N. J., Goater, R. J., Abudahab, K., Aanensen, D. M., and Harris, S. R. (2018). Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* 34, 292–293. doi: 10.1093/bioinformatics/btx610

Heun, M., Schäfer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., et al. (1997). Site of einkorn wheat domestication identified by DNA fingerprinting. *Science* 278, 1312–1314. doi: 10.1126/science.278.5341.1312

Jacques, M.-A., Arlat, M., Boulanger, A., Boureau, T., Carrère, S., Cesbron, S., et al. (2016). Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annu. Rev. Phytopathol.* 54, 163–187. doi: 10.1146/annurev-phyto-080615-100147

Jaenicke, S., Bunk, B., Wibberg, D., Spröer, C., Hersemann, L., Blom, J., et al. (2016). Complete genome sequence of the barley pathogen *Xanthomonas translucens* pv. *translucens* DSM 18974T (ATCC 19319T). *Genome Announc.* 4:e01334-16.

Ji, C., Ji, Z., Liu, B., Cheng, H., Liu, H., Liu, S., et al. (2020). Xa1 allelic R genes activate rice blight resistance suppressed by interfering TAL effectors. *Plant Commun.* 1:100087. doi: 10.1016/j.xplc.2020.100087

Ji, Z., Ji, C., Liu, B., Zou, L., Chen, G., and Yang, B. (2016). Interfering TAL effectors of *Xanthomonas oryzae* neutralize R-gene-mediated plant disease resistance. *Nat. Commun.* 7:13435.

Khojasteh, M., Shah, S. M. A., Haq, F., Xu, X., Taghavi, S. M., Osdaghi, E., et al. (2020). Transcription activator-like effectors diversity in Iranian strains of *Xanthomonas translucens*. *Phytopathology* 110, 758–767. doi: 10.1094/phyto-11-19-0428-r

Khojasteh, M., Taghavi, S. M., Khodaygan, P., Hamzehzarghani, H., Chen, G., Bragard, C., et al. (2019). Molecular typing reveals high genetic diversity of *Xanthomonas translucens* strains infecting small-grain cereals in Iran. *Appl. Environ. Microbiol.* 85:e01518-19.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32, 2103–2110. doi: 10.1093/bioinformatics/btw152

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Niu, X.-N., Wei, Z.-Q., Zou, H.-F., Xie, G.-G., Wu, F., Li, K.-J., et al. (2015). Complete sequence and detailed analysis of the first indigenous plasmid from *Xanthomonas oryzae* pv. *oryzicola*. *BMC Microbiol.* 15:233. doi: 10.1186/s12866-015-0562-x

Osdaghi, E., Rahimi, T., Taghavi, S. M., Ansari, M., Zarei, S., Portier, P., et al. (2020). Comparative genomics and phylogenetic analyses suggest several novel species within the genus *Clavibacter*, including nonpathogenic tomato-associated strains. *Appl. Environ. Microbiol.* 86:e02873-19.

Osdaghi, E., Taghavi, S. M., Hamzehzarghani, H., Fazliarab, A., and Lamichhane, J. R. (2017). Monitoring the occurrence of tomato bacterial spot and range of the causal agent *Xanthomonas perforans* in Iran. *Plant Pathol.* 66, 990–1002. doi: 10.1111/ppa.12642

Osdaghi, E., Taghavi, S. M., Hamzehzarghani, H., and Lamichhane, J. R. (2016). Occurrence and characterization of the bacterial spot pathogen *Xanthomonas euvesicatoria* on Pepper in Iran. *J. Phytopathol.* 164, 722–734. doi: 10.1111/jph.12493

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinformatics* 42, 3.1.1–3.1.8.

Peng, Z., Hu, Y., Xie, J., Potnis, N., Akhunova, A., Jones, J., et al. (2016). Long read and single molecule DNA sequencing simplifies genome assembly and TAL effector gene analysis of *Xanthomonas translucens*. *BMC Genomics* 17:21. doi: 10.1186/s12864-015-2348-9

Peng, Z., Hu, Y., Zhang, J., Huguet-Tapia, J. C., Block, A. K., Park, S., et al. (2019). *Xanthomonas translucens* commandeers the host rate-limiting step in ABA biosynthesis for disease susceptibility. *Proc. Natl. Acad. Sci. U.S.A.* 116, 20938–20946. doi: 10.1073/pnas.1911660116

Pérez-Quintero, A. L., Lamy, L., Gordon, J., Escalon, A., Cunnac, S., Szurek, B., et al. (2015). QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. *Front. Plant Sci.* 6:545. doi: 10.3389/fpls.2015.00545

Pesce, C., Bolot, S., Cunnac, S., Portier, P., Fischer-Le Saux, M., Jacques, M.-A., et al. (2015). High-quality draft genome sequence of the *Xanthomonas translucens* pv. *cerealis* pathotype strain CFBP 2541. *Genome Announc.* 3:e01574-14.

Pesce, C., Jacobs, J. M., Berthelot, E., Perret, M., Vancheva, T., Bragard, C., et al. (2017). Comparative genomics identifies a novel conserved protein, HpaT, in Proteobacterial Type III secretion systems that do not possess the putative Translocon Protein HrpF. *Front. Microbiol.* 8:1177. doi: 10.3389/fmicb.2017.01177

Potnis, N., Timilsina, S., Strayer, A., Shantharaj, D., Barak, J. D., Paret, M. L., et al. (2015). Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Mol. Plant Pathol.* 16, 907–920. doi: 10.1111/mpp.12244

Rambaut, A., and Drummond, A. J. (2012). *FigTree version 1.4*.0. http://tree.bio.ed.ac.uk/software/figtree/

Read, A. C., Rinaldi, F. C., Hutin, M., He, Y.-Q., Triplett, L. R., and Bogdanove, A. J. (2016). Suppression of Xo1-mediated disease resistance in rice by a truncated, non-DNA-binding TAL effector of *Xanthomonas oryzae*. *Front. Plant Sci.* 7:1516. doi: 10.3389/fpls.2016.01516

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi: 10.1093/bioinformatics/btv681

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1

Sapkota, S., Mergoum, M., and Liu, Z. (2020). The translucens group of *Xanthomonas translucens*: complicated and important pathogens causing bacterial leaf streak on cereals. *Mol. Plant Pathol.* 21, 291–302. doi: 10.1111/mpp.12909

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shah, S. M. A., Haq, F., Ma, W., Xu, X., Wang, S., Xu, Z., et al. (2019). Tal1NXtc01 in *Xanthomonas translucens* pv. *cerealis* contributes to virulence in bacterial leaf streak of wheat. *Front. Microbiol.* 10:2040. doi: 10.3389/fmicb.2019.02040

Shane, W., Baumer, J., and Teng, P. (1987). Crop losses caused by *Xanthomonas* streak on spring wheat and barley. *Plant Dis.* 71, 927–930. doi: 10.1094/pd-71-0927

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* 25, 4.10.11–14.10.14.

Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F. F., et al. (2020). *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nat. Rev. Microbiol.* 18, 415–427.

Vaser, R., Šovic', I., Nagarajan, N., and Šikic', M. (2017). Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wang, C., Zhang, X., Fan, Y., Gao, Y., Zhu, Q., Zheng, C., et al. (2015). XA23 is an executor R protein and confers broad-spectrum disease resistance in rice. *Mol. Plant* 8, 290–302. doi: 10.1016/j.molp.2014.10.010

Wichmann, F., Vorhölter, F. J., Hersemann, L., Widmer, F., Blom, J., Niehaus, K., et al. (2013). The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol. Plant Pathol.* 14, 576–588. doi: 10.1111/mpp.12030

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686

Yoon, S. H., Ha, S. M., Lim, J. M., Kwon, S. J., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 02 July 2021
doi: 10.3389/fmicb.2021.679102

Analysis of the Taxonomy and Pathogenic Factors of *Pectobacterium aroidearum* L6 Using Whole-Genome Sequencing and Comparative Genomics

Peidong Xu[1,2†], Huanwei Wang[1†], Chunxiu Qin[1], Zengping Li[1], Chunhua Lin[1], Wenbo Liu[1] and Weiguo Miao[1]**

[1]Key Laboratory of Green Prevention and Control of Tropical Plant Diseases and Pests, Ministry of Education, College of Plant Protection, Hainan University, Haikou, China, [2]School of Life Sciences, Hainan University, Haikou, China

OPEN ACCESS

Edited by:
Ralf Koebnik,
Plant Health Institute, France

Reviewed by:
Mohammad Arif,
University of Hawaii at Manoa, United States
Melanie J. Filiatrault,
United States Department of Agriculture, United States
Ewa Lojkowska,
University of Gdansk, Poland

***Correspondence:**
Wenbo Liu
saucher@hainanu.edu.cn
Weiguo Miao
miao@hainanu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbe and Virus Interactions With Plants, a section of the journal Frontiers in Microbiology

Received: 22 March 2021
Accepted: 07 June 2021
Published: 02 July 2021

Citation:
Xu P, Wang H, Qin C, Li Z, Lin C, Liu W and Miao W (2021) Analysis of the Taxonomy and Pathogenic Factors of Pectobacterium aroidearum L6 Using Whole-Genome Sequencing and Comparative Genomics. Front. Microbiol. 12:679102. doi: 10.3389/fmicb.2021.679102

Soft rot pectobacteria are devastating plant pathogens with a global distribution and a broad host range. *Pectobacterium aroidearum* L6, previously isolated from leaves of *Syngonium podophyllum*, is a pectolytic bacterial pathogen that causes typical soft rot on *S. podophyllum*. There is a shortage for genome data of *P. aroidearum*, which seriously hinders research on classification and pathogenesis of *Pectobacterium*. We present here the complete genome sequence of *P. aroidearum* L6. The L6 strain carries a single 4,995,896-bp chromosome with 53.10% G + C content and harbors 4,306 predicted protein-coding genes. We estimated *in silico* DNA–DNA hybridization and average nucleotide identity values in combination with the whole-genome-based phylogeny from 19 *Pectobacterium* strains including *P. aroidearum* L6. The results showed that L6 and PC1 formed a population distinct from other populations of the *Pectobacterium* genus. Phylogenetic analysis based on 16S rRNA and genome sequences showed a close evolutionary relationship among *Pectobacterium* species. Overall, evolutionary analysis showed that L6 was in the same branch with PC1. In comparison with 18 *Pectobacterium* spp. reference pathogens, strain L6 had 2,712 gene families, among which 1,632 gene families were identified as orthologous to those strains, as well as 1 putative unique gene family. We discovered 478 genes, 10.4% of the total of predicted genes, that were potentially related to pathogenesis using the Virulence Factors of Pathogenic Bacteria database. A total of 25 genes were related to toxins, 35 encoded plant cell-wall degrading enzymes, and 122 were involved in secretion systems. This study provides a foundation for a better understanding of the genomic structure of *P. aroidearum* and particularly offers information for the discovery of potential pathogenic factors and the development of more effective strategies against this pathogen.

Keywords: genome sequence, taxonomy, pathogenic gene, *Pectobacterium aroidearum*, comparative genomics

INTRODUCTION

Soft rot *Pectobacteriaceae* are considered to be one of the top ten most important agricultural phytopathogens (Mansfield et al., 2012). The family *Pectobacteriaceae* consisted of *Pectobacterium* spp. and *Dickeya* spp., formerly characterized as pectinolytic *Erwinia* spp. (Pérombelon, 2002; Charkowski et al., 2012; Adeolu et al., 2016). They cause destructive soft rot on a variety of field crops, fruits, ornamentals, and vegetables, including the staple food crop potato (Toth et al., 2003, 2011; Ma et al., 2007). *Pectobacterium* was first established in 1945 (Waldee, 1945; Ma et al., 2007). However, the classification of *Pectobacterium* spp. is not clear. Hauben et al. (1998) established the genus *Pectobacterium* that included three species and five subspecies: *P. cacticidum, P. chrysanthemi, P. cypripedii, P. carotovorum* subsp. *atrosepticum, P. carotovorum subsp. betavasculorum, P. carotovorum* subsp. *carotovorum, P. carotovorum* subsp. *odoriferum*, and *P. carotovorum* subsp. *wasabiae*. Subsequently, three subspecies of *P. carotovorum* were elevated into the species level, namely *P. atrosepticum, P. odoriferum*, and *P. wasabiae* (Gardan et al., 2003; Duarte et al., 2004; Nykyri et al., 2012). The classification of the genus *Pectobacterium* has been subjected to wide revision over the last decade, and it is likely that some of the genome-sequenced strains have been incorrectly assigned to *P. carotovorum* (Gardan et al., 2003; Khayi et al., 2016; Pritchard et al., 2016; Zhang et al., 2016). For instance, *P. carotovorum* subsp. *carotovorum* tends to serve as a catchall for pectobacteria isolates differing from the specific descriptions of the other pectobacteria taxa, and *P. aroidearum* was classified as a novel species in 2013 (Nabhan et al., 2013). The genome-sequenced strain PC1 (formerly classified as *P. carotovorum* subsp. *carotovorum*) is actually *P. aroidearum* under the new classification (Nabhan et al., 2013).

In the age of genomics, the *Pectobacterium* genus has been subjected to revision based on the development of bioinformatics. Several pectolytic bacterial strains were thought to belong to a novel *Pectobacterium* species after several taxonomic analyses including 16S rRNA gene sequence, DNA–DNA hybridization (DDH), genomics, and comparative genomics. These include *P. actinidiae* (Koh et al., 2012), *P. polaris* (Dees et al., 2017), *P. peruviense* (Waleron et al., 2018), *P. zantedeschiae* (Waleron et al., 2019), *P. punjabense* (Sarfraz et al., 2018, 2020), and *P. aroidearum* (Nabhan et al., 2013). The species of the genus Pectobacterium has 18 of child taxa with a validly published with correct name and four proposed species not yet validated based on The List of Prokaryotic names with Standing in Nomenclature[1] (**Table 1**; Adeolu et al., 2016; Parte et al., 2020). Eleven species had complete genome data, and 7 had not complete assembly based on the National Center for Biotechnology Information (NCBI) genome database[2] (**Table 1**). Only strain PC1 of *P. aroidearum* has its whole genome sequenced. Thus, there is a shortage of whole-genome data for *P. aroidearum*.

Currently, no methods and chemicals are effective in controlling *Pectobacterium* disease or preventing the spread of these pathogens (Zhang et al., 2017). In addition, planting patterns and storage conditions are not applicable for control of the disease (Yaganza et al., 2014). Bacterial strain L6, isolated from *Syngonium podophyllum* soft rot samples in Hainan Province, was recognized as *P. aroidearum* (Xu et al., 2020). There is a shortage for whole-genome data of *P. aroidearum*, which seriously hinders research on classification and pathogenesis of *Pectobacterium*. Genome

[1]https://www.bacterio.net/genus/pectobacterium
[2]https://www.ncbi.nlm.nih.gov/genome/?term=Pectobacterium

TABLE 1 | Nomenclatural status and type strain of genome data of the *Pectobacterium* genus.

Name	Nomenclatural status	Origin	Type strain of genome data
Pectobacterium actinidiae	Validly published	Portier et al. (2019)	KKH3
Pectobacterium aquaticum	Validly published	Pédron et al. (2019)	No complete assembly
Pectobacterium aroidearum	Validly published	Nabhan et al. (2013)	PC1(formerly classified as *P. carotovorum*, not corrected in the NCBI database)
Pectobacterium atrosepticum	Validly published	Gardan et al. (2003)	JG1008, 21A
Pectobacterium betavasculorum	Validly published	Gardan et al. (2003)	No complete assembly
Pectobacterium brasiliense	Validly published	Portier et al. (2019)	SX309, 1,692
Pectobacterium cacticida	Validly published	Hauben et al. (1998)	No complete assembly
Pectobacterium carnegieana	Validly published	Brenner et al. (1973)	No complete assembly
Pectobacterium carotovorum	Validly published	Waldee (1945)	JR1.1, 67
Pectobacterium fontis	Validly published	Oulghazi et al. (2019)	No complete assembly
Pectobacterium odoriferum	Validly published	Portier et al. (2019)	JK2.1, BC S7
Pectobacterium parmentieri	Validly published	Khayi et al. (2016)	HC, RNS 08–42-1A
Pectobacterium parvum	Validly published	Pasanen et al. (2020)	No complete assembly
Pectobacterium polaris	Validly published	Dees et al. (2017)	PZ1, NIBIO1006
Pectobacterium polonicum	Validly published	Waleron et al. (2019)	No complete assembly
Pectobacterium punjabense	Validly published	Sarfraz et al. (2018)	SS95
Pectobacterium versatile	Validly published	Portier et al. (2019)	14A, 3–2
Pectobacterium wasabiae	Validly published	Gardan et al. (2003)	CFBP 3304
Pectobacterium delphinii	Not validly published	Waldee (1945)	
Pectobacterium melonis	Not validly published	Waldee (1945)	
Pectobacterium peruviense	Not validly published	Waleron et al. (2018)	
Pectobacterium zantedeschiae	Not validly published	Waleron et al. (2019)	

comparison revealed that most virulence genes are highly conserved in the *Pectobacterium* strains, especially for the key virulence determinants involved in the biosynthesis of extracellular enzymes and secretion systems (Li et al., 2019). The functional genomics methods are the effective ways to elucidate that this pathogen interacts with plants and causes disease (Toth et al., 2015). In this study, we sequenced the whole genome of *P. aroidearum* L6. Then, we compared it with genome analyses of 18 *Pectobacterium* reference strains. Furthermore, the genome annotation and comparative genomics analysis provided a foundation for a better understanding of the genomic structure of *P. aroidearum* and particularly offered information for the discovery of potential pathogenic factors and the development of more effective strategies against this pathogen.

MATERIALS AND METHODS

Strain and Type Strain Genome Sequences

Strain L6 was previously isolated from *S. podophyllum* soft rot samples in a plant nursery in the Haidian campus of Hainan University, Haikou, Hainan Province, China, in July 2019. Samples were collected from symptomatic *S. podophyllum* for bacterial isolation. Internal fragments containing symptomatic tissues were transferred to 1 mL of sterile distilled water, after 20 min, rinsed with 70% alcohol and sterile distilled water, and cultured onto Luria-Bertani (LB) medium for 48 h at 28°C to differentiate and characterize the bacterial pathogen. A total of 10 bacterial colonies were isolated from infected tissues. The isolated colonies were subcultured until the pure culture of the suspected bacterium was obtained. Two representative isolates (L5 and L6) were selected for further tests and one isolate preserved in Key Laboratory of Green Prevention and Control of Tropical Plant Diseases and Pests (Hainan University), Ministry of Education as *P. aroidearum* L6.

The pathogenicity of *P. aroidearum* L6 was previously reported (Xu et al., 2020). The putative pathogen was re-inoculated to confirm its pathogenicity in the incubator and field on the leaves of *S. podophyllum*. Typical symptoms of soft rot were observed 12–24 h after inoculation. *P. aroidearum* was re-isolated from a diseased leaf, fulfilling Koch's postulates. L6 was grown on LB broth for 12–24 h at 28°C. And genomic DNA of L6 was extracted by the Bacteria Genomic DNA Extraction Kit (Tiangen Biotech Co. Ltd., Beijing, China). Successful extraction of genomic DNA was confirmed by 0.8% agarose gels and quantified by Nanodrop ND-2000 (Thermo Fisher Scientific, United States). All complete genome sequences of *Pectobacterium* were retrieved from NCBI. Type strain genome sequences of 11 species which had complete assembly genome data were used, including *P. aroidearum* PC1 (GCF_000023605.1), *P. carotovorum* subsp. *carotovorum* JR1.1 (NZ_CP034237.1), *P. carotovorum* subsp. *carotovorum* 67 (NZ_CP034211.1), *P. atrosepticum* JG10-08 (NZ_CP007744.1), *P. atrosepticum* 21A (NZ_CP009125.1), *P. brasiliense* SX309 (NZ_CP020350.1), *P. brasiliense* 1,692 (NZ_CP047495.1), *P. odoriferum* JK2.1 (NZ_CP034938.1), *P. odoriferum* BC S7 (NZ_CP009678.1), *P. polaris* PZ1 (NZ_CP046377.1), *P. polaris* NIBIO1006 (NZ_CP017481.1), *P. actinidiae* KKH3 (NZ_JRMH01000001.1), *P. wasabiae* CFBP 3304 (NZ_CP015750.1), *P. parmentieri* HC (NZ_CP046376.1), *P. parmentieri* RNS 08–42-1A (NZ_CP015749.1), *P. versatile* 14A (NZ_CP034276.1), *P. versatile* 3–2 (NZ_CP024842.1), and *P. punjabense* SS95 (NZ_CP03 8498.1).

Genome Sequencing and Assembly

The *P. aroidearum* L6 genome was sequenced using a PacBio RS II platform and Illumina HiSeq 4,000 platform. For Illumina HiSeq sequencing, the fragments of 470 bp (with the approximate insert size of 350 bp) from adaptor-ligated DNA were recovered according to standard protocols. The libraries with different indices were multiplexed and loaded on an Illumina HiSeq instrument. Cutadapt (v1.9.1; Martin, 2011) was employed for quality control, and reads whose base groups have quality value below 20 at both ends, sequences containing more than 10% N base, or less than 75 bp in length were removed. The Illumina data were used for estimate and correction. Four SMRT cells zero-mode waveguide (a nano-optical device used to confine light to a small observation volume) arrays of sequencing were used in the PacBio platform to generate the subreads set. PacBio subreads (length < 1 kb) were removed. The program pbdagcon[3] was used for self-correction. Draft genomic unitigs, which are uncontested groups of fragments, were assembled using the Celera Assembler (Myers et al., 2000) against a high-quality corrected circular consensus sequence subreads set. To improve accuracy of the genome sequences, the Genome Analysis Toolkit[4] (Mckenna et al., 2010) and SOAP tool packages (SOAP2, SOAPsnp, SOAPindel; Li et al., 2013) were used to make single-base corrections. To trace the presence of any plasmid, the filtered Illumina reads were mapped using SOAP to the bacterial plasmid database.[5]

Genome Component Prediction

Gene prediction was performed on the L6 genome assembly using NCBI Prokaryotic Genome Annotation Pipeline[6] (Tatusova et al., 2016; Haft et al., 2018; Li et al., 2021). The tRNA, rRNA, and sRNA recognition made use of tRNAscan-SE (Lowe and Eddy, 1997), RNAmmer (Lagesen et al., 2007), and the Rfam database (Gardner et al., 2009). Tandem repeats annotation was obtained using the Tandem Repeat Finder[7] (Wirawan et al., 2010), and the minisatellite DNA and microsatellite DNA were selected based on the number and length of repeat units. The Genomic Island Suite of Tools GIST v1.0 was used for genomic island analysis with Island Path-DIOMB, SIGI-HMM, and Island Picker method (Hasan et al., 2012). Prophage regions were predicted using the PHAge Search Tool web server[8] (Zhou et al., 2011) and CRISPR identification using CRISPR Finder (Grissa et al., 2007).

[3]https://github.com/PacificBiosciences/pbdagcon
[4]https://www.broadinstitute.org/gatk/
[5]http://www.ebi.ac.uk/genomes/plasmid.html; last accessed July 8, 2016
[6]https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
[7]http://tandem.bu.edu/trf/trf.html
[8]http://phast.wishartlab.com/

Phylogenetic Analysis

The maximum-likelihood (ML) phylogenetic analysis was inferred with FastME 2.1.6.1 including SPR postprocessing (Lefort et al., 2015) from the Genome BLAST Distance Phylogeny approach (GBDP) distances calculated using default settings from 16S rRNA gene sequences and genome sequences. Branch support was inferred from 1,000 pseudo-bootstrap replicates each. The trees were rooted at the midpoint (Farris, 1972) and visualized with PhyD3 (Kreft et al., 2017).

Average Nucleotide Identity and *in silico* DNA–DNA Hybridization Analysis

Pairwise comparison of their Average Nucleotide Identity (ANI) was based on BLAST+ (ANIb) from JSpeciesWS[9] (Richter et al., 2015), while *in silico* DNA–DNA Hybridization (*is*DDH) was conducted using GBDP by the Type (Strain) Genome Server[10] (Meier-Kolthoff and Göker, 2019). One-hundred distance replicates were calculated each. The DDH values and confidence intervals were calculated using the recommended settings of the Genome-to-Genome Distance Calculator (GGDC 2.1) for GGDC formula 2 (Meier-Kolthoff et al., 2013).

Comparative Genomics

Pectobacterium aroidearum L6 was compared with reference 18 *Pectobacterium* strains by using their genome sequence and gene sequence. BLAST core/pan genes of the six strains were clustered using CD-HIT (Fu et al., 2012) rapid clustering of similar proteins software with a threshold of 50% pairwise identity and 0.7 length difference cutoff in amino acids. Gene families were constructed using genes of L6 and reference strains. We carried out gene family TreeFam clustering treatment for the alignment results by Hcluster_sg software (Maqbool and Babri, 2007) and multiple sequence alignment with the clustered gene family using Muscle software (Edgar, 2004a,b). The phylogenetic tree was constructed using multiple sequence alignment results based on Muscle by the TreeBeST (Nandi et al., 2010) using the ML method.

Gene Annotation and Protein Classification

The function annotation is accomplished by analysis of protein sequences. We align genes with databases to obtain the highest quality corresponding annotations. Seven databases were used for general function annotation: Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa et al., 2016), Clusters of Orthologous Groups (COG; Galperin et al., 2015; Makarova et al., 2015), Non-Redundant Protein Database (NR), Swiss-Prot (UniProt Consortium, 2015), Gene Ontology (GO; Ashburner et al., 2000; Philip et al., 2014), TrEMBL (Apweiler et al., 2004), and Evolutionary Genealogy of Genes (EggNOG: Non-supervised Orthologous Groups; Huerta-Cepas et al., 2016) databases. Four databases were used for pathogenicity and drug resistance analysis. Virulence factors and resistance genes were identified based on the core dataset in Virulence Factors of Pathogenic Bacteria (VFDB; Chen et al., 2016) and the Antibiotic Resistance Genes Database (Liu and Pop, 2009); the other two databases were Pathogen Host Interactions (PHI; Winnenburg et al., 2006) and Carbohydrate-Active enZYmes Database (CAZy; Levasseur et al., 2013). Type III secretion system (T3SS) effector proteins were detected using EffectiveT3 (Vargas et al., 2012).

[9]http://jspecies.ribohost.com/jspeciesws/
[10]http://tygs.dsmz.de/

RESULTS

Genomic Features Among *P. aroidearum* L6 and Reference Strains

The *P. aroidearum* L6 genome was sequenced using a PacBio RS II platform and Illumina HiSeq 4,000 platform. The raw data were filtered, and this generated 1, 274 Mb of clean data. The reads were assembled into a genome and then assessed. We obtained a genome size of 4,995,896 bp, which consisted of a circular chromosome with no plasmid (**Figure 1**). GC-Depth analysis was performed on the assembly results to show the assembly length (4,995,896 bp), coverage length (4,995,089), coverage (99.98%), reads usage percent (95.3%), and depth (260) of *P. aroidearum* L6. The average G + C content of the genome was 53.10%. A total of 4,306 putative protein-coding sequences, with total length of 4,298,622 bp (86.04% of total genome length; **Table 2**) and average length of 998.29 bp, were annotated on the *P. aroidearum* L6 genome. The genome encoded 77 tRNA operons and 40 sRNA genes. In addition, a total of 22 rRNA operons were present on the chromosome: eight 5S rRNAs, seven 16S rRNAs, and seven 23S rRNAs.

Phylogenetic Analysis of *P. aroidearum* L6

The 16S rRNA is the most useful and is commonly used as a molecular clock in the systematic classification of bacteria. Its evolution has good clock properties, being highly conserved in structure and function, and can well reflect the differences between different bacteria (Coenye and Vandamme, 2010). Thus, the accessibility of a large quantity of completely sequenced bacterial genomes allows the speedy and reliable determination of intragenomic sequence heterogeneity of 16S rRNA genes. The phylogenetic tree was inferred with FastME 2.1.6.1 from GBDP distances calculated from *Pectobacterium* species 16S rRNA and genome sequences. The branch lengths were scaled in terms of GBDP distance formula d5. The numbers on branches are GBDP pseudo-bootstrap support values >60% from 1,000 replications, with an average branch support of 91.1%. The delta value was 0.315 based on 16S rRNA and 0.156 based on genome sequences. A phylogram based on computing of the 16S rRNA suggested a close relationship between both *P. aroidearum* L6 and PC1 genomes (**Figure 2A**). The phylogenetic relationships among *P. aroidearum* L6 and reference strains were determined based on genome sequence results (**Figure 2B**). The whole-genome-based phylogenetic tree showed that L6 was most closely related to *P. aroidearum* PC1.

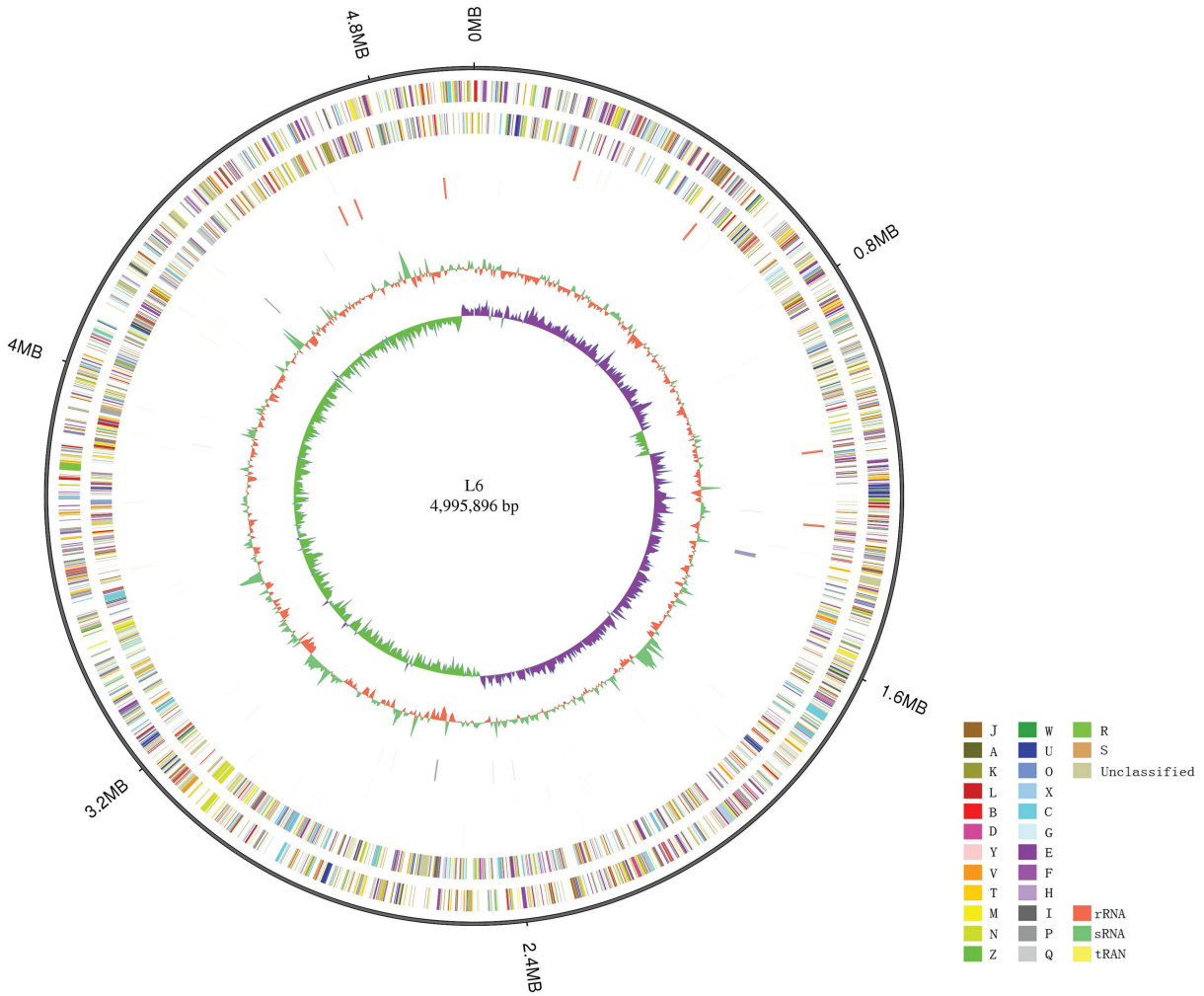


FIGURE 1 | Circular representation of the *Pectobacterium aroidearum* L6 genome. From outer to inner: first circle is genome size; second and third circles are forward and reverse strand gene, respectively, colored according to cluster of COG classification (A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control, cell division, and chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure, and biogenesis; K, transcription; L, replication, recombination, and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover, and chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport, and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; W, extracellular structures; X, mobilome: prophages, transposons; Y, nuclear structure; Z, cytoskeleton); fourth and fifth circles are ncRNA (yellow indicates tRNA, orange indicates rRNA, and green indicates sRNA) and repeat, respectively; seventh circle is GC content (green indicates greater than average value, and orange indicates less than average value); and eighth circle is GC-SKEW (GC-SKEW = (G – C)/(G + C), purple indicates >0, and green indicates <0).

Assessment of Taxonomy of *Pectobacterium* Species Using *is*DDH and ANI

We used *is*DDH and ANI to determine species delineation of *Pectobacterium* spp. Using empirical evidence based on classified species and their comparisons with DDH and ANI values, the same species were set at ≥70% identity in *is*DDH and ≥95% identity in ANI. The *is*DDH and ANI values (**Figure 3**) were consistent with their phylogenetic relationships (**Figure 2**), and members of the same phylogenetic clade also showed high *is*DDH and ANI values. The ranges of isDDH and ANI values for 18 *Pectobacterium* strains were 88–100 and 36–100. The *P. aroidearum* L6 and PC1 were evaluated to species level with *is*DDH = 83 and ANI = 98; furthermore, *P. carotovorum* subsp. *carotovorum* JR1.1 and 67 were not the same species with *is*DDH = 51 and ANI = 93, indicating that these two strains were misidentified.

Analysis of the Core Genome Among *Pectobacterium* Species and Reference Strains

The genomes of 18 *Pectobacterium* strains with *P. aroidearum* L6 were compared. The dispensable gene heatmap showed percentage of dispensable genes among strains. The identity matrix was

TABLE 2 | Genomic features of the *P. aroidearum* L6 genome and comparison with genomes of reference strains.

	Genome size (bp)	G+C content (mol%)	Gene number	Clustered gene number	Number of rRNAs	Number of tRNAs	Family number	Unique family number
P. aroidearum L6	4,995,896	53.1	4,306	4,209	22	77	2,712	1
P. aroidearum PC1	4,862,913	51.9	4,201	4,132	22	78	2,670	1
P. carotovorum subsp. *carotovorum* JR1.1	4,872,902	52.0	4,086	4,019	22	76	2,667	0
P. carotovorum subsp. *carotovorum* 67	4,909,824	51.3	3,532	3,436	22	75	2,361	8
P. atrosepticum JG10-08	5,004,926	51.1	4,245	4,215	22	76	2,805	0
P. atrosepticum 21A	4,991,806	51.1	4,323	4,296	22	77	2,850	0
P. brasiliense SX309	4,966,299	52.2	4,209	4,137	22	76	2,733	1
P. brasiliense 1,692	4,851,982	52.2	4,145	4,044	22	77	2,649	0
P. odoriferum JK2.1	4,997,932	51.5	4,356	4,158	22	77	2,786	4
P. odoriferum BC S7	4,933,575	51.8	3,912	3,830	22	77	2,570	2
P. polaris PZ1	4,994,870	51.0	4,115	3,923	22	77	2,621	8
P. polaris NIBIO1006	4,826,824	52.0	4,088	4,010	22	77	2,645	1
P. actinidiae KKH3	4,068,673	51.5	4,152	4,079	21	76	2,624	1
P. wasabiae CFBP 3304	5,043,228	50.6	4,369	4,203	22	78	2,805	2
P. parmentieri HC	5,208,618	50.4	4,494	4,366	22	77	2,935	1
P. parmentieri RNS 08–42-1A	5,030,841	50.4	4,423	4,343	22	77	2,922	4
P. versatile 14A	4,997,114	51.8	4,304	4,250	22	77	2,778	2
P. versatile 3–2	4,975,878	51.8	4,266	4,191	22	78	2,737	1
P. punjabense SS95	4,793,778	50.7	4,152	4,036	22	76	2,639	0

The genes were taken from reference genome as gene pool. The blast results were filtered by their length and identity. Gene number, the gene number in each strain; Clustered gene number, the gene number that can be clustered in gene family; Family number, the gene family number in strain; Unique family number, the unique gene family number in strain.

calculated based on BLASTP. The strains L6 and PC1 had the highest genetic similarity. Otherwise, *P. punjabense* SS95 also grouped together with *P. aroidearum* L6 and PC1 based on the dispensable gene heat map (**Figure 4A**). Analysis of pan genes among L6 strain and reference strains was carried out. There were 1944 genes shared by all of the bacteria. Among them, 132 genes were unique to L6 and 100 genes were unique to PC1 (**Figure 4B**). Research on special genes and core genes is important for the detection of functional differences and similarities between samples and provides molecular evidence for phenotype differences and similarities. A gene family is a group of genes that have the same ancestor and comprises more than two gene copies. The members of a gene family have similarity in structure and function, and the proteins produced are also similar. Gene families can be used to detect evolutionary history and gene differentiation. The gene family statistics showed that the final core genome was 1,632 gene families. One gene family was unique to L6, and one gene family was unique to PC1 (**Figures 4C,D**).

Analysis of Gene Function Annotation of *P. aroidearum* L6

To further determine the difference in functions encoded by 4,306 genes of *P. aroidearum* L6, we analyzed the data using GO, COG, and KEGG. A total of 3,013 (65.59%) genes could be annotated to one or more of the GO definitions. In our study, 6,763 genes were annotated to biological processes, 2,441 to cellular components, and 3,691 to molecular functions in GO analysis (**Figure 5A**). Most gene functions focused on cellular process (1,640), metabolic process (1,678), single-organism process (1,411), membrane (772), binding (1,231), and catalytic activity (1,601). There were 3,625 (78.92%) predicted genes assigned to COG categories (**Figure 5B**): 43.78% (1,871) of the genes were related to metabolism, 26.11% (1,116) to cellular processes, 17.92% (766) to information, and 12.19% (521) to poorly. A total of 3,081 genes were annotated using the KEGG database (**Figure 5C**). Among the categories, metabolism was the largest group, containing metabolic pathway (690 genes, 22.40%), biosynthesis of secondary metabolites (321 genes, 10.42%), biosynthesis of antibiotics (234 genes, 7.59%), microbial metabolism in diverse environments (229 genes, 7.43%), and others. The cluster of environmental information processing primarily consisted of ABC transporters (279 genes, 9.06%) and two-component system (173 genes, 5.62%).

Pathogenic Candidate Genes Obtained Through Gene Annotation Screening

We predicted genes associated with pathogenicity using GO, COG, KEGG, and especially VFDB, PHI, and T3SS, which are important databases for predicting bacterial pathogenicity. The VFDB database mainly focuses on the infectious agents among bacteria, mycoplasma, and chlamydia. A total of 478 (10.40%) genes were annotated to the VFDB definitions. The PHI database contains relationships between pathogens and hosts, and it predicted 432 (9.40%) genes. The T3SS has close relationship with Gram-negative pathogens and aids in determining infection mechanisms and toxicity at the molecular level. There were 723 (15.74%) predicted genes assigned to T3SS categories. Moreover, we screened the relevant genes

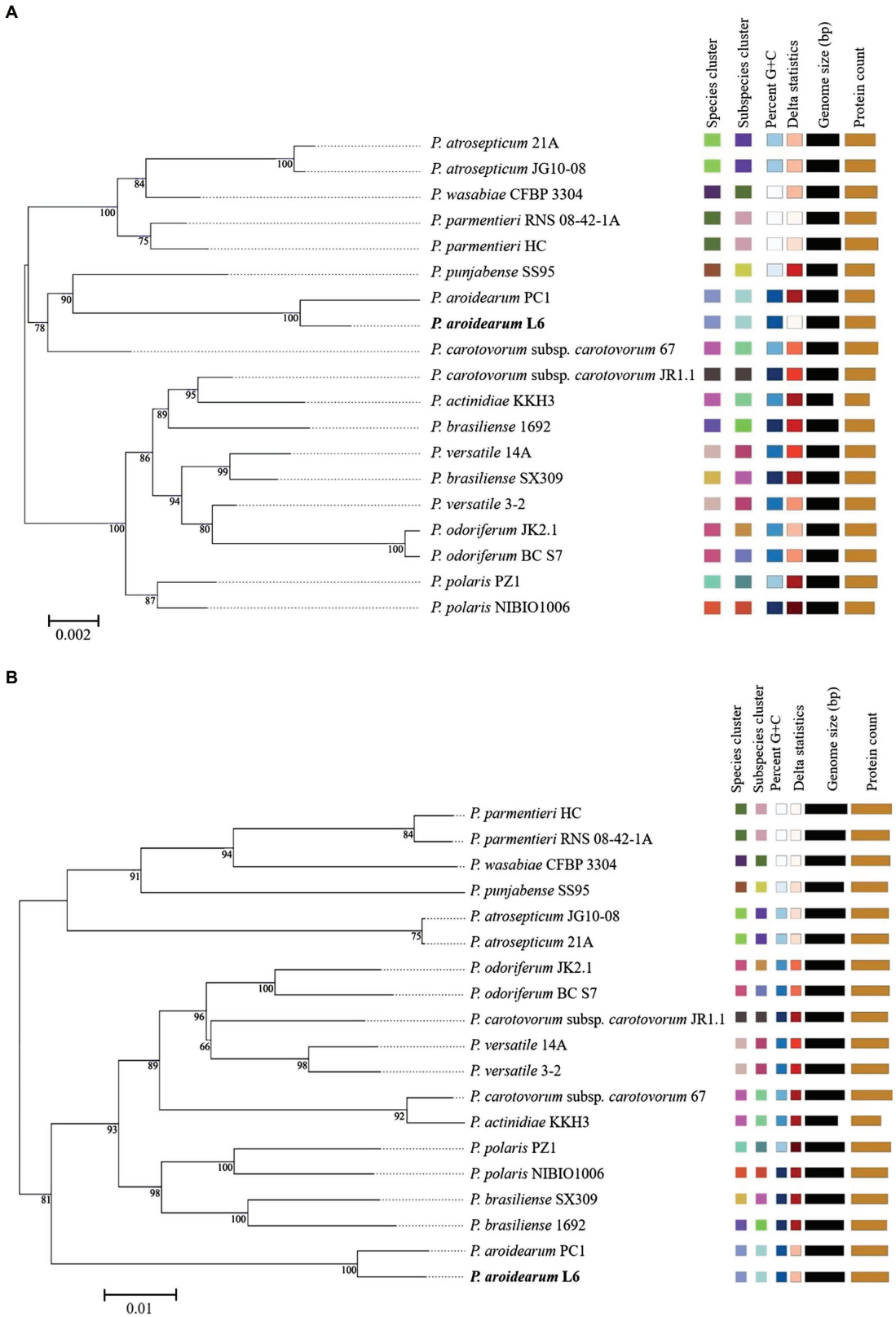
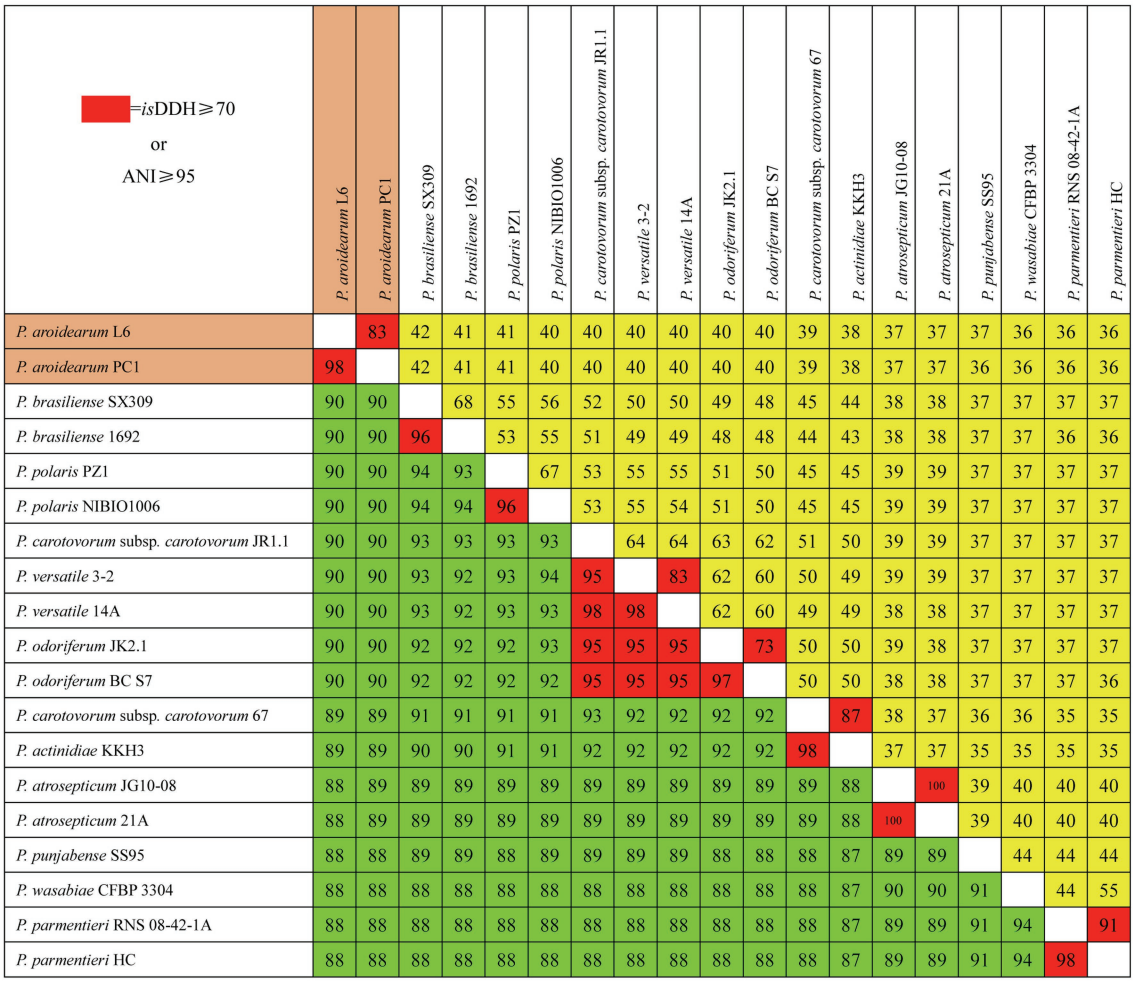


FIGURE 2 | ML phylogenetic analysis for *P. aroidearum* L6 and reference strains. **(A)** Phylogenetic tree drawn using the 16S rRNA of *Pectobacterium* spp. **(B)** Phylogenetic tree drawn using the genomes of *Pectobacterium* spp. Bootstrap values are indicated in % of repetitions.



=*is*DDH≥70 or ANI≥95	*P. aroidearum* L6	*P. aroidearum* PC1	*P. brasiliense* SX309	*P. brasiliense* 1692	*P. polaris* PZ1	*P. polaris* NIBIO1006	*P. carotovorum* subsp. *carotovorum* JR1.1	*P. versatile* 3-2	*P. versatile* 14A	*P. odoriferum* JK2.1	*P. odoriferum* BC S7	*P. carotovorum* subsp. *carotovorum* 67	*P. actinidiae* KKH3	*P. atrosepticum* JG10-08	*P. atrosepticum* 21A	*P. punjabense* SS95	*P. wasabiae* CFBP 3304	*P. parmentieri* RNS 08-42-1A	*P. parmentieri* HC
P. aroidearum L6		83	42	41	41	40	40	40	40	40	40	39	38	37	37	37	36	36	36
P. aroidearum PC1	98		42	41	41	40	40	40	40	40	40	39	38	37	37	36	36	36	36
P. brasiliense SX309	90	90		68	55	56	52	50	50	49	48	45	44	38	38	37	37	37	37
P. brasiliense 1692	90	90	96		53	55	51	49	49	48	48	44	43	38	38	37	37	36	36
P. polaris PZ1	90	90	94	93		67	53	55	55	51	50	45	45	39	39	37	37	37	37
P. polaris NIBIO1006	90	90	94	94	96		53	55	54	51	50	45	45	39	39	37	37	37	37
P. carotovorum subsp. *carotovorum* JR1.1	90	90	93	93	93	93		64	64	63	62	51	50	39	39	37	37	37	37
P. versatile 3-2	90	90	93	92	93	94	95		83	62	60	50	49	39	39	37	37	37	37
P. versatile 14A	90	90	93	92	93	93	98	98		62	60	49	49	38	38	37	37	37	37
P. odoriferum JK2.1	90	90	92	92	92	93	95	95	95		73	50	50	39	38	37	37	37	37
P. odoriferum BC S7	90	90	92	92	92	92	95	95	95	97		50	50	38	38	37	37	37	36
P. carotovorum subsp. *carotovorum* 67	89	89	91	91	91	91	93	92	92	92	92		87	38	37	36	36	35	35
P. actinidiae KKH3	89	89	90	90	91	91	92	92	92	92	92	98		37	37	35	35	35	35
P. atrosepticum JG10-08	88	89	89	89	89	89	89	89	89	89	89	89	88		100	39	40	40	40
P. atrosepticum 21A	88	89	89	89	89	89	89	89	89	89	89	89	88	100		39	40	40	40
P. punjabense SS95	88	88	89	89	88	89	89	89	89	88	88	88	87	89	89		44	44	44
P. wasabiae CFBP 3304	88	88	88	88	88	88	88	88	88	88	88	88	87	90	90	91		44	55
P. parmentieri RNS 08-42-1A	88	88	88	88	88	88	88	88	88	88	88	88	87	89	89	91	94		91
P. parmentieri HC	88	88	88	88	88	88	88	88	88	88	88	88	87	89	89	91	94	98	

FIGURE 3 | Pairwise comparisons of *is*DDH and ANI values of *Pectobacterium* species. The upper triangle (yellow portion) displays *is*DDH values (%), and the lower triangle (green portion) displays ANI values (%). Boxes with *is*DDH ≥ 70% or ANI ≥ 95% are colored red.

encoding plant cell-wall degrading enzymes (PCWDEs), toxins, and secretion systems for *P. aroidearum* L6 (**Table 3**). There were 25 genes related to toxins, and 35 genes encoded PCWDEs, including 28 encoding pectinases, three encoding cellulases, and four encoding proteinases. In addition, 122 genes were involved in six types of secretion systems: 14 genes in Type I (T1SS), 23 in Type II (T2SS), 29 in Type III (T3SS), 25 in Type IV (T4SS), none in Type V (T5SS), and 31 in Type VI (T6SS) secretion systems.

DISCUSSION

Plant bacterial soft rot is one of the destructive diseases of cabbage, tomato, and potato (Meng et al., 2017; Wang et al., 2017; Cui et al., 2019). And it always can cause more serious losses than any other bacterial disease (Qi et al., 2021). Most of the soft rot disease of vegetables is caused by *Pectobacterium* spp., one of the top ten bacterial plant pathogens. *Pectobacterium* spp. has attracted more attention about its wide distribution and diversity (He et al., 2021). *Pectobacterium* usually exists in soils with a broad range of hosts; thus, *Pectobacterium* species cause soft rot disease in plants of at least 16 dicotyledonous and 11 monocotyledonous angiosperm families (Ma et al., 2007; Nabhan et al., 2013). *Pectobacterium aroidearum* was classified as a novel species in 2013 (Nabhan et al., 2013). In previous studies, bacterial soft rot disease was caused by *P. aroidearum* in calla (*Zantedeschia aethiopica*; Nabhan et al., 2013), potato (*Solanum tuberosum*; Moretti et al., 2016), Chinese cabbage (*Brassica rapa*; Xie et al., 2017), zucchini (*Cucurbita pepo*; Moraes et al., 2017), konjac (*Amorphophallus konjac*; Sun et al., 2019), pepper (*Capsicum annuum*; Moraes et al., 2020), and carrot (*Daucus carota*; Tang et al., 2021). In our previous research, *P. aroidearum* as a pathogen on *S. podophyllum* was found in China (Xu et al., 2020). It is important to prevent the spread of this pathogen because many ornamental and edible plant species are susceptible to *P. aroidearum*.

The classification of the genus *Pectobacterium* has been subject to wide revision over the last decade. *Pectobacterium* spp. are highly phenotypically, genetically, and pathogenically

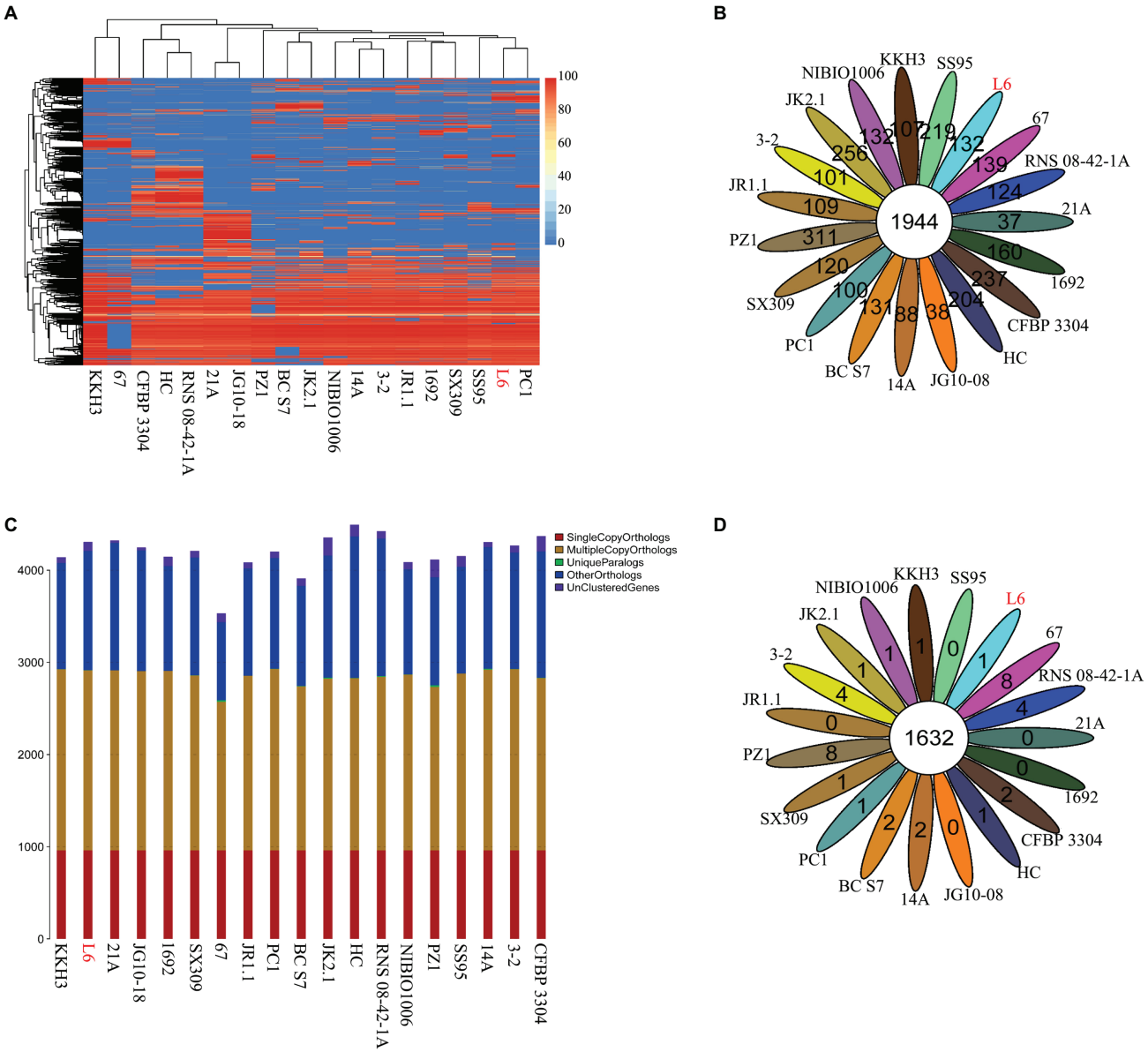


FIGURE 4 | Core genes and dispensable genes and orthologs in the *P. aroidearum* L6 genome and reference strain genomes. **(A)** Dispensable gene heat map (left, dispensable gene cluster; top, strain cluster; red gradient bar represents the scale of similarity percentage); **(B)** Venn diagram of pan genes (each ellipse represents one strain, and the number in the ellipse is the cluster number. One cluster has genes of >50% identity and <0.3 length diversity); **(C)** Ortholog number (Single Copy Orthologs, the number of single-copy homologous genes in the species common gene families; Multiple Copy Orthologs, the number of multiple-copy homologous genes in the species common gene families; Unique Paralogs, genes in specific gene families; Other Orthologs, other genes; and Unclustered Genes, genes that have not been clustered into any families); **(D)** Venn diagram of orthologs in gene family (each ellipse represents one strain, and the number in the ellipse is the family number).

particularly heterogeneous, indicating a need for re-evaluation and a better classification of these species (Dees et al., 2017). Three subspecies of *P. carotovorum* were reclassified as one subspecies (*P. carotovorum* subsp. *carotovorum*), and *P. carotovorum* subsp. *odoriferum* and *P. carotovorum* subsp. *brasiliense* were reclassified as *P. odoriferum* and *P. brasiliense*, respectively, based on genomics (Liu and Filiatrault, 2020; Liu et al., 2020). A lot of genomes of *Pectobacterium* have been sequenced, annotated, and analyzed previously (Oulghazi et al., 2020; Pedersen et al., 2020; Jonkheer et al., 2021). But there was no information about the whole genome of *P. aroidearum*-type strain SCRI 109. At present, only strain PC1 of *P. aroidearum* has sequenced its whole genome (PC1 formerly classified as *P. carotovorum* subsp. *carotovorum*, the classification has not been corrected in the NCBI database; He et al., 2021). To understand the molecular mechanisms of taxonomy and pathogenic factors in *P. aroidearum*, the whole genome of L6 was successively sequenced in this study.

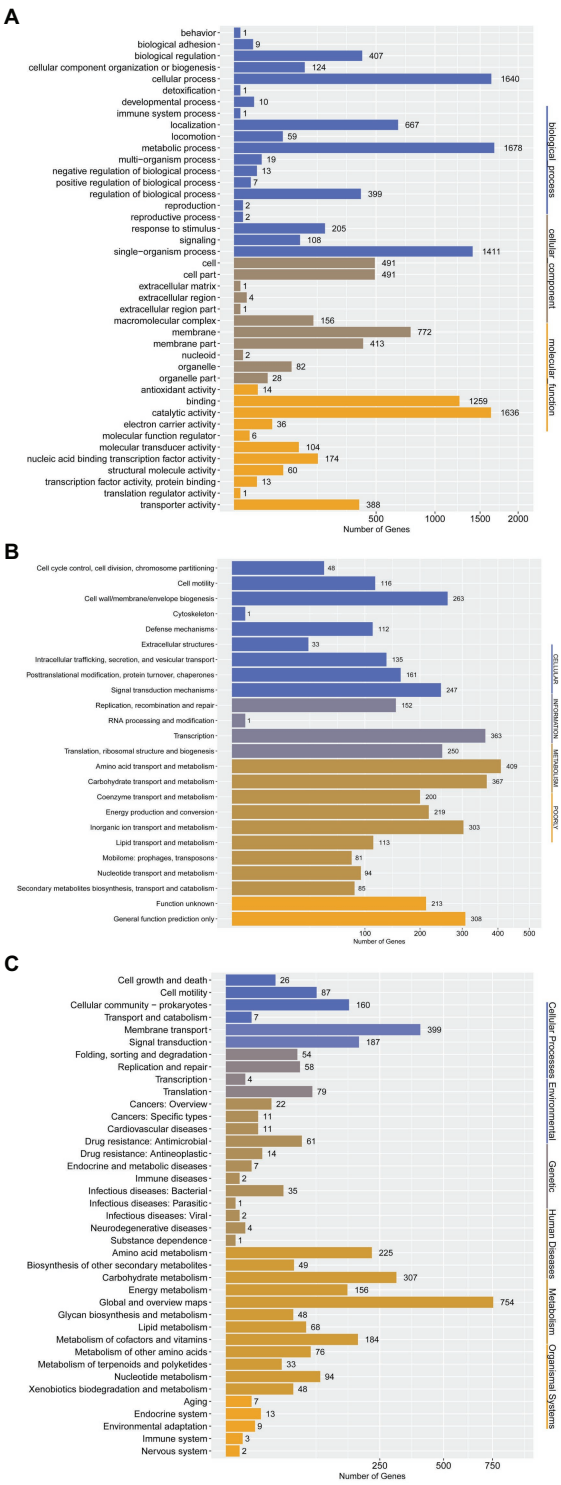


FIGURE 5 | Gene annotation by GO, COG, and KEGG for *P. aroidearum* L6. **(A)** GO function classification of genes in L6. GO analysis was performed for three main categories: cellular components, molecular function, and biological processes. **(B)** COG function classification of genes in L6, grouped into four main parts: metabolism, cellular processes, information, and poorly. **(C)** The KEGG pathway classification of genes in L6 contains six groups: cellular processes, environmental, genetic, human diseases, metabolism, and organismal systems.

And it will be the first public report on the genome of *P. roidearum*.

Comparison of genomes is an efficient method for classification and detection of bacterial and fungal pathogens (Pritchard et al., 2012; Bühlmann et al., 2013; Malapi-Wight et al., 2016; Tang et al., 2017; Van Dam et al., 2018). Zoledowska et al. (2018) researched the comparative genomic of 15 *P. parmentieri* strains and found the high genomic variation among *P. parmentieri* strains. Pédron et al. (2019) established the taxonomic status of six *Pectobacterium* strains based on phylogenetic data, ANI values, and isDDH results by comparative genomics and identified a novel species of the genus *Pectobacterium* named *Pectobacterium aquaticum*. He et al. (2021) analyzed comparative genomics of four *Pectobacterium* strains and obtained three kinds of highly conserved key pathogenic genes related to cell-wall degrading enzymes in *Pectobacterium* strain PC1, including 19 pectinase genes, 25 cellulase genes, and 22 protease genes. Zhang et al. (2016) compared 85 genomes of the genera *Dickeya* and *Pectobacterium* and found that at least ten tested genomes from these genera were misnamed in GenBank based on ANI, *is*DDH, and whole genome. In our study, L6 and PC1 were grouped in one population distinct from other populations of the *Pectobacterium* genus and we also found some strains were misnamed in GenBank. (*P. carotovorum* subsp. *carotovorum* JR1.1 and 67 were not the same species.) It is effective for re-evaluating current prokaryotic species definition and establishing a unified prokaryotic species definition frame by using whole-genome sequences for taxonomically challenging genera (Zhang et al., 2016; Qi et al., 2021).

Currently, once plants are infected, there is no effective method to control bacterial soft rot (Sun et al., 2019). It is also very possible to develop new control methods. By screening pathogenic genes based on whole-genome sequences of *Pectobacterium* species and analyzing the pathogenic mechanism at the molecular level, Zhang et al. (2017) discovered a total of 168 genes related to pathogenesis including nine specific genes encoding toxins on the genome of *P. atroseptica* JG10-08. Huang et al. (2019) selected five putative effectors from the genome of *P. carotovorum* subsp. *brasiliense* BZA12 and discovered that candidate effector A12GL002483 was localized in the cell nucleus and induced cell death. We discovered 478 genes, 10.4% of total predicted genes, that were potentially related to pathogenesis according to the VFDB database. Previous research has shown that soft rot pathogenesis basically relies on toxins, PCWDEs, and the secretion system. Toxins play a key role in the pathogenicity of *Pectobacterium* species. We discovered 29 genes related to toxins in *P. aroidearum* L6. Moreover, PCWDEs are crucial in three distinct pathogenic functions: degradation, nutrition, and feedback regulation (Franza et al., 2002; Yang et al., 2007). The pathogens benefit from the nutrients produced after degradation; these degradation products accumulate in the host and can induce bacterium to generate more enzymes (Zhang et al., 2017). Therefore, the production of PCWDEs is characteristic of infection by *Pectobacterium* species. The PCWDEs consist of pectinases, cellulases, and proteinases. In this study, we identified a total

TABLE 3 | Pathogenic candidate genes of *P. aroidearum* L6 identified through gene annotation screening.

Type		Number	Gene
Toxins		25	*higB21*, *relE*, *rhaS*, *symE*, *cvpA*, *prtC*, *y4kP*, *higB22*, *hlyC*, *rtxC*, *ortT*, *cbtA*, *abiEii*, *parE1*, *ccdB*, *aebG*, *pasT*, *cptA*, *ccdB*, *yoeB*, *parE3*, *higB23*, *tabA*, *pinD*, *stbE*
PCWDEs	Pectinases	28	*pel1*, *pel2*, *pel3*, *pelB*, *pel4*, *pel5*, *pelL*, *pelW*, *pelP*, *ply1*, *pelD*, *ply2*, *pelX*, *fhaB*, *pehK*, *paxE*, *pehX*, *ogl*, *paaE*, *kduI*, *kdgF*, *plpb*, *fhaB1*, *fhaB2*, *ppbH*, *pglR1*, *ssp., pmeB*
	Cellulases	3	*bcsZ*, *celB*, *exlX*
	Proteinase	4	*btlcP*, *nprE*, *pi38*, *ps53*
Secretion systems	T1SS	14	*lssD*, *lssB*, *lapE*, *cttD*, *hasD*, *hasE*, *hasF*, *aprD*, *prtE*, *prtF*, *tolC*, *lapB*, *lassD*, *mdsABC*
	T2SS	23	*hofQ*, *hofC*, *hofB*, *ppdD*, *pilT*, *gspB*, *gspC*, *gspD*, *gspE*, *gspF*, *gspG*, *gspH*, *gspI*, *gspJ*, *gspK*, *gspL*, *gspM*, *gspN*, *gspO*, *tadC*, *tadB*, *cpaF*, *cpaC*
	T3SS	29	*hrtA*, *ycgR*, *hrtB*, *hrpT*, *hrcC*, *hrtC*, *hrpF*, *fliH*, *hrtD*, *hrcJ*, *hrpB*, *hrpJ*, *hrcV*, *hrpQ*, *hrcN*, *hrtE*, *hrtF*, *hrcQ*, *hrcR*, *hrcS*, *hrcT*, *hrcU*, *fliH*, *fliI*, *fliN*, *flip*, *fliR*, *flhA*, *flhB*
	T4SS	25	*rhsA*, *traC*, *rsmE*, *pilS*, *virB11*, *virB10*, *virB9*, *virB8*, *virB6*, *virB5*, *virB4*, *virB2*, *virB1*, *rhsB*, *pFL4*, *lysM*, *triB*, *trbD*, *trbE*, *trbL*, *ntf2*, *trbG*, *trbI*, *trbK*, *yjgA*
	T5SS	0	
	T6SS	31	*rhsGE*, *hcpA*, *aec32*, *hcpB*, *vgrGA*, *hcpC*, *vgrGB*, *paar1*, *hcpD*, *impB*, *impC*, *iraD*, *vasA*, *vasB*, *vasC*, *vasD*, *vase*, *vasF*, *vasG*, *vasI*, *vasJ*, *vasK*, *vasL*, *hcp1*, *vgrGC*, *paar2*, *hcpE*, *hcpF*, *vgrGD*, *hcpG*, *hcpH*

of 36 genes encoding PCWDEs: 29 encoding pectinases, three encoding cellulases, and four encoding proteinases.

In addition to toxins and PCWDEs, secretion systems play a critical role in plant bacterial disease development. There are six types of secretion systems to export extracellular enzymes and effector proteins in bacteria (Lory, 1998). Through secretion systems, effectors can be transported inside the plant cell and promote rapid infection of the host plant (Holst et al., 1996). Among them, the Hrp gene cluster encoding the T3SS is particular important in many Gram-negative pathogens; this is a multi-protein complex bacterial structure to deliver virulence effector proteins directly into plant cells (Tampakaki et al., 2004; Tam et al., 2007; Xie et al., 2019). In our study, there were 723 predicted genes assigned to T3SS categories. Furthermore, 122 genes were involved in the six types of secretion system in *P. aroidearum* L6 based on GO, COG, and KEGG: 14 in T1SS, 23 in T2SS, 29 in T3SS, 25 in T4SS, and 31 in T6SS. Pasanen et al. (2020) identified a novel species *Pectobacterium parvum* and found it contained SPI-1-type Type III secretion island by comparing between the genomes of *Pectobacterium* species. In the genome of *Pseudomonas syringae*, Kang et al. (2014) researched the role of T3SS effectors in the disruption of actin cytoskeleton and inhibition of endocytosis. In the genome of *Shewanella* sp., Alex and Antunes (2019) detected the genes encoding for T3SS core components and four copies of homologs of T3SS effector. Currently, the effectors of *P. aroidearum* pathogenesis have not been studied. All these genes in *P. aroidearum* L6 have potential virulence functions. Thus, further research on the pathogenic factors in L6 may reveal the mechanism of *Pectobacterium* species infection of plants.

CONCLUSION

The classification of the genus *Pectobacterium* has long been unclear. *Pectobacterium* spp. are highly phenotypically, genetically, and pathogenically particularly heterogeneous, can cause severe soft rot in plant hosts, and have a wide host range. Our results suggest that *P. aroidearum* L6 synthesizes and transports virulence factors. Moreover, 182 genes were involved in toxins, PCWDEs, and the secretion system. The results of this research will serve as a foundation for a better understanding of the genomic structure of *P. aroidearum*. The discovery of potential pathogenic factors can help in preventing spread and outbreak of this pathogen and providing effective biological measures against it.

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed in the current study are available from the corresponding author on reasonable request. The genome sequence of *P. aroidearum* L6, including all assemblies and annotations, generated for this study is available at NCBI GenBank with accession number CP065044. The 16S ribosomal RNA sequence is available with accession code MT120309.

AUTHOR CONTRIBUTIONS

WL and PX designed the experiments and wrote the manuscript. PX, HW, CQ, and ZL performed collection and bioinformatics analysis. CL and WM revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was financially supported by the Hainan Major Research Fund of Science and Technology (grant no. ZDKJ201817), the National Natural Science Foundation of China (grant nos. 31560495, 31760499, and 31660033), the National Natural Rubber Industry Technical System (grant no. CARS-34-GW8), the Opening Project Fund of Key Laboratory of Biology and Genetic Resources of Rubber Tree, Ministry of Agriculture and Rural Affairs, P. R. China (grant no. RRI-KLOF201903), and the Hainan Province Graduate Student Innovation Research Project (grant nos. Hyb2018-07 and Hyb2019-02).

REFERENCES

Adeolu, M., Alnajar, S., Naushad, S., and Gupta, R. S. (2016). Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae, Erwiniaceaefam*. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* 66, 5575–5599. doi: 10.1099/ijsem.0.001485

Alex, A., and Antunes, A. (2019). Whole-genome comparisons among the genus *Shewanella* reveal the enrichment of genes encoding ankyrin-repeats containing proteins in sponge-associated bacteria. *Front. Microbiol.* 10:5. doi: 10.3389/fmicb.2019.00005

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 32, D115–D119. doi: 10.1093/nar/gkh131

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Brenner, D. J., Steigerwalt, A. G., Miklos, G. V., and Fanning, G. R. (1973). Deoxyribonucleic acid relatedness among Erwiniae and other *Enterobacteriaceae*: the soft-rot organisms (genus *Pectobacterium* Waldee). *Int. J. Syst. Bacteriol.* 23, 205–216. doi: 10.1099/00207713-23-3-205

Bühlmann, A., Pothier, J. F., Tomlinson, J. A., Frey, J. E., Boonham, N., Smits, T. H. M., et al. (2013). Genomics-informed design of loop-mediated isothermal amplification for detection of phytopathogenic *Xanthomonas arboricola* pv. *pruni* at the intraspecific level. *Plant Pathol.* 62, 475–484. doi: 10.1111/j.1365-3059.2012.02654.x

Charkowski, A., Blanco, C., Condemine, G., Expert, D., Franza, T., Hayes, C., et al. (2012). The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. *Annu. Rev. Phytopathol.* 50, 425–449. doi: 10.1146/annurev-phyto-081211-173013

Chen, L. H., Zheng, D. D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239

Coenye, T., and Vandamme, P. (2010). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol. Lett.* 228, 45–49. doi: 10.1016/S0378-1097(03)00717-1

Cui, W., He, P., Munir, S., He, P., He, Y., Li, X., et al. (2019). Biocontrol of soft rot of chinese cabbage using an endophytic bacterial strain. *Front. Microbiol.* 10:1471. doi: 10.3389/fmicb.2019.01471

Dees, M. W., Lysøe, E., Rossmann, S., Perminow, J., and Brurberg, M. B. (2017). *Pectobacterium polaris* sp. nov., isolated from potato (*Solanum tuberosum*). *Int. J. Syst. Evol. Microbiol.* 67, 5222–5229. doi: 10.1099/ijsem.0.002448

Duarte, V., de Boer, S. H., Ward, L. J., and de Oliveira, A. M. R. (2004). Characterization of atypical *Erwinia carotovora* strains causing blackleg of potato in Brazil. *J. Appl. Microbiol.* 96, 535–545. doi: 10.1111/j.1365-2672.2004.02173.x

Edgar, R. C. (2004a). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Edgar, R. C. (2004b). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf.* 5:113. doi: 10.1186/1471-2105-5-113

Farris, J. S. (1972). Estimating phylogenetic trees from distance matrices. *Am. Nat.* 106, 645–668. doi: 10.2307/2459725

Franza, T., Michaud-Soret, I., Piquerel, P., and Expert, D. (2002). Coupling of iron assimilation and pectinolysis in *Erwinia chrysanthemi* 3937. *Mol. Plant-Microbe Interact.* 15, 1181–1191. doi: 10.1094/MPMI.2002.15.11.1181

Fu, L., Niu, B., Zhu, Z., Zhu, Z. W., Wu, S., and Li, W. Z. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Eugene, V. K. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. doi: 10.1093/nar/gku1223

Gardan, L., Gouy, C., Christen, R., and Samson, R. (2003). Elevation of three subspecies of *Pectobacterium carotovorum* to species level: *Pectobacterium atrosepticum* sp. nov., *Pectobacterium betavasculorum* sp. nov. and *Pectobacterium wasabiae* sp. nov. *Int. J. Syst. Evol. Microbiol.* 53, 381–391. doi: 10.1099/ijs.0.02423-0

Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., et al. (2009). Rfam: updates to the RNA families database. *Nucleic Acids Res.* 37, D136–D140. doi: 10.1093/nar/gkn766

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360

Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., et al. (2018). RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46, D851–D860. doi: 10.1093/nar/gkx1068

Hasan, M. S., Liu, Q., Wang, H., Fazekas, J., Chen, B., and Che, D. S. (2012). GIST: genomic island suite of tools for predicting genomic islands in genomic sequences. *Bioinformation* 8, 203–205. doi: 10.6026/97320630008203

Hauben, L., Moore, E. R., Vauterin, L., Steenackers, M., Mergaert, J., Verdonck, L., et al. (1998). Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. *Syst. Appl. Microbiol.* 21, 384–397. doi: 10.1016/S0723-2020(98)80048-9

He, X., Lu, T., and Zhou, X. (2021). Whole genome sequencing and comparative genomics analysis of *Pectobacterium carotovorum* identifies key pathogenic genes, molecular phylogenetics and evolution. *Mol. Phylogenet. Evol.* 17:107114. doi: 10.1016/j.ympev.2021.107114

Holst, O., Ulmer, A. J., Brade, H., Flad, H. D., and Rietschel, E. T. (1996). Biochemistry and cell biology of bacterial endotoxins. *FEMS Immunol. Med. Microbiol.* 16, 83–104. doi: 10.1111/j.1574-695X.1996.tb00126.x

Huang, Y. F., Liu, C. Y., Wang, H., Guan, T. S., Liu, L., and Yu, S. Y. (2019). Bioinformatic analysis of the complete genome sequence of *Pectobacterium carotovorum* subsp. *brasiliense* BZA12 and candidate effector screening. *J. Plant Path.* 101, 39–49. doi: 10.1007/s42161-018-0126-7

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* D1, D286–D293. doi: 10.1093/nar/gkv1248

Jonkheer, E. M., Balázs Br ankovics, Houwers, I. M., Wolf, J. M. V. D., and Lee, T. A. J. V. D. (2021). The *pectobacterium* pangenome, with a focus on *pectobacterium brasiliense*, shows a robust core and extensive exchange of genes from a shared gene pool. *BMC Genomics* 22:265. doi: 10.1186/s12864-021-07583-5

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kang, Y., Jelenska, J., Cecchini, N. M., Li, Y., Lee, M. W., Kovar, D. R., et al. (2014). HopW1 from *Pseudomonas syringae* disrupts the actin cytoskeleton to promote virulence in *Arabidopsis*. *PLoS Pathog.* 10:e1004232. doi: 10.1371/journal.ppat.1004232

Khayi, S., Cigna, J., Chong, T. M., Quêtu-Laurent, A., Chan, K., Hélias, V., et al. (2016). Transfer of the potato plant isolates of *Pectobacterium wasabiae* to *Pectobacterium parmentieri* sp. nov. *Int. J. Syst. Evol. Microbiol.* 66, 5379–5383. doi: 10.1099/ijsem.0.001524

Koh, Y. J., Kim, G. H., Lee, Y. S., Sohn, S. H., Koh, H. S., Kwon, S., et al. (2012). *Pectobacterium carotovorum* subsp. *actinidiae* subsp. nov. a new bacterial pathogen causing canker-like symptoms in yellow kiwifruit, *Actinidia chinensis*. *N. Z. J. Crop Hortic. Sci.* 40, 269–279. doi: 10.1080/01140671.2012.707129

Kreft, L., Botzki, A., Coppens, F., Vandepoele, K., and Van, B. M. (2017). PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics* 33, 2946–2947. doi: 10.1093/bioinformatics/btx324

Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* 32, 2798–2800. doi: 10.1093/molbev/msv150

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels.* 6:41. doi: 10.1186/1754-6834-6-41

Li, S. T., Li, R. Q., Li, H., Lu, J. L., Li, Y. R., Bolund, L., et al. (2013). SOAPindel: efficient identification of indels from short paired reads. *Genome Res.* 23, 195–200. doi: 10.1101/gr.132480.111

Li, W. J., O'Neill, K. R., Haft, D. H., DiCuccio, M., Chetvernin, V., Badretdin, A., et al. (2021). RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res.* 49, D1020–D1028. doi: 10.1093/nar/gkaa1105

Li, L., Yuan, L. F., Shi, Y. X., Xie, X. W., Chai, A., Wang, Q., et al. (2019). Comparative genomic analysis of *Pectobacterium carotovorum* subsp. *brasiliense* SX309 provides novel insights into its genetic and phenotypic features. *BMC Genomics* 20:486. doi: 10.1186/s12864-019-5831-x

Liu, Y. Y., and Filiatrault, M. J. (2020). Complete genome sequence of the necrotrophic plant-pathogenic bacterium *Pectobacterium brasiliense* 1692. *Microbiol. Resour. Announc.* 9:e00037-20. doi: 10.1128/MRA.00037-20

Liu, Y. Y., Helmann, T., Stodghill, P., and Filiatrault, M. (2020). Complete genome sequence resource for the necrotrophic plant-pathogenic bacterium *Pectobacterium carotovorum* WPP14. *Plant Dis.* 105, 196–198. doi: 10.1094/PDIS-05-20-1059-A

Liu, B., and Pop, M. (2009). ARDB-antibiotic resistance genes database. *Nucleic Acids Res.* 37, D443–D447. doi: 10.1093/nar/gkn656

Lory, S. (1998). Secretion of proteins and assembly of bacterial surface organelles: shared pathways of extracellular protein targeting. *Curr. Opin. Microbiol.* 1, 27–35. doi: 10.1016/s1369-5274(98)80139-2

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Ma, B., Hibbing, M. E., Kim, H. S., Reedy, R. M., Yedidia, I., and Breuer, J. (2007). Host range and molecular phylogenies of the soft rot Enterobacterial genera *Pectobacterium* and *Dickeya*. *Phytopathology* 97, 1150–1163. doi: 10.1094/PHYTO-97-9-1150

Makarova, K., Wolf, Y., and Koonin, E. (2015). Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between Thermococcales, Methanococcales, and Methanobacteriales. *Life* 5, 818–840. doi: 10.3390/life5010818

Malapi-Wight, M., Demers, J. E., Veltri, D., Marra, R. E., and Crouch, J. A. (2016). LAMP detection assays for boxwood blight pathogens: a comparative genomics approach. *Sci. Rep.* 6:26140. doi: 10.1038/srep26140

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, M., et al. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13, 614–629. doi: 10.1111/j.1364-3703.2012.00804.x

Maqbool, O., and Babri, H. (2007). Hierarchical clustering for software architecture recovery. *IEEE Trans. Softw. Eng.* 33, 759–780. doi: 10.1109/TSE.2007.70732

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinf.* 14:60. doi: 10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., and Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* 10:2182. doi: 10.1038/s41467-019-10210-3

Meng, X. L., Chai, A., Shi, Y. X., Xie, X. W., Ma, Z. H., and Li, B. J. (2017). Emergence of bacterial soft rot in cucumber caused by *Pectobacterium carotovorum* subsp. *brasiliense* in China. *Plant Dis.* 101, 279–287. doi: 10.1094/PDIS-05-16-0763-RE

Moraes, A. J. G., Baia, A. D. B., Souza, E. B., Peixoto, A. R., and Gama, M. A. S. (2020). First report of *Pectobacterium aroidearum* causing soft rot of pepper (*Capsicum annuum*) fruits in Brazil. *Plant Dis.* 104:3054. doi: 10.1094/PDIS-02-20-0403-PDN

Moraes, A. J. G., Souza, E. B., Mariano, R. L. R., Silva, A. M. F., Lima, N. B., Peixoto, A. R., et al. (2017). First report of *Pectobacterium aroidearum* and *Pectobacterium carotovorum* subsp. *brasiliensis* causing soft rot of *Cucurbita pepo* in Brazil. *Plant Dis.* 101, 379–380. doi: 10.1094/PDIS-08-16-1168-PDN

Moretti, C., Fakhr, R., Cortese, C., De, V. P., Cerri, M., Geagea, L., et al. (2016). *Pectobacterium aroidearum* and *Pectobacterium carotovorum* subsp. *carotovorumas* causal agents of potato soft rot in Lebanon. *Eur. J. Plant Pathol.* 144, 205–211. doi: 10.1007/s10658-015-0743-3

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., et al. (2000). A whole-genome assembly of *Drosophila*. *Science* 287, 2196–2204. doi: 10.1126/science.287.5461.2196

Nabhan, S., De Boer, S. H., Maiss, E., and Wydra, K. (2013). *Pectobacterium aroidearum* sp. nov. a soft rot pathogen with preference for monocotyledonous plants. *Int. J. Syst. Evol. Microbiol.* 63, 2520–2525. doi: 10.1099/ijs.0.046011-0

Nandi, T., Ong, C., Singh, A. P., Justin, B., Timothy, A., Mitali, S. T., et al. (2010). A genomic survey of positive selection in *Burkholderia pseudomallei* provides insights into the evolution of accidental virulence. *PLoS Pathog.* 6:e1000845. doi: 10.1371/journal.ppat.1000845

Nykyri, J., Niemi, O., Koskinen, P., Nokso-Koivisto, J., Pasanen, M., Broberg, M., et al. (2012). Revised phylogeny and novel horizontally acquired virulence determinants of the model soft rot phytopathogen *pectobacterium* wasabiae scc3193. *PLoS Pathog.* 8:e1003013. doi: 10.1371/journal.ppat.1003013

Oulghazi, S., Cigna, J., Lau, Y. Y., Moumni, M., Chan, K. G., and Faure, D. (2019). Transfer of the waterfall source isolate *Pectobacterium carotovorum* M022 to *Pectobacterium* fontis sp. nov., a deep-branching species within the genus *Pectobacterium*. *Int. J. Syst. Evol. Microbiol.* 69, 470–475. doi: 10.1099/ijsem.0.003180

Oulghazi, S., Moumni, M., Khayi, S., Robic, K., and Faure, D. (2020). Diversity of *Pectobacteriaceae* species in potato growing regions in northern Morocco. *Microorganisms* 8:895. doi: 10.3390/microorganisms8060895

Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020). List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* 70, 5607–5612. doi: 10.1099/ijsem.0.004332

Pasanen, M., Waleron, M., Schott, T., Cleenwerck, I., Misztak, A., Waleron, K., et al. (2020). *Pectobacterium parvum* sp. nov., having a *Salmonella* SPI-1-like type III secretion system and low virulence. *Int. J. Syst. Evol. Microbiol.* 70, 2440–2448. doi: 10.1099/ijsem.0.004057

Pedersen, J. S., Carstens, A. B., Djurhuus, A. M., Kot, W., and Hansen, L. H. (2020). Isolation and complete genome sequencing of *Pectobacterium* phage Jarilo, representing a novel genus of bacteriophages within the subfamily *Autographivirinae*. bioRxiv [Preprint]. doi:10.1101/2020.01.16.909176

Pédron, J., Claire, B., Taghouti, G., Portier, P., and Barny, M.-A. (2019). *Pectobacterium aquaticum* sp. nov., isolated from waterways. *Int. J. Syst. Evol. Microbiol.* 69, 745–751. doi: 10.1099/ijsem.0.003229

Pérombelon, M. C. M. (2002). Potato diseases caused by soft rot erwinias: an overview of pathogenesis. *Plant Pathol.* 51, 1–12. doi: 10.1046/j.0032-0862.2001.Shorttitle.doc.x

Philip, J., David, B., Chang, H. Y., Fraser, M., Li, W. Z., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Portier, P., Pedron, J., Taghouti, G., Fischer-Le Saux, M., Caullireau, E., Bertrand, C., et al. (2019). Elevation of *Pectobacterium carotovorum* subsp. *odoriferum* to species level as *Pectobacterium odoriferum* sp. nov., proposal of *Pectobacterium brasiliense* sp. nov. and *Pectobacterium actinidiae* sp. nov., emended description of *Pectobacterium carotovorum* and description of *Pectobacterium versatile* sp. nov., isolated from streams and symptoms on diverse plants. *Genomics Taxon.* 69, 3214–3223. doi: 10.1099/ijsem.0.003611

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods.* 8, 12–24. doi: 10.1039/c5ay02550h

Pritchard, L., Humphris, S., Saddler, G. S., Parkinson, N. M., Bertrand, V., Elphinstone, J. G., et al. (2012). Detection of phytopathogens of the genus *Dickeya* using a PCR primer prediction pipeline for draft bacterial genome sequences. *Plant Pathol.* 62, 587–596. doi: 10.1111/j.1365-3059.2012.02678.x

Qi, T., Wang, S., Deng, L. L., Yi, L. H., and Zeng, K. F. (2021). Controlling pepper soft rot by *Lactobacillus paracasei* WX322 and identification of multiple bacteriocins by complete genome sequencing. *Food Control.* 121:107629. doi: 10.1016/j.foodcont.2020.107629

Richter, M., Rosselló-Móra, R., Glöckner, F. O., and Peplies, J. (2015). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi: 10.1093/bioinformatics/btv681

Sarfraz, S., Oulghazi, S., Cigna, J., Sahi, S. T., Riaz, K., Tufail, M. R., et al. (2020). Complete genome sequence of the type strain *Pectobacterium punjabense* SS95, isolated from a potato plant with blackleg symptoms. *Microbiol. Res. Announc.* 9:e00420. doi: 10.1128/MRA.00420-20

Sarfraz, S., Riaz, K., Oulghazi, S., Cigna, J., Sahi, S. T., Khan, S. H., et al. (2018). *Pectobacterium punjabense* sp. nov., isolated from blackleg symptoms of potato plants in Pakistan. *Int. J. Syst. Evol. Microbiol.* 68, 3551–3556. doi: 10.1099/ijsem.0.003029

Sun, M. M., Liu, H., Huang, J. B., Peng, J. B., Fei, F. H., Zhang, Y., et al. (2019). A loop-mediated isothermal amplification assay for rapid detection of *Pectobacterium aroidearum* that causes soft rot in konjac. *Int. J. Mol. Sci.* 20:1937. doi: 10.3390/ijms20081937

Tam, V. C., Serruto, D., Dziejman, M., Brieher, W., and Mekalanos, J. J. (2007). A type III secretion system in *Vibrio cholerae* translocates a formin/spire hybrid-like actin nucleator to promote intestinal colonization. *Cell Host Microbe* 1, 95–107. doi: 10.1016/j.chom.2007.03.005

Tampakaki, A. P., Fadouloglou, V. E., Gazi, A. D., Panopoulos, N. J., and Kokkinidis, M. (2004). Conserved features of type III secretion. *Cell. Microbiol.* 6, 805–816. doi: 10.1111/j.1462-5822.2004.00432.x

Tang, W. Q., Chang, C. Y., Lee, Y. J., and Chu, C. C. (2021). First report of *Pectobacterium aroidearum* causing bacterial soft rot of carrot in Taiwan. *Plant Dis.* 105:695. doi: 10.1094/PDIS-08-20-1824-PDN

Tang, J. T., Zheng, L., Jia, Q., Liu, H., Hsiang, T., and Huang, J. B. (2017). PCR markers derived from comparative genomics for detection and identification of the rice pathogen *Ustilaginoidea virens* in plant tissues. *Plant Dis.* 101, 1515–1521. doi: 10.1094/PDIS-08-16-1088-RE

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

Toth, I. K., Bell, K. S., Holeva, M. C., and Birch, P. R. J. (2003). Soft rot erwiniae: from genes to genomes. *Mol. Plant Pathol.* 4, 17–30. doi: 10.1046/j.1364-3703.2003.00149.x

Toth, I., Humphris, S., Campbell, E., and Pritchard, L. (2015). Why genomics research on *Pectobacterium* and *Dickeya* makes a difference. *Am. J. Potato Res.* 92, 218–222. doi: 10.1007/s12230-015-9446-8

Toth, I. K., van der Wolf, J. M., Saddler, G., Lojkowska, E., Helias, V., Pirhonen, M., et al. (2011). *Dickeya* species: an emerging problem for potato production in Europe. *Plant Pathol.* 60, 385–399. doi: 10.1111/j.1365-3059.2011.02427.x

UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989

Van Dam, P., de Sain, M., ter Horst, A., van der Gragt, M., and Rep, M. (2018). Use of comparative genomics-based markers for discrimination of host specificity in *Fusarium oxysporum*. *Appl. Environ. Microbiol.* 84:e01868–17. doi: 10.1128/AEM.01868-17

Vargas, W. A., Martin, J. M. S., Rech, G. E., Rivera, L. P., Benito, E. P., Diaz-Minguez, J. M., et al. (2012). Plant defense mechanisms are activated during biotrophic and necrotrophic development of *Colletotricum graminicola* in maize. *Plant Physiol.* 158, 1342–1358. doi: 10.1104/pp.111.190397

Waldee, E. L. (1945). Comparative studies of some peritrichous phytopathogenic bacteria. *Iowa State J. Sci.* 19, 435–484.

Waleron, M., Misztak, A., Waleron, M., Franczuk, M., Jonca, J., Wielgomas, B., et al. (2019). *Pectobacterium zantedeschiae* sp. nov. a new species of a soft rot pathogen isolated from Calla lily (*Zantedeschia* spp.). *Syst. Appl. Microbiol.* 42:153. doi: 10.1016/j.syapm.2018.08.004

Waleron, M., Misztak, A., Waleron, M., Franczuk, M., Wielgomas, B., and Waleron, K. (2018). Transfer of *Pectobacterium carotovorum* subsp. *carotovorum* strains isolated from potatoes grown at high altitudes to *Pectobacterium peruviense* sp. nov. *Syst. Appl. Microbiol.* 41, 85–93. doi: 10.1016/j.syapm.2017.11.005

Wang, L., L, X. B., Suo, H. C., An, K., Luo, H. M., and Liu, X. J. (2017). Soft rot of potatoes caused by *Bacillus amyloliquefaciens* in Guangdong province China. *Can. J. Plant pathol.* 39, 533–539. doi: 10.1080/07060661.2017.1381994

Winnenburg, R., Baldwin, T. K., Urban, M., Rawlings, C., Köhler, J., Hammond-Kosack, K. E., et al. (2006). PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res.* 34, D459–D464. doi: 10.1093/nar/gkj047

Wirawan, A., Kwoh, C. K., Hsu, L. Y., and Koh, T. H. (2010). "INVERTER: integrated variable number tandem repeat finder," in *Computational Systems-Biology and Bioinformatics*. eds. J. H. Chan, Y.-S. Ong and S.-B. Cho (Berlin: Springer), 151–164.

Xie, H., Li, X. Y., Ma, Y. L., and Tian, Y. (2017). First report of *Pectobacterium aroidearum* causing soft rot of Chinese cabbage in China. *Plant Dis.* 102:674. doi: 10.1094/PDIS-07-17-1059-PDN

Xie, Y. P., Shao, X. L., and Deng, X. (2019). Regulation of type III secretion system in *Pseudomonas syringae*. *Environ. Microbiol.* 21, 4465–4477. doi: 10.1111/1462-2920.14779

Xu, P. D., Wei, D. D., Li, Z. P., Qin, C. X., Li, X., Lin, C. H., et al. (2020). First report of bacterial soft rot on *Syngonium podophyllum* caused by *Pectobacterium aroidearum* in China. *Plant Dis.* 104:2720. doi: 10.1094/PDIS-03-20-0552-PDN

Yaganza, E. S., Tweddell, R. J., and Arul, J. (2014). Postharvest application of organic and inorganic salts to control potato (*Solanum tuberosum* L.) storage soft rot: plant tissue-salt physicochemical interactions. *J. Agric. Food Chem.* 62, 9223–9231. doi: 10.1021/jf5017863

Yang, J. K., Tian, B. Y., Liang, L. L. M., and Zhang, K. Q. (2007). Extracellular enzymes and the pathogenesis of nematophagous fungi. *Appl. Microbiol. Biotechnol.* 75, 21–31. doi: 10.1007/s00253-007-0881-4

Zhang, Y., Fan, Q., and Loria, R. (2016). A re-evaluation of the taxonomy of phytopathogenic genera *Dickeya* and *Pectobacterium* using whole genome sequencing data. *Syst. Appl. Microbiol.* 39, 252–259. doi: 10.1016/j.syapm.2016.04.001

Zhang, D., Zhou, Y., Zhao, D. M., Zhou, J. M., Yang, Z. H., and Zhu, M. M. (2017). Complete genome sequence and pathogenic genes analysis of *Pectobacterium atroseptica* JG10-08. *Genes Genom.* 39, 945–955. doi: 10.1007/s13258-017-0559-y

Zhou, Y., Liang, Y. J., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39(Suppl. 2), W347–W352. doi: 10.1093/nar/gkr485

Zoledowska, S., Motyka-Pomagruk, A., Sledz, W., Mengoni, A., and Lojkowska, E. (2018). High genomic variability in the plant pathogenic bacterium *Pectobacterium parmentieri* deciphered from de novo assembled complete genomes. *BMC Genomics* 19:751. doi: 10.1186/s12864-018-5140-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fmicb.2021.684756

Pan-Genome-Wide Analysis of *Pantoea ananatis* Identified Genes Linked to Pathogenicity in Onion

Gaurav Agarwal[1], Divya Choudhary[1], Shaun P. Stice[2], Brendon K. Myers[1], Ronald D. Gitaitis[1], Stephanus N. Venter[3], Brian H. Kvitko[2] and Bhabesh Dutta[1]**

[1] Department of Plant Pathology, Coastal Plain Experimental Station, University of Georgia, Tifton, GA, United States, [2] Department of Plant Pathology, University of Georgia, Athens, GA, United States, [3] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa

OPEN ACCESS

Edited by:
Neha Potnis,
Auburn University, United States

Reviewed by:
Flavia Figueira Aburjaile,
Federal University of Minas Gerais, Brazil
Siomar De Castro Soares,
Universidade Federal do Triângulo Mineiro, Brazil
Veronica Roman-Reyna,
The Ohio State University, United States

***Correspondence:**
Bhabesh Dutta
bhabesh@uga.edu
Gaurav Agarwal
gaurav.agarwal@uga.edu

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: 23 March 2021
Accepted: 28 July 2021
Published: 19 August 2021

Citation:
Agarwal G, Choudhary D, Stice SP, Myers BK, Gitaitis RD, Venter SN, Kvitko BH and Dutta B (2021) Pan-Genome-Wide Analysis of Pantoea ananatis Identified Genes Linked to Pathogenicity in Onion. Front. Microbiol. 12:684756. doi: 10.3389/fmicb.2021.684756

Pantoea ananatis, a gram negative and facultative anaerobic bacterium is a member of a *Pantoea* spp. complex that causes center rot of onion, which significantly affects onion yield and quality. This pathogen does not have typical virulence factors like type II or type III secretion systems but appears to require a biosynthetic gene-cluster, HiVir/PASVIL (located chromosomally comprised of 14 genes), for a phosphonate secondary metabolite, and the '*alt*' gene cluster (located in plasmid and comprised of 11 genes) that aids in bacterial colonization in onion bulbs by imparting tolerance to thiosulfinates. We conducted a deep pan-genome-wide association study (pan-GWAS) to predict additional genes associated with pathogenicity in *P. ananatis* using a panel of diverse strains (n = 81). We utilized a red-onion scale necrosis assay as an indicator of pathogenicity. Based on this assay, we differentiated pathogenic (n = 51)- vs. non-pathogenic (n = 30)-strains phenotypically. Pan-genome analysis revealed a large core genome of 3,153 genes and a flexible accessory genome. Pan-GWAS using the presence and absence variants (PAVs) predicted 42 genes, including 14 from the previously identified HiVir/PASVIL cluster associated with pathogenicity, and 28 novel genes that were not previously associated with pathogenicity in onion. Of the 28 novel genes identified, eight have annotated functions of site-specific tyrosine kinase, N-acetylmuramoyl-L-alanine amidase, conjugal transfer, and HTH-type transcriptional regulator. The remaining 20 genes are currently hypothetical. Further, a core-genome SNPs-based phylogeny and horizontal gene transfer (HGT) studies were also conducted to assess the extent of lateral gene transfer among diverse *P. ananatis* strains. Phylogenetic analysis based on PAVs and whole genome multi locus sequence typing (wgMLST) rather than core-genome SNPs distinguished red-scale necrosis inducing (pathogenic) strains from non-scale necrosis inducing (non-pathogenic) strains of *P. ananatis*. A total of 1182 HGT events including the HiVir/PASVIL and *alt* cluster genes were identified. These events could be regarded as a major contributing factor to the diversification, niche-adaptation and potential acquisition of pathogenicity/virulence genes in *P. ananatis*.

Keywords: pan-genome, horizontal gene transfer, genome-wide association study, SNPs, presence and absence variants

INTRODUCTION

The genus *Pantoea* currently has 27 recognized species; five of which are known to cause disease-associated losses in several crops (Arnold et al., 2003; Cruz et al., 2007; Coutinho and Venter, 2009; Bankevich et al., 2012; De Maayer et al., 2014). Three species of *Pantoea*, namely, *P. ananatis, P. agglomerans* and *P. stewartii* subsp. *indologenes* are responsible for more than 80% of the reported cases of disease in onions (Kini et al., 2019). *Pantoea ananatis*, a Gram-negative and facultative anaerobic bacterium that belongs to the Erwiniaceae (previously assigned to enterobacteriaceae) is part of a *Pantoea* spp. complex (also including *P. allii, P. agglomerans*, and *P. stewartii* subs. *indologenes*), which causes center rot of onion (Gitaitis et al., 2002; Walcott et al., 2002; Stumpf et al., 2018). Foliar symptoms primarily appear as white streaks and water-soaked lesions, and more advanced infections result in complete collapse of foliar tissues, discoloration and softening of specific scale layers in the bulb. Under favorable conditions, center rot can result in 100% losses in the field. In Georgia, late maturing varieties are more susceptible to center rot than early maturing varieties (Gitaitis and Gay, 1997; Carr et al., 2013; Agarwal et al., 2019). Out of the four species in the *Pantoea* spp. complex, *P. ananatis* has been associated predominantly with center rot, in Georgia (Dutta et al., 2014). *Pantoea ananatis* has been identified in other onion-growing regions of the United States, including Colorado (Schwartz and Otto, 2000), Michigan (Schwartz and Mohan, 2008), New York (Carr et al., 2010), and Pennsylvania (Pfeufer et al., 2015). The bacterium can be found as an epiphyte on crop and weed plants (Gitaitis et al., 2002) or as an endophyte in maize kernels and rice seeds (Okunishi et al., 2005; Rijavec et al., 2007). Apart from its epiphytic and endophytic niche in crops and weeds, *P. ananatis* can be disseminated through infected onion seed and contaminated insect vectors (thrips) to onion crops (Gitaitis et al., 2003; Dutta et al., 2014).

Pantoea ananatis, unlike many other phytopathogenic bacteria, lacks genes that code for type II, III and IV protein secretory systems that are associated with pathogenicity and virulence (De Maayer et al., 2014). Recent studies utilized whole genome/small RNA sequencing that aided in identifying some virulence factors associated with *P. ananatis* in onion. These virulence factors are flagellar and pilin motility factors (Weller-Stuart et al., 2017) and a global virulence regulator (Hfq, an RNA chaperone) associated with quorum sensing and biofilm production (Shin et al., 2019). However, these genetic factors are present in both onion-pathogenic and non-pathogenic strains. Previous comparative genomics studies on *P. ananatis* were carried out to identify pathogenicity-related regions in the genome using a small sub-set of strains (2 to 10 strains) (De Maayer et al., 2014; Sheibani-Tezerji et al., 2015; Asselin et al., 2018; Stice et al., 2018). Asselin et al. (2018) proposed a biosynthetic gene-cluster, HiVir/PASVIL (High Virulence also known as PASVIL; *Pantoea ananatis* specific virulence locus) that encodes a proposed phosphonate or phosphonate secondary metabolite cluster located on the chromosome, has been demonstrated to be associated with onion foliar and bulb necrosis (Asselin et al., 2018; Takikawa and Kubota, 2018). Recently, Polidore et al. (2021) identified this gene cluster to encode for at least three phosphonates two of which are characterized and named as pantaphos 2-(hydroxy[phosphono]methyl) maleate and 2- (phosphonomethyl) maleate.

In addition, Stice et al. (2018) showed that a megaplasmid-borne onion virulence region (OVR) in *P. ananatis* is correlated with onion virulence. In a recent study (Stice et al., 2020) showed that the OVRA cluster contains 11 genes that are critical for colonizing necrotized bulb tissue. This gene cluster was described as the '*alt*' cluster that imparts bacterial tolerance to the thiosulfinate 'allicin' in onion bulbs. Presence or absence of genes in these gene clusters (14 genes in HiVir/PASVIL and 11 genes in '*alt*') may dictate the pathogenic potential of *P. ananatis* in onion. However, a large-scale genomic study utilizing diverse *P. ananatis* strains with pathogenic or non-pathogenic phenotype has not been done to evaluate this aspect.

The pan-genome is defined as the total number of non-redundant genes present in a given clade, amounting to a given clade's entire genomic repertoire, and encodes for all possible ecological niches of the strains examined (Tettelin et al., 2005; Snipen et al., 2009). A pan-genome typically contains core genes, accessory genes and the strain-specific genes. Defining the soft-core genes helps in obtaining a robust estimate of the core genes. The core and soft-core clusters together represented a pool of highly conserved genes, which may provide insights into the evolutionary history of a bacterial pathogen. The cloud cluster include genes that are strain specific or unique in the pan-genome. Both the shell and cloud clusters represent a subset of flexible (accessory) genome that reflects life-style and adaptation of bacterial strains to the environments in which they reside (Nelson and Stegen, 2015). Another significance of a pan-genome is that it can provide a greater resolution and aids in reconstructing bacterial phylogeny in a more reliable way than single or multiple gene-based phylogeny. The pan-genome provides an overview of the entire gene set (100% of the genomes) of a given population, unlike a 16S rRNA phylogeny that represents only a tiny fraction of the genome (∼0.07%), or multi-locus sequence analysis (MLSA) involving house-keeping genes (∼0.2%). De Maayer et al. (2014) analyzed the pan-genomes of eight sequenced *P. ananatis* strains isolated from different sources and identified factors that can potentially explain their ecological niche and their interactions with the environment and host. Asselin et al. (2018) and Stice et al. (2018, 2020) utilized comparative genomic approaches to identify HiVir/PASVIL and the *alt* gene clusters, respectively in onion-pathogenic *P. ananatis*. Further pan-genomic studies with large strain collection from diverse isolation sources and onion-pathogenic potential may shed further light on the novel pathogenic factors responsible for center rot disease in onion. In addition, the openness of a bacterial pan-genome reflects the diversity of the gene pool among the strains of the same bacterial species. The addition of new genomes to an existing pan-genome can significantly alter the size of an open pan-genome in contrast to a closed pan-genome (Vernikos et al., 2015). This aspect was evaluated earlier in *P. ananatis* using a small sub-set of pan-genomes (n = 8), which needs further evaluation using a large and diverse set of *P. ananatis* strains (De Maayer et al., 2014).

Bacterial genomes are labile entities, fluctuating both in size and gene content through time. Such genome fluctuations are maintained by the counteracting processes of gene gain and loss (Touchon et al., 2009; Nowell et al., 2014). Horizontal gene transfer (HGT) can result in the replacement of genetic segments with donor homologs, often within species via homologous recombination, or via. acquisition of new genetic material. The gene-presence-absence variants (PAVs) are primarily the result of various HGT events in bacterial evolution. Such HGT events have not been fully explored in *P. ananatis* particularly in onion-pathosystem. In this manuscript, we therefore explored HGT events among *P. ananatis* strains isolated from diverse sources including onion, weeds and thrips with or without pathogenic potential on onion.

Genomic variants identified using pan-genomic analysis can be associated with the unique phenotypic characteristics of an organism using computational methods to identify the genetic basis of phenotypic variations. Such computational association methods have been utilized in research of humans (Visscher et al., 2017; Chen et al., 2018; Gong et al., 2019), domesticated plants (Varshney et al., 2017; Varshney et al., 2019), animals (Bolormaa et al., 2011), and bacteria (Chen and Shapiro, 2015; Falush, 2016; Epstein et al., 2018). This approach is widely used as a genome wide association study (GWAS) that associates genomic characteristics such as single nucleotide polymorphisms (SNPs), insertion and deletions (InDels) or copy number variants (CNVs) or PAVs with the phenotypic characteristics of an organism. This approach has not been used in *P. ananatis* and may aid in identifying novel factors responsible for pathogenicity in onion. In this manuscript, we describe the use of a pan-GWAS approach that predicted novel genes associated with onion-pathogenicity in *P. ananatis* and provide genomic evidence that presence of entire HiVir/PASVIL and '*alt*' clusters do not warrant an onion pathogenic phenotypic.

MATERIALS AND METHODS

Bacterial Strains, Identification, and Culture Preparation

Eighty-one *P. ananatis* strains used in this study were isolated from onion foliage, bulb and seeds as well as from weeds and thrips throughout the state of Georgia (**Table 1**). These strains were stored in a 15% aqueous glycerol solution at -80°C. The source, year of isolation, and county of origin in Georgia for these strains are listed (**Table 1**). Strains were identified as *P. ananatis* by their colony morphology and physiological characteristics as Gram-negative, facultatively anaerobic, positive for indole production, negative for nitrate reductase, and phenylalanine deaminase and using *P. ananatis*-specific primers (Walcott et al., 2002). Strains that were isolated from onion plants were designated as "PNA" and strains from non-onion sources (e.g., weeds or thrips) were identified as "PANS."

Inoculum was prepared by transferring single colonies of each bacterial strain from 48-h-old cultures on nutrient agar (NA) medium to nutrient broth (NB). The broth was shaken overnight on a rotary shaker (Thermo Fisher Scientific, Gainesville, FL, United States) at 180 rpm. After 12 h of incubation, 3 ml of each bacterial suspension were centrifuged at 6,000 × *g* (Eppendorf, Westbury, NY, United States) for 2 mins. The supernatant was discarded and the pellet was re-suspended in sterile water. Inoculum concentration was adjusted using a spectrophotometer (Eppendorf, Westbury, NY, United States) to an optical density of 0.3 at 600 nm [$\approx 1 \times 10^8$ colony forming unit (CFU)/ml]. The bacterial suspension was diluted serially in sterile distilled water to obtain the desired concentration.

Pantoea ananatis Phenotyping: Red Onion Scale Necrosis Assay as a Measure of Onion-Pathogenicity

The red onion scale necrosis assay was conducted as described by Stice et al. (2018). Briefly, red onion bulbs (cv. Red Creole) were each surface-disinfested by removing the outer, dry scales and then spraying the outer exposed fleshy scales with a 3% sodium hypochlorite solution using a spray bottle. This was followed by wiping the outer scales dry with a sterile paper towel. Later, the bulbs were sprayed with sterile distilled water and wiped again with a sterile paper towel. Surface-disinfested onion scales were each sliced into approximately 6 cm × 3 cm (length × width) segments using a surface-disinfested knife. Each onion scale piece was pierced on the inner surface with a pipette tip, and 10 μl of the relevant bacterial suspension (1×10^6 CFU/ml) was deposited on the wounded tissue. Scales were maintained in petri plates containing autoclaved paper towels moistened with sterile water. Petri plates were kept in an aluminum tray covered with a plastic lid. The inoculated onion scales were incubated at room temperature for 96 h in the dark, after which the size of the necrotic, pigment-clearing zone was recorded as a measure of pathogenicity. Strains that cleared the red anthocyanin pigment and caused necrosis were classified as pathogenic, and those that did not were classified as non-pathogenic to onion (**Supplementary Figure 1**). Three replicates of onion-scale pieces were used to test each strain, and the experiment was repeated twice (a total three experiments). Onion scales inoculated with sterile water and a known pathogenic strain of *P. ananatis* (PNA 97-1) served as negative and positive control treatments, respectively.

To confirm that symptoms on the onion scale were caused by *P. ananatis*, bacteria from symptomatic scale tissue ($n = 3$) were isolated from the margins of the necrotic area and healthy tissue, and streaked onto Tryptic soy broth agar (TSBA) and incubated for 48 h at 28°C. Yellow-pigmented colonies were selected to test for genus and species identity using physiological tests and a species-specific TaqMan-based polymerase-chain reaction (PCR) assay (Dutta et al., 2014) for *P. ananatis*. Briefly, presumptive *P. ananatis* colonies were picked using a sterile inoculation loop and suspended separately in 2 ml micro-centrifuge tubes, each containing 25 μl of sterile deionized water. The bacterial suspension was heated (Modular Dry Block Heaters, Cole Parmer, IL, United States) for 3 min at 95°C. A suspension (5 μl) was amplified in 20 μl of PCR master-mix containing 10 mMmTris-HCl (pH 9.0), 50 mM KCl, 0.1% Triton X-100, 1.5 mM $MgCl_2$, and 0.2 mM of each

TABLE 1 | List of *Pantoea ananatis* strains used in this study along with their pan-genomes details, geographical locations, pathogenic phenotypes and isolation sources.

Strains	BioSample ID	Genome accession	Size (Mb)	Contigs (>300 bp)	CDSs	Genes/ mRNAs	GC (%)	Location	Red-onion scale necrosis (±)*	Source
PNA_99_9	SAMN14604903	JABEBS000000000	5.13	151	4696	4766	53.32	Tattnall Co.	+	Onion
PNA_99_8	SAMN14604904	JABEBR000000000	4.99	56	4549	4617	53.41	Wheeler Co.	+	Onion
PNA_99_7	SAMN14604905	JABEBQ000000000	5.05	36	4662	4728	53.42	Tattnall Co.	+	Onion
PNA_99_6	SAMN14604906	JABEBP000000000	5.1	43	4710	4778	53.36	Toombs Co.	+	Onion
PNA_99_3	SAMN14604907	JABEBO000000000	5.14	39	4659	4726	53.29	Tift Co.	+	Onion
PNA_99_2	SAMN14604908	JABEBN000000000	4.98	38	4540	4607	53.49	Tattnall Co.	+	Onion
PNA_99_14	SAMN14604909	JABEBM000000000	4.96	67	4549	4625	53.44	Toombs Co.	+	Onion
PNA_99_1	SAMN14604910	JABEBL000000000	4.91	67	4489	4556	53.48	MT Vernon	+	Onion
PNA_98_8	SAMN14604911	JABEBK000000000	4.93	31	4547	4618	53.45	Vidalia Region	+	Onion
PNA_98_7	SAMN14604912	JABEBJ000000000	5.08	43	4678	4748	53.64	Tift Co.	−	Onion
PNA_98_3	SAMN14604913	JABEBI000000000	5.47	413	4997	5067	53.18	Dougherty	−	Onion
PNA_98_2	SAMN14604914	JABEBH000000000	5.14	38	4693	4757	53.3	Tift Co.	+	Onion
PNA_98_12	SAMN14604915	JABEBG000000000	5.73	942	4990	5061	53.43	Toombs Co.	+	Onion
PNA_98_11	SAMN14604916	JABEBF000000000	5.8	954	5065	5136	53.59	Evans Co.	−	Onion
PNA_98_1	SAMN14604917	JABEBE000000000	5.67	796	4938	5016	53.47	Tattnall Co.	+	Onion
PNA_97_3	SAMN14604918	JABEBD000000000	4.96	36	4545	4613	53.25	Toombs Co.	+	Onion
PNA_97_11	SAMN14604919	JABEBC000000000	5.09	38	4695	4762	53.4	Toombs Co.	+	Onion
PNA_92_7	SAMN14604920	JABEBB000000000	5.31	373	4714	4780	53.44	Vidalia Region	+	Onion
PNA_200_8	SAMN14604921	JABEBA000000000	5	36	4606	4680	53.52	Tift Co.	−	Onion
PNA_200_7	SAMN14604922	JABEAZ000000000	5.01	70	4567	4637	53.27	Tift Co.	+	Onion
PNA_200_3	SAMN14604923	JABEAY000000000	5.01	39	4610	4686	53.52	Tift Co.	−	Onion
PNA_200_12	SAMN14604924	JABEAX000000000	4.92	34	4517	4591	53.48	Tift Co.	+	Onion
PNA_200_11	SAMN14604925	JABEAW000000000	4.96	32	4545	4616	53.49	Tift Co.	+	Onion
PNA_200_10	SAMN14604926	JABEAV000000000	5.01	25	4589	4655	53.48	Tift Co.	+	Onion
PNA_18_9s	SAMN14604927	JABEAU000000000	5.52	342	5110	5178	52.94	Vidalia Region	+	Onion
PNA_18_8s	SAMN14604928	JABEAT000000000	5.01	34	4651	4714	53.42	Vidalia Region	−	Onion
PNA_18_7s	SAMN14604929	JABEAS000000000	4.88	52	4422	4492	53.49	Vidalia Region	+	Onion
PNA_18_6s	SAMN14604930	JABEAR000000000	5.02	46	4648	4719	53.42	Vidalia Region	−	Onion
PNA_18_5s	SAMN14604931	JABEAQ000000000	5.61	1027	5054	5182	53.03	Vidalia Region	+	Onion
PNA_18_5	SAMN14604932	JABEAP000000000	4.86	46	4429	4499	53.49	Vidalia Region	+	Onion
PNA_18_3s	SAMN14604933	JABEAO000000000	4.95	34	4524	4591	53.64	Vidalia Region	+	Onion
PNA_18_2	SAMN14604934	JABEAN000000000	4.95	26	4526	4591	53.25	Vidalia Region	+	Onion
PNA_18_10s	SAMN14604935	JABEAM000000000	5.08	80	4668	4738	53.42	Vidalia Region	−	Onion
PNA_18_10	SAMN14604936	JABEAL000000000	5.53	433	4932	5006	53.56	Vidalia Region	−	Onion
PNA_18_1	SAMN14604937	JABEAK000000000	4.96	27	4530	4603	53.25	Vidalia Region	+	Onion
PNA_15_3	SAMN14604938	JABEAJ000000000	4.93	33	4544	4619	53.24	Tattnall Co.	−	Onion
PNA_15_1	SAMN14604939	JABEAI000000000	4.99	31	4588	4662	53.34	Tattnall Co.	+	Onion
PNA_14_2	SAMN14604940	JABEAH000000000	4.98	46	4615	4691	53.35	Lyons	−	Onion
PNA_13_1	SAMN14604941	JABEAG000000000	4.87	46	4498	4564	53.39	Lyons	−	Onion
PNA_11_1	SAMN14604942	JABEAF000000000	4.91	30	4476	4549	53.41	Vidalia Region	−	Onion
PNA_08_1	SAMN14604943	JABEAE000000000	4.88	31	4486	4561	53.59	Tattnall Co.	+	Onion
PNA_07_7	SAMN14604944	JABEAD000000000	4.9	26	4495	4568	53.43	Toombs Co.	+	Onion
PNA_07_5	SAMN14604945	JABEAC000000000	4.94	27	4551	4621	53.6	Wayne Co.	+	Onion
PNA_07_22	SAMN14604946	JABEAB000000000	4.91	62	4534	4598	53.6	Tift Co.	−	Onion
PNA_07_14	SAMN14604947	JABEAA000000000	4.93	32	4503	4571	53.52	Toombs Co.	−	Onion
PNA_07_13	SAMN14604948	JABDZZ000000000	4.93	29	4527	4594	53.48	Toombs Co.	−	Onion
PNA_07_10	SAMN14604949	JABDZY000000000	4.91	19	4505	4570	53.41	Toombs Co.	+	Onion
PNA_07_1	SAMN14604950	JABDZX000000000	4.92	33	4492	4557	53.47	Tattnall Co.	+	Onion
PNA_06_4	SAMN14604951	JABDZW000000000	4.96	49	4557	4620	53.6	Wayne Co.	+	Onion
PNA_05_1	SAMN14604952	JABDZV000000000	5.07	41	4655	4726	53.35	Vidalia Region	+	Onion
PNA_03_2	SAMN14604953	JABDZU000000000	5.06	108	4626	4698	53.51	Tift Co.	−	Onion
PNA_03_1	SAMN14604954	JABDZT000000000	4.96	27	4588	4659	53.55	Tift Co.	+	Onion
PNA_02_12	SAMN14604955	JABDZS000000000	4.73	34	4300	4373	53.58	Tift Co.	+	Onion
PANS_99_9	SAMN14604956	JABDZR000000000	5.1	61	4684	4759	53.53	Tift Co.	−	Prairie verbena
PANS_99_5	SAMN14604957	JABDZQ000000000	4.99	46	4612	4683	53.51	Tift Co.	−	Prairie verbena
PANS_99_4	SAMN14604958	JABDZP000000000	5.12	39	4654	4719	53.29	Tift Co.	+	Florida pusley
PANS_99_36	SAMN14604959	JABDZO000000000	4.87	41	4458	4524	53.5	Terrell Co.	−	Florida pusley
PANS_99_33	SAMN14604960	JABDZN000000000	4.98	37	4553	4621	53.28	Coffee Co.	+	Florida pusley
PANS_99_32	SAMN14604961	JABDZM000000000	4.99	28	4576	4653	53.53	Vidalia Region	−	Florida pusley
PANS_99_31	SAMN14604962	JABDZL000000000	5.67	938	5105	5234	52.63	Tattnal Co.	+	Texas millet
PANS_99_29	SAMN14604963	JABDZK000000000	4.86	37	4447	4516	53.46	Tift Co.	+	Crab grass
PANS_99_27	SAMN14604964	JABDZJ000000000	5.35	621	4720	4797	53.54	Vidalia Region	+	Florida beggarweed

(Continued)

TABLE 1 | Continued

Strains	BioSample ID	Genome accession	Size (Mb)	Contigs (>300 bp)	CDSs	Genes/ mRNAs	GC (%)	Location	Red-onion scale necrosis (±)*	Source
PANS_99_26	SAMN14604965	JABDZI000000000	4.95	33	4481	4559	53.17	Vidalia Region	–	Hyssop spurge
PANS_99_25	SAMN14604966	JABDZH000000000	4.83	24	4418	4487	53.54	Vidalia Region	+	Bristly starbur
PANS_99_23	SAMN14604967	JABDZG000000000	4.94	32	4527	4597	53.3	Vidalia Region	–	Yellow nut sedge
PANS_99_22	SAMN14604968	JABDZF000000000	4.96	27	4570	4638	53.55	Tift Co.	–	Crab grass
PANS_99_11	SAMN14604969	JABDZE000000000	5.08	46	4698	4767	53.22	Tift Co.	+	Crab grass
PANS_4_2	SAMN14604970	JABDZD000000000	4.91	30	4498	4564	53.52	Tift Co.	–	Adult tobacco thrip from peanut
PANS_200_2	SAMN14604971	JABDZC000000000	4.9	113	4418	4489	53.39	Reidsville	+	Pink purslane
PANS_200_1	SAMN14604972	JABDZB000000000	4.83	20	4429	4498	53.22	Reidsville	–	Slender amaranth
PANS_2_8	SAMN14604973	JABDZA000000000	4.84	28	4462	4531	53.41	Tift Co.	+	Thrips infected peanut leaf
PANS_2_7	SAMN14604974	JABDYZ000000000	4.84	28	4466	4534	53.42	Tift Co.	+	Thrips from peanut blossoms
PANS_2_6	SAMN14604975	JABDYY000000000	5.1	31	4682	4747	53.44	Tift Co.	+	Thrips from peanut blossoms
PANS_2_5	SAMN14604976	JABDYX000000000	4.86	39	4468	4534	53.41	Tift Co.	+	Thrips from peanut blossoms
PANS_2_1	SAMN14604977	JABDYW000000000	5.01	38	4610	4679	53.55	Tift Co.	–	Adult tobacco thrip from peanut
PANS_1_9	SAMN14604978	JABDYV000000000	4.96	22	4576	4647	53.51	Tift Co.	–	Thrip feces from peanut leaf
PANS_1_8	SAMN14604979	JABDYU000000000	5	31	4596	4664	53.45	Tift Co.	–	Adult tobacco thrip
PANS_1_6	SAMN14604980	JABDYT000000000	4.81	21	4395	4466	53.45	Tift Co.	+	Adult tobacco thrip
PANS_1_5	SAMN14604981	JABDYS000000000	4.81	23	4398	4476	53.45	Tift Co.	+	Adult tobacco thrip
PANS_1_2	SAMN14604982	JABDYR000000000	4.92	32	4520	4590	53.48	Tift Co.	+	Thrips from onion leaf
PANS_1_10	SAMN14604983	JABDYQ000000000	4.96	26	4578	4651	53.51	Tift Co.	–	Thrip feces from peanut leaf

*'+' indicates red-onion scale necrosis causing strain '-' indicates non-red-onion-scale necrosis causing strain.

nucleotide (dATP, dCTP, dGTP, and dTTP), 25 μM each of primer PanITS1 (5′-GTCTGATAGAAAGATAAAGAC-3′) and EC5 (5′-CGGTGGATGCCCTGGCA-3′) and 10 μM of TaqMan probe 6-FAM TAGCGGTTAGGACTCCGCCCTTTCA-BHQ. The PCR reaction was conducted in a Cepheid Smart Cycler (Sunnyvale, CA, United States) using the following thermal profile: denaturation at 95°C for 180 s, 35 cycles each of denaturation at 95°C for 15 s, and annealing at 60°C for 40 s. Samples with cycle threshold (Ct) values <35 were considered positive for *P. ananatis*.

DNA Isolation and Library Preparation for Whole Genome Sequencing of *P. ananatis*

Pantoea ananatis strains (n = 81) were revived from the 15% aqueous glycerol solution at −80°C by streaking individually onto NA and incubated for 48-h at 28°C. After incubation, a single colony of each strain was picked and placed into 3 ml of Luria Bertani broth. The resulting broth was shaken overnight on a rotatory shaker (Thermo Fisher Scientific, Gainesville, FL, United States) at 180 rpm at 30°C. After incubation, broth (1.5 ml) cultures were each centrifuged in a microcentrifuge tube (2 ml) at 6,000 × g. The supernatant was discarded and the bacterial pellet suspended in 1 ml of sterile water, from which DNA was isolated. DNA isolation was done using a DNeasy ultra clean microbial DNA extraction kit (Qiagen, Germany) using the manufacturer's instructions. DNA samples were quantified (ng/μl) using a Nanodrop (Thermo Fisher Scientific, Gainesville, FL, United States).

For Illumina Nextera library preparation, a total of 100 ng DNA from each bacterial strain was used according to KAPA Hyper Prep kit (KAPA Biosystems, MA, United States) at the Georgia Genomics and Bioinformatics Core (GGBC), University of Georgia, Athens, GA, United States. Briefly, the genomic DNA from each bacterial strain was fragmented followed by end repairing and A-tailing, which produced end-repaired 5′-phosphorylated, 3′-dA-tailed DNA fragments. Adapters were ligated resulting in a 3′-dTMP overhang ligated to 3′-dA-tailed molecules. Post-ligation cleanup was performed to remove un-ligated adapter and/or adapter-dimer molecules from the library before library amplification. Library amplification was done employing a high-fidelity, low-bias PCR assay to amplify library fragments with appropriate adapters on the ends. Dual indexing was done during the library preparation and PCR amplification. Dual indexing (by introducing indexes into both library adapters) was done to overcome the occurrence of mixed clusters on the flow as this is a predominant source of error while multiplex sequencing. Libraries for the 81 bacterial strains were pooled and sequenced on the Illumina Nextseq500 using a high output run. All samples were sequenced to produce 150 bp paired end reads.

Read Data Filtering

FastQC was run to assess the raw fastq files. Overall, data quality was good with typical drop off in quality at the ends of the reads. The number of raw reads ranged from 8.4 million to 18.2 million PE reads. The read data were filtered to remove low quality reads/bases and trimmed for reads containing primer/adaptor sequences using Trimmomatic (v 0.36) in paired end mode (Bolger et al., 2014). Further, all 5′ and 3′ stretches of ambiguous

'N' nucleotides were clipped to ensure high quality reads for downstream analysis. Trimmed data were re-assessed using FastQC. These data were used for genome assembly followed by pan-genome analyses. A total of 675.6 million raw reads were generated and, after stringent quality filtering, 657.6 million high-quality reads were obtained including a lowest of 3.55 million and a highest of 16.4 million reads (**Supplementary Table 1**).

Genome Assembly and Pan-Genome Analyses

Trimmed reads were assembled using the SPAdes (v 3.11.1) assembler (Bankevich et al., 2012). Both the paired and unpaired data were used in assembly at default settings. Assembly files were submitted to NCBI under the bio-project identity PRJNA624643. The assembled contigs were also submitted to LINbase to obtain life identification numbers (LIN) (**Supplementary Table 18**). The scaffolds of the respective 81 *P. ananatis* strains were annotated using prokka (v 1.13) (Seemann, 2014) to produce gff3 (general feature format) and gbk (gene bank format) files. The gbk files were used for pan-genome analyses by using get_homologues (Contreras-Moreira and Vinuesa, 2013). These gbk files were used to get the syntenic sequence clusters by get_homologues.pl using OrthoMCL (OMCL) algorithm. In order to check the openness/closeness of the pan-genome, the theoretical estimation of pan-genome size was carried out using an exponential model of Tettelin et al. (2005), which was fitted to the OMCL accessory gene clusters. The syntenic clusters generated were used to a develop a pan-genome matrix showing presence and absence variants (PAVs) using compare_clusters.pl, and the pan-genome matrix was used to classify the genes into core, soft-core, shell and cloud genes using parse_pangenome_matrix.pl (auxiliary script of get_homologues.pl). Core genes (part of soft-core) were defined as those present in all 81 genomes and accessory genes were present in a subset of the 81 strains. The accessory gene cluster was further divided into shell and cloud gene clusters. Soft-core genes occurred in 95% of the genomes. Cloud genes were present in $\leq$2 genomes and shell genes comprised of remaining genes (Contreras-Moreira and Vinuesa, 2013). Distribution of cluster sizes as a function of the number of genomes these clusters contained was displayed using R with parse_pangenome_matrix.pl. Gower's distance matrix was generated using the tab delimited pan-genome PAV file as input by executing the shell script hcluster_pangenome_matrix.sh (auxiliary script of get_homologues) when used to call R function hclust. The presence and absence of 14 HiVir/PASVIL and 11 *alt* cluster genes were determined by looking at the cluster of genes using compare_clusters.pl program and blastn search, respectively. HiVir/PASVIL genes that were clustered from different strains were considered present in those strains. Each of the *alt* cluster sequence was subjected to blastn against each of the 81 genome assemblies. If a blast hit was found against a genome, that gene was considered present. The *alt* cluster was considered conserved in a given strain if all 11 genes were present, and was considered absent if anywhere from 9 to 11 genes were absent. Similarly, if all 14 genes of the HiVir/PASVIL cluster were present in a strain, then it was considered conserved, otherwise it was rated as absent if all some or all of the genes were absent in a particular strain. The presence and absence of these genes was plotted as heat maps.

Horizontal Gene Transfer (HGT) and Prediction of Genomic Islands, and Pathogenicity and Symbiotic Factors in *P. ananatis*

The horizontal gene transfer (HGT) study was conducted in two steps. Broadly, the first step involved phylogenetic classification of genomes based on conserved genes to identify the closely clustered genomes. The second step used the genomes that clustered together in groups (A-E) to find the HGT events among these groups. In the first step, all *P. ananatis* draft genomes were scanned for the presence of 120 conserved bacterial marker genes or single copy genes (SCG) using GTDB toolkit (Chaumeil et al., 2020). Genomes were assigned to the domain with the highest proportion of identified marker genes. The selected domain-specific markers were aligned with HMMER, concatenated into a single multiple sequence alignment (MSA). Based on MSA, strains (81 genomes) were classified into a reference phylogenetic tree. The tree file was visualized in iTOL[1]. The resulted tree was resolved into five groups based on phylogenetically closely related genomes that grouped in the same cluster. In the second step, these grouped genomes were used as input for MetaCHIP (Song et al., 2019) algorithm to identify donor and recipient genes among the customized user defined phylogenetic groups (**Figures 3A,B**). MetaCHIP identified putative donor and recipient transfer events within the 81 *P. ananatis* strains based on combined similarity and phylogenetic incongruency (Douglas and Langille, 2019).

We used *P. ananatis* strains (PNA_99_3 and PNA_99_14) to identify the genomic islands, and pathogenicity and symbiotic factors in respective genomes as they shared maximum number of HGT events (n = 67). SIGI-HMM and IslandPath-DIMOB prediction methods in IslandViewer 4 was used to predict the genomic islands (Bertelli et al., 2017). GIPSy was used to predict the pathogenicity and symbiotic factors (Soares et al., 2016).

Presence and Absence Variations, Core Genome SNPs, and Whole Genome Multi Locus Sequence Typing (wgMLST) Based Phylogeny

To carry out the phylogenetic analysis of *P. ananatis* strains (n = 81), PAVs and core SNPs were identified using Panseq pipeline (Laing et al., 2010). Core genome threshold of 81 and sequence identity of 85% was used. SNPs identified within the shared core genome regions were obtained in Phylip multiple sequence alignment format. PAVs were identified in a binary Phylip format. Phylogenetic trees using PAVs and SNPs were constructed using RAxML (Stamatakis, 2014). RAxML with rapid bootstrapping returned the best ML tree with support

[1]https://github.com/songweizhi/BioSAK/tree/master/BioSAK_tutorial/Demo_tree_visualization_with_iTol

values after 1000 bootstrap searches. PAVs- and SNPs-based phylogenetic analysis involved 81 sequences with 16,190 and 2112 positions, respectively in the final dataset in each sequence. PAVs along with SNPs were identified using pan-seq to conduct comparative phylogenetic analysis. Further, whole genome multi locus sequence typing (wgMLST) tree was constructed (Liu et al., 2016). To carry out this analysis the assembled contigs file of each of the 81 *P. ananatis* strains were uploaded to PGAdb-builder and a pan-genome allele database was constructed. This database was used to build wgMLSTtree using PGAdb-builder as described earlier (Liu et al., 2016).

Average Nucleotide Identity (ANI)/Average Amino Acid Identity (AAI), Pan-Genome Wide Association and Annotation

Get_homologues perl program was used to estimate the average nucleotide (ANI) and amino acid (AAI) identities of CDSs and the proteins coded by the CDSs among all individual strains of the pan-genome. The resulting ANI (-a 'CDS') and AAI pan-genome matrices obtained using get_homologues.pl were used to plot the heat maps with the shell script plot_matrix_heatmap.sh. For AAI, BLASTP scores were calculated among protein sequences. For association analysis, the pan-genome matrix was used with the phenotyping data using Scoary, a python program (Brynildsrud et al., 2016). Scoary was used to calculate associations among genes in the pan-genome and the red-scale necrosis assay (a qualitative; pathogenic- vs. non-pathogenic association). The output of this program comprised of a list of genes sorted by strength of association with these traits. Genes with a naïve p-Value < 0.05, a Bonferroni p-Value < 0.05 and corrected p-Value (Benjamini-Hochberg) of association < 0.05 were considered significant. The predicted core, soft-core, shell and cloud gene sequences were annotated by conducting a blastx search (E-value $\leq$ 1e-5) against the NR database. The blast output was generated in an XML format that was used as input for Blast2GO to assign gene ontology (GO) terms and assignment of KEGG pathways (Conesa and Götz, 2008). Based on the assigned GO categories each component of pan-genome was categorized into biological process (BP), molecular function (MF) and cellular component (CC).

RESULTS

The *P. ananatis* Genome and Pan-Genome Architecture

Overall, more than 97% of the read data were retained after quality filtering, amounting to 98.6 Gb (**Supplementary Figure 2**). These quality-filtered reads were used to construct draft assemblies. We conducted pan-genome analyses on the draft genomes of the 81 selected *P. ananatis* strains collected from onions, weeds and thrips from different regions of Georgia, United States (**Table 1**). The size of the draft assemblies ranged from 4.7 to 5.8 Mb. The strains PNA 02-12 and PNA 98-11 had minimum and maximum assembly sizes of 4.7 and 5.8 Mb, respectively (**Table 1**). In all the assemblies, the total number of protein-coding sequences ranged from 4,300 to 5,110, and the number of mRNAs ranged from 4,373 to 5,234 (**Table 1**).

The full spectrum of the pan-genome contained 14,452 protein coding genes. Among these, 3,799 genes including 3,153 genes present in all 81 strains (core) and 646 genes present in $\geq$76 genomes <81 genomes defined the soft-core. The largest group of 6,808 genes ($\leq$2 genomes) were cloud and the remaining 3,845 genes belonged to shell ($\leq$3 and =75 genomes) gene cluster (**Supplementary Figures 3A,B**). Cloud genes are considered to be unique genes contributed by each strain of *P. ananatis*. In this study, we found 6,808 genes that are unique to two or less genomes. Details of the number of core and accessory genes contributed by each strain are listed in **Supplementary Table 2**.

The ANI and AAI are widely practiced genome-based characteristics based on pairwise genome comparisons and averaging the sequence identities of shared orthologus genes (amino acids for AAI and nucleotide for ANI) and are determinant for the circumscription of prokaryotic species (Rosselló-Mora, 2005). We, therefore, estimated ANI- and AAI-based all vs. all matrices and constructed clustered tree-based heat-maps. Overall, the ANI and AAI among 81 strains varied from 99.0 to 99.9% and 96.0 to 99.8%, respectively, which confirmed that all strains used in this study belonged to the species *P. ananatis* (**Supplementary Figure 4**).

Phenotyping

All the strains used were identified as *P. ananatis* based on PCR assays, prior to phenotyping. *Pantoea ananatis* strains were classified as pathogenic and non-pathogenic based on the ability to necrotize red onion-scales and clear the anthocyanin pigment. Strains that caused red onion scale necrosis and cleared the anthocyanin pigment were classified as pathogenic (+) and those that did not were identified as non-pathogenic (−) (**Table 1**). Based on the red-scale necrosis assay, 62.9% (n = 51) and 37.1% (n = 30) of the strains were identified as pathogenic and non-pathogenic, respectively. Among the strains that were isolated only from onions (regarded as PNA), 68% (36 of 53 onion strains) were able to cause red-onion scale necrosis (**Table 1**). In contrast, among the strains (n = 28) that were isolated from weeds and thrips (regarded as PANS), 28.5% (8 of 28) and 25% (7 of 28) were able to cause necrosis on the red-onion scale, respectively. A majority of strains (13 of 28) from non-onion sources [weeds: 28.5% (8 of 28) and thrips: 17.8% (5 of 28)] were classified as non-pathogenic as they were not able to cause red-onion scale necrosis (**Table 1**; **Supplementary Figure 1**).

Pantoea ananatis Has an Open Pan-Genome

The pan-genome architecture of the 81 *P. ananatis* genomes analyzed is characterized in **Figure 1**. We used the exponential decay model of Willenbrock et al. (2007) that fitted the core gene clusters generated using the OMCL algorithm, which predicted a theoretical core genome of 2,935 genes. In addition, the core genome was not continuous because of the draft assemblies (not the finished sequence assembly) used in this study (**Figure 1A**).

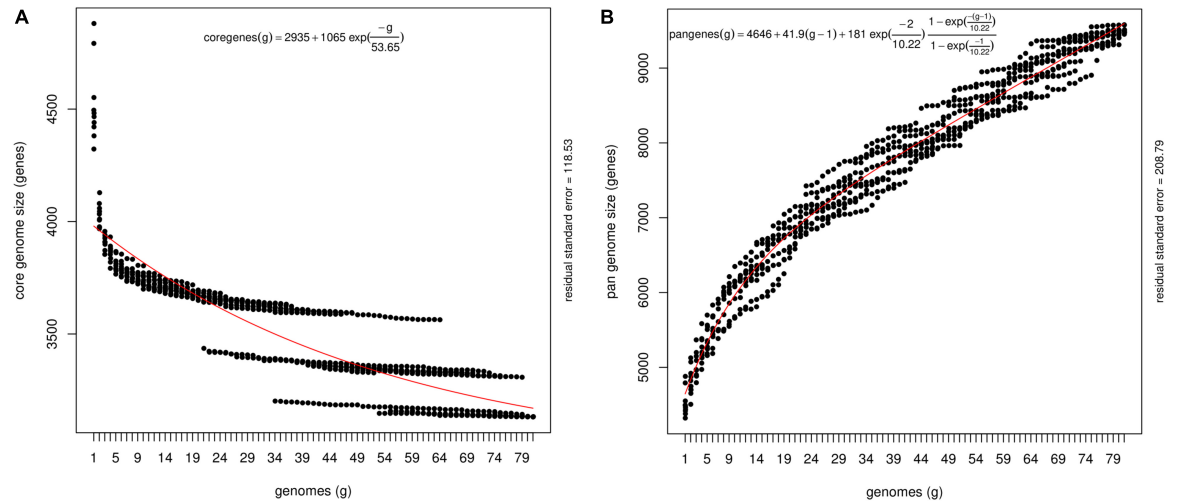


FIGURE 1 | Theoretical estimation of the core and pan-genome sizes based on the exponential decay model. **(A)** Estimation of core genome size based on Willenbrock model fit to OMCL clusters. The equation for the core gene estimation include theoretical estimate of the number of core genes (2395), free parameters for amplitude of exponential decay (1065) and decay constant (53.65). Decay constant measures the speed at which the core genes converges to its asymptotic value. Size of the core genome is shown here as '2395' for the number of genomes $\rightarrow \infty$, where number of genomes display continuous extrapolation of the number of strains considered. The core genome, that is, the number of conserved genes present in all genomes, was estimated by fitting an exponential decay function by non-linear least squares. **(B)** Estimation of pan-genome size based on Tettelin model fit to OMCL clusters. The equation for the theoretical estimation of pan-genome include average number of genes per sequenced genome (4646). The other components in the equation depict a linear term [41.9 (g-1)] that represents extrapolated angle of growth of the size of pan-genome size without introducing new free parameters, amplitude of exponential decay (181), and decay constant (10.88).

In order to check the openness/closeness of the pan-genome, the theoretical estimation of pan-genome size was carried out using an exponential model of Tettelin et al. (2005), which was fitted to the OMCL accessory gene clusters. The pan-genome samples appeared to converge to linear growth with >10,000 genes, with ∼50 new genes being added on average to the pan-genome with each new *P. ananatis* genome sequenced (**Figure 1B**). Our results indicate that the pan-genome of *P. ananatis* is open.

Analysis of *P. ananatis* Strains Revealed Pathogenic and Non-pathogenic Differentiation

We used 51 onion pathogenic and 30 non-pathogenic strains as identified using the red-scale necrosis assay. The pathogenic strains displayed typical scale-clearing phenotype as opposed to no symptoms observed in case of non-pathogenic strains (**Supplementary Figure 1**). Based on pan-genome enabled phylogenetic-tree, strains were distributed in six clusters. Further, we also constructed dendrograms using only shell and cloud PAVs and compared the clustering of pathogenic- and non-pathogenic strains. A soft-core genome based (including 3153 conserved core genes and 646 genes present in ≥76 genomes) dendrogram was constructed to understand the relationship among the *P. ananatis* strains (**Figure 2**). Soft-core genes resulted in six clusters with cluster I, II and III mainly consisting of non-pathogenic (colored green in figure) strains. Cluster III contained eight pathogenic (colored red in figure) out of total 24 strains. Cluster IV comprised of all pathogenic strains except for the presence of one non-pathogenic strain. Cluster V and VI comprised of mainly pathogenic strains with one exception (one non-pathogenic strain was present in cluster V) (**Figure 2**).

Clustering using shell PAVs resulted in five broad clusters; cluster I with all 7 pathogenic strains, cluster II- with five non-pathogenic (green) strains, cluster III with all 23 pathogenic (red), cluster IV with 17 pathogenic (red) out of total 25 strains and cluster V with four pathogenic (red) out of the total 21 strains (**Supplementary Figure 5**). Cloud PAVs showed a mixed pattern of clustering of pathogenic- and non-pathogenic strains (**Supplementary Figure 6**).

Identification of Horizontal Gene Transfer (HGT), Genomic Islands, and Pathogenicity and Symbiotic Factors in *P. ananatis*

Phylogenetic analyses based on multiple sequence alignment resulted in five predominant clusters (cluster A-E) (**Figure 3**). Groupings were assigned to each genome based on the phylogenetic classification. Further, based on the assigned groupings, HGT study was conducted to assess the transfer of genes from putative donor strains to putative recipient strains (**Figure 3** and **Supplementary Table 3**) to the putative recipient strains. The HGT events in *P. ananatis* were found to be extensive. We found 1,182 HGT events among the 77 strains with 68 putative donor and 70 recipient strains (**Supplementary Table 3**). A maximum of 114 putative gene transfers occurred from PNA_98_12 (cluster A) followed by 106 from PNA_99_3 (cluster B), 61 from PANS_99_9 (cluster B), and 60 from PNA_07_22 (cluster E) (**Figure 3** and **Supplementary Table 3**).

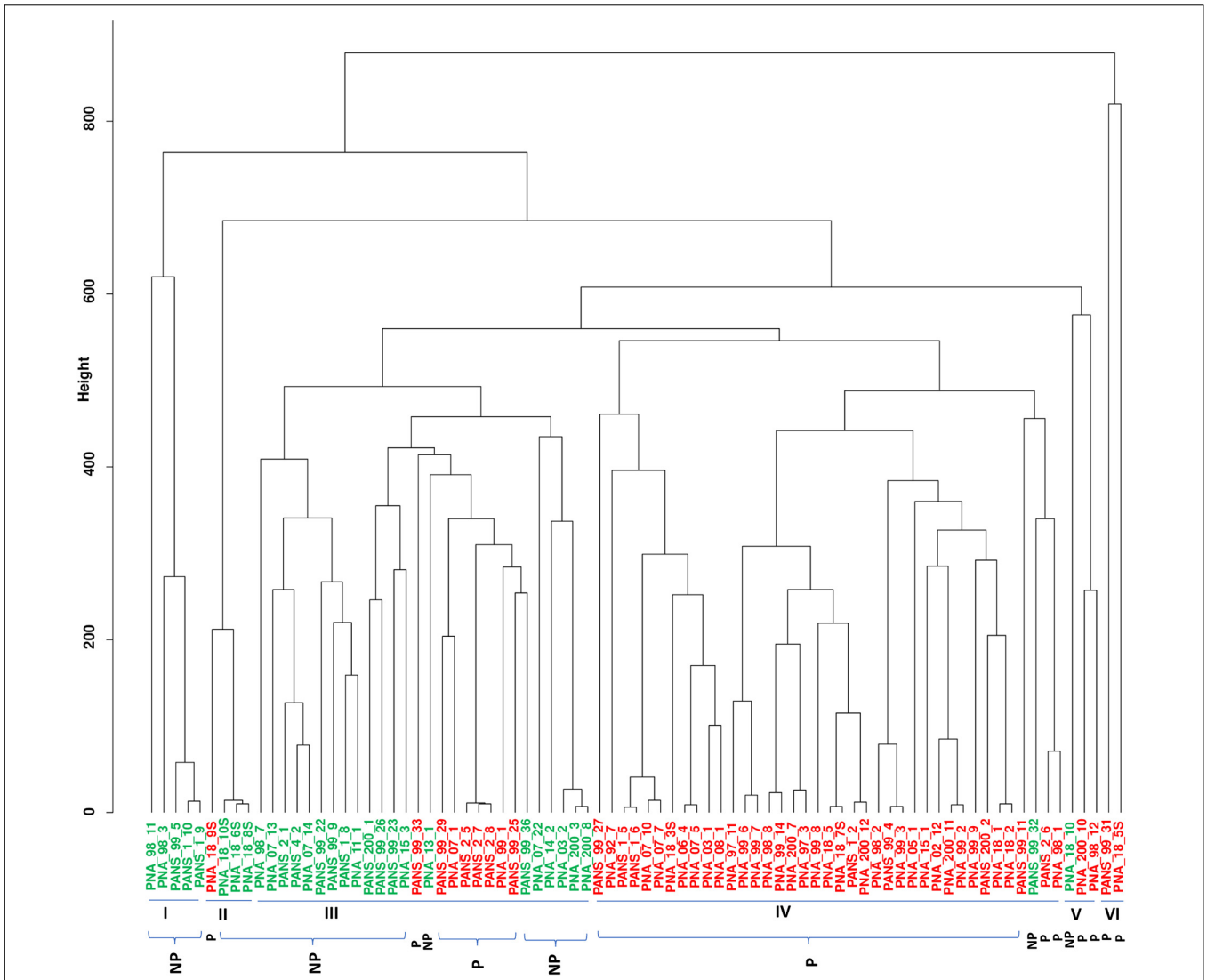


FIGURE 2 | Dendrogram of 81 pathogenic and non-pathogenic strains of *Pantoea ananatis* based on the soft-core genes. Dendrogram to detect outliers with cut-off height is shown in the y-axis. The height scale is the distance metrics between the clusters, and strains displaying low merging height are highly related. Strains highlighted in green are non-pathogenic (NP) and ones in red are pathogenic (P). Bacterial strains of onion and non-onion origin are designated with a prefix as "PNA" and "PANS", respectively.

Rest of the HGTs between genomes ranged from 1 gene to 30 genes (**Supplementary Table 3**). Three strains that received the greatest number of gene transfers were: PNA_98_3 (cluster C) that received 125 gene transfers, followed by 84 genes received by PNA_99_14 (cluster E) and 64 genes received by PANS_99_4 (cluster B) (**Figure 3** and **Supplementary Table 3**). Highest number of 67 HGTs occurred from PNA_99_3 to PNA_99_14 followed by 32 HGTs from PNA_98_12 to PNA_98_3 and 31 HGTs from PNA_99_3 to PNA_98_3. Notably, the maximum number of HGT events occurred between the donor and the recipient strains that were identified in the years 1997, 1998 and 1999. Out of the total 114 genes transferred by PNA 98_12, a maximum of 32 were received by PNA 98_3 followed by PNA_97_11 (21 genes). Similarly, for the second highest donor PNA_99_3, out of 106 HGT events, 67 were transferred to PNA_99_14 and 31 to PNA_98_3. However, PNA_07_22 with third most number (60) of donor genes had PANS_99_4 as the recipient with most number (31) of HGTs.

Four genes of the HiVir/PASVIL cluster were found to be horizontally transferred between the strains. These genes were flavin-dependent oxidoreductase (*pavC*), SAM-dependent methyltransferase (*pavG*), GNAT family N-acetyltransferase (*pavH*) and MFS transporter (*pavJ*). Flavin-dependent oxidoreductase (*pavC*) was transferred from PNA_200_3 (00224) and PNA_200_8 (00224) to PNA_98_11 (01052), *pavG* was transferred from PNA_99_1 (01588) and PNA_99_8 (02166) to PNA_05_1 (01633) and PNA_99_14 (02553) respectively. In case of *pavH*, 25 HGTs were identified and for *pavJ,* a total of 16 HGTs were identified. Another homologous HiVir/PASVIL gene involved in analogous tricarboxylic acid

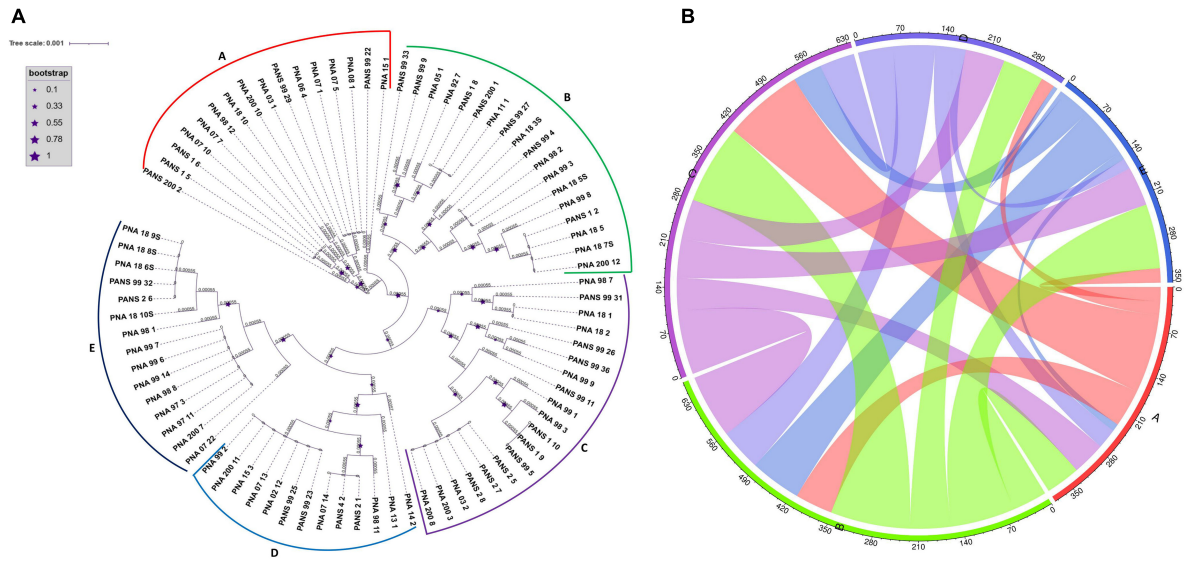


FIGURE 3 | Phylogeny based horizontal gene transfer among 81 *Pantoea ananatis* strains. **(A)** Phylogenetic tree of 81 *P. ananatis* strains based on multiple sequence alignment of 120 single copy genes. Phylogenetic tree resulted in five clusters **(A–E)**. Size of stars represents the bootstrap values in that order. **(B)** Predicted gene flow within the five phylogenetic clusters of *P. ananatis*. Bands connect donors and recipients, with the width of the band correlating to the number of HGTs and the color corresponding to the donors. Numbers on the circumference of circos plot represent the number of genes that undergo horizontal gene transfers. The alphabets on five arcs **(A–E)** of circos represent the five phylogenetic clusters of *P. ananatis* strains.

cycle reaction was homocitrate synthase that was involved in HGT between PNA_99_3 (01851) and PNA_99_12 (02485) (**Supplementary Table 4**).

We further investigated the *alt* cluster genes to be involved in HGT. Seven *alt* genes namely, *altA* (alkene reductase), *altB* (SDR family oxidoreductase), *altC* (DsbA family oxidoreductase), *altE* (carboxymuconolactone decarboxylase family protein), *altH* (DNMT family transporter), *altI* (aminotransferase) and *altR* (TetR/AcrR family transcriptional regulator) were horizontally transferred between strains. Six HGTs were identified for *altH,* five for *altB,* two each for *altE* and *altI* and one each *altA*, *altE* and *altR* (**Supplementary Table 4**).

We used *P. ananatis* strains, PNA_99_3 and PNA_99_14 to identify the genomic islands, and pathogenicity and symbiotic factors in respective genomes. These two strains were selected as they shared maximum number of HGT events (n = 67). In PNA_99_3, SGI-HMM and IslandPath-DIMOB predicted a total of 25 (length = 4.3 kb) and 14 (length = 197.2 kb) genomic islands, respectively (**Supplementary Tables 5-8**). In case of PNA_99_14, SGI-HMM and DIMOB predicted 21 (length = 4.06 kb) and 13 (length = 210.6 kb) genomic islands, respectively (**Supplementary Tables 9-12**). A total of 380 pathogenicity and 150 symbiotic factors were identified in PNA_99_3 (**Supplementary Tables 5-8**) and 410 pathogenicity and 169 symbiotic factors were identified in PNA_99_14 (**Supplementary Tables 9-12**). Among several genes present in the genomic islands and, pathogenicity and symbiotic factors, HiVir and *alt* cluster genes were identified. HiVir cluster genes that were identified include 3-isopropylmalate dehydratase large sub-unit (*pavE*), 3-isopropylmalate dehydratase small sub-unit (*pavF*), nitrilotriacetate monooxygenase (flavin monooxygenase) (*pavC*), SAM dependent methyltransferase (*pavG*), MFS transporter (*pavJ*) phosphoenolpyruvate mutase (*pepM*), as a part of genomic islands in both PNA_99_3 and PNA_99_14 (**Supplementary Tables 5, 6, 9, 10**). Similarly, two *alt* cluster genes were identified in the predicted pathogenicity factors. The *alt* genes identified were tetR family transcriptional regulator (*altR*) in PNA_99_14 (**Supplementary Table 11**) and thioredoxin reductase (*altD*) in PNA_99_3 and PNA_99_14 (**Supplementary Tables 7,11**).

Presence and Absence Variations, Core Genome SNPs, and Whole Genome Multi Locus Sequence Typing (wgMLST) Based Phylogeny

The PAVs were identified using Pan-seq pipeline along with SNPs to carry out a comparative phylogenetic analysis based on PAVs and SNPs. PAVs-based phylogeny as compared to core-SNPs-based phylogeny distinguished the pathogenic strains (P) from the non-pathogenic (NP) strains (**Supplementary Figure 7**). Considering the strictest constrain of 81 genomes, SNPs were identified that were present in all genomes used in this study. The 51 pathogenic and 30 non-pathogenic strains were distributed in several groups with both pathogenic and non-pathogenic strains grouped together (**Supplementary Figure 7A**). However, PAVs-based phylogenetic tree showed that the 30 non-pathogenic strains clustered separately from the pathogenic strains in three groups of 16, five and nine strains. Largely, the strains were separated out based on their pathogenicity in PAVs-based phylogeny. SNPs- and PAVs-based phylogenetic analysis identified PNA_98_11 (NP), PNA_98_12 (P), PANS_99_27 (P), and PNA_92_7 (P) as the most diverse strain based on its

branch length. PAVs-based phylogeny identified PNA_98_1 (P) and PNA_18_10 (NP) as other diverse strains that were not as diverse with SNPs-based phylogenetic analysis (**Supplementary Figures 7A,B**). We also used wgMLST based approach to construct a phylogenomic tree using the assembled scaffolds of 81 *P. ananatis* strains. In this approach first, a pan-genome allele database for *P. ananatis* strains were established using PGAdb-builder. The database consisted of 3,370 alleles in 81 *P. ananatis* strains. Using this allelic distribution, a genetic relatedness tree was constructed (**Figure 4**). There were six different clusters of non-pathogenic (NP) strains that were grouped separately from the pathogenic (P) strains. However, two non-pathogenic strains (PANS_99_36 and PANS_99_32) grouped arbitrarily with pathogenic strains (**Figure 4**).

Pan-Genome-Wide Association Study

Presence and absence of each candidate genes in the accessory genomes was screened and scored. Further, Scoary was used to identify genes that were significantly associated with red-onion scale necrosis (indicative of pathogenicity to onion) using the 81 pathogenic- and non-pathogenic *P. ananatis* strains. Scoary predicted 42 genes, including the 14 HiVir/PASVIL cluster genes (*hvaA*, *pepM, pavC-pavN*) that are shown to be associated strongly with red-onion scale necrosis or pathogenicity to onion based on stringent *p*-Values (**Supplementary Table 13**). Earlier the same cluster of HiVir genes was predicted to be responsible for onion pathogenicity (Asselin et al., 2018; Takikawa and Kubota, 2018). A total of 28 genes were identified outside the HiVir/PASVIL cluster. Eight of the 28 genes are annotated and the remaining 20 are hypothetical. Annotated functions of eight genes include: site-specific tyrosine recombinase (*xerD6*), pyridoxal-4-dehydrogenase (*pld_1*), murein tetrapeptide carboxypeptidase, sporulation initiation inhibitor protein (*soj3*), N-acetylmuramoyl-L-alanine amidase (*amiD_4*), conjugal transfer protein (*traR_3*), adhesin/invasin TibA autotransporter (*tibA*), and helix turn helix-type transcriptional regulator (*dmlR_11*) (**Supplementary Table 11**). Out of the 20 hypothetical genes, 13 genes (four, five and four genes) could possibly be constituting separate operons in *P. ananatis* because of their contiguity (**Supplementary Table 13** highlighted in colors blue, green and red).

Out of 14 HiVir/PASVIL cluster genes associated with pathogenicity to onion, two are annotated as hypothetical (*hvaA* and *pavK*) and the remaining 12 are annotated with functions in metabolite production (**Table 2**). These 12 genes include *pepM*, coding for phosphonopyruvate mutase; *pavC*, that encodes nitrilotriacetate monooxygenase component A (catalyzes plant-derived aromatic compounds); *pavD* for homocitrate synthase; two genes associated with leucine biosynthesis associated metabolites, including *pavE* for 3-isopropylmalate dehydratase large subunit and *pavF* for 3-isopropylmalate dehydratase small subunit; *pavG*, which codes for SAM dependent methyltransferase; *pavH* for N-acetyltransferase; *pavI* for the ATP-grasp domain containing protein; *pavJ* for MFS transporter; *pavL* which encodes flavin reductase; *pavM* for carboxylate-amine ligase; and *pavN* as transposase.

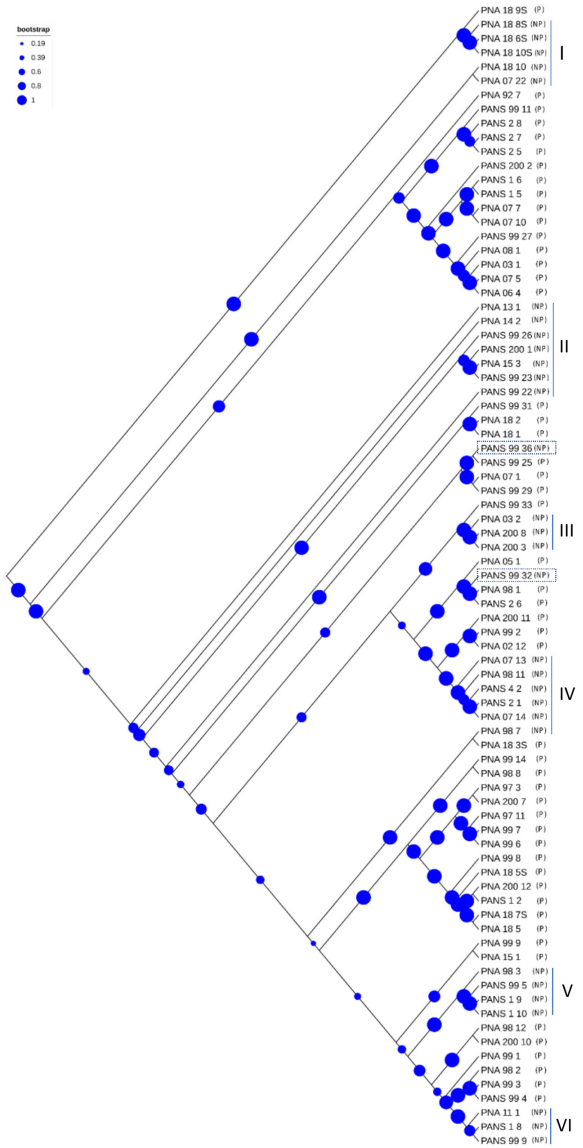


FIGURE 4 | Phylogenetic tree based on whole genome multi-locus sequence typing (WgMLST). Dendrogram for 81 *Pantoea ananatis* strains was constructed using assembled genome contigs. The letter 'P' represents pathogenic and 'NP' represents non-pathogenic strains. Strains labeled as 'PNA' represent onion as a source of isolation whereas strains labeled as 'PANS' represent source of isolation other than onion. Size of blue circles represent the bootstrap values. Six 'NP' clusters that grouped separately from the 'P' strains are numbered as I to VI and the two NP strains that grouped with 'P' strains are highlighted in rectangular blocks.

Annotation

Prokka cannot distinguish between complete and truncated genes during annotation, which may potentially result in some genes to be wrongly annotated. Hence, once the pan-genome was defined, we carried out a blast-based annotations of sequences that constituted the core and accessory genomes. We annotated core, soft-core, shell and cloud genes. A total of 2,705 core, 3,293 soft-core, 2,058 shell and 3,503 cloud

TABLE 2 | Highest-ranking genes associated with red-onion scale necrosis and corresponding statistics in *Pantoea ananatis*.

Gene	Annotation	Naive_p*	Bonferroni _p*	Benjamini _H_p*
hvaA	Hypothetical	2.13E-14	2.41E-10	2.68E-11
pepM	Phosphoenolpyruvate mutase	4.53E-13	5.12E-09	1.46E-10
pavC	Flavin-dependent monooxygenase	2.13E-14	2.41E-10	2.68E-11
pavD	Phosphonomethyl malate synthase	2.13E-14	2.41E-10	2.68E-11
pavE	3-isopropylmalate dehydratase large subunit	4.53E-13	5.12E-09	1.46E-10
pavF	3-isopropylmalate dehydratase small subunit	2.13E-14	2.41E-10	2.68E-11
pavG	SAM dependent methyltransferase	2.58E-15	2.91E-11	9.70E-12
pavH	GNAT family N-acetyltransferase	1.57E-13	1.78E-09	1.37E-10
pavI	ATP-grasp domain containing protein	2.13E-14	2.41E-10	2.68E-11
pavJ	MFS transporter	2.13E-14	2.41E-10	2.68E-11
pavK	Hypothetical	2.13E-14	2.41E-10	2.68E-11
pavL	Flavin reductase	1.57E-13	1.78E-09	1.37E-10
pavM	ATP-grasp domain containing protein	5.79E-14	6.54E-10	6.54E-11
pavN	Transposases	1.57E-13	1.78E-09	1.37E-10

**p-Value.*

sequences were annotated successfully, mapped and assigned at least one gene ontology (GO) id and GO slim category (**Supplementary Tables 14 - 17**). GO analyses of the top terms revealed that metabolic process represented the most abundant category, followed by cellular processes under BP (**Supplementary Figure 8**). Under BP, the cellular amino acid metabolic process was specific to core genes and not present in soft-core, shell and cloud genes (**Supplementary Figure 8**). For genes to which MFs could be assigned, catalytic activity was the most abundant category, followed by binding. Transmembrane transporter and oxidoreductase activity; however, was not observed in the shell genes. GO analyses showed that cellular anatomical entities were represented the most abundantly under the CC category (**Supplementary Figure 8**).

High Virulence/PASVIL and *alt* Genes Presence and Absence

Overall, 80.3% (41 of 51) of the pathogenic strains had both HiVir/PASVIL (n = 14 genes) and *alt* (n = 11 genes) clusters. However, none of the non-pathogenic strains had both of these gene clusters. Alternatively, none of the pathogenic strains showed the absence of both gene clusters (**Figure 5**).

The HiVir/PASVIL cluster was conserved in 98% (50 of 51) of the pathogenic strains. The pathogenic strain (PNA_18_9s) showed a partial loss of HiVir/PASVIL genes (*pepM, pavE, pavJ* and *pavK* genes). Absence of these genes could either be due to assembly artifacts or attributed to inconsistent and weak phenotypes (negligible red-scale clearing). Among the 30 non-pathogenic strains, 73.3% (n = 22) lacked a complete HiVir/PASVIL cluster and 6.6% (n = 2) of the strains showed the presence of only a subset (one or more) of the genes in the HiVir/PASVIL cluster. Interestingly, 20% (n = 6) of the non-pathogenic strains possessed a conserved complete HiVir/PASVIL cluster (**Figure 5**).

Among the pathogenic strains, the *alt* cluster was conserved in 80.3% (n = 41), absent in 15.6% (n = 8) strains and partially present in two strains (with just one gene present). However, among the non-pathogenic strains, the *alt* cluster was present in 33.3% (n = 10), absent in 43.3% (n = 13) and partially present (one or two out of 11 *alt* genes present) in 23.3% (n = 7) of the strains. Since only one or two genes in the *alt* gene cluster were present in the seven strains (PANS_2_1, PANS_4_2, PANS_99_32, PNA_7_13, PNA_7_14, PNA_98_11, PNA_98_7), these strains were considered as negative for the presence of the conserved *alt* cluster (**Figure 5**). As a result, 66.6% (n = 20) of the non-pathogenic strains lacked the *alt* cluster.

DISCUSSION

Pan-Genome of *P. ananatis*

Building a pan-genome is important as it tremendously aids in gene discovery and helps in understanding the genome architecture of a species. A pan-genome represents the genomic repertoire of a species that can address questions related to varied phenotypes exhibited by individuals of that species. Current pan-genome study of *P. ananatis* that causes center rot in onions has helped in defining a conserved core genome and a dynamic accessory genome represented by PAVs. Pan-genome analysis of 81 genomes (n = 51 pathogenic, n = 30 non-pathogenic) revealed that the core genome was stabilized with 3,153 gene clusters while the pan-genome expanded continuously as a result of the addition of gene clusters indicating an open-pangenome. Earlier pan-genome studies identified similar but a little higher number of core genes while using a much smaller number of *P. ananatis* genomes (De Maayer et al., 2014; Sheibani-Tezerji et al., 2015; Stice et al., 2018). The ANI of $\geq$ 95% is a benchmark to classify organisms of the same species (Richter and Rosselló-Móra, 2009; Sangal et al., 2016), whereas, genomes of organisms with ANI values $<$ 94% suggests that the organisms belong to different species (Sangal et al., 2016). In this study, ANI ranged from 99.0-99.9% and AAI ranged from 96-99%, which not only suggested low core genome diversity among the 81 *P. ananatis* strains but also confirmed that the strains belonged to *P. ananatis* species.

Phylogenetic Analysis and Identification of HGTs, Genomic Islands, Pathogenicity and Symbiotic Factors in *P. ananatis*

Earlier, MLSA and rep-PCR assays showed limited genetic diversity, despite high phenotypic variation, among 50 *P. ananatis* strains (Stice et al., 2018). In the same study, the authors demonstrated that PAVs from the pan-genome

FIGURE 5 | Pictorial representation of presence and absence of HiVir/PASVIL genes associated with red onion scale necrosis and *alt* genes associated with colonization in necrotic zone in pathogenic- and non-pathogenic strains of *Pantoea ananatis*. Presence and absence matrices of genes in the two gene clusters are aligned with each *P. ananatis* strain, '+' represent the pathogenic and '–' represent the non-pathogenic strains.

analysis of 10 *P. ananatis* strains separated the pathogenic strains from the non-pathogenic strains, which was not observed when the core genome was used. Phylogenetic studies conducted so far in *P. ananatis* have relied only on PAVs, which are mainly horizontally acquired (Welch et al., 2002) contrary to the vertically inherited SNPs (Straub et al., 2021). In the current study, we therefore conducted a comparative phylogenetic study using core genome SNPs, PAVs and wgMLST approach. Core

SNPs are relied to infer phylogeny as they represent the vertically inherited portion of the genome in the population (Didelot and Wilson, 2015; Straub et al., 2021). It was however observed that pathogenic strains clustered together with the non-pathogenic strains (in some groups in the phylogenetic tree) based on core SNPs. We expected core genome SNP variations to distinguish the pathogenic and non-pathogenic strains based on their vertically inherited evolutionary history. It could be argued that presence of homologous recombination and HGT events in the core genome could have distorted the phylogenetic relationship thus resulting in an unexpected phylogenetic classification of pathogenic and non-pathogenic strains. However, the possibility of HGT based distortion can be ruled out because our core SNP analysis was not based on a single reference genome. Use of a reference genome-based SNP calling approach can result in SNPs that are present in core genome of some strains that have horizontally acquired genes similar to reference genome but not present in other strains that do not have such horizontally acquired genes (Straub et al., 2021). In the current study, the core genome region of 81 strains was aligned (non-reference based) to identify the variant and invariant sites in all the strains. Pathogenicity phenotype in *P. ananatis* is therefore concluded to be a result of PAVs of pathogenicity-related genes in strains, which may be a result of various HGT events during bacterial evolution more than the vertically inherited SNPs.

Whole genome multi locus sequence typing (wgMLST) (Maiden et al., 2013) is an extended concept of the traditional MLST (Aanensen and Spratt, 2005) and is considered as an ideal approach to sort out WGS data and generate genetic layouts that are portable and comparable among laboratories (Liu et al., 2016). Using wgMLST, pathogenic and non-pathogenic *P. ananatis* strains were resolved. The wgMLST approach is based on the presence and absence of an allele. This approach in a way is complementary to phylogenomic analysis based on PAVs. Therefore, it further supports our findings that PAVs alone could provide a better phylogenomic resolution of pathogenic and non-pathogenic strains than SNPs.

Unrelated bacteria sharing a common environment are known to engage in frequent HGT (Smillie et al., 2011). HGT and gene loss are the key processes in bacterial evolution. HGT primarily occurs through lateral gene transfer, which drives both phenotyping changes and subsequent adaptation when the acquired genes confer new traits resulting in diversification of that lineage into a new environment (Treangen and Rocha, 2011; Nowell et al., 2014). The dynamics of genome fluctuation in *P. ananatis* can be attributed to HGT. Further, existence of pathogenic non-onion host strains (designated as PANS) could be the result of acquired pathogenicity from the pathogenic *P. ananatis* through HGT. HGT can increase the genetic variability if the donor is dispersed from a foreign population or is distantly related, conversely, HGT can homogenize a population in terms of gene content if it spreads the genetic material throughout the population (Van Rossum et al., 2020). The strain, PNA_98_12 in our study possessed maximum branch length (PAV based phylogeny) and also resulted in highest number of HGT events suggesting that the strains might possess novel gene content, which is distributed to the population in order to homogenize the population. Other strains with similar longer branch lengths, PNA_98_11, PNA_98_1, PANS_99_27 and PNA_18_10 could also possibly be involved in similar phenomenon of homogenizing the population based on lateral transfer of genes via HGT implying soft selective sweeps (multiple beneficial alleles at a locus gain prevalence, replacing standing genetic variation in the population). However, further studies are required to confirm the occurrence in *P. ananatis*.

Genomic islands, also known as clusters of bacterial or archaeal genes are of probable horizontal origin and are of interest as they contain the genes for virulence (pathogenicity), symbiosis and metabolism (Bertelli et al., 2019). We could predict the genomic islands, and pathogenicity and symbiotic factors for two *P. ananatis* strains (PNA_99_3 and PNA_99_14) that were involved in the maximum number of HGT events. We found that some of the pathogenic genes that were identified to play a significant role in causing red-scale necrosis of onions using pan-GWAS were also the part of the predicted genomic islands and pathogenicity factors.

Pan-GWAS of *P. ananatis*

Bacterial phenotypes, in general, can be linked to the presence or absence of genes that are inherited through either descent or lateral gene transfer (Tettelin et al., 2005). We conducted pan-GWAS analysis to predict and associate genes related to pathogenicity in onion. Pan-genome PAVs were utilized to associate onion pathogenicity phenotypes (determined using a red-onion scale necrosis assay). The 14 strongly associated genes that were identified as a part of the HiVir/PASVIL cluster were annotated as *hvaA*, *pepM*, *pavC-N* (Takikawa and Kubota, 2018). These associated genes coded for phosphonate metabolism, metabolism of plant-derived aromatic compounds, monooxygenases, a methyltransferase, leucine biosynthesis and an L-amino acid ligase. Association of HiVir/PASVIL genes using this pan-genome *in silico* approach corroborated earlier findings of the roles of these genes in *P. ananatis*-pathogenicity in onion (Asselin et al., 2018; Polidore et al., 2021). Rest of the 28 associated genes were regarded as 'novel' as these were not identified in any of the previous pan-genome investigations. The eight annotated genes (part of 28 novel genes) code for: murein tetrapeptide carboxypeptidase, which is involved in the peptidoglycan recycling pathway (Perna et al., 2001); pyridoxal-4-dehydrogenase (*pld_1*) that belongs to aldo/keto reductase family and is involved in synthesis of 4-pyridoxate from pyridoxal (Yokochi et al., 2004); chromosome partitioning ATPase (coded by *soj3*) involved in the genome maintenance (Charaka and Misra, 2012); XerD protein is a site-specific tyrosine recombinase involved in cell division (Cerro et al., 2013); N-acetylmuramoyl-L-alanine amidase (coded by *amiD*) that belongs to glycosyl hydrolase family and is involved in cell wall macromolecule and peptidoglycan catabolic process (Wilkes et al., 2010); conjugal transfer protein (coded by *traR*) with zinc ion binding ability (Chen et al., 2019); the tibA adhesin, which induces bacterial aggregation and biofilm formation (Sherlock et al., 2005); and LysR transcriptional regulatory family protein involved in DNA-binding transcription factor activity (coded by *dmlR*) (Moura et al., 2017). Further

functional analysis is warranted to understand their roles in onion pathogenicity.

Comparative Genomics of HiVir/PASVIL Cluster and Role of Individual Genes in the Cluster

Comparative genomic analysis showed a trend in the presence or absence of complete/conserved HiVir/PASVIL cluster genes in pathogenic vs. non-pathogenic *P. ananatis* strains. Interestingly, 20% ($n = 6$) of the non-pathogenic strains possessed a conserved complete HiVir/PASVIL cluster. If the presence of a complete cluster is correlated with onion pathogenicity, then it is difficult to explain the presence of a complete cluster of genes in these non-pathogenic strains. It is possible that the HiVir/PASVIL genes in the cluster in non-pathogenic strains are not expressed or are non-functional, which may require further investigation and confirmation.

Phosphoenolpyruvate mutase (*pepM*) is involved in phosphonate biosynthesis. Organophosphonates are synthesized as secondary metabolites in certain prokaryotes to function as antibiotics, and can have specialized roles in pathogenesis or signaling (Hilderbrand, 1983). Phosphonate metabolites are derived from phosphonopyruvate, which in turn is formed from phosphoenolpyruvate (PEP) by the action of PEP mutase (PepM). Asselin et al. (Asselin et al., 2018) identified a *pepM* gene as the first pathogenicity factor associated with the fitness of *P. ananatis* as well as with symptom development in infected onion leaves and bulbs. Deletion of *pepM* or inactivation of *pavJ* gene resulted in loss of the ability to cause lesions on onion foliage and bulbs. Furthermore, growth of the deletion mutant in onion leaves was significantly reduced compared with the wild-type *P. ananatis* strain. This pan-genome *in silico* study corroborated the association of *pepM* gene with onion pathogenicity, using a diverse panel of *P. ananatis* strains. The *pepM* gene was present in 50 of 51 pathogenic strains, with the exception of a strain PNA 18-9 s. This strain also lacked *pavE*, *pavJ*, and *pavK*. If it is not an assembly artifact, the absence of *pepM* along with four other genes in the HiVir/PASVIL cluster in this strain could be the reason for a compromised red scale necrosis phenotype (weak pathogenicity). This observation also indicated the presence of a potential alternative pathogenicity factor than *pepM*, which requires further investigation. For the non-pathogenic strains of *P. ananatis*, *pepM* gene was absent in 23 of 30 strains. Despite the presence of *pepM* gene and, in some cases, the entire HiVir/PASVIL cluster (six of 30 strains), these strains displayed a non-pathogenic phenotype. These observations suggest that these genes may be non-functional in these strains, which warrants further research.

A monooxygenase and a flavin reductase enzyme belonging to the two-component non-hemeflavin-diffusible monooxygenase (TC-FDM) family were found to be associated with onion pathogenicity using pan-GWAS study. The monooxygenase and the reductase associated are nitrilotriacetate monooxygenase coded by *nta*A (similar to *pavC* in the HiVir/PASVIL cluster of *P. ananatis*) and flavin reductase, a flavin:NADH oxidoreductase component of 4-hydroxyphenylacetate (4-HPA) 3-monooxygenase coded by *hpa*C (similar to *pavL* in the HiVir/PASVIL cluster of *P. ananatis*). Nitrilotriacetate monooxygenase was reported previously in the genomic region referred as WHOP (woody host of *Pseudomonas* spp.) in a *Pseudomonas syringae* complex (Caballo-Ponce et al., 2017b). This region is associated with strains of *P. syringae* that infect woody host plants, and is absent in strains infecting herbaceous tissues. This gene, along with other genes present in the WHOP region, is responsible for the fitness and virulence of *Pseudomomas savastanoi* pv. *savastanoi* in woody olive trees, but not in non-woody olive trees (Caballo-Ponce et al., 2017a; Caballo-Ponce et al., 2017b). Nitrilotriacetate monooxygenase is known to catabolize plant-derived aromatic compounds and help bacteria to adapt to woody host tissues (Ramos et al., 2012). On the contrary, *P. ananatis* colonize foliar and bulb tissue in onion, which are non-woody, therefore, it was intriguing to find this gene associated with pathogenicity in onion, an herbaceous plant.

The HiVir/PASVIL gene *pavG* in *P. ananatis* has an annotated function for a class-I S-adenosyl-L-methionine (SAM)-dependent methyltransferase. We presume that *pavG* is responsible for the esterification of phosphonates synthesized in *P. ananatis*, led by *pepM,* based on the fact that the di-anionic form of phosphonates interferes with the metabolic intermediates and carboxylates of antibacterial compounds (Metcalf and van der Donk, 2009; Lee et al., 2010; Yu et al., 2013). To counteract this problem, microbes either synthesize phosphinites (with a double bond between the C and P instead of a single bond) or carry out esterification of phosphonates. Phosphonate esterification appears to be an obvious mechanism operational in *P. ananatis* because of the presence of *pavG* in the HiVir/PASVIL cluster. SAM dependent O-methyltransferase has been shown to methylate a variety of phosphonates (1-hydroxyethylphosphonate, 1, 2-dihydroxyethylphosphonate, and acetyl-1- aminoethylphosphonate) (Lee et al., 2010). Therefore, there is a high possibility of involvement of SAM methyltransferase in methylation of the phosphonate produced in *P. ananatis*. Further studies are required to characterize the type of phosphonate and its methylation in order to understand the mechanism of SAM methyltransferase and *pepM* in causing red-scale necrosis. Another role that *pavG* could be playing is methylation of other HiVir/PASVIL genes that renders them inactive despite their presence in the cluster. We hypothesize that the inactivity of HiVir/PASVIL genes may be due to methylation of genes carried out by SAM dependent methyltransferase in non-pathogenic strains of *P. ananatis*, implying a secondary role of *pavG* in strains non-pathogenic to onion. Methylation profiling will help evaluate this hypothesis.

The HiVir/PASVIL gene *pavI* is similar to RizA an L-amino-acid ligase (LAL) from *Bacillus subtilis* that participates in the biosynthesis of rhizocticin, a phosphonate oligopeptide antibiotic and possess L-arginyl-L-2-amino-5-phosphono-3-cis-pentenoic acid (Kino et al., 2009). Although, the functional role of *pavI* is yet to be characterized in *P. ananatis*, it may play a role in the formation of anti-microbial secondary metabolites of "phosphonate derivatives." LAL is a member of the ATP-dependent carboxylate–amine/thiol ligase superfamily

(Galperin and Koonin, 1997), and catalyzes the ligation reaction, which involves an aminoacyl-phosphate intermediate, in an ATP-dependent manner (Fan et al., 1995). LALs contain the ATP-grasp fold, which is composed of three conserved domains referred to as the A-domain (N-terminal domain), the B-domain (central domain) and the C-domain (C-terminal domain). These three domains commonly grasp the ATP molecule, and also provide binding sites for the Mg2 + ion and the amino-acid substrate (Kagawa et al., 2015).

The pan-GWAS approach used in this study did not associate the genes in *alt* cluster with the onion pathogenic phenotype using the red-scale necrosis assay. This may be because of the type of phenotyping assay utilized in this study. The red-scale necrotic assay has been shown to be induced by the HiVir/PASVIL cluster (Stice et al., 2020). However, the *alt* cluster comes into play after the onset of necrosis, when endogenous antimicrobial sulfur compounds are produced by damaged onion cells. In this scenario, the *alt* cluster helps *P. ananatis* to survive and colonize onion tissue. *Pantoea ananatis* uses 11 *alt* cluster genes associated with the sulfur metabolism that impart tolerance to the thiosulfinate 'allicin' produced by damaged onion cells (Stice et al., 2020). The presence of the *alt* cluster in 80% (n = 41) of the onion pathogenic strains, and its absence in 67% (n = 20) of the non-pathogenic strains, suggests a potential involvement in bacterial virulence. However, the *alt* cluster alone is not sufficient for the onion pathogenic phenotype, as 33% (n = 10) of the non-pathogenic strains did not exhibit any evidence of pathogenicity in the onion red scale assay despite possessing a completely conserved *alt* cluster. These ten strains; however, did not contain complete or partial HiVir cluster except for one strain (PNA_98_3), which possessed a partial HiVir cluster.

CONCLUSION

In this study, we used the pan-GWAS approach to predict genes associated with onion-pathogenicity in *P. ananatis*. We concluded that the HiVir/PASVIL genes are associated with onion-pathogenicity as determined by the red-scale necrosis assay, and the *alt* gene cluster alone is not sufficient for pathogenic phenotype. Also, HiVir/PASVIL gene expression is potentially regulated, and the mere presence of the HiVir/PASVIL cluster does not guarantee a strain to be pathogenic on onion. In addition, a large repertoire of accessory genes identified in these strains may aid *P. ananatis* in diverse niche-adaptation and potentially in host-range expansion. The pan-GWAS pipeline can be deployed to characterize *P. ananatis* strains pathogenic to other plant hosts. We observed HGT events as major contributing factor for PAVs resulting in diversification of *P. ananatis* strains. Further integration of 'omics' technologies will provide deeper insights into the identification of novel pathogenicity and virulence factors in *P. ananatis* populations that cause center rot of onion. Whole transcriptome and proteome studies are required to understand the expression and function of identified pathogenicity and virulence factors in *P. ananatis*. A time-course based transcriptomic studies will reveal the temporal expression of pathogenicity and virulence genes as infection progresses. Proteomic studies will be used to validate transcriptomic expression of these genes and identify gene products.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

GA and BD conceived the project. GA performed the bioinformatics analyses and complied the manuscript. DC maintained the bacterial cultures, isolation of strains, and phenotyping of the 81 strains. SS and BK contributed in planning, designing the experiment, and manuscript revision. BM and SV contributed to the discussion. GA, RG, and BD designed and finalized the manuscript. BD planned the project, secured extramural funds, and revised and submitted manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia Office of the Vice President for Research and Office of the Vice President for Information Technology. This work was partially supported by the Specialty Crop Block Grant AWD00009682. This work was also partially supported by the Specialty Crops Research Initiative Award 2019-51181-30013 from the USDA National Institute of Food and Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture. The University of Georgia is an equal opportunity provider and employer.

ACKNOWLEDGMENTS

We thank Matthew Tyler Garrick, Medora Hoopes, Walt Sanders, and David Abgott for their technical assistance in the field and laboratory.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.684756/full#supplementary-material

Supplementary Figure 1 | Onion red scale necrosis shown by pathogenic strains of *P. ananatis*. Three replicates (rep 1, rep 3, and rep 3) are shown for

representative pathogenic stains (PNA 99-14, PNA 99-2, PNA 97-11, 99-1, and PANS 200-1) used to inoculate the red onion scales. Pathogenic strains showed scale clearing (bleaching out phenotype) whereas non-pathogenic strain did not show any symptoms. The positive control (PNA 97-1) and negative control (sterile water) is represented by "+" and "−", respectively.

Supplementary Figure 2 | Raw-read data generated, and filtered reads retained after stringent quality filtering using Trimmomatic.

Supplementary Figure 3 | Pan-genome analysis of 81 *Pantoea ananatis* genomes. **(A)** Genes contributed to pan-genome by individual genomes. **(B)** Distribution of gene (cluster) sizes as a function of the number of genomes they contain showing the partition of OMCL pan-genomic matrix into shell, cloud, soft-core and core compartments.

Supplementary Figure 4 | Average nucleotide identities (ANI) of the coding DNA sequences and average amino acid identities (AAI) of the protein coding genes of 81 *Pantoea ananatis*. **(A)** Heatmap was generated using the identity matrix calculated by get_homologues.pl using the BLAST scores representing the degree of similarity of the genomes based on ANI. **(B)** Heatmap showing the degree of similarity of the genomes based on the AAI calculated using BLASTP scores implemented in get_homologues.pl. Heatmaps were derived from the ANI and AAI matrix based on the pan-genome matrices. High similarity is represented by lighter color (light yellow to white) and low similarity is represented by dark orange to light orange shade. Vertical and horizontal axes are labeled with names of 81 strains of *P. ananatis*.

Supplementary Figure 5 | Dendrogram of 51 red-onion scale necrotic and 30 red-onion scale non-necrotic strains of *P. ananatis* based on shell genes. Strains highlighted in green are non-pathogenic and ones in red are pathogenic.

Supplementary Figure 6 | Dendrogram of 51 red-onion scale necrotic and 30 red-onion scale non-necrotic strains of *P. ananatis* based on cloud genes. Strains highlighted in green are non-pathogenic and ones in red are pathogenic.

Supplementary Figure 7 | Comparative phylogeny of 81 pathogenic and non-pathogenic strains of *Pantoea ananatis* based on core genome SNPs and presence and absence variations. **(A)** Phylogenetic tree constructed using core SNPs using RAxML. **(B)** Phylogenetic tree constructed using PAVs using RAxML. Numerical values in decimal represent the branch length. Longer branch length mean higher genetic divergence. 'P' represents pathogenic and 'NP' represents non-pathogenic strains. Strains labeled as PANS represent their non-onion origin. The rest were identified from onions (PNA).

Supplementary Figure 8 | Top GO terms: A bar chart representing the GO terms according to the number of annotated sequences. Panel A-D shows the function of genes assigned to biological process; panel E-H shows the function of genes assigned to molecular function and panel I-L represent the function of genes assigned to cellular component.

Supplementary Table 1 | Details of raw and filtered data generated for each sample of *P. ananatis* used in this study.

Supplementary Table 2 | Total number of core and accessory genes shared by each strain of *P. ananatis* used in this study.

Supplementary Table 3 | List of putative horizontal gene transfers, donor and recipient genomes and, gene transfers between two genomes.

Supplementary Table 4 | List of HGTs involving HiVir/PASVIL/analogous TCA genes.

Supplementary Table 5 | Genomic islands in PNA 99_3 predicted by SGI-HMM implemented by IslandViewer 4.

Supplementary Table 6 | Genomic islands in PNA 99_3 predicted by IslandPath-DIMOB implemented by IslandViewer 4.

Supplementary Table 7 | Pathogenicity factors (virulence factors) in PNA_99_3 as predicted by GIPSy.

Supplementary Table 8 | Symbiotic factors in PNA_99_3 as predicted by GIPSy.

Supplementary Table 9 | Genomic islands in PNA 99_14 as predicted by SGI-HMM implemented by IslandViewer 4.

Supplementary Table 10 | Genomic islands in PNA 99_14 predicted by IslandPath-DIMOB implemented by IslandViewer 4.

Supplementary Table 11 | Pathogenicity factors (virulence factors) in PNA_99_14 as predicted by GIPSy.

Supplementary Table 12 | Symbiotic factors in PNA_99_14 as predicted by GIPSy.

Supplementary Table 13 | List of 42 genes associated with red-onion scale necrosis caused by pathogenic strains of *P. ananatis* based on *P*-values.

Supplementary Table 14 | Annotation and mapping of representative core genes.

Supplementary Table 15 | Annotation and mapping of representative soft-core genes.

Supplementary Table 16 | Annotation and mapping of representative shell genes.

Supplementary Table 17 | Annotation and mapping of representative cloud genes.

Supplementary Table 18 | Life Identification Number (LIN) Ids assigned to *P. ananatis* strains submitted to LIN base.

REFERENCES

Aanensen, D. M., and Spratt, B. G. (2005). The multilocus sequence typing network: mlst. net. *Nucleic Acids Res.* 33(Suppl._2), W728–W733.

Agarwal, G., Stumpf, S., Kvitko, B. H., and Dutta, B. (2019). Center rot of onion. *Plant Health Instructor* doi: 10.1094/PHI-I-2019-0603-01

Arnold, D. L., Pitman, A., and Jackson, R. W. (2003). Pathogenicity and other genomic islands in plant pathogenic bacteria. *Mol. Plant Pathol.* 4, 407–420. doi: 10.1046/j.1364-3703.2003.00187.x

Asselin, J. A. E., Bonasera, J. M., and Beer, S. V. (2018). Center rot of onion (*Allium cepa*) caused by Pantoea ananatis requires pepM, a predicted phosphonate-related gene. *Mol. Plant Microbe Interact.* 31, 1291–1300. doi: 10.1094/mpmi-04-18-0077-r

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bertelli, C., Laird, M. R., Williams, K. P., Group, S. F. U. R. C., Lau, B. Y., Hoad, G., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35.

Bertelli, C., Tilley, K. E., and Brinkman, F. S. (2019). Microbial genomic island discovery, visualization and analysis. *Brief. Bioinform.* 20, 1685–1698. doi: 10.1093/bib/bby042

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bolormaa, S., Neto, L. P., Zhang, Y., Bunch, R., Harrison, B., Goddard, M., et al. (2011). A genome-wide association study of meat and carcass traits in Australian cattle. *J. Anim. Sci.* 89, 2297–2309. doi: 10.2527/jas.2010-3138

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 17:238.

Caballo-Ponce, E., Murillo, J., Martínez-Gil, M., Moreno-Pérez, A., Pintado, A., and Ramos, C. (2017a). Knots untie: molecular determinants involved in knot formation induced by *Pseudomonas savastanoi* in woody hosts. *Front. Plant Sci.* 8:1089. doi: 10.3389/fpls.2017.01089

Caballo-Ponce, E., van Dillewijn, P., Wittich, R. M., and Ramos, C. (2017b). WHOP, a genomic region associated with woody hosts in the *Pseudomonas syringae* complex contributes to the virulence and fitness of *Pseudomonas savastanoi* pv. savastanoi in olive plants. *Mol. Plant Microbe Interact.* 30, 113–126. doi: 10.1094/mpmi-11-16-0233-r

Carr, E. A., Zaid, A. M., Bonasera, J. M., Lorbeer, J. W., and Beer, S. V. (2013). Infection of onion leaves by Pantoea ananatis leads to bulb infection. *Plant Dis.* 97, 1524–1528. doi: 10.1094/pdis-06-12-0597-re

Carr, E., Bonasera, J., Zaid, A., Lorbeer, J., and Beer, S. (2010). First report of bulb disease of onion caused by Pantoea ananatis in New York. *Plant Dis.* 94, 916–916. doi: 10.1094/pdis-94-7-0916b

Cerro, C. D., Felpeto-Santaro, C., Rojas, A., Tortajada, M., Ramon, D., and Garcia, J. L. (2013). Genome sequence of the butanol hyperproducer clostridium saccharoperbutylacetonicum N1-4. *Genome Announc.* 1:e00070-13. doi: 10.1128/genomeA.00070-13.

Charaka, V. K., and Misra, H. S. (2012). Functional characterization of the role of the chromosome I partitioning system in genome segregation in *Deinococcus radiodurans*. *J. Bacteriol.* 194, 5739–5748. doi: 10.1128/jb.00610-12

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927.

Chen, J. A., Chen, Z., Won, H., Huang, A. Y., Lowe, J. K., Wojta, K., et al. (2018). Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol. Neurodegener.* 13:41.

Chen, J., Gopalkrishnan, S., Chiu, C., Chen, A. Y., Campbell, E. A., Gourse, R. L., et al. (2019). *E. coli* TraR allosterically regulates transcription initiation by altering RNA polymerase conformation. *Elife* 8:e49375.

Chen, P. E., and Shapiro, B. J. (2015). The advent of genome-wide association studies for bacteria. *Curr. Opin. Microbiol.* 25, 17–24. doi: 10.1016/j.mib.2015.03.002

Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832.

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/aem.02411-13

Coutinho, T. A., and Venter, S. N. (2009). Pantoea ananatis: an unconventional plant pathogen. *Mol. Plant Pathol.* 10, 325–335. doi: 10.1111/j.1364-3703.2009.00542.x

Cruz, A. T., Cazacu, A. C., and Allen, C. H. (2007). Pantoea agglomerans, a plant pathogen causing human disease. *J. Clin. Microbiol.* 45, 1989–1992. doi: 10.1128/jcm.00632-07

De Maayer, P., Chan, W. Y., Rubagotti, E., Venter, S. N., Toth, I. K., Birch, P. R., et al. (2014). Analysis of the Pantoea ananatis pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics* 15:404. doi: 10.1186/1471-2164-15-404

Didelot, X., and Wilson, D. J. (2015). ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* 11:e1004041. doi: 10.1371/journal.pcbi.1004041

Douglas, G. M., and Langille, M. G. (2019). Current and promising approaches to identify horizontal gene transfer events in metagenomes. *Genome Biol. Evol.* 11, 2750–2766. doi: 10.1093/gbe/evz184

Dutta, B., Barman, A., Srinivasan, R., Avci, U., Ullman, D., Langston, D., et al. (2014). Transmission of Pantoea ananatis and *P. agglomerans*, causal agents of center rot of onion (*Allium cepa*), by onion thrips (*Thrips tabaci*) through feces. *Phytopathology* 104, 812–819. doi: 10.1094/phyto-07-13-0199-r

Epstein, B., Abou-Shanab, R. A., Shamseldin, A., Taylor, M. R., Guhlin, J., Burghardt, L. T., et al. (2018). Genome-wide association analyses in the model Rhizobium Ensifer meliloti. *MSphere* 3, e00386–18.

Falush, D. (2016). Bacterial genomics: microbial GWAS coming of age. *Nat. Microbiol.* 1:16059.

Fan, C., Moews, P. C., Shi, Y., Walsh, C. T., and Knox, J. R. (1995). A common fold for peptide synthetases cleaving ATP to ADP: glutathione synthetase and D-alanine: D-alanine ligase of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 1172–1176. doi: 10.1073/pnas.92.4.1172

Galperin, M. Y., and Koonin, E. V. (1997). A diverse superfamily of enzymes with ATP-dependent carboxylate—amine/thiol ligase activity. *Protein Sci.* 6, 2639–2643. doi: 10.1002/pro.5560061218

Gitaitis, R., and Gay, J. (1997). First report of a leaf blight, seed stalk rot, and bulb decay of onion by Pantoea ananas in Georgia. *Plant Dis.* 81, 1096–1096. doi: 10.1094/pdis.1997.81.9.1096c

Gitaitis, R., Walcott, R., Culpepper, S., Sanders, H., Zolobowska, L., and Langston, D. (2002). Recovery of Pantoea ananatis, causal agent of center rot of onion, from weeds and crops in Georgia, USA. *Crop Prot.* 21, 983–989. doi: 10.1016/s0261-2194(02)00078-9

Gitaitis, R., Walcott, R., Wells, M., Perez, J. D., and Sanders, F. (2003). Transmission of Pantoea ananatis, causal agent of center rot of onion, by tobacco thrips, Frankliniella fusca. *Plant Dis.* 87, 675–678. doi: 10.1094/pdis.2003.87.6.675

Gong, J., Wang, F., Xiao, B., Panjwani, N., Lin, F., Keenan, K., et al. (2019). Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. *PLoS Genet.* 15:e1008007. doi: 10.1371/journal.pgen.1008007

Hilderbrand, R. L. (1983). *Role of Phosphonates in Living Systems*. Boca Raton, FL: CRC Press.

Kagawa, W., Arai, T., Ishikura, S., Kino, K., and Kurumizaka, H. (2015). Structure of RizA, an L-amino-acid ligase from *Bacillus subtilis*. *Acta Crystallogr. F Struct. Biol. Commun.* 71, 1125–1130.

Kini, K., Dossa, R., Dossou, B., Mariko, M., Koebnik, R., and Silué, D. (2019). A semi-selective medium to isolate and identify bacteria of the genus Pantoea. *J. Gen. Plant Pathol.* 85, 424–427. doi: 10.1007/s10327-019-00862-w

Kino, K., Kotanaka, Y., Arai, T., and Yagasaki, M. (2009). A novel L-amino acid ligase from *Bacillus subtilis* NBRC3134, a microorganism producing peptide-antibiotic rhizocticin. *Biosci. Biotechnol. Biochem.* 73, 901–907. doi: 10.1271/bbb.80842

Laing, C., Buchanan, C., Taboada, E. N., Zhang, Y., Kropinski, A., Villegas, A., et al. (2010). Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. *BMC bioinformatics* 11:461. doi: 10.1186/1471-2105-11-461

Lee, J.-H., Bae, B., Kuemin, M., Circello, B. T., Metcalf, W. W., Nair, S. K., et al. (2010). Characterization and structure of DhpI, a phosphonate O-methyltransferase involved in dehydrophos biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17557–17562. doi: 10.1073/pnas.1006848107

Liu, Y.-Y., Chiou, C.-S., and Chen, C.-C. (2016). PGAdb-builder: a web service tool for creating pan-genome allele database for molecular fine typing. *Sci. Rep.* 6:36213.

Maiden, M. C., Van Rensburg, M. J. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., et al. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736. doi: 10.1038/nrmicro3093

Metcalf, W. W., and van der Donk, W. A. (2009). Biosynthesis of phosphonic and phosphinic acid natural products. *Annu. Rev. Biochem.* 78, 65–94. doi: 10.1146/annurev.biochem.78.091707.100215

Moura, Q., Fernandes, M. R., Cerdeira, L., Nhambe, L. F., Ienne, S., Souza, T. A., et al. (2017). Draft genome sequence of a multidrug-resistant KPC-2-producing *Enterobacter aerogenes* isolated from a hospitalised patient in Brazil. *J. Glob. Antimicrob. Resist.* 10, 277–278. doi: 10.1016/j.jgar.2017.07.016

Nelson, W. C., and Stegen, J. C. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Front. Microbiol.* 6:713. doi: 10.3389/fmicb.2015.00713

Nowell, R. W., Green, S., Laue, B. E., and Sharp, P. M. (2014). The extent of genome flux and its role in the differentiation of bacterial lineages. *Genome Biol. Evol.* 6, 1514–1529. doi: 10.1093/gbe/evu123

Okunishi, S., Sako, K., Mano, H., Imamura, A., and Morisaki, H. (2005). Bacterial flora of endophytes in the maturing seed of cultivated rice (*Oryza sativa*). *Microbes Environ.* 20, 168–177. doi: 10.1264/jsme2.20.168

Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., et al. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157: H7. *Nature* 409, 529–533.

Pfeufer, E., Hoepting, C., and Gugino, B. (2015). Advances in managing onion bacterial diseases in the Northeastern US. *Onion World* 31, 22–27.

Polidore, A. L., Furiassi, L., Hergenrother, P. J., and Metcalf, W. W. (2021). A phosphonate natural product made by pantoea ananatis is necessary and sufficient for the hallmark lesions of onion center rot. *Mbio* 12:e03402–20.

Ramos, C., Matas, I. M., Bardaji, L., Aragón, I. M., and Murillo, J. (2012). P seudomonas savastanoi pv. savastanoi: some like it knot. *Mol. Plant Pathol.* 13, 998–1009.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Rijavec, T., Lapanje, A., Dermastia, M., and Rupnik, M. (2007). Isolation of bacterial endophytes from germinated maize kernels. *Can. J. Microbiol.* 53, 802–808. doi: 10.1139/w07-048

Rosselló-Mora, R. (2005). Updating prokaryotic taxonomy. *J. Bacteriol.* 187, 6255–6257. doi: 10.1128/jb.187.18.6255-6257.2005

Sangal, V., Goodfellow, M., Jones, A. L., Schwalbe, E. C., Blom, J., Hoskisson, P. A., et al. (2016). Next-generation systematics: an innovative approach to resolve the structure of complex prokaryotic taxa. *Sci. Rep.* 6:38392.

Schwartz, H. F., and Mohan, S. K. (2008). *Compendium of Onion and Garlic Diseases and Pests*. Saint Paul, MN: American Phytopathological Society.

Schwartz, H., and Otto, K. (2000). First report of a leaf blight and bulb decay of onion by Pantoea ananitas in Colorado. *Plant Dis.* 84, 808–808. doi: 10.1094/pdis.2000.84.7.808a

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sheibani-Tezerji, R., Naveed, M., Jehl, M.-A., Sessitsch, A., Rattei, T., and Mitter, B. (2015). The genomes of closely related Pantoea ananatis maize seed endophytes having different effects on the host plant differ in secretion system genes and mobile genetic elements. *Front. Microbiol.* 6:440. doi: 10.3389/fmicb.2015.00440

Sherlock, O., Vejborg, R. M., and Klemm, P. (2005). The TibA adhesin/invasin from enterotoxigenic *Escherichia coli* is self recognizing and induces bacterial aggregation and biofilm formation. *Infect. Immun.* 73, 1954–1963. doi: 10.1128/iai.73.4.1954-1963.2005

Shin, G. Y., Schachterle, J. K., Shyntum, D. Y., Moleleki, L. N., Coutinho, T. A., and Sundin, G. W. (2019). Functional characterization of a global virulence regulator hfq and identification of Hfq-dependent sRNAs in the plant pathogen Pantoea ananatis. *Front. Microbiol.* 10:2075. doi: 10.3389/fmicb.2019.02075

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244. doi: 10.1038/nature10571

Snipen, L., Almøy, T., and Ussery, D. W. (2009). Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics* 10:385. doi: 10.1186/1471-2164-10-385

Soares, S. C., Geyik, H., Ramos, R. T., de Sá, P. H., Barbosa, E. G., Baumbach, J., et al. (2016). GIPSy: genomic island prediction software. *J. Biotechnol.* 232, 2–11. doi: 10.1016/j.jbiotec.2015.09.008

Song, W., Wemheuer, B., Zhang, S., Steensen, K., and Thomas, T. (2019). MetaCHIP: community-level horizontal gene transfer identification through the combination of best-match and phylogenetic approaches. *Microbiome* 7:36.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stice, S. P., Stumpf, S. D., Gitaitis, R. D., Kvitko, B. H., and Dutta, B. (2018). Pantoea ananatis genetic diversity analysis reveals limited genomic diversity as well as accessory genes correlated with onion pathogenicity. *Front. Microbiol.* 9:184. doi: 10.3389/fmicb.2018.00184

Stice, S. P., Thao, K. K., Khang, C. H., Baltrus, D. A., Dutta, B., and Kvitko, B. H. (2020). Thiosulfinate tolerance is a virulence strategy of an atypical bacterial pathogen of onion. *Curr. Biol.* 30, 3130–3140. doi: 10.1016/j.cub.2020.05.092

Straub, C., Colombi, E., and McCann, H. C. (2021). Population genomics of bacterial plant pathogens. *Phytopathology* 111, 23–31. doi: 10.1094/phyto-09-20-0412-rvw

Stumpf, S., Kvitko, B., Gitaitis, R., and Dutta, B. (2018). Isolation and characterization of novel *Pantoea stewartii* subsp. indologenes strains exhibiting center rot in onion. *Plant Dis.* 102, 727–733. doi: 10.1094/pdis-08-17-1321-re

Takikawa, Y., and Kubota, Y. (2018). *A Genetic Locus Determining Pathogenicity of Pantoea ananatis (Abstr.). Phytopathology* 108 (212). Available online at: https://doi.org/10.1094/PHYTO-108-10-S1.1 (accessed October 2018).

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., et al. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* 5:e1000344. doi: 10.1371/journal.pgen.1000344

Treangen, T. J., and Rocha, E. P. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* 7:e1001284. doi: 10.1371/journal.pgen.1001284

Van Rossum, T., Ferretti, P., Maistrenko, O. M., and Bork, P. (2020). Diversity within species: interpreting strains in microbiomes. *Nat. Rev. Microbiol.* 18, 491–506. doi: 10.1038/s41579-020-0368-1

Varshney, R. K., Saxena, R. K., Upadhyaya, H. D., Khan, A. W., Yu, Y., Kim, C., et al. (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* 49:1082. doi: 10.1038/ng.3872

Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864. doi: 10.1038/s41588-019-0401-3

Vernikos, G., Medini, D., Riley, D. R., and Tettelin, H. (2015). Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* 23, 148–154. doi: 10.1016/j.mib.2014.11.016

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22. doi: 10.1016/j.ajhg.2017.06.005

Walcott, R., Gitaitis, R., Castro, A., Sanders, F. Jr., and Diaz-Perez, J. (2002). Natural infestation of onion seed by Pantoea ananatis, causal agent of center rot. *Plant Dis.* 86, 106–111. doi: 10.1094/pdis.2002.86.2.106

Welch, R. A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17020–17024.

Weller-Stuart, T., De Maayer, P., and Coutinho, T. (2017). Pantoea ananatis: genomic insights into a versatile pathogen. *Mol. Plant Pathol.* 18, 1191–1198. doi: 10.1111/mpp.12517

Wilkes, T., Darby, A., Choi, J. H., Colbourne, J., Werren, J., and Hurst, G. (2010). The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis. *Insect Mol. Biol.* 19, 59–73. doi: 10.1111/j.1365-2583.2009.00963.x

Willenbrock, H., Hallin, P. F., Wassenaar, T. M., and Ussery, D. W. (2007). Characterization of probiotic *Escherichia coli*isolates with a novel pan-genome microarray. *Genome Biol.* 8:R267.

Yokochi, N., Yoshikane, Y., Trongpanich, Y., Ohnishi, K., and Yagi, T. (2004). Molecular cloning, expression, and properties of an unusual aldo-keto reductase family enzyme, pyridoxal 4-dehydrogenase, that catalyzes irreversible oxidation of pyridoxal. *J. Biol. Chem.* 279, 37377–37384. doi: 10.1074/jbc.m405344200

Yu, X., Doroghazi, J. R., Janga, S. C., Zhang, J. K., Circello, B., Griffin, B. M., et al. (2013). Diversity and abundance of phosphonate biosynthetic genes in nature. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20759–20764. doi: 10.1073/pnas.1315107110

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 October 2021
doi: 10.3389/fmicb.2021.752760

Microbial Diversity Analysis and Genome Sequencing Identify *Xanthomonas perforans* as the Pathogen of Bacterial Leaf Canker of Water Spinach (*Ipomoea aquatic*)

Ming Hu[1†], Chuhao Li[1†], Xiaofan Zhou[1†], Yang Xue[1], Si Wang[1], Anqun Hu[1], Shanshan Chen[1], Xiuwen Mo[2] and Jianuan Zhou[1]*

[1] Guangdong Laboratory for Lingnan Modern Agriculture, Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Centre, South China Agricultural University, Guangzhou, China, [2] Agricultural Technology Service Centre of Daojiao Town, Dongguan, China

OPEN ACCESS

Edited by:
Marco Scortichini, Council for Agricultural and Economics Research (CREA), Italy

Reviewed by:
Chiaraluce Moretti, University of Perugia, Italy
Veronica Ancona, Texas A&M University Kingsville, United States

***Correspondence:**
Jianuan Zhou
jianuanzhou@scau.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: 04 August 2021
Accepted: 22 September 2021
Published: 27 October 2021

Citation:
Hu M, Li C, Zhou X, Xue Y, Wang S, Hu A, Chen S, Mo X and Zhou J (2021) Microbial Diversity Analysis and Genome Sequencing Identify Xanthomonas perforans as the Pathogen of Bacterial Leaf Canker of Water Spinach (Ipomoea aquatic). Front. Microbiol. 12:752760. doi: 10.3389/fmicb.2021.752760

Ipomoea aquatica is a leafy vegetable widely cultivated in tropical Asia, Africa, and Oceania. Bacterial leaf canker disease has been attacking the planting fields and seriously affecting the quality of *I. aquatica* in epidemic areas in China. This study examined the microbial composition of *I. aquatica* leaves with classical symptoms of spot disease. The results showed that *Xanthomonas* was overwhelmingly dominant in all four diseased leaf samples but rarely present in rhizospheric soil or irrigation water samples. In addition, *Pantoea* was also detected in two of the diseased leaf samples. Pathogen isolation, identification, and inoculation revealed that both *Xanthomonas* sp. TC2-1 and *P. ananatis* were pathogenic to the leaves of *I. aquatic*, causing crater-shaped ulcerative spots and yellowing with big brown rot lesions on leaves, respectively. We further sequenced the whole genome of strain TC2-1 and showed that it is a member of *X. perforans*. Overall, this study identified *X. perforans* as the causal pathogen of *I. aquatica* bacterial leaf canker, and *P. ananatis* as a companion pathogen causing yellowing and brown rot on leaves. The correct identification of the pathogens will provide important basis for future efforts to formulate targeted application strategy for bacterial disease control.

Keywords: bacterial leaf canker of water spinach, 16S rDNA amplicon sequencing, MLSA analysis, pathogenicity tests, genome sequencing

INTRODUCTION

Ipomoea aquatica Forsk, usually called water spinach, water convolvulus, kangkong, and swamp cabbage, is native to East Asia and has been widely cultivated as a popular green leafy vegetable in tropical and subtropical regions (Ismail and Fun, 2003). In China, it is the only species of the genus *Ipomoea* that is grown in paddy fields. It can grow in all provinces of the Yangtze River Basin from April to October, and is suitable to grow in warm and humid climate, as well as fertile and wet soil. In addition to its high nutritive value to humans and animals, *I. aquatica* also possesses medicinal importance (Hu et al., 2010; Lawal et al., 2015, 2017). It can be taken orally to alleviate nervous

debility, jaundice (Lawal et al., 2017), nosebleeds, and high blood pressure (Duke and Ayensu, 1985). The extract has been reported to inhibit prostaglandin synthesis (Tseng et al., 1992) and possess antidiabetic (Malalavidhane et al., 2000), antioxidant (Prasad et al., 2005b), antinematodal (Mackeen et al., 1997), and anticancer activities (Prasad et al., 2005a).

Water spinach is affected by a variety of foliar diseases, such as the white rust caused by *Albugo ipomoeae-aquaticae*, which happens in almost all the planting areas (Sawada, 1922; Safeeulla and Thrumalachar, 1953; Ho and Edie, 1969; Gao et al., 1985; Yu et al., 2015), the foliar diseases caused by *Phyllosticta ipomoeae*, *Cerocospora ipomoeae*, and *Pseudomonas syrigae* pv. *Syringae*, which were observed in commercial greenhouses in Ontario and California during the 1990s (Cerkauskas et al., 2006), spot blight caused by *Stagonosporopsis cucurbitacearum* in China (Liu et al., 2017), and leaf spot caused by *Myrothecium roridum* in Jianshui County, Yunnan Province, China, in May 2015 (Wang et al., 2017).

Bacterial leaf spot of water spinach was first observed in Bang Pai, Pasee Charoen, Bangkok, Thailand, in 1990, and the pathogen was initially identified as a pathovar of *Xanthomonas campestris* by biochemical characters and xanthomonadin pigment (Leksomboon et al., 1991). Records in the NCBI GenBank database indicate that the disease also happened in Fuzhou, Fujian Province, China, in 2015 where *X. perforans* was identified as the pathogen based on the *dnaK* (MN626337.1), *gyrB* (MN626338.1), *recA* (MN626339.1), and *rpoD* (MN626340.1) gene sequence similarities.

Since 2010, bacterial leaf canker (also called bacterial leaf spot) has occurred in Dongguan City, Guangdong Province, China. The scattered land scale of Dongguan City and the high cost of land rent bring a demand for the plantation of high-value economic crops. During the flood season from April to September every year, due to the continuous rainfall and the small land blocks among industrial and living regions, farmers are inclined to plant water spinach in the pattern of basically paddy field planting for its tolerance to waterlogging and high temperature. Farmers vividly called bacterial leaf canker "pockmarked disease" according to the crater shape around the leaf lesions. The disease is becoming increasingly serious and has significantly affected the production of water spinach, causing great economic losses to vegetable farmers. The pathogen of the disease in Dongguan City is unclear. Farmers mistakenly attributed the pockmarks to white rust and, therefore, often applied broad-spectrum protective fungicides such as mancozeb in the field, but received no controlling effect. Most farmers are helpless and have to temporarily give up field management and harvest until the weather becomes cool.

In the summer of 2019 and 2020, we surveyed the incidence of bacterial leaf canker of water spinach in Dongguan City and collected diseased leaves, water and soil samples for analysis of the microbial diversity using 16S rDNA amplicon sequencing. We also isolated and identified the pathogens using culture-based approaches. Results revealed that the pathogen responsible for the classical symptoms of pockmarked spots on leaves was *Xanthomonas* sp., and *Pantoea ananatis* was a companion bacterium causing yellowing with big brown rot lesions on leaves. Further identification of the *Xanthomonas* strain by genome sequencing revealed that it belongs to *X. perforans*.

MATERIALS AND METHODS

Disease Investigation and Sample Collection

Dongguan City is located at 22°39′ to 23°09′ North latitude and 113°31′ to 114°15′ East longitude, in the south-central part of Guangdong Province, the East Bank of the Pearl River Estuary and the Pearl River Delta downstream of the Dongjiang River.

Since 2010, water spinach in towns and districts of Qingxi, Shatian, Chashan, Daojiao, and Wanjiang and other places in Dongguan City has suffered from a new disease (vividly called "pockmark" by farmers). In this study, the five sampling method was used to investigate the incidence rate of the disease in the summer of 2019 and 2020. Briefly, five points in each field and 50 clumps of water spinach at each point were examined. A total of 2,500 clumps in 10 fields were investigated. The incidence of disease was calculated as the percentage of infected clumps in total clumps surveyed. For the analysis of microbial diversity and pathogen isolation, samples from Wanjiang District were collected.

Microbiota Analysis

Genomic DNA of symptomatic leaf (1 g), water (10 ml), and soils (5 g) were, respectively, extracted using ALFA-SEQ Advanced Plant DNA Kit (for the leaf) and ALFA-SEQ Advanced Soil DNA Kit (for water and soil samples) (Guangdong Magigene Biotechnology Co., Guangzhou, China) according to the manual protocols. Integrity, purity, and concentration of the DNA were examined using 1% agarose gel electrophoresis. The V3–V4 region of the 16S rDNA of the sample DNA was amplified using TaKaRa Premix Taq® version 2.0 (TaKaRa Biotechnology Co., Dalian, China) with barcoded primers (338F 5′-ACTCCTACGGGAGGCAGCA-3′; 806R 5′-GGACTACHVGGGTWTCTAAT-3′) in conditions of pre-denaturation at 94°C for 5 min; 30 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 30 s; extension at 72°C for 10 min on BioRad S1000 (Bio-Rad Laboratory, CA, United States). Three replicates were amplified for each sample and combined for further analysis. The amplicons were checked using 1% agarose gel electrophoresis, and fragments in size of 470 bp were purified using E.Z.N.A.® Gel Extraction Kit (Omega, United States). A library was built according to the standard process of NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, United States). The products were sequenced by Illumina Nova 6000 PE250 technology at Guangdong Magigene Biotechnology Co., Guangzhou, China.

Paired-end clean reads were obtained after quality control by removing the primer sequences of the paired-end raw reads using the cutadapt software[1] and assembled using arch-fastq_mergepairs V10 (parameters include the minimum overlap length as 16 bp and the maximum allowable mismatch of

[1]https://github.com/marcelm/cutadapt/

overlap region of splicing sequence as 5 bp)[2] to obtain raw tags, which were subjected to quality tailoring (-W 4 -M 20) using fastp version 0.14.1[3] to get the clean tags. The raw sequencing reads were deposited in the NCBI Sequence Read Archive database under the accession number: PRJNA739543. Operational taxonomic units (OTUs) were sorted and counted using the UPARSE software. Representative sequences of each OUT were annotated by the Ribosomal Database Project (RDP) to get the species sources. Community analysis and different abundance of OTUs were performed using STAMP 2.0.8[4], vegan (R package).

Microbial Isolation and Purification

Microbial separation from the symptomatic leaves was carried out using the method previously described (Li et al., 2020). First, diseased samples were cut into small pieces (0.5–1 cm^2), and the surface was disinfected in sequence with 70% ethanol solution for 30 s, 5% sodium hypochlorite for 1 min, and sterile water washing for three times. Second, tissues were placed onto Luria–Bertani (LB, containing 10 g/L of typtone, 5 g/L of yeast extract, and 10 g/L of NaCl) agar (1.5% w/v) plates and potato dextrose agar (PDA, TOPBIO, Zhaoyuan, China) plates for incubation at 28°C for 24 h and 3 days, respectively. Colonies from the plates were, respectively, streaked onto fresh LB medium plates for purification and grown in LB liquid medium with shaking at 200 rpm overnight to preserve for further study.

16S rDNA Gene Sequencing of Single Colony

Bacteria were grown in LB medium until OD_{600} = 1.0, and genomic DNAs were extracted using the EasyPure Bacteria Genomic DNA Kit (TransGen Biotech, Beijing, China). The 16S rDNA gene sequences were amplified using the primers 27f and 1492r (Coenye et al., 1999) listed in **Supplementary Table 1**. The products were examined using 1% agarose gel electrophoresis and sent to Sangon Biotech Company in Shanghai, China, for sequencing. SeqMan V.5.00 was used to assemble sequences generated from forward and reverse primers.

Pathogenicity Tests of the Isolated Strains

Single colonies of the tested strains were grown in 10 ml of LB medium at 28°C overnight and adjusted to OD_{600} = 1.5. Since the leaves of water spinach are very thin and tender, inoculation by acupuncture or injection *in vitro* or *in vivo* resulted in rapid rot (data not shown). Therefore, we have created a new inoculation method suitable for pathogenicity tests on tender leaves. Briefly, leaves with a similar size on the plants grown in a pot were chosen and wiped with 70% ethanol solution on the back. After drying, sterile toothbrushes dipped in each bacterial culture were used to brush gently on the lower side of three leaves. LB liquid medium was used as a negative control. The pots were kept in a growth chamber (Shanghai YiHeng Scientific Instruments Co., Ltd) with controlled conditions as 28 ± 2°C, 75 ± 15% relative humidity, and 12-h white light illuminance (7,350 lx). Photos were taken after 2- and 7-day post inoculation. The experiment was repeated twice.

To fulfill Koch's postulates, the bacteria were re-isolated from the diseased leaves, and the 16S rDNA gene sequences were amplified and compared with the corresponding sequences of the inoculated ones.

For co-inoculation of TC2-1 and TC3-1, sterile toothbrushes were used to gently brush on the right half lower side of the blades to make microwounds, and 100 μl of bacterial cultures (grown in LB medium until OD_{600} = 1.5) of TC2-1, TC3-1, or TC2-1 + TC3-1 mixture was evenly spread onto the whole lower side of each blade. The pots were kept in the growth chamber. Photos were taken after 1- and 3-day post inoculation. The experiment was repeated three times.

To test the pathogenicity of *X. perforans* on different plants, the tested strains were grown in LB medium until OD_{600} = 1.0, and different plant organs were selected using different inoculation methods. For pepper (*Capsicum annuum*) and tomato (*Lycopersicon esculentum*) leaves, and mango (*Mangifera indica*) fruit, sterile toothbrushes dipped in TC2-1 culture were used to brush gently on the inoculated sites. For citrus (*Citrus reticulata*) leaves, TC2-1 was inoculated on the lower side by pressing a sterile puncher (5 mm) dipped with bacterial culture and incubated at 28°C. All trays were kept at 28°C until symptoms appeared. The same volume of LB medium was inoculated as a negative control. Each assay was repeated three times.

Hypersensitive Response Assay

Bacteria were grown in LB medium until OD_{600} = 1.0 (greater than 10^9 CFU/ml) (Kido et al., 2010). Leaves of *Nicotiana tabacum* variant K326 (Chen et al., 2018) were inoculated on the lower side by pressing a sterilized puncher (5 mm) dipped with bacterial culture and incubated at 28°C. A positive hypersensitive response (HR) reaction was recorded when the inoculated leaf tissue collapsed or light brown necrosis occurred within 48 h after inoculation (Lana et al., 2012). Each assay was repeated three times.

Pathogen Identification by Multilocus Sequence Analysis

Pathogenic isolates that fulfilled Koch's postulates were further identified using the multilocus sequence analysis (MLSA). For the *Xanthomonas* isolates, the *atpD* (ATP synthase β subunit) (Brady et al., 2008), *avrBs2* (Type III secretion effector AvrBs2) (Hajri et al., 2009), *gyrB* (DNA gyrase subunit B) (Parkinson et al., 2007), *cpn60* (60-kDa chaperonin protein subunits) (Hill et al., 2004), and *rpoD* (RNA polymerase sigma factor) (Tian, 2018) genes were selected for PCR amplification; for the *Pantoea* isolates, the *atpD*, *gyrB*, *infB* (initiation translation factor 2), and *rpoB* (DNA-directed RNA polymerase subunit beta) (Brady et al., 2008) gene sequences were, respectively, amplified using the primers listed in **Supplementary Table 1**. Amplicons were

[2]http://www.drive5.com/usearch/

[3]https://github.com/OpenGene/fastp

[4]http://www.r-project.org/

FIGURE 1 | Symptoms of bacterial leaf canker of water spinach in the field. **(A–C)** Diseased water spinach plants in the field. **(D,E)** Early symptoms on the lower and upper side of the leaf. **(F,G)** Leaf symptoms in the following course of disease development on the lower and upper side of leaf. **(H,I)** Leaf symptoms in the later stage on the lower and upper side of the leaf. Scale bar = 1 cm.

purified with a PCR Purification Kit (TransGen Biotech, Beijing, China) and sequenced by Sangon Biotech Company in Shanghai, China. Sequences were submitted to the GenBank database with accession nos. indicated in **Supplementary Table 2**.

In view of the classification confusion and evolution inconsistency of different housekeeping genes in *Xanthomonas* genus, phylogenetic trees of isolates TC1-1 and TC2-1 were built based on each gene sequence to reveal their taxonomic status; for isolates TC3-1 and TC3-2, phylogenetic analysis was performed based on the joint sequences of *atpD*, *gyrB*, *infB*, and *rpoB* genes of the two isolates and their corresponding close-related strains obtained from NCBI database. Sequences were aligned with ClustalW, and trees were constructed using the MEGA 6.0 software by the maximum-likelihood method with 1,000 bootstrap replicates.

Genome Sequencing, Assembly, and Annotation

Genomic DNA was extracted from TC2-1 in LB medium culture using the SDS method. The harvested DNA was detected by the agarose gel electrophoresis and quantified by Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States). The complete genome was sequenced by Novogene (Tianjin, China) using the Nanopore PromethION platform and Illumina NovaSeq platform. The sequence was first assembled with long read data and then polished with both long and short read data by medaka v1.2.5 and unicycler v0.4.8 (Wick et al., 2017), respectively. The sequencing data and genome assembly have been deposited in the NCBI database under the accession no. PRJNA742079.

PGAP (2021-01-11.build5132) (Tatusova et al., 2016) was used to predict gene structure and gene function of the complete genome. Islandviewer4 (Bertelli et al., 2017) was used to predict genomic islands. Phaster (Arndt et al., 2016) was used to predict prophage. TXSScan was used to annotate secretion systems (Abby and Rocha, 2017). Effectors were predicted by aligning all protein sequence of TC2-1 using BLASTP with validated effectors in the T3SE (Hu et al., 2017), T4SE (Bi et al., 2013; Zou et al., 2013; An et al., 2018), and T6SE databases (Li et al., 2015; An et al., 2018). Hits with e-value less than $1e^{-5}$ and qcov_hsp_perc more than 50 were identified as effectors.

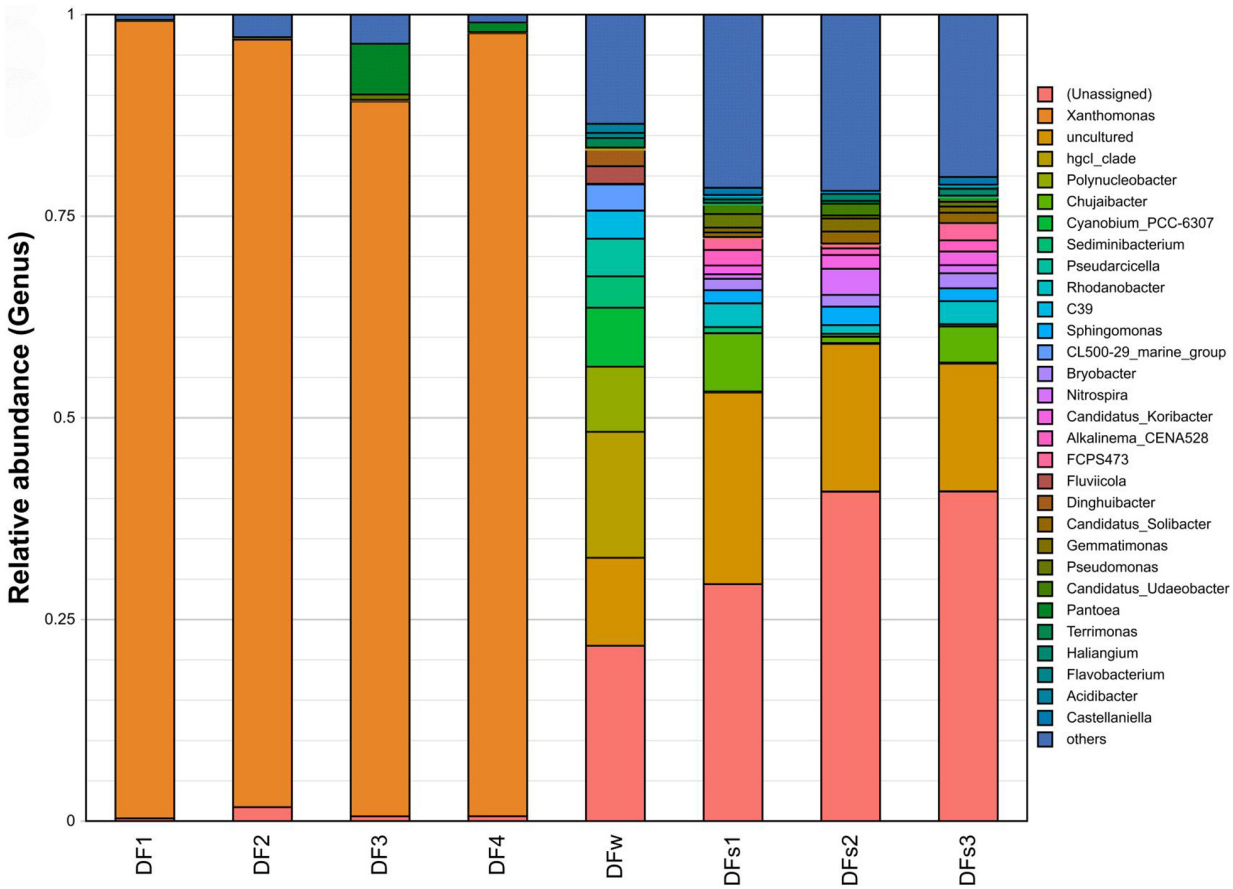


FIGURE 2 | Relative abundance of the top 30 16S rDNA-based operational taxonomic units (OUTs) at the genus level. Samples are labeled to indicate their source of collection: DF1 to DF4 indicate four leaf samples of diseased plants from Dafen Zone. DFs1 to DFs3 indicate three rhizosphere soil samples of diseased plants from Dafen Zone. DFw indicates nearby irrigation water of diseased plants from Dafen Zone.

Phylogenetic Analysis of TC2-1 and Genomic Comparison of Available Genomes of *Xanthomonas perforans* Strains

Phylogenetic analysis was performed by GTDB-tk v1.4.0 (Chaumeil et al., 2020). In detail, TC2-1 was placed to a pre-built bacteria-wide phylogenetic tree according to the sequence identity of 120 single-copy conserved genes, and then ANI values between TC2-1 and its closely related representative strains on the tree was calculated by fastANI. The taxonomic identity of TC2-1 was determined based on both the phylogenetic and ANI analyses. A separate phylogenetic tree including TC2-1 representative *Xanthomonas* strains was extract from the output of GTDB-tk, and the ANI value between strains was calculated again by the fastANI. Tree files were operated by ete3 (Huerta-Cepas et al., 2016), and phylogenetic tree was visualized by iTOL (Letunic and Bork, 2019).

To identify TC2-1-specific genes, a phylogenetic tree of TC2-1 and 141 *X. perforans* retrieved from the NCBI RefSeq database was built by the method described previously (Pierre-Alain et al., 2018). Then TC2-1 and five representative strains were selected based on the phylogenetic tree, namely, *X. perforans* NI1 (GCF_003136155.1), *X. perforans* Xp11-2 (GCF_001009445.1), *X. perforans* AL66 (GCF_007714065.1), *X. perforans* GEV2392 (GCF_006979735.1), and *X. perforans* CFBP 7293 (GCF_001976075.1), which were analyzed by orthofinder v2.5.2 (Emms and Kelly, 2015). TC2-1-specific genes were then screened based on the orthofinder results after filtering the protein sequences according to the parameters (qcovhsps > 80% and identity > 65%) compared with those of *X. perforans* in the NCBI database.

RESULTS

Disease Incidence and Symptoms of Bacterial Leaf Canker of Water Spinach in Dongguan City

Since 2010, bacterial leaf canker disease has attacked water spinach cultivated in many towns and districts of

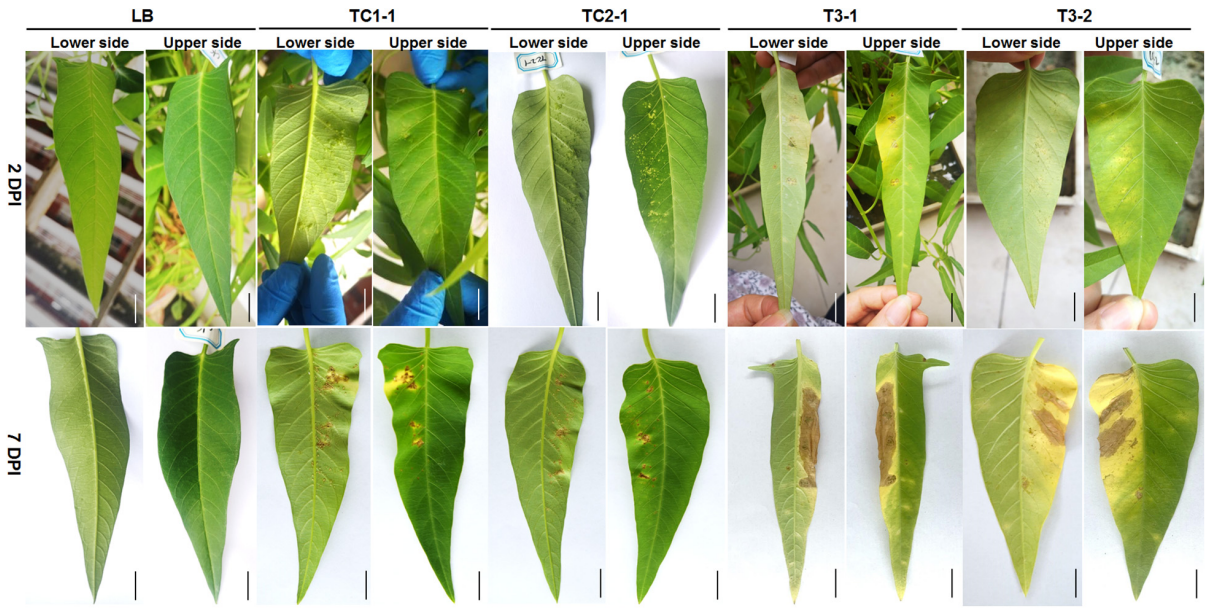


FIGURE 3 | Symptoms of water spinach after inoculation with LB, strains TC1-1, TC2-1, TC3-1, and TC3-2, respectively. Photos were taken after 2 and 7 DPI (days post inoculation), respectively. Scale bar = 1 cm.

Dongguan City, Guangdong Province, China. Due to scattered field planting, it is difficult to calculate the total planting area of water spinach. Therefore, we investigated some cultivated fields in Daojiao Town, Gaobu Town, and Wanjiang District in the summer of 2019 and 2020. The occurrences of the disease on "Baijun 311" and "Baijun 611" cultivars are 100% in all the surveyed places, while the occurrences on "Qingtong 311" cultivar are 16.7, 86, and 100% in Daojiao, Wanjiang, and Gaobu, respectively (**Supplementary Table 3**).

When the disease occurs, some leaves become yellow (**Figure 1A**), while some remain green, with brown or chlorotic spots observed on both (**Figure 1B**). Small raised black spots are observed on the lower side of the leaf (**Figure 1C**). In the early phase of infection, water-soaked transparent spots emerge on the lower side of the leaves, protuberant, edge clear (**Figure 1D**). Chlorotic spots appear on the upper side of the leaves, flat, edge dim, and irregular, and the middle of the lesions gradually turns brown (**Figure 1E**). In the following course of disease development, more and more watery spots grow on the lower side of the leaves, gradually expand, and turn brown, with a yellow halo on the edge (**Figure 1F**). Chlorotic spots turn larger on the leaf surface with brown in the middle and irregular yellow halo on the edge (**Figure 1G**). Finally, some disease spots expand and become connected; some leaves curl up and deform; the edge of the spots protrudes into crater -like cankers on the lower side (**Figure 1H**), which appear as brown necrosis with a yellow halo around on the upper side (**Figure 1I**). During the whole period of the disease, the roots and the stems remain healthy, indicating that the pathogens preferentially colonize mesophyll tissue.

Taxa Abundance and Diversity of the Samples

Leaves and rhizospheric soil of diseased plants and irrigation water near the planting field were collected from Dafen Zone, Wanjiang District, and the microbiomes were determined by 16S rDNA amplicon sequencing. A total of about 84,000 raw paired-end reads were generated for each sample (**Supplementary Table 4**). The proportions of high-quality reads were between 3.23 and 25.74% for leaf samples, between 49.38 and 55.40% for soil samples, and as high as 84.83% for the water sample (**Supplementary Table 4** and **Supplementary Figure 1**). A combined analysis of high-quality reads from all samples identified a total of 4,903 OTUs, among which 333 were detected in the four leaf tissues, 1,100 were detected in the water sample, and 4,268 were detected in the three soil samples.

The most abundant OTUs are vastly different among samples (**Supplementary Table 5**). In leaf samples, *Xanthomonas* was the genus with the highest relative abundance in the community (ranging from 88.64 to 98.88%), and *Pantoea* ranked second in two of the samples. In contrast, in the water and soil samples, the proportion of *Xanthomonas* was between 0.06 and 0.39%, and the proportion of *Pantoea* was between 0 and 0.15% (**Figure 2** and **Supplementary Table 5**).

Molecular Identification of Bacterial Strains

Diseased leaf samples collected from Wanjiang District were subjected to pathogen isolation. As a result, a total of 23 isolates were obtained from the leaf tissues, including 5 *Xanthomonas* sp.

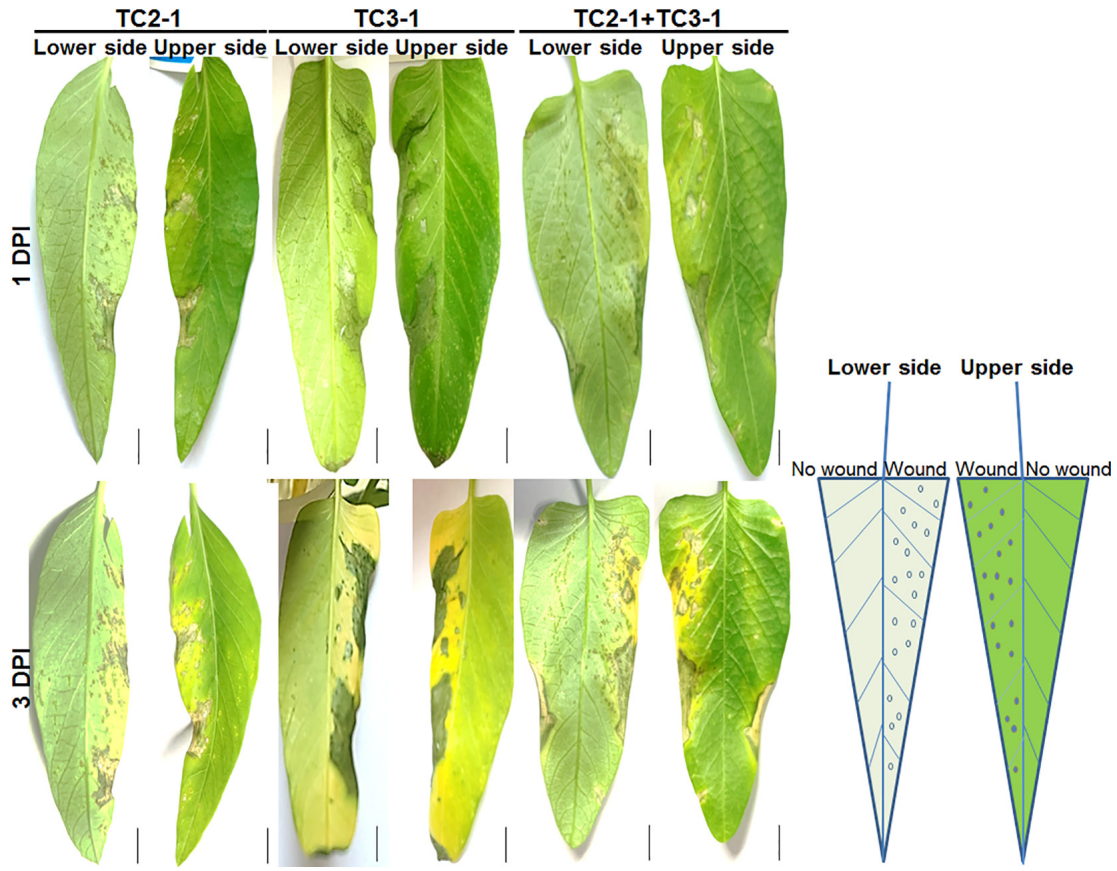


FIGURE 4 | Co-inoculation of TC2-1 and TC3-1 on water spinach leaves. Sterile toothbrushes were used to gently brush on the right half lower side of blades to make micro wounds, 100 μl of bacterial cultures (grown in LB medium until OD_{600} = 1.5) of TC2-1, TC3-1, or TC2-1 + TC3-1 mixture were evenly spread onto the whole lower side of each blade. Photos were taken after 1 and 3 DPI (days post inoculation), respectively. Scale bar = 1 cm.

isolates (TC1-1, TC1-2, TC2-1, TC2-3, and TCX21), 3 *Pantoea* sp. isolates (TC3-1, TC3-2, and TC3-5), 1 *Pseudomonas aeruginosa* isolate (TC11), and 3 additional *Pseudomonas* sp. isolates (TC1-3, TC3-3, and TC3-4), as well as 1 *Bacillus aquimaris* isolate (TC1-4) and 10 additional *Bacillus* spp. isolates (TC2-2, TC2-4, TCX22, TCX31, TCX32, TCX33, WJ2-1, WJ1, WJ2-2, and WJ3) (**Supplementary Table 6**). For *Xanthomonas* isolates, all the 16S rDNA gene sequences are completely identical to those of *X. axonopodis* pv. *commiphoreae* strain LMG26789, *X. perforans* strains 91-118 and LH3, and 99.93% identical to that of *X. axonopodis* pv. *citrumelo* F1. For three *Pantoea* isolates, the 16S rDNA gene sequences are completely identical to that of *Pantoea ananatis* SGAir0210, and 99.73% are identical to that of *P. ananatis* PP1.

To determine which isolates are pathogenic, we performed pathogenicity tests of TC1-1, TC2-1, TC1-3, TC1-4, TCX11, WJ2-1, TC3-1, TC3-2, WJ1, WJ2-2, and WJ3 on leaves of water spinach, and found that only isolates TC1-1, TC2-1, TC3-1, and TC3-2 are pathogenic (**Figure 3**). Two days after inoculation, dense water-soaked spots raised from the back of the leaves inoculated with *Xanthomonas* isolates TC1-1 and TC2-1, and the spots are chlorotic and yellow with clear edge on the leaf surface (**Figure 3**). On the other hand, small chlorotic spots appeared on the back of the leaves inoculated with *Pantoea* isolates TC3-1 and TC3-2; lesions were yellow, irregular, and indistinct with non-swollen mesophyll tissue on the leaf surface (**Figure 3**). At 7 days post inoculation, the protuberant spots in the TC1-1- and TC2-1-inoculated groups turned brown, and some expanded together with a yellow halo around the lesions. In general, the number of spots did not increase significantly (**Figure 3**). In comparison, in the TC3-1-and TC3-2-inoculated groups, the disease spots rapidly spread and connected into blocks with mesophyll tissue rotted and turned brown. The leaves were seriously yellowing on the inoculated site (**Figure 3**). According to the disease symptoms of the inoculated leaves, those inoculated with isolates TC1-1 and TC2-1 were similar to the typical symptoms of the bacterial canker disease. Thus, we observed *Xanthomonas* isolates TC1-1 and TC2-1 as the pathogens of the bacterial leaf canker of water spinach.

The simultaneous isolation of both *Pantoea* and *Xanthomonas* from diseased plants raises the possibility of synergistic or antagonistic interaction between both pathogens. Thus, co-inoculation of TC2-1 and TC3-1 were performed, and results showed that in the condition of wound pre-treatment, the

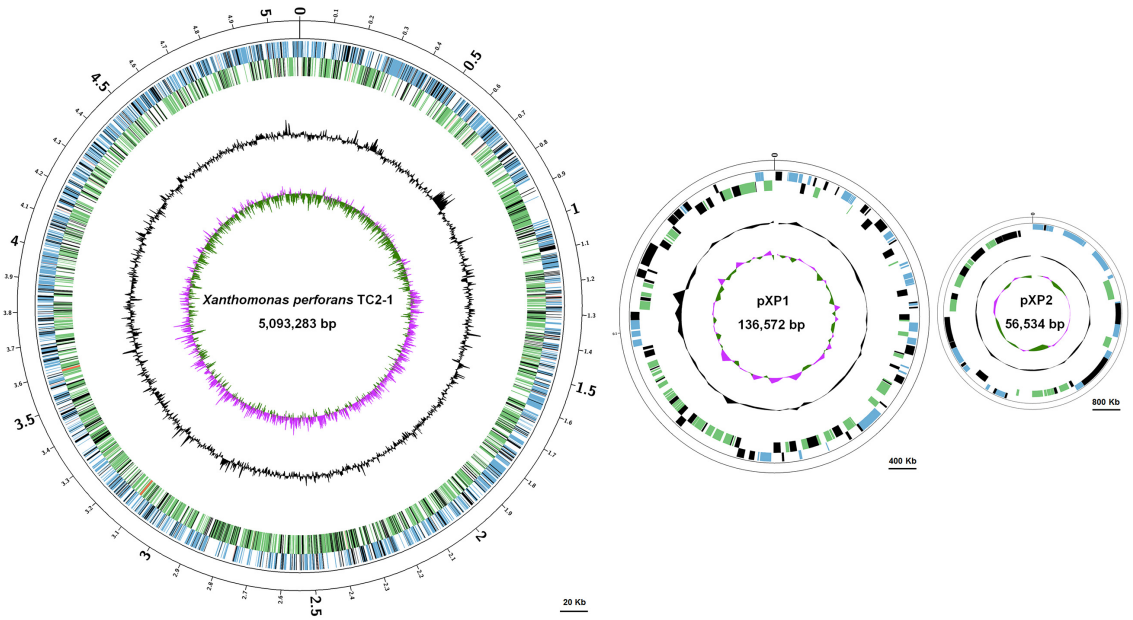


FIGURE 5 | Circular map of the chromosome and two plasmids pXP1 and pXP2 of *Xanthomonas perforans* strain TC2-1. The circles from outside to inside represent feature of the positive strand, CDS (blue), rRNA (red), pseudogene (black); feature of the negative strand, CDS (green), rRNA (red), pseudogene (black); GC content; and GC-skew. Bars were indicated under each of the circular maps.

TABLE 1 | Genomic features of *Xanthomonas perforans* TC2-1.

Features	Chromosome	pXP1	pXP2	Total
Size (bp)	5,093,283	136,572	46,534	5,276,389
GC content (%)	64.76	60.96	62.10	64.64
Gene	3,820	81	37	3,938
CDS	3,730	81	37	3,848
RNA genes	90	0	0	90
rRNA	6	0	0	6
tRNA	51	0	0	51
ncRNA	30	0	0	30
RNase_P_RNA	1	0	0	1
SRP_RNA	1	0	0	1
Pseudogene	599	60	12	671
Riboswitch	6	0	0	6
tmRNA	1	0	0	1
Transposase	289	40	5	334

co-inoculation of both pathogens led to similar symptoms as the inoculation of TC2-1 alone. However, for non-wound pre-treatment, which resemble the condition of natural infection, co-inoculation of both pathogens resulted in accelerated disease development and more severe symptoms compared with the inoculation of either pathogen individually (**Figure 4**).

To refine the taxonomic position of the pathogenic isolates, we performed MLSA analysis based on five housekeeping genes (*atpD*, *avrBs2*, *cpn60*, *gyrB*, and *rpoD*) for isolates TC1-1 and TC2-1, and four housekeeping genes (*atpD*, *gyrB*, *infB*, and *rpoB*) for isolates TC3-1 and TC3-2. As a result, the single-gene phylogenies of *atpD*, *cpn60*, *gyrB*, and *rpoD*, as well as the joint phylogenetic tree of all five genes, all contain a monophyletic clade consisting of TC1-1, TC2-1, and six other strains including *X. axonopodis* pv. *citrumelo* F1, *X. axonopodis* pv. *commiphoreae* strain LMG26789, *X. perforans* strains 91-118 and LH3, *X. euvesicatoria* LMG930, and *X. campestris* pv. *vesicatirua* 85-10 (**Supplementary Figure 2**). However, the phylogenetic results were inconclusive on the closest relatives of TC1-1 and TC2-1, leaving the taxonomic status of both strains at the species level uncertain. A high similarity between strains in different species indicates inconsistency in the current *Xanthomonas* classification situation. Therefore, we decided to sequence the whole genome of one of the isolates TC2-1 to determine the exact taxonomy of bacterial leaf canker pathogen of water spinach.

In addition, TC3-1 and TC3-2 have completely identical sequences of *atpD*, *gyrB*, *infB*, and *rpoB*, which are most similar to those of strain SGAir0210, sharing 99.11, 99.18, 99.38, and 99.36% identities, respectively. Joint phylogenetic analysis of *atpD*, *gyrB*, *infB*, and *rpoB* also indicated that strains TC3-1 and TC3-2 are grouped together with *P. ananatis* strains (**Supplementary Figure 3**). Thus, based on these results, we can conclude that strains TC3-1 and TC3-2 belong to *Pantoea ananatis*.

Species Classification of Isolate TC2-1 Based on Its Complete Genome Information

The TC2-1 genome was sequenced by a combination of Illumina short-read and Nanopore long-read sequencing technologies,

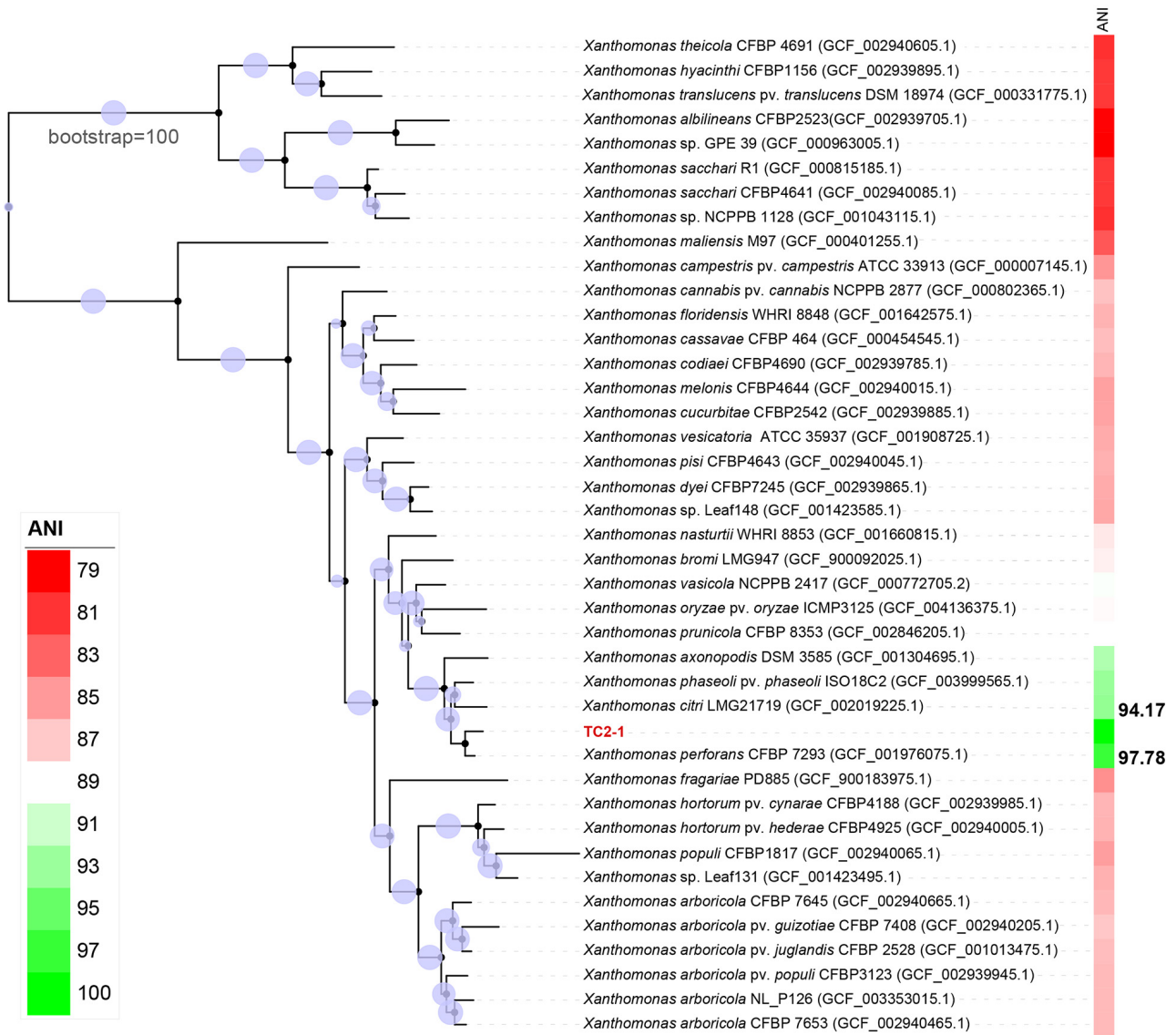


FIGURE 6 | Phylogenetic tree of strain TC2-1 and the representative strains in genus *Xanthomonas* listed in the GTDB. ANI analysis was used by means of GTDB-Tk (Chaumeil et al., 2020). The closest strain of TC2-1 is *X. perforans* CFBP 7293 with a 97.78% ANI value.

and assembled into one circular chromosome (size: 5,093,283 bp; GC content: 64.76%) and two plasmids (sizes: 136,572 and 46,534 bp; GC contents: 60.96 and 62.10%) (**Figure 5** and **Table 1**). A total of 3,848 protein-coding genes, 90 RNA genes (6 rRNAs, 51 tRNAs, and 30 ncRNAs), 671 pseudogenes, and 334 transposases (some are predicted as CDSs, some are pseudogenes) are annotated in the TC2-1 genome (GenBank no. CP077785.1) (**Table 1**).

The genome sequence of TC2-1 was analyzed with GTDB-Tk, which first placed TC2-1 in a pre-built bacterial phylogeny based on 120 proteins and then calculated the average nucleotide identity (ANI) values between TC2-1 and closely related type strains on the tree (Chaumeil et al., 2020). The result showed that TC2-1 is a member of *X. perforans*; the ANI value between TC2-1 and the type strain of *X. perforans* (strain CFBP 7293) is 97.78%, which is above the 95% threshold commonly used for bacterial species classification (Jain et al., 2018; **Figure 6** and **Supplementary Table 7**). Additionally, the ANI values between TC2-1 and all 141 *X. perforans* genomes in the NCBI RefSeq database are above 97% (**Supplementary Figure 4**). Therefore, the water spinach pathogenic strain TC2-1 belongs to *Xanthomonas perforans*. To the best of our knowledge, water spinach is the fourth reported natural host plant of *X. perforans* besides tomato, pepper, and *Eucalyptus pellita*.

TABLE 2 | Homolog of T3SE genes in *X. perforans* TC2-1.

Gene loci	Protein	Organism of the validated protein	Sequence similarity (%)
TC2-1_000053	AvrBs2	*Xanthomonas oryzae* pv. *oryzae* KACC10331	89.747
TC2-1_000255	XopR	*Xanthomonas campestris* pv. *vesicatoria* 85-10	99.476
TC2-1_000349	HrpF	*Xanthomonas campestris* pv. *vesicatoria* 85-10	98.025
TC2-1_000355	HrpE	*Xanthomonas campestris* pv. *vesicatoria* 85-10	77.419
TC2-1_000366	HrpB2	*Xanthomonas campestris* pv. *vesicatoria* 85-10	99.231
TC2-1_000381	XopA	*Xanthomonas campestris* pv. *vesicatoria* 85-10	97.436
TC2-1_000382	BapC	*Burkholderia pseudomallei* K96243	44.355
TC2-1_000383	XopM	*Xanthomonas campestris* pv. *vesicatoria* 85-10	98.266
TC2-1_000509	XopX	*Xanthomonas oryzae* pv. *oryzae* KACC10331	59.556
TC2-1_000682	XopV	*Xanthomonas campestris* pv. *vesicatoria* 85-10	94.207
TC2-1_001172	Cpn0490	*Chlamydophila pneumoniae* CWL029	30.405
TC2-1_001178	RipTPS	*Ralstonia solanacearum* GMI1000	40.839
TC2-1_001236	CT610	*Chlamydia trachomatis* D/UW-3/CX	23.077
TC2-1_001263	XAC3090	*Xanthomonas axonopodis* pv. *citri* str. 306	81.559
TC2-1_001267	XopK	*Xanthomonas campestris* pv. *vesicatoria* 85-10	98.118
TC2-1_001324	Mlr6361	*Mesorhizobium loti* MAFF303099	39.568
TC2-1_001337	AvrBsT	*Xanthomonas euvesicatoria* Bv5-4a	23.427
TC2-1_001346	RipAL	*Ralstonia solanacearum* UW551	46.779
TC2-1_001847	RipAP	*Ralstonia solanacearum* UW551	29.338
TC2-1_002077	XopZ	*Xanthomonas fuscans* subsp. *aurantifolii* str. ICPB 10535	91.354
TC2-1_002411	Pgl	*Bradyrhizobium diazoefficiens* USDA 110	38.235
TC2-1_002958	XopF2	*Xanthomonas campestris* pv. *vesicatoria* 85-10	99.279
TC2-1_002961	XopN	*Xanthomonas oryzae* pv. *oryzae* KACC10331	63.15
TC2-1_003035	T3SE	*Xanthomonas campestris* pv. *badrii* NEB122	90.72
TC2-1_003236	RipC1	*Ralstonia solanacearum* GMI1000	47.291
TC2-1_003276	SrfJ	*Salmonella enterica* subsp. *enterica* serovar Typhimurium str. SL1344	33.556
TC2-1_003677	HopAK1	*Pseudomonas syringae* pv. *tomato* str. DC3000	23.333
TC2-1_003780	HopK1	*Pseudomonas syringae* pv. *tomato* str. DC3000	37.681
TC2-1_003954	HopG1	*Pseudomonas syringae* pv. *tomato* str. DC3000	47.216
TC2-1_004171	AvrXacE1	*Xanthomonas axonopodis* pv. *citri* str. 306	92.269
TC2-1_004204	XopS	*Xanthomonas campestris* pv. *vesicatoria* 85-10	95.548
TC2-1_004427	XopT	*Xanthomonas fragariae* NBC2815	90.78
TC2-1_004430	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	65.672
TC2-1_004432	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	67.164
TC2-1_004435	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	67.164
TC2-1_004439	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	68.657
TC2-1_004442	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	65.672
TC2-1_004445	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	69.231
TC2-1_004449	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	68.657
TC2-1_004455	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	67.164
TC2-1_004461	XopT	*Xanthomonas fragariae* NBC2815	81.667
TC2-1_004471	XopT	*Xanthomonas fragariae* NBC2815	85
TC2-1_004490	XopT	*Xanthomonas fragariae* NBC2815	80
TC2-1_004495	XopT	*Xanthomonas fragariae* NBC2815	85
TC2-1_004498	AvrXccC	*Xanthomonas campestris* pv. *campestris* 8004	92.145
TC2-1_004521	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	70.769
TC2-1_004526	HopAO1	*Pseudomonas syringae* pv. *tomato* str. DC3000	32.845
TC2-1_004532	HopF4	*Pseudomonas savastanoi* NCPPB 3335	77.193
TC2-1_004536	AvrBs3	*Xanthomonas campestris* pv. *vesicatoria* 85-10	92.354
TC2-1_004538	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	67.164
TC2-1_004547	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	70.149
TC2-1_004562	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	64.179
TC2-1_004564	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	65.672

(Continued)

TABLE 2 | (Continued)

Gene loci	Protein	Organism of the validated protein	Sequence similarity (%)
TC2-1_004566	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	65.672
TC2-1_004580	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	74.346
TC2-1_004589	XOO4824	*Xanthomonas oryzae* pv. *oryzae* KACC10331	68.254

Genome Comparison Between Strain TC2-1 and Other *Xanthomonas perforans* Strains

To identify TC2-1-specific genes related to host specialization, we performed a phylogenetic analysis of TC2-1 and all the 141 sequenced *X. perforans* strains in the NCBI RefSeq database based on 120 conserved, single-copy genes (**Supplementary Figure 4**) and selected five representative strains for further comparative analysis, including four strains isolated from tomato leaves (strain NI1 from Northern guinea savannah, and strains Xp11-2, GEV2392, and CFBP 7293 from different regions of the US) and one (strain AL66) isolated from pepper. Proteins present only in TC2-1, but not the other five strains, were then searched against the NCBI NR database to remove those with matches of high coverage (>80%) and identity (>65%) in any *X. perforans* strains. As a result, we identified a total of 277 genes that are unique to TC2-1, encoding 121 hypothetical proteins, 101 transposases, and 7 transcriptional regulators (**Supplementary Table 8**).

It is generally believed that there is a correlation between the repertoire of T3SEs and the host specificity of pathogenic *Xanthomonas* strains (Hajri et al., 2009, 2012a,b; Jacques et al., 2016). In the TC2-1 genome, we found an intact Type-III secretion system and predicted 56 T3SEs that are homologous to validated T3SEs in *Xanthomonas* and other bacteria (**Table 2**). Notably, 15 of these predicted T3SEs share low levels of sequence similarity with known T3SEs and, thus, were classified as TC2-1 unique. Among the predicted T3SEs, 19 encode short fragments of XOO4824 or XopT (**Table 2**), sharing high identity to each other; most of them were found to be adjacent to transposable elements, resulting in multiple copies in the genome. Actually, the real XopT and XOO4824 homologs in TC2-1 are TC2-1_004427 and TC2-1_004580, respectively, with 98.80% identity to each other. In summary, the number of predicted intact T3SEs in *X. perforans* strain TC2-1 is 37 indeed.

DISCUSSION

The rapid advancements of high-throughput sequencing technologies and metagenomics in recent years have greatly promoted many types of biological studies, including plant pathology. In this study, we employed 16S rDNA gene-based metagenomic sequencing to understand the composition of microbiota associated with bacterial leaf canker of water spinach, which provides an important guide for pathogen identification. Metagenomic analysis showed that *Xanthomonas* is overwhelmingly dominant in all the diseased leaf samples, and *Pantoea* is also present in two of the leaf samples, which is in sharp contrast with rhizospheric soil and irrigation water samples (**Figure 2**). Species identification unveiled them as *X. perforans* and *P. ananatis* through complete genome sequencing and MLSA analysis, respectively.

Results from the metagenomic analysis revealed very low abundance of *Xanthomonas* (0.06 and 0.39%) and *Pantoea* (0∼0.15%) in water and soil samples. We cannot determine whether the *Xanthomonas* and *Pantoea* in water and soil can contribute to the disease since no isolation work has been performed from soil and water. However, from our observation on field planting, farmers prefer to reserve stumps from the previous year and plant them directly in the field; these old stumps could more likely be the primary infection source of the disease, and the pathogens can spread in large numbers in the field with wind and rain.

Inoculation of the bacteria revealed that both *X. perforans* and *P. ananatis* are pathogenic, causing distinctive diseased symptomatic features on water spinach leaves. Previously, *X. campestris* and *X. perforans* have been isolated from water spinach in Bang Pai, Thailand (Leksomboon et al., 1991), and Fuzhou, China (information from NCBI GenBank sequence records MN626337.1 to MN626340.1), respectively, causing bacterial leaf spot. Due to the lack of description of disease symptoms, we could not judge whether the disease symptoms caused by these two pathogens are consistent with the bacterial leaf canker caused by *X. perforans* strains reported in this study. Nevertheless, *X. perforans* has been reported to infect *Eucalyptus pellita* causing bacterial leaf blight and dieback (Bophela et al., 2019), with similar necrosis symptoms to the bacterial leaf canker of water spinach. Interestingly, this disease has also been reported to be caused by *P. ananatis* previously (Coutinho et al., 2002). Inoculation of either *X. perforans* or *P. ananatis* resulted in similar diseased symptoms on *Eucalyptus* leaves (Bophela et al., 2019), suggesting that both of the pathogens can infect the same host plant, although they have not been isolated from the same host sample simultaneously before this study.

In this study, both *X. perforans* and *P. ananatis* were isolated from the same sample of water spinach. Notably, there have been an increasing number of reports of co-infection in plant diseases. For instance, *P. ananatis* was found to synergistically infect *Eucalyptus* along with a fungal pathogen, *Corniothyrium zuluense*, causing serious canker disease (Van Zyl, 1999). Recently, we also isolated two bacterial pathogens from the same diseased rice sample, namely, *Enterobacter asburiae* and *P. ananatis*, both of which were determined as the causal agents of rice bacterial blight with no detectable synergistic or antagonistic interactions (Xue et al., 2020). Moreover, *Dickeya zeae* and *Stenotrophomonas maltophilia* have been isolated from the same tissue of soft rot leaf of *Clivia miniata* with no

synergism (Hu et al., 2018, 2021). These, and many other cases, all suggest the common phenomenon of co-infection of two or more pathogens on host plants in nature.

The interaction between co-infected pathogens is synergistic, mutualistic, or antagonistic. Synergistic interactions of plant pathogens usually result in increased disease severity, the most typical example of which is the tomato pith necrosis caused by pathogen complexes including *Pseudomonas corrugata* (Scarlett et al., 1978), *Ps. mediterranea* (Catara et al., 2002), *Ps. marginalis* (Bella and Catara, 2009), *Ps. fluorescens*, *Ps. putida*, *Ps. citronellolis*, *Ps. straminea*, *Pantoea agglomerans*, and *X. perforans* (Aiello et al., 2013). The disease is greatly aggravated when co-inoculation with *Ps. corrugata* and *Ps. mediterranea* (Moura et al., 2005). In some cases, the synergism among pathogens could be indirect as for instance, some pathogens can suppress host immunity and promote colonization by other ones. On the other hand, antagonistic interaction between mixed pathogens often results from nutrition competition. In a study, growth interference was observed between different strains of *Ps. syringae* in a mixed infection (Bartoli et al., 2015). In our case, *X. perforans* is most likely the major pathogen based on its dramatically higher abundance and isolation rate, as well as its typical characteristics of crater-shaped ulcerative spots on leaves, whereas, *P. ananatis* is possibly a companion pathogen causing yellowing and brown rot lesions on leaves. In this study, we performed co-inoculation of *X. perforans* and *P. ananatis* on the host plant, and results showed that co-inoculation of both pathogens could aggravate the disease development under the condition of non-wound inoculation (**Figure 4**), suggesting that there may be synergistic interaction between *P. ananatis* TC3-1 and *X. perforans* TC2-1 under natural infection. From our observation, the main disease symptom in the field is the leaf canker caused by *Xanthomonas*, whereas *Pantoea* is also present in the field according to the results of 16S rDNA amplicon sequencing and laboratory microbial isolation toward symptomatic leaves, suggesting that *P. anantis* might behave as a latent pathogen within a susceptible leaf tissue. To further verify the pathogenicity of *P. ananatis* TC3-1, we inoculated it into tobacco leaves, and results showed that *P. ananatis* TC3-1 is able to induce hypersensitivity reaction (HR), although the HR response was not as strong as *X. perforans* TC2-1 (**Supplementary Figure 5**).

The genus *Xanthomonas* currently comprises over 35 species, most of which cause plant diseases in more than 400 different hosts (Timilsina et al., 2020). In 1991, *X. perforans* was first identified in tomato fields (Jones, 1995) and has outcompeted *X. euvesicatoria* by 2006 in Florida (Horvath et al., 2012). Until a single-strain Xp2010 was isolated from an infected pepper sample in 2010 in Florida (Schwartz et al., 2015), the host range of *X. perforans* was believed to be restricted to tomato. Findings in this study expand the host range of both *X. perforans* and *P. ananatis*. We also tested the pathogenicity of *X. perforans* TC2-1 on pepper, tomato, citrus leaves, and mango fruit, and found that strain TC2-1 could infect all of the above plants (**Supplementary Figure 6**). To identify genes that might be responsible for the host specificity of *X. perforans* TC2-1, its genome was sequenced and compared with the genomes of other *X. perforans*, which were all isolated from tomato and pepper. In total, 277 TC2-1-specific proteins were found, most of which encode transposases, suggesting a frequent genomic recombination of the strain, and more informatively, 23 predicted T3SEs were found to be absent in the other *X. perforans*-released genomes (**Supplementary Table 8**). The repertoire of T3SEs has been believed to be correlated with the host specificity of pathogenic *Xanthomonas* strains (Hajri et al., 2009, 2012a,b; Jacques et al., 2016). Whether these TC2-1 unique T3SEs play an important role in host specialization awaits further functional investigation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JZ conceived and designed the experiments. MH isolated and identified the pathogens and performed the pathogenicity tests. MH, YX, SW, AH, SC, and XM investigated the disease incidence. MH, CL, and XZ analyzed the metagenomic and genome data. MH, CL, JZ, and XZ wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Key-Area Research and Development Program of Guangdong Province (2018B020205003 and 2020B0202090001), the National Natural Science Foundation of China (31972230), the Natural Science Foundation of Guangdong Province, China (2020A1515011534), and the China Scholarship Council fellowship program grant (202108440367) to MH.

ACKNOWLEDGMENTS

We thank Zhuojun Xi of the Dongguan Agro-Tech Extension and Management Office for his arrangement of the disease surveys in Dongguan City.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.752760/full#supplementary-material

REFERENCES

Abby, S. S., and Rocha, E. P. (2017). "Identification of protein secretion systems in bacterial genomes using MacSyFinder," in *Bacterial Protein Secretion Systems*, eds L. Journet and E. Cascales (New York, NY: Humana Press), 1–21. doi: 10.1007/978-1-4939-7033-9_1

Aiello, D., Scuderi, G., Vitale, A., Firrao, G., Polizzi, G., and Cirvilleri, G. (2013). A pith necrosis caused by *Xanthomonas perforans* on tomato plants. *Eur. J. Plant Pathol.* 137, 29–41. doi: 10.1007/s10658-013-0214-7

An, Y., Wang, J., Li, C., Leier, A., Marquez-Lago, T., Wilksch, J., et al. (2018). Comprehensive assessment and performance improvement of effector protein predictors for bacterial secretion systems III, IV and VI. *Brief. Bioinform.* 19, 148–161.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44, W16–W21. doi: 10.1093/nar/gkw387

Bartoli, C., Lamichhane, J. R., Berge, O., Guilbaud, C., Varvaro, L., Balestra, G. M., et al. (2015). A framework to gauge the epidemic potential of plant pathogens in environmental reservoirs: the example of kiwifruit canker. *Mol. Plant Pathol.* 16, 137–149. doi: 10.1111/mpp.12167

Bella, P., and Catara, V. (2009). Occurrence of tomato pith necrosis caused by *Pseudomonas marginalis* in Italy. *New Dis. Rep.* 19:58.

Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing Group, Lau, B. Y., Hoad, G., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343

Bi, D., Liu, L., Tai, C., Deng, Z., Rajakumar, K., and Ou, H. Y. (2013). SecReT4: a web-based bacterial type IV secretion system resource. *Nucleic Acids Res.* 41, D660–D665. doi: 10.1093/nar/gks1248

Bophela, K. N., Venter, S. N., Wingfield, M. J., Duran, A., Tarigan, M., and Coutinho, T. A. (2019). *Xanthomonas perforans*: a tomato and pepper pathogen associated with bacterial blight and dieback of *Eucalyptus pellita* seedlings in Indonesia. *Australas. Plant Pathol.* 48, 543–551. doi: 10.1007/s13313-019-00657-9

Brady, C., Cleenwerck, I., Venter, S. N., Vancanneyt, M., Swings, J., and Coutinho, T. A. (2008). Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Syst. Appl. Microbiol.* 31, 447–460. doi: 10.1016/j.syapm.2008.09.004

Catara, V., Sutra, L., Morineau, A., Achouak, W., Christen, R., and Gardan, L. (2002). Phenotypic and genomic evidence for the revision of *Pseudomonas corrugata* and proposal of *Pseudomonas mediterranea* sp. nov. *Int. J. Syst. Evo. Microbiol.* 52, 1749–1758. doi: 10.1099/00207713-52-5-1749

Cerkauskas, R. F., Koike, S. T., Azad, H. R., Lowery, D. T., and Stobbs, L. W. (2006). Diseases, pests, and abiotic disorders of greenhouse-grown water spinach (*Ipomoea aquatica*) in Ontario and California. *Can. J. Plant Pathol.* 28, 63–70. doi: 10.1080/07060660609507271

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chen, Y., Dong, Y., Liang, Z., Zhang, L., and Deng, Y. (2018). Enhanced vascular activity of a new chimeric promoter containing the full CaMV 35S promoter and the plant xylogen protein 1 promoter. *3 Biotech* 8:380. doi: 10.1007/s13205-018-1379-8

Coenye, T., Falsent, E., Vananneyt, M., Hostef, B., Govant, J. R. W., Kerster, S. K., et al. (1999). Classification of *Alcaligenes faecalis*-like isolates from the environment and human clinical samples as *Ralstonia gilardii* sp. nov. *Int. J. Syst. Bacteriol.* 49, 405–413. doi: 10.1099/00207713-49-2-405

Coutinho, T. A., Preisig, O., Mergaert, J., Cnockaert, M. C., Riedel, K.-H., Swings, J., et al. (2002). Bacterial blight and die-back of *Eucalyptus* species, hybrids and clones in South Africa. *Plant Dis.* 86, 20–25. doi: 10.1094/PDIS.2002.86.1.20

Duke, J. A., and Ayensu, E. S. (1985). *Medicinal Plants of China*, Vol. 2. Algonac, MI: Reference Publications.

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 1–14. doi: 10.1186/s13059-015-0721-2

Gao, R. X., Zhang, C. Q., and Chen, M. Y. (1985). A preliminary report on the integrated control of Kankon White rust in Fuzhou. *J. Fujian Agri. Coll.* 14, 81–86.

Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., et al. (2009). A «repertoire for repertoire» hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS One* 4:e6632. doi: 10.1371/journal.pone.0006632

Hajri, A., Brin, C., Zhao, S., David, P., Feng, J. X., Koebnik, R., et al. (2012a). Multilocus sequence analysis and type III effector repertoire mining provide new insights into the evolutionary history and virulence of *Xanthomonas oryzae*: phylogeny and effector genes of *Xanthomonas oryzae*. *Mol. Plant Pathol.* 13, 288–302. doi: 10.1111/j.1364-3703.2011.00745.x

Hajri, A., Pothier, J. F., Fischer-Le Saux, M., Bonneau, S., Poussier, S., Boureau, T., et al. (2012b). Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. *Appl. Environ. Microbiol.* 78, 371–384. doi: 10.1128/AEM.06119-11

Hill, J. E., Penny, S. L., Crowell, K. G., Goh, S. H., and Hemmingsen, S. M. (2004). cpnDB: a chaperonin sequence database. *Genome Res.* 14, 1669–1675. doi: 10.1101/gr.2649204

Ho, B. W., and Edie, H. H. (1969). White rust (*Albugo* ipomoeae-aquaticae) of *Ipomoeae aquatica* in Hongkong. *Plant Dis. Repir.* 53, 959–962.

Horvath, D. M., Stall, R. E., Jones, J. B., Pauly, M. H., Vallad, G. E., Dahlbeck, D., et al. (2012). Transgenic resistance confers effective field level control of bacterial spot disease in tomato. *PLoS One* 7:e42036. doi: 10.1371/journal.pone.0042036

Hu, M. H., Yuan, J. H., Yang, X. E., and Jiang, H. (2010). Study on nutraceutical properties of different cultivars *Ipomoea aquatica* Forsskal ('Chunbai' and 'Liulv') in an eutrophic water body. *Sci. Hort.* 124, 419–422. doi: 10.1016/j.scienta.2010.01.011

Hu, M., Li, C. H., Xue, Y., Hu, A. Q., Chen, S. H., Chen, Y. F., et al. (2021). Isolation, characterization, and genomic investigation of a phytopathogenic strain of *Stenotrophomonas* maltophilia. *Phytopathology*. doi: 10.1094/PHYTO-11-20-0501-R

Hu, M., Li, J. L., Chen, R. T., Li, W. J., Feng, L. W., Shi, L., et al. (2018). Dickeya zeae strains isolated from rice, banana and clivia rot plants show great virulence differentials. *BMC Microbiol.* 18:136. doi: 10.1186/s12866-018-1300-y

Hu, Y., Huang, H., Cheng, X., Shu, X., White, A. P., Stavrinides, J., et al. (2017). A global survey of bacterial type III secretion systems and their effectors. *Environ. Microbiol.* 19, 3879–3895. doi: 10.1111/1462-2920.13755

Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* 33, 1635–1638. doi: 10.1093/molbev/msw046

Ismail, A., and Fun, C. S. (2003). Determination of vitamin C, β-carotene and riboflavin contents in five green vegetables organically and conventionally grown. *Mal. J. Nutr.* 9, 31–39.

Jacques, M. A., Arlat, M., Boulanger, A., Boureau, T., Carrère, S., Cesbron, S., et al. (2016). Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Ann. Rev. Phytopathol.* 54, 163–187. doi: 10.1146/annurev-phyto-080615-100147

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9, 1–8. doi: 10.1038/s41467-018-07641-9

Jones, J. B. (1995). A third tomato race of *Xanthomonas campestris* pv. vesicatoria. *Plant Dis.* 79, 395. doi: 10.1094/PD-79-0395

Kido, K., Hasegawa, M., Matsumoto, H., Kobayashi, M., and Takikawa, Y. (2010). Pantoea ananatis strains are differentiated into three groups based on reactions of tobacco and welsh onion and on genetic characteristics. *J. Gen. Plant Pathol.* 76, 208–218. doi: 10.1007/s10327-010-0230-9

Lana, U. G. D. P., Gomes, E. A., Silva, D. D., Costa, R. V., Cota, L. V., Parreira, D. F., et al. (2012). Detection and molecular diversity of *Pantoea ananatis* associated with white spot disease in maize, sorghum and crabgrass in brazil. *J. Phytopathol.* 160, 441–448. doi: 10.1111/j.1439-0434.2012.01924.x

Lawal, U., Maulidiani, M., Shari, K., Ismail, I. S., Khatib, A., and Abas, F. (2017). Discrimination of Ipomoea aquatica cultivars and bioactivity correlations using NMR-based metabolomics approach. *Plant Biosyst.* 151, 833–843. doi: 10.1080/11263504.2016.1211198

Lawal, U., Mediani, A., Maulidiani, H., Shaari, K., Ismail, I. S., Khatib, A., et al. (2015). Metabolite profiling of *Ipomoea aquatica* at different growth stages in correlation to the antioxidant and α-glucosidase inhibitory activities elucidated by 1H NMR-based metabolomics. *Sci. Hort.* 192, 400–408. doi: 10.1016/j.scienta.2015.06.036

Leksomboon, C., Thaveechai, N., and Kositratana, W. (1991). "Bacterial leaf spot of Chinese convolvulus [Ipomoea aquatica]," in *Proceedings of the 29th Kasetsart University: Plant Science*, (Bangkok: Kasetsart University), 277–283.

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, J., Yao, Y., Xu, H. H., Hao, L., Deng, Z., Rajakumar, K., et al. (2015). SecReT6: a web-based resource for type VI secretion systems found in bacteria. *Environ. Microbiol.* 17, 2196–2202. doi: 10.1111/1462-2920.12794

Li, W. J., Hu, M., Xue, Y., Li, Z. J., Zhang, Y. F., Zheng, D. X., et al. (2020). Five fungal pathogens are responsible for bayberry twig blight and fungicides were screened for disease control. *Microorganisms* 8:689. doi: 10.3390/microorganisms8050689

Liu, P. Q., Wei, M. Y., Zhu, L., and Li, B. J. (2017). First report of spot blight on water spinach (*Ipomoea aquatica*) caused by *Stagonosporopsis cucurbitacearum* in China. *Plant Dis.* 101:838. doi: 10.1094/PDIS-10-16-1485-PDN

Mackeen, M. M., Ali, A. M., Abdullah, M. A., Nasir, R. M., Mat, N. B., Razak, A. R., et al. (1997). Antinematodal activity of some Malaysian plant extracts against the pine wood nematode, *Bursaphelenchus xylophilus*. *Pest Sci.* 51, 165–170. doi: 10.1002/(SICI)1096-9063(199710)51:2<165::AID-PS623>3.0.CO;2-2

Malalavidhane, T. S., Wickramasinghe, S. N., and Jansz, E. R. (2000). Oral hypoglycaemic activity of *Ipomoea aquatica*. *J. Ethnopharmacol.* 72, 293–298. doi: 10.1016/S0378-8741(00)00217-8

Moura, M. L., Jacques, L. A., Brito, L. M., Mourao, I. M., and Duclos, J. (2005). Tomato pith necrosis caused by P-corrugata and P-mediterranea: severity of damages and crop loss assessment. *Acta Hort.* 695, 365–372. doi: 10.17660/ActaHortic.2005.695.45

Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J., and Stead, D. (2007). Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *Int. J. Syst. Evol. Microbiol.* 57, 2881–2887. doi: 10.1099/ijs.0.65220-0

Pierre-Alain, C., Parks, D. H., Christian, R., Phillip, H., Maria, C., Adam, S., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Prasad, K. N., Divakar, S., Shivamurthy, G. R., and Aradhya, S. M. (2005b). Isolation of a free radical-scavenging antioxidant from water spinach (*Ipomoea aquatica* Forsk). *J. Sci. Food Agric.* 85, 1461–1468. doi: 10.1002/jsfa.2125

Prasad, K. N., Ashok, G., Raghu, C., Shivamurthy, G. R., Vijayan, P., and Aradhya, S. M. (2005a). In vitro cytotoxic properties of Ipomoea aquatica leaf. *Indian J. Pharmacol.* 37:397. doi: 10.4103/0253-7613.19079

Safeeulla, K. M., and Thrumalachar, M. J. (1953). Morphological and cytological studies in *Albugo* speicies on *Ipomoea aquatic* and *Merremia emarginata*. *Cellule* 55, 225–232.

Sawada, K. (1922). *Descriptive Catalogue of Formosan Fungi*, Vol. 2. Amsterdam: Elsevier Science B.V, 27–31.

Scarlett, C. M., Fletcher, J. T., Roberts, P., and Lelliott, R. A. (1978). Tomato pith necrosis caused by *Pseudomonas corrugata* n. sp. *Ann. Appl. Biol.* 88, 105–114. doi: 10.1111/j.1744-7348.1978.tb00684.x

Schwartz, A. R., Potnis, N., Timilsina, S., Wilson, M., Patané, J., Martins, J. J., et al. (2015). Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front. Microbiol.* 6:535. doi: 10.3389/fmicb.2015.00535

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

Tian, Q. (2018). *Screening of DNA Barcode for Xanthomonas and Detection Techniques for Important Pathovars*. Ph.D. thesis. Beijing: Chinese Academy of Agricultural Sciences Dissertation.

Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F. F., et al. (2020). Xanthomonas diversity, virulence and plant-pathogen interactions. *Nat. Rev. Microbiol.* 18, 415–427. doi: 10.1038/s41579-020-0361-8

Tseng, C. F., Iwakami, S., Mikajiri, A., Shibuya, M., Hanaoka, F., Ebizuka, Y., et al. (1992). Inhibition of in vitro prostaglandin and leukotriene biosyntheses by cinnamoyl-beta-phenethylamine and N-acyldopamine derivatives. *Chem. Pharm. Bull.* 40, 396–400. doi: 10.1248/cpb.40.396

Van Zyl, L. M. (1999). *Factors Associated With Coniothyrium Canker Of Eucalyptus In South Africa*. Ph.D. thesis. Bloemfontein: University of the Orange Free State.

Wang, Y. Y., Shi, Y. X., Chai, A. L., Xie, X. W., and Li, B. J. (2017). Identification of *Myrothecium roridum* causing leaf spot on water spinach. *Acta Phytopathol. Sin.* 1, 117–121.

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:e1005595. doi: 10.1371/journal.pcbi.1005595

Xue, Y., Hu, M., Chen, S. S., Hu, A. Q., Li, S. M., Han, H. Y., et al. (2020). *Enterobacter* asburiae and *Pantoea ananatis* causing rice bacterial blight in China. *Plant Dis.* 105. doi: 10.1094/PDIS-10-20-2292-RE

Yu, Y., Chen, J., Liu, C., Gao, Y., Wang, J. Y., Yang, Y. J., et al. (2015). Study on occurring characteristics of *Albugo* ipomoeae-aquaticae on *Ipomoea aquatica* Forsk and its controlling techniques in Shanghai. *China Plant Prot.* 35, 38–42.

Zou, L., Nan, C., and Hu, F. (2013). Accurate prediction of bacterial type IV secreted effectors using amino acid composition and PSSM profiles. *Bioinformatics* 29, 3135–3142. doi: 10.1093/bioinformatics/btt554

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fmicb.2021.815911

Genome Mining Shows Ubiquitous Presence and Extensive Diversity of Toxin-Antitoxin Systems in *Pseudomonas syringae*

Prem P. Kandel[1], Marina Naumova[1], Chad Fautt[1], Ravikumar R. Patel[2], Lindsay R. Triplett[2] and Kevin L. Hockett[1,3]**

[1] Department of Plant Pathology and Environmental Microbiology, Pennsylvania State University, University Park, PA, United States, [2] Department of Plant Pathology and Ecology, The Connecticut Agricultural Experiment Station, New Haven, CT, United States, [3] The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA, United States

OPEN ACCESS

Edited by:
Ralf Koebnik, Plant Health Institute, France

Reviewed by:
Marie-Laure PINEL-MARIE, INSERM U1230 ARN Régulateurs Bactériens et Médecine (BRM), France
Marcus Vinicius Merfa, Auburn University, United States

***Correspondence:**
Prem P. kandel
puk155@psu.edu
Kevin L. Hockett
klh450@psu.edu

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: *16 November 2021*
Accepted: *13 December 2021*
Published: *12 January 2022*

Citation:
Kandel PP, Naumova M, Fautt C, Patel RR, Triplett LR and Hockett KL (2022) Genome Mining Shows Ubiquitous Presence and Extensive Diversity of Toxin-Antitoxin Systems in Pseudomonas syringae. Front. Microbiol. 12:815911. doi: 10.3389/fmicb.2021.815911

Bacterial toxin-antitoxin (TA) systems consist of two or more adjacent genes, encoding a toxin and an antitoxin. TA systems are implicated in evolutionary and physiological functions including genome maintenance, antibiotics persistence, phage defense, and virulence. Eight classes of TA systems have been described, based on the mechanism of toxin neutralization by the antitoxin. Although studied well in model species of clinical significance, little is known about the TA system abundance and diversity, and their potential roles in stress tolerance and virulence of plant pathogens. In this study, we screened the genomes of 339 strains representing the genetic and lifestyle diversity of the *Pseudomonas syringae* species complex for TA systems. Using bioinformatic search and prediction tools, including SLING, BLAST, HMMER, TADB2.0, and T1TAdb, we show that *P. syringae* strains encode 26 different families of TA systems targeting diverse cellular functions. TA systems in this species are almost exclusively type II. We predicted a median of 15 TA systems per genome, and we identified six type II TA families that are found in more than 80% of strains, while others are more sporadic. The majority of predicted TA genes are chromosomally encoded. Further functional characterization of the predicted TA systems could reveal how these widely prevalent gene modules potentially impact *P. syringae* ecology, virulence, and disease management practices.

Keywords: *Pseudomonas syringae*, toxin-antitoxin systems, phylogroup, genomics, stress

INTRODUCTION

The *Pseudomonas syringae* species complex consists of a monophyletic group of plant-pathogenic, plant-commensal, and environmental isolates within the major clade of the genus *Pseudomonas* (Xin et al., 2018). The strains are grouped into 13 phylogroups by multilocus sequence analysis (Berge et al., 2014) and phylogenomics (Gomila et al., 2017; Dillon et al., 2019a; Newberry et al., 2019). Based on host range and symptomatology, *P. syringae* strains have been also classified into over 60 pathovars that together affect almost all known crop plants of economic importance. Periodic outbreaks, including a recent devastating epidemic of bacterial canker in kiwifruit in

New Zealand (McCann et al., 2013), establish these plant pathogens as a serious threat to global food production. In addition to being a pathogen, *P. syringae* survives well in the environment and is commonly found in precipitation and natural bodies of water (Morris et al., 2008, 2013). Epiphytic and environmental populations face extreme fluctuations in UV light, temperature, humidity, and nutrient availability, as well as antimicrobial compounds produced by plants and their microbiota (Hirano and Upper, 2000; Lindow and Brandl, 2003; Craig et al., 2021). Previous studies have reported induction of viable but not culturable (VBNC) state and a concomitant alteration in the gene expression profile upon exposure of *P. syringae* cells to conditions that mimic oxidative burst of early stage plant infection (Mock et al., 2015; Postnikova et al., 2015). We recently reported that a plant pathogenic *P. syringae* can survive bacteriocin and antibiotics exposure by entering another dormant, tolerant state known as persistence (Kandel et al., 2020; Patel et al., 2021). Although several stress response pathways have been identified, it is still not well understood how *P. syringae* can cope with environmental stresses in epiphytic and free-living conditions or maintain virulence genes in the extended absence of host selection pressure.

Toxin-antitoxin (TA) systems are small, self-regulating genetic elements, comprised of two to three adjacent genes. The toxin component interferes with an essential cellular function such as translation, DNA replication, and cell wall biosynthesis, whereas the cognate antitoxin neutralizes toxin activity (Unterholzner et al., 2013b). Because the antitoxin is typically less stable than the toxin, conditions that affect antitoxin expression or degradation can increase the pool of unbound toxin, inhibiting cellular processes (Unterholzner et al., 2013b). TA systems have been broadly classified into eight types depending on the nature of antitoxin and its mode of interaction with the toxin (Song and Wood, 2020b). Most well-studied TA systems are Type II, characterized by protein antitoxins that directly bind and neutralize toxin active sites (unlike the type I and III systems, where antitoxins are non-coding RNAs). TA systems were initially implicated as a plasmid maintenance mechanism that suppressed the growth of plasmid-free daughter cells in the absence of antitoxin expression (Ogura and Hiraga, 1983). Subsequent studies have identified TA genes in bacterial chromosomes, which can similarly prevent loss of genomic islands (Wozniak and Waldor, 2009; Yao et al., 2015). TA loci have been implicated in a wide range of roles in addition to plasmid and genome maintenance, including physiological tolerance to antibiotics (Moyed and Bertrand, 1983), phage defense (Sberro et al., 2013; Song and Wood, 2020a), biofilm formation, and virulence (Wood and Wood, 2016). Recent studies have demonstrated that TA systems can impact the ecology, plasmid maintenance, chemical control tolerance, and virulence capabilities of plant pathogens. For example, in *Acidovorax citrulli,* which causes bacterial fruit blotch of cucurbits, expression of a TA operon was significantly induced during plant infection (Shavit et al., 2016). In *Xylella fastidiosa,* overexpression of a TA toxin decreased virulence and increased persistence to copper exposure (Merfa et al., 2016), while in a separate study, deletion of another TA system caused a hypervirulence phenotype suggestive of a regulatory role governing systemic infection (Burbank and Stenger, 2017). In *Xanthomonas oryzae* pv. *oryzicola*, the toxin of one TA system functions as a virulence effector delivered into host cells via the type III secretion system (Triplett et al., 2016). Moreover, in *Erwinia amylovora*, a 10-fold greater tolerance to streptomycin was reported when a TA toxin gene was expressed (Peng et al., 2021). TA systems were also shown to play a significant role in maintenance of *P. syringae* virulence plasmids (Bardaji et al., 2019). While TA systems have been predicted in several plant pathogenic species, including in *Xanthomonas citri* and *Erwinia amylovora* (Martins et al., 2016; Shidore et al., 2019), TA system composition and diversity has not been described in the *P. syringae* species complex.

In this study, we used genome mining to predict TA systems in 339 *P. syrinage* strains representing the broad diversity of the species complex, examined their abundance, diversity, and association with strain phylogeny, source of isolation, and virulence determinants. We show that homologs of well-characterized TA systems are ubiquitous and present in multiple copies in *P. syringae* genomes. We identified six major TA families conserved across species, and others that have more sporadic distributions. Most TA systems occurred in the chromosome and were likely associated to mobile elements and virulence factors. Further experimental characterization of the predicted TA system will reveal their significance in the context of *Pseudomonas syringae* biology, ecology, and virulence, and can provide important insights on how these bacteria can survive in very diverse and harsh environments. These insights could be important in designing new management options for the pathogens.

MATERIALS AND METHODS

Selection of Strains, Phylogenomics, and Phylogroup Designation

We retrieved a set of 2,345 genomes identified as *Pseudomonas syringae* complex in the NCBI database as of October 2020, and generated a preliminary phylogenetic tree using the evolutionary distance estimation method of andi with default settings (Haubold et al., 2014). From this tree, we manually extracted 357 genomes representing the genetic diversity within the *P. syringae* complex. We then assessed the completeness of the assembly and annotation with the benchmarking tool BUSCO (Simão et al., 2015), using the Pseudomonadales single-copy gene ortholog dataset. Genomes with a BUSCO score of less than 95% complete genes were filtered out, and 339 genomes were retained for further analysis (**Supplementary Table 1**).

A phylogenomic tree of the 339 strains was constructed using GToTree v1.5.46 using Gammaproteobacteria gene profiles (Lee, 2019). Sequences of 172 homologous gene sets were used to construct the tree. The clade containing *P. fuscovaginae* SE-1 was used as an outgroup to root the tree. Clustering of strains with previously known phylogroup designation was used as the basis for assigning phylogroups to unknown strains, after confirming that strains of the same phylogroup clustered

together. Together, we included 77, 104, 38, 31, 11, 2, 61, 2, 4, 4, 2 strains from phylogroups 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, and 13, respectively. The *P. fuscovaginae* group clade was not assigned to any phylogroup. Our genome list did not have any representatives from phylogroups 8 and 12 due to unavailability of quality genomes in the public database.

Prediction of Toxin-Antitoxin Systems

Type II and IV TA systems were predicted for whole genomes of all 339 strains using SLING 2.0.1 (Horesh et al., 2018), a program that identifies gene pairs based on Hidden Markov Models (HMM) alignments to annotated genes or unannotated putative ORFs, as well as gene pair structural requirements. Nucleotide sequence, annotations, and toxin profile HMMs included with the program were used as inputs with default parameters. SLING was run on Roar high-performance computing platform using 16 processors at Penn State. SLING output consisted of 260 hits that contained a gene with a known toxin domain (**Supplementary Figure 1**). A hit is a CDS with a HMMER bit score of at least 20 for overall sequence/profile comparison with a previously characterized and/or predicted toxin CDS included in the program. Hits with the same toxin gene domain were given separate designations using an *e*-value cutoff of 0.01 and an identity score of 30% as reported previously (Horesh et al., 2018). Toxin hits sharing common domain names were considered homologs and combined into one category to simplify the results. For example, the *parE* toxin gene included 47 hits that were combined to form a single ParDE TA system. The homologs with the same hit label (eg., 4H-*parE*) across the strains (i.e., orthologs) were highly conserved (> 70% amino acid identity) in their amino acid sequence, while hits within a strain (i.e., paralogs) with different labels (eg., 4H-*parE* and 11H-*parE*) were divergent (< 30% identity) (**Supplementary Figure 2**). To check for completeness of the SLING prediction, we acquired the TA systems of the three model strains (DC3000, B728a, and 1448A) predicted in TADB2.0, an updated database of bacterial type II TA loci (Xie et al., 2018). The TAFinder tool of TADB2.0 was also run with default parameters using the Refseq accession numbers of the model strains.

To predict TA systems not included in the training sets of above programs, an HMM alignment approach was used to screen genomes for type I, III, V, and VI systems. For each family of toxin and protein antitoxin to be searched, a minimum of 20 amino acid sequences representing that family across diverse bacterial species were retrieved from UniProt. Sequences were aligned with MAFFT v 7.450 using default settings and an HMM profile was built using default settings of the hmmbuild function in the HMMER webserver (Finn et al., 2011). A locally created target database of 339 *P. syringae* genomes translated into six reading frames was used in HMMER's HMMsearch. The TA systems screened this way were type I (*hok*/*sok*, *ldrD*/*rdlD*, *tisB*/*istR-1*, *symE*/*symR*, *ibsC*/*sibC*, *txpA*/*ratA*, *fst*/*RNAI-RNAII*, *pndA*/*pndB*, *shoB*/*ohsC*); type III (*toxN*/*toxI*, *abiQ*/*antiQ*, *cptI*/*cptN*, *tenpN*/*tenpI*); type V (*ghoT*/*ghoS*); and type VI (*socB*/*socA*). A type I TA system predicted for strain DC3000 in the type I toxin-antitoxin database T1TAdb (Tourasse and Darfeuille, 2021) was included in the analysis after confirming its presence in our genomes by nucleotide and protein BLAST.

Heatmaps were created using the Interactive Tree of Life (iTOL) v5 (Letunic and Bork, 2021) using the phylogenomic tree and TA system abundance used as annotation. Predicted TA systems were categorized based on their frequency among strains, as prevalent (if at least one homolog of the given toxin domain is present in ≥ 80% of strains), common (present in 20–79% of strains), or rare (present in < 20% of the strains).

Association of Toxin-Antitoxin System Abundance With Phylogroup, Genome Size, Plasmid Content, Source of Isolation, and Presence of Virulence Determinants

Differences between phylogroups in TA system abundance were determined from the SLING output and uniquely predicted TA system of TADB2.0 (*psyrTA*), and T1TAdb (*hok-sok*). Genome size was defined as the assembly size and was recorded for all strains, while plasmid counts were based on analysis of 28 closed genomes with resolved plasmids. Chromosome or plasmid localization of genes was determined by manually examining the start and end nucleotide positions of predicted toxin genes. Isolation sources were obtained from strain information within NCBI annotation files, with all isolates from non-plant sources classified as "other" (**Supplementary Table 1**). To identify correlations between TA system abundance and the abundance of predicted type III effectors, we used bioinformatically predicted effector repertoires of 140 of the 339 of our genomes as computed in a previous study (Dillon et al., 2019b). Putative secretion signals in toxin and antitoxin genes of *P. syringae* strain DC3000 was predicted using EffectiveDB (Eichinger et al., 2016).

GC Content, Codon Usage Variation, Gene Neighborhood, and Strain Grouping by Toxin-Antitoxin Abundance

To identify signs of TA system acquisition through horizontal transfer, GC content of top 20 abundant toxin hits was extracted and was compared to that of the genomic GC content. Codon usage was extracted from the Codon Usage Database.[1] The gene neighborhood of the most ubiquitous three TA system homologs (*pasTI*, *RES-Xre*, and *vapBC*) for the three strains DC3000, B728a, and 1448A was visualized using gene graphics.[2] Grouping of strains based on similar TA system content was performed using SLING outputs of the toxin hits using the average linkage method for clustering and the Euclidean method for distance measurement in heatmapper (Babicki et al., 2016).

Sequence Search, Blast, Alignment, and Phylogenetics

Sequences of the orthologs and paralogs of the same toxin hits (e.g., *parE* had 47 total homologs with up to 12 paralogs predicted

[1]https://www.kazusa.or.jp/codon/

[2]https://katlabs.cc/genegraphics/app

in a single genome) were retrieved from the model strains DC3000, B728a, and 1448A. Alignment of the sequences was performed in Muscle 3.8.425 within Geneious using 10 maximum iterations. A phylogenetic tree of the homologs was constructed using FastTree 2.1.11 using default parameters.

Literature Search for Toxin-Antitoxin System Expression Patterns

We identified 10 papers for which RNA-Seq or microarray data were reported for strains DC3000, B728a, and 1448A under environmental or mutant conditions. Fold change or $\log_2$ fold change data was obtained from supplementary materials, and locus IDs corresponding to SLING-predicted TA systems were manually extracted using the start and end nucleotide positions of TA genes in the genome. Significantly differentially expressed loci were reported here as reported by the authors; no new data analysis was performed. Moreover, we used data from a recent study on genome-wide identification of fitness conferring genes in strain B728a and analyzed the fitness profiles of TA genes from their supplementary data (Helmann et al., 2019).

Statistical Analyses

Statistical analysis of difference in TA counts by phylogroup was performed in JMP Pro 15 using the non-parametric Kruskal-Wallis test ($P < 0.05$) followed by each pair multiple comparisons after confirming that assumptions of parametric tests are not satisfied as per the goodness of fit test and unequal variance test in JMP. Phylogroups having fewer than 10 strains were excluded from this analysis. Correlation analysis between the TA counts and genome size, plasmid content, and type III effector count was also performed in JMP Pro 15 using the multivariate option and Spearman's ρ. GC content differences were assessed using Kruskal-Wallis test with a genomic GC used as control group in Steel method.

RESULTS

P. syringae Genomes Encode 26 Toxin-Antitoxin System Families Targeting Diverse Cellular Functions, of Which Six Are Prevalent Throughout the Species

As a basic step in understanding the roles of TA systems in *P. syringae* ecology, we aimed to obtain a comprehensive list and abundance patterns of TA systems across the species complex using bioinformatics prediction tools including SLING, BLAST, and HMMER. We predicted 26 distinct TA system families throughout the *P. syringae* genomospecies (**Figure 1** and **Supplementary Figure 3**). Of these, 24 belonged to the type II TA system families, and one each belonged to type I (*hok-sok*) and type IV system (a*biEii*) families. The majority of TA system toxins predicted are orthologs of toxins in other species, which have been previously reported to inhibit translation through mRNA cleavage or through modification of translation factors, while others are reported to inhibit replication, cell wall biosynthesis, NAD + homeostasis, and other essential functions (**Table 1**). To confirm the comprehensiveness of the SLING method, predictions from three *P. syringae* model strains were compared to that from the database TADB2.0. The majority of the SLING predictions were also identified in TADB2.0 (7 of 12 in B728a, 14 of 19 in DC3000, and 1448A), including five of the six most prevalent systems described below. Six TA systems, including the highly conserved *pasTI* module, were only predicted using SLING. A TA system from a validated family, *psyrTA* (Sberro et al., 2013), was predicted in TADB2.0 but not SLING. Few TA genes were uniquely predicted in TADB2.0 and TAFinder (**Supplementary Table 2**).

Genomes varied substantially in TA system repertoires, and only six TA system families were highly prevalent (i.e., present in ≥ 80% of the genomes analyzed (**Figures 1A,B**). These included systems encoding putative translation inhibitor toxins (*pasTI*, *vapBC*, *hipAB*, and *higBA*), a replication inhibitor (*parDE*), and a RES domain toxin proposed to interfere with cellular NAD^+ homeostasis (**Table 1**). Eight additional predicted systems present in more than 20% of the strains were categorized as common TA systems, and the remaining 12 were rare (**Figure 1B**). Prevalent TA system families included sets of toxin orthologs that diverged across strains despite sharing the common toxin domain, and some included multiple distinct paralogs within a genome. For example, for the *parE* toxin gene of the *parDE* system, 47 variants were predicted (**Supplementary Figure 1**). Similarly, the *vapBC* family had 21 distinct variants. A conserved copy of *pasTI*, *RES-xre*, and *vapBC* systems were predicted in all or most genomes (100, 94, and 87%, respectively), suggesting an important conserved function.

A *hok-sok* module of type I TA system was predicted in the strain DC3000 in the T1TAdb database (Tourasse and Darfeuille, 2021). Nucleotide BLAST of the predicted sequence in the locally constructed database of our 339 genomes confirmed presence of full complement of this system in strains of phylogroup 1 (**Figure 1A**). Other genomes contained orthologs of only the toxin component. Alignment of the predicted Hok sequence to previously characterized Hok homologs showed low identity (15% to that of pR1 plasmid and 13% to that of *Erwinia amylovora*). No TA toxin genes for any of the type I, III, V, and VI systems were detected using BLAST and HMMER alignment methods. All strains contained a homolog to the proposed type IV TA system *cptB/sdhE* antitoxin that is encoded in an operon with a hypothetical protein of unknown function (DC3000 locus ID: PSPTO_4227), but no *cptA* toxin was identified, and *sdhE* is an essential flavinylation factor in bacteria, so this is not a likely TA system. A few potential homologs were detected for antitoxins *abiGii* (type IV), *ghoS* (type V), and *socA* (type VI). Since these did not occur in the type strains and no toxin homolog was detected, no further assessment was performed.

Toxin-Antitoxin System Abundance Differs by Phylogroup of *P. syringae*

As TA system abundance greatly varied across genomes with a minimum of 5 to a maximum of 35 TA systems (median of 15) predicted per genome of *P. syringae*, we

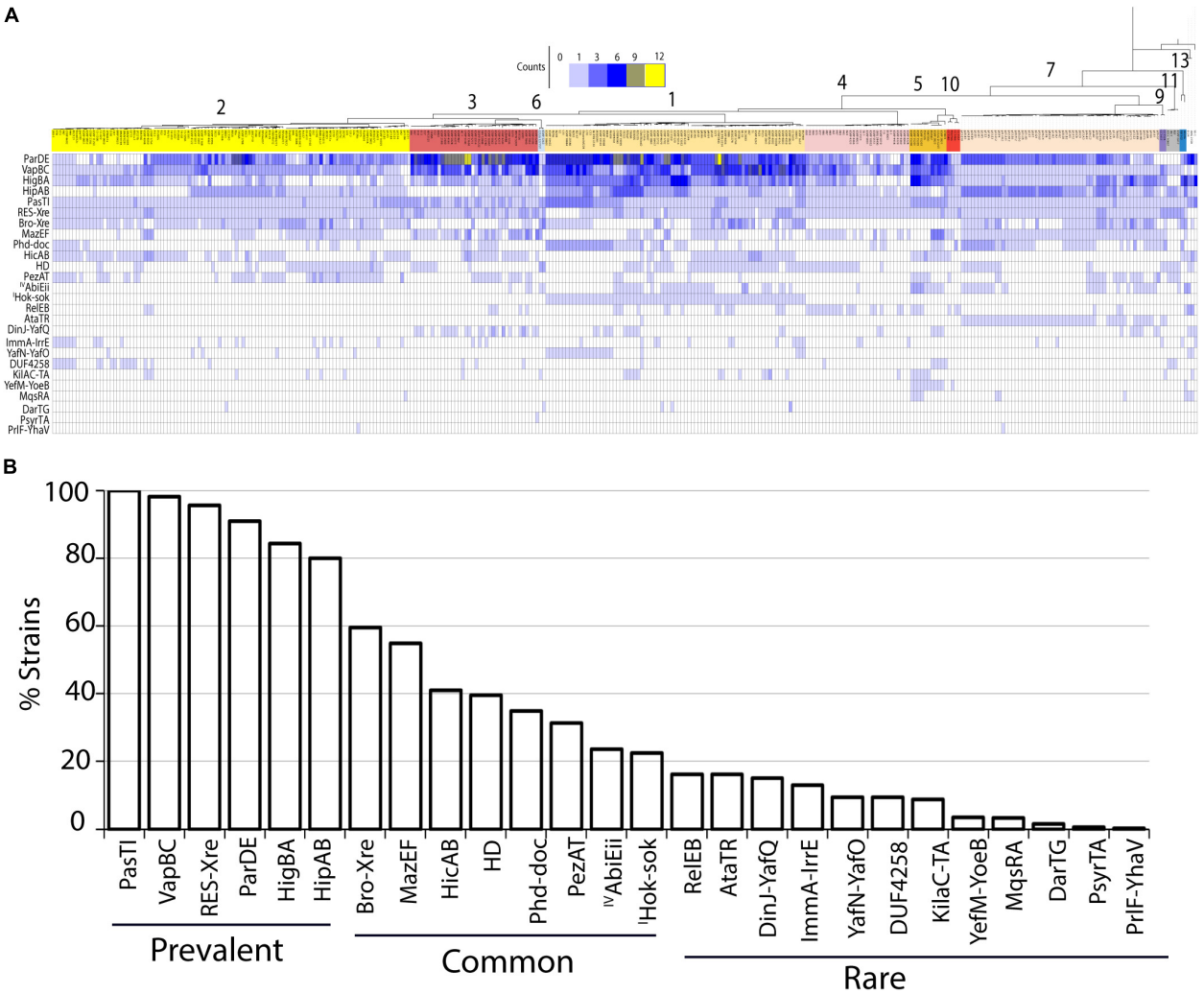


FIGURE 1 | (A) Heatmap showing TA system content and distribution across *P. syringae* species complex. Names of TA systems are indicated for each row. Number of homologs of each TA systems are indicated by the color key at the top. Phylogenetic tree color coded according to phylogroups are indicated by numerals. **(B)** Percentage of strains containing the TA systems. List of all the TA systems found in *P. syringae* are shown in the x-axis, and the fraction of strains carrying a given TA system out of the 339 strains is shown in the y-axis. TA systems were categorized as prevalent, common, and rare based on their abundance as shown. Except *abiEii* (type IV), and *hok-sok* (type I) all belong to type II TA families.

aimed to understand the determinants that influence TA system abundance. TA system abundance varied significantly across phylogroups (**Figure 2A**). Phylogroups 1, 3, and 5 had significantly more TA systems per genome than other phylogroups (**Figure 2A**). These differences were mainly driven by a large variation in the within-genome abundance of three of the prevalent TA systems, *parDE*, *vapBC*, and *higBA* (**Figure 2B**). In addition to the copy number differences, phylogroups also differed in composition of TA system families. For example, putative *bro-xre* systems were common in phylogroup 2, but absent in phylogroups 4, 5, and 10. The *mazEF* system was ubiquitously predicted in phylogroup 7 strains but was sporadic in others (**Figures 1A**, **2B**). Similarly, full complement of the type I *hok-sok* system was predicted only in phylogroup 1 (**Figure 1A**). Phylogroups could also be distinguished by divergent orthologs of the same families; for example, the HipA-5H homologs were predominantly found in phylogroups 7, while the HipA-6H group was concentrated in phylogroups 1 and 4 (**Supplementary Figure 1**).

In a clustering analysis based on TA system repertoires, TA system composition followed phylogroup structure in *P. syringae*, with >90% of strains clustering with their phylogroup, suggesting that most TA system acquisition happened prior to phylogroup diversification (**Supplementary Figure 4**). However, at least 26 strains clustered outside their phylogroups, and phylogroups 1, 4, and 5 were significantly split into two or more clusters, potentially indicative of more recent TA system gain or loss in these strains. Overall, genetic similarity (i.e., phylogroup association) is a major factor governing similarities in TA system abundance.

TABLE 1 | Toxin-antitoxin systems predicted in *P. syringae*, toxin function and physiological effect.

TA system	Toxin domains	Toxin function	Function affected	References
ParDE^{P}	ParE	DNA Gyrase inhibitor	Replication	Jiang et al., 2002
VapBC^{P}	PIN, PIN_3, DUF4411	Cleavage of tRNA^{fMet}	Translation	Winther and Gerdes, 2011
HigBA^{P}	HigB, HigB-like, Gp49	50S ribosome dependent mRNA cleavage	Translation	Hurley and Woychik, 2009
HipAB^{P}	HipA_C, HipA_C-HipA_N-Couple_hipA	Phosphorylation of glutamyl tRNA synthase (GltX)	Translation	Germain et al., 2013
PasTI^{P} (YfjG-YfjF)	RatA, polyketide_cyc2	Binds 50S ribosomal subunit and blocks 70S ribosome formation	Translation	Zhang and Inouye, 2011
RES-*xre*^{P}	RES	ADP ribosylates phosphoribosyl pyrophosphate synthetase (Prs)	NAD + homeostasis	Piscotta et al., 2019
Bro-TA^{C}*	Bro-N	?	?	
MazEF^{C}	PemK	Ribosome independent mRNA and rRNA cleavage	Translation	Zhang et al., 2003
Phd-doc^{C}	Fic, DUF4172	Phosphorylates elongation factor Tu (EF-TU)	Translation	Cruz et al., 2014
HicAB^{C}	HicA	Ribosome independent mRNA cleavage	Translation	Jørgensen et al., 2009
HD-TA^{C}	HD	c-di-AMP hydrolysis	Signaling?	Huynh et al., 2015
PezAT^{C} (Epsilon-Zeta)	Zeta	Phosphorylate uridine diphosphate-N-acetylglucosamine (UNAG)	Cell wall biosynthesis	Mutschler et al., 2011
AbiEii^{C} (MenTA)	AbiEii NTP_transf_2	Nucleotidyltransferase on serine tRNA	Translation	Cai et al., 2020
Hok-sok^{C}	Hok	Membrane depolarization	Membrane	Gerdes et al., 1986
RelEB^{R}	RelE_like	Ribosome bound mRNA cleavage	Translation	Christensen and Gerdes, 2003
AtaTR^{R}	GNAT_acetyltran	Transfers an acetyl group from acetyl coenzyme A to Met-tRNAfMet.	Translation	Jurënas et al., 2017
DinJ-YafQ^{R}	YafQ	Ribosome dependent mRNA cleavage	Translation	Prysak et al., 2009
ImmA-IrrE^{R}	Peptidase_M78, DUF955	Metalloprotease, activation causes transcriptional induction of genes required for repair and survival after radiation exposure	Cleaves repressor, and induces transcription	Ludanyi et al., 2014
YafN-YafO^{R}	YafO	Ribosome dependent mRNA cleavage	Translation	Zhang et al., 2009
DUF4258^{R}	DUF4258, CdiA	DNA double strand break (contact dependent inhibition protein)	DNA damage	Roussin et al., 2019
KilAC-TA^{R}*	ANT	? (Could be involved in phage repression)	?	
YefM-YoeB^{R}	YoeB	50S ribosomal subunit dependent mRNA cleavage at A site	Translation	Zhang and Inouye, 2009
MqsRA^{R}	MqsR	Ribosome independent mRNA cleavage	Translation	Yamaguchi et al., 2009
DarTG^{R}	DUF4433	ADP ribosylates DNA	Replication	Jankevicius et al., 2016
PsyrTA^{R}	RecQ	ATP-dependent DNA helicase	?	Sberro et al., 2013
PrlF-YhaV^{R}	Toxin_YhaV	Ribosome dependent mRNA cleavage	Translation	Choi et al., 2017

**These systems were predicted by bioinformatics analysis. Overexpression was toxic to cell, but TA function was not confirmed (Chan et al., 2014).* ^{P}*prevalent,* ^{C}*common,* ^{R}*rare.*

Toxin-Antitoxin System Abundance Is Positively Linked to Genome Size and Plasmid Number, and Most Toxin-Antitoxin Systems Are Chromosomally Encoded

TA systems frequently occur in mobile genetic elements (MGE), that are known to be horizontally transferred. As the genome sizes increase due to integration of MGEs or presence of plasmids, we hypothesize that the number of TA systems will also increase. We performed a correlation analysis on 339 strains to determine whether predicted TA system abundance is associated with genome size. Genome size was weakly, though significantly, correlated to TA system count ($\rho = 0.52$, $P < 0.0001$, **Figure 3A**). There was a similar positive correlation observed between TA system count and the number of plasmids ($\rho = 0.44$, $P = 0.018$, **Figure 3B**).

Next, we examined whether the TA systems predicted in the 28 complete *P. syringae* genomes were chromosome- or plasmid-borne. The majority (45–100%, mean = 87.7 ± 14.3) of predicted TA systems were encoded in the chromosome (**Figure 3C** and **Supplementary Figure 3**). The most widely distributed TA systems, with *pasT*, *vapC*, and RES domain toxins, were exclusively or almost exclusively detected on chromosomes, though the common and rare *vapC* homologs also occurred in plasmids (**Figure 3C**). Of the 40 plasmids analyzed, 33 contained at least one TA system with a mean count of 1.8 TA systems per plasmid, and 7 plasmids did not contain TA systems. A maximum

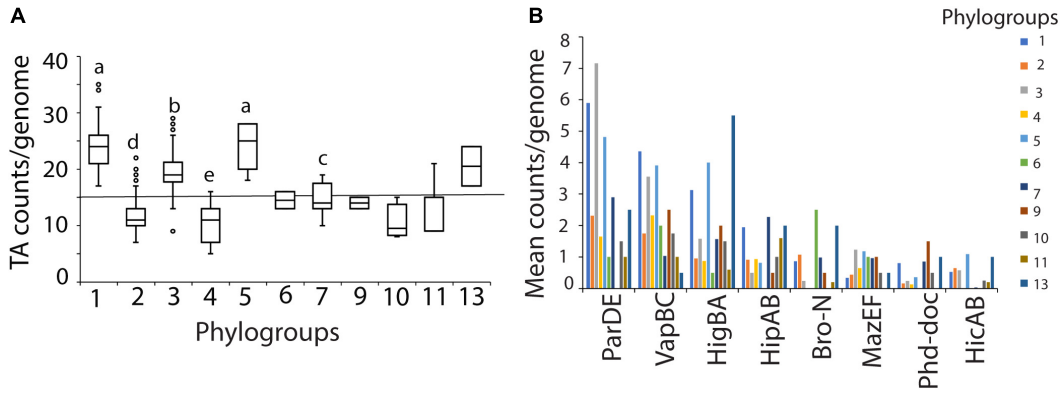


FIGURE 2 | (A) TA counts per genome by phylogroups. The central line shows median number of TA system genes from the 339 strains used. Different letters indicate significant difference in counts as analyzed by Kruskal-Wallis test at $P < 0.05$. Phylogroups with less than 10 strains were not used in the analysis. **(B)** Differences in major TA system content among phylogroups. Mean counts per genome by phylogroup are shown. Phylogroups differed in copy number per genome of *parDE*, *vapBC* and *higBA* families (also refer heatmap in **Figure 1A**).

of eight TA systems were predicted in the large plasmid of the strain 1448A.

We next analyzed GC content, codon usage variation, and genomic context of TA genes to detect if they contain signs of horizontal transfer. By comparing the GC content of top 20 toxin hits to that of the genome, we showed that GC content of 17 of the 20 toxin genes significantly differed from that of the genome. GC content varied by > 5% in 8 of the 20, and by > 3% in 12 of the 20 toxin genes (**Supplementary Figure 5**). Although we did not detect a noticeable codon usage variation compared to the genome in strain DC3000, few toxin genes such as *parE*, *hicA* and *higB* showed minor differences (**Supplementary Figure 6**). Upon examining the genomic context of the three most abundant and highly conserved TA system variants (i.e., *pasT*, RES, and *vapC*), we observed that these systems are immediately flanked by or lie within 20 kb region from MGE in at least one of the strains (**Supplementary Figure 7**). Together, these results suggest that most TA systems in *P. syringae* occur in mobile elements and are likely to be acquired by horizontal transfer.

Isolation Source of Strains Does Not Affect Toxin-Antitoxin System Abundance, and Toxin-Antitoxin Gene Counts Are Weakly Correlated to Type III Effector Gene Counts

A previous study reported higher abundance of TA systems in free-living than in host-associated isolates (Pandey and Gerdes, 2005). We therefore analyzed if the source of isolation of the strains influences the abundance of TA system in *P. syringae*. We did not detect any difference in TA abundance based on the environment from which strains were isolated (**Figure 4A** and **Supplementary Figure 8**). Notably, strain Psy642, which forms a separate subclade 2c within phylogroup 2, lacks the major virulence determinant of canonical type III secretion system, and was reported to be avirulent in a previous study (Demba Diallo et al., 2012), showed a different pattern of TA content than other strains of the same phylogroup (**Supplementary Figure 1**). Psy642 and two other strains in this clade did not contain the *vapBC* genes (**Figure 1A**). The highly prevalent *vapC*-3H, *parE*-4H, and *hipA*-5H homologs are missing in these strains (**Supplementary Figure 1**). Since TA systems are known to stabilize genomic islands that contain virulence and antibiotics resistant determinants (Wozniak and Waldor, 2009; Yao et al., 2015), we hypothesized that the predicted type III secretion system effector counts are correlated to TA gene counts. To this end, we examined the association between the number of TA loci and type III effectors, as predicted in a previous study (Dillon et al., 2019b). Although a similar trend was observed in the TA gene counts and type III effectors in the three phylogroups (1, 2, and 3) that are dominated by plant pathogens, overall, a weak positive correlation ($\rho = 0.38$, $P < 0.0001$) was detected between the two variables (**Figure 4B**). We also attempted to examine the association of TA genes to markers of copper and streptomycin resistance, bactericides that are frequently used in control of *P. syringae* diseases. However, these markers could be predicted only in a limited number of strains based on BLAST search of the resistant genes as query against our local database of 339 genomes.

EffectiveT3 predicted four TA gene products of DC3000 as type III secreted proteins. These included antitoxins of *vapBC*-3H (PSPTO_2000), *vapBC*-26H (PSPTO_1058), and toxins of HD-15H (PSPTO_2479), and *ataTR* (PSPTO_RS175570).

Published Gene Expression and Fitness Profiles for Predicted Toxin-Antitoxin Systems Indicate Distinct Regulatory Patterns

Although TA systems are posttranslationally regulated, some antitoxins are repressors of the TA operon, and conditions leading to antitoxin degradation and excess free toxin might lead to induction of the TA operon (Chan et al., 2016). Therefore, expression patterns could point to conditions in

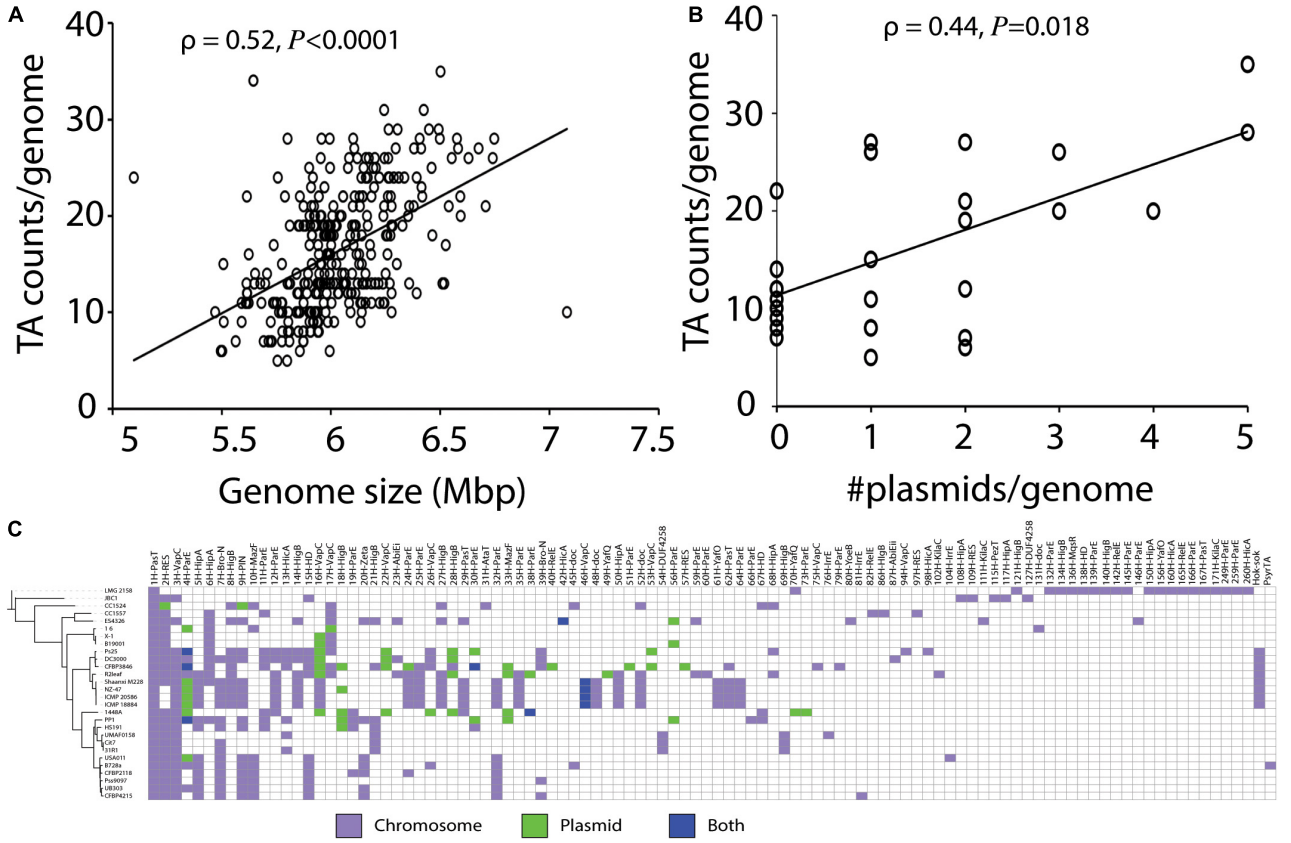


FIGURE 3 | (A) Association of TA counts by genome size and **(B)** plasmid counts. **(C)** Heatmap showing occurrence of all 260 homologs of TA systems in the 28 completely sequenced strains in chromosome or plasmids. Most TA systems were encoded in the chromosome.

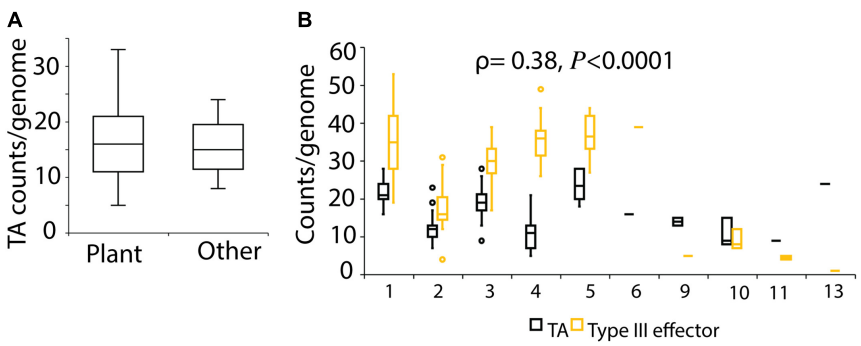


FIGURE 4 | (A) TA counts by isolation source of strains. Strains isolated outside of plant environment (rain, snow, stream, lake water, epilithon) were categorized as other (*n* = 13). **(B)** TA counts and type III effector counts by phylogroups (x-axis). Type III secretion system effector counts predicted in 140 genomes from a previous study (Dillon et al., 2019b) were used. Strains of phylogroup 7 were not used in this analysis as the strains used here were not used in Dillon et al. (2019b).

which certain toxins might be functional. We compiled reported transcriptional differences from nine RNA-seq studies and one well-replicated microarray study, all performed on one of the model strains DC3000, B728a, or 1448A. Seven of the studies reported significant differential expression of one or more of the four highly conserved TA system genes (*pasTI, RES-xre, vapBC, and parDE)* under host or abiotic stress, or in virulence regulatory mutants (**Table 2**), while three studies found no difference. In microarray analysis of strain B728a, epiphytic growth and nitrogen starvation induced 3 TA systems. *vapC* toxin genes were reported to be substantially induced during nitrogen starvation (**Supplementary Table 3**). The expression of *RES-Xre* was suppressed during early infection of DC3000 on *Arabidopsis* plants and was conditionally affected by several regulatory gene

deletions in 1448A (**Table 2**). While these expression patterns do not provide conclusive evidence of a function or lack thereof, they indicate that the different TA system families are regulated in distinct and complex pathogen and environmental contexts.

In a recent Rb-TnSeq study aimed at identifying genome-wide fitness profile of B728a genes, three antitoxin gene mutants (that of two *parDE* and a *vapBC*) showed below threshold abundance in the BarSeq library (**Supplementary Figure 9**). The authors suggested that these genes could be important for *in vitro* growth and survival. Similarly, although not reported to be significant in the paper, the plant fitness score was lower for antitoxin mutants of *RES-xre*, *hipAB*, *psyrTA*, and a *vapBC* system (**Supplementary Figure 9**). We speculate that the disruption of antitoxin genes by transposon insertion could have freed their cognate toxins. The freed toxins subsequently caused growth inhibition or death of the mutants lowering their abundance in the library. Interestingly, toxin gene mutants of the *pasTI* system showed below threshold abundance in the library, which could suggest a different regulation of this TA system or multiple cellular functions of the predicted toxin gene. Moreover, in the study that validated TA system function of *psyrTA* mentioned above, a *parDE* system is reported to be functional based on cloning success of predicted TA genes during Sanger-based whole-genome shotgun sequencing (Sberro et al., 2013).

DISCUSSION

Using *in silico* prediction from a broad-scale genomic data, we show that TA systems are prevalent and occur in diverse patterns in *Pseudomonas syringae* species complex. We found that 100% of screened genomes contain at least five TA systems with a median count of 15 per genome across the species. We predicted 26 distinct TA families that almost exclusively belong to the type II group, except for a type IV system that was sporadically detected in some strains and a type I system detected in phylogroup 1. Of the 26 TA systems, six are highly abundant across the species implying potentially conserved roles in *P. syringae* ecology. The most prevalent among the TA systems was that of the *pasTI* system, occurring mostly as a single copy in all strains with a high degree of sequence conservation. Although limited data exists regarding the functional characterization, a previous study in *Escherichia coli* CFT073 showed that this TA system is involved in stress tolerance under nutrient limitation, oxidative stress, and antibiotic persistence (Norton and Mulvey, 2012). Another recent study has confirmed its role in stress survival but contradicted its function as a TA system (Fino et al., 2020). Interestingly, The *pasT* gene was recently shown to be important for *in vitro* fitness in *P. syringae*, indicating a functional role (Helmann et al., 2019).

Orthologs of other prevalent TA loci (*vapBC*, *RES-xre*, *parDE*, *higBA*, *hipAB*) have been widely described to function as TA systems. We predicted multiple paralogs of *parDE*, *vapBC*, and *higBA* systems in some phylogroups and strains. We determined that sequences of paralogs are highly divergent. Whether all paralogs are active and if they are also functionally diverse is a subject of further research. The *parDE* and *vapBC* homologs were also among the most prevalent in another plant pathogen, *Erwinia amylovora* (Shidore et al., 2019). A *vapBC* ortholog was characterized as a TA system in the plant pathogen *Acidovorax citruli*, and was shown to be overexpressed during plant infection (Shavit et al., 2016). We predicted that some of the *vapB* antitoxin sequences contain a type III secretion signal, indicating a potential role in virulence. The role of *vapBC* in stress responses is reported in several other species (Arcus et al., 2010). The *parDE* system has been implicated in persistence against DNA gyrase inhibiting antibiotics in *P. aeruginosa* (Muthuramalingam et al., 2019). Homologs are also known for plasmid and genome maintenance in several other species including plant pathogens (Roberts et al., 1994; Yuan et al., 2011; Unterholzner et al., 2013a). Similarly, TA systems containing the RES domain toxin was also prevalent, and a copy was highly conserved across species. Although this system is previously characterized as a bona fide TA system including in *P. putida* (Skjerning et al., 2019), its physiological role remains to be determined. *higBA* system, another prevalent TA system in *P. syringae*, was shown to regulate virulence factor production in *P. aeruginosa* (Wood and Wood, 2016). Similarly, *hipAB* system has been shown to function in antibiotics persistence *in vitro* (Moyed and Bertrand, 1983; Germain et al., 2013) and also in the natural host environment (Schumacher et al., 2015). Taken together, the prevelance of these TA systems in *P. syringae* and their characterized roles in other taxa indicate that they may play multiple roles in the ecology of this phytopathogen. Further experiments will be able to reveal more specific details on their roles and the conditions under which they are activated.

TA genes that were less prevelant but were commonly found in *P. syringae* included toxin families of well characterized TA systems in other species. Some notable examples in this category were *mazEF*, *hicAB*, *phd-doc*, *pezAT*, *abiEii,* and *hok-sok*. Previous studies have reported stress and virulence associated function of these TA genes. For example, MazF toxin of the *mazEF* system was shown to contribute *Mycobacterium tuberculosis* cells survive oxidative stress, nutrient depletion, and virulence in the host (Tiwari et al., 2015). Rare TA systems also contained TA candidates that have been experimentally verified in other systems, including toxin families RelE, YafO, YafQ, AtaT (GNAT), YoeB, MqsR, and YhaV. Some candidates of both common and rare TA systems predicted in this study are less well studied and have not been demonstrated to function as TA systems experimentally, such as the Bro-N and KilaC(ANT) toxins (Chan et al., 2014). Only further experimental research can validate if they function as a TA system in *P. syringae* or are involved in other functions.

Our results show that TA system abundance differs by phylogroups. In addition to the genetic similarity, strains belonging to the same phylogroup are generally expected to have similar phenotypic traits consistent with their unique ecology. For instance, phylogroup 2 strains were shown to be more aggressive in cantaloupe seedling, were most consistently ice-nucleation active, and more likely to produce syringomycin-like toxin than strains of other phylogroups. Similarly, phylogroup 3 strains are rarely known to possess ice-nucleation activity

TABLE 2 | Reported expression patterns of TA system genes in three model strains B728a, 1448A, DC3000.

Strain	Condition	Vs.	PasT		RES		VapC-3H		ParE-4H		References
			A	T	A	T	A	T	A	T	
B728a	H_2O_2	HMM									Yu et al., 2013
	Apoplast	HMM									
	Epiphytic	HMM									
	Low Fe	HMM									
	Low N	HMM									
	Low NaCl	HMM									
B728a	30°C	28°C									Hockett et al., 2013
1448A	Δ*lon*	wt									Hua et al., 2020
1448A (MM)	Δ*OmpR*	Δ*ompS*									
	Δ*phoP*	Δ*phoQ*									
	Δ*pilR*	Δ*pilS*									
	Δ*psrA*	wt									
	Δ*rhpR*	Δ*rhpS*									
	Δ*vfr*	wt									
	Δ*algU*	wt									
	Δ*gacA*	wt									Shao et al., 2021
	Δ*OmpR*	Δ*ompR*									
	Δ*phoP*	Δ*phoQ*									
1448A (KB)	Δ*pilR*	Δ*pilS*									
	Δ*psrA*	wt									
	Δ*rhpR*	Δ*rhpS*									
	Δ*rpoN*	wt									
	Δ*vfr*	wt									
DC3000	Δ*algU*	wt									Markel et al., 2016
	At PTI (1h)	Naive 1									
	PTI 3 h	Naive 3 h									
	PTI 5 h	Naive 5 h									
	Naïve 1 h	inoculum									
DC3000	Naive 3 h	inoculum									
	Naive 5 h	inoculum									
	PTI 3 h	inoculum									Lovelace et al., 2018
	PTI 1 h	inoculum									
	PTI 5 h	inoculum									
	Naive 1 h	Naive 3 h									
	Naive 1 h	Naive 5 h									
	Naive 3 h	Naive 5 h									
	PTI 1 h	PTI 3 h									
	PTI 1 h	PTI 5 h									
	PTI 3 h	PTI 5 h									

Green box indicate significantly upregulated gene, blue box indicate the significantly downregulated gene and white box indicate no significant difference in the expression level in comparison with control.

Bronstein et al. (2008), (DC3000): RES Antitoxin, down regulated and ParE-1 Toxin, upregulated due to ferric citrate and sodium citrate treatment at 4 h compared to control respectively.

Lam et al. (2014), (DC3000)- hrpRS and HrpL mutants: no effect on any of these four TA systems.

Filiatrault et al. (2010), (DC3000): no effect on any of these four TA systems.

Santamaría-Hernando et al. (2020)20, (DC3000): no effect on TA systems of blue-light perception.

Greenwald et al. (2012), (B728a)-B728a ΔacsS vs. B728a: no effect on any of these four TA systems.

Liu et al. (2020), (DC3000, B728a) DC3000 Δ(p)ppGpp vs. DC3000 and B728a Δ(p)ppGpp vs. B728a: None of these TA systems are differentially expressed.

A, antitoxin; T, toxin; HMM, HRP MM-mannitol medium; Δlon, Lon knockout mutant; ΔOmpR, ΔompS, ΔphoP, ΔphoQ, ΔpilR, ΔpilS, ΔpsrA, ΔrhpR, ΔrhpS, Δvfr, ΔalgU, ΔgacA, and ΔrpoN, deletion mutant of Type 3 secretion system regulators; MM, M9 minimum medium; KB, King's B medium; PTI, pattern-triggered immunity; 1 h, 1-h post inoculation; 3 h, 3-h post inoculation; 5 h, 5-h post inoculation; hrpRS$^-$ and hrpL$^-$, deletion mutant of hrpRS and hrpL gene; and DC3000 Δ(p)ppGpp and B728a Δ(p)ppGpp, ppGpp gene knockout mutant of DC3000 and B728a, respectively.

and are negative for syringomycin-like toxins (Berge et al., 2014). Previous studies have also reported phylogroup dependent differences in the content of virulence factors (Baltrus et al., 2012; Dillon et al., 2019b). While it is impossible to speculate the significance of these difference at this point, similar sporadic distribution of TA system linked to genetic similarity has been also described previously both in clinical (Horesh et al., 2020) and plant pathogens (Shavit et al., 2016; Shidore et al., 2019).

TA systems are mostly known to occur in the accessory genome and are associated to MGEs or plasmids (Fraikin et al., 2020). Although we predicted that 82.5% (33 out of 40) of plasmids contain at least one TA system, a vast majority of TA systems are encoded in the chromosome in *P. syringae*. Chromosomally encoded TA systems are also reported to contribute to stability of genomic islands (Wozniak and Waldor, 2009; Yao et al., 2015). Moreover, virulence and antimicrobial resistance determinants are known to occur in laterally acquired genomic islands including in *P. syringae* (Kim et al., 1998; Feil et al., 2005). A comparison of the counts of type III effectors and TA loci showed similar trends in three of the phylogroups that are dominated by plant pathogens. Phylogroup 2 strains, which are known to be broad-host range and contain relatively smaller repertoire of type III effectors compared to other phylogroups (Baltrus et al., 2012), also contained lower counts of TA genes in our study. Moreover, strain CC1557 that was shown to contain unusually lower counts of type III effectors in a previous study (Hockett et al., 2014) was predicted to encode fewer of TA system genes (8 TA systems encoded compared to a median of 15). It could be possible that the TA systems occur in same genomic islands as the virulence and resistant determinants and could prevent their loss, imparting adaptational benefits. Alternatively, these results might indicate that strains that rely on larger repertoires of type III effectors, also rely on larger TA system repertoires in a manner related to their ecology. Testing these alternative hypotheses will require a combination of additional bioinformatic analyses and experimentation.

Taken together, we show that *P. syringae* genomes encode multiple TA systems that target diverse cellular functions upon activation. Based on our results of the abundance patterns and evidence of the ecological roles discussed above, we speculate that TA systems could be important as virulence effectors, in maintaining genes of virulence and antimicrobial resistance, in antimicrobial and environmental stress tolerance, in defending phage predation among others in *P. syrinage*. Further research focusing on the experimental characterization and ecological roles of the predicted TA systems can provide important insights in the epidemiology and management of plant disease caused by *P. syringae*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PK, LT, and KH: conceptualization and writing manuscript drafts. PK, MN, CF, and RP: data acquisition and analysis. LT and KH: supervision and funding acquisition. All authors: revision of drafts.

FUNDING

This research was supported by the USDA NIFA AFRI grant (2019-67013-29353) awarded to LT and KH. KH was also supported by the USDA National Institute of Food and Agriculture and Federal Hatch appropriations under project PEN04648 and accession no. 1015871.

ACKNOWLEDGMENTS

We thank David S. Guttman from University of Toronto for sharing the raw data on the predicted type III secretion system effectors of *Pseudomonas syringae*.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.815911/full#supplementary-material

Supplementary Table 1 | *Pseudomonas syringae* strain genomes used in this study.

Supplementary Table 2 | Sequences and other details on toxin antitoxin systems of the three model strains DC3000, B728a, and 1448A.

Supplementary Table 3 | Data on transcriptomics study analyzed for TA systems in this study.

REFERENCES

Arcus, V. L., McKenzie, J. L., Robson, J., and Cook, G. M. (2010). The PIN-domain ribonucleases and the prokaryotic VapBC toxin–antitoxin array. *Protein Eng. Des. Select.* 24, 33–40. doi: 10.1093/protein/gzq081

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., et al. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* 44, W147–W153. doi: 10.1093/nar/gkw419

Baltrus, D. A., Nishimura, M. T., Dougherty, K. M., Biswas, S., Mukhtar, M. S., Vicente, J., et al. (2012). The molecular basis of host specialization in bean pathovars of *Pseudomonas* syringae. *Mol. Plant Microbe Interact* 25, 877–888. doi: 10.1094/MPMI-08-11-0218

Bardaji, L., Añorga, M., Echeverría, M., Ramos, C., and Murillo, J. (2019). The toxic guardians — multiple toxin-antitoxin systems provide stability, avoid deletions and maintain virulence genes of *Pseudomonas* syringae virulence plasmids. *Mobile DNA* 10:7. doi: 10.1186/s13100-019-0149-4

Berge, O., Monteil, C. L., Bartoli, C., Chandeysson, C., Guilbaud, C., Sands, D. C., et al. (2014). A user's guide to a data base of the diversity of *Pseudomonas* syringae and its application to classifying strains in this phylogenetic complex. *PLoS One* 9:e105547. doi: 10.1371/journal.pone.0105547

Bronstein, P. A., Filiatrault, M. J., Myers, C. R., Rutzke, M., Schneider, D. J., and Cartinhour, S. W. (2008). Global transcriptional responses of *Pseudomonas* syringae DC3000 to changes in iron bioavailability *in vitro*. *BMC Microbiol.* 8:209. doi: 10.1186/1471-2180-8-209

Burbank, L. P., and Stenger, D. C. (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

Cai, Y., Usher, B., Gutierrez, C., Tolcan, A., Mansour, M., Fineran, P. C., et al. (2020). A nucleotidyltransferase toxin inhibits growth of Mycobacterium

tuberculosis through inactivation of tRNA acceptor stems. *Sci. Adv.* 6:eabb6651. doi: 10.1126/sciadv.abb6651

Chan, W. T., Espinosa, M., and Yeo, C. C. (2016). Keeping the wolves at bay: antitoxins of prokaryotic type II toxin-antitoxin systems. *Front. Mol. Biosci.* 3:9. doi: 10.3389/fmolb.2016.00009

Chan, W. T., Yeo, C. C., Sadowy, E., and Espinosa, M. (2014). Functional validation of putative toxin-antitoxin genes from the Gram-positive pathogen Streptococcus pneumoniae: phd-doc is the fourth bona-fide operon. *Front. Microbiol.* 5:677. doi: 10.3389/fmicb.2014.00677

Choi, W., Yamaguchi, Y., Lee, J.-W., Jang, K.-M., Inouye, M., Kim, S.-G., et al. (2017). Translation-dependent mRNA cleavage by YhaV in *Escherichia coli*. *FEBS Lett.* 591, 1853–1861. doi: 10.1002/1873-3468.12705

Christensen, S. K., and Gerdes, K. (2003). RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. *Mol. Microbiol.* 48, 1389–1400. doi: 10.1046/j.1365-2958.2003.03512.x

Craig, K., Johnson, B. R., and Grunden, A. (2021). Leveraging *Pseudomonas* stress response mechanisms for industrial applications. *Front. Microbiol.* 12:660134. doi: 10.3389/fmicb.2021.660134

Cruz, J. W., Rothenbacher, F. P., Maehigashi, T., Lane, W. S., Dunham, C. M., and Woychik, N. A. (2014). Doc toxin is a kinase that inactivates elongation factor Tu. *J. Biol. Chem.* 289, 7788–7798. doi: 10.1074/jbc.M113.544429

Demba Diallo, M., Monteil, C. L., Vinatzer, B. A., Clarke, C. R., Glaux, C., Guilbaud, C., et al. (2012). *Pseudomonas* syringae naturally lacking the canonical type III secretion system are ubiquitous in nonagricultural habitats, are phylogenetically diverse and can be pathogenic. *ISME J.* 6, 1325–1335. doi: 10.1038/ismej.2011.202

Dillon, M. M., Thakur, S., Almeida, R. N. D., Wang, P. W., Weir, B. S., and Guttman, D. S. (2019a). Recombination of ecologically and evolutionarily significant loci maintains genetic cohesion in the *Pseudomonas* syringae species complex. *Genome Biol.* 20:3. doi: 10.1186/s13059-018-1606-y

Dillon, M. M., Almeida, R. N. D., Laflamme, B., Martel, A., Weir, B. S., Desveaux, D., et al. (2019b). Molecular evolution of *Pseudomonas* syringae type III secreted effector proteins. *Front. Plant Sci.* 10:418. doi: 10.3389/fpls.2019.00418

Eichinger, V., Nussbaumer, T., Platzer, A., Jehl, M.-A., Arnold, R., and Rattei, T. (2016). EffectiveDB–updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res.* 44, D669–D674. doi: 10.1093/nar/gkv1269

Feil, H., Feil, W. S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., et al. (2005). Comparison of the complete genome sequences of *Pseudomonas* syringae pv. syringae B728a and pv. tomato DC3000. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11064–11069. doi: 10.1073/pnas.0504930102

Filiatrault, M. J., Stodghill, P. V., Bronstein, P. A., Moll, S., Lindeberg, M., Grills, G., et al. (2010). Transcriptome analysis of *Pseudomonas* syringae identifies new genes, noncoding RNAs, and antisense activity. *J. Bacteriol.* 192, 2359–2372. doi: 10.1128/JB.01445-09

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Fino, C., Vestergaard, M., Ingmer, H., Pierrel, F., Gerdes, K., and Harms, A. (2020). PasT of *Escherichia coli* sustains antibiotic tolerance and aerobic respiration as a bacterial homolog of mitochondrial Coq10. *Microbiol. Open* 9:e1064. doi: 10.1002/mbo3.1064

Fraikin, N., Goormaghtigh, F., Melderen, L. V., and Margolin, W. (2020). Type II toxin-antitoxin systems: evolution and revolutions. *J. Bacteriol.* 202, e763–e719. doi: 10.1128/JB.00763-19

Gerdes, K., Bech, F. W., Jørgensen, S. T., Løbner-Olesen, A., Rasmussen, P. B., Atlung, T., et al. (1986). Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon. *Embo J.* 5, 2023–2029.

Germain, E., Castro-Roa, D., Zenkin, N., and Gerdes, K. (2013). Molecular mechanism of bacterial persistence by HipA. *Mol. Cell.* 52, 248–254. doi: 10.1016/j.molcel.2013.08.045

Gomila, M., Busquets, A., Mulet, M., García-Valdés, E., and Lalucat, J. (2017). Clarification of taxonomic status within the *Pseudomonas* syringae species group based on a phylogenomic analysis. *Front. Microbiol.* 8:2422. doi: 10.3389/fmicb.2017.02422

Greenwald, J. W., Greenwald, C. J., Philmus, B. J., Begley, T. P., and Gross, D. C. (2012). RNA-seq analysis reveals that an ECF σ Factor, AcsS, regulates achromobactin biosynthesis in *Pseudomonas* syringae pv. syringae B728a. *PLoS One* 7:e34804. doi: 10.1371/journal.pone.0034804

Haubold, B., Klötzl, F., and Pfaffelhuber, P. (2014). andi: fast and accurate estimation of evolutionary distances between closely related genomes. *Bioinformatics* 31, 1169–1175. doi: 10.1093/bioinformatics/btu815

Helmann, T. C., Deutschbauer, A. M., and Lindow, S. E. (2019). Genome-wide identification of *Pseudomonas* syringae genes required for fitness during colonization of the leaf surface and apoplast. *Proc. Natl. Acad. Sci.* 116, 18900–18910. doi: 10.1073/pnas.1908858116

Hirano, S. S., and Upper, C. D. (2000). Bacteria in the leaf ecosystem with emphasis on *Pseudomonas* syringae-a pathogen, ice nucleus, and epiphyte. *Microbiol. Mol. Biol. Rev.* 64, 624–653. doi: 10.1128/MMBR.64.3.624-653.2000

Hockett, K. L., Burch, A. Y., and Lindow, S. E. (2013). Thermo-regulation of genes mediating motility and plant interactions in *Pseudomonas syringae*. *PLoS One* 8:e59850. doi: 10.1371/journal.pone.0059850

Hockett, K. L., Nishimura, M. T., Karlsrud, E., Dougherty, K., and Baltrus, D. A. (2014). *Pseudomonas* syringae CC1557: a highly virulent strain with an unusually small type III effector repertoire that includes a novel effector. *Mol. Plant Microbe Interact* 27, 923–932. doi: 10.1094/MPMI-11-13-0354-R

Horesh, G., Fino, C., Harms, A., Dorman, M. J., Parts, L., Gerdes, K., et al. (2020). Type II and type IV toxin-antitoxin systems show different evolutionary patterns in the global *Klebsiella pneumoniae* population. *Nucleic Acids Res.* 48, 4357–4370. doi: 10.1093/nar/gkaa198

Horesh, G., Harms, A., Fino, C., Parts, L., Gerdes, K., Heinz, E., et al. (2018). SLING: a tool to search for linked genes in bacterial datasets. *Nucleic Acids Res.* 46:e128. doi: 10.1093/nar/gky738

Hua, C., Wang, T., Shao, X., Yingpeng, X., Hao, H., Jingui, L., et al. (2020). *Pseudomonas* syringae dual-function protein Lon switches between virulence and metabolism by acting as both DNA-binding transcriptional regulator and protease in different environments. *Environ. Microbiol.* 22, 2968–2988. doi: 10.1111/1462-2920.15067

Hurley, J. M., and Woychik, N. A. (2009). Bacterial toxin HigB associates with ribosomes and mediates translation-dependent mRNA cleavage at A-rich sites. *J. Biol. Chem.* 284, 18605–18613. doi: 10.1074/jbc.M109.008763

Huynh, T. N., Luo, S., Pensinger, D., Sauer, J. D., Tong, L., and Woodward, J. J. (2015). An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence. *Proc. Natl. Acad. Sci. U. S. A.* 112, E747–E756. doi: 10.1073/pnas.1416485112

Jankevicius, G., Ariza, A., Ahel, M., and Ahel, I. (2016). The toxin-antitoxin system DarTG catalyzes reversible ADP-Ribosylation of DNA. *Mol. Cell.* 64, 1109–1116. doi: 10.1016/j.molcel.2016.11.014

Jiang, Y., Pogliano, J., Helinski, D. R., and Konieczny, I. (2002). ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of *Escherichia coli* gyrase. *Mol. Microbiol.* 44, 971–979. doi: 10.1046/j.1365-2958.2002.02921.x

Jørgensen, M. G., Pandey, D. P., Jaskolska, M., and Gerdes, K. (2009). HicA of *Escherichia coli* defines a novel family of translation-independent mRNA interferases in bacteria and archaea. *J. Bacteriol.* 191, 1191–1199. doi: 10.1128/JB.01013-08

Jurënas, D., Chatterjee, S., Konijnenberg, A., Sobott, F., Droogmans, L., Garcia-Pino, A., et al. (2017). AtaT blocks translation initiation by N-acetylation of the initiator tRNA(fMet). *Nat. Chem. Biol.* 13, 640–646. doi: 10.1038/nchembio.2346

Kandel, P. P., Baltrus, D. A., and Hockett, K. L. (2020). *Pseudomonas* can survive tailocin killing via persistence-like and heterogenous resistance mechanisms. *J. Bacteriol.* 202:13. doi: 10.1128/JB.00142-20

Kim, J. F., Charkowski, A. O., Alfano, J. R., Collmer, A., and Beer, S. V. (1998). Sequences related to transposable elements and bacteriophages flank avirulence genes of *Pseudomonas* syringae. *Mol. Plant Microbe Interact* 11, 1247–1252. doi: 10.1094/mpmi.1998.11.12.1247

Lam, H. N., Chakravarthy, S., Wei, H.-L., BuiNguyen, H., Stodghill, P. V., Collmer, A., et al. (2014). Global analysis of the HrpL regulon in the plant pathogen *Pseudomonas* syringae pv. tomato DC3000 reveals new regulon members with diverse functions. *PLoS One* 9:e106115. doi: 10.1371/journal.pone.0106115

Lee, M. D. (2019). GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35, 4162–4164.

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Lindow, S. E., and Brandl, M. T. (2003). Microbiology of the phyllosphere. *Appl. Environ. Microbiol.* 69, 1875–1883. doi: 10.1128/aem.69.4.1875-1883.2003

Liu, J., Yu, M., Chatnaparat, T., Lee, J. H., Tian, Y., Hu, B., et al. (2020). Comparative transcriptomic analysis of global gene expression mediated by (p) ppGpp reveals common regulatory networks in *Pseudomonas* syringae. *BMC Genomics* 21:296. doi: 10.1186/s12864-020-6701-2

Lovelace, A. H., Smith, A., and Kvitko, B. H. (2018). Pattern-triggered immunity alters the transcriptional regulation of virulence-associated genes and induces the sulfur starvation response in *Pseudomonas* syringae pv. tomato DC3000. *Mol. Plant Microbe Interact* 31, 750–765. doi: 10.1094/MPMI-01-18-0008-R

Ludanyi, M., Blanchard, L., Dulermo, R., Brandelet, G., Bellanger, L., Pignol, D., et al. (2014). Radiation response in Deinococcus deserti: irrE is a metalloprotease that cleaves repressor protein DdrO. *Mol. Microbiol.* 94, 434–449. doi: 10.1111/mmi.12774

Markel, E., Stodghill, P., Bao, Z., Myers, C. R., and Swingle, B. (2016). AlgU controls expression of virulence genes in *Pseudomonas* syringae pv. tomato DC3000. *J. Bacteriol* 198, 2330–2344. doi: 10.1128/JB.00276-16

Martins, P. M. M., Machado, M. A., Silva, N. V., Takita, M. A., and de Souza, A. A. (2016). Type II Toxin-Antitoxin Distribution and Adaptive Aspects on Xanthomonas Genomes: focus on Xanthomonas citri. *Front Microbiol.* 7:652. doi: 10.3389/fmicb.2016.00652

McCann, H. C., Rikkerink, E. H. A., Bertels, F., Fiers, M., Lu, A., Rees-George, J., et al. (2013). Genomic analysis of the kiwifruit pathogen *Pseudomonas* syringae pv. actinidiae provides insight into the origins of an emergent plant disease. *PLoS Pathog.* 9:e1003503. doi: 10.1371/journal.ppat.1003503

Merfa, M. V., Niza, B., Takita, M. A., and De Souza, A. A. (2016). The MqsRA toxin-antitoxin system from Xylella fastidiosa plays a key role in bacterial fitness, pathogenicity, and persister cell formation. *Front Microbiol.* 7:904. doi: 10.3389/fmicb.2016.00904

Mock, N. M., Baker, C. J., and Aver'yanov, A. A. (2015). Induction of a viable but not culturable (VBNC) state in some *Pseudomonas* syringae pathovars upon exposure to oxidation of an apoplastic phenolic, acetosyringone. *Physiol. Mol. Plant Pathol.* 89, 16–24.

Morris, C. E., Monteil, C. L., and Berge, O. (2013). The life history of *Pseudomonas* syringae: linking agriculture to earth system processes. *Annu. Rev. Phytopathol.* 51, 85–104. doi: 10.1146/annurev-phyto-082712-102402

Morris, C. E., Sands, D. C., Vinatzer, B. A., Glaux, C., Guilbaud, C., Buffière, A., et al. (2008). The life history of the plant pathogen *Pseudomonas* syringae is linked to the water cycle. *ISME J.* 2, 321–334. doi: 10.1038/ismej.2007.113

Moyed, H. S., and Bertrand, K. P. (1983). hipA, a newly recognized gene of *Escherichia coli* K-12 that affects frequency of persistence after inhibition of murein synthesis. *J. Bacteriol.* 155, 768–775. doi: 10.1128/jb.155.2.768-775.1983

Muthuramalingam, M., White, J. C., Murphy, T., Ames, J. R., and Bourne, C. R. (2019). The toxin from a ParDE toxin-antitoxin system found in *Pseudomonas aeruginosa* offers protection to cells challenged with anti-gyrase antibiotics. *Mol. Microbiol.* 111, 441–454. doi: 10.1111/mmi.14165

Mutschler, H., Gebhardt, M., Shoeman, R. L., and Meinhart, A. (2011). A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLoS Biol.* 9:e1001033. doi: 10.1371/journal.pbio.1001033

Newberry, E. A., Ebrahim, M., Timilsina, S., Zlatkovič, N., Obradović, A., Bull, C. T., et al. (2019). Inference of convergent gene acquisition among *Pseudomonas syringae* strains isolated from watermelon, cantaloupe, and squash. *Front Microbiol.* 10:270. doi: 10.3389/fmicb.2019.00270

Norton, J. P., and Mulvey, M. A. (2012). Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic *Escherichia coli*. *PLoS Pathog.* 8:e1002954. doi: 10.1371/journal.ppat.1002954

Ogura, T., and Hiraga, S. (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc. Natl. Acad. Sci. U. S. A.* 80, 4784–4788. doi: 10.1073/pnas.80.15.4784

Pandey, D. P., and Gerdes, K. (2005). Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.* 33, 966–976. doi: 10.1093/nar/gki201

Patel, R. R., Kandel, P. P., Traverso, E., Hockett, K. L., Triplett, L. R., and Vidaver, A. K. (2021). *Pseudomonas* syringae pv. phaseolicola uses distinct modes of stationary-phase persistence to survive bacteriocin and streptomycin treatments. *Mbio* 12, e00161–21. doi: 10.1128/mBio.00161-21

Peng, J., Triplett, L. R., and Sundin, G. W. (2021). Activation of metabolic and stress responses during subtoxic expression of the type I toxin hok in Erwinia amylovora. *BMC Genomics* 22:74. doi: 10.1186/s12864-021-07376-w

Piscotta, F. J., Jeffrey, P. D., and Link, A. J. (2019). ParST is a widespread toxin–antitoxin module that targets nucleotide metabolism. *Proc. Natl. Acad. Sci.* 116:826. doi: 10.1073/pnas.1814633116

Postnikova, O. A., Shao, J., Mock, N. M., Baker, C. J., and Nemchinov, L. G. (2015). Gene expression profiling in viable but nonculturable (VBNC) cells of *Pseudomonas* syringae pv. syringae. *Front. Microbiol.* 6:1419. doi: 10.3389/fmicb.2015.01419

Prysak, M. H., Mozdzierz, C. J., Cook, A. M., Zhu, L., Zhang, Y., Inouye, M., et al. (2009). Bacterial toxin YafQ is an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific and frame-dependent mRNA cleavage. *Mol. Microbiol.* 71, 1071–1087. doi: 10.1111/j.1365-2958.2008.06572.x

Roberts, R. C., Ström, A. R., and Helinski, D. R. (1994). The parDE operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss. *J. Mol. Biol.* 237, 35–51. doi: 10.1006/jmbi.1994.1207

Roussin, M., Rabarioelina, S., Cluzeau, L., Cayron, J., Lesterlin, C., Salcedo, S. P., et al. (2019). Identification of a contact-dependent growth inhibition (CDI) system that reduces biofilm formation and host cell adhesion of *Acinetobacter baumannii* DSM30011 strain. *Front. Microbiol.* 10:2450. doi: 10.3389/fmicb.2019.02450

Santamaría-Hernando, S., Cerna-Vargas, J. P., Martínez-García, P. M., de Francisco-de Polanco, S., Nebreda, S., Rodríguez-Palenzuela, P., et al. (2020). Blue-light perception by epiphytic *Pseudomonas* syringae drives chemoreceptor expression, enabling efficient plant infection. *Mol. Plant Pathol.* 21, 1606–1619. doi: 10.1111/mpp.13001

Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U., et al. (2013). Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. *Mol. Cell.* 50, 136–148. doi: 10.1016/j.molcel.2013.02.002

Schumacher, M. A., Balani, P., Min, J., Chinnam, N. B., Hansen, S., Vulić, M., et al. (2015). HipBA-promoter structures reveal the basis of heritable multidrug tolerance. *Nature* 524, 59–64. doi: 10.1038/nature14662

Shao, X., Tan, M., Xie, Y., Yao, C., Wang, T., Huang, H., et al. (2021). Integrated regulatory network in *Pseudomonas* syringae reveals dynamics of virulence. *Cell Rep.* 34:108920. doi: 10.1016/j.celrep.2021.108920

Shavit, R., Lebendiker, M., Pasternak, Z., Burdman, S., and Helman, Y. (2016). The vapb–vapc operon of Acidovorax citrulli functions as a bona-fide toxin–antitoxin module. *Front Microbiol.* 6:1499. doi: 10.3389/fmicb.2015.01499

Shidore, T., Zeng, Q., and Triplett, L. R. (2019). survey of toxin–antitoxin systems in Erwinia amylovora reveals insights into diversity and functional specificity. *Toxins* 11:206. doi: 10.3390/toxins11040206

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Skjerning, R. B., Senissar, M., Winther, K. S., Gerdes, K., and Brodersen, D. E. (2019). The RES domain toxins of RES-Xre toxin-antitoxin modules induce cell stasis by degrading NAD+. *Mol. Microbiol.* 111, 221–236. doi: 10.1111/mmi.14150

Song, S., and Wood, T. K. (2020b). Toxin/antitoxin system paradigms: toxins bound to antitoxins are not likely activated by preferential antitoxin degradation. *Adv. Biosyst.* 4:1900290. doi: 10.1002/adbi.201900290

Song, S., and Wood, T. K. (2020a). A primary physiological role of toxin/antitoxin systems is phage inhibition. *Front Microbiol.* 11:1895. doi: 10.3389/fmicb.2020.0189

Tiwari, P., Arora, G., Singh, M., Kidwai, S., Narayan, O. P., and Singh, R. (2015). MazF ribonucleases promote Mycobacterium tuberculosis drug tolerance and virulence in guinea pigs. *Nat. Commun.* 6:6059. doi: 10.1038/ncomms7059

Tourasse, N. J., and Darfeuille, F. (2021). T1TAdb: the database of type I toxin-antitoxin systems. *Rna* 27, 1471–1481. doi: 10.1261/rna.078802.121

Triplett, L. R., Shidore, T., Long, J., Jiamin, M., Wu, S., Han, Q., et al. (2016). AvrRxo1 is a bifunctional type III secreted effector and toxin-antitoxin system component with homologs in diverse environmental contexts. *PLoS One* 11:e0158856. doi: 10.1371/journal.pone.0158856

Unterholzner, S. J., Poppenberger, B., and Rozhon, W. (2013b). Toxin-antitoxin systems: biology, identification, and application. *Mob. Genet. Elements* 3:e26219. doi: 10.4161/mge.26219

Unterholzner, S. J., Hailer, B., Poppenberger, B., and Rozhon, W. (2013a). Characterisation of the stbD/E toxin-antitoxin system of pEP36, a plasmid of the plant pathogen Erwinia pyrifoliae. *Plasmid* 70, 216–225. doi: 10.1016/j.plasmid.2013.04.002

Winther, K. S., and Gerdes, K. (2011). Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7403–7407. doi: 10.1073/pnas.1019587108

Wood, T. L., and Wood, T. K. (2016). The HigB/HigA toxin/antitoxin system of *Pseudomonas aeruginosa* influences the virulence factors pyochelin, pyocyanin, and biofilm formation. *Microbiol. Open* 5, 499–511. doi: 10.1002/mbo3.346

Wozniak, R. A. F., and Waldor, M. K. (2009). A toxin–antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet.* 5:e1000439. doi: 10.1371/journal.pgen.1000439

Xie, Y., Wei, Y., Shen, Y., Li, X., Zhou, H., Tai, C., Deng, Z. et al. (2018). TADB 2.0: an updated database of bacterial type II toxin-antitoxin loci. *Nucleic Acids Res.* 46, D749–D753. doi: 10.1093/nar/gkx1033

Xin, X.-F., Kvitko, B., and He, S. Y. (2018). *Pseudomonas* syringae: what it takes to be a pathogen. *Nat. Rev. Microbiol.* 16, 316–328. doi: 10.1038/nrmicro.2018.17

Yamaguchi, Y., Park, J. H., and Inouye, M. (2009). MqsR, a crucial regulator for quorum sensing and biofilm formation, is a GCU-specific mRNA interferase in *Escherichia coli*. *J. Biol. Chem.* 284, 28746–28753. doi: 10.1074/jbc.M109.032904

Yao, X., Chen, T., Shen, X., Zhao, Y., Wang, M., Rao, X., et al. (2015). The chromosomal SezAT toxin–antitoxin system promotes the maintenance of the SsPI-1 pathogenicity island in epidemic Streptococcus suis. *Mol. Microbiol.* 98, 243–257. doi: 10.1111/mmi.13116

Yu, X., Lund, S. P., Scott, R. A., Greenwald, J. W., Records, A. H., Nettleton, D., et al. (2013). Transcriptional responses of *Pseudomonas* syringae to growth in epiphytic versus apoplastic leaf sites. *Proc. Natl. Acad. Sci. U. S. A.* 110, E425–E434. doi: 10.1073/pnas.1221892110

Yuan, J., Yamaichi, Y., and Waldor, M. K. (2011). The three *Vibrio cholerae* chromosome II-encoded ParE toxins degrade chromosome I following loss of chromosome II. *J. Bacteriol.* 193, 611–619. doi: 10.1128/JB.01185-10

Zhang, Y., and Inouye, M. (2009). The inhibitory mechanism of protein synthesis by YoeB, an *Escherichia coli* toxin. *J. Biol. Chem.* 284, 6627–6638. doi: 10.1074/jbc.M808779200

Zhang, Y., and Inouye, M. (2011). RatA (YfjG), an *Escherichia coli* toxin, inhibits 70S ribosome association to block translation initiation. *Mol. Microbiol.* 79, 1418–1429. doi: 10.1111/j.1365-2958.2010.07506.x

Zhang, Y., Yamaguchi, Y., and Inouye, M. (2009). Characterization of YafO, an *Escherichia coli* toxin. *J. Biol. Chem.* 284, 25522–25531. doi: 10.1074/jbc.m109.036624

Zhang, Y., Zhang, J., Hoeflich, K. P., Ikura, M., Qing, G., and Inouye, M. (2003). MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*. *Mol. Cell* 12, 913–923. doi: 10.1016/s1097-2765(03)00402-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 January 2022
doi: 10.3389/fmicb.2022.778927

Genome-Wide Identification of Genes Important for Growth of *Dickeya dadantii* and *Dickeya dianthicola* in Potato (*Solanum tuberosum*) Tubers

Tyler C. Helmann[1†], Melanie J. Filiatrault[1,2†] and Paul V. Stodghill[1,2†]*

[1] Emerging Pests and Pathogens Research Unit, Agricultural Research Service, United States Department of Agriculture, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, United States, [2] Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:
Jeffrey Jones,
University of Florida, United States

Reviewed by:
Robert Czajkowski,
University of Gdańsk, Poland
Daniel Bellieny Rabelo,
University College Dublin, Ireland

***Correspondence:**
Paul V. Stodghill
paul.stodghill@usda.gov

†ORCID:
Tyler C. Helmann
orcid.org/0000-0002-8431-6461
Melanie J. Filiatrault
orcid.org/0000-0001-7704-9097
Paul V. Stodghill
orcid.org/0000-0003-3875-8450

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: 17 September 2021
Accepted: 10 January 2022
Published: 25 January 2022

Citation:
Helmann TC, Filiatrault MJ and Stodghill PV (2022) Genome-Wide Identification of Genes Important for Growth of Dickeya dadantii and Dickeya dianthicola in Potato (Solanum tuberosum) Tubers. Front. Microbiol. 13:778927. doi: 10.3389/fmicb.2022.778927

Dickeya species are causal agents of soft rot diseases in many economically important crops, including soft rot disease of potato (*Solanum tuberosum*). Using random barcode transposon-site sequencing (RB-TnSeq), we generated genome-wide mutant fitness profiles of *Dickeya dadantii* 3937, *Dickeya dianthicola* ME23, and *Dickeya dianthicola* 67-19 isolates collected after passage through several *in vitro* and *in vivo* conditions. Though all three strains are pathogenic on potato, *D. dadantii* 3937 is a well-characterized model while *D. dianthicola* strains ME23 and 67-19 are recent isolates. Strain ME23 specifically was identified as a representative strain from a 2014 outbreak on potato. This study generated comparable gene fitness measurements across ecologically relevant conditions for both model and non-model strains. Tubers from the potato cultivars "Atlantic," "Dark Red Norland," and "Upstate Abundance" provided highly similar conditions for bacterial growth. Using the homolog detection software PyParanoid, we matched fitness values for orthologous genes in the three bacterial strains. Direct comparison of fitness among the strains highlighted shared and variable traits important for growth. Bacterial growth in minimal medium required many metabolic traits that were also essential for competitive growth *in planta*, such as amino acid, carbohydrate, and nucleotide biosynthesis. Growth in tubers specifically required the pectin degradation gene *kduD*. Disruption in three putative DNA-binding proteins had strain-specific effects on competitive fitness in tubers. Though the Soft Rot *Pectobacteriaceae* can cause disease with little host specificity, it remains to be seen the extent to which strain-level variation impacts virulence.

Keywords: potato, soft rot, RB-TnSeq, TnSeq, *Dickeya dadantii*, *Dickeya dianthicola*

INTRODUCTION

The Soft Rot *Pectobacteriaceae* comprise *Dickeya* and *Pectobacterium* species that are the causal agents of bacterial soft rot diseases on economically important vegetables and ornamentals (Adeolu et al., 2016; Motyka et al., 2017). These necrotrophic pathogens rely on numerous traits to survive the various stress conditions encountered in the host, including oxidative stress, osmotic stress, iron starvation, and toxic compounds (Jiang et al., 2016; Reverchon et al., 2016). During initial

plant colonization, cells encounter a slightly acidic apoplast (pH 5.0–6.5) or tuber (pH 5.5–6.2), and pH levels increase to slightly basic (pH 8) by the late stage of infection (Grignon and Sentenac, 1991; Nachin and Barras, 2000; Kiszonas and Bamberg, 2010). At high cell densities, the production of plant cell wall degrading enzymes is induced, leading to the maceration of plant tissues and the formation of visible symptoms (Reverchon et al., 2016). In response to these changing conditions, cells must respond to environmental cues to adapt global expression of stress response genes and virulence factors (Jiang et al., 2016; Raoul des Essarts et al., 2019).

The taxonomy of the *Dickeya* genus has undergone substantial revision with the addition of novel species *D. solani*, *D. aquatica*, and *D. fangzhongdai* (Samson et al., 2005; Parkinson et al., 2014; Van Der Wolf et al., 2014; Tian et al., 2016). However, an increase in available whole-genome sequence data has improved species-level identification based on pairwise average nucleotide identity (ANI), *in silico* DNA-DNA hybridization (*is*DDH), and core genome multilocus sequence analysis (MLSA) (Zhang et al., 2016). There is little known about host-specific traits, as these species generally have broad host ranges (Van Gijsegem et al., 2021). In addition, there are no known resistance genes for potato soft rot, and it is therefore impossible to predict cultivar resistance without testing (Lyon, 1989; Czajkowski et al., 2011; Chung et al., 2013). Without gene-for-gene resistance, potato cultivar tolerance is reliant on physical barriers and antimicrobial small molecules such as phenolics or the phytoalexin rishitin (Lyon, 1989). An alternative strategy being explored is the use of bacteriophage-based biocontrol for potato plants and tubers, particularly of the highly virulent *D. solani* (Adriaenssens et al., 2012; Czajkowski et al., 2017). *Dickeya* virulence factors and transcriptional regulators of virulence genes are generally conserved. Studies in *D. solani* have suggested a closed pangenome with many conserved virulence factors and transcriptional regulators (Golanowska et al., 2018; Motyka-Pomagruk et al., 2020). However, virulence regulon differences indicate some virulence genes could have differential expression among strains (Golanowska et al., 2018). Pangenomic analysis of *D. dianthicola* also reflects a closed pangenome, though almost all sequenced strains were originally isolated from potato (Ge et al., 2021c).

Transposon mutagenesis followed by high-throughput sequencing (TnSeq) is a valuable screening tool to identify genes important for growth in a given condition. Gene fitness is functionally defined as the change in relative mutant abundance within a population and is a quantitative measure of growth rate (van Opijnen et al., 2009). TnSeq has been used to identify *D. dadantii* genes important for growth in chicory (Royet et al., 2019). This work identified several metabolic pathways essential for *in planta* growth, primarily those involved in biosynthesis of nucleotides, amino acids, and some vitamins (Royet et al., 2019). A modification of TnSeq to add 20-nucleotide DNA "barcodes" to transposon donor plasmids, known as random barcode transposon-site sequencing (RB-TnSeq) enables highly scalable TnSeq assays (Wetmore et al., 2015). This method has been applied to over 44 bacterial strains to date (Price et al., 2018), including plant pathogenic *Pseudomonas* spp. and *Ralstonia* spp. (Cole et al., 2017; Helmann et al., 2019; Georgoulis et al., 2021). A motivating factor for this study was to demonstrate the value of RB-TnSeq to characterize necrotrophic plant pathogenic bacteria *in planta*; using both an established model strain as well as recently isolated and uncharacterized strains (Liu and Deutschbauer, 2018).

To identify bacterial traits important for growth in potato (*Solanum tuberosum*) tubers, we examined three strains across two species: *D. dadantii* 3937 (*Dda*3937), *D. dianthicola* ME23 (*Ddia*ME23), and *D. dianthicola* 67-19 (*Ddia*6719). While these three strains are all pathogenic on potato, *Dda*3937 was originally isolated from *Saintpaulia ionantha* (Lemattre and Narcy, 1972), and *Ddia*6719 was recently isolated from New Guinea impatiens (*Impatiens hawkeri*) (Liu et al., 2020a,b). *Dda*3937 has been a model strain used for molecular studies since its isolation in 1972 (Lemattre and Narcy, 1972), while *Ddia*ME23 was isolated as a representative strain for a 2014 potato disease outbreak (Ma et al., 2019). Pairwise ANI between *Dda*3937 and *Ddia*ME23 is 92.8% (Chen et al., 2019). Using barcoded transposon insertion mutant libraries in these three strains, we screened for genes that contributed to competitive fitness during growth *in vitro* and in potato tubers. By leveraging RB-TnSeq in a shared susceptible host for *D. dadantii* and *D. dianthicola*, we aimed to identify common and unique virulence factors among representative strains for these two species.

MATERIALS AND METHODS

PyParanoid Gene Ortholog Group Assignments

Gene ortholog groups were generated using the PyParanoid analysis pipeline v0.4.1 (Melnyk et al., 2019). Peptide sequences from the following RefSeq genome assemblies were used to construct ortholog groups: GCF_000147055.1 (*Dda*3937), GCF_003403135.1 (*Ddia*ME23), and GCF_014893095.1 (*Ddia*6719). From these assemblies, RefSeq gene loci were then matched to their corresponding protein names to allow comparison to the Barcode Sequencing (BarSeq) fitness data. Additionally, Clusters of Orthologous Groups (COG) categories for *Dda*3937 were downloaded from the IMG database (Chen et al., 2019), GenBank gene names were replaced with their corresponding RefSeq names, and added to this ortholog table, allowing putative COG assignments for orthologous genes in *D. dianthicola* strains.

Barcoded Transposon Library Construction

Strains used in this study are described in **Supplementary Table 1**. All bacteria were cultured in Luria-Bertani (LB) medium (10 g tryptone, 5 g yeast extract, and 10 g NaCl per 1 L) (Bertani, 1951) at 28°C, except for pure culture *E. coli* grown at 37°C. When noted, kanamycin was used at a final concentration of 50 μg/ml. Barcoded transposon libraries were constructed by conjugating the barcoded *mariner* transposon plasmid pKMW3 from the *E. coli* WM3064 donor library APA752

(Wetmore et al., 2015) into wild-type *Dda*3937, *Dia*ME23, and *Ddia*6719. Recipient *Dickeya* strains were grown as 3 ml LB liquid cultures overnight and 1.5 ml of each culture was subcultured into 30 ml LB without antibiotics. An entire 1 ml freezer aliquot of the *E. coli* donor library was thawed and used to inoculate 25 ml LB containing 300 μM diaminopimelic acid (DAP) (Sigma-Aldrich, United States) and kanamycin. All cultures were grown to OD_{600} of approximately 1.0, washed with 10 mM KPO_4, and combined in equal amounts before plating donor-recipient pairs each on 50 LB plates containing 300 μM DAP. Conjugations were incubated at 28°C overnight, and exconjugants were then scraped into 40 ml 10 mM KPO_4. This conjugation mixture was then vortexed, spread onto 200 LB kanamycin plates per strain, and incubated at 28°C for 3 days. All colonies were resuspended in 200 ml LB with kanamycin, diluted back to a starting OD_{600} of 0.2, and grown at 28°C with shaking for 6–8 h, to a concentration of approximately 10^9 CFU/ml (measured as a 1/10 dilution at $OD_{600} = 0.15–0.3$). Glycerol was added to the library to a final concentration of 15%, and 1 ml aliquots were frozen at –80°C.

DNA Library Preparation and Sequencing

For DNA library preparation, genomic DNA from each library was purified from an entire 1 ml cell pellet using the Monarch Genomic DNA Purification Kit (New England Biolabs, United States). Samples were eluted in 50 μl nuclease-free water. Purified DNA was quantified on a NanoDrop One (Thermo Fischer Scientific, United States), and 500 ng DNA was used as input for the NEBNext Ultra II FS DNA Library Prep kit (New England Biolabs, United States), following the manufacturer's instructions with modifications as follows. For enzymatic DNA fragmentation, a 12-min incubation time was used. DNA fragments were size selected using AMPure XP magnetic beads (Beckman Coulter, United States) at the recommended ratios 0.4X and 0.2X. We used a modified version of the protocol described in Wetmore et al. (2015), with a two-step PCR used to enrich for transposon insertion sites, based on (Rubin et al., 2021). A custom splinkerette adapter was ligated to fragmented DNA, prepared by annealing oligos:/5Phos/G*ATCGGAAGAGCACACGTCTGGGTTTTTT TTTTCAAAAAAA*A and G*AGATCGGTCTCGGCATTCCC AGACGTGTGCTCTTCCGATC*T (Rubin et al., 2021). Between rounds of PCR and before submitting for sequencing, DNA was cleaned by binding to AMPure XP magnetic beads, using a bead ratio of 0.9X and eluted in 15 μl 0.1X TE buffer for intermediate steps and 30 μl 0.1X TE for sequencing. Finally, the sequencing library was quantified using a Qubit dsDNA HS assay kit (Thermo Fischer Scientific, United States). DNA libraries were submitted for sequencing at the Biotechnology Resource Center (BRC) Genomics Facility at the Cornell Institute of Biotechnology on an Illumina MiSeq to check library quality, followed by sequencing on a NextSeq 500 (Illumina, Inc., United States). All mapping used single-end sequencing for 150 bp fragments.

Transposon Library Mapping

Sequence data were analyzed using the scripts from the FEBA package v1.3.1 (Wetmore et al., 2015). MapTnSeq.pl was used to identify the barcode and location in the genome of each read with identifiable transposon sequence from both MiSeq and NextSeq reads, based on the "model_pKMW3.2" sequence. DesignRandomPool.pl was used to assemble the mutant pool using barcodes seen in a single location 10 or more times. All TnSeq mapping and BarSeq fitness calculation code is available at http://bitbucket.org/berkeleylab/feba/ (Wetmore et al., 2015). Mapping scripts were run on a Cornell University BioHPC Cloud 40-core Linux (CentOS 7.6). server with 256 GB RAM.

Gene Essentiality Predictions

Using the output from MapTnSeq.pl, gene essentiality predictions were made using https://bitbucket.org/berkeleylab/feba/src/master/bin/Essentiality.pl and the function "Essentials" from https://bitbucket.org/berkeleylab/feba/src/master/lib/comb.R (Wetmore et al., 2015). Using the median insertion density and the median length of genes > 100 bp, this method calculates how short a gene can be and still be unlikely to have no insertions by chance ($P < 0.02$, Poisson distribution); genes shorter than this threshold are then excluded (Price et al., 2018). For the *Dickeya* strains examined here, the minimum gene length for a gene to be predicted as essential for growth in LB was 175 bp (*Dda*3937) or 150 bp (*Ddia*ME23 and *Ddia*6719). Protein-coding genes are then considered to be essential or nearly essential if there are no fitness values and the normalized central insertion density score and normalized read density score as computed by the FEBA package were < 0.2 (Price et al., 2018).

Library Pre-culture

For a given BarSeq experiment, a single transposon library freezer aliquot was thawed and recovered in 25 ml LB containing kanamycin at 28°C until $OD_{600} = 0.5–0.7$, approximately 6–8 h. At this point, two 1 ml cell pellets were frozen as time0 measurements, and the remaining culture was washed in 10 mM KPO_4 and used to inoculate experimental samples.

In vitro Samples

All *in vitro* cultures were grown in 1 ml volumes in 24-well plates. In each well, 50 μl starter culture at 0.3 OD_{600} was added to 950 μl medium containing kanamycin. Media tested were LB, Potato Dextrose Broth (PDB)(pH 5) (Sigma-Aldrich, United States), and M9 minimal medium (M9) as described in M9 Minimal Medium (Standard) (2010) but containing 0.4% glycerol instead of 0.4% glucose. Plates were incubated at 28°C with shaking at 200 rpm. After 1 day (LB and PDB) or 2 days (M9), each 1 ml sample was pelleted and frozen prior to genomic DNA extraction.

Tuber Samples

Prior to inoculation, all tubers were rinsed and then surface sterilized by submerging in 70% ethanol for 10 min, followed by two washes with distilled water. Inoculum was standardized to 10^9 CFU/ml by measuring a 1/10 culture dilution at $OD_{600} = 0.3$ (corresponding to 10^8 CFU/ml), and 10 μl was inoculated in two replicate stab wounds created by pushing a 200 μl pipet tip roughly 3 mm into each tuber. Six replicate tubers were used for each bacterial strain and potato cultivar. Inoculated tubers were

stored in plastic bags at 28°C. Two days after inoculation, ∼2 cm length cores were taken at each site of inoculation using a 1 cm diameter cork borer. Duplicate cores from each tuber were pooled in 8 ml 10 mM KPO_4 and shaken at 200 rpm at 28°C for 10 min. For each sample, 2 ml of bacterial suspension were pelleted and frozen prior to DNA extraction.

BarSeq PCR and Sequencing

Genomic DNA was purified from cell pellets using the Monarch Genomic DNA Purification Kit (New England Biolabs, United States). Purified DNA samples were eluted in 30 μl nuclease-free water and quantified on a Nanodrop One (Thermo Fischer Scientific, United States). After gDNA extraction, the 98°C BarSeq PCR as described in Wetmore et al. (2015) was used to specifically amplify the barcode region of each sample. The PCR for each sample was performed in 50 μl total volume: containing 0.5 μl Q5 High-Fidelity DNA polymerase (New England Biolabs, United States), 10 μl 5X Q5 buffer, 10 μl 5X GC enhancer, 1 μl 10 mM dNTPs, 150–200 ng template gDNA, 2.5 μl common reverse primer (BarSeq_P1), and 2.5 μl of forward primer from one of the 96 indexed forward primers (BarSeq_P2_ITXXX), both at 10 μM (Wetmore et al., 2015). Following the BarSeq PCR, 10 μl of each reaction was pooled (46–49 samples per pool), and 200 μl of this DNA pool was subsampled and purified using the DNA Clean and Concentrator Kit (Zymo Research, United States). The final DNA sequencing library was eluted in 30 μl nuclease-free water, quantified on a Nanodrop One, and submitted for sequencing at the BRC Genomics Facility at the Cornell Institute of Biotechnology. Prior to sequencing, the quality of each amplicon pool was assessed using a Bioanalyzer. Each sequencing pool was run on a single NextSeq 500 (Illumina, Inc., United States) lane for 75 bp single-end reads.

Gene Fitness Calculations

Sequencing reads were used to calculate genome-wide gene fitness using the FEBA scripts MultiCodes.pl, combineBarSeq.pl, and BarSeqR.pl (Wetmore et al., 2015). Scripts to calculate gene fitness values were run on a Cornell University BioHPC Cloud 40-core Linux (CentOS 7.6). Server with 256 GB RAM. From the raw fastq reads for each sample, individual barcode sequences were identified and counted using MultiCodes.pl, and these counts were combined across samples for a given transposon library using combineBarSeq.pl. Using BarSeqR.pl, fitness values for each insertion strain were calculated as the $\log_2$ ratio of barcode abundance following library growth under a given experimental condition divided by the abundance in the time0 sample. The fitness of each gene is the weighted average of strain fitness values based on the "central" transposon insertions only, i.e., those within the central 10–90% of a gene. Barcode counts were summed between replicate time0 samples. For analysis, genes were required to have at least 30 reads per gene in the time0 sample, and 3 reads per individual strain (Wetmore et al., 2015). Gene fitness values were normalized across the chromosome so that the median gene fitness value was 0. All experiments described here passed previously described quality control metrics (Wetmore et al., 2015; Price et al., 2018).

Fitness Analysis and Plotting

We focused on genes having fitness values > 1 or < -1 and absolute t-like test statistic > 4. This t score is an estimate of the reliability of the fitness measurement for a gene, and is equal to the fitness value divided by the square root of the maximum variance calculated in two ways (Wetmore et al., 2015). With these cutoffs, we also calculated gene fitness values comparing replicate time0 samples (Price et al., 2018; Liu et al., 2021). Across 6 (each *Dda*3937 and *Ddia*ME23) and 2 (*Ddia*6719) replicate time0 samples, 0 gene fitness values had fitness > 1 or < -1 and absolute $t > 4$. Data were analyzed in R v4.0.3 (R Core Team, 2017) and visualized using the package ggplot2 v3.3.5 (Wickham, 2016). The principal components analysis was performed on the gene fitness matrix for each *Dickeya* strain using the R function prcomp, which performs centered singular value decomposition.

Data Availability Statement

All raw Illumina reads used for mapping and fitness assays have been deposited in the Sequence Read Archive under BioProject accession number PRJNA692477. Individual sample accession numbers are listed in **Supplementary Table 2**. Annotated scripts used for computational analysis are available at http://github.com/tylerhelmann/dickeya-barseq-2021. Experimental fitness values are publicly available on the interactive Fitness Browser at http://fit.genomics.lbl.gov.

RESULTS

Identification of Homologous Gene Families in *Dickeya dadantii* 3937, *Dickeya dianthicola* ME23 and *Dickeya dianthicola* 67-19

To enable direct comparison of gene fitness measurements between strains, we constructed a database of homologous gene families using the PyParanoid analysis pipeline (Melnyk et al., 2019). Based on clustering of all predicted protein sequences from *Dda*3937, *Ddia*ME23, and *Ddia*6719, 3,821 total homolog groups were identified, representing 88.1% of the total input sequences. Of these, 3,310 groups contained single-copy genes in all three strains. For each group, gene loci, protein identifiers, and gene descriptions are listed in **Supplementary Table 3**. This table also contains COG assignments matched from the *Dda*3937 IMG genomic annotation (Chen et al., 2019).

Creation of Barcoded Transposon Libraries in *Dickeya* spp.

To measure contributions of individual genes to fitness, we constructed barcoded transposon mutant libraries in the *Dickeya* strains using a barcoded *mariner E. coli* donor library (Wetmore et al., 2015). These libraries ranged in size from 334,893 to 541,278 mapped genomic insertional strains, with 37–62 median strains per hit gene (**Table 1**). Of the three strains tested, only one gene in the *Ddia*6719 genome did not contain any TA dinucleotide sites and was therefore inaccessible to the *mariner*

transposon. Mapped insertions were evenly distributed across the chromosome of each strain (**Supplementary Figure 1**).

Identification of Essential Gene Sets in *Dickeya dadantii* and *Dickeya dianthicola*

Based on analysis of the TnSeq mapping data, essential genes were predicted using the FEBA RB-TnSeq analysis pipeline (Wetmore et al., 2015; Price et al., 2018). We identified 374–426 genes per strain that are likely to encode essential or near-essential genes for growth in LB (**Supplementary Table 4**). Using the ortholog group assignments for these genes, 316 of these predicted essential genes (74–84%) are shared among all three strains (**Supplementary Figure 2**). Most predicted essential genes are in the functional categories of "Translation, ribosomal structure, and biogenesis," "Cell wall/membrane/envelope

TABLE 1 | Characteristics of the *mariner* transposon libraries.

Library	Insertions in genome	Central insertion strains	Genes with central insertions (Total)	Median strains per hit gene
D. dadantii 3937	337,541	193,696	3,882 (4,213)	37
D. dianthicola ME23	541,278	321,087	3,805 (4,182)	62
D. dianthicola 67-19	334,893	200,170	3,728 (4,110)	41

"Central" insertions are those within the central 10–90% of a gene.

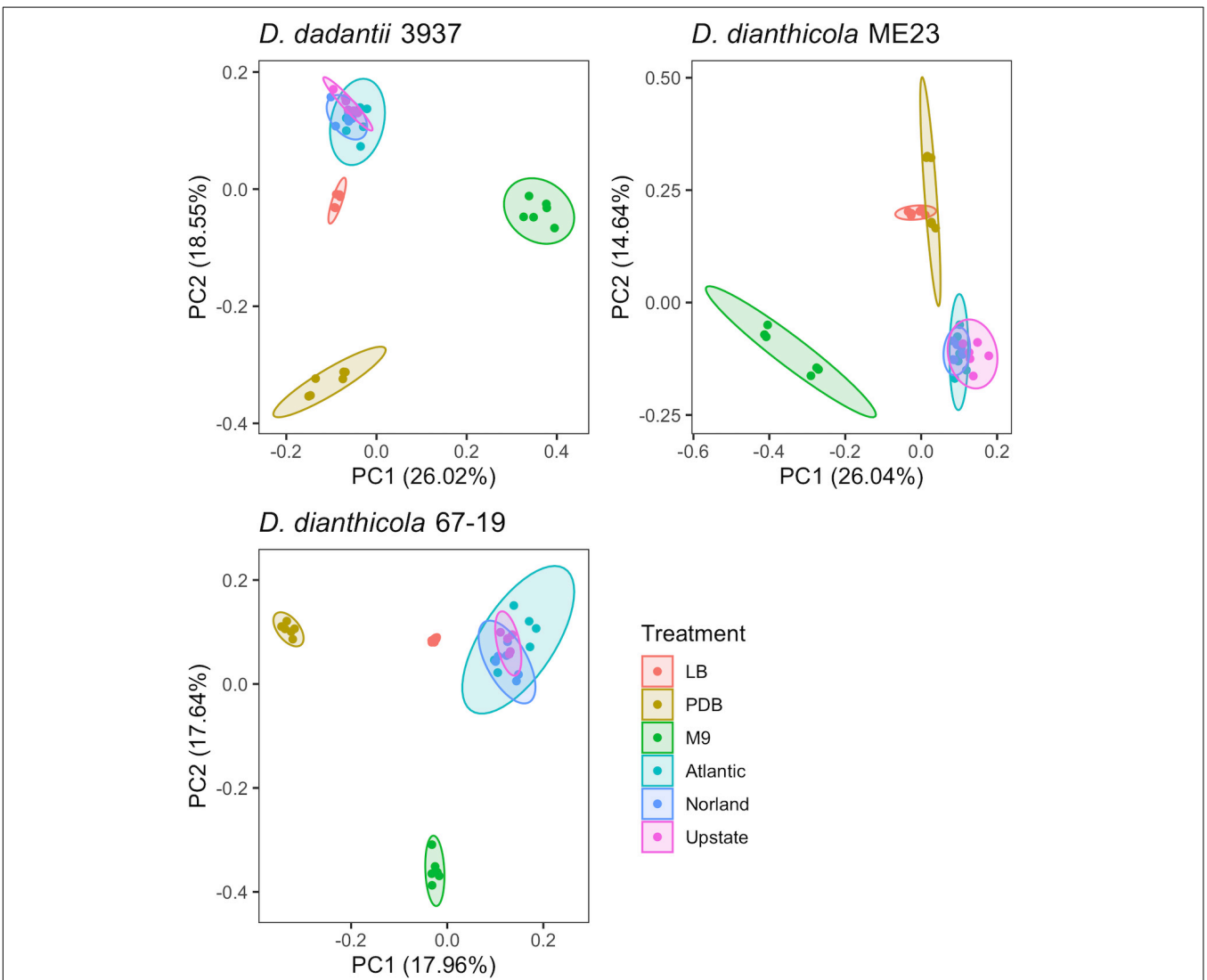


FIGURE 1 | PCA of experimental samples based on fitness values calculated for *D. dadantii* 3937, *D. dianthicola* ME23, and *D. dianthicola* 67–19. Available fitness values for each sample, respectively: *N* = 3,705; 3,761; 3,528. Superimposed ellipses are based on a multivariate t-distribution.

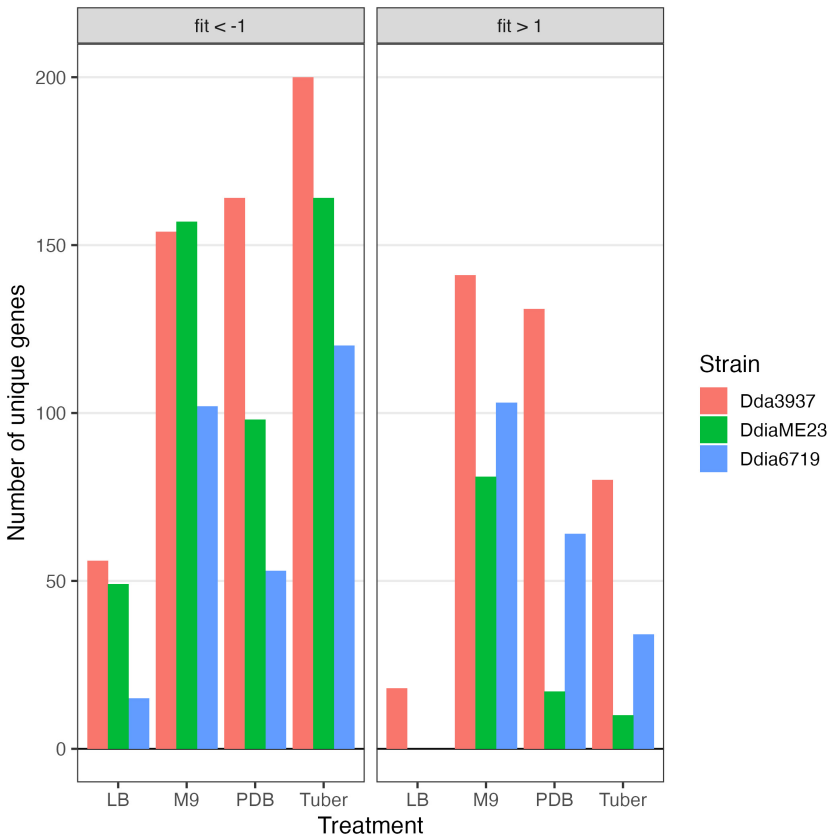


FIGURE 2 | Number of unique genes for each condition with fitness values of < -1 or > 1, and absolute t-like test statistic > 4 in at least one replicate sample.

biogenesis," "Coenzyme transport and metabolism," "Energy production and conversions," and "Replication, recombination, and repair" (**Supplementary Table 5**).

Conducting Pooled Growth Assays to Measure Relative Mutant Fitness

To generate genome-wide gene fitness values for the barcoded transposon libraries, each strain was grown in the rich media LB and Potato Dextrose Broth (PDB) as well as M9 minimal medium supplemented with 0.4% glycerol (**Supplementary Figure 3**). Strain fitness values were calculated as a log2 ratio of barcode abundance following sample growth with barcode abundance measured in the time0 duplicate samples. Gene fitness is the weighted average of individual strain fitness values (Wetmore et al., 2015). For fitness calculations, insertions in the first and last 10% of coding regions were excluded, with insertions in the remaining 80% of the gene considered "central." While 91–92% of genes in all strains contained centrally mapped insertions, not all genes were used in fitness calculations due to low read or insertion abundance. We focused our analysis on genes with fitness values > 1 or < -1, and absolute t-score > 4 (**Supplementary Table 6**). Across all conditions, we calculated fitness values for 3,705 (*Dda*3937), 3,761 (*Ddia*ME23), and 3,528 (*Ddia*6719) genes, representing 88, 90, and 86% of the total genes in each strain, respectively.

Principal component analysis showed gene fitness values of the three tuber conditions overlapped (**Figure 1**), and so these samples were jointly considered as a single "Tuber" condition for some subsequent analyses.

Disruption Mutants With Fitness Defects in Rich Media

As the libraries were constructed on LB medium, relatively few mutations deleterious in LB were maintained in the populations (**Figure 2**). **Figure 3** presents these data split by COG category. Limited mutations in genes categorized as "cell wall/membrane/envelope biogenesis" (*mdoGH*) and "cell cycle control, cell division, chromosome partitioning" (*ftsX*) were present in the mapped populations but generally detrimental in LB for all three strains (**Supplementary Figure 4**). Even in LB, some variation was apparent between strains, such as disruptions in the gene encoding the cell division protein ZapB which decreased competitive fitness in *Dda*3937 but not *Ddia*ME23 or *Ddia*6719 (**Supplementary Figure 4**).

The rich medium PDB provided a very different gene fitness profile than LB, likely due to nutritional differences and its slightly acidic nature (pH 5). For all three strains, more genes were detrimental (fit > 1) and beneficial (fit < -1) in PDB relative to LB (**Figure 2**). Genes in diverse metabolic categories contributed to competitive fitness, including "amino

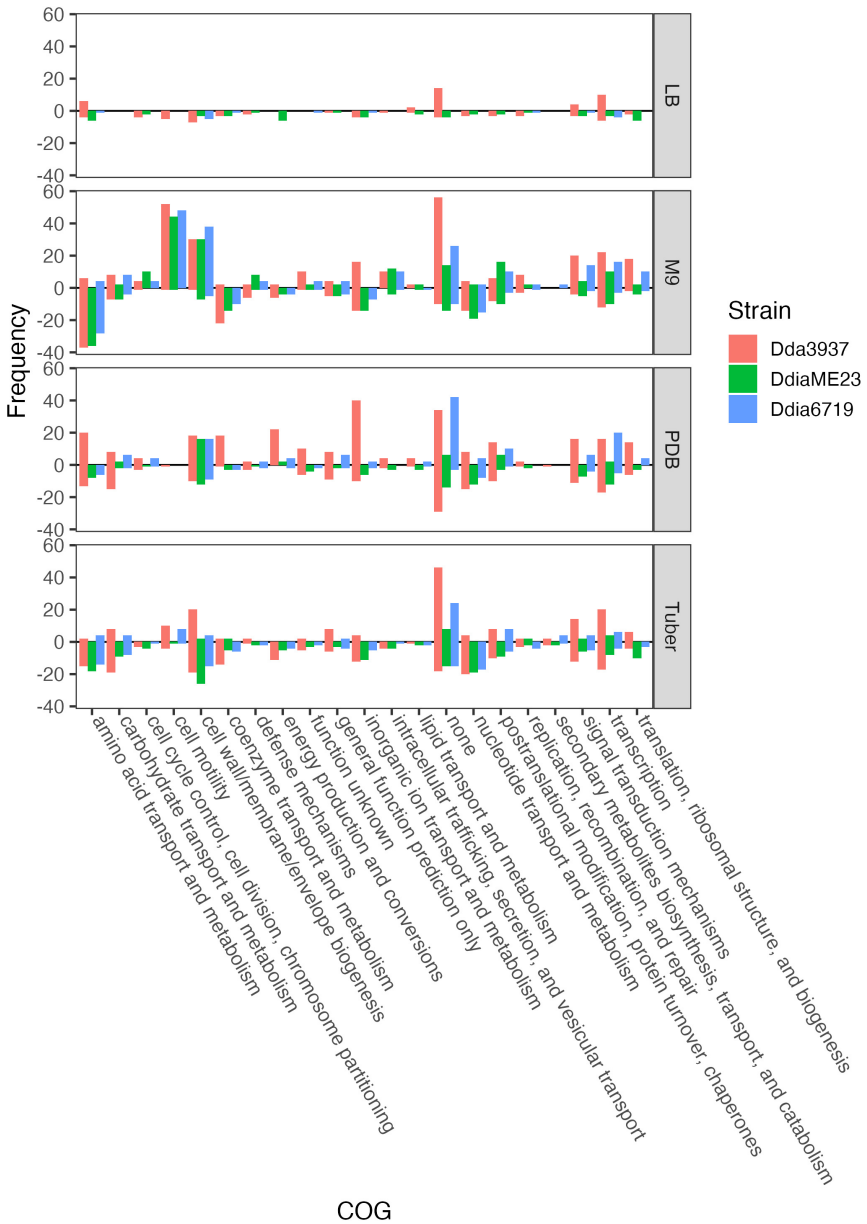


FIGURE 3 | Number of unique genes for each condition with fitness values of < -1 or > 1, and absolute t-like test statistic > 4 in at least one replicate sample. Genes where fit < -1 are show below the line $y = 0$, while genes where fit > 1 are shown above. COG assignments are based on the *D. dadantii* 3937 annotation in the IMG database (Chen et al., 2019), and extrapolated to *D. dianthicola* ME23 and *D. dianthicola* 67-19 based on PyParanoid-generated ortholog groups.

acid transport and metabolism," "carbohydrate transport and metabolism," "cell wall/membrane/envelope biogenesis," "coenzyme transport and metabolism," "inorganic ion transport and metabolism," "nucleotide transport and metabolism," "signal transduction mechanisms," "transcription," and "translation, ribosomal structure, and biogenesis" (**Figure 3**). For example, in all three strains oligopeptidase A and the low affinity potassium transporter Kup were specifically important in PDB for growth (**Supplementary Figure 5**). Disruptions in the two-component system RtsAB were specifically beneficial for *Dda*3937 in PDB, as were disruptions in the zinc uptake transcriptional repressor Zur (**Supplementary Figure 5**). Though LB and PDB are both complex rich media, pH and/or specific available nutrients differed enough to clearly separate the gene fitness profiles for *Dda*3937 and *Ddia*6719, though not for *Ddia*ME23 (**Figure 1**).

Disruption Mutants With Fitness Defects in Minimal Medium

In the minimal medium M9 containing glycerol as a carbon source, important genes included categories such as "amino acid transport and metabolism," "carbohydrate transport and

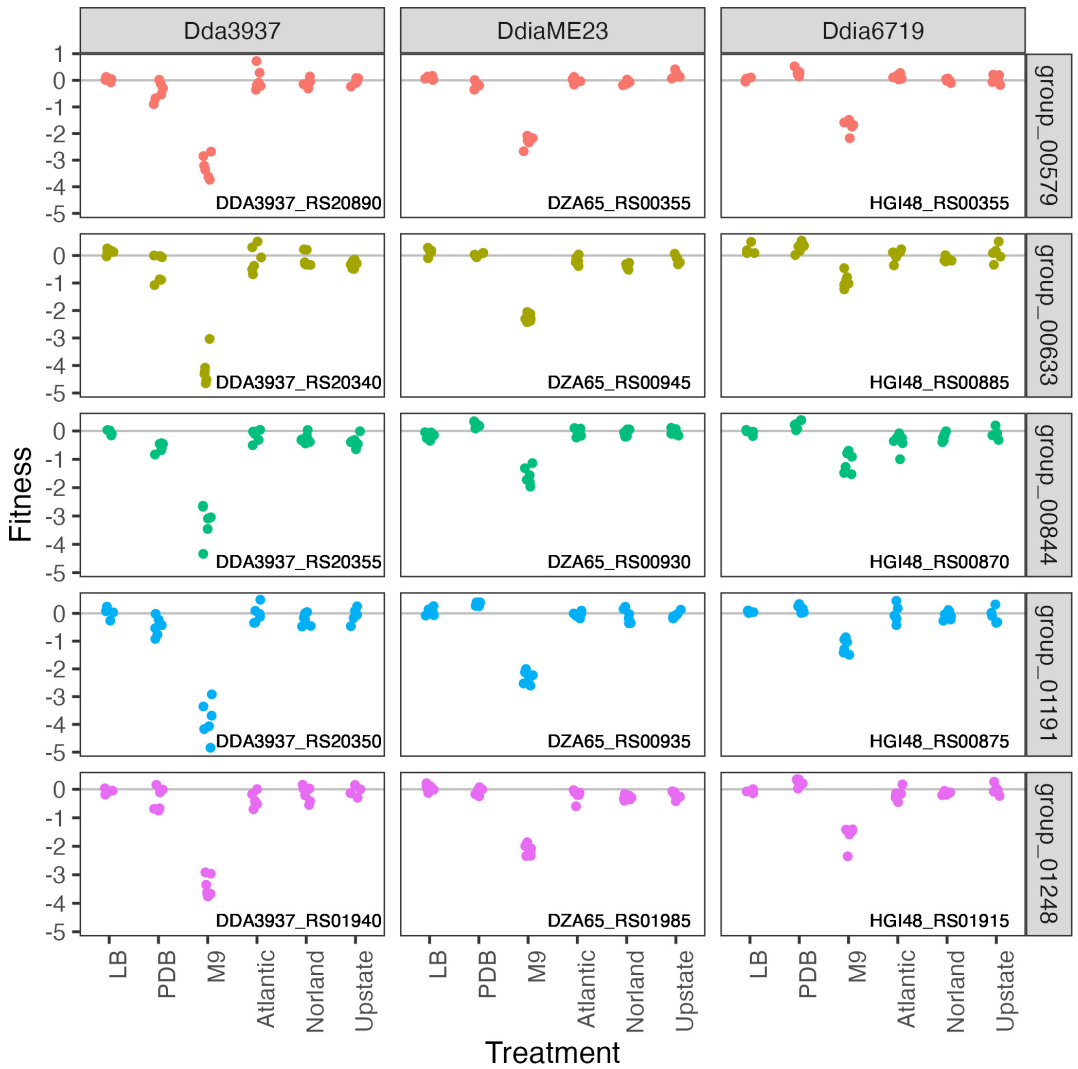


FIGURE 4 | Mutations in arginine biosynthetic genes are detrimental only in M9 minimal medium. Gene fitness values for argininosuccinate synthase ArgG (group 00579), argininosuccinate lyase ArgH (group 00633), acetylornithine deacetylase ArgE (group 00844), N-acetyl-gamma-glutamyl-phosphate reductase ArgC (group 01191), and ornithine carbamoyltransferase ArgF (group 01248).

metabolism," "coenzyme transport and metabolism," and "nucleotide transport and metabolism." While many amino acids were limiting in both M9 and tuber samples, arginine biosynthetic genes (*argCEFGH*) were uniquely important in M9, suggesting the presence of available arginine in tubers (**Figure 4**). Conversely, mutations in many "cell motility" genes had a large positive effect, though this effect was often limited to *Dda*3937 (**Figure 5**). This is indicative of the limited energy available in minimal medium, and the high energy cost of motility.

Genes Contributing to Growth in Tubers

To calculate genome-wide gene fitness values in an ecologically and economically relevant condition, we inoculated the transposon libraries into tubers of three potato cultivars: "Atlantic," "Dark Red Norland," and "Upstate Abundance." As each transposon library contains over 300,000 unique strains, we inoculated approximately 10^7 cells into each tuber (10 µl of a 10^9 CFU/ml solution). After 2 days incubation at high humidity, we recovered cells by streaming for barcode sequencing and calculation of gene fitness values. Many genes involved in amino acid biosynthesis that were important for growth in M9 were also important in tubers (*leuAC, thrC, serB*), highlighting potentially limiting factors for growth during potato soft rot (**Figure 6**).

The pectin degradation protein 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase KduD was specifically important for growth in tubers (**Figure 7**). Interestingly, we identified several putative DNA-binding or helix-turn-helix transcriptional regulators where mutant strains had strain-specific increased fitness in tubers (**Figure 7**). Insertions in the *Ddia*6719 helix-turn-helix transcriptional regulator *HGI48_RS01985* increased fitness in tubers, while insertions in the paralog *HGI48_RS02000* had no effect on fitness. There is no ortholog for this gene in

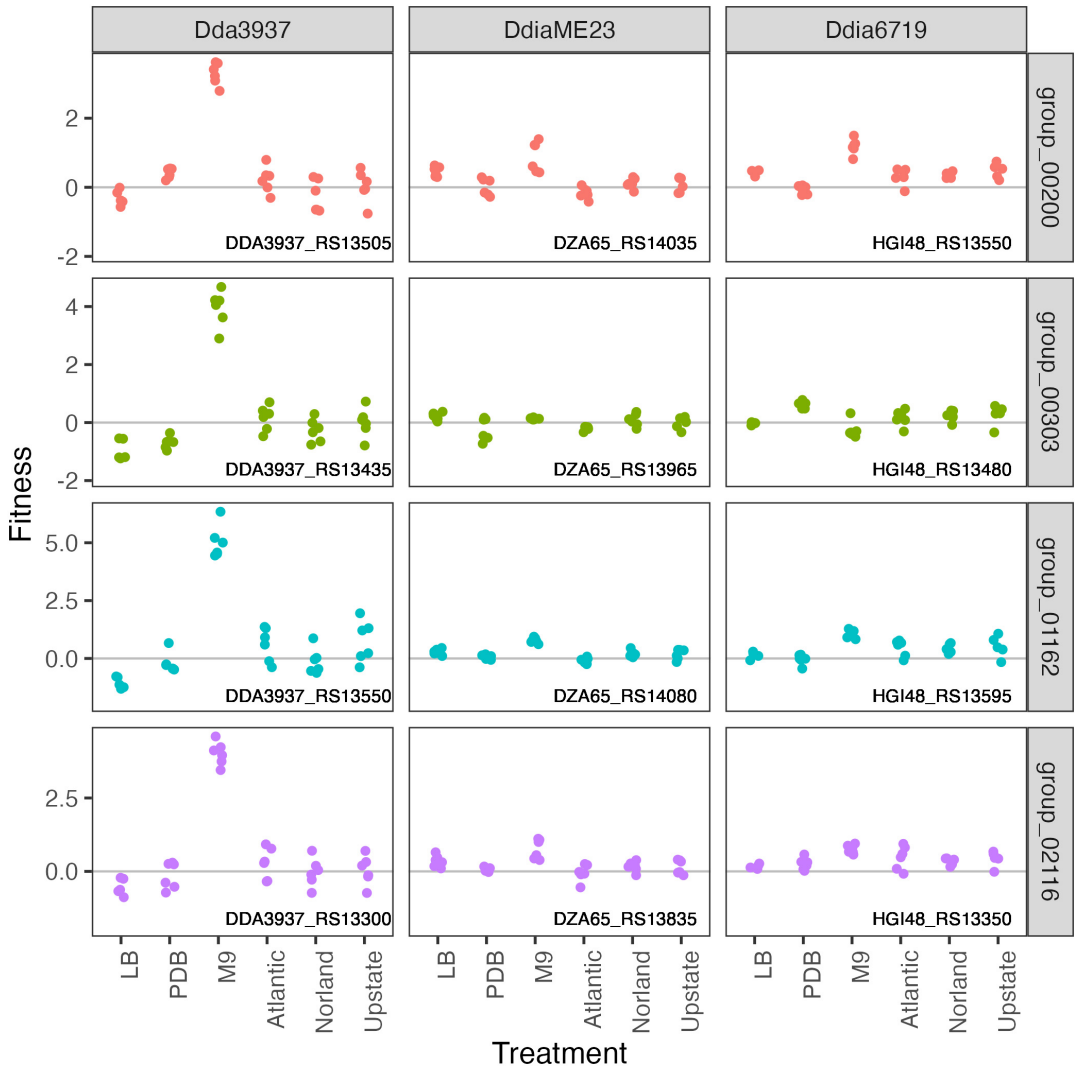


FIGURE 5 | Mutations in flagellar-associated genes increase competitive fitness of *D. dadantii* 3937 in M9 minimal medium. Gene fitness values for flagellar biosynthesis protein FlhA (group 00200), flagellar hook-associated protein FlgK (group 00303), flagellar motor protein MotB (group 001162), and RNA polymerase sigma factor FliA (group 02116).

Dda3937, and the ortholog in *Ddia*ME23 had no disruption phenotype in any condition tested (**Figure 7**).

DISCUSSION

The soft rot infection process is complex, requiring tolerance of stress conditions and coordination of disease-stage specific virulence traits (Jiang et al., 2016; Reverchon et al., 2016). Despite taxonomic and diagnostic improvements within the soft rot *Pectobacteriaceae*, our molecular understanding of these pathogens remains highly reliant on the model strain *D. dadantii* (previously *Erwinia chrysanthemi*) 3937 (Toth et al., 2011; Van Gijsegem et al., 2021). However, *D. dianthicola* strains have been among those more recently responsible for outbreaks in the United States and globally (Toth et al., 2011; van der Wolf et al., 2021). It is therefore important to consider common and variable virulence traits of *Dickeya* species; to better understand the infection processes and associated risk factors.

In this study, we focused specifically on bacterial colonization of potato tubers, an ecologically important infection model. *Dickeya* species share many common virulence strategies, including the production of plant cell wall degrading enzymes such as pectate lyases (Reverchon et al., 2016). However, virulence traits can vary even among closely related strains. For example, Ge et al. (2021b,c) reported the *D. dianthicola* strains isolated in the United States comprise multiple genetic groups, with differing type IV and type VI secretion system genes. These differences, determined by pan-genomic analysis, suggest driving forces for population level changes in outbreaks which can be explored by molecular characterization (Ge et al., 2021c). By constructing barcoded transposon insertion libraries in *D. dadantii* and

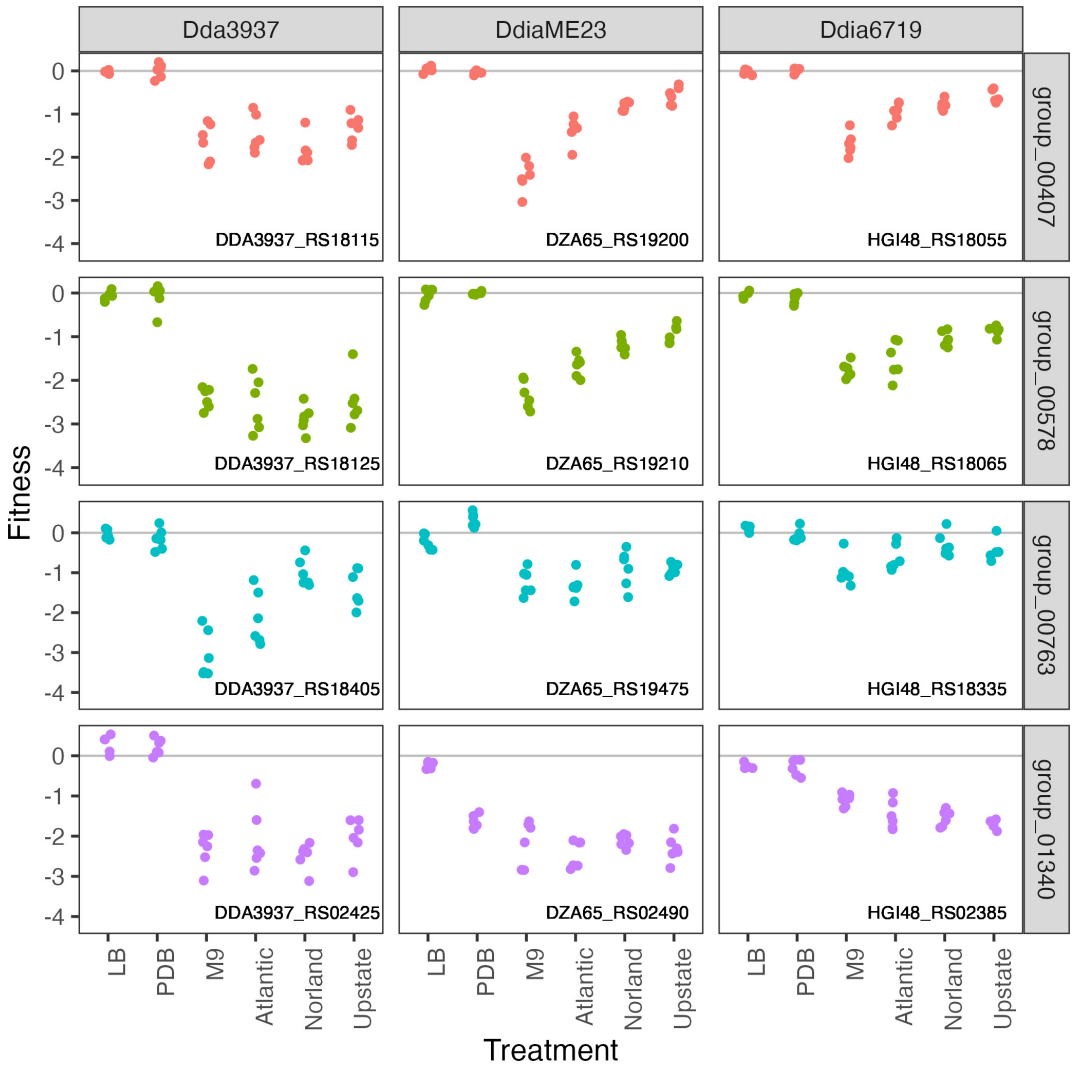


FIGURE 6 | Amino acid biosynthetic genes important in M9 minimal medium as well as growth in potato tubers. Gene fitness values for 2-isopropylmalate synthase LeuA (group 00407), 3-isopropylmalate dehydratase large subunit LeuC (group 00578), threonine synthase ThrC (group 00763), and phosphoserine phosphatase SerB (group 01340).

D. dianthicola, and testing these mutant populations in parallel conditions, we were able to directly compare gene contributions to fitness among these related strains. The scalability of RB-TnSeq, paired with ortholog prediction or gene co-fitness measurements, has proven to be a useful method to compare ortholog presence and function (Price et al., 2018). By measuring genome-wide gene contributions to growth, we identified a comprehensive list of genes in three *Dickeya* strains that contribute to fitness in three *in vitro* conditions and tubers. This approach could be applied to any (genetically tractable) soft rot causing strains, particularly building upon pan-genomic analysis to identify key strains of interest.

In the strains tested here, many core metabolic processes were highly important in both minimal medium and tuber samples, such as biosynthesis of many amino acids. Notably, potato tubers contain higher arginine concentrations than other essential amino acids (Bártová et al., 2015), correlating with the dispensability of the arginine biosynthetic genes (*argCEFGH*) in tubers (**Figure 4**). Interestingly, though an *Erwinia amylovora argD* mutant is auxotrophic and non-pathogenic in apple (Ramos et al., 2014), *argD* is present in two apparently redundant copies in *Dda*3937 (DDA3937_RS19450 and DDA3937_RS03635), and mutants in these genes had no phenotype in the conditions tested. Previous studies have shown the important of chemotaxis and motility for early stage virulence (Jahn et al., 2008), but these traits were dispensable for growth in tubers with the inoculation and sampling methods used here. We used potato dextrose broth as an *in vitro* condition to mimic the nutritional composition of potato tubers more closely. This medium was slightly acidic, reflecting conditions encountered in the tuber or apoplast (Grignon and Sentenac, 1991; Kiszonas and Bamberg, 2010). Of the PDB-specific fitness

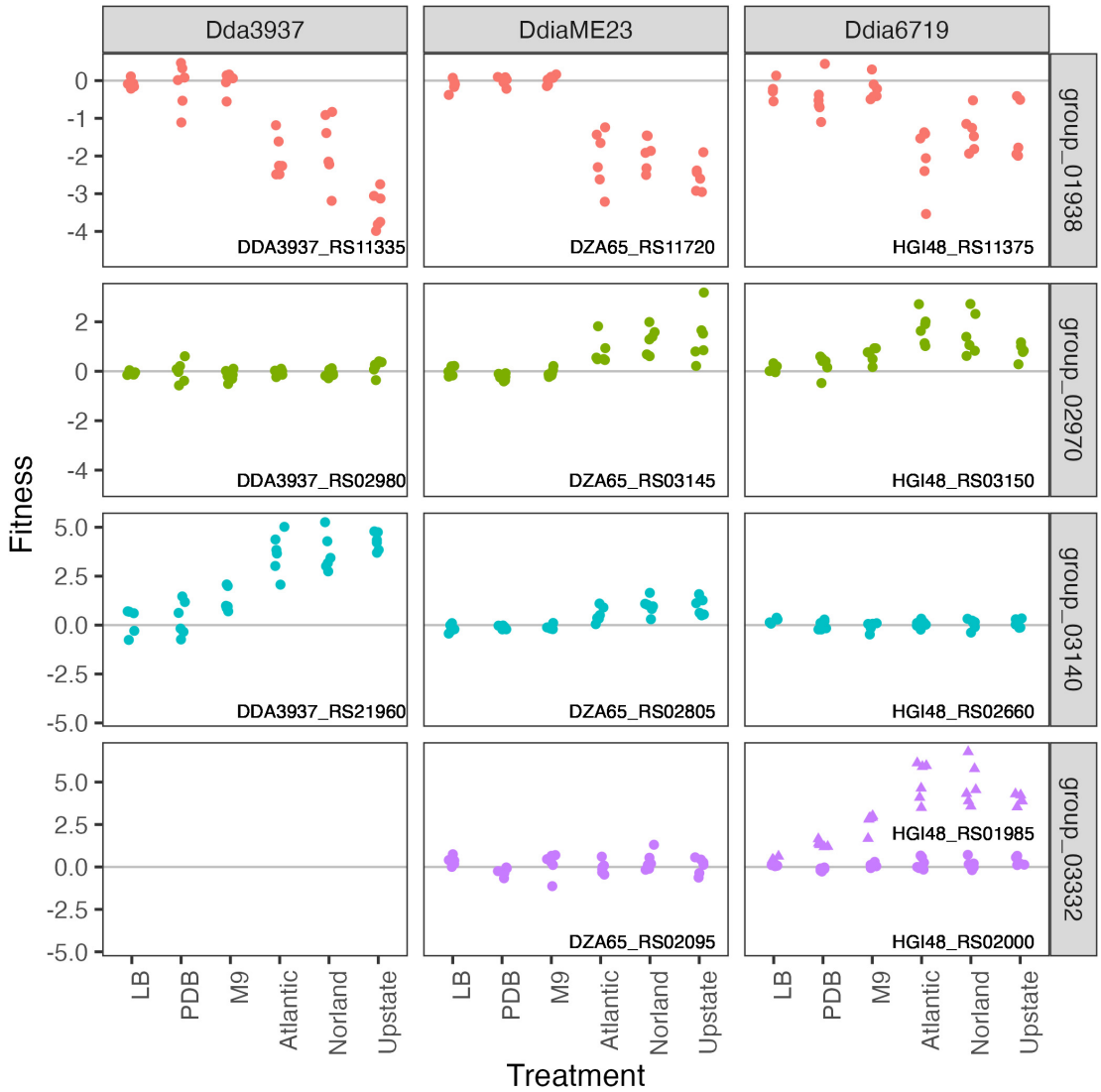


FIGURE 7 | In tubers, pectin degradation contributes to growth, while the role of putative DNA-binding proteins varies among strains. Pectin degradation by 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase KduD (group 01938) is specifically important for growth in tubers. Gene fitness values shown for three putative DNA binding proteins: ortholog groups 02970, 03140, and 03332. No ortholog was detected in *D. dadantii* 3937 for group 03332 (helix-turn-helix transcriptional regulators). In *D. dianthicola* 67-19, there are two genes in group 03332 (HGI48_RS01985, triangles; HGI48_RS02000, circles). All other orthogroup members are single-copy in each strain.

factors, the low affinity potassium transporter Kup was likely required because of its role in ion transport particularly at low pH (Trchounian and Kobayashi, 1999).

Bacterial growth on potato tubers requires plant-specific virulence traits. Interestingly, the gene fitness profiles were indistinguishable among the potato cultivars tested. It is possible that tubers of alternative cultivars might sufficiently differ in their chemical composition to alter the metabolic genes needed for colonization. It would be useful to see if these results are consistent across additional plant tissues, such as the stem. Our data generally support *in planta* findings from other groups, from the importance of core metabolic capabilities during tuber colonization to the production of the pectin degradation protein KduD (Condemine and Robert-Baudouy, 1991; Royet et al., 2019; Czajkowski et al., 2020). Transposon mutagenesis of *P. carotovorum* followed by screening for altered soft rot symptoms in Chinese cabbage identified homologous genes involved in nutrient utilization, production of plant cell wall degrading enzymes, motility, biofilm formation, and toxin susceptibility (Lee et al., 2013). In chicory, *Dda*3937 required core metabolic and nucleic acid biosynthetic genes, flagellar motility, and iron uptake genes for full virulence (Royet et al., 2019). As in Royet et al. (2019) study, type III secretion system genes were dispensable for virulence when inoculated at high concentrations here. Type III effectors may promote initial multiplication within the host, but their role in *Dickeya*

necrotrophic growth is secondary to the production of plant cell wall degrading enzymes (Yang et al., 2010). Though pectate lyases are collectively important for virulence (Condemine and Robert-Baudouy, 1991; Nachin and Barras, 2000), their redundancy makes mutagenesis-based study difficult (Schoedel and Collmer, 1986; Lojkowska et al., 1995). In all three *Dickeya* strains tested here, each of six pectate lyase genes were individually dispensable, supporting functional redundancy and/or complementation *in trans* (**Supplementary Figure 6**).

While *D. dadantii* and *D. dianthicola* can cause soft rot on potato tubers, variation in some key genes under these conditions suggests species- and strain-level differences in virulence strategies and stress responses. General strategies for environmental growth and host colonization are consistent, such as general metabolic capabilities and stress tolerance. However, gene fitness data suggest variation in gene regulation, such as the helix-turn-helix and other putative DNA-binding proteins. Large positive fitness values in *Dda*3937 flagellar mutants *in vitro* suggest that in liquid minimal medium flagella are costly. Since the same phenotype is not observed for *Ddia*ME23 or *Ddia*6719, this might imply flagella are controlled differently in liquid media in these species. Further characterization of the regulation of these traits is needed.

The scalability of RB-TnSeq, paired with ortholog identification, has proven to be a useful method to directly compare gene fitness between related strains. *Dickeya* species generally have common virulence strategies, primarily the production of plant cell wall degrading enzymes such as pectate lyases (Reverchon et al., 2016). However, genomic and transcriptomic variation at the strain and species level highlights distinctive virulence traits (Raoul des Essarts et al., 2019). This supports the intriguing possibility that while enzymatic virulence traits are shared across pathogens, there exists strain-specific virulence regulation. This idea has been proposed, but not directly tested, in *D. solani* based on predicted binding sites for transcriptional regulators (Golanowska et al., 2018). In the case of our study, while *Dda*3937 is pathogenic on potato, it was originally isolated from *Saintpaulia ionantha* (Lemattre and Narcy, 1972), suggesting potato infection is simply opportunistic. *Ddia*6719 was originally isolated from New Guinea impatiens (*Impatiens hawkeri*) (Liu et al., 2020a), but observed symptoms in tubers were similar to those caused by *Ddia*ME23.

This study focused on isolated strain growth, to generate a comprehensive dataset of likely essential genes and soft rot virulence traits in *D. dadantii* and *D. dianthicola*. Testing other *Dickeya* species, and closely related pathogens such as *Pectobacterium* spp., will more broadly expand our understanding of soft rot pathogens. With interest in using biocontrol strains or applying antimicrobial small molecules directly for control of soft rot (Czajkowski et al., 2011), it would be useful to screen pathogenic strains for potential resistance mechanisms. Similarly, in the field, soft rot symptoms can be the result of complex community interactions, with *Dickeya* and *Pectobacterium* co-infections frequently observed (Ge et al., 2021a). This presence of additional community members might change the composition of genes required for full competitive fitness. Understanding bacterial virulence strategies will aid in breeding efforts, as well as identify potential bacterial traits that could enable overcoming host tolerance or exacerbating disease at all stages of production. Similarly, the ability to rapidly predict and identify specific virulence traits in novel isolates will be key to addressing emerging outbreak pathogens.

DATA AVAILABILITY STATEMENT

All raw Illumina reads used for mapping and fitness assays have been deposited in the Sequence Read Archive under BioProject accession number PRJNA692477. Individual sample accession numbers are listed in **Supplementary Table 2**. Annotated scripts used for computational analysis are available at github.com/tylerhelmann/dickeya-barseq-2021. Experimental fitness values are publicly available at fit.genomics.lbl.gov.

AUTHOR CONTRIBUTIONS

TH, MF, and PS contributed to conception and designed of the study. TH performed the experiments and wrote the first draft of the manuscript. TH and PS performed data analysis. MF and PS supervised the project. All authors contributed to data interpretation, contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This project was funded by the USDA-ARS project no. 8062-2100-042.

ACKNOWLEDGMENTS

We would like to thank Mary Ann Karp for technical assistance; Walter De Jong and the Cornell Potato Breeding Program for providing potato tubers; Adam Deutschbauer for providing the donor *E. coli* library APA752 containing the barcoded *mariner* vector; and Morgan Price for assisting with the FEBA pipeline and data upload to the Fitness Browser. Sequencing was performed by the Biotechnology Resource Center (BRC) Genomics Facility at the Cornell Institute of Biotechnology. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture. USDA is an equal opportunity provider and employer.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.778927/full#supplementary-material

REFERENCES

Adeolu, M., Alnajar, S., Naushad, S., and Gupta, R. S. (2016). Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* 66, 5575–5599. doi: 10.1099/ijsem.0.001485

Adriaenssens, E. M., van Vaerenbergh, J., Vandenheuvel, D., Dunon, V., Ceyssens, P. J., de Proft, M., et al. (2012). T4-related bacteriophage LIMEstone isolates for the control of soft rot on potato caused by "Dickeya solani.". *PLoS One* 7:e33227. doi: 10.1371/journal.pone.0033227

Bártová, V., Bárta, J., Brabcová, A., Zdráhal, Z., and Horáčková, V. (2015). Amino acid composition and nutritional value of four cultivated South American potato species. *J. Food Compos. Anal.* 40, 78–85. doi: 10.1016/j.jfca.2014.12.006

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* 62, 293–300. doi: 10.1128/JB.62.3.293-300.1951

Chen, I. M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., et al. (2019). IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* 47, D666–D677. doi: 10.1093/nar/gky901

Chung, Y. S., Goeser, N. J., Cai, X., and Jansky, S. (2013). The effect of long term storage on bacterial soft rot resistance in potato. *Am. J. Potato Res.* 90, 351–356. doi: 10.1007/s12230-013-9311-6

Cole, B. J., Feltcher, M. E., Waters, R. J., Wetmore, K. M., Mucyn, T. S., Ryan, E. M., et al. (2017). Genome-wide identification of bacterial plant colonization genes. *PLoS Biol.* 15:e2002860. doi: 10.1371/journal.pbio.2002860

Condemine, G., and Robert-Baudouy, J. (1991). Analysis of an *Erwinia chrysanthemi* gene cluster involved in pectin degradation. *Mol. Microbiol.* 5, 2191–2202. doi: 10.1111/j.1365-2958.1991.tb02149.x

Czajkowski, R., Fikowicz-Krosko, J., Maciag, T., Rabalski, L., Czaplewska, P., Jafra, S., et al. (2020). Genome-wide identification of *Dickeya solani* transcriptional units up-regulated in response to plant tissues from a crop-host *Solanum tuberosum* and a weed-host *Solanum dulcamara*. *Front. Plant Sci.* 11:580330. doi: 10.3389/fpls.2020.580330

Czajkowski, R., Pérombelon, M. C. M., Van Veen, J. A., and Van der Wolf, J. M. (2011). Control of blackleg and tuber soft rot of potato caused by *Pectobacterium* and *Dickeya* species: a review. *Plant Pathol.* 60, 999–1013. doi: 10.1111/j.1365-3059.2011.02470.x

Czajkowski, R., Smolarska, A., and Ozymko, Z. (2017). The viability of lytic bacteriophage ΦD5 in potato-associated environments and its effect on *Dickeya solani* in potato (*Solanum tuberosum* L.) plants. *PLoS One* 12:e183200. doi: 10.1371/journal.pone.0183200

Ge, T., Jiang, H., Tan, E. H., Johnson, S. B., Larkin, R., Charkowski, A. O., et al. (2021c). Pangenomic analysis of *Dickeya dianthicola* strains related to the outbreak of blackleg and soft rot of potato in USA. *Plant Dis. [Online ahead of print]* DIS03210587RE. doi: 10.1094/PDIS-03-21-0587-RE

Ge, T., Jiang, H., Johnson, S. B., Larkin, R. P., Charkowski, A. O., Secor, G., et al. (2021b). Genotyping *Dickeya dianthicola* causing potato blackleg and soft rot outbreak associated with inoculum geography in the United States. *Plant Dis.* 105, 1976–1983. doi: 10.1094/PDIS-10-20-2138-RE

Ge, T., Ekbataniamiri, F., Johnson, S. B., Larkin, R. P., and Hao, J. (2021a). Interaction between *Dickeya dianthicola* and *Pectobacterium parmentieri* in potato infection under field conditions. *Microorganisms* 9, 1–10. doi: 10.3390/microorganisms9020316

Georgoulis, S., Shalvarjian, K. E., Helmann, T. C., Hamilton, C. D., Carlson, H. K., Deutschbauer, A. M., et al. (2021). Genome-wide identification of tomato xylem sap fitness factors for three plant-pathogenic *Ralstonia* species. *mSystems* 6, e1229–e1221. doi: 10.1101/2020.08.31.276741

Golanowska, M., Potrykus, M., Motyka-Pomagruk, A., Kabza, M., Bacci, G., Galardini, M., et al. (2018). Comparison of highly and weakly virulent *Dickeya solani* strains, with a view on the pangenome and panregulon of this species. *Front. Microbiol.* 9:1940. doi: 10.3389/fmicb.2018.01940

Grignon, C., and Sentenac, H. (1991). pH and ionic conditions in the apoplast. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 42, 103–128. doi: 10.1146/annurev.pp.42.060191.000535

Helmann, T. C., Deutschbauer, A. M., and Lindow, S. E. (2019). Genome-wide identification of *Pseudomonas syringae* genes required for fitness during colonization of the leaf surface and apoplast. *Proc. Natl. Acad. Sci. U.S.A.* 116, 18900–18910. doi: 10.1073/pnas.1908858116

Jahn, C. E., Willis, D. K., and Charkowski, A. O. (2008). The flagellar sigma factor FliA is required for *Dickeya dadantii* virulence. *Mol. Plant-Microbe Interact.* 21, 1431–1442. doi: 10.1094/MPMI-21-11-1431

Jiang, X., Zghidi-Abouzid, O., Oger-Desfeux, C., Hommais, F., Greliche, N., Muskhelishvili, G., et al. (2016). Global transcriptional response of *Dickeya dadantii* to environmental stimuli relevant to the plant infection. *Environ. Microbiol.* 18, 3651–3672. doi: 10.1111/1462-2920.13267

Kiszonas, A. M., and Bamberg, J. B. (2010). Survey of tuber pH variation in potato (*Solanum*) species. *Am. J. Potato Res.* 87, 167–176. doi: 10.1007/s12230-009-9120-0

Lee, D. H., Lim, J. A., Lee, J., Roh, E., Jung, K., Choi, M., et al. (2013). Characterization of genes required for the pathogenicity of *Pectobacterium carotovorum* subsp. *carotovorum* Pcc21 in Chinese cabbage. *Microbiology (United Kingdom)* 159, 1487–1496. doi: 10.1099/mic.0.067280-0

Lemattre, M., and Narcy, J. P. (1972). Une affection bacterienne nouvelle du Saintpaulia due a *Erwinia chrysanthemi*. *C. R. Acad. Sci* 58, 227–231.

Liu, H., and Deutschbauer, A. M. (2018). Rapidly moving new bacteria to model-organism status. *Curr. Opin. Biotechnol.* 51, 116–122. doi: 10.1016/j.copbio.2017.12.006

Liu, H., Shiver, A. L., Price, M. N., Carlson, H. K., Trotter, V. V., Chen, Y., et al. (2021). Functional genetics of human gut commensal *Bacteroides thetaiotaomicron* reveals metabolic requirements for growth across environments. *Cell Rep.* 34:108789. doi: 10.1016/j.celrep.2021.108789

Liu, Y., Helmann, T., Stodghill, P., and Filiatrault, M. (2020a). Complete genome sequence resource for the necrotrophic plant-pathogenic bacterium *Dickeya dianthicola* 67-19 isolated from New Guinea Impatiens. *Plant Dis.* 105, 1174–1176. doi: 10.1094/PDIS-09-20-1968-A

Liu, Y., Vasiu, S., Daughtrey, M. L., and Filiatrault, M. (2020b). First Report of *Dickeya dianthicola* causing blackleg on New Guinea Impatiens (*Impatiens hawkeri*) in New York State, USA. *Plant Dis. [Online ahead of print]* doi: 10.1094/pdis-09-20-2020-pdn

Lojkowska, E., Masclaux, C., Boccara, M., Robert-Baudouy, J., and Hugouvieux-Cotte-Pattat, N. (1995). Characterization of the *pelL* gene encoding a novel pectate lyase of *Erwinia chrysanthemi* 3937. *Mol. Microbiol.* 16, 1183–1195.

Lyon, G. D. (1989). The biochemical basis of resistance of potatoes to soft rot *Erwinia* spp.—a review. *Plant Pathol.* 38, 313–339. doi: 10.1111/j.1365-3059.1989.tb02152.x

M9 Minimal Medium (Standard) (2010). M9 minimal medium (standard). *Cold Spring Harb. Protoc.* 2010:db.rec12295. doi: 10.1101/pdb.rec12295

Ma, X., Perna, N. T., Glasner, J. D., Hao, J., Johnson, S., Nasaruddin, A. S., et al. (2019). Complete genome sequence of *Dickeya dianthicola* ME23, a pathogen causing blackleg and soft rot diseases of potato. *Microbiol. Resour. Announc.* 8, 14–15. doi: 10.1128/mra.01526-18

Melnyk, R. A., Hossain, S. S., and Haney, C. H. (2019). Convergent gain and loss of genomic islands drives lifestyle changes in plant-associated *Pseudomonas*. *ISME J.* 13, 1575–1588. doi: 10.1101/345488

Motyka, A., Zoledowska, S., Sledz, W., and Lojkowska, E. (2017). Molecular methods as tools to control plant diseases caused by *Dickeya* and *Pectobacterium* spp: a minireview. *N. Biotechnol.* 39, 181–189. doi: 10.1016/j.nbt.2017.08.010

Motyka-Pomagruk, A., Zoledowska, S., Misztak, A. E., Sledz, W., Mengoni, A., and Lojkowska, E. (2020). Comparative genomics and pangenome-oriented studies reveal high homogeneity of the agronomically relevant enterobacterial plant pathogen *Dickeya solani*. *BMC Genomics* 21:449. doi: 10.1186/s12864-020-06863-w

Nachin, L., and Barras, F. (2000). External pH: an environmental signal that helps to rationalize pel gene duplication in *Erwinia chrysanthemi*. *Mol. Plant-Microbe Interact.* 13, 882–886. doi: 10.1094/MPMI.2000.13.8.882

Parkinson, N., DeVos, P., Pirhonen, M., and Elphinstone, J. (2014). *Dickeya aquatica* sp. nov., isolated from waterways. *Int. J. Syst. Evol. Microbiol.* 64, 2264–2266. doi: 10.1099/ijs.0.058693-0

Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, M., Ray, J., Liu, H., et al. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 557, 503–509. doi: 10.1038/s41586-018-0124-0

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramos, L. S., Lehman, B. L., Peter, K. A., and McNellis, T. W. (2014). Mutation of the *Erwinia amylovora argD* gene causes arginine auxotrophy, nonpathogenicity in apples, and reduced virulence in pears. *Appl. Environ. Microbiol.* 80, 6739–6749. doi: 10.1128/AEM.02404-14

Raoul des Essarts, Y., Pédron, J., Blin, P., Van Dijk, E., Faure, D., and Van Gijsegem, F. (2019). Common and distinctive adaptive traits expressed in *Dickeya dianthicola* and *Dickeya solani* pathogens when exploiting potato plant host. *Environ. Microbiol.* 21, 1004–1018. doi: 10.1111/1462-2920.14519

Reverchon, S., Muskhelisvili, G., and Nasser, W. (2016). Virulence program of a bacterial plant pathogen: the Dickeya model. *Progr. Mol. Biol. Transl. Sci.* 142, 51–92. doi: 10.1016/bs.pmbts.2016.05.005

Royet, K., Parisot, N., Rodrigue, A., Gueguen, E., and Condemine, G. (2019). Identification by Tn-seq of *Dickeya dadantii* genes required for survival in chicory plants. *Mol. Plant Pathol.* 20, 287–306. doi: 10.1111/mpp.12754

Rubin, B. E., Diamond, S., Cress, B. F., Crits-Christoph, A., Lou, Y. C., Borges, A. L., et al. (2021). Species- and site-specific genome editing in complex bacterial communities. *Nat. Microbiol.* 7, 34–47. doi: 10.1038/s41564-021-01014-7

Samson, R., Legendre, J. B., Christen, R., Fischer-Le Saux, M., Achouak, W., and Gardan, L. (2005). Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. *Int. J. Syst. Evol. Microbiol.* 55, 1415–1427. doi: 10.1099/ijs.0.02791-0

Schoedel, C., and Collmer, A. (1986). Evidence of homology between the pectate lyase-encoding *pelB* and *pelC* genes in *Erwinia chrysanthemi*. *J. Bacteriol.* 167, 117–123. doi: 10.1128/jb.167.1.117-123.1986

Tian, Y., Zhao, Y., Yuan, X., Yi, J., Fan, J., Xu, Z., et al. (2016). *Dickeya fangzhongdai* sp. nov., a plant-pathogenic bacterium isolated from pear trees (*Pyrus pyrifolia*). *Int. J. Syst. Evol. Microbiol.* 66, 2831–2835. doi: 10.1099/ijsem.0.001060

Toth, I. K., van der Wolf, J. M., Saddler, G., Lojkowska, E., Hélias, V., Pirhonen, M., et al. (2011). *Dickeya* species: an emerging problem for potato production in Europe. *Plant Pathol.* 60, 385–399. doi: 10.1111/j.1365-3059.2011.02427.x

Trchounian, A., and Kobayashi, H. (1999). Kup is the major K+ uptake system in *Escherichia coli* upon hyper-osmotic stress at a low pH. *FEBS Lett.* 447, 144–148. doi: 10.1016/S0014-5793(99)00288-4

van der Wolf, J. M., Acuña, I., De Boer, S. H., Brurberg, M. B., Cahill, G., Charkowski, A. O., et al. (2021). "Diseases caused by *Pectobacterium* and *Dickeya* species around the world," in *Plant Diseases Caused by Dickeya and Pectobacterium Species*, eds F. Van Gijsegem, J. M. van der Wolf, and I. K. Toth (Cham: Springer International Publishing), 215–261.

Van Der Wolf, J. M., Nijhuis, E. H., Kowalewska, M. J., Saddler, G. S., Parkinson, N., Elphinstone, J. G., et al. (2014). *Dickeya solani* sp . nov ., a pectinolytic plant-pathogenic bacterium isolated from potato (*Solanum tuberosum*). *Int. J. Syst. Evol. Microbiol.* 64, 768–774. doi: 10.1099/ijs.0.052944-0

Van Gijsegem, F., Toth, I. K., and van der Wolf, J. M. (2021). "Outlook-Challenges and perspectives for management of diseases caused by *Pectobacterium* and *Dickeya* species," in *Plant Diseases Caused by Dickeya and Pectobacterium Species*, eds F. Van Gijsegem, J. M. van der Wolf, and I. K. Toth (Cham: Springer International Publishing), 283–289.

van Opijnen, T., Bodi, K. L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* 6, 767–772. doi: 10.1038/nmeth.1377

Wetmore, K. M., Price, M. N., Waters, R. J., Lamson, J. S., He, J., Hoover, C. A., et al. (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* 6, 1–15. doi: 10.1128/mBio.00306-15

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

Yang, S., Peng, Q., Zhang, Q., Zou, L., Li, Y., Robert, C., et al. (2010). Genome-wide identification of HrpL-regulated genes in the necrotrophic phytopathogen *Dickeya dadantii* 3937. *PLoS One* 5:e13472. doi: 10.1371/journal.pone.0013472

Zhang, Y., Fan, Q., and Loria, R. (2016). A re-evaluation of the taxonomy of phytopathogenic genera *Dickeya* and *Pectobacterium* using whole-genome sequencing data. *Syst. Appl. Microbiol.* 39, 252–259. doi: 10.1016/j.syapm.2016.04.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 01 February 2022
doi: 10.3389/fmicb.2021.793345

Comparative Genome Analyses of *Clavibacter michiganensis* Type Strain LMG7333^T Reveal Distinct Gene Contents in Plasmids From Other *Clavibacter* Species

*Eom-Ji Oh[1], In Sun Hwang[1], In Woong Park[1] and Chang-Sik Oh[1,2]**

[1] Department of Horticultural Biotechnology, College of Life Science, Kyung Hee University, Yongin, South Korea, [2] Graduate School of Biotechnology, Kyung Hee University, Yongin, South Korea

Clavibacter michiganensis, a Gram-positive, plant-pathogenic bacterium belonging to Actinobacteria, is a causal agent of bacterial canker in tomatoes. Although LMG7333^T is the type strain of *C. michiganensis*, it has not been used in many studies, probably because of a lack of the complete genome sequence being available. Therefore, in this study, the complete genome sequence of this type strain was obtained, and comparative genome analysis was conducted with the genome sequences of two other *C. michiganensis* strains and type strains of *Clavibacter* species, of which their complete genome sequences are available. *C. michiganensis* LMG7333^T carries one chromosome and two plasmids, pCM1 and pCM2, like two other *C. michiganensis* strains. All three chromosomal DNA sequences were almost identical. However, the DNA sequences of two plasmids of LMG7333^T are similar to those of UF1, but different from those of NCPPB382, indicating that both plasmids carry distinct gene content among *C. michiganensis* strains. Moreover, 216 protein-coding sequences (CDSs) were only present in the LMG7333^T genome compared with type strains of other *Clavibacter* species. Among these 216 CDSs, approximately 83% were in the chromosome, whereas others were in both plasmids (more than 6% in pCM1 and 11% in pCM2). However, the ratio of unique CDSs of the total CDSs in both plasmids were approximately 38% in pCM1 and 30% in pCM2, indicating that the high gene content percentage in both plasmids of *C. michiganensis* are different from those of other *Clavibacter* species, and plasmid DNAs might be derived from different origins. A virulence assay with *C. michiganensis* LMG7333^T using three different inoculation methods, root-dipping, leaf-clipping, and stem injection, resulted in typical disease symptoms, including wilting and canker in tomato. Altogether, our results indicate that two plasmids of *C. michiganensis* carry distinct gene content, and the genome information of the type strain LMG7333^T will help to understand the genetic diversity of the two plasmids of *Clavibacter* species, including *C. michiganensis*.

Keywords: *Clavibacter michiganensis*, comparative genomics, complete genome sequence, Gram-positive bacteria, type strain

OPEN ACCESS

Edited by:
Ralf Koebnik,
Plant Health Institute, France

Reviewed by:
Mohammad Arif,
University of Hawai'i at Mānoa
Ebrahim Osdaghi,
University of Tehran, Iran

***Correspondence:**
Chang-Sik Oh
co35@khu.ac.kr

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: *12 October 2021*
Accepted: *16 December 2021*
Published: *01 February 2022*

Citation:
Oh E-J, Hwang IS, Park IW and Oh C-S (2022) Comparative Genome Analyses of Clavibacter michiganensis Type Strain LMG7333^T Reveal Distinct Gene Contents in Plasmids From Other Clavibacter Species. Front. Microbiol. 12:793345. doi: 10.3389/fmicb.2021.793345

INTRODUCTION

Plant pathogenic bacteria belonging to the genus *Clavibacter* are Gram-positive and often cause severe damage to crops (Eichenlaub and Gartemann, 2011). This genus was reclassified from *Corynebacterium* and newly identified as a new genus, including several phytopathogenic bacteria (Davis et al., 1984). The genus *Clavibacter* contained the single species, *C. michiganensis* with nine subspecies, according to its host specificity (Eichenlaub et al., 2007; Gonzalez and Trapiello, 2014; Yasuhara-Bell and Alvarez, 2015; Oh et al., 2016). Recently, seven out of eight subspecies except *C. michiganensis* subsp. *chilensis* (non-pathogenic bacterium) were reclassified as separate species based on the genomic analysis in addition to its host specificity (Li et al., 2018; Méndez et al., 2020): *C. michiganensis* (bacterial canker and wilting in tomato) (Smith, 1910), *C. sepedonicus* (ring rot in potato) (Manzer and Genereux, 1981), *C. nebraskensis* (Goss's leaf blight and wilting in maize) (Vidaver and Mandel, 1974), *C. insidiosus* (bacterial wilt in alfalfa) (McCulloch, 1925), *C. tessellarius* (leaf freckles and leaf spots in wheat) (Carlson and Vidaver, 1982), *C. phaseoli* (bacterial leaf yellowing in beans) (Gonzalez and Trapiello, 2014), and *C. capsici* (bacterial canker in pepper) (Oh et al., 2016) and a non-pathogenic *C. californiensis* (Yasuhara-Bell and Alvarez, 2015). Although Méndez et al. (2020) also propose reclassifying *C. michiganensis* subsp. *chilensis* to *C. phaseoli*, we kept the original subspecies name to distinguish this strain from the original *C. phaseoli* in this study.

Disease through *Clavibacter* bacterial infection resulted in economic damage due to the significant yield loss. For example, *C. michiganensis* was first reported in 1910 in North America (Smith, 1910), and since then, this pathogen has caused significant economic loss (Xu et al., 2012; Peritore-Galve et al., 2020). Additionally, *C. sepedonicus* was designated a quarantine pathogen in Europe owing to its severe economic damage annually (Van der Wolf et al., 2005). *C. michiganensis* has been studied relatively well among *Clavibacter* species. Although virulence mechanisms of this pathogen are not well known yet, previous research shows that two plasmid-borne genes, *celA* and *pat-1*, play essential roles in *C. michiganensis* pathogenicity (Dreier et al., 1997; Jahr et al., 2000; Burger et al., 2005; Hwang et al., 2019). Moreover, a pathogenicity island (PAI) region, which contains putative virulence genes such as the *ppa*, *chp*, and *pel* gene families, was found in the chromosome (Gartemann et al., 2008; Stork et al., 2008). In tomato seeds, *C. michiganensis* ssp. *chilensis* and *C. californiensis* were isolated and exhibited non-pathogenic effects to tomato plants (Yasuhara-Bell and Alvarez, 2015). Notably, these two non-pathogenic bacteria missed most virulence genes and/or putative virulence genes possessed by pathogenic *C. michiganensis* (Méndez et al., 2020).

So far, the complete genome sequences of type strains of only three *Clavibacter* species have been reported: *C. sepedonicus* ATCC33113^T (Bentley et al., 2008), *C. nebraskensis* NCPPB2581^T (GenBank accession #: HE614873), and *C. capsici* PF008^T (Bae et al., 2015). In the case of *C. insidiosus* species, the complete genome sequence of strain R1-1, which is not a type strain, is available in the public database (Lu et al., 2015). Based on these whole genome sequences of *Clavibacter* species in the public database, studies on *Clavibacter* species at the genomic level have recently increased. Moreover, newly sequenced genome data of *Clavibacter* strains have allowed us to conduct comparative genome analysis at the species or strain level (Tambong, 2017; Osdaghi et al., 2020). Indeed, several comparative genomic approaches have been conducted to target *Clavibacter* species isolated from specific geographic regions using genome database (Thapa et al., 2017; Hwang et al., 2020a; Méndez et al., 2020), and the genomic analysis is becoming more important to the genus *Clavibacter* research.

In the case of a tomato pathogen *C. michiganensis* species, only four strains, NCPPB382 (Gartemann et al., 2008), UF1, and two Chile isolates, VL527 and MSF322 (Méndez et al., 2020), were completely sequenced. Most of *C. michiganensis* genome sequences have not been completely assembled (Thapa et al., 2017). Although the draft sequence of *C. michiganensis* type strain LMG7333^T (= NCPPB 2979^T, CFBP 2352^T, DSM 46364^T) was already reported, it has not yet been completed (Li and Yuan, 2017; Li et al., 2018). The draft genome sequencing of *C. michiganensis* LMG7333^T was conducted previously by PacBio-SMRT sequencing technology (Li and Yuan, 2017), resulting in five contigs. Because this genome sequence was not complete, there was a limitation to use it as a reference genome of *C. michiganensis* species. For these reasons, NCPPB382 rather than a type strain of *C. michiganensis* was mainly used for many studies that have focused on molecular mechanisms to understand virulence mechanism(s) even though this type strain can induce typical disease as severe as *C. michiganensis* NCPPB382 on tomato plants (Hwang et al., 2019).

In this study, we report the completed genome sequence of *C. michiganensis* type strain LMG7333^T and use this sequence for comparison with other previously published genomes of *C. michiganensis* to obtain more valid insights onto genetic diversity among pathogenic *Clavibacter* species. Additionally, we conducted a virulence assay with this type strain using three inoculation methods. Our results suggest that comparative genome analysis with the complete genome sequence of *C. michiganensis* type strains LMG7333^T can provide a genomic resource to understand the genetic diversity among *Clavibacter* species better.

MATERIALS AND METHODS

Genome Sequencing, Assembly, and Annotation of *Clavibacter michiganensis* LMG7333^T

The genomic DNA (gDNA) of *C. michiganensis* LMG7333^T was extracted using the solution-based HiGene Genomic DNA Prep kit (BIOFACT, Daejeon, Korea), according to the manufacturer's instructions. Using extracted gDNA, whole genome sequencing of LMG7333^T was conducted using combined methods of two sequencing platforms, PacBio RSII and Illumina HiSeqXten

(Macrogen, Seoul, Korea), for genome sequence completion. The reads obtained from PacBio RSII were assembled by Hierarchical Genome Assembly Process (Chin et al., 2013). HiseqXten reads were also used to correct assembly by Pilon v1.21 (Walker et al., 2014). Based on the assembled sequences, gene prediction and annotation were conducted by Prokka v1.13 (Seemann, 2014). Then, InterProScan v5.30–69.0 (Jones et al., 2014) and psiblast v2.4.0 (Camacho et al., 2009) with EggNOG DB v4.5 (Huerta-Cepas et al., 2016) and the Clustering Orthologous Groups (COG) database v2014 (Tatusov et al., 2000) were used to confirm the predicted proteins. Circos program v0.69.3 (Krzywinski et al., 2009) was used for drawing circular maps of the genome.

Genome Comparison of *Clavibacter michiganensis* LMG7333^T

The completed whole genome sequence of *C. michiganensis* LMG7333^T type strain was first compared by aligning the EasyFig program (Sullivan et al., 2011) and NCBI blast program with those of *C. michiganensis* NCPPB382 (GenBank accession #: AM711865, AM711866, and AM711867) (Gartemann et al., 2008) and UF1 (GenBank accession #: CP033724, CP033725, and CP033726). Then, it was also compared with completed genome sequences of nine other *Clavibacter* species, including several type strains such as *C. sepedonicus* ATCC33113^T (GenBank accession #: AM849034) (Bentley et al., 2008), *C. capsici* PF008^T (GenBank accession #: CP012573, CP012574, and CP012575) (Bae et al., 2015), *C. insidiosus* R1-1 (GenBank accession #: CP011043, CP011044, CP011045, and CP011046) (Lu et al., 2015), *C. nebraskensis* NCPPB2581^T (GenBank accession #: HE614873), *C. insidiosus* LMG3663^T (GenBank accession #: MZMO01000001, MZMO01000002, and MZMO01000003), *C. tessellarius* ATCC33566^T (GenBank accession #: MZMQ01000001 and MZMQ01000002) (Li and Yuan, 2017; Li et al., 2018), *C. phaseoli* CFBP8627^T (GenBank accession #: QWGV00000000), *C. michiganensis* subsp. *chilensis* CFBP8217^T (GenBank accession #: QWGS01000000), and *C. californiensis* CFBP8216^T (GenBank accession #: QWEE01000000) (Méndez et al., 2020). The genome sequence of *Leifsonia xyli* subsp. *cydontis* DSM46306^T (GenBank accession #: CP006734) (Monteiro-Vitorello et al., 2013), which belongs to the same family as genus *Clavibacter*, was used as an out group. To understand the phylogenetic relationships among the analyzed strains, OrthoANI values were calculated, and UPGMA dendrograms were generated using the Orthologous ANI Tool of ChunLab (Lee et al., 2016). Moreover, the pan-genome analysis was conducted, and results were shown through a Venn diagram (Bardou et al., 2014).

Virulence Assay in Tomato Plants

C. michiganensis strain LMG7333^T was grown at 26°C in KB medium (20 g proteose peptone No. 3, 1.5 g K_2HPO_4, 6 mL 1 M $MgSO_4$, 16 mL 50% glycerol, and 15 g agar in the total volume of 1 L) for 3 days and inoculated into tomato plants using three different inoculation methods, which are leaf-clipping inoculation, stem injection, and root-dipping inoculation. Ten-day-old and 3- and 4-week-old tomato plants (*Solanum lycopersicum* L. cv. "Betatini") were used for the root-dipping inoculation, leaf-clipping inoculation, and stem injection, respectively. For the leaf-clipping method, the bacterial inoculum was prepared and adjusted to a concentration of 2×10^8 CFU/mL (OD_{600} = 0.4). Sterilized scissors were dipped into this inoculum, and then half of the leaves were cut using scissors (Hwang et al., 2020b). Three leaves located in different directions of each plant and at least five plants were used for each treatment. For stem injection, a bacterial concentration was adjusted to 10^9 CFU/mL (OD_{600} = 2.0), and the 20 μL bacterial suspension was injected into tomato stems above 2 cm from the cotyledon after making a hole with a needle. At least four plants were used for each treatment. For the root-dipping inoculation method, 10-day-old seedlings were pulled out from soil, and their roots were dipped in 1 mL of bacterial suspension (10^9 CFU/mL) for 30 min. The seedlings were then replanted into pots. Five plants were used for each treatment. *C. michiganensis* LMG7333^T was resuspended in sterilized 10 mM $MgCl_2$ buffer, and this buffer was used as a negative control in all inoculation methods.

Bacterial Growth Measurement *in planta*

After inoculation with *C. michiganensis* LMG7333^T by stem injection, tomato stems were collected at 0, 7, 14, 21, and 28 days after inoculation (dai). For sample collection, 3 cm lengths of inoculated stems were harvested from four different sites, which are the inoculated site and 10, 15, and 20 cm above from the inoculated site. Then, 2 mL of 10 mM $MgCl_2$ was added to ground samples and the supernatant was dotted onto KB media after 10-fold serial dilution. Three plants were used for each repetition, and experiments were repeated at least three times.

Statistical Analysis

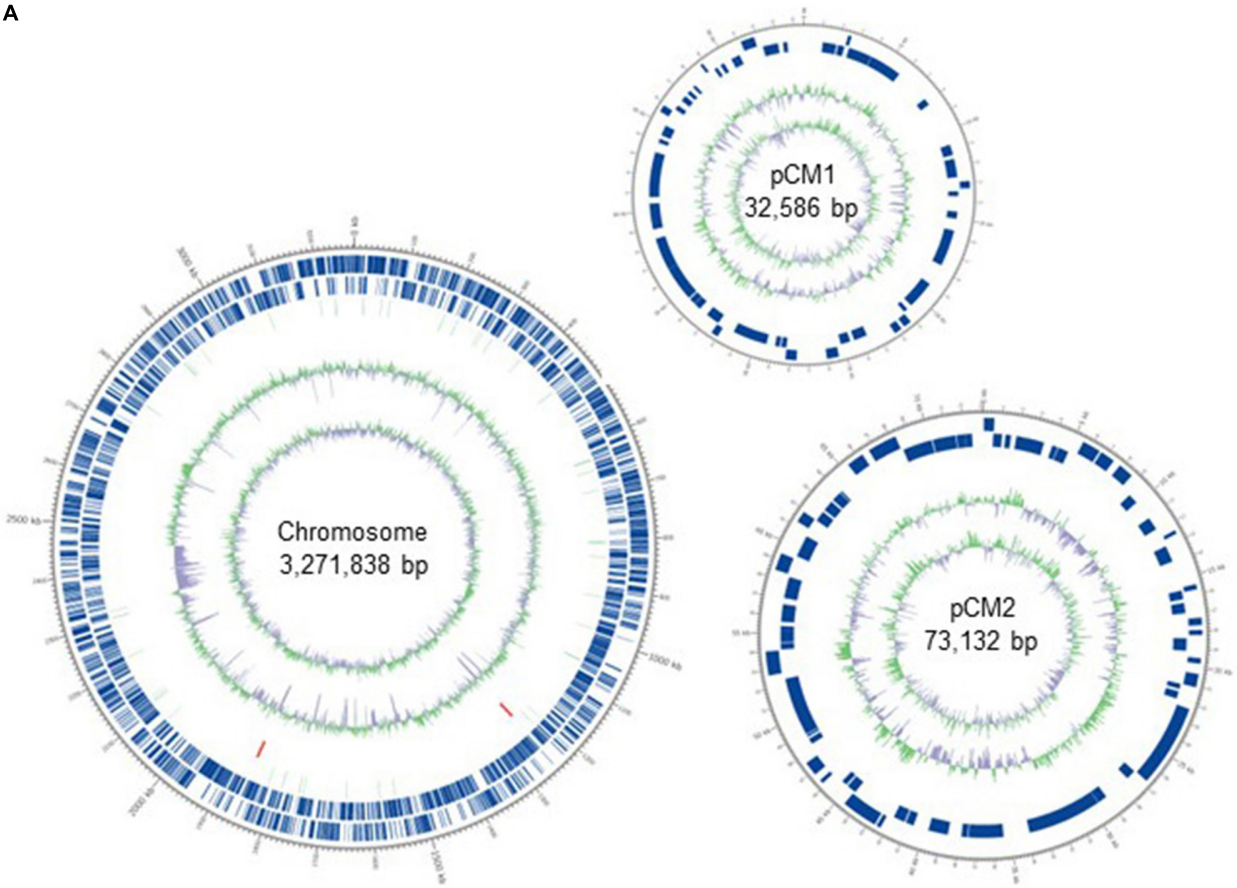
For statistical analysis of bacterial growth of *Clavibacter michiganensis* LMG7333^T *in planta*, Duncan's multiple range test was conducted ($p < 0.05$).

RESULTS

Complete Genome Information of *Clavibacter michiganensis* LMG7333^T

The complete genome sequence of *C. michiganensis* LMG7333^T was obtained through the combination of PacBio RSII and Illumina HiSeqXten sequencing methods, and the genome coverage was 320X. From HiSeqXten, we got 9,154,590 reads, and after assembly, the N50 value was 3,271,838. Finally, we obtained three contigs, one circular chromosome (GenBank accession #: CP080437), and two plasmids, pCM1 and pCM2 (GenBank accession #: CP080438 and CP080439). Their genome sizes and G + C contents are 3.27 Mb with 72.74% G + C content, 32 kb with 67.75% G + C content, and 73 kb with 66.99% G + C content, respectively (**Figures 1A,B**). This genomic composition of *C. michiganensis* LMG7333^T is similar to those of NCPPB382 and UF1 strains, of which the complete genome sequences are

A



B

Strain	LMG7333^T			NCPPB382			UF1		
Features	Chr	pCM1	pCM2	Chr	pCM1	pCM2	Chr	pCM1	pCM2
Genome size (bp)	3,271,838	32,586	73,132	3,297,891	27,357	69,989	3,271,848	20,560	73,141
GC content (%)	72.74	67.75	66.99	72.66	67.56	66.50	72.74	67.95	66.99
No. of CDS	3,033	32	67	2,984	28	68	3,023	20	72
No. of tRNA	45	0	0	45	0	0	45	0	0
No. of rRNA	6	0	0	2	0	0	6	0	0

FIGURE 1 | Genome information of *C. michiganensis* type strain LMG7333^T. **(A)** A circular map drawn by Circos program. Circles outside to the center are represented as CDS on the forward strand, CDS on the reverse strand, tRNA, rRNA, GC content, and GC skew. **(B)** Genome summary of three *C. michiganensis* strains, LMG7333^T, NCPPB382, and UF1. Chr, chromosome; pCM1 and pCM2, plasmids; CDS, coding sequences.

available. However, their sizes are different from one another, particularly, plasmid sizes (**Figure 1B**). Genome annotation revealed 3,033 protein-coding sequences (CDSs) as well as 45 tRNAs and 6 rRNAs in the chromosome and 32 and 67 CDSs in pCM1 and pCM2, respectively (**Figure 1B**). The annotated proteins were categorized into 21 COG. Approximately 40% of CDSs were categorized into two groups for the function unknown (S; 891 CDSs) and the general function only (R; 349 CDSs), which are poorly characterized, in the COG database (**Figure 2**). COG groups for carbohydrate transport and metabolism (G; 255 CDSs); transcription (K; 205 CDSs); amino acid transport and metabolism (E; 198 CDSs); inorganic ion transport and metabolism (P; 179 CDSs); translation, ribosomal structure, and biogenesis (J; 146 CDSs); and cell wall/membrane/envelope biogenesis (M; 137 CDSs) were found to be dominant groups. In contrast, only two CDSs were categorized into cell motility groups (N) (**Figure 2**). Some CDSs were not grouped into specific COG or grouped to two or more COG groups. There are no CDSs

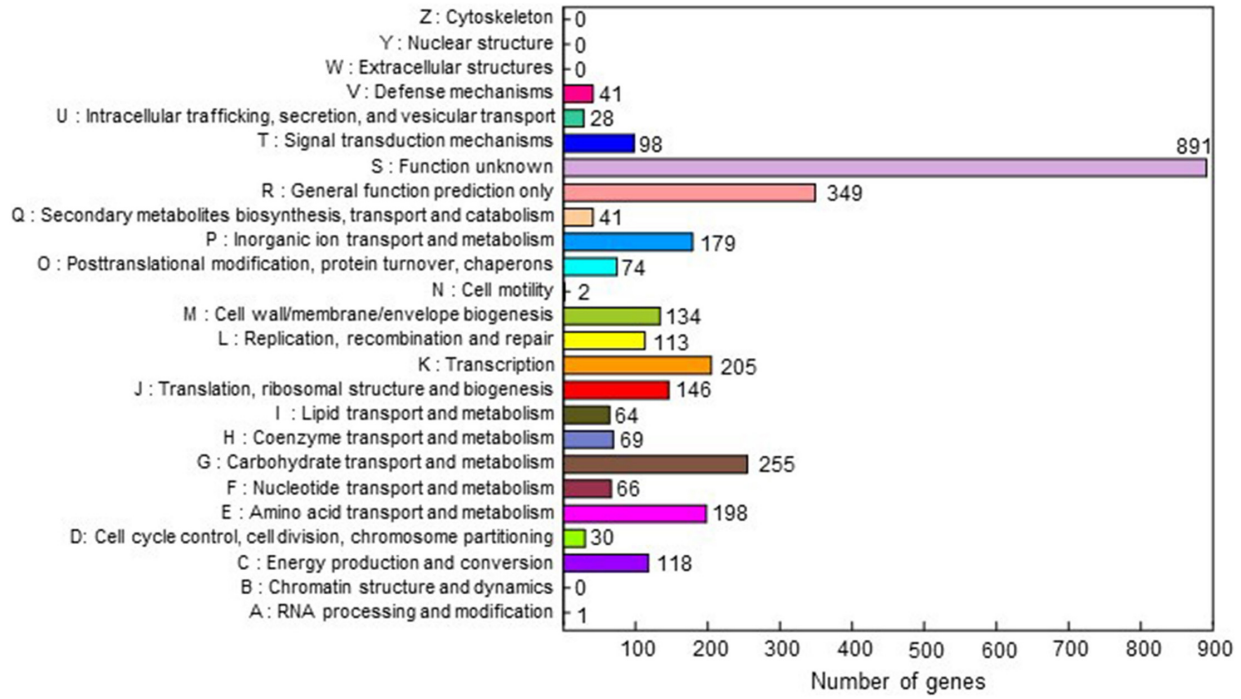


FIGURE 2 | Clusters of orthologous groups (COG) of the annotated CDSs of *C. michiganensis* type strain LMG7333^T. COG database v2014 was used for analysis. Annotated genes were classified into 21 categories of 25. The *X*-axis represents the number of genes, and the *Y*-axis indicates each COG cluster. The numbers next to each bar indicate the number of genes in each COG cluster.

for flagella formation, which is consistent with a previous study reporting that *C. michiganensis* LMG7333^T did not have flagella (Oh et al., 2016).

Genome Comparison Among Complete Genome Sequences of Three *Clavibacter michiganensis* Strains

As shown in **Figure 1B**, three *C. michiganensis* strains differed in the sizes of a chromosome and two plasmids. Here, three genome sequences were compared to examine how different they were. First, the sequence identity of chromosomal DNA among the three strains was almost 100% identical although the query cover between LMG7333^T and NCPPB382 strains was 97% (**Figure 3**). As with other *C. michiganensis* strains, *C. michiganensis* LMG7333^T has about 120 kb of pathogenicity island (PAI) region from 2,334,274 to 2,464,147 bp on its chromosome (**Figure 3** and **Supplementary Figure 1**). In the *chp* region of PAI, the putative virulence genes, such as the *chp*, *ppa*, and *pel* gene families, were intact like those in the NCPPB382 strain. Moreover, the *tomA* gene was found in the *tomA* region of PAI (**Supplementary Figure 1**). Two other important virulence genes, *celA*, and *pat-1* genes are located from 7,321 to 9,561 bp in pCM1 and from 7,833 to 8,675 bp in pCM2 of *C. michiganensis* LMG7333^T (**Figure 3**). The presence of an intact PAI region, *celA* and *pat-1* genes, suggests that *C. michiganensis* LMG7333^T can also infect tomato plants to cause typical disease symptoms like the NCPPB382 stain.

The size and sequence query cover of the pCM1 plasmid of LMG7333^T was the most different from those of two other strains although the sequence identity in the matched regions was almost 100% identical. Its size was the biggest among the three strains, and the matched regions were only 63% with UF1 pCM1 and only 39% with NCPPB382 pCM1 (**Figure 3**). Among unmatched regions in pCM1 of LMG7333^T, some regions were unique to this strain, whereas others were only matched with that of NCPPB382. Most of the genes in these unmatched regions are annotated as hypothetical proteins, indicating that the origin and functions of these genes are unknown. In the public database, *C. michiganensis* NCPPB382 pCM1 has genes for cellulase (*celA*), portioning protein (*parA* and *parB*), and replication origin (*repA*) and several hypothetical proteins. When the sequence of pCM1 was compared with these annotated genes of pCM1 in *C. michiganensis* NCPPB382, we found that *C. michiganensis* LMG7333^T pCM1 also contained these annotated genes except the *traG* gene encoding putative conjugal transfer protein (Załuga et al., 2014). Gene annotation also revealed that the *espI* gene, which is a component of the type VII ESX-1 secretion system (Zhang et al., 2014), was found in the pCM1 of LMG7333^T but not NCPPB382, whereas both pCM2 plasmids have the *espI* homologous gene.

The size and sequence of LMG7333^T pCM2 plasmid were almost 100% identical to that of the UF1 strain, whereas the matched regions with NCPPB382 pCM2 were only approximately 37 kb of 73 kb, which is only 55% of the total plasmid size with 91–99% sequence identity (**Figure 3**). In the

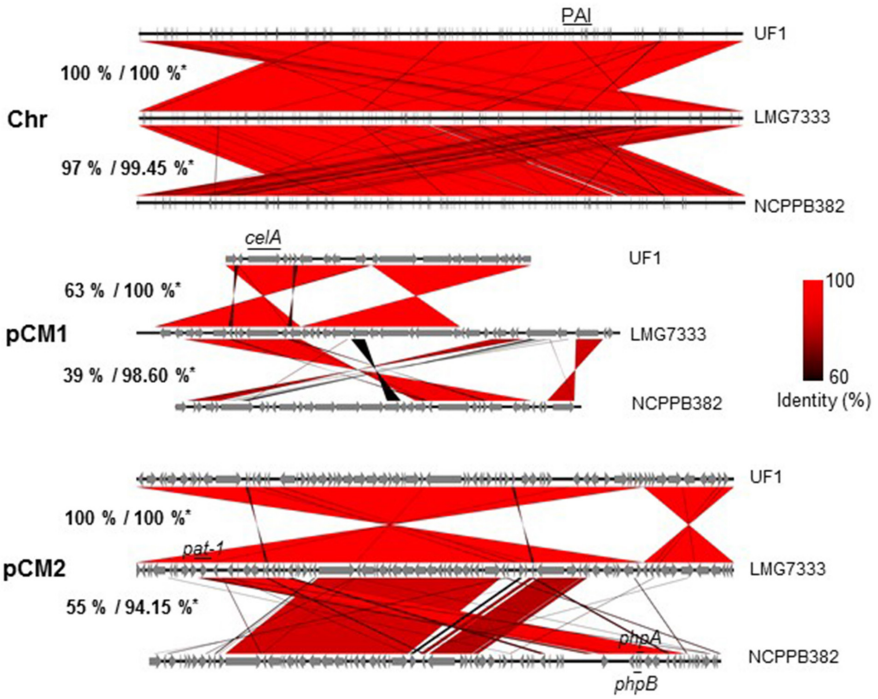


FIGURE 3 | Genome comparison of three *C. michiganensis* strains. Genome alignment of three *C. michiganensis* strains. *Percentages indicate represented query cover/identity of a chromosome (Chr) and two plasmids, pCM1 and pCM2, of *C. michiganensis* NCPPB382 and UF1 based on *C. michiganensis* LMG7333^T. The identical parts were aligned in red, and the key virulence genes such as *celA* and *pat-1* genes or a pathogenicity island (PAI) are shown.

matched region of pCM2 plasmids, all three *C. michiganensis* strains possess *pat-1* genes, encoding a putative serine protease, which was known as a plasmid-borne virulence factor of *C. michiganensis* (Burger et al., 2005). In the case of *C. michiganensis* NCPPB382, there are more *pat-1* paralogs, *phpA* and *phpB,* in pCM2. However, *C. michiganensis* LMG7333^T lacks these two genes. The unmatched region covered by 45% of the total plasmid size was unique to each strain.

Overall, these results indicate that, although three *C. michiganensis* strains have similar genome size and major virulence factors, like PAI, *celA*, and *pat-1*, in their genomes, the plasmid sequences are varied, and both pCM1 and pCM2 of LMG7333^T contain unique DNA sequences mostly annotated as hypothetical proteins. Moreover, genome comparison among three *C. michiganensis* strains using complete genome sequences indicates that *C. michiganensis* LMG7333^T genome sequence can be used as another reference genome of *C. michiganensis*.

Genome Comparison of *Clavibacter michiganensis* LMG7333^T With Other *Clavibacter* Species

Based on genomic information, average nucleotide identity (ANI) values were calculated among *C. michiganensis* LMG7333^T and nine other *Clavibacter* bacteria, including type strains of other *Clavibacter* species and *Leifsonia xyli* subsp. *cydontis* as an outgroup to measure the genetic distance among genomes. Based on ANI data, three *C. michiganensis* strains were clustered into the same clade with more than 99% ANI value, and non-pathogenic *Clavibacter californiensis* was the closest to *C. michiganensis* with almost 95% ANI value (**Figure 4** and **Supplementary Table 1**). There was the second clade, including *C. insidiosus*, *C. nebraskensis*, *C. phaseoli*, and *C. michiganensis* subsp. *chilensis*, which is close to the clade of *C. michiganensis*. Type strains of *C. capsici* PF008^T, *C. sepedonicus* ATCC33113^T, and *C. tessellarius* ATCC33566^T are phylogenetically distinct from the *C. michiganensis* clade.

Using genome sequences of pathogenic *Clavibacter* strains, pan-genome analysis was also conducted. Three *C. michiganensis* strains were first compared, and they have 2,959 common genes. *C. michiganensis* LMG7333^T only has seven unique genes in its genome. Moreover, it has 8 and 84 genes overlapped with *C. michiganensis* NCPPB382 and UF1, respectively (**Figure 5A**). Among the seven unique genes in *C. michiganensis* LMG7333^T, only one gene was annotated as RNA helicase, and the other six genes were annotated as hypothetical proteins. The specific genes in other two *C. michiganensis* strains were also annotated as hypothetical proteins. *C. michiganensis* LMG7333^T has 91 unique genes compared with *C. michiganensis* NCPPB382. Among these genes, 50 genes are located in the chromosome of the type strain, and the other 41 genes exist in two plasmids, 17 and 24 genes in pCM1 and pCM2, respectively. Among these 91 genes, only 17 genes were annotated (**Supplementary Table 2**).

BLAST analysis with 91 genes unique to *C. michiganensis* LMG7333^T compared with *C. michiganensis* NCPPB382 was conducted to identify the possible origin of these genes. Based

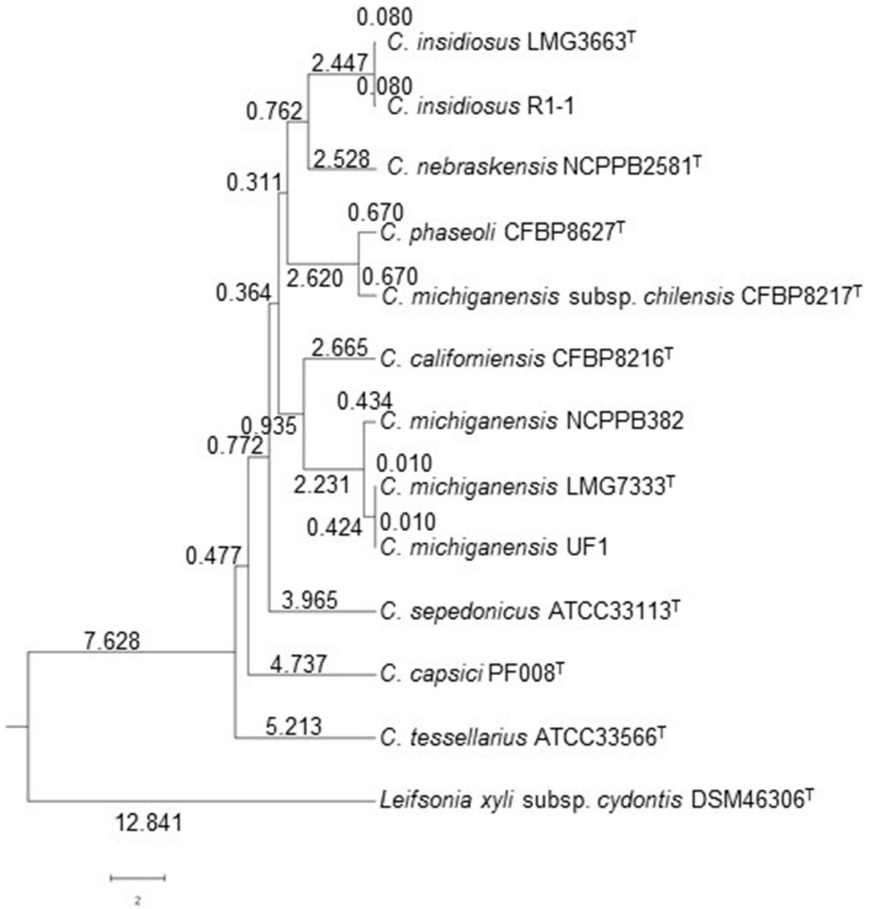


FIGURE 4 | Phylogenetic tree of *C. michiganensis* type strain LMG7333^T with other *Clavibacter* species and subspecies based on ANI. This analysis was conducted using OrthoANI software. *Leifsonia xyli* subsp. *cydontis* DSM46306^T was used for outgroup. Numbers indicate genetic similarity (GS) coefficient value. T, type strain.

on the analysis results, these 91 genes exhibited significant homology with diverse organisms, including other *Clavibacter* species (**Supplementary Table 2**). One gene exists in only *C. michiganensis* LMG7333^T, and nine genes are found in only two *C. michiganensis* strains, LMG7333^T and UF1. Twenty-seven genes were shared with other *C. michiganensis* strains, MSF322 and VL527 (Méndez et al., 2020). Twenty-two genes are shared with other *Clavibacter* bacteria, such as *C. insidiosus* (10), *C. capsici* (7), *C. nebraskensis* (4), and *C. sepedonicus* (1). The remaining 32 genes are shared with diverse organisms, 11 families, and 18 genera. Most of these genes are shared with bacteria, including four families, *Actinomycetaceae* (1), *Cellulomonadaceae* (1), *Microbacteriaceae* (18), and *Micrococcaceae* (2), which belong to the phylum *Actinobacteria*. In particular, the 18 genes shared with the *Microbacteriaceae* family, including *Clavibacter* genus, were shared with six genera, *Microbacterium* (5), *Cryobacterium* (1), *Curtobacterium* (8), *Leifsonia* (2), *Leucobacter* (1), and *Rathayibacter* (1). Finally, among 91 *C. michiganensis* LMG7333^T genes, 77 genes are found only in the *Microbacteriaceae* family. These results indicate that the *C. michiganensis* LMG7333^T might obtain most foreign genes from the same bacterial family.

The *C. michiganensis* LMG7333^T genome was also compared with those of four other *Clavibacter* type strains. Type strains of five *Clavibacter* species have 2,345 core genes. Interestingly, each *Clavibacter* has 138–316 specific genes in their genomes (**Figure 5B**). Most of these genes were annotated as hypothetical proteins and categorized mostly into unknown function genes, carbohydrate transport, and metabolism as well as transcription proteins in the COG database (**Supplementary Table 3**). When pan-genome analysis was conducted, duplicated genes were regarded as one gene. Therefore, the total CDS number (**Figure 1B**) and CDS number of pan-genome analysis (**Figure 5B**) were slightly different.

Disease Development in Tomato Plants by *Clavibacter michiganensis* LMG7333^T

In our previous study, *C. michiganensis* LMG7333^T and its mutant strain were used to investigate the gene function involved in virulence (Hwang et al., 2019). For the virulence assay or mutant screening, the root-dipping method was the best way to observe typical wilting symptoms within a short period. Here, we examined the virulence of *C. michiganensis*

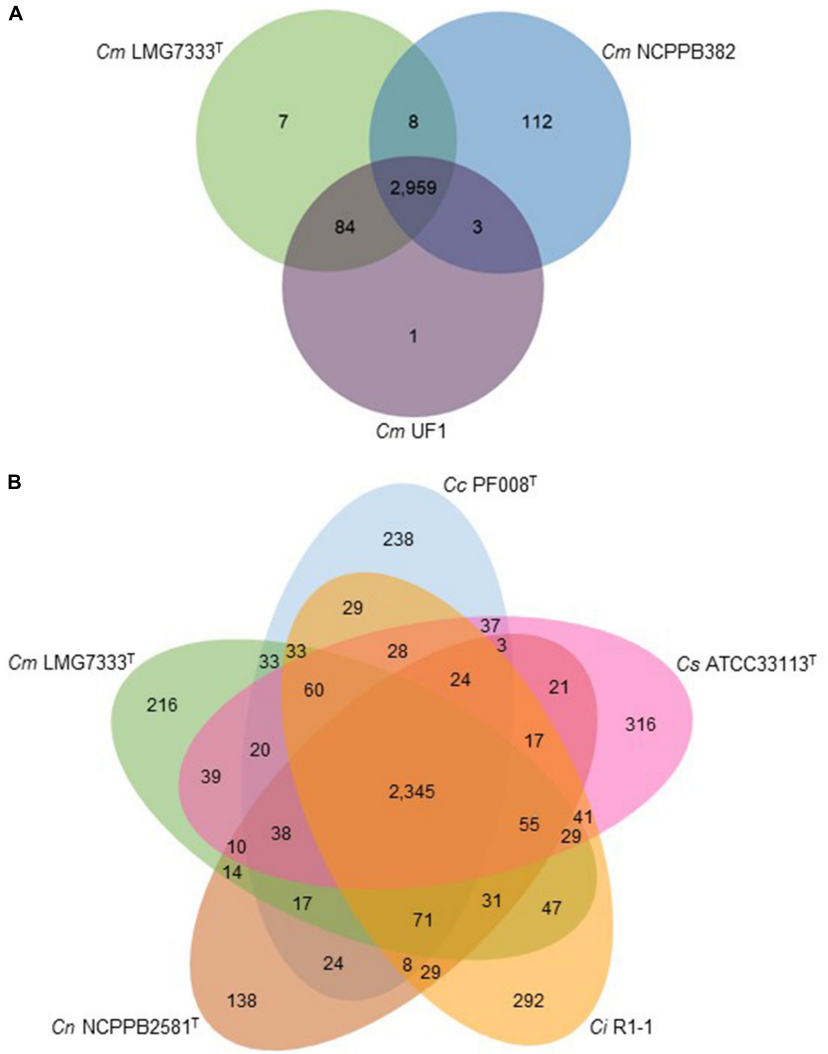


FIGURE 5 | Comparative genome analysis of *C. michiganensis* LMG7333^T and other *Clavibacter* strains. **(A)** Venn diagram of three *C. michiganensis* strains. **(B)** Venn diagram of *C. michiganensis* LMG7333^T and four other type strains of *Clavibacter* species. Numbers represent gene numbers, which are common or unique to each strain. *Cm*, *C. michiganensis*; *Cc*, *C. capsici*; *Cs*, *C. sepedonicus*; *Ci*, *C. insidiosus*; *Cn*, *C. nebraskensis*.

LMG7333^T on tomato plants in various ways. Tomato leaves infected with *C. michiganensis* LMG7333^T using the leaf-clipping method exhibited wilting symptoms from the cut part of leaves (**Figure 6A**). As bacteria are directly inoculated into the plants, it was successfully spread along the leaf veins and induced wilting symptoms on the inoculated leaves. Eventually, this bacterium spread throughout the plant and caused typical wilting symptoms even on non-inoculated leaves and branches (**Figure 6A**). In contrast, no disease symptoms and no physiological disorders were found in buffer-treated tomato plants.

Stem inoculation after artificial wounding, which is a conventional method for studying *C. michiganensis*, induced large canker lesions at the site of inoculation (**Figure 6A**). The inoculated bacteria were gradually spread along the xylem toward the tip, and approximately from 21 dai, leaves on the tomato tip exhibited wilting symptoms (**Figure 6B**). Movement of *C. michiganensis* LMG7333^T within the stem was confirmed by bacterial growth *in planta*. From 7 dai, bacteria were detected at 10 cm above the inoculation site, and from 14 dai, it was isolated at the topper site of the stem, indicating that *C. michiganensis* LMG7333^T could move and colonize in areas higher than the inoculated site (**Figure 6B**). Altogether, typical bacterial cankers at the site of injection as well as wilting symptoms at the branch and leaves were observed in stem-inoculated tomato plants.

Disease symptoms through this bacterial infection were also caused by the root-dipping methods (**Figure 6A**). Although this method has not been widely used, disease symptoms were shown within 10 days by the root-dipping method in 8–10-day-old tomato plants. Because this pathogen can enter the tomato plant through its roots during natural infection, the virulence or pathogenicity assay in tomato seedlings by root infection may be more significant than the stem inoculation (**Figure 6A**).

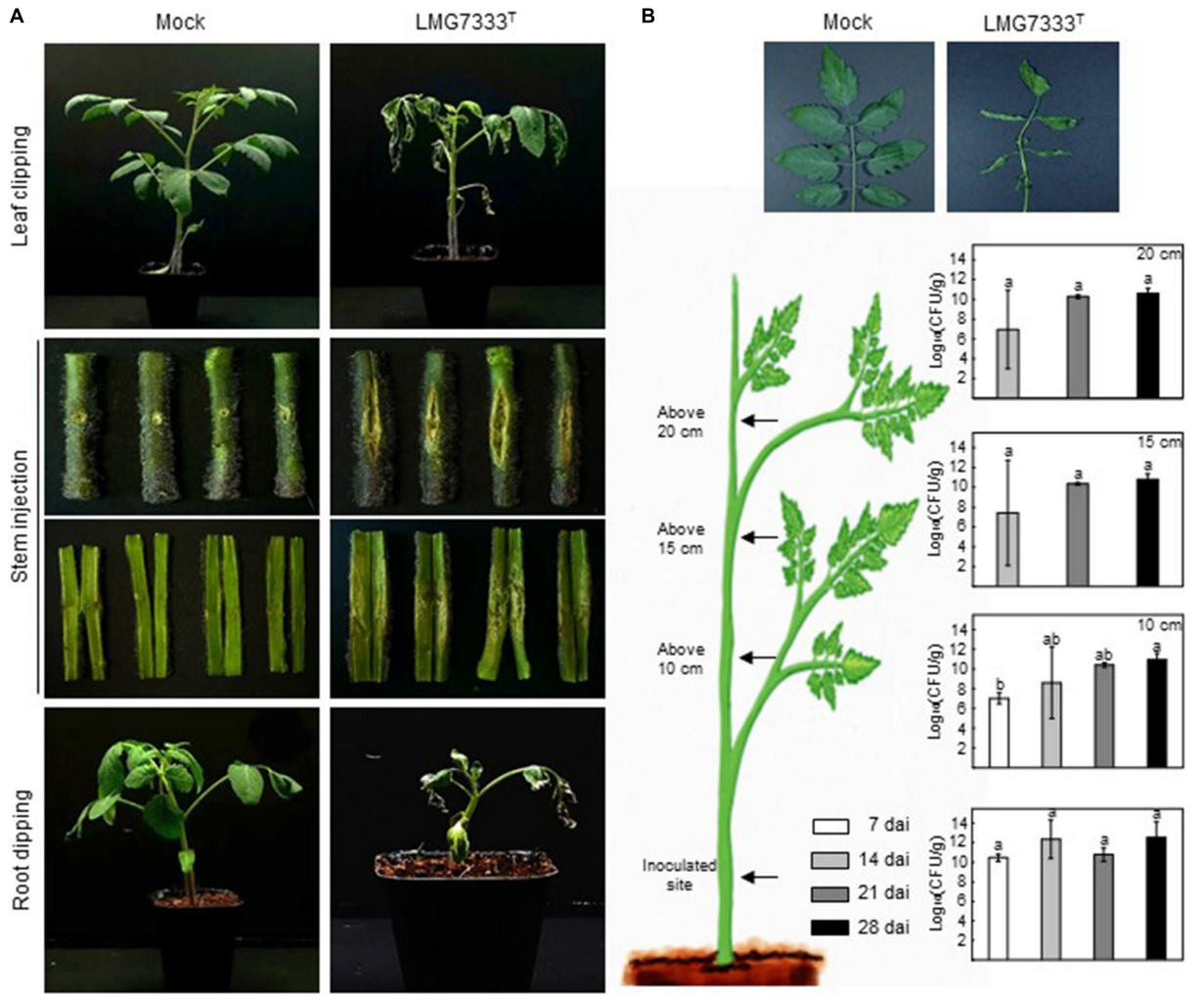


FIGURE 6 | Virulence of *C. michiganensis* LMG7333^T on tomato plants. **(A)** Typical disease symptoms of *C. michiganensis* LMG7333^T, according to inoculation methods. Photos were taken at 30 days after inoculation (dai) for leaf clipping and stem inoculation and for root-dipping inoculation at 10 dai. **(B)** Bacterial numbers after stem inoculation at the indicated time points. The graphs from bottom to top indicate cfu/g at the inoculation site, 10, 15, and 20 cm above site from the inoculation point. Error bars indicate standard deviation ($n = 3$). The different letters indicate statistically significant differences as determined by Duncan's multiple range test ($p < 0.05$).

The observation that *C. michiganensis* LMG7333^T infection in tomato plants resulted in typical disease symptoms by three different inoculation methods indicates that this pathogen with the completed whole genome sequence might be valuable as a standard strain for *C. michiganensis* study and its interaction with the plants.

DISCUSSION

So far, genome sequences of many *Clavibacter* bacteria have been reported and used for understanding these bacteria at the genomic level. For *C. michiganensis*, the complete whole genome sequence of *C. michiganensis* strain NCPPB382 was first reported (Gartemann et al., 2008), and this genomic information has provided the opportunity to reveal many putative virulence genes in a PAI region, such as the *chp*, *ppa*, and *pel* gene families, and also the presence of two plasmids, pCM1 and pCM2. Since the NCPPB382 genome sequence became available, many draft genome sequences of *C. michiganensis* strains have been reported (Thapa et al., 2017; Méndez et al., 2020). Genome sequencing of *C. michiganensis* type strain *C. michiganensis* LMG7333^T was previously conducted with PacBio-SMRT sequencing technology, and its draft genome sequence was reported (Li and Yuan, 2017). However, this report showed five contigs. Because this genome sequence was not complete, there was a limitation in using it as a reference genome of *C. michiganensis* species to understand the genetic diversity in either chromosome or plasmids among strains of this species. In this study, for completion of genome sequencing of type strain LMG7333^T, we combined two sequencing platforms, PacBio RSII and Illumina HiSeqXten, and could successfully obtain the complete genome

sequence with three contigs, one chromosome and two plasmids (**Figure 1**). This could make it possible to perform comparative genomic analyses among strains within *C. michiganensis* species and among other *Clavibacter* species.

Comparative genomic analyses revealed that type strain LMG7333^T has a genome composition similar to NCPPB382, one chromosome and two plasmids. Overall G + C content and the presence of a PAI region and major virulence genes such as *celA* and *pat-1* were also similar. However, the plasmid size and gene contents were different between the two strains: only 39 and 55% of plasmid DNA of pCM1 and pCM2, respectively, were identical. These findings indicate that high genetic diversity is more present in plasmids than in a chromosome (almost 100% identical between two strains). Moreover, some genes, such as *phpA* and *phpB* genes in pCM2 of NCPPB382, which are *pat-1* paralogs, were missing in pCM2 of type strains LMG7333^T (**Figure 3**), indicating that these two strains might use different virulence factors interaction with host plants.

Among *C. michiganensis*, of which their genome sequences are available in the public genome database, the UF1 strain was most similar to *C. michiganensis* LMG7333^T (**Figures 1**, **3**). Only 4 and 15 genes within the whole genome were specific to the UF1 strain and a type strain LMG7333^T, respectively (**Figure 5**). Based on the genome comparison, the UF1 strain might be pathogenic although its pathogenicity has not been reported. Moreover, the genetic diversity between type strain LMG7333^T and NCPPB382 was bigger because 91 and 115 genes were specific to each strain, respectively. More analysis of 91 type strain LMG7333^T-specific genes showed that most of them encode hypothetical proteins, and their homologous are primarily present in bacteria belonging to *Microbacteriaceae* family, including the genus *Clavibacter*. This might indicate that gene exchange or transfer among bacteria in this family occurs.

As expected, the genetic difference among *Clavibacter* species was bigger than that of strains within the single species, *C. michiganensis* (**Figure 5**). The number of core genes among five *Clavibacter* species is 2,345, whereas its number among strains of *C. michiganensis* is 2,959. Moreover, in each species, the number of unique genes ranged from 138 to 316. Notably, among *C. michiganensis*-specific genes, *tomA* encodes tomatinase. Although the function of these genes is shown to be critical to *C. michiganensis* virulence (Kaup et al., 2005), this gene has been defined as a virulence factor in *C. michiganensis*. Seipke and Loria (2008) show that *Streptomyces scabies* have tomatinase, and it is related to plant-microbe interactions rather than the pathogenicity of bacterial pathogen, implying that *tomA* in *C. michiganensis* might affect the interaction with host plants. There are many genes specific to each *Clavibacter* species (**Figure 5B** and **Supplementary Table 3**). However, many of them encode hypothetical proteins with unknown functions. The function of these genes remains determined.

Infection with *C. michiganensis* LMG7333^T induced typical disease symptoms, canker, and wilting on tomato plants (**Figure 6**) as observed on tomato plants infected with other pathogenic *C. michiganensis*. This means that type strain LMG7333^T is also fully pathogenic, and it can be used to understand *C. michiganensis*–host plant interactions. Furthermore, the availability of the complete genome information of type strain LMG7333^T would make this strain more useful for studies to reveal virulence genes and understand virulence mechanisms against host plants.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

E-JO, IP, IH, and C-SO: investigation, writing—review and editing. E-JO, IP, and IH: methodology and formal analysis. C-SO: validation, project administration, and funding acquisition. All authors have read and agreed to the submitted version of the manuscript.

FUNDING

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (2019R1A2C2004568).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.793345/full#supplementary-material

REFERENCES

Bae, C., Oh, E.-J., Lee, H.-B., Kim, B.-Y., and Oh, C.-S. (2015). Complete genome sequence of the cellulase-producing bacterium *Clavibacter michiganensis* PF008. *J. Biotechnol.* 214, 103–104. doi: 10.1016/j.jbiotec.2015.09.021

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15:293. doi: 10.1186/1471-2105-15-293

Bentley, S. D., Corton, C., Brown, S. E., Barron, A., Clark, L., Doggett, J., et al. (2008). Genome of the actinomycete plant pathogen *Clavibacter michiganensis* subsp. *sepedonicus* suggests recent niche adaptation. *J. Bacteriol.* 190, 2150–2160. doi: 10.1128/JB.01598-07

Burger, A., Gräfen, I., Engemann, J., Niermann, E., Pieper, M., Kirchner, O., et al. (2005). Identification of homologues to the pathogenicity factor Pat-1, a putative serine protease of *Clavibacter michiganensis* subsp. *michiganensis*. *Microbiol. Res.* 160, 417–427. doi: 10.1016/j.micres.2005.03.006

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

Carlson, R. R., and Vidaver, A. K. (1982). Taxonomy of *Corynebacterium* plant pathogens, including a new pathogen of wheat, based on polyacrylamide gel electrophoresis of cellular proteins. *Int. J. Syst. Evol. Microbiol.* 32, 315–326. doi: 10.1099/00207713-32-3-315

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Davis, M. J., Gillaspie, A. G. Jr., Vidaver, A. K., and Harris, R. W. (1984). *Clavibacter*: a new genus containing some phytopathogenic coryneform bacteria, including *Clavibacter xyli* subsp. *xyli* sp. nov., subsp. nov. and *Clavibacter xyli* subsp. *cynodontis* subsp. nov., pathogens that cause ratoon stunting disease of sugarcane and bermudagrass stunting disease. *Int. J. Syst. Evol. Microbiol.* 34, 107–117. doi: 10.1099/00207713-34-2-107

Dreier, J., Meletzus, D., and Eichenlaub, R. (1997). Characterization of the plasmid encoded virulence region *pat-1* of phytopathogenic *Clavibacter michiganensis* subsp. *michiganensis*. *Mol. Plant Microbe Interact.* 10, 195–206. doi: 10.1094/MPMI.1997.10.2.195

Eichenlaub, R., and Gartemann, K.-H. (2011). The *Clavibacter michiganensis* subspecies: molecular investigation of gram-positive bacterial plant pathogens. *Annu. Rev. Phytopathol.* 49, 445–464. doi: 10.1146/annurev-phyto-072910-095258

Eichenlaub, R., Gartemann, K.-H., and Burger, A. (2007). "*Clavibacter michiganensis*, a group of Gram-positive phytopathogenic bacteria," in *Plant-associated bacteria*, ed. S. S. Gnanamanickam (Dordrecht: Springer), 385–421. doi: 10.1007/978-1-4020-4538-7_12

Gartemann, K.-H., Abt, B., Bekel, T., Burger, A., Engemann, J., Flügel, M., et al. (2008). The genome sequence of the tomato-pathogenic actinomycete *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382 reveals a large island involved in pathogenicity. *J. Bacteriol.* 190, 2138–2149. doi: 10.1128/JB.01595-07

Gonzalez, A. J., and Trapiello, E. (2014). *Clavibacter michiganensis* subsp. *phaseoli* subsp. nov., pathogenic in bean. *Int. J. Syst. Evol. Microbiol.* 64, 1752–1755. doi: 10.1099/ijs.0.058099-0

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248

Hwang, I. S., Lee, H. M., Oh, E. J., Lee, S., Heu, S., and Oh, C. S. (2020a). Plasmid composition and the *chpG* gene determine the virulence level of *Clavibacter capsici* natural isolates in pepper. *Mol. Plant Pathol.* 21, 808–819. doi: 10.1111/mpp.12932

Hwang, I. S., Oh, E.-J., and Oh, C.-S. (2020b). Transcriptional Changes of Plant Defense-Related Genes in Response to *Clavibacter* Infection in Pepper and Tomato. *Plant Pathol. J.* 36, 450–458. doi: 10.5423/PPJ.OA.07.2020.0124

Hwang, I. S., Oh, E.-J., Lee, H. B., and Oh, C.-S. (2019). Functional characterization of two cellulase genes in the Gram-positive pathogenic bacterium *Clavibacter michiganensis* for wilting in tomato. *Mol. Plant Microbe Interact.* 32, 491–501. doi: 10.1094/MPMI-08-18-0227-R

Jahr, H., Dreier, J., Meletzus, D., Bahro, R., and Eichenlaub, R. (2000). The endo-β-1, 4-glucanase CelA of *Clavibacter michiganensis* subsp. *michiganensis* is a pathogenicity determinant required for induction of bacterial wilt of tomato. *Mol. Plant Microbe Interact.* 13, 703–714. doi: 10.1094/MPMI.2000.13.7.703

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kaup, O., Gräfen, I., Zellermann, E.-M., Eichenlaub, R., and Gartemann, K.-H. (2005). Identification of a tomatinase in the tomato-pathogenic actinomycete *Clavibacter michiganensis* subsp. *michiganensis* NCPPB382. *Mol. Plant Microbe Interact.* 18, 1090–1098. doi: 10.1094/MPMI-18-1090

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Lee, I., Kim, Y. O., Park, S.-C., and Chun, J. (2016). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* 66, 1100–1103. doi: 10.1099/ijsem.0.000760

Li, X., Tambong, J., Yuan, K. X., Chen, W., Xu, H., Lévesque, C. A., et al. (2018). Re-classification of *Clavibacter michiganensis* subspecies on the basis of whole-genome and multi-locus sequence analyses. *Int. J. Syst. Evol. Microbiol.* 68, 234–240. doi: 10.1099/ijsem.0.002492

Li, X. S., and Yuan, X. K. (2017). Genome sequences for multiple *Clavibacter* strains from different subspecies. *Genome Announc.* 5, e00721–17. doi: 10.1128/genomeA.00721-17

Lu, Y., Samac, D. A., Glazebrook, J., and Ishimaru, C. A. (2015). Complete genome sequence of *Clavibacter michiganensis* subsp. *insidiosus* R1-1 using PacBio single-molecule real-time technology. *Genome Announc.* 3, e00396–15. doi: 10.1128/genomeA.00396-15

Manzer, F., and Genereux, H. (1981). "Ring rot," in *Compendium of potato disease*, ed. W. J. Hooker (Saint Paul, MN: American Phytopathological Society Press), 31–32.

McCulloch, L. (1925). *Aplanobacter insidiosum* n. sp., the cause of an Alfalfa disease. *Phytopathology* 15, 496–497.

Méndez, V., Valenzuela, M., Salvà-Serra, F., Jaén-Luchoro, D., Besoain, X., Moore, E. R., et al. (2020). Comparative genomics of pathogenic *Clavibacter michiganensis* subsp. *michiganensis* strains from Chile reveals potential virulence features for tomato plants. *Microorganisms* 8:1679. doi: 10.3390/microorganisms8111679

Monteiro-Vitorello, C. B., Zerillo, M. M., Van Sluys, M.-A., Camargo, L. E. A., and Kitajima, J. P. (2013). Complete genome sequence of *Leifsonia xyli* subsp. *cynodontis* strain DSM46306, a gram-positive bacterial pathogen of grasses. *Genome Announc.* 1, e00915–13. doi: 10.1128/genomeA.00915-13

Oh, E.-J., Bae, C., Lee, H.-B., Hwang, I. S., Lee, H.-I., Yea, M. C., et al. (2016). *Clavibacter michiganensis* subsp. *capsici* subsp. nov., causing bacterial canker disease in pepper. *Int. J. Syst. Evol. Microbiol.* 66, 4065–4070. doi: 10.1099/ijsem.0.001311

Osdaghi, E., Rahimi, T., Taghavi, S. M., Ansari, M., Zarei, S., Portier, P., et al. (2020). Comparative genomics and phylogenetic analyses suggest several novel species within the genus *Clavibacter*, including nonpathogenic tomato-associated strains. *Appl. Environ. Microbiol.* 86, e02873–19. doi: 10.1128/AEM.02873-19

Peritore-Galve, F. C., Tancos, M. A., and Smart, C. D. (2020). Bacterial canker of tomato: revisiting a global and economically damaging seedborne pathogen. *Plant Dis.* 105, 1581–1595. doi: 10.1094/PDIS-08-20-1732-FE

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Seipke, R. F., and Loria, R. (2008). *Streptomyces scabies* 87-22 possesses a functional tomatinase. *J. Bacteriol.* 190, 7684–7692. doi: 10.1128/JB.01010-08

Smith, E. (1910). A new tomato disease of economic importance. *Science* 31, 794–796.

Stork, I., Gartemann, K. H., Burger, A., and Eichenlaub, R. (2008). A family of serine proteases of *Clavibacter michiganensis* subsp. *michiganensis*: *chpC* plays a role in colonization of the host plant tomato. *Mol. Plant Pathol.* 9, 599–608. doi: 10.1111/j.1364-3703.2008.00484.x

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Tambong, J. T. (2017). Comparative genomics of *Clavibacter michiganensis* subspecies, pathogens of important agricultural crops. *PLoS One* 12:e0172295. doi: 10.1371/journal.pone.0172295

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33

Thapa, S. P., Pattathil, S., Hahn, M. G., Jacques, M.-A., Gilbertson, R. L., and Coaker, G. (2017). Genomic analysis of *Clavibacter michiganensis* reveals insight into virulence strategies and genetic diversity of a Gram-positive bacterial pathogen. *Mol. Plant Microbe Interact.* 30, 786–802. doi: 10.1094/MPMI-06-17-0146-R

Van der Wolf, J., Elphinstone, J., Stead, D., Metzler, M., Müller, P., Hukkanen, A., et al. (2005). *Epidemiology of Clavibacter michiganensis subsp. sepedonicus in relation to control of bacterial ring rot. Report no. 95*. Wageningen: Plant Research International,

Vidaver, A. K., and Mandel, M. (1974). *Corynebacterium nebraskense*, a New, Orange-Pigmented Phytopathogenic Species1. *Int. J. Syst. Evol. Microbiol.* 24, 482–485. doi: 10.1099/00207713-24-4-482

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Xu, X., Rajashekara, G., Paul, P. A., and Miller, S. A. (2012). Colonization of tomato seedlings by bioluminescent *Clavibacter michiganensis* subsp. *michiganensis* under different humidity regimes. *Phytopathology* 102, 177–184. doi: 10.1094/PHYTO-03-11-0090

Yasuhara-Bell, J., and Alvarez, A. M. (2015). Seed-associated subspecies of the genus *Clavibacter* are clearly distinguishable from *Clavibacter michiganensis* subsp. *michiganensis*. *Int. J. Syst. Evol. Microbiol.* 65, 811–826. doi: 10.1099/ijs.0.000022

Załuga, J., Stragier, P., Baeyen, S., Haegeman, A., Van Vaerenbergh, J., Maes, M., et al. (2014). Comparative genome analysis of pathogenic and non-pathogenic *Clavibacter* strains reveals adaptations to their lifestyle. *BMC Genomics* 15:392. doi: 10.1186/1471-2164-15-392

Zhang, M., Chen, J. M., Sala, C., Rybniker, J., Dhar, N., and Cole, S. T. (2014). EspI regulates the ESX-1 secretion system in response to ATP levels in *Mycobacterium tuberculosis*. *Mol. Microbiol.* 93, 1057–1065. doi: 10.1111/mmi.12718

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 February 2022
doi: 10.3389/fmicb.2022.826365

Genome Context Influences Evolutionary Flexibility of Nearly Identical Type III Effectors in Two Phytopathogenic Pseudomonads

David A. Baltrus[1,2], Qian Feng[3] and Brian H. Kvitko[4]**

[1] School of Plant Sciences, University of Arizona, Tucson, AZ, United States, [2] School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ, United States, [3] Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA, United States, [4] Department of Plant Pathology, University of Georgia, Athens, GA, United States

Integrative Conjugative Elements (ICEs) are replicons that can insert and excise from chromosomal locations in a site-specific manner, can conjugate across strains, and which often carry a variety of genes useful for bacterial growth and survival under specific conditions. Although ICEs have been identified and vetted within certain clades of the agricultural pathogen *Pseudomonas syringae,* the impact of ICE carriage and transfer across the entire *P. syringae* species complex remains underexplored. Here we identify and vet an ICE (PmaICE-DQ) from *P. syringae* pv. maculicola ES4326, a strain commonly used for laboratory virulence experiments, demonstrate that this element can excise and conjugate across strains, and highlight that this element contains loci encoding multiple type III effector proteins. Moreover, genome context suggests that another ICE (PmaICE-AOAB) is highly similar in comparison with and found immediately adjacent to PmaICE-DQ within the chromosome of strain ES4326, and also contains multiple type III effectors. Lastly, we present passage data from *in planta* experiments that suggests that genomic plasticity associated with ICEs may enable strains to more rapidly lose type III effectors that trigger R-gene mediated resistance in comparison to strains where nearly isogenic effectors are not present in active ICEs. Taken together, our study sheds light on a set of ICE elements from *P. syringae* pv. maculicola ES4326 and suggests how genomic context may lead to different evolutionary dynamics for shared virulence genes between strains.

Keywords: *Pseudomonas syringae*, integrative conjugative element (ICE), type III effector, phytopathogen, *Nicotiana benthamiana*

OPEN ACCESS

Edited by:
Sujan Timilsina,
University of Florida, United States

Reviewed by:
Marcus Michael Dillon,
University of Toronto, Canada
Roshni R. Kharadi,
Michigan State University, United States

****Correspondence:***
David A. Baltrus
baltrus@email.arizona.edu
Brian H. Kvitko
bkvitko@uga.edu

Specialty section:
This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: *30 November 2021*
Accepted: *19 January 2022*
Published: *18 February 2022*

Citation:
Baltrus DA, Feng Q and Kvitko BH (2022) Genome Context Influences Evolutionary Flexibility of Nearly Identical Type III Effectors in Two Phytopathogenic Pseudomonads. Front. Microbiol. 13:826365. doi: 10.3389/fmicb.2022.826365

INTRODUCTION

Genome fluidity is crucial for survival of bacterial phytopathogens, as selection pressures on the presence and function of specific virulence genes can dramatically change from host to host (Dillon et al., 2019a; de Vries et al., 2020). Much research characterizing gene composition of bacterial pathogens focuses on presence and absence of specific virulence genes, and therefore often extrapolates from lists of loci to predict virulence and evolutionary potential (Baltrus et al., 2011; Dillon et al., 2019b). However, even if virulence genes are conserved across strains of a particular pathogen, differences in genomic flexibility for shared virulence genes could potentiate different evolutionary outcomes between strains under conditions of strong selection. Here we characterize one instance where such differences in genomic context and fluidity for a set of type III effectors

affects evolutionary potential for *Pseudomonas syringae*, and we speculate about the ability of such systems to shift evolutionary dynamics for phytopathogens moving forward.

Integrative Conjugative Elements (ICEs) are mobile replicons that blend characteristics of both plasmids and prophage and are well known for their ability to harbor antibiotic resistance genes and other niche-association traits across bacteria (Johnson and Grossman, 2015). Like conjugative plasmids, ICEs contain all genes and pathways required for conjugation between bacterial cells. Like prophage, ICEs contain site specific recombinases that enable recombination into chromosomal locations and repress the genes responsible for conjugation and replication while in this quiescent state. Most importantly, ICEs also contain cargo regions that can house genes and pathways that directly contribute to dramatic changes in bacterial phenotypes (Lovell et al., 2009; Johnson and Grossman, 2015). Not only can these elements modify phenotypes in their current host cells, but the ability of ICEs to transfer throughout populations and across communities through horizontal gene transfer means that cargo genes present on ICEs can rapidly proliferate across strains if they are beneficial. The rise of sequencing technologies enabling closed bacterial genomes has reinforced the importance of mobile elements in genomes and increased awareness of the prevalence of ICEs throughout bacteria.

The phytopathogen *Pseudomonas syringae* (sensu lato) is a recognized agricultural pest for many crops throughout the world, with numerous strains well established as laboratory systems for understanding virulence of plant pathogens *in vitro* and *in planta* (Baltrus et al., 2017). The presence of a type III secretion system is critical for virulence *in planta* for many strains of this pathogen, and this system is used to translocate upwards of 40 effectors proteins per strain from the bacterial cytoplasm into plant cells (Dillon et al., 2019a; Laflamme et al., 2020). Once inside the plant cells, effector proteins can disrupt plant immune responses in a variety of ways to promote bacterial growth and infection. The presence of effector proteins can also be monitored by plant immune responses through the action of R-genes, with recognition of effector protein functions leading to an overarching immune reaction termed effector triggered immunity (ETI) (Collmer et al., 2000; Jones and Dangl, 2006; Jayaraman et al., 2021). Triggering of the ETI response can quickly shut down nascent infections in resistant plant cultivars and has thus formed the basis of future plans to engineer durable crop resistance to infection through genetic modification and selective breeding (Dong and Ronald, 2019; Laflamme et al., 2020). Thus, type III effectors sit at an evolutionary inflection point where they can be highly beneficial for bacterial growth in some host backgrounds and highly detrimental in others.

Throughout this manuscript, we use the phrase "genome context" as a catch-all term to represent both the placement and relative location of specific genes within a genome compared to orthologs and homologs of these specific genes. We intend this as an open-ended term to describe how sequence characteristics (gene order, location, etc.) can affect parameters that contribute to evolutionary potential. For instance, imagine two antibiotic resistance genes that are sequence identical and found within a single genome but where one copy is present on a chromosome and stably maintained whereas the other copy is found in a region that is much more easily lost (e.g., on a small mobilizable plasmid). While these two genes may play similar if not redundant phenotypic roles for cells containing this particular version of the genome, differences between copies in rates of loss and horizontal transfer could significantly influence the evolutionary potential of each copy across the population and microbial community by altering population genetic parameters like population size and deletion rates.

The presence of virulence genes, including type III effectors, on plasmids and consequent movement across strains can dramatically alter trajectories of virulence evolution for *P. syringae* on different host species (Cazorla et al., 2002; Schierstaedt et al., 2019). Plasmids can be lost or modified if effector proteins are recognized by a potential host, which could enable infections to proceed at a population level despite the ETI response (Grant et al., 2006; Bardaji et al., 2019). Plasmids are not the only element within *P. syringae* genomes displaying plasticity, though. For instance, the effector AvrPphB (aka HopAR1) triggers HR in bean plants that contain the R-gene RPS5 (Qi et al., 2014). AvrPphB is found on an genomic island in certain strains of *P. syringae* pathovar phaseolicola, and under selective pressure from ETI responses, bacteria with the island excised from the chromosome and maintained independently as an episome rapidly dominate the population (Godfrey et al., 2011; Neale et al., 2018). Excision of the genomic island containing AvrPphB prevents ETI because this Avr gene is downregulated under episomal replication, but enables this gene to be maintained within a population for infection of host plants where it may be beneficial. ICE elements have also been identified and characterized in *P. syringae* pathovar actinidae strains where their presence/absence is one of the most glaring differences between closely related strains and where they contribute to large scale differences in strain metabolism and resistance to antibacterial compounds (Colombi et al., 2017; McCann et al., 2017; Poulter et al., 2018).

Here we describe how genomic context in a locus encoding an effector protein, HopQ, differentially affects evolutionary flexibility for strains containing this effector. Recognition of HopQ by the plant R-gene Roq1 triggers ETI in the host plant *Nicotiana benthamiana*, and thus limits growth of strain *P. syringae* pv. tomato DC3000 (*Pto)* in this host (Wei et al., 2007; Schultink et al., 2017). We characterize the genomic context for *hopQ* and other linked virulence genes in both *Pto*DC3000 and another pathogen, *Pseudomonas syringae* pv. maculicola ES4326 (*Pma*ES4326). We demonstrate that this and other effectors are differentially present in an active ICE element in *Pma*ES4326 but not in *Pto*DC3000, and confirm that recognition of HopQ also limits growth of *Pma*ES4326 in *N. benthamiana*. These differences are somewhat surprising because at a broad scale these regions appear syntenic. We further show how this differential genomic context allows for differential genomic plasticity for *hopQ* between these two strains, by demonstrating that passage of strains in *Nicotiana benthamiana* leads *hopQ* to be more readily lost in *Pma*ES4326 than in *Pto*DC3000. Thus, we directly demonstrate how genomic context for homologous virulence

genes in two closely related pathogens can contribute to different evolutionary trajectories for these strains.

MATERIALS AND METHODS

Bacterial Strains and Culturing Conditions

E. coli was routinely cultured at 37°C in LB (Lysogeny broth; per 1 L = 10 g tryptone, 5 g yeast extract, 10 g NaCl, pH 7.5 + with 15 g agar for solidified media). *P. syringae* was routinely cultured at 25°C or 28°C in either LM (LB modified; per 1 L = 10.0 g tryptone, 6.0 g yeast extract, 2.0 g $K_2HPO_4{\cdot}3H_2O$, 0.6 g NaCl, 0.4 g $MgSO_4{\cdot}7H_2O$, with 18 g agar for solidified media), KB (King's B; per 1 L = 20.0 g proteose peptone 3, 0.4 g $MgSO_4{\cdot}7H_2O$, glycerol 10 mL, 2.0 g $K_2HPO_4{\cdot}3H_2O$, with 18 g agar for solidified media) or KBC (KB amended with boric acid 1.5 g/L and cephalexin 80 mg/L). Liquid cultures were incubated with shaking at 200 rpm. Where appropriate, media was augmented at final concentration with rifampicin (Rf) 40–60 μg/mL, gentamicin (Gm) 10 μg/mL, kanamycin (Km) 50 μg/mL, spectinomycin (Sp) 50 μg/mL, and diaminopimelic acid (DAP) 200 μg/mL for liquid media, 400 μg/mL for solid media for the growth of DAP-auxotrophic *E. coli* strains.

Genome Sequencing and Assembly

For each strain with a genome sequence reported herein, a frozen stock was streaked to single colonies on King's B (KB) agar plates, at which point a single colony was picked to 2 mL KB liquid and grown overnight on a shaker at 27°C. Genomic DNA was extracted from these overnight cultures using a Promega (Madison, WI) Wizard kit and including the optional RNAse step. Each strain was sequenced using multiple technologies, and in each case independent DNA isolations were used to prepare libraries for different sequencing platforms. Illumina sequencing for strains was performed by MiGS (Pittsburgh, PA) using their standard workflow. As described in Baym et al. (2015), this workflow uses an Illumina tagmentation kit for library generation, followed by sequencing on a NextSeq 550 instrument with 150-bp paired-end reads. Trimmomatic was used for adaptor trimming (Bolger et al., 2014). For nanopore sequencing for strain *Pma*ES4326-D, a library was prepared from unsheared DNA using the Rapid sequencing kit (SQK-RAD004) and sequenced on a Flongle flowcell. For nanopore sequencing for strains *Pma*ES4326-C, *Pma*ES4326ΔDQ3, and *Pma*ES4326-C-LA-P5-20-1, each library was prepared using unsheared DNA as an input to the LSK109 ligation sequencing kit and was sequenced on R9.4 MinION flowcells. Nucleotide bases were called during sequencing using Guppy v3.2.6 in Fast-Mode. All genomes were assembled using Unicycler v.0.4.8 (Wick et al., 2017). The public facing genomes for *Pma*ES4326-D and *Pma*ES4326-C were annotated using PGAP (Tatusova et al., 2016). Genomes for *Pma*ES4326ΔDQ3, and *Pma*ES4326-C-LA-P5-20-1 found in Figshare doi: 10.6084/m9.figshare.17064080 were annotated using Prokka v 1.14.6 (Seemann, 2014). We used breseq v. 0.35.7 (Deatherage and Barrick, 2014) to identify evolutionary changes that occurred in the passage strain *Pma*ES4326 LA-P5-20-1 (hereafter *Pma*ES4326 LAP5-20). Default parameters were used for all software unless otherwise specified.

Integrative Conjugative Element Identification

Potential ICE elements were identified by manual searches of gene annotations for loci that could code for proteins critical for processing of conjugative elements. These include a site specific recombinase, a TraG-like NTPase, an ATPase, and numerous proteins involved in creation of the conjugation pilus. To be present within an ICE, these elements must all be present in a contiguous segment of the genome which is bordered by a potential attachment site (often tRNA loci).

DNA Manipulation

Plasmid DNA was routinely purified using the GeneJet plasmid miniprep kit (Thermo Fisher Scientific). PCR was conducted with Phusion HiFi polymerase (Thermo Fisher Scientific). PCR/reaction cleanup and gel extraction were conducted with the Monarch PCR and DNA cleanup kit and Monarch DNA gel extraction kits (NEB). *E. coli* transformation was conducted by either by preparing competent cells using the Mix and Go! *E. coli* transformation kit (Zymo Research) or standard electro-transformation protocols. *Pma*ES4326 strains were transformed with pCPP5372hopQ (pHopQ) (Wei et al., 2018) plasmid DNA via electro-transformation after washing and concentration of overnight liquid cultures with 300 mM sucrose (Choi et al., 2006). Restriction enzymes, T4 ligase, and Gibson Assembly Mastermix were purchased from NEB. Oligonucleotide primers were synthesized by IDT. Commercial molecular biology reagents were used in accordance with their manufacturer's recommendations.

To create the site-specific Tn*7* 3xmCherry labeling vector pURR25DK-3xmcherry, pURR25 (Teal et al., 2006) was first digested with *Pst*I and recircularized with T4 ligase to remove the *nptII* Km^R marker gene creating pURR25DK. To replace the *gfp* gene in pURR25DK, the 3xmCherry cassette was PCR amplified from pGGC026 (pGGC026 was a gift from Jan Lohmann (Addgene plasmid # 48831[1]; RRID:Addgene_48831) using primers bko374 (5′**ACATCTAGAATTAAAGAGGAGAAATTAAGCATGGTG AGCAAGGGCGAGGAGGATAACATG 3′**) **and** bko375 (5′**CAGGAGTCCAAGCTCAGCTAATTAAGCTTACTTGTAC AACTCATCCATACCACCTGTTGA 3′**) to introduce 30 bp 5′ overlaps corresponding to the *gfp* flanking regions. Gibson assembly was used to join the 3xmCherry PCR amplicon with pURR25DK backbone digested with *Bse*RI and partially digested with *Hin*dIII.

Bacterial Conjugation and Creation of Mutant Strains

All genetic manipulations of *Pma*ES4326 were conducted with the *Pma*ES4326-C strain. Conjugations were performed by mixing 15 μL of fivefold concentrated, washed, overnight LM liquid cultures of each parent strain and co-culturing at 28°C

[1]http://n2t.net/addgene:48831

overnight on sterile nitrocellulose membranes on either LM or LM + DAP plates (for conjugations with DAP-auxotrophic *E. coli* donor strains). Tn*7* transposition conjugations always included the *E. coli* RHO3 pTNS3 Tn*7* transposase helper strain as a third parent. For all conjugation experiments cultures of each parent strain were included separately as controls. *Pma*ES4326 merodiploid exconjugants of pCPP5729 (pK18msGmΔ*hopQ1)* were recovered on LM Km (Kvitko et al., 2009). Resolved Pma ES4326 pCPP5729 merodiploids were recovered via counter-selection on LM Rf + 10% sucrose and sucrose resistant clones were screened for kanamycin sensitivity by patch plating indicating the loss of the pK18ms plasmid backbone (Kvitko and Collmer, 2011). Derivative *att*Tn*7*-3xmCherry transposant strains were recovered on LM Sp and pink clones were selected after 4 days incubation at 4°C and restreaked to isolation. To test the native mobility of ICE-DQ, conjugation was conducted as described above with the *Pma*ES4326 pCPP5729 merodiploid as the donor parent and *att*Tn*7*-3xmCherry derivatives of *Pto*DC3000 and *Pma*ES4326 ΔICE-DQ strains (*Pma* DQ3 and *Pma* LAP5-20) as recipients. ICE-DQ exconjugants were recovered on LM Rf Sp Km. For all conjugation experiments cultures of each parent strain were included separately as controls. Conjugation frequency was calculated as the number of Sp^RKm^R colonies recovered per recipient CFU as determined by dilution plating.

Nicotiana benthamiana Growth, Inoculation and Bacterial Passage Assays

Nicotiana benthamiana, WT LAB accession, and *roq1-1* (Qi et al., 2018) were grown in a Conviron Adaptis growth chamber with 12 h light (125 μmol/m^2/s) at 26°C and 12 h dark at 23°C. Plants were used at 5-7 weeks post germination. To prepare inoculum, *P. syringae* cultures were recovered from fresh KB plate cultures, resuspended in 0.25 mM $MgCl_2$, standardized to OD_{600} 0.2, and serially diluted 10,000X to ∼3 × 10^4 CFU/mL. Cell suspensions were infiltrated with a blunt syringe into either the 3rd, 4^{th}, or 5th leaves. Infiltrated spots were allowed to dry fully and then plants were covered with a humidity dome and kept at 100% humidity for 6–8 days to allow symptoms to develop. At end point, leaves were photographed to document symptoms and four 4 mm diameter leaf punches (∼0.5 cm^2 total) were collected with a 4 mm diameter biopsy punch from each infiltrated area. Discs were macerated in 0.1 mL of 0.25 mM $MgCl_2$ using an Analytik Jena SpeedMill Plus homogenizer and the bacterial CFU/cm^2 leaf tissue was determined by serial dilution spot plating from 10 μL volumes on LM Rf.

For *P. syringae* passaging in *N. benthamiana*, single colonies of *Pto* DC3000 WT and *Pma*ES4326 were inoculated into LM Rf liquid cultures. Samples of the initial cultures (P0) were cryo-preserved at −80°C in 15% final strength sterile glycerol. The liquid cultures were diluted 1000X in 0.25 mM $MgCl_2$ to approximately 5 × 10^5 CFU/mL prior to syringe infiltration into three leaf areas establishing three lineages (A, B, C) each for *Pto*DC3000 and *Pma*ES4326. Tissue samples were collected 6–7 days post inoculation and processed as described above. Bacteria cultured from the 10^{-1} tissue macerate dilutions of each lineage were directly scraped from the dilution plate with an inoculation loop and suspended in 1 mL 0.25 mM $MgCl_2$. These suspensions were then sub-cultured 5 μL into 5 mL LM Rf and diluted as described above to create inoculum for the next passage. This passaging scheme was repeated five times and samples of both the post-passage recovered bacteria (P1) and the corresponding sub-cultured bacteria (P1c) were cryo-preserved for each lineage and each passage. To screen for changes in *N. benthamiana* disease compatibility with passaged strains in a medium-throughput format, bacteria from the P5 cryo-preserved samples were streaked to isolation on KBC plates and isolated "P5" colonies were cultured in 200 μL of LM Rf in sterile 96 well microtiter plates along with Pto DC3000, *Pto* DC3000Δ*hopQ*, *Pma* ES4326 and *Pma*ES4326ΔDQ3 control strains. Cultures were serially diluted 5000X in 0.25 mM $MgCl_2$ to ∼3 × 10^4 CFU/mL and inoculated into *N. benthamiana* leaves as described. Inoculum concentration was verified via serial dilution spot plating. Inoculated areas were monitored visually for qualitative changes to disease symptoms compared to their respective WT and targeted *hopQ* deletion control strains over the course to 6–8 days. The CFU/cm^2 leaf tissue bacterial load of select clones was determined for each strain as described above. Isolated colonies of strains that displayed bacterial load and symptoms comparable to their respective Δ*hopQ* backgrounds were subcultured from the dilution spot count plates and retained.

We estimate that N_e of these passage populations to be on the order of 2.5 × 10^4 over the course of five passages. This number was calculated by taking the inoculum of each passage (5 × 10^5 CFU/mL in 50 μL) and estimating that approximately 10 μL were inoculated per cm^2 for a total of 2.5 × 10^4 for the area inoculated (∼5 cm^2). We further estimate that this is the smallest bottleneck over the course of the passages, and that populations undergo ∼10 generations of division, as these populations grow from 10^4 to 10^7 cm^2 *in planta*). Given these assumptions, N_e can be estimated by taking the harmonic mean of fluctuating population sizes (Sjödin et al., 2005) and will be highly skewed by smaller numbers. Since 1 cm^2 is harvested at the end of each cycle, a conservative estimate for N_e over five passages is therefore approximately 2.5 × 10^4 CFU.

RESULTS

Complete Genome Sequences for Multiple Isolates of *Pma*ES4326

We previously reported a draft genome sequence for an isolate of *Pma*ES4326 acquired from the lab of Jeff Dangl (Baltrus et al., 2011), and our first goal was to generate a complete genome sequence for this strain (referred to herein as *Pma*ES4326-D, **Table 1**). We used MiGS (Pittsburgh, PA) to generate Illumina reads for *Pma*ES4326-D, and their workflow generated a total of 3,284,990 paired reads and 418 Mb (∼64× coverage) of sequence. We also independently isolated genomic DNA and sequenced using an Oxford Nanopore MinION to generate 32,291 reads for a total of 465 Mb (∼71× coverage) of sequence with a read N50 of 30,656 bp. Assembly of these reads resulted in a complete circular chromosome and four separate plasmids.

TABLE 1 | Sequencing of *Pma*ES4326 strains.

Strain	Reads (Illumina/Nanopore)	Total bp sequenced Mb (Illumina/Nanopore)	Read N50 Nanopore bp	Total sequencing depth (Illumina/Nanopore)	SRA accession (Illumina/Nanopore)
*Pma*ES4326-D	3,284,990/32,291	418.372/465.212	30,656	64×/71×	SRR1598872 SRR15988724
*Pma*ES4326-C	3,757,284/55,845	485.469/700.483	23,669	74×/107×	SRR15988571/ SRR15988568
*Pma*ES4326ΔDQ3	2,479,212/15,275	331.276/178.118	31,491	51×/27×	SRR15988570/ SRR15988567
*Pma*ES4326-C-LA-P5-20-1	1,372,720/9,934	189.971/146.897	31,020	29×/22×	SRR17005742/ SRR15988569

TABLE 2 | Assembly of *Pma*ES4326 strains.

Strain	Complete genome	Number of contigs in assembly	Replicons present	Assembly accession
*Pma*ES4326-D	Yes	5	Chromosome (circular), pPmaES4326ABEF	GCA_000145845.2
*Pma*ES4326-C	Yes	7	Chromosome (circular), pPmaES4326ABCDEF	GCA_020309905.1
*Pma*ES4326ΔDQ3	Yes	6	Chromosome (circular), pPmaES4326ABCF	doi: 10.6084/m9.figshare.17064080
*Pma*ES4326-C-LA-P5-20-1	No	7	Chromosome (not circular), pPmaES4326ABCDEF	doi: 10.6084/m9.figshare.17064080

Notably, our isolate of *Pma*ES4326-D contains three previously reported plasmids (pPma4326A, pPma4326B, pPma4326E) but also appears to have lost two different plasmids (pPma4326C and pPma4326D) first reported as present in this strain by Stavrinides and Guttman (2004). The assembly for this strain appears to contain an additional ∼350 kb plasmid that was not reported by Stavrinides and Guttman (2004) and which we name pPma4326F following earlier naming conventions.

Given the absence of two plasmids from *Pma*ES4326-D, we sought to generate a genome assembly from a different lab isolate, acquired from the lab of Alan Collmer referred to here as *Pma*ES4326-C (**Table 1**). We used MiGS (Pittsburgh, PA) to generate Illumina reads for the Collmer lab version of *Pma*ES4326, and their workflow generated a total of 3,757,284 paired reads and 485 Mb (∼74× coverage) of sequence. We also independently isolated genomic DNA and sequenced using an Oxford Nanopore MinION to generate 55,845 reads for a total of 700 Mb (∼107× coverage) of sequence with a read N50 of 24,669 bp. Hybrid assembly of all read types resulted in a single chromosome that appears nearly complete but was not circularized by Unicycler. However, this assembly does contain all predicted circular plasmids as well as pPmaES4326F.

Sequencing and assembly characteristics for all strains can be found in **Tables 1**, **2**, respectfully.

An Integrative Conjugative Element Hotspot in *Pma*ES4326

Genomic inspection of the region containing type III effectors *hopQ* and *hopD* in strains *Pma*ES4326-C and *Pma*ES4326-D indicate that this area is a potential hotspot for genomic plasticity. Specifically, in a region bordered by *clpB* and the type III effector *hopR,* gene content characterization strongly suggests the presence of two independent integrative conjugative elements (ICEs, **Figure 1**). Both ICEs are approximately 70 kb in length, and are found adjacent to tRNA loci encoding a proline anticodon. Moreover, roughly 70% of each element is composed of sequences with > 95% nucleotide similarity and which encode many of the structural genes predicted to be involved in ICE proliferation and integration. These conserved genes include predicted integrases/recombinases, pilus proteins and ATPases, regulator proteins, a topoisomerase (*topB*) and helicase, chromosome partitioning proteins (*parB*), DNA coupling proteins (*traD*), an NTPase (*traG*), and numerous loci annotated as "integrative conjugative element proteins." We have included annotations for all genes within these two ICEs in **Table 3**.

Despite relatively high sequence similarity across these two closely related ICEs positioned successively in the genome of *Pma*ES4326, gene comparisons suggest the presence of three distinct cargo regions that clearly differentiate these elements (**Figure 1**). The first ICE contains type III effector proteins *hopAO* and *hopAB3-1* (aka *hopPmaL*) in two different predicted cargo regions, and so we name this region ICE-ABAO. The second ICE contains loci for type III effectors *hopD* and *hopQ* as well as the lytic transglycosylase *hopAJ1*, and so we name this ICE ICE-DQ. Cargo region one contains four predicted ORFs in *Pma*ECE-ABAO and fourteen predicted ORFs in ICE-DQ. Cargo region two contains two predicted ORFs in ICE-ABAO and seven predicted ORFs in ICE-DQ, including the effectors *hopAO*, *hopD*, and *hopQ*. Cargo region three contains six predicted ORFs in ICE-ABAO and two in ICE-DQ. Aside from the type III effectors, many of loci potentially encoded by these cargo regions are annotated as hypothetical proteins or as parts of IS elements and transposons.

Differential Contexts for the Genomic Island Containing *hopD* and *hopQ* Across *Pma*ES4326 and *Pto*DC3000

The type III effectors HopD and HopQ are nearly sequence identical in two phytopathogens commonly used as laboratory

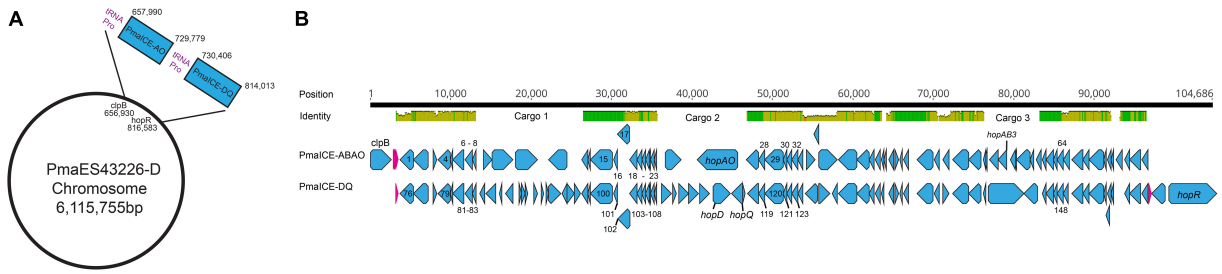


FIGURE 1 | Comparison of PmaICE-ABAO and PmaICE-DQ. **(A)** A representation of the circular chromosome of strain *Pma*ES4326-D, highlighting the region where both ICE elements are found as well as the tRNA-Pro loci that border these ICE elements. **(B)** We aligned nucleotide sequences from the pair of potential ICEs found within the *clpB-hopR* region in strain *Pma*ES4326 against each other, and visualize the results here. The top line of the figure displays overall nucleotide identity between these two ICEs, with green bars representing conserved nucleotides and yellow representing slight to modest divergence. If no colored bars on this top line, the sequences are completely divergent, and we highlight the three potential cargo regions that differ between these ICE elements. Predicted loci for each ICE are shown on the lines immediately below the nucleotide identity comparison. *Pma*ICE-ABAQ is positioned immediately adjacent to *clpB* in the *Pma*ES4326 genome while *Pma*ICE-DQ is positioned immediately adjacent to *hopR* in the same genome. Each predicted locus is represented in the figure in blue, while predicted tRNA loci are represented in magenta. We labeled all identified type III effector loci within these regions as well as *clpB*. All loci that are implicated in ICE function, by annotation, have been labeled with numbers that correspond to their annotations within **Table 3**.

models for studying *P. syringae* virulence, *Pto*DC3000 and *Pma*ES4326 (Baltrus et al., 2011). Furthermore, these effectors are found in roughly the same genomic locations in the two strains, bordered on one side by *clpB* and on the other by the type III effector *hopR* (**Figures 1**, **2**). Broad comparisons between *Pto*DC3000 and *Pma*ES4326 further suggest that these effectors form a genomic island along with a third effector (*hopR*) across these two strains (**Figure 2**), an island which has been referred to as effector cluster IV in in the *Pto*DC3000 genome (Kvitko et al., 2009). Moreover, the size of the region between *clpB* and *hopR* in *Pto*DC3000 is roughly 70 kb, which closely approximates the size of each ICE in *Pma*ES4326. While this region within PtoDC3000 does contain hints that it previously housed a functional ICE, there are other sections that are quite divergent from that in *Pma*ES4326 (**Figure 2**). Specifically, the predicted recombinase in *Pto*DC3000 is truncated due to a frameshift mutation and multiple other loci that encode functions important for ICE conjugation and transfer are disrupted by IS elements (**Figure 2**). Lastly, we note that conservation of nucleotide and protein sequences between PmaICEs coupled with nucleotide diversity in *Pto*DC3000 at this genomic location renders accurate evaluation and comparison of evolutionary histories between *Pto*DC3000 and the ICEs of *Pma*ES4326 challenging. However, the most parsimonious explanation does appear that an ICE element housed both *hopQ* and *hopD* within an ancestor of strain *Pto*DC3000 and that this ICE was locked in place through both IS element insertions and loss of recombinase activity.

An Evictable Integrative Conjugative Element Containing HopQ and HopD in *Pma*ES4326

ICEs are categorized by their ability to cleanly excise from the genome, and we confirmed the prediction that the type three effectors *hopQ* and *hopD* are contained in an excisable region in *Pma*ES4326 in two distinct ways. As a first step, we characterized the size of the region excised during intentional creation of a *hopQ*- mutant. To do this, we generated a merodiploid strain in which plasmid pCPP5729 was recombined into a region adjacent to *hopQ*, and then selected for resolution of the merodiploid through plating on sucrose. Presence of *sacB* in plasmid pCPP5729 renders the merodiploid sensitive to killing by sucrose. Notably, this merodiploid also contains regions sequence identical to those surrounding *hopQ* in *Pma*ES4326 and so we originally expected that the *hopQ* gene could be locally deleted through RecA-dependent homologous recombination. We isolated sucrose resistant isolates after plating this merodiploid, and identified strains that were neither WT reverants nor clean *hopQ* deletions by PCR genotyping using previously validated genotyping primers. Interestingly, while no clean *hopQ* deletion mutants were identified after sucrose counter-selection of merodiploid strains, WT revertants were. We then performed whole genome sequencing to confirm whether the ICE-DQ genomic region was lost under these selective conditions. We refer to this strain recovered by *in vitro* selection hereafter as *Pma* DQ3. As one can see in **Figure 3**, *hopQ* and the surrounding regions that are predicted to be part of ICE-DQ have been lost in the targeted deletion strain *Pma* DQ3, confirming that this region can be cleanly and completely excised from the chromosome in a manner consistent with the action of site-specific recombinases contained by ICEs. Searches of both the raw Illumina and Nanopore reads for remnants of *hopQ* or *hopD* yielded no hits despite extensive depth (∼51× for Illumina reads and ∼27× for Nanopore reads), which suggests that ICE-DQ was fully lost and not retained as an episome in *Pma* DQ3 (data not shown).

Loss of PmaICE-DQ Enables Virulence of *Pma*ES436 on *Nicotiana benthamiana*

The presence of *hopQ* in *Pma* would be expected to elicit the ETI response in *Nicotiana benthamiana* accessions containing the R-gene *Roq1*, and a lack of compatible infection and disease. Given this information, we tested whether *Pma* DQ3 would

TABLE 3 | Gene annotations in *Pma*ES4326 ICEs.

Number	Name	Type	Minimum	Maximum	Length	Direction
	tRNA-Asn	tRNA	2,862	2,937	76	Forward
	tRNA-Pro	tRNA	2,967	3,043	77	Forward
	tRNA-Lys	tRNA	3,049	3,124	76	Forward
	tRNA-Pro	tRNA	3,220	3,296	77	Forward
1	Tyrosine-type recombinase/integrase	CDS	3,625	5,064	1,440	Reverse
2	DNA-binding domain-containing protein	CDS	5,061	6,992	1,932	Reverse
3	DUF3742 family protein	CDS	7,604	8,005	402	Forward
4	*traG* conjugal transfer protein	CDS	8,047	9,585	1,539	Reverse
5	Hypothetical protein	CDS	9,582	9,950	369	Reverse
6	Integrating conjugative element protein	CDS	9,953	11,323	1,371	Reverse
7	TIGR03756 family integrating conjugative elementprotein	CDS	11,344	12,282	939	Reverse
8	TIGR03757 family integrating conjugative elementprotein	CDS	12,279	12,755	477	Reverse
9	Hypothetical protein	CDS	13,053	13,241	189	Forward
10	Hypothetical protein	CDS	13,259	13,798	540	Forward
11	Hypothetical protein	CDS	13,839	14,393	555	Forward
12	Hypothetical protein	CDS	14,752	15,297	546	Reverse
13	Thioredoxin domain-containing protein	CDS	15,609	16,307	699	Reverse
14	Hypothetical protein	CDS	16,304	16,603	300	Reverse
15	Conjugative transfer ATPase	CDS	16,600	19,347	2,748	Reverse
16	TIGR03751 family conjugal transfer lipoprotein	CDS	19,557	20,012	456	Reverse
17	TIGR03752 family integrating conjugative elementprotein	CDS	19,990	21,486	1,497	Reverse
18	TIGR03749 family integrating conjugative elementprotein	CDS	21,476	22,399	924	Reverse
19	TIGR03746 family integrating conjugative elementprotein	CDS	22,396	23,064	669	Reverse
20	TIGR03750 family conjugal transfer protein	CDS	23,061	23,453	393	Reverse
21	TIGR03745 family integrating conjugative elementmembrane protein	CDS	23,469	23,837	369	Reverse
22	TIGR03758 family integrating conjugative elementprotein	CDS	23,857	24,096	240	Reverse
23	Conjugal transfer protein	CDS	24,093	24,458	366	Reverse
24	DUF4177 domain-containing protein	CDS	24,538	24,870	333	Reverse
25	Hypothetical protein	CDS	25,041	25,472	432	Forward
26	*hopAO* type III effector	CDS	25,861	26,877	1,017	Reverse
27	UvrD-helicase domain-containing protein	CDS	27,062	28,519	1,458	Reverse
28	TIGR03747 family integrating conjugative elementmembrane protein	CDS	28,529	29,278	750	Reverse
29	*traD*	CDS	29,318	31,477	2,160	Reverse
30	Integrating conjugative element protein	CDS	31,489	31,992	504	Reverse
31	Transglycosylase SLT domain-containing protein	CDS	31,989	32,546	558	Reverse
32	TIGR03759 family integrating conjugative element protein	CDS	32,531	33,277	747	Reverse
33	Hypothetical protein	CDS	33,286	33,990	705	Reverse
34	*dcm*	CDS	34,335	35,390	1,056	Forward
35	*vsr*	CDS	35,329	35,766	438	Reverse
36	DNA mismatch repair protein	CDS	35,850	37,871	2,022	Forward
37	DEAD/DEAH box helicase family protein	CDS	37,973	40,231	2,259	Reverse
38	Class I SAM-dependent methyltransferase	CDS	40,335	41,786	1,452	Reverse
39	Hypothetical protein	CDS	41,816	42,415	600	Reverse
40	DUF3275 family protein	CDS	42,536	43,147	612	Reverse
41	Hypothetical protein	CDS	43,248	43,610	363	Reverse
42	Hypothetical protein	CDS	43,672	43,986	315	Reverse
43	DUF3577 domain-containing protein	CDS	44,290	44,859	570	Reverse
44	Hypothetical protein	CDS	44,879	45,403	525	Forward
45	Hypothetical protein	CDS	45,480	45,848	369	Reverse
46	Regulator	CDS	46,056	46,589	534	Reverse
47	*topB*	CDS	47,650	49,671	2,022	Reverse
48	Hypothetical protein	CDS	49,794	50,363	570	Reverse
49	Hypothetical protein	CDS	50,657	51,094	438	Reverse
50	ATP-dependent helicase	CDS	51,329	53,281	1,953	Reverse
51	Hypothetical protein	CDS	53,284	54,849	1,566	Reverse
52	CrpP family protein	CDS	55,287	55,466	180	Reverse
53	Hypothetical protein	CDS	55,471	55,644	174	Reverse
54	AAA family ATPase	CDS	55,991	56,941	951	Reverse
55	*hopAB* type III effector	CDS	56,958	58,115	1,158	Reverse
56	IS91 family transposase	CDS	58,481	59,101	621	Reverse
57	RulB protein	CDS	59,132	59,200	69	Reverse
58	IS481 family transposase	CDS	59,217	60,878	1,662	Reverse
59	Recombinase family protein	CDS	60,859	61,434	576	Reverse
60	Hypothetical protein	CDS	61,751	62,224	474	Reverse

(Continued)

TABLE 3 | (Continued)

Number	Name	Type	Minimum	Maximum	Length	Direction
61	Single-stranded DNA-binding protein	CDS	62,425	62,886	462	Reverse
62	Phage regulatory protein	CDS	62,898	63,730	833	Reverse
63	DUF3158 family protein	CDS	63,736	64,257	522	Reverse
64	TIGR03761 family integrating conjugative elementprotein	CDS	64,294	65,073	780	Reverse
65	Hypothetical protein	CDS	65,066	65,662	597	Reverse
66	Hypothetical protein	CDS	65,992	67,353	1,362	Reverse
67	DUF2857 family protein	CDS	67,350	68,087	738	Reverse
68	*parB* chromosome partitioning protein	CDS	68,124	69,884	1,761	Reverse
69	Arc family DNA-binding protein	CDS	69,887	70,204	318	Reverse
70	Hypothetical protein	CDS	70,382	70,786	405	Reverse
71	Hypothetical protein	CDS	70,788	71,117	330	Reverse
72	Hypothetical protein	CDS	71,437	72,123	687	Reverse
73	Hypothetical protein	CDS	72,563	73,153	591	Reverse
74	*dnaB*	CDS	73,150	74,496	1,347	Reverse
75	AAA family ATPase	CDS	74,554	75,414	861	Reverse
	tRNA-Pro	tRNA	75,651	75,727	77	Forward
76	Tyrosine-type recombinase/integrase	CDS	76,041	77,495	1,455	Reverse
77	DNA-binding domain-containing protein	CDS	77,492	79,423	1,932	Reverse
78	DUF3742 family protein	CDS	80,037	80,438	402	Forward
79	*traG* conjugal transfer protein	CDS	80,499	82,031	1,533	Reverse
80	Hypothetical protein	CDS	82,028	82,402	375	Reverse
81	Integrating conjugative element protein	CDS	82,405	83,772	1,368	Reverse
82	TIGR03756 family integrating conjugative elementprotein	CDS	83,793	84,731	939	Reverse
83	TIGR03757 family integrating conjugative elementprotein	CDS	84,728	85,159	432	Reverse
84	DUF1440 domain-containing protein	CDS	85,713	86,201	489	Reverse
85	Lytic murein transglycosylase	CDS	86,453	87,694	1,242	Reverse
86	Hypothetical protein	CDS	88,186	88,395	210	Reverse
87	IS5 family transposase	CDS	88,841	89,818	978	Reverse
88	Carbon storage regulator	CDS	90,586	90,810	225	Forward
89	TraR/DksA family transcriptional regulator	CDS	90,846	91,262	417	Forward
90	Hypothetical protein	CDS	91,297	91,518	222	Forward
91	GNAT family N-acetyltransferase	CDS	92,339	92,830	492	Forward
92	AAA family ATPase	CDS	93,208	93,489	282	Forward
93	*tnpB*	CDS	93,794	94,150	357	Forward
94	IS66 family transposase	CDS	94,167	95,699	1,533	Forward
95	IS5 family transposase	CDS	95,741	96,382	642	Reverse
96	AAA family ATPase	CDS	96,453	97,271	819	Forward
97	(p)ppGpp synthetase	CDS	97,274	98,092	819	Forward
98	Thioredoxin domain-containing protein	CDS	98,266	98,964	699	Reverse
99	Hypothetical protein	CDS	98,961	99,260	300	Reverse
100	Conjugative transfer ATPase	CDS	99,257	102,004	2,748	Reverse
101	TIGR03751 family conjugal transfer lipoprotein	CDS	102,214	102,669	456	Reverse
102	TIGR03752 family integrating conjugative elementprotein	CDS	102,647	104,143	1,497	Reverse
103	TIGR03749 family integrating conjugative elementprotein	CDS	104,133	105,056	924	Reverse
104	TIGR03746 family integrating conjugative elementprotein	CDS	105,053	105,721	669	Reverse
105	TIGR03750 family conjugal transfer protein	CDS	105,718	106,110	393	Reverse
106	TIGR03745 family integrating conjugative elementmembrane protein	CDS	106,126	106,494	369	Reverse
107	TIGR03758 family integrating conjugative elementprotein	CDS	106,514	106,753	240	Reverse
108	Conjugal transfer protein	CDS	106,750	107,115	366	Reverse
109	DUF4177 domain-containing protein	CDS	107,195	107,525	331	Reverse
110	MFS transporter	CDS	108,102	109,286	1,185	Forward
111	Transposase	CDS	109,422	109,937	516	Forward
112	Amidinotransferase	CDS	110,235	111,335	1,101	Forward
113	Serine kinase - nikkomycin	CDS	111,422	112,675	1,254	Forward
114	Serine kinase	CDS	112,728	113,972	1,245	Forward
115	*hopD* type III effector protein	CDS	114,438	116,555	2,118	Forward
116	*hopQ* type III effector protein	CDS	116,671	118,014	1,344	Reverse
117	Phosphoribulokinase	CDS	118,048	118,365	318	Forward
118	UvrD-helicase domain-containing protein	CDS	118,603	120,060	1,458	Reverse
119	TIGR03747 family integrating conjugative elementmembrane protein	CDS	120,070	120,819	750	Reverse
120	*traD*	CDS	120,859	123,018	2,160	Reverse
121	Integrating conjugative element protein	CDS	123,030	123,536	507	Reverse
122	Tansglycosylase SLT domain-containing protein	CDS	123,533	124,090	558	Reverse
123	TIGR03759 family integrating conjugative elementprotein	CDS	124,075	124,821	747	Reverse

(Continued)

TABLE 3 | (Continued)

Number	Name	Type	Minimum	Maximum	Length	Direction
124	Hypothetical protein	CDS	124,830	125,534	705	Reverse
125	IS5 family transposase	CDS	125,711	126,561	851	Reverse
126	Hypothetical protein	CDS	126,700	127,581	882	Forward
127	Hypothetical protein	CDS	127,745	128,107	363	Reverse
128	DEAD/DEAH box helicase family protein	CDS	128,202	130,460	2,259	Reverse
129	Class I SAM-dependent methyltransferase	CDS	130,564	132,015	1,452	Reverse
130	Hypothetical protein	CDS	132,045	132,644	600	Reverse
131	DUF3275 family protein	CDS	132,770	133,381	612	Reverse
132	Hypothetical protein	CDS	133,482	133,844	363	Reverse
133	DUF3577 domain-containing protein	CDS	134,172	134,726	555	Reverse
134	Hypothetical protein	CDS	134,746	135,270	525	Forward
135	Hypothetical protein	CDS	135,347	135,715	369	Reverse
136	Regulator	CDS	135,923	136,456	534	Reverse
137	*topB*	CDS	137,517	139,538	2,022	Reverse
138	Hypothetical protein	CDS	139,661	140,230	570	Reverse
139	Hypothetical protein	CDS	140,774	141,457	684	Forward
140	ATP-dependent helicase	CDS	141,779	143,731	1,953	Reverse
141	Hypothetical protein	CDS	143,734	145,259	1,526	Reverse
142	CrpP family protein	CDS	145,697	145,897	201	Reverse
143	IS3 family transposase	CDS	146,020	147,260	1,241	Forward
144	Hypothetical protein	CDS	147,274	147,519	246	Reverse
145	Single-stranded DNA-binding protein	CDS	147,725	148,186	462	Reverse
146	Phage regulatory protein	CDS	148,198	149,031	834	Reverse
147	DUF3158 family protein	CDS	149,037	149,600	564	Reverse
148	TIGR03761 family integrating conjugative elementprotein	CDS	149,597	150,376	780	Reverse
149	Hypothetical protein	CDS	150,369	150,918	550	Reverse
150	Hypothetical protein	CDS	151,362	152,723	1,362	Reverse
151	DUF2857 family protein	CDS	152,720	153,457	738	Reverse
152	*parB* chromosome partitioning protein	CDS	153,494	155,254	1,761	Reverse
153	Arc family DNA-binding protein	CDS	155,257	155,574	318	Reverse
154	Hypothetical protein	CDS	155,510	155,755	246	Reverse
155	Hypothetical protein	CDS	155,752	156,156	405	Reverse
156	Hypothetical protein	CDS	156,158	156,487	330	Reverse
157	Hypothetical protein	CDS	156,797	157,387	591	Reverse
158	*dnaB*	CDS	15,7384	158,730	1,347	Reverse
159	AAA family ATPase	CDS	158,788	159,648	861	Reverse
	tRNA-Pro	tRNA	159,887	159,963	77	Forward

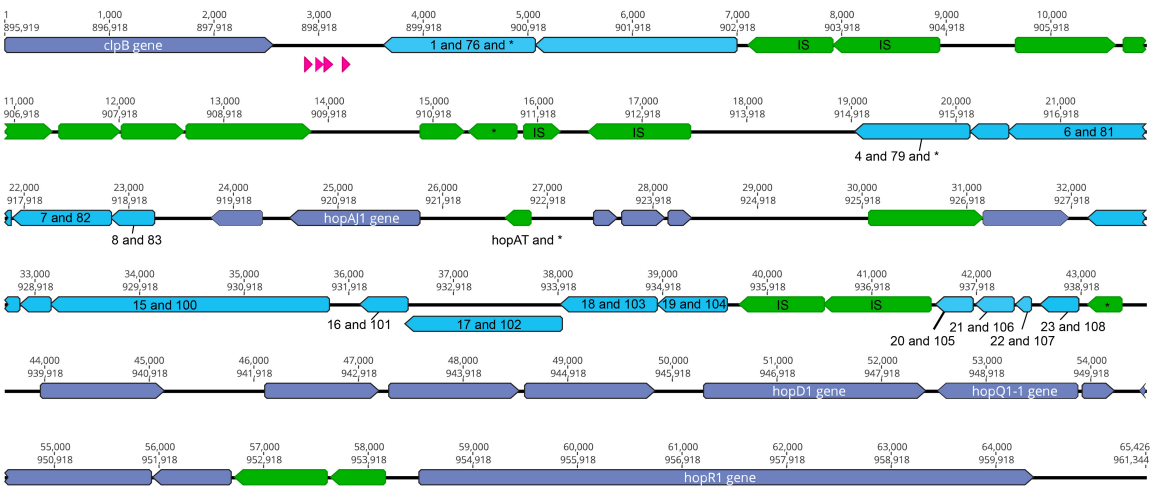


FIGURE 2 | The *clpB-hopR* Region in *Pto*DC3000. Here we visualize the region between *clpB* and *hopR* from the *Pto*DC3000 genome, and compare this region to both predicted ICE elements from *Pma*ES4326. Each predicted locus is colored as purple (present in one of the *Pma*ES4326 ICE elements), blue (present in both *Pma*ES4326 ICE elements as per nucleotide identity), and green (only present in *Pto*DC3000). Predicted tRNA loci are shown in magenta. We also show two nucleotide positions within this region: top number is from the start of *clpB* and the bottom number is from the start of the genome (from *dnaA*). We label all identified type III effector loci and *clpB*, as well as all annotated IS elements (with an "IS") and genes that have predicted frameshifts (with an "*"). All genes potentially implicated in ICE function have been labeled with numbers corresponding to potential orthologs from strain *Pma*ES4326 ICEs and listed in **Table 3**.

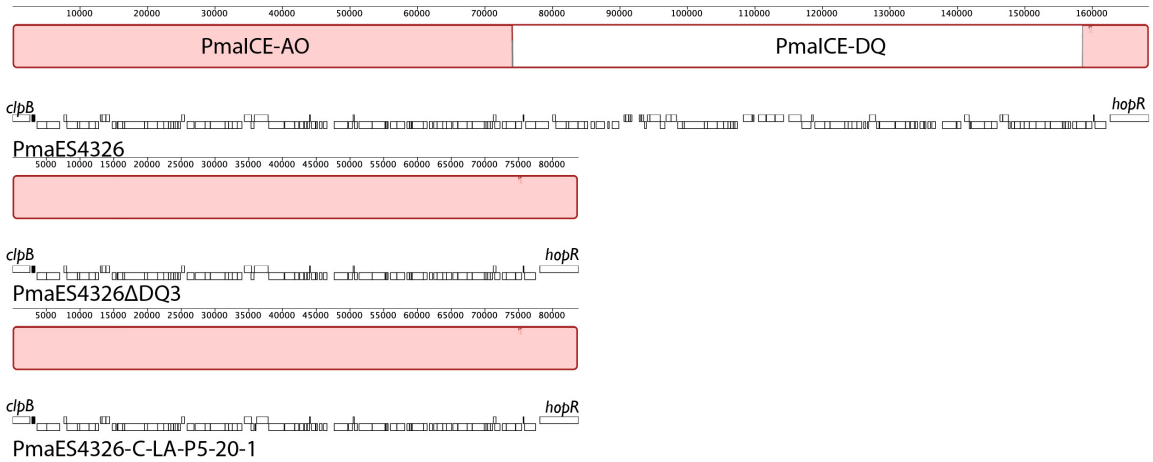


FIGURE 3 | Identical deletions of PmaICE-DQ following laboratory selection and passage under selective conditions *in planta*. We compared genomic regions bordering ICE-DQ in a strain where the region containing *hopQ* was intentionally deleted (*Pma* DQ3) as well as a strain that arose after five passages *in planta* in *N. benthamiana* (*Pma* LAP5-20). Genomic segments from each of the three strains are arranged vertically, with each line labeled by strain name and genes within each strain outlined in white boxes underneath the alignment. Each contiguous segment is oriented so that *clpB* is found on the left side of the segment and so that *hopR* is found on the right side. Sections within the alignment colored red are those that are aligned into contiguous blocks by Mauve and are shared by all three strains. A white section within the wild type *Pma*ES4326 alignment represents a region (PmaICE-DQ) that is not present in the other two strains. Mauve alignments of these regions demonstrate that identical evictions of ICE-DQ occurred in each strain.

gain compatibility with *N. benthamiana*. The *Pma* DQ3 strain did in fact gain disease compatibility with *N. benthamiana* in a manner similar to the gain-of-compatibility phenotypes previously observed for *Pto*DC3000 Δ*hopQ1-1* mutants and *Xanthomonas euvesicatoria* 85–10 Δ*xopQ* mutants (Wei et al., 2007; Adlung et al., 2016). Furthermore, *N. benthamiana* incompatibility could be restored in *Pma*ICEΔDQ3 by *hopQ* complementation *in trans* and *Pma*ES4326 incompatibility was not observed in *roq1-1* CRISPR-edited *N. benthamiana* as shown in **Figure 4** (Qi et al., 2018).

Loss of PmaICE-DQ Under Selection *in planta*

Observing that we could readily recover strains containing spontaneous evictions of the ICE-DQ containing a single crossover pK18 merodiploid under the selective pressure of sucrose counter-selection, and that the targeted *Pma* DQ3 strain gained disease compatibility with *N. benthamiana,* we were curious whether the selective pressure of ETI would also result in recovery of strains with ICE-DQ evicted. We inoculated *N. benthamiana* with *Pma*ES4236-C and *Pto*DC3000 at approximately 5×10^5 CFU/mL establishing three lineages for each strain (Lineage A, LA; Lineage B, LB; Lineage C, LC). Six to seven days post inoculation bacteria from each lineage were recovered and used to create new inoculum to passage the bacteria through *N. benthamiana* a total of five times. Twenty-two isolated clones of each passage 5 (P5) lineage were screened for altered *N. benthamiana* disease compatibility phenotypes. For *Pto*DC3000, none of the 66 P5 clones tested produced disease symptoms consistent with the *Pto*DC3000Δ*hopQ N. benthamiana* compatible strain and thus were not examined further. However, for *Pma*ES4326, 7 of 66 (2 Lineage A, 3 Lineage B, 2 Lineage C) P5 clones were able to cause disease symptoms similar to the targeted deletion strain *Pma* DQ3 (see **Figure 5A**). For a subset of clones bacterial populations were determined to correlate qualitative differences in symptoms with differences in bacterial load (see **Figure 5B**). Genome sequencing of *Pma*ES4326 clone (*Pma*ES43226 LAP5-20) confirmed a loss of ICE-DQ in this strain identical to that observed in the *Pma* DQ3 targeted deletion strain recovered after sucrose counter-selection (**Table 3** and **Figure 3**). We have further queried for additional evolutionary changes in *Pma*ES4326 LAP5-20 by using breseq (Deatherage and Barrick, 2014) to map Illumina reads to the *Pma*ES4326-C genome sequence. While it is possible, but not entirely clear, that a handful of additional changes have occurred in intergenic regions of this genome, there is no compelling data that any changes occurred in known virulence genes (see breseq output file on Figshare doi: 10.6084/m9.figshare.17064080).

Transfer of PmaICE-DQ Between Strains

Aside from being able to excise from the genome, ICE elements are also categorized by their ability to transfer across bacterial strains, and we therefore tested whether ICE-DQ could undergo conjugation to a naive strain. To do this, ICE-DQ was marked with kanamycin resistance by generating a single-crossover merodiploid with a pK18ms-derivative plasmid (*Pma* ICE-DQ:pK18). Three recipient strains, *Pto*DC3000 and two independent PmaΔICE-DQ strains (the targeted *Pma* DQ3 strain and the passage derived *Pma* LAP5-20) were labeled with a Tn*7* 3xmCherry resistant transposon to confer spectinomycin resistance as well as visually differentiate them from the donor strain. The *Pma* ICE-DQ:pK18, Km^R marked donor strain was co-cultured with each of the three Tn*7*

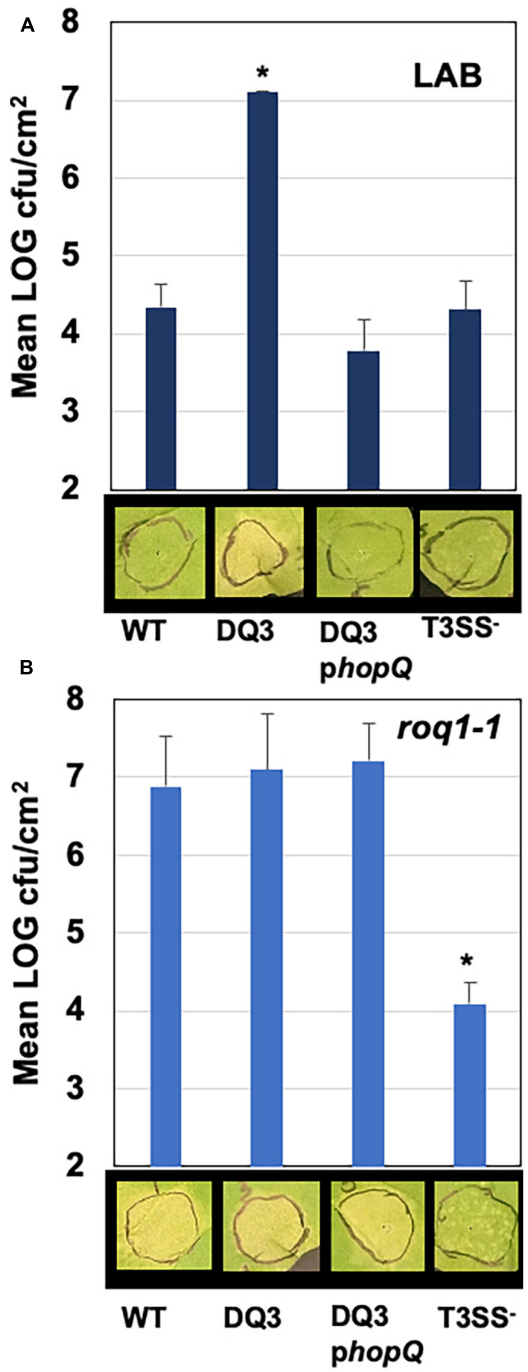


FIGURE 4 | PmaES4326 gains disease compatibility with *N. benthamiana* in the absence of *hopQ*/*Roq1*-mediated ETI. The *Pma*ES4326 WT strain, *Pma* DQ3 (ΔICE-DQ), DQ3 complemented with pCPP5372hopQ (DQ3pQ), and the non-pathogenic T3SS- strain *hrcN*:Tn*5* were infiltrated into leaves of the *N. benthamiana* Lab accession **(A)** or the *roq1-1* CRISPR mutant derivative line **(B)** at 3 × 10^4 CFU/mL. Eight days post inoculation leaves were photographed to document symptoms and bacterial load from four 4 mm diameter leaf discs from each infiltrated area by dilution plating was determined as CFUs/cm^2 leaf tissue. Values are the means and standard deviations of LOG transformed CFUs/cm^2 from three inoculated plants. *$p < 0.05$ compared to *Pma*ES4326 WT strain as determined by unpaired one-tailed *t*-test.

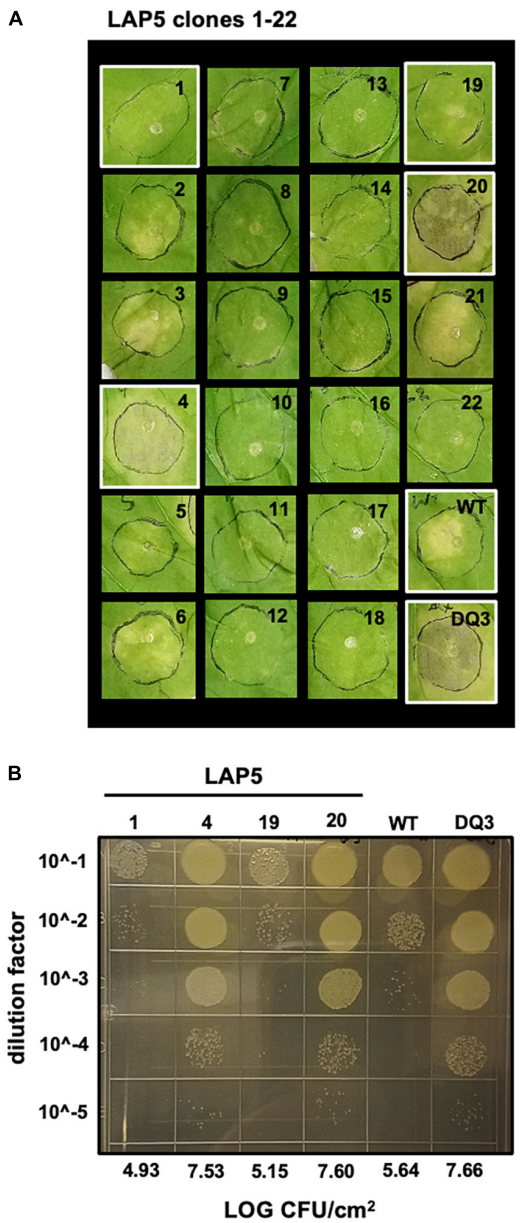


FIGURE 5 | A subset of *Pma*ES4326 isolates passaged through *N. benthamiana* gain disease compatibility. **(A)** Two isolates out of 22 *Pma*ES4326 lineage A (LA) passage 5 (P5) isolates produce disease symptoms in the *N. benthamiana* LAB accession comparable to the targeted *Pma* DQ3 (ΔICE-DQ) strain. Leaves were photographed eight days post inoculation with 3 × 10^4 CFU/mL bacterial suspensions. **(B)** Tenfold dilution series 10 μL spot plate for assessing bacterial load of select strains highlighted in white in panel A.

3xmCherry Sp^R recipients. Conjugations between these strains were then plated on kanamycin and spectinomycin selective media (**Figure 6A**). No kanamycin and spectinomycin resistant mutants or exconjugants were recovered from the *Pto*DC3000 conjugation or from the single parent control experiments. However, we were able to recover kanamycin and spectinomycin

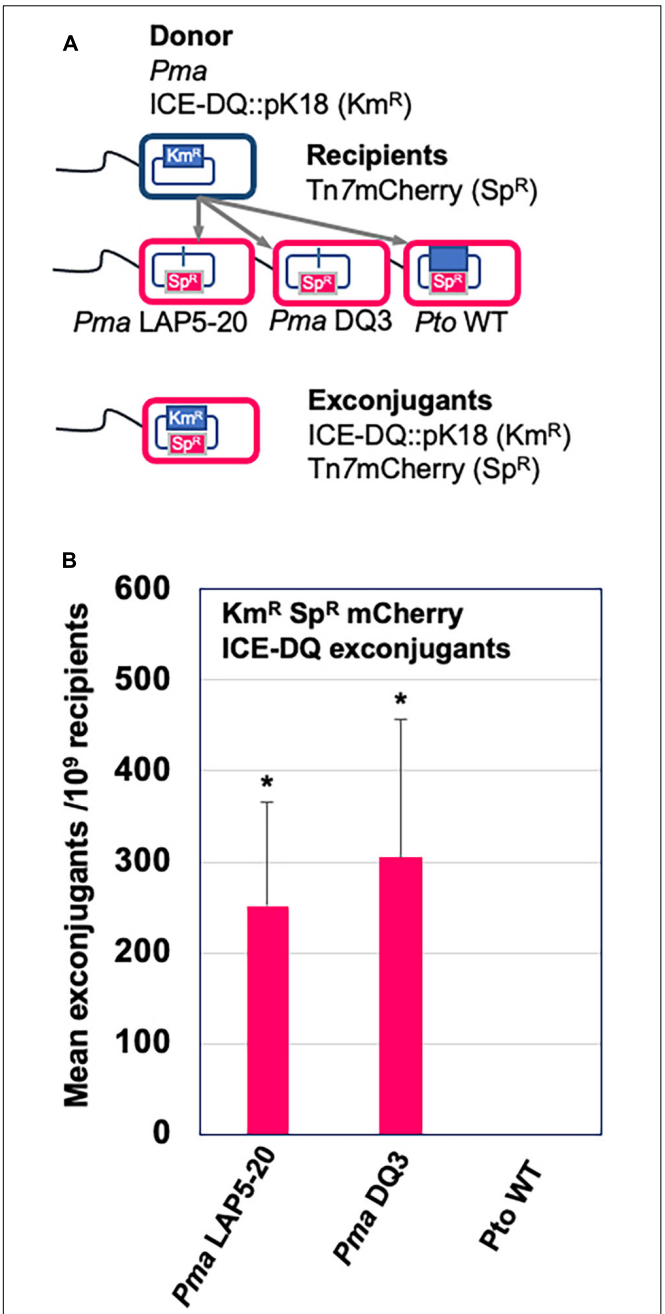


FIGURE 6 | ICE-DQ can be readily transferred into *Pma*ES4326 strains that lack the ICE element. **(A)** Schematic representation of the conjugation strategy for monitoring ICE-DQ transfer between strains. The ICE-DQ of the donor strain was marked with kanamycin resistance by single crossover integration the pK18ms plasmid (*Pma* ICE-DQ:pK18). The recipient strains *Pma* DQ3 (ΔICE-DQ targeted), *Pma* LAP5-20 (ΔICE-DQ passage-derived) and *Pto*DC3000 were marked by introduction of a spectinomycin resistance Tn*7*3xmCherry transposon. Donor and recipients were co-cultured on LM media and ICE-DQ:pK18 mCherry expressing ICE-DQ exconjugants were recovered on kanamycin and spectinomycin. **(B)** ICE-DQ:pK18 exconjugant recovery frequencies/10^9 recipients were determined based on kanamycin and spectinomycin exconjugants per spectinomycin resistant recipients as determined by dilution plating. Values are means and standard deviations of three biological replicates. *$p < 0.05$ compared to Pto WT strain as determined by unpaired one-tailed *t*-test.

resistant mCherry + exconjugants in *Pma*ES4326ΔICE-DQ strains at rates of 2.50×10^{-7} stdev $\pm$ 1.23^{-7} and 3.05×10^{-7} stdev $\pm$ 1.51^{-7} exconjugants per recipient respectively in *Pma* DQ3and *Pma* LAP5-20 recipient strains (**Figure 6B**) which strongly suggests that the ICE-DQ element is readily transmissible into *Pma*ES4326 strains lacking ICE-DQ.

DISCUSSION

Integrative Conjugative Elements (ICEs) are important drivers of evolutionary dynamics within and across bacterial populations and communities because of their potential to disseminate genes and pathways through horizontal transfer (Johnson and Grossman, 2015). Genes encoded by ICEs often impart phenotypes critical for survival under specific environmental conditions, including antibiotic resistance loci as well as phage defense systems (Botelho and Schulenburg, 2021; LeGault et al., 2021). In the phytopathogen *Pseudomonas syringae*, plasticity of ICE elements encoding genes involved in heavy metal resistance, type III effectors, and carbon metabolism has been identified as the main difference between epidemic strains causing disease across kiwi orchards in New Zealand and with previous outbreaks (Colombi et al., 2017; McCann et al., 2017; Poulter et al., 2018). Despite the accumulation of numerous complete genomes sequences for *P. syringae* and related species and the likelihood of ICEs to impart traits that affect growth in agricultural settings, to date there have been relatively few ICEs identified and vetted for this pathogen.

Placement of genes and pathways in ICEs may be especially important under conditions of fluctuating selection pressures where selection pressures on cargo genes in an ICE switch from positive to negative depending on the environment. One such scenario involves the presence of type three effector genes, which are bacterial proteins delivered from symbionts into eukaryotic host cells and which are critical for pathogenesis of *P. syringae* strains *in planta* (Grant et al., 2006). Type three effectors are particularly well studied in phytopathogens such as *P. syringae*, where they can be of exceptional benefit on some host plants by enabling strains to overcome host immune responses but may also directly trigger immune responses based on the presence of cognate immune receptors on other host genotypes (Dillon et al., 2019a; Laflamme et al., 2020). The presence of type three effectors on ICEs would enable the rapid transmission of these critical virulence genes across strains, while also enabling genes to be rapidly lost from lineages if host genotypes shift such that the effectors are recognized by R gene pathways. Extending this thought, effectors in ICEs that are recognized by hosts are potentially lost more readily through ICE element excision than through mutation (given dedicated excision machinery of the ICEs), which may facilitate adaptation if a strain encounters a resistant host background. While not an extensive experiment, our passage experiment where a subset of *Pma*ES4326 but not *Pto*DC3000 clones becomes compatible with *N. benthamiana* does demonstrate that parameter space exists for such a scenario. Moreover, effectors that are inactivated through mutation can only be reactivated by reversion mutations

or through horizontal transfer and acquisition from different strains. In contrast, if effectors are found on ICEs and lost through excision, it's possible that a small resident pool of strains containing these ICEs will remain on other host plants or throughout the environment, facilitating rapid reacquisition of effectors that could be beneficial under different contexts. Lastly, there are a plethora of type III effectors in *P. syringae* that carry out a variety of functions across different host plants and resistance backgrounds (Dillon et al., 2019a). Some of these effectors can be considered "core" and found in syntenic locations across strains while others are more variably present. It's possible that presence of effectors in ICEs could itself reflect something about the characteristics of these effectors, in that the proteins found as cargo on ICEs may be subject to higher levels of fluctuating selection across hosts than those found in the core set.

HopQ and HopD are often found together on genomic islands in *P. syringae* genomes, and have been associated with adaptation to agriculturally important crop plants (Wei et al., 2007; Baltrus et al., 2011; Monteil et al., 2016). Presence of *hopQ* in an active ICE in strain *Pma*ES4326 sets up an interesting scenario because a nearly identical version of this effector is present in a somewhat syntenic position in a relatively closely related strain *Pto*DC3000 except that this effector is not part of a functional ICE. Therefore, this genomic context suggests that *hopQ* would experience more evolutionary flexibility in *Pma*ES4326 than in *Pto*DC3000 because it can be more easily lost and/or transferred from this strain due to its presence in a region of the chromosome that can excised at a relatively high rate. To test for differences in evolutionary flexibility, we passaged replicate populations of both *Pma*ES4326 and *Pto*DC3000 in an accession of *N. benthamiana* which can recognize and mount an immune response to HopQ. As such, wild type versions of *Pma*ES4326 and *Pto*DC3000 are recognized by this accession and trigger an immune response which prevents disease. However, we found that passage of *Pma*ES4326 from plant to plant resulted in recovery of strains of *Pma*ES4326 that didn't trigger the immune response. Therefore, presence of this type III effector inside of an ICE enables rapid loss of the recognized effector under conditions of negative selection and there is relatively less flexibility for this effector to be lost in a strain where it is not present in an active ICE.

One other curious feature of the genome of *Pma*ES4326 is that it encodes two distinct but highly similar ICE elements (**Figure 1**). The configuration is particularly interesting as two distinct ICE elements have integrated successively into the *Pma*ES4326 genome and many of the genes involved in ICE functions are quite conserved between the two. It also appears as though integration of the first element recreated a tRNA with a proline anticodon and that the second ICE element can use this as a further integration point. Indeed, excision of ICE-DQ in both the targeted *Pma* DQ3 and passage-derived *Pma* LAP5-20 strains reported here also deletes one of the predicted proline tRNAs. It may therefore be no coincidence that two ICE elements that are highly similar (but which contain different cargo regions) can integrate successively into the genome as it's straightforward to imagine that highly specific recombinases would have similar target sequences.

Although not the main focus of this manuscript, we also report on plasticity of the *Pma*ES4326 genome writ large across lab derived strains (**Table 2**). As the number of complete *P. syringae* genome sequences accumulates, it is becoming more apparent that strain *Pma*ES4326 is notable compared to the rest of the species complex because it contains so many secondary replicons as well as additional elements that contribute to genomic plasticity (Stavrinides and Guttman, 2004; Stavrinides et al., 2012). Many of these plasmids appear to contain genes involved in virulence and so acquisition of these replicons through horizontal gene transfer has likely contributed to the pathogenic ability of this strain across multiple hosts (Stavrinides and Guttman, 2004). However, there is a clear downside to the large plasmid repertoire of *Pma*ES4326 as demonstrated by genome sequences reported here. Although we can't definitively catalog passage histories, genome sequences from isolates from the Dangl and Collmer labs are likely not diverged by $>$ 10 passages total. In this time, the Dangl isolate (*Pma*ES4326-D) has lost two different plasmids that were originally reported by Stavrinides et al. and which are contained in the Collmer lab isolate (*Pma*ES4326-C). Indeed, genome assemblies for the targeted ICE-DQ excision strain *Pma* DQ3 suggest that this strain has also lost at least one plasmid in addition to the ICE-DQ after only three additional passages in culture. Lastly, even though this strain was extensively surveyed for plasmids through gel electrophoresis and Sanger sequencing, we report the existence of an additional large plasmid present within all sequenced isolates of this strain.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

AUTHOR CONTRIBUTIONS

DB and BK conceived, conceptualized, analyzed the experiments, and co-wrote the manuscript. QF conducted a subset of experiments described in the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the University of Georgia Office of Research as well as the College of Agricultural and Environmental Sciences' Research Office.

ACKNOWLEDGMENTS

We would like to thank Karl Schreiber and Darrell Desveaux for providing *Pma*ES4326 *hrcN*:Tn*5* as well as Alex Schultink and Brian Staskawicz for providing the *roq1-1 N. benthamiana* line.

REFERENCES

Adlung, N., Prochaska, H., Thieme, S., Banik, A., Blüher, D., John, P., et al. (2016). Non-host resistance induced by the *Xanthomonas* effector XopQ Is Widespread within the Genus *Nicotiana* and functionally depends on EDS1. *Front. Plant Sci.* 7:1796. doi: 10.3389/fpls.2016.01796

Baltrus, D. A., McCann, H. C., and Guttman, D. S. (2017). Evolution, genomics and epidemiology of *Pseudomonas* syringae: challenges in bacterial molecular plant pathology. *Mol. Plant Pathol.* 18, 152–168. doi: 10.1111/mpp.12506

Baltrus, D. A., Nishimura, M. T., Romanchuk, A., Chang, J. H., Mukhtar, M. S., Cherkis, K., et al. (2011). Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas* syringae isolates. *PLoS Pathog.* 7:e1002132. doi: 10.1371/journal.ppat.1002132

Bardaji, L., Añorga, M., Echeverría, M., Ramos, C., and Murillo, J. (2019). The toxic guardians — multiple toxin-antitoxin systems provide stability, avoid deletions and maintain virulence genes of *Pseudomonas* syringae virulence plasmids. *Mob. DNA* 10:7. doi: 10.1186/s13100-019-0149-4

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., and Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* 10:e0128036. doi: 10.1371/journal.pone.0128036

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Botelho, J., and Schulenburg, H. (2021). The role of integrative and conjugative elements in antibiotic resistance evolution. *Trends Microbiol.* 29, 8–18. doi: 10.1016/j.tim.2020.05.011

Cazorla, F. M., Arrebola, E., Sesma, A., Pérez-García, A., Codina, J. C., Murillo, J., et al. (2002). Copper resistance in *Pseudomonas* syringae strains isolated from mango is encoded mainly by plasmids. *Phytopathology* 92, 909–916. doi: 10.1094/PHYTO.2002.92.8.909

Choi, K. H., Kumar, A., and Schweizer, H. P. (2006). A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: application for DNA fragment transfer between chromosomes and plasmid transformation. *J. Microbiol. Methods* 64, 391–397. doi: 10.1016/j.mimet.2005.06.001

Collmer, A., Badel, J. L., Charkowski, A. O., Deng, W.-L., Fouts, D. E., Ramos, A. R., et al. (2000). *Pseudomonas* syringae Hrp type III secretion system and effector proteins. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8770–8777. doi: 10.1073/pnas.97.16.8770

Colombi, E., Straub, C., Künzel, S., Templeton, M. D., McCann, H. C., and Rainey, P. B. (2017). Evolution of copper resistance in the kiwifruit pathogen *Pseudomonas* syringae pv. actinidiae through acquisition of integrative conjugative elements and plasmids. *Environ. Microbiol.* 19, 819–832. doi: 10.1111/1462-2920.13662

de Vries, S., Stukenbrock, E. H., and Rose, L. E. (2020). Rapid evolution in plant–microbe interactions – an evolutionary genomics perspective. *New Phytol.* 226, 1256–1262. doi: 10.1111/nph.16458

Deatherage, D. E., and Barrick, J. E. (2014). Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using Breseq. *Methods Mol. Biol.* 1151, 165–188. doi: 10.1007/978-1-4939-0554-6_12

Dillon, M. M., Almeida, R. N. D., Laflamme, B., Martel, A., Weir, B. S., Desveaux, D., et al. (2019a). Molecular evolution of *Pseudomonas* syringae Type III secreted effector proteins. *Front. Plant Sci.* 10:418. doi: 10.3389/fpls.2019.00418

Dillon, M. M., Thakur, S., Almeida, R. N. D., Wang, P. W., Weir, B. S., and Guttman, D. S. (2019b). Recombination of ecologically and evolutionarily significant loci maintains genetic cohesion in the *Pseudomonas* syringae species complex. *Genome Biol.* 20, 1–28. doi: 10.1186/s13059-018-1606-y

Dong, O. X., and Ronald, P. C. (2019). Genetic engineering for disease resistance in plants: recent progress and future perspectives. *Plant Physiol.* 180, 26–38. doi: 10.1104/pp.18.01224

Godfrey, S. A. C., Lovell, H. C., Mansfield, J. W., Corry, D. S., Jackson, R. W., and Arnold, D. L. (2011). The stealth episome: suppression of gene expression on the excised genomic island PPHGI-1 from *Pseudomonas* syringae pv. phaseolicola. *PLoS Pathog.* 7:e1002010. doi: 10.1371/journal.ppat.1002010

Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M., and Dangl, J. L. (2006). Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* 60, 425–449. doi: 10.1146/annurev.micro.60.080805.142251

Jayaraman, J., Chatterjee, A., Hunter, S., Chen, R., Stroud, E. A., Saei, H., et al. (2021). Rapid methodologies for assessing *Pseudomonas* syringae pv. actinidiae colonization and effector-mediated hypersensitive response in Kiwifruit. *Mol. Plant Microbe Interact.* 34, 880–890. doi: 10.1094/MPMI-02-21-0043-R

Johnson, C. M., and Grossman, A. D. (2015). Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.* 49, 577–601. doi: 10.1146/annurev-genet-112414-055018

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329.

Kvitko, B. H., and Collmer, A. (2011). Construction of *Pseudomonas* syringae pv. tomato DC3000 mutant and polymutant strains. *Methods Mol. Biol.* 712, 109–128. doi: 10.1007/978-1-61737-998-7_10

Kvitko, B. H., Park, D. H., Velásquez, A. C., Wei, C.-F., Russell, A. B., Martin, G. B., et al. (2009). Deletions in the Repertoire of *Pseudomonas* syringae pv. tomato DC3000 Type III secretion effector genes reveal functional overlap among effectors. *PLoS Pathog.* 5:e1000388. doi: 10.1371/journal.ppat.1000388

Laflamme, B., Dillon, M. M., Martel, A., Almeida, R. N. D., Desveaux, D., and Guttman, D. S. (2020). The pan-genome effector-triggered immunity landscape of a host-pathogen interaction. *Science* 367, 763–768. doi: 10.1126/science.aax4079

LeGault, K. N., Hays, S. G., Angermeyer, A., McKitterick, A. C., Johura, F. T., Sultana, M., et al. (2021). Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science* 373, eabg2166. doi: 10.1126/science.abg2166

Lovell, H. C., Mansfield, J. W., Godfrey, S. A. C., Jackson, R. W., Hancock, J. T., and Arnold, D. L. (2009). Bacterial evolution by genomic Island transfer occurs via DNA transformation in planta. *Curr. Biol.* 19, 1586–1590. doi: 10.1016/j.cub.2009.08.018

McCann, H. C., Li, L., Liu, Y., Li, D., Pan, H., Zhong, C., et al. (2017). Origin and evolution of the Kiwifruit canker pandemic. *Genome Biol. Evol.* 9:932. doi: 10.1093/gbe/evx055

Monteil, C. L., Yahara, K., Studholme, D. J., Mageiros, L., Méric, G., Swingle, B., et al. (2016). Population-genomic insights into emergence, crop adaptation and dissemination of *Pseudomonas* syringae pathogens. *Microb. Genom.* 2:e000089. doi: 10.1099/mgen.0.000089

Neale, H. C., Jackson, R. W., Preston, G. M., and Arnold, D. L. (2018). Supercoiling of an excised genomic island represses effector gene expression to prevent activation of host resistance. *Mol. Microbiol.* 110, 444–454. doi: 10.1111/mmi.14111

Poulter, R. T. M., Ho, J., Handley, T., Taiaroa, G., and Butler, M. I. (2018). Comparison between complete genomes of an isolate of *Pseudomonas* syringae pv. actinidiae from Japan and a New Zealand isolate of the pandemic lineage. *Sci. Rep.* 8:10915. doi: 10.1038/s41598-018-29261-5

Qi, D., Dubiella, U., Kim, S. H., Sloss, D. I., Dowen, R. H., Dixon, J. E., et al. (2014). Recognition of the protein kinase AVRPPHB SUSCEPTIBLE1 by the disease resistance protein RESISTANCE TO *PSEUDOMONAS* SYRINGAE5 is dependent on s-acylation and an exposed loop in AVRPPHB SUSCEPTIBLE1. *Plant Physiol.* 164, 340–351. doi: 10.1104/pp.113.227686

Qi, T., Seong, K., Thomazella, D. P. T., Kim, J. R., Pham, J., Seo, E., et al. (2018). NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in Nicotiana benthamiana. *Proc. Natl. Acad. Sci. U.S.A.* 115, E10979–E10987. doi: 10.1073/pnas.1814856115

Schierstaedt, J., Bziuk, N., Kuzmanović, N., Blau, K., Smalla, K., and Jechalke, S. (2019). Role of plasmids in plant-bacteria interactions. *Curr. Issues Mol. Biol.* 30, 17–38. doi: 10.21775/cimb.030.017

Schultink, A., Qi, T., Lee, A., Steinbrenner, A. D., and Staskawicz, B. (2017). Roq1 mediates recognition of the *Xanthomonas* and *Pseudomonas* effector proteins XopQ and HopQ1. *Plant J.* 92, 787–795. doi: 10.1111/tpj.13715

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sjödin, P., Kaj, I., Krone, S., Lascoux, M., and Nordborg, M. (2005). On the meaning and existence of an effective population size. *Genetics* 169, 1061–1070. doi: 10.1534/genetics.104.026799

Stavrinides, J., and Guttman, D. S. (2004). Nucleotide sequence and evolution of the five-plasmid complement of the phytopathogen *Pseudomonas* syringae pv.

maculicola ES4326. *J. Bacteriol.* 186, 5101–5115. doi: 10.1128/JB.186.15.5101-5115.2004

Stavrinides, J., Kirzinger, M. W. B., Beasley, F. C., and Guttman, D. S. (2012). E622, a miniature, virulence-associated mobile element. *J. Bacteriol.* 194:509. doi: 10.1128/JB.06211-11

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.

Teal, T. K., Lies, D. P., Wold, B. J., and Newman, D. K. (2006). Spatiometabolic Stratification of *Shewanella oneidensis* biofilms. *Appl. Environ. Microbiol.* 72:7324. doi: 10.1128/aem.01163-06

Wei, C. F., Kvitko, B. H., Shimizu, R., Crabill, E., Alfano, J. R., Lin, N. C., et al. (2007). A *Pseudomonas* syringae pv. tomato DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant *Nicotiana benthamiana*. *Plant J.* 51, 32–46. doi: 10.1111/j.1365-313X.2007.03126.x

Wei, H. L., Zhang, W., and Collmer, A. (2018). Modular study of the Type III effector repertoire in *Pseudomonas* Syringae Pv. Tomato DC3000 reveals a matrix of effector interplay in pathogenesis. *Cell Rep.* 23, 1630–1638. doi: 10.1016/j.celrep.2018.04.037

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:e1005595. doi: 10.1371/journal.pcbi.1005595

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 02 March 2022
doi: 10.3389/fmicb.2021.817815

Complete Genome Assemblies of All *Xanthomonas translucens* Pathotype Strains Reveal Three Genetically Distinct Clades

*Florian Goettelmann[1], Veronica Roman-Reyna[2,3], Sébastien Cunnac[4], Jonathan M. Jacobs[2,3], Claude Bragard[5], Bruno Studer[1], Ralf Koebnik[4] and Roland Kölliker[1]**

[1] Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zürich, Zurich, Switzerland, [2] Department of Plant Pathology, The Ohio State University, Columbus, OH, United States, [3] Infectious Diseases Institute, The Ohio State University, Columbus, OH, United States, [4] Plant Health Institute of Montpellier, University of Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France, [5] Earth and Life Institute, UCLouvain, Louvain-la-Neuve, Belgium

OPEN ACCESS

Edited by:
Feng Gao, Tianjin University, China

Reviewed by:
Meng Yuan, Huazhong Agricultural University, China
Toni A. Chapman, New South Wales Department of Primary Industries, Australia

***Correspondence:**
Roland Kölliker
roland.koelliker@usys.ethz.ch

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *18 November 2021*
Accepted: *29 December 2021*
Published: *02 March 2022*

Citation:
Goettelmann F, Roman-Reyna V, Cunnac S, Jacobs JM, Bragard C, Studer B, Koebnik R and Kölliker R (2022) Complete Genome Assemblies of All Xanthomonas translucens Pathotype Strains Reveal Three Genetically Distinct Clades. Front. Microbiol. 12:817815. doi: 10.3389/fmicb.2021.817815

The *Xanthomonas translucens* species comprises phytopathogenic bacteria that can cause serious damage to cereals and to forage grasses. So far, the genomic resources for *X. translucens* were limited, which hindered further understanding of the host–pathogen interactions at the molecular level and the development of disease-resistant cultivars. To this end, we complemented the available complete genome sequence of the *X. translucens* pv. *translucens* pathotype strain DSM 18974 by sequencing the genomes of all the other 10 *X. translucens* pathotype strains using PacBio long-read technology and assembled complete genome sequences. Phylogeny based on average nucleotide identity (ANI) revealed three distinct clades within the species, which we propose to classify as clades Xt-I, Xt-II, and Xt-III. In addition to 2,181 core *X. translucens* genes, a total of 190, 588, and 168 genes were found to be exclusive to each clade, respectively. Moreover, 29 non-transcription activator-like effector (TALE) and 21 TALE type III effector classes were found, and clade- or strain-specific effectors were identified. Further investigation of these genes could help to identify genes that are critically involved in pathogenicity and/or host adaptation, setting the grounds for the development of new resistant cultivars.

Keywords: *Xanthomonas translucens*, complete genomes, phylogeny, host adaptation, comparative genomics, virulence factors

INTRODUCTION

Xanthomonas translucens is a species of Gram-negative phytopathogenic bacteria that causes serious damage in gramineous plants. To date, a total of 11 pathovars have been defined based on their host range, each represented by one strain referred to as the pathotype strain and deposited in the appropriate bacterial strain collections. Two major groups are historically distinguished within the species: the "translucens" group and the "graminis" group. The translucens group consists of pathovars *cerealis*, *hordei*, *secalis*, *translucens*, and *undulosa*, which cause leaf streak and black chaff in economically important cereals such as wheat (*Triticum* spp.) and barley

(*Hordeum vulgare*) (Sapkota et al., 2020). The graminis group consists of pathovars *arrhenatheri*, *graminis*, *phlei*, *phleipratensis*, and *poae*, which cause bacterial wilt in forage grasses (Egli et al., 1975; Egli and Schmidt, 1982; Stead, 1989). Although recent genome sequence data loosely supported this phenotypic classification (Peng et al., 2016; Hersemann et al., 2017; Langlois et al., 2017; Shah et al., 2021), these groups are not clearly defined based on the genetic relationships between the strains. Moreover, genomic studies have previously suggested that pv. *cerealis* could be genetically distinct from the other "translucens" group pathovars (Bragard et al., 1997; Rademaker et al., 2006; Peng et al., 2016; Langlois et al., 2017). Furthermore, *Xanthomonas* strains causing dieback in pistachio (*Pistacia vera* L.) were found to belong to the *X. translucens* species, a species so far thought to infect only monocotyledonous plants (Giblot-Ducray et al., 2009). These strains were classified as pv. *pistaciae*, in which two groups were distinguished by rep-PCR and *gyrB*-based phylogeny and are referred to as groups A and B (Facelli et al., 2005; Marefat et al., 2006; Giblot-Ducray et al., 2009). The group A strain was found to be closely related to pathovars *translucens*, *secalis*, and *undulosa*, while the group B strain was closest to pv. *cerealis*, suggesting that the ability to infect pistachio may have evolved two separate times in *X. translucens*.

X. translucens pathovars were initially classified solely based on their host range as pathovars of the *X. campestris* species. However, a classification based on DNA–DNA hybridization was proposed by Vauterin et al. (1995), who amended these pathovars of *X. campestris* to the species level, forming the *X. translucens* species. Nonetheless, though these pathovars are now considered pathovars of the *X. translucens* species, the definition of these pathovars is still based on their host range as before. Phylogeny is crucial to better understand the evolutionary history of each pathovar, in order to determine how each strain has adapted to infect its hosts. This knowledge is key in defining targets for breeding resistant cultivars. The current definition of pathovars of *X. translucens* might not properly reflect these processes of host adaptation, and a better definition based on the genetic relationships between the strains would improve the comprehension of the host–pathogen relationships. Recent advances in sequencing technology now allow for the rapid sequencing of complete genomes of bacterial strains, which provide a better basis for phylogenetic analyses. At the time of this study, complete genome sequences were publicly available for pv. *cerealis* strain NXtc01 (Shah et al., 2019); pv. *translucens* strains DSM 18974 (Jaenicke et al., 2016), XtKm7, XtKm8, XtKm9, XtKm33, XtKm34 (Shah et al., 2021); and pv. *undulosa* strains XT4699 (Peng et al., 2016), ICMP 11055 (Falahi Charkhabi et al., 2017), LW16, P3 (Peng et al., 2019), XtFa1, XtLr8, XtKm12, and XtKm15 (Shah et al., 2021). Though the number of available complete genome sequences is continuously growing, these sequences represent only three pathovars to date, with a complete genome sequence being available for only one pathotype strain.

High-quality genome sequences also help in the identification of genes that are directly linked to the pathogen's virulence and host range. Indeed, successful infection by pathogenic bacteria is mediated by a set of virulence factors, including degradative enzymes and effector proteins. These are often host specific, and the repertoire of virulence factors that a bacterial strain possesses defines which hosts it is able to infect. In *Xanthomonas* species, virulence factors generally depend on the type II and type III secretion systems (T2SS and T3SS, respectively). These secretion systems allow the bacteria to export virulence factors in order to enable or facilitate their proliferation and survival in the host by targeting specific host components (Büttner and Bonas, 2010; Alvarez-Martinez et al., 2021). Identifying which virulence factors are involved in the pathogen's virulence and which host components are the targets of these virulence factors is crucial, as these host components can then be the focus of resistance breeding.

The T3SS is encoded by the "hypersensitive reaction and pathogenicity" (*hrp*) gene cluster (Bonas et al., 1991). Some *hrp* genes that are highly conserved are referred to as "*hrp*-conserved" (*hrc*) genes (Bogdanove et al., 1996). Moreover, some genes in the cluster that are involved in, but not necessary to, the host–plant interaction are called "*hrp*-associated" (*hpa*) genes. In the *X. translucens* species, the core *hrp* cluster consists of 23 genes, with 8 *hrp* genes, 11 *hrc* genes, and 4 *hpa* genes (Wichmann et al., 2013; Pesce et al., 2017).

This secretion system injects effector proteins into the host cell. In *Xanthomonas*, these effectors are generally called "*Xanthomonas* outer proteins" (Xop), with 53 classes from XopA to XopBA (White et al., 2009). Other effectors are named according to their avirulence characteristics, causing hypersensitive response in the host, such as AvrBs1 to AvrBs3. These effectors are key virulence factors, as, when translocated into the host cell, they are able to target the different pathways of the host, allowing the pathogen, for example, to acquire nutrients or to evade or suppress host defenses. A specific type of effectors secreted by the T3SS are the "transcription activator-like" effectors (TALEs). Their amino acid sequences contain highly conserved repetitive sequences of ∼34 amino acids, with only the 12th and 13th residues being hypervariable and referred to as the "repeat variable di-residue" (RVD). The RVD array of each TALE allows it to bind to specific nucleotide sequences in the host DNA, thus activating the expression of the neighboring genes to the pathogen's advantage (Streubel et al., 2017; Wang et al., 2017).

Due to their repetitive sequences, TALE genes are difficult to assemble using short-read sequencing technology. Complete genome sequences based on either long-read sequencing, a very high coverage of short-read sequencing, or a mixture of both, are thus necessary to properly identify these effectors. Such high-quality genome sequences allowed the identification of eight and five TALE genes in pv. *translucens* strains DSM 18974 and UPB886, respectively (Jaenicke et al., 2016; Roman-Reyna et al., 2020), eight and seven in pv. *undulosa* strains XT4699 and ICMP 11055, respectively (Peng et al., 2016; Falahi Charkhabi et al., 2017), and two in pv. *cerealis* strains CFBP 2541 and NXtc01 (Pesce et al., 2015; Shah et al., 2019). However, of these, only four were functionally characterized and found to play a role in virulence to date (Falahi Charkhabi et al., 2017; Peng et al., 2019; Shah et al., 2019). No TALE has yet been identified in other pathovars of *X. translucens*.

Similar to the T3SS, the T2SS is responsible for the export of virulence factors, most of which are cell wall-degradative enzymes, into the host apoplasm (Jha et al., 2005). In *Xanthomonas*, two types of clusters encoding the T2SS can be found. The *xps* cluster is conserved across *Xanthomonas* species, while the *xcs* cluster is only found in some species such as *X. citri* and *X. campestris* (Szczesny et al., 2010).

In addition to the T2SS and T3SS, the type IV and type VI secretion systems (T4SS and T6SS, respectively) also secrete proteins that may affect the virulence of the pathogen. However, while the T2SS and the T3SS target host components, the T4SS and the T6SS are involved in the defense against microbial predators such as amoeba, as well as in the competition with other microorganisms (Büttner and Bonas, 2010; Alvarez-Martinez et al., 2021). The T4SS is evolutionarily related to bacterial conjugation systems and is involved in the competition with other bacteria by injecting protein effectors or protein–DNA complexes into their cells (Sgro et al., 2019). The T6SS is related to the tail of bacteriophages and, similarly to the T4SS, is able to inject effector proteins into prokaryotic as well as eukaryotic cells (Bayer-Santos et al., 2019). In the Xanthomonadales order, three subtypes of T6SS have been found: subtypes 1, 3, and 4. Moreover, subtype 3 is further subdivided into subgroups 3*, 3**, and 3***. Although not directly related to virulence, the T4SS and T6SS could be key elements in the proliferation and survival of the bacteria on the host plant (Souza et al., 2015; Choi et al., 2020).

In this study, in order to complement the available complete genome sequence of the pv. *translucens* pathotype strain DSM 18974 (Jaenicke et al., 2016), we sequenced the whole genomes of all the other 10 pathotype strains of the *X. translucens* species, as well as a representative strain of the *X. translucens* pv. *pistaciae* group B, to produce high-quality genome sequences. Using these, we built a phylogeny of these strains to clarify the taxonomy of the *X. translucens* species. We then scrutinized the genomes for major virulence features of these strains to identify genes that might be important for pathogenicity and in defining their host range.

MATERIALS AND METHODS

Bacterial Strains, Growth Conditions, and DNA Extraction

The relevant data for all *X. translucens* strains used in this study are listed in **Table 1**. The genome sequence of *X. translucens* pv. *translucens* strain DSM 18974 was retrieved from the National Center for Biotechnology Information (NCBI) GenBank database (accession number LT604072). Strains LMG 726, LMG 727, LMG 728, LMG 730, LMG 843, and UPB458 were grown at 28°C on YDC agar medium (2% dextrose, 1% yeast extract, 2% CaCO3, 1.5% agar) for 48 h. Bacteria were then dissolved in 10 ml washing buffer (50 mM TRIS-HCl pH 8.0, 50 mM Ethylenediaminetetraacetic acid (EDTA) pH 8.0, 150 mM NaCl). The genomic DNA was then extracted with the NucleoSpin® Microbial DNA kit (Macherey Nagel, Duren, Germany), according to the manufacturer's recommendations. Strains CFBP 2055, CFBP 2539, CFBP 2541, and CFBP 8304 were grown at 28°C on PSA medium (0.5% peptone, 2% sucrose, 1.5% agar) for 24 h. Bacteria were then resuspended in 10 mM $MgCl_2$ and diluted to an optical density at 600 nm of 1.0. Cells from 2 ml were harvested by centrifugation and washed once with 10 mM $MgCl_2$, and genomic DNA was isolated using QIAGEN Genomic tip 100/G (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. The genomic DNA from strain ICMP 16317 was extracted following a standard phenol/chloroform method (Booher et al., 2015).

Sequencing, Genome Assembly, and Annotation

The library preparation and DNA sequencing of strains LMG 726, LMG 727, LMG 728, LMG 730, LMG 843, and UPB458 were done at the Functional Genomics Center Zurich. For these strains, as well as for ICMP 16317, libraries were prepared and multiplexed with the PacBio SMRTbell® Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, United States) according to the published protocol[1]. Multiplex library preparation in pools of eight strains, including strains CFBP 2055, CFBP 2539, CFBP 2541, and CFBP 8304, and simultaneous sequencing of eight strains on one SMRTCell was conducted at the GENTYANE genotyping platform (INRA Clermont-Ferrand, France). All strains were sequenced with the PacBio Sequel technology.

Genomic sequences were then *de novo* assembled with Flye 2.7 for strains CFBP 2055, CFBP 2539, CFBP 2541, and CFBP 8304; Flye 2.8.1 for LMG 726, LMG 728, LMG 730, and UPB458; Flye 2.9 for LMG 727, and LMG 843; and HGAP 4 for ICMP 16317 (Chin et al., 2013; Kolmogorov et al., 2019). Assemblies produced with Flye were done using the "--plasmids --iterations 2" parameters. Genomes were functionally annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) or with Prokka 1.13 (Seemann, 2014).

Phylogeny and Comparative Analysis

Average nucleotide identity (ANI) was calculated using Pyani 0.2.11 with default parameters (Pritchard et al., 2016), and a phylogeny dendrogram was constructed using Ward's hierarchical clustering method in R version 4.1.0. The gene content of the strains used in this study was compared with Roary 3.7.0 using the Prokka annotation with default parameters (Page et al., 2015). To identify genomic rearrangements and conserved genomic regions within the three clades that we defined, we compared the genomic structure of the complete genome sequences used in this study with Mauve v20150226 (Darling, 2004).

The presence of T2SS, T3SS, and T4SS was determined by tBLASTn using the amino acid sequences of the main components of each cluster retrieved from UniProt as query (**Supplementary Dataset 1**) and the genomic sequences of each strain as subject, with a 0.01 *e*-value threshold. The presence of each secretion system was then validated by looking for the presence of the cluster in the PGAP annotation at the predicted locus. The presence of T6SS was determined using SecReT6 3.0

[1]https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-Multiplexed-Microbial-Libraries-Using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf

TABLE 1 | *X. translucens* pathotype strains used in this study and characteristics of the obtained genome assemblies.

Pathovar	Strain	Country	Isolated from	Origin[a]	Coverage	Assembly length	Plasmid length	Accession
arrhenatheri	LMG 727	Switzerland	*Arrhenatherum elatius*	BCCM	103	4,843,101	NA	CP086333
cerealis	CFBP 2541	United States	*Bromus inermis*	CFBP	404	4,504,942	NA	CP074364
graminis	LMG 726	Switzerland	*Dactylis glomerata*	BCCM	81	4,678,781	NA	CP076254
hordei	UPB458[b]	India	*Hordeum vulgare*	BCCM	94	4,679,124	NA	CP076249
phlei	LMG 730	Norway	*Phleum pratense*	BCCM	87	4,569,024	NA	CP076251
phleipratensis	LMG 843	United States	*Phleum pratense*	BCCM	121	4,902,099	NA	CP086332
pistaciae (group A)	CFBP 8304	Australia	*Pistacia vera*	CFBP	301	4,599,174	NA	CP074365
pistaciae (group B)	ICMP 16317	Australia	*Pistacia vera*	ICMP	161	4,386,175	NA	CP083804
poae	LMG 728	Switzerland	*Poa trivialis*	BCCM	86	4,792,655	NA	CP076250
secalis	CFBP 2539	Canada	*Secale cereale*	CFBP	186	4,565,955	NA	CP074363
translucens	DSM 18947[c]	United States	*Hordeum vulgare*	DSMZ	223	4,715,357	NA	LT604072
undulosa	CFBP 2055	Canada	*Triticum turgidum* subsp. *durum*	CFBP	379	4,653,288	46,036	CP074361-CP074362

[a] *DSMZ, German Collection of Microorganisms and Cell Cultures; CFBP, Collection of Plant Pathogenic Bacteria; ICMP, International Collection of Microorganisms from Plants; BCCM, Belgian Coordinated Collections of Microorganisms; UPB, Collection of plant pathogenic bacteria, Earth and Life Institute, UCLouvain.*
[b] *UPB458 is the name of the strain in the UPB collection from where it was obtained and corresponds to LMG 737 in the BCCM collection.*
[c] *Jaenicke et al. (2016).*

with default settings (Li et al., 2015). The T3SS gene cluster was analyzed further by comparing the sequences of the PGAP-annotated genes using Clinker 0.0.21 with default parameters (Gilchrist and Chooi, 2021). In strains CFBP 8304 and LMG 726, the *xopF* gene found in the cluster was incorrectly annotated with PGAP, but was correctly annotated with Prokka, and was thus manually corrected in the PGAP annotation for the cluster comparison. The same was true for the *xopM* gene in strains UPB458, CFBP 8304, and LMG 728.

The presence of type III effectors was determined by BLASTp using the amino acid sequences of the effectors retrieved from http://xanthomonas.org/ as query (**Supplementary Dataset 2**) and the amino acid sequences of the genes annotated with Prokka as subject, with a 0.01 *e*-value threshold. Only hits that had > 30% identity over > 70% query sequence length were retained. To validate the presence of each effector, the amino acid sequences of the selected genes were then extracted and used as a query for a BLASTp against the type III effectors' sequences as subject. Hits that had > 30% identity over > 70% query sequence length were then considered to be putative type III effectors, as discussed in this article. The presence of TALEs, their RVD sequence, and their classification were determined using AnnoTALE 1.5 (Grau et al., 2016).

RESULTS

Genome Assembly

For all the sequenced strains, sequence coverages between 81- and 404-fold were obtained. This allowed to assemble complete genome sequences, consisting of one single circular chromosome for all strains (**Table 1**). Genome sizes ranged from 4,386,175 bp for ICMP 1317 to 4,902,099 for LMG 843. In strain CFBP 2055, an additional circular contig of 46,036 bp was assembled. A comparison by BLAST to the NCBI non-redundant nucleotide sequences database showed homology to plasmid sequences. This contig was thus considered to represent a plasmid of pv. *undulosa* strain CFBP 2055.

Phylogeny and Comparative Analysis

Phylogeny based on ANI revealed that though all strains shared > 95% ANI, three distinct groups could be observed, with ANI values above 97% (**Figure 1**). The first group consisted of strains CFBP 2055, CFBP 2539, UPB458, DSM 18794, and CFBP 8304; the second consisted of strains ICMP 16317 and CFBP 2541, and the third consisted of strains LMG 730, LMG 843, LMG 728, LMG 727, and LMG 726. These three groups will thereafter be referred to as clades Xt-I, Xt-II, and Xt-III, respectively. These three clades could also be observed when including all the publicly available complete genome sequences of *X. translucens* strains, with ANI values above 97% (**Supplementary Figure 1**). Interestingly, the two strains regarded as pv. *pistaciae* were genetically distinct and were found in two separate clades.

A comparison of the genomic structure within each clade showed 34–48 locally collinear blocks (LCB) in clade Xt-I (**Figure 2A**), 11 LCB in clade Xt-II (**Figure 2B**), and 175–312 LCB in clade Xt-III (**Figure 2C**). These LCB correspond to genomic regions that are conserved between the compared strains, showing no rearrangement. Thus, the low number of LCB in clade Xt-II showed that there are very few genomic rearrangements between the two strains of the clade, while there are more rearrangements in clade Xt-I and the most in clade Xt-III.

The number of genes found in each strain with the Prokka annotation ranged from 3,735 genes in ICMP 16317 to 4,262 genes in LMG 726. The pangenome of the 12 *X. translucens* strains used in this study consisted of 9,772 genes, while the core genome consisted of 2,181 genes (**Figure 3**). A total of 190 genes were exclusive to clade Xt-I, 588 genes to clade Xt-II, and 168 genes to clade Xt-III. Moreover, a total of 3,681 genes were

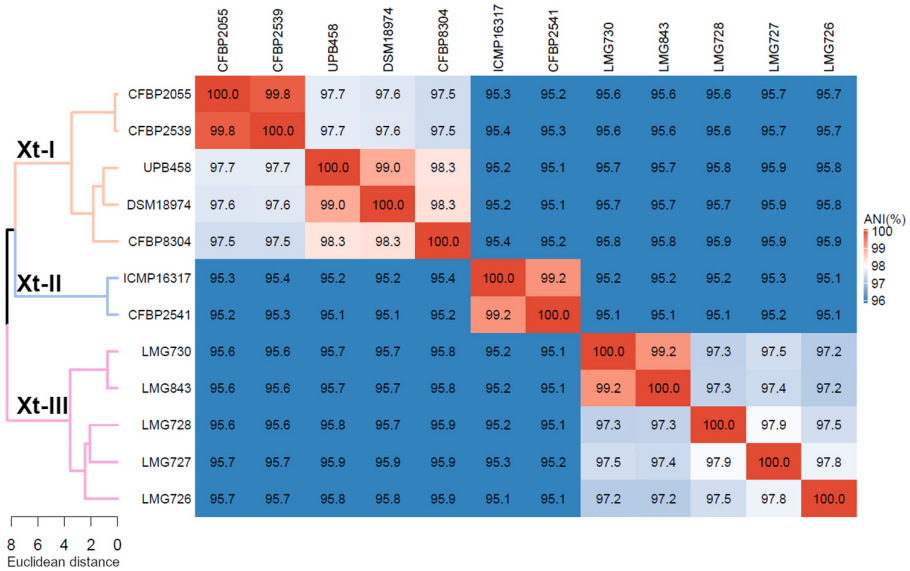


FIGURE 1 | Average nucleotide identity (ANI) of *X. translucens* pathotype strains and ANI-based phylogeny constructed with Ward's hierarchical clustering method. Distance in the dendrogram represents ANI dissimilarity between nodes. Orange: clade Xt-I, blue: clade Xt-II, pink: clade Xt-III. ANI is depicted as a gradient from blue (< 96%) to white (98%) to red (100%).

exclusive to one strain only, ranging from 48 genes in strain CFBP 2539 to 966 genes in LMG 726.

In all strains, an *xps* T2SS and a T3SS were identified, while no *xcs* T2SS was identified in any strain (**Figure 4A**). A T6SS-i3*** was identified in all strains of clades Xt-I and Xt-II, while a T6SS-i4 was identified in strains CFBP 2055, CFBP 2539, UPB458, ICMP 16317, CFBP 2541, and LMG 843. A T4SS was identified in strains DSM 18974, CFBP 8304, LMG 730, LMG 728, and LMG 727 which all lacked a T6SS-i4. The strain LMG 726 was the only one that did not possess a T4SS or a T6SS.

A direct comparison of the *hrp* cluster showed that all strains share the same genetic organization of the cluster (**Figure 5**). Moreover, most of the main components of the T3SS, from *hrcC* to *hrpD*, as well as *hpaH* and *xopF* were very conserved, with > 80% identity across the species. However, the *hrpE* structural component was the most variable, with as little as 60% identity within clades Xt-I and Xt-III (**Figure 6**). In strains LMG 728 and LMG 726, an additional gene was found between *hrpX* and *hrcT*, in opposite directions in each strain, but showing 100% identity. This gene showed a high identity with insertion sequence (IS) 5 family transposases found in other *Xanthomonas* species by BLASTx against the NCBI non-redundant protein sequence database. Moreover, in strains LMG 727 and LMG 726, one additional gene was predicted between *hpaB* and *hrpG*, with 61% identity between the two strains. However, no similarity to a known gene was found by BLASTx. Additionally, *hpaC* was found to be truncated in DSM 18974 due to an early stop codon, but still showed between 71 and 86% identity with the other strains (**Figure 6**).

The T3SS-associated components upstream and downstream of the core *hrp* cluster (*hpaT*, *hpa1*, *xopM*, and *hgiB*) showed more variability. The putative translocon *hpaT* was very conserved between the two clade Xt-II strains and showed between 63 and 89% identity in clade Xt-I (**Figure 6**). It was much more variable in clade Xt-III, where it showed between 39 and 67% identity. The second putative translocon component, *hpa1*, was very conserved among the two clade Xt-II strains (98% identity) and among the UPB458, DSM 18974, and CFBP 8304 strains in clade Xt-I. It was, however, more dissimilar between these three strains and the other two of the clade, with 45–47% identity. Within the Xt-III clade, there was a high variability, with 50–79% identity between strains. Moreover, one additional gene was predicted between *hpaT* and *hpa1* in UPB458, but no similarity to any known gene was found by BLASTx.

In clade Xt-I, the XopM effector was only present in UPB458, DSM 18974, and CFBP 8304 and was 99% identical between the three strains. The effector was also found in LMG 728, where it was 92–93% identical to the ones found in clade Xt-I. In the other four strains of clade Xt-III, the gene was disrupted by early stop codons. The *hgiB* gene was also very variable, with as low as 43% identity within clade Xt-I, and 37% identity within clade Xt-III, but with 99% identity between the two clade Xt-II strains. In LMG 728, it was found to be truncated. Furthermore, an additional gene was found between *xopM* and *hgiB* in LMG 726, which showed 100% identity to an IS4 family transposase by BLASTx.

Together with XopF and XopM, a total of 29 putative non-TALE type III effector classes were found to be present in the 12 *X. translucens* strains used in this study, ranging from 21 effectors in LMG 726 to 31 in CFBP 2541 and DSM 18974 (**Figure 4B**). Among the predicted type III effectors, AvrBs2, as well as effectors of classes XopC, XopF, XopK, XopL, XopN, XopP, XopQ, XopR, XopV, XopX, XopZ, and XopAM were found in all strains. The effectors of classes XopB, XopG, XopE, XopAA, and XopAF were conserved in all strains of clades Xt-I and Xt-II

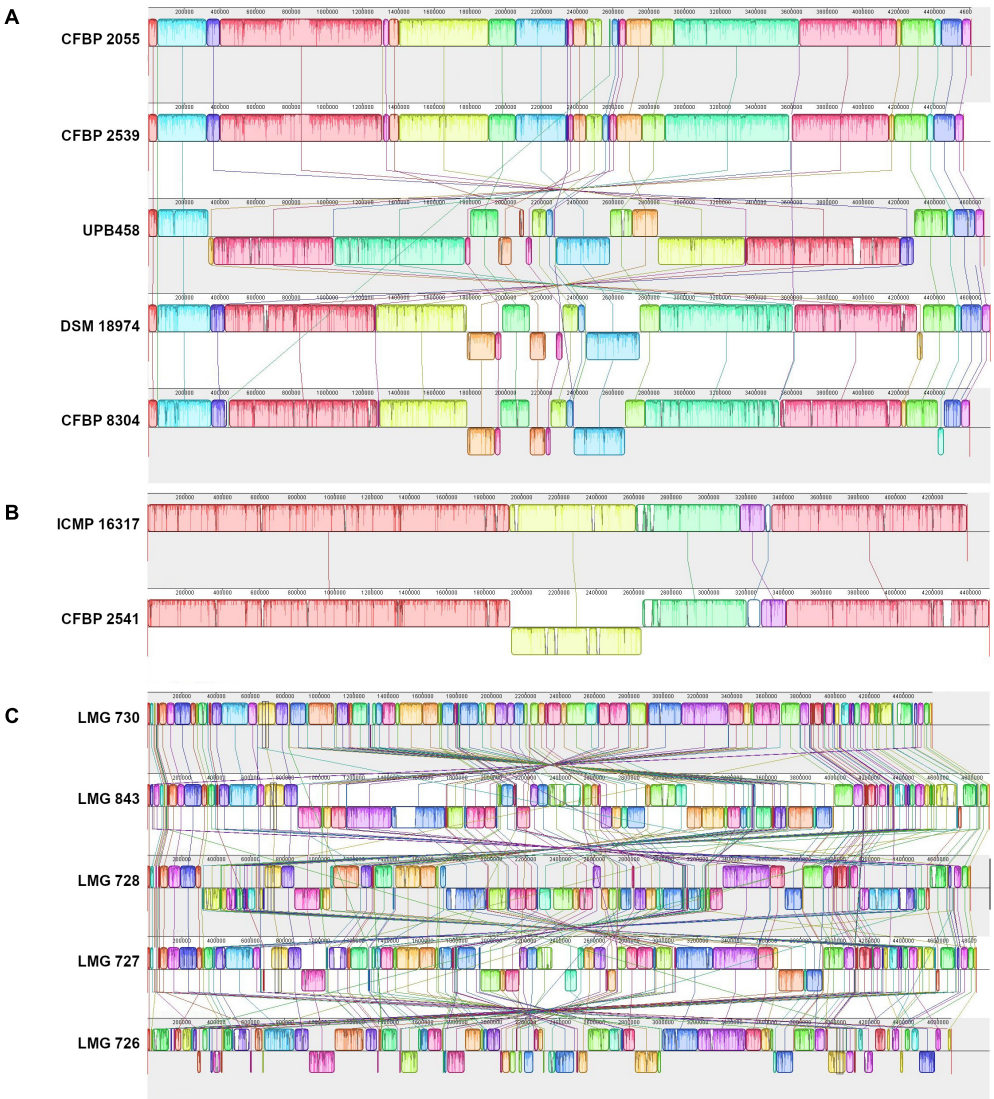


FIGURE 2 | Pairwise comparisons of the genomic structure within the three *X. translucens* clades with progressive Mauve (Darling, 2004). **(A)** Clade Xt-I, **(B)** clade Xt-II, **(C)** clade Xt-III. Colors represent conserved genomic regions (locally collinear blocks, LCBs), i.e., regions with no rearrangement across all the compared genome sequences. Lines between strains link LCBs that are orthologous between two genome sequences. LCBs found on the bottom part represent regions that are in reverse orientation compared to the reference. In each comparison, the sequence on top is used as reference.

but present in only some strains of clade Xt-III. XopJ was found only in clade Xt-I, while XopAD was found only in clades Xt-II and Xt-III. Interestingly, the *xopM* gene located downstream of the *hrp* cluster was not found in this analysis, the XopM effector identified in LMG 728 being a different one. However, this XopM was found to be present in some strains by visual inspection of the gene clusters (**Figure 5**).

A total of 21 TALE classes were identified in the 12 strains used in this study, with up to 8 in DSM 18974 (**Table 2**). Clade Xt-I was the clade with the most TALEs. However, the TALE repertoires of the clade were very diverse, with only TalDA being present in all strains of the clade. Interestingly, TalDA was also found in ICMP 16317. Clade Xt-II and Xt-III strains had a much smaller set of TALEs, ranging from zero to three, with most of them being exclusive to one strain, except for TalIT, which was found in both LMG 843 and LMG 727. Additionally, a potential pseudo-TALE was identified in LMG 727, with only two repeats. However, the second repeat is only 19 amino acids long and many stop codons were found in the N-terminus. LMG 726 was the only strain in which no TALEs were identified.

DISCUSSION

In this study, we generated high-quality complete genome sequences of all pathotype strains of *X. translucens* and make them available as a community resource for in-depth comparative genome analyses within one of the most important pathogenic

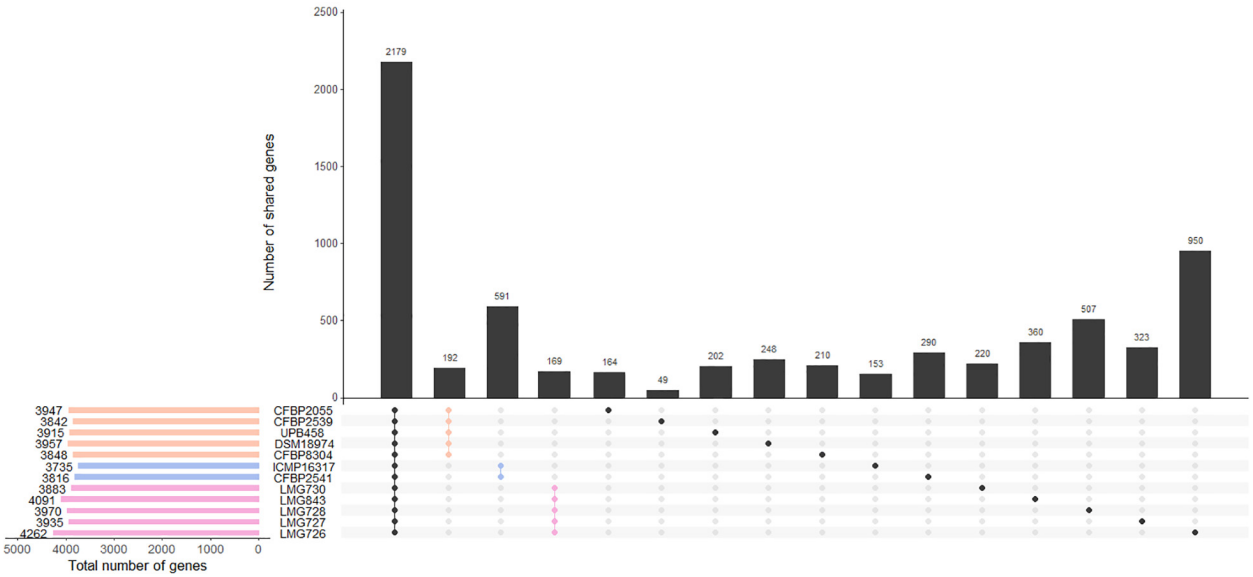


FIGURE 3 | Gene content of the 12 *X. translucens* pathotype strains used in this study. Vertical bars show the number of genes exclusive to the strains marked in the matrix below. The first group represents the core genome of the species. The second, third, and fourth groups represent genes that are exclusive to the three clades, as indicated by their respective color. Other groups represent genes that are exclusive to one strain. Horizontal bars show the total number of genes in each strain. Orange: clade Xt-I, blue: clade Xt-II, pink: clade Xt-III.

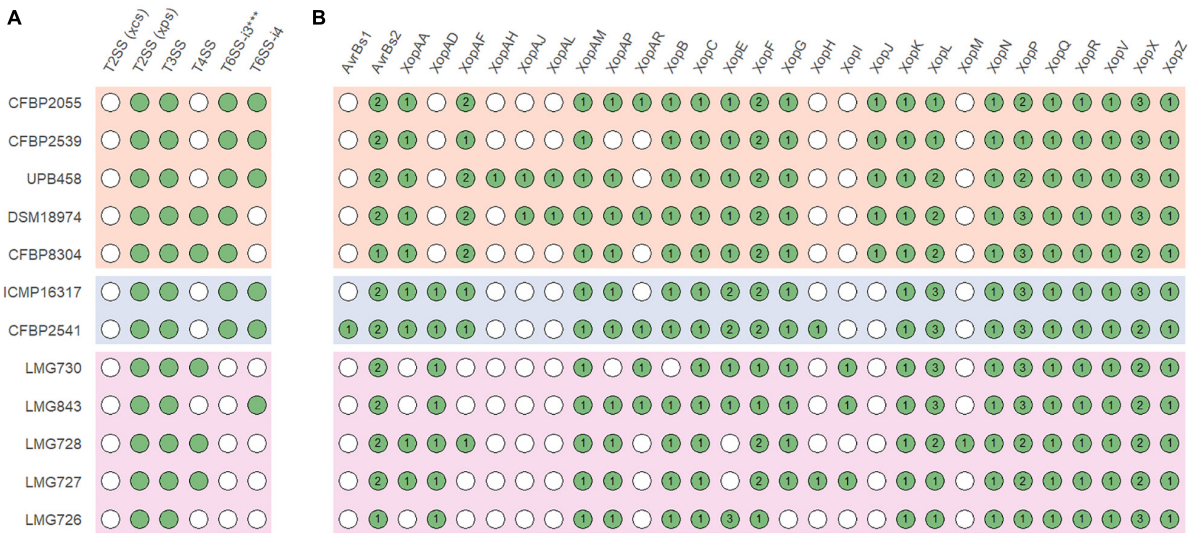


FIGURE 4 | (A) Presence of the *xps* and *xcs* type II secretion systems (T2SS), type III secretion system (T3SS), type IV secretion system (T4SS), subtype 3*** type VI secretion system (T6SS-i3***), and subtype 4 type VI secretion system (T6SS-i4) gene clusters in *X. translucens* pathotype strains. **(B)** Presence of putative type III effectors. Green circles indicate presence, and white circles indicate absence. Numbers denote the number of putative effectors of each class. Orange: clade Xt-I, blue: clade Xt-II, pink: clade Xt-III.

bacterial species. These are the first complete genome sequences for strains of the pathovars *arrhenatheri*, *graminis*, *hordei*, *phlei*, *phleipratensis*, *pistaciae*, *poae*, and *secalis*. These resources complement the complete genome of the pv. *translucens* pathotype strain DSM 18974 (Jaenicke et al., 2016), as well as the already available genome sequences of pv. *cerealis, translucens*, and *undulosa* strains.

Phylogeny based on ANI revealed that three genetically distinct groups can be identified, with members of each clade being less than 96% identical to members of the two other clades. This is in contrast with the usual distinction of only two groups, the "translucens" group and the "graminis" group but is in line with the previous studies suggesting that pv. *cerealis* could be genetically distinct from the other pathovars (Peng et al., 2016;

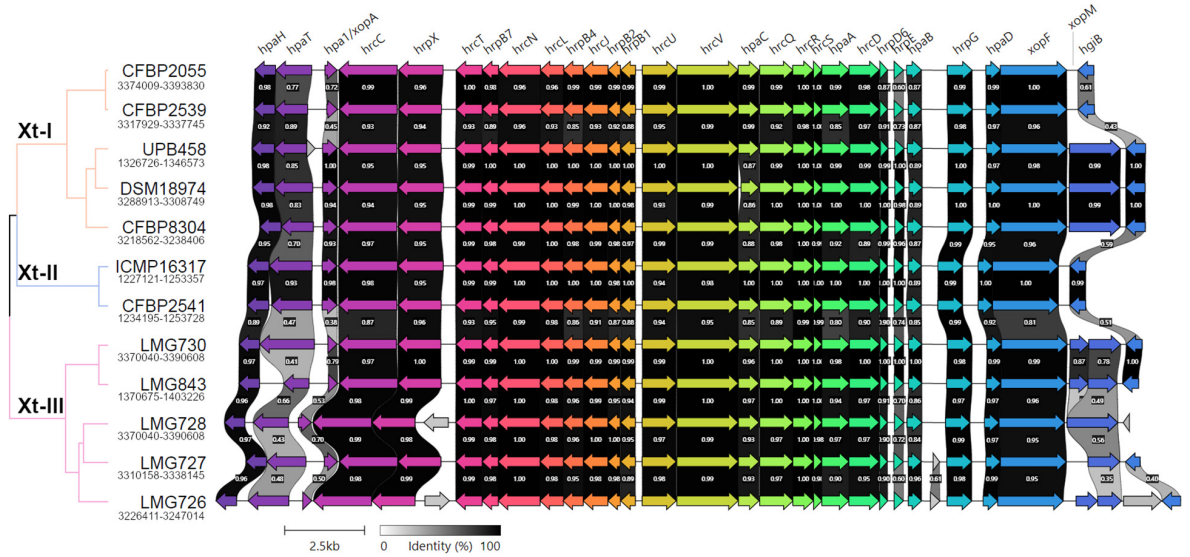


FIGURE 5 | Pairwise comparisons of the genetic content of the type III secretion system cluster in the *X. translucens* pathotype strains obtained with Clinker (Gilchrist and Chooi, 2021). The phylogenetic tree is based on ANI as shown in **Figure 1**, orange: Clade Xt-I, blue: clade Xt-II, pink: clade Xt-III. Colors of the arrows represent groups of similar genes. Clusters for strains ICMP 16317, CFBP 2541, and LMG 727 are found in opposite direction and were reversed in the figure.

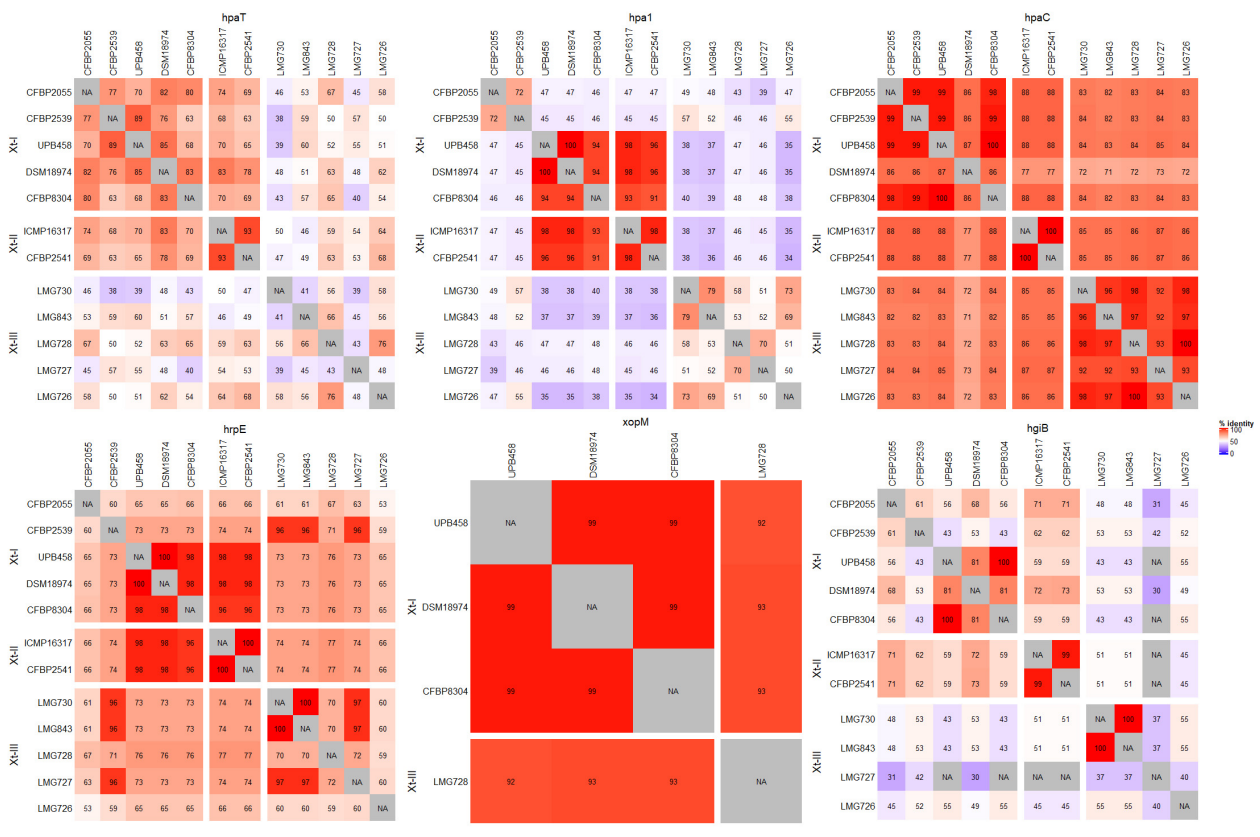


FIGURE 6 | Pairwise comparisons of the nucleotide sequences of the most variable components of the type III secretion system across all *X. translucens* pathotype strains. For *xopM*, only UPB458, DSM 18974, CFBP 8304, and LMG 728 were considered, as they were the only strains where a complete gene was identified.

TABLE 2 | Classes and repeat variable diresidue (RVD) sequences of transcription activator-like effectors (TALEs) found in the *X. translucens* pathotype strains.

TALE class[a]	Strain	RVD sequence	Chromosome position (location:strand)
TalCT	CFBP 2055	HN-HD-HD-HD-NI-NI-NI-HN-HD-HD-NN-NN-NI-NN-HD	1,324,426–1,327,651:1
	CFBP 2539	NN-HD-HD-HD-NI-NI-NI-HN-HD-HD-NN-NN-NI-NN-HD	1,325,514–1,328,739:1
	DSM 18974	NN-HD-HD-HD-NI-NI-NI-NN-HD-HD-NN-NN-NI-NN-HD	671,336–674,561:1
	UPB458	NN-HD-HD-HD-NI-NI-NI-NN-HD-HD-NN-NN-NI-NN-HD	3,936,481–3,939,712:-1
TalCU	DSM 18974	NG-HD-HD-HN-NG-NI-HG-HG-HD-ND-NN-NN-NI-NH-QD	3,565,727–3,568,955:-1
TalCV	DSM 18974	NG-NN-HD-HD-NN-NI-HG-HD-ND-HG-NI-NN-HD	3,562,396–3,565,408:-1
	UPB458	NG-NN-HD-HD-NN-NI-HG-HD-ND-HG-NI-NN-HD	1,087,442–1,090,460:1
TalCW	DSM 18974	NN-NI-HN-HD-NI-NH-NG-HN-HD-HD-HD-NI-QD	660,311–663,329:1
TalCX	CFBP 8304	NN-HD-NG-NN-HN-KG-NI-HD-NI-NK-HD-HD-HD-HD-NI-HN-NH-HD-QD	674,110–677,491:1
	DSM 18974	NN-HD-NG-NI-HN-KG-NI-HD-NI-NH-NG-NN-HD-HD-NI-NN-NI-HD-QD	663,647–667,295:1
	UPB458	NN-QD-NG-NN-HN-KG-NI-HD-NI-NH-NG-HN-HD-HD-NI-NN-HD	3,939,840–3,943,281:-1
TalCY	DSM 18974	NI-NG-HN-NN-HD-NG-ND-NK-QD-NH-QD	668,858–671,207:1
TalCZ	CFBP 2055	NH-NN-HD-NN-HD-NH-HD-YK-NG-NH-Y*-HD-NN-NI-NG-QD	1,916,521–1,919,857:-1
	CFBP 2539	NH-NN-HD-NN-HD-NH-HD-YK-NG-NH-Y*-HD-NN-NI-NG-QD	1,917,023–1,920,359:-1
	DSM 18974	NH-NN-HD-NN-HD-NH-HD-YK-NG-NH-Y*-HD-NN-NI-NG-QD	1,995,271–1,998,607:-1
TalDA	CFBP 2055	HD-YD-NI-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-QD-NH-HD	593,184–596,364:1
	CFBP 2539	HD-YD-NI-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-QD-NH-HD	594,296–597,476:1
	CFBP 8304	HD-YD-NI-NG-NG-NN-YK-NG-HD-NG-NG-ND-NK-QD-NH-QD	653,260–656,587:1
	DSM 18974	NN-HD-NG-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-HD-NH-HD	627,243–630,573:1
	ICMP 16317	HD-YD-NI-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-QD-NH-HD	586,137–589,323:1
	UPB458	NN-HD-NG-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-HD-NH-HD	4,001,560–4,004,884:-1
TalDC	CFBP 2055	NN-NG-HD-HD-HD-KG-NN-Y*-NG-HD-HD-QD-HN	1,321,073–1,324,088:1
TalDD	CFBP 2055	NN-HD-NG-NN-HN-KG-NI-HD-NI-NN-HD-HN-HD-HD-NI-HN-HD-QD	612,552–616,092:1
	CFBP 2539	NN-HD-NG-NN-HN-KG-NI-HD-NI-NN-HD-HN-HD-HD-NI-HN-HD-QD	613,664–617,204:1
	CFBP 2539	NN-HD-NG-NN-HN-KG-NI-HD-NI-NN-HD-HD-NN-NN-NI-HN-HD	1,321,735–1,325,176:1
TalDE	CFBP 2539	NN-HD-NG-NN-HN-HN-NI-NI-NI-NH-NN-HD-NN-NH-HD-HD	1,717,212–1,720,551:1
	CFBP 2055	NN-HD-NG-NN-HN-HN-NI-NI-NI-NH-NN-HD-NN-NH-HD-HD	1,716,713–1,720,052:1
TalDF	CFBP 2055	HD-HN-HN-HD-NH-NH-HG-HD-ND-NN-Y*-NG-HD-NI-NH-NG-HD-HN	1,711,799–1,715,336:1
	CFBP 2539	HD-HN-HN-HD-NH-NH-HG-HD-KG-NN-Y*-NG-HD-NI-NH-NG-HD-HN	1,712,298–1,715,835:1
TalDO	CFBP 2541	NN-NN-KI-NN-HD-NG-HD-NG-NG-NK-HD-HD-NN-QD-NG-QD	2,549,947–2,553,289:1
TalDP	CFBP 2541	NS-KI-NI-HD-NK-GI-HD-NK-HD-NN-HD-NK	590,471–593,369:1
TalJU	LMG 843	NS-NG-HD-HD-NN	1,670,023–1,673,728:-1
TalJS	LMG 727	NN-NG-NN-NG-HD-NK-NG-NI-DD-NK-HD-NG-NN-NI-NG-NN-HD-HD-QD	3,139,389–3,143,094:1
	LMG 843	NN-NG-NN-NG-HD-NK-NG-NK-DD-NK-DD-NG-NN-NI-NG-NN-HD-HD-QD	2,289,805–2,292,265:1
TalIY	UPB 458	NI-NG-HN-NK-HD-NH-HN-HD-HD-HD-HD-QD	1,816,261–1,819,174:-1
TalJB	CFBP 8304	NH-NN-HE-NK-HD-NK-HD-YK-NG-NH-Y*-HE-NI-NI-NG-QD	1,986,200–1,989,539:-1
TalJT	LMG 730	NN-NN-HK-HK-HD-HN-HD-NN	2,152,104–2,154,627:1
TalJV	LMG 728	NN-NI-HD-HD-HD-HD-KT-NG-NN-NN-KT-NI-NN-HD-NG-NG-NN-HD-NK	2,784,954–2,788,575:-1
TalJW	LMG 728	HK-HD-HN-NI-NG-HD-HN-NI-NG-HD-NG-NN-HN-HD-NG-NN-NI-HD-NG-NN-HD-HD-QD	2,798,245–2,802,361:-1
TalJX	LMG 728	NK-NG-NI-HD-NG-NN-NG-HD-NK-N*-NK-HD-NN-HD-NG-NN-NI-NG-HD-HD-NN-HD-HD-NN-HD-HD-QD	1,903,396–1,907,938:-1
Pseudo TALE	LMG 727	NI-NN	2,236,495–2,238,688:1

[a] *TALE classes based on AnnoTALE (Grau et al., 2016).*
**Thirteenth residue is missing.*

Langlois et al., 2017; Shah et al., 2019). Based on this phylogeny, we propose to classify these groups as clade Xt-I, containing pathovars *hordei, translucens*, *undulosa*, and *secalis*; clade Xt-II, containing pv. *cerealis*; and clade Xt-III, containing pathovars *arrhenatheri*, *graminis*, *phlei*, *phleipratensis*, and *poae*.

Moreover, the two strains of pv. *pistaciae* were grouped in two different clades, with the group A strain found in clade Xt-I and the group B strain found in clade Xt-II, the two strains being only 95.4% identical. This confirmed the previous phylogeny based on *gyrB* sequences where the group A strain was more closely related to pathovars *translucens, secalis*, and *undulosa*, while the group B strain was closest to pv. *cerealis* (Giblot-Ducray et al., 2009). There is clear evidence that these strains are not directly related, although they share the same host.

These results raise some limitations of the classical pathovar classification used in *Xanthomonas*. Indeed, the pathotype strains

of pathovars *secalis* and *undulosa* are 99.8% identical and share similar genomic organization and virulence features. As they also have a similar host range, this could lead to them being considered as the same taxonomic entity. On the other hand, the two groups of pv. *pistaciae* are very different genetically and considering them under the same pathovar could hinder the better understanding of their respective biology. Despite these limitations, the pathovar classification was surprisingly robust for the other pathovars. Nonetheless, this work suggests that this classification, currently based solely on pathogenicity tests, should be rethought to better reflect the genetic relationships between pathovars and their evolutionary history.

Within the three clades identified by the ANI-based phylogeny, we found 190, 588, and 168 genes that were specific to clade Xt-I, Xt-II, and Xt-III, respectively, as well as 48–966 genes that were strain specific. However, these numbers are probably biased by the small number of strains included in the comparison and including more strains would result in a smaller number of clade- and strain-specific genes. Nonetheless, these constitute a valuable list of genes that could shape the host range of each clade and/or strain.

All strains had an *xps* T2SS and no *xcs* T2SS, as well as a T3SS with a similar genetic organization, although a few genes of the *hrp* cluster were variable in sequence between strains. Interestingly, strains that possessed a T4SS did not possess a T6SS-i4 and vice versa. As these two secretion systems are both involved in antimicrobial activity, it is possible that their function is redundant in *X. translucens* and only one of them is required to play this role. However, the presence or absence of these two secretion systems does not necessarily reflect the phylogenetic relationships between the strains investigated in this study. Additionally, a T6SS-i3*** was found in all strains of clades Xt-I and Xt-II. However, it was previously hypothesized that this subgroup of T6SS could be non-functional due to the lack of a PAAR and could be complemented by the presence of another subtype of T6SS (Bayer-Santos et al., 2019). Interestingly, no T4SS or T6SS was identified in strain LMG 726, which confirms the previous research that showed that many pv. *graminis* strains lack a T6SS, with only strains Xtg2, Xtg9, Xtg10, and NCPPB3709 found to harbor one (Hersemann et al., 2017).

We have identified a total of 13 type III non-TALE effector classes that could constitute a core set of effectors in *X. translucens*. Moreover, XopAD was not found in the strains of clade Xt-I in our analysis but was previously identified in strain DSM 18974 in different analyses, indicating that it could also be part of the *X. translucens* core set of type III effectors (Peng et al., 2016; Koebnik et al., 2021; Shah et al., 2021). Five additional effector classes constitute a core set of effectors in clades Xt-I and Xt-II, as well as XopJ, which was specific to clade Xt-I. Some effector classes were specific to one strain, such as AvrBs1 in pv. *cerealis* strain CFBP 2541, or XopAH in pv. *hordei* strain UPB458. Furthermore, XopAJ and XopAL class effectors were found in both pv. *translucens* and *hordei*. These effectors could have a role in the host specificity of each clade and/or pathovar. However, no type III effector has yet been functionally characterized in *X. translucens*, and the effectors identified in this study will need to be functionally validated to confirm their role in the pathogenicity of each pathovar.

High-quality genome assemblies allowed for the identification of the first TALE to be reported in grass-infecting *X. translucens* strains, as well as in pathovars *secalis*, *hordei*, and *pistaciae*. No TALE was found in pv. *graminis* in our analysis, confirming previous research where no TALE was identified in draft genomes of the pathovar (Wichmann et al., 2013). The TALEs identified in pv. *undulosa* correspond to those previously identified in other strains of this pathovar, except for Xt4699-Tal3, for which no similar TALE was found, as was the case in strain ICMP 11055 (Peng et al., 2016; Falahi Charkhabi et al., 2017). The TALE of class TalDC that we identified has the same RVD sequence as Xt4699-Tal8 and a similar sequence to ICMP 11055-Tal4b. Additionally, the TALE of class TalDD identified in CFBP 2055 and CFBP 2539 has a similar RVD sequence to ICMP 11055-Tal2. Furthermore, the TALEs identified in pv. *cerealis* strain CFBP 2541 correspond to the TALE previously identified in strain Nxtc01, with the TALE of class TalDP having a similar sequence to Nxtc01-Tal1 (Shah et al., 2019). As these four TALEs have previously been shown to have a role in virulence, the corresponding effectors we identified in this study could play a similar role in their respective strain (Falahi Charkhabi et al., 2017; Peng et al., 2019; Shah et al., 2019). However, these are the only TALEs that have been functionally characterized in *X. translucens* to date. Nonetheless, as most of the TALEs identified in this study are clade or strain specific, they could be essential components of host adaptation. Indeed, TALEs activate the transcription of plant genes by binding to their promoter region. As these sequences vary between plant species, TALEs must be adapted to target the specific sequence of that promoter in their host and thus reflect processes of co-adaptation between the bacteria and its host (Jacques et al., 2016). However, the lack of TALEs in pv. *graminis* indicates that other mechanisms play a role in *X. translucens* host speciation.

The pv. *pistaciae* group A pathotype strain shared a very similar genomic organization, T3SS cluster organization, and gene identity, as well as type III effector and TALE repertoires with the pathotype strains of pv. *translucens* and *hordei*. The same was true for the group B pv. *pistaciae* strain and the pv. *cerealis* pathotype strain, which were even more closely related. This could indicate that a very small set of genes is responsible for their ability to infect pistachio, as a previous study showed that pv. *translucens* strain DAR 35705 was unable to cause symptoms in pistachio and other Anacardiaceaea, while the pv. *pistaciae* strains of both groups were pathogenic in Anacardiaceaea and Poaceae (Marefat et al., 2006). However, no type III effector or TALE exclusive to either strain of pv. *pistaciae* has been identified in this study.

Although it had the biggest set of total genes, the pv. *graminis* strain LMG 726 was found to lack many virulence features such as the T4SS, the T6SS, as well as TALEs and had the smallest set of type III effectors. This is surprising, as this pathovar is known to be the most widespread and virulent grass-infecting *X. translucens* pathovar, with the largest host range in clade Xt-III. Indeed, while pv. *phlei* and pv.

phleipratensis, pv. *arrhenatheri*, and pv. *poae* are restricted to the genera *Phleum*, *Arrhenatherum*, and *Poa*, respectively, pv. *graminis* can infect many grasses from genera such as *Agrostis*, *Alopecurus*, *Dactylis*, *Deschampsia*, *Festuca*, *Lolium*, *Phalaris*, *Phleum*, *Poa*, and *Trisetum* (Egli et al., 1975; Egli and Schmidt, 1982). Nonetheless, these results go along previous research that showed that strains of pv. *graminis* also lack a flagellum and have a distinct type IV pilus compared to other *X. translucens* pathovars (Hersemann et al., 2017). It was hypothesized that since pv. *graminis* is usually spread by mowing tools, the lack of a flagellum might not hinder its ability to spread in the plant, as flagellar motility is mostly necessary for the bacteria to reach points of entry in the plant. Similarly, it might encounter less competition with epiphytic microorganisms, and the presence of a T4SS and/or a T6SS might not be crucial for its survival. Additionally, as these features can act as elicitors of plant defense, their absence in pv. *graminis* could help to evade such defense mechanisms—an intriguing hypothesis that, however, has not yet been tested.

CONCLUSION

In conclusion, our study substantially increased the number of complete genome sequences available for *X. translucens*, providing high-quality genomic resources for all the pathovars of the species. These sequences constitute a valuable basis for future studies investigating the phylogenetic relationships between *X. translucens* pathovars and other *Xanthomonas* species, as well as their key genetic features. The virulence features of *X. translucens* that we identified will help to better understand the biology of each clade and/or pathovar, and processes of adaptation to their respective hosts. The inclusion of the high-quality genome sequences of the additional strains of each pathovar in further comparative genomics studies will help in refining this list, laying the foundations for the development of new resistant cereal crop and forage grass cultivars.

DATA AVAILABILITY STATEMENT

The genome sequences generated in this study can be found on https://www.ncbi.nlm.nih.gov with accession numbers CP074361-CP074365, CP076249-CP076251, CP076254, CP083804, CP086332-CP086333, and LT604072

AUTHOR CONTRIBUTIONS

FG and RoK: design of the study with the help of RaK, CB, JJ, VR-R, and BS. FG, VR-R, RaK, and SC: genome sequencing and assemblies. FG: data analysis with assistance from VR-R. FG, RoK, RaK, JJ, VR-R, CB, SC, and BS: writing and reviewing. All authors contributed to the article and approved the submitted version.

FUNDING

Funding was provided by the Swiss National Science Foundation (Grant No. IZCOZO_177062). This research was also supported by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) (Award No. 2018-67013-28490) through the Joint National Science Foundation-NIFA Plant Biotic Interactions Program and the USDA NIFA Food and Agriculture Cyberinformatics and Tools grant program (Grant No. 2021-67021-34343) to JJ. Work in RaK's laboratory was supported by a grant from the French National Research Agency (Grant No. CROpTAL ANR-14-CE19-0002).

ACKNOWLEDGMENTS

Library preparation and sequencing were done by the Functional Genomics Center Zurich (Zurich, Switzerland; https://fgcz.ch/) and the GENTYANE genotyping platform (INRA Clermont-Ferrand, France; https://gentyane.clermont.inrae.fr). Technical assistance by Ildiko Katalin Nagy (Enviroinvest Corp. Pecs, Hungary) is greatly appreciated. RaK and RoK are grateful for an international mobility EXPLORE grant from the Montpellier University of Excellence (MUSE). This article is based upon work from COST Action CA16107 EuroXanth, supported by COST (European Cooperation in Science and Technology).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.817815/full#supplementary-material

Supplementary Figure 1 | Average nucleotide identity (ANI) of all available complete genome sequences of *Xanthomonas translucens* strains and ANI-based phylogeny constructed with Ward's hierarchical clustering method. Distance in the dendrogram represents ANI dissimilarity between nodes. Orange: clade Xt-I, blue: clade Xt-II, pink: clade Xt-III. ANI is depicted as a gradient from blue (<96 %) to white (98%) to red (100%).

Supplementary Dataset 1 | Amino acid sequences of all secretion system components used to predict the presence of these secretion systems by tBLASTn.

Supplementary Dataset 2 | Amino acid sequences of all type III effectors (T3Es) used to predict the presence of these effectors by BLASTp.

REFERENCES

Alvarez-Martinez, C. E., Sgro, G. G., Araujo, G. G., Paiva, M. R. N., Matsuyama, B. Y., Guzzo, C. R., et al. (2021). Secrete or perish: the role of secretion systems in *Xanthomonas* biology. *Comput. Struct. Biotechnol. J.* 19, 279–302. doi: 10.1016/j.csbj.2020.12.020

Bayer-Santos, E., Ceseti, L., de, M., Farah, C. S., and Alvarez-Martinez, C. E. (2019). Distribution, function and regulation of type 6 secretion systems of Xanthomonadales. *Front. Microbiol.* 10:1635. doi: 10.3389/fmicb.2019.01635

Bogdanove, A. J., Beer, S. V., Bonas, U., Boucher, C. A., Collmer, A., Coplin, D. L., et al. (1996). Unified nomenclature for broadly conserved *hrp* genes

of phytopathogenic bacteria. *Mol. Microbiol.* 20, 681–683. doi: 10.1046/j.1365-2958.1996.5731077.x

Bonas, U., Schulte, R., Fenselau, S., Mainsavage, G. V., Staskawicz, B. J., and Stall, R. E. (1991). Isolation of a gene cluster from *Xanthomonas campestris* pv. *vesicatoria* that determines pathogenicity and the hypersensitive response on and the hypersensitive response on pepper and tomato. *Mol. Plant Microbe Interact.* 4, 81–88.

Booher, N. J., Carpenter, S. C. D., Sebra, R. P., Wang, L., Salzberg, S. L., Leach, J. E., et al. (2015). Single molecule real-time sequencing of *Xanthomonas oryzae* genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. *Microb. Genom.* 1:e000032. doi: 10.1099/mgen.0.000032

Bragard, C., Singer, E., Alizadeh, A., Vauterin, L., Maraite, H., and Swings, J. (1997). *Xanthomonas translucens* from small grains: diversity and phytopathological relevance. *Phytopathology®* 87, 1111–1117. doi: 10.1094/PHYTO.1997.87.11.1111

Büttner, D., and Bonas, U. (2010). Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* 34, 107–133. doi: 10.1111/j.1574-6976.2009.00192.x

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Choi, Y., Kim, N., Mannaa, M., Kim, H., Park, J., Jung, H., et al. (2020). Characterization of type VI secretion system in *Xanthomonas oryzae* pv. *oryzae* and its role in virulence to rice. *Plant. Pathol. J.* 36, 289–296. doi: 10.5423/PPJ.NT.02.2020.0026

Darling, A. C. E. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Egli, T., Goto, M., and Schmidt, D. (1975). Bacterial wilt, a new forage grass disease. *J. Phytopathol.* 82, 111–121. doi: 10.1111/j.1439-0434.1975.tb02825.x

Egli, T., and Schmidt, D. (1982). Pathogenic variation among the causal agents of bacterial wilt of forage grasses. *J. Phytopathol.* 104, 138–150. doi: 10.1111/j.1439-0434.1982.tb00520.x

Facelli, E., Taylor, C., Scott, E., Fegan, M., Huys, G., Noble, R. D., et al. (2005). Identification of the causal agent of pistachio dieback in Australia. *Eur. J. Plant Pathol.* 112, 155–165. doi: 10.1007/s10658-005-3120-9

Falahi Charkhabi, N., Booher, N. J., Peng, Z., Wang, L., Rahimian, H., Shams-Bakhsh, M., et al. (2017). Complete genome sequencing and targeted mutagenesis reveal virulence contributions of Tal2 and Tal4b of *Xanthomonas translucens* pv. *undulosa* ICMP11055 in bacterial leaf streak of wheat. *Front. Microbiol.* 8:1488. doi: 10.3389/fmicb.2017.01488

Giblot-Ducray, D., Marefat, A., Gillings, M. R., Parkinson, N. M., Bowman, J. P., Ophel-Keller, K., et al. (2009). Proposal of *Xanthomonas translucens* pv. *pistaciae* pv. nov., pathogenic to pistachio (*Pistacia vera*). *Syst. Appl. Microbiol.* 32, 549–557. doi: 10.1016/j.syapm.2009.08.001

Gilchrist, C. L. M., and Chooi, Y.-H. (2021). Clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 37, 2473–2475. doi: 10.1093/bioinformatics/btab007

Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R. D., Wilson, G. G., et al. (2016). AnnoTALE: bioinformatics tools for identification, annotation and nomenclature of TALEs from *Xanthomonas* genomic sequences. *Sci. Rep.* 6:21077. doi: 10.1038/srep21077

Hersemann, L., Wibberg, D., Blom, J., Goesmann, A., Widmer, F., Vorhölter, F.-J., et al. (2017). Comparative genomics of host adaptive traits in *Xanthomonas translucens* pv. *graminis*. *BMC Genomics* 18:35. doi: 10.1186/s12864-016-3422-7

Jacques, M.-A., Arlat, M., Boulanger, A., Boureau, T., Carrère, S., Cesbron, S., et al. (2016). Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annu. Rev. Phytopathol.* 54, 163–187. doi: 10.1146/annurev-phyto-080615-100147

Jaenicke, S., Bunk, B., Wibberg, D., Spröer, C., Hersemann, L., Blom, J., et al. (2016). Complete genome sequence of the barley pathogen *Xanthomonas translucens* pv. *translucens* DSM 18974 ^{T} (ATCC 19319 ^{T}). *Genome Announc.* 4:e01334–16. doi: 10.1128/genomeA.01334-16

Jha, G., Rajeshwari, R., and Sonti, R. V. (2005). Bacterial type two secretion system secreted proteins: double-edged swords for plant pathogens. *Mol. Plant Microbe Interact®.* 18, 891–898. doi: 10.1094/MPMI-18-0891

Koebnik, R., Burokiene, D., Bragard, C., Chang, C., Saux, M. F.-L., Kölliker, R., et al. (2021). The complete genome sequence of *Xanthomonas theicola*, the causal agent of canker on tea plants, reveals novel secretion systems in clade-1 xanthomonads. *Phytopathology®* 111, 611–616. doi: 10.1094/PHYTO-07-20-0273-SC

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8

Langlois, P. A., Snelling, J., Hamilton, J. P., Bragard, C., Koebnik, R., Verdier, V., et al. (2017). Characterization of the *Xanthomonas translucens* complex using draft genomes, comparative genomics, phylogenetic analysis, and diagnostic lamp assays. *Phytopathology®* 107, 519–527. doi: 10.1094/PHYTO-08-16-0286-R

Li, J., Yao, Y., Xu, H. H., Hao, L., Deng, Z., Rajakumar, K., et al. (2015). SecReT6: a web-based resource for type VI secretion systems found in bacteria: genomics update. *Environ. Microbiol.* 17, 2196–2202. doi: 10.1111/1462-2920.12794

Marefat, A., Scott, E. S., Ophel-Keller, K., and Sedgley, M. (2006). Genetic, phenotypic and pathogenic diversity among xanthomonads isolated from pistachio (*Pistacia vera*) in Australia. *Plant Pathol.* 55, 639–649. doi: 10.1111/j.1365-3059.2006.01437.x

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Peng, Z., Hu, Y., Xie, J., Potnis, N., Akhunova, A., Jones, J., et al. (2016). Long read and single molecule DNA sequencing simplifies genome assembly and TAL effector gene analysis of *Xanthomonas translucens*. *BMC Genomics* 17:21. doi: 10.1186/s12864-015-2348-9

Peng, Z., Hu, Y., Zhang, J., Huguet-Tapia, J. C., Block, A. K., Park, S., et al. (2019). *Xanthomonas translucens* commandeers the host rate-limiting step in ABA biosynthesis for disease susceptibility. *Proc. Natl. Acad. Sci. U.S.A* 116, 20938–20946. doi: 10.1073/pnas.1911660116

Pesce, C., Bolot, S., Cunnac, S., Portier, P., Fischer-Le Saux, M., Jacques, M.-A., et al. (2015). High-quality draft genome sequence of the *Xanthomonas translucens* pv. *cerealis* pathotype strain cfbp 2541. *Genome Announc.* 3:e01574-14. doi: 10.1128/genomeA.01574-14

Pesce, C., Jacobs, J. M., Berthelot, E., Perret, M., Vancheva, T., Bragard, C., et al. (2017). Comparative genomics identifies a novel conserved protein, hpaT, in proteobacterial type III secretion systems that do not possess the putative translocon protein hrpF. *Front. Microbiol.* 8:1177. doi: 10.3389/fmicb.2017.01177

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/C5AY02550H

Rademaker, J. L. W., Norman, D. J., Forster, R. L., Louws, F. J., Schultz, M. H., and de Bruijn, F. J. (2006). Classification and identification of *Xanthomonas translucens* isolates, including those pathogenic to ornamental asparagus. *Phytopathology®* 96, 876–884. doi: 10.1094/PHYTO-96-0876

Roman-Reyna, V., Luna, E. K., Pesce, C., Vancheva, T., Chang, C., Ziegle, J., et al. (2020). Genome resource of barley bacterial blight and leaf streak pathogen *Xanthomonas translucens* pv. *translucens* strain UPB886. *Plant Dis.* 104, 13–15. doi: 10.1094/PDIS-05-19-1103-A

Sapkota, S., Mergoum, M., and Liu, Z. (2020). The translucens group of *Xanthomonas translucens*: complicated and important pathogens causing bacterial leaf streak on cereals. *Mol. Plant Pathol.* 21, 291–302. doi: 10.1111/mpp.12909

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sgro, G. G., Oka, G. U., Souza, D. P., Cenens, W., Bayer-Santos, E., Matsuyama, B. Y., et al. (2019). Bacteria-killing type IV secretion systems. *Front. Microbiol.* 10:1078. doi: 10.3389/fmicb.2019.01078

Shah, S. M. A., Haq, F., Ma, W., Xu, X., Wang, S., Xu, Z., et al. (2019). Tal1NXtc01 in *Xanthomonas translucens* pv. *cerealis* contributes to virulence in bacterial leaf streak of wheat. *Front. Microbiol.* 10:2040. doi: 10.3389/fmicb.2019.02040

Shah, S. M. A., Khojasteh, M., Wang, Q., Taghavi, S. M., Xu, Z., Khodaygan, P., et al. (2021). Genomics-enabled novel insight into the pathovar-specific population structure of the bacterial leaf streak pathogen *Xanthomonas translucens* in small grain cereals. *Front. Microbiol.* 12:674952. doi: 10.3389/fmicb.2021.674952

Souza, D. P., Oka, G. U., Alvarez-Martinez, C. E., Bisson-Filho, A. W., Dunger, G., Hobeika, L., et al. (2015). Bacterial killing via a type IV secretion system. *Nat. Commun.* 6:6453. doi: 10.1038/ncomms7453

Stead, D. E. (1989). Grouping of *Xanthomonas campestris* pathovars of cereals and grasses by fatty acid profiling 1. *EPPO Bull.* 19, 57–68.

Streubel, J., Baum, H., Grau, J., Stuttman, J., and Boch, J. (2017). Dissection of TALE-dependent gene activation reveals that they induce transcription cooperatively and in both orientations. *PLoS One* 12:e0173580. doi: 10.1371/journal.pone.0173580

Szczesny, R., Jordan, M., Schramm, C., Schulz, S., Cogez, V., Bonas, U., et al. (2010). Functional characterization of the Xcs and Xps type II secretion systems from the plant pathogenic bacterium *Xanthomonas campestris* pv *vesicatoria*. *New Phytol.* 187, 983–1002. doi: 10.1111/j.1469-8137.2010.03312.x

Vauterin, L., Hoste, B., Kersters, K., and Swings, J. (1995). Reclassification of *Xanthomonas*. *Int. J. Syst. Bacteriol.* 45, 472–489. doi: 10.1099/00207713-45-3-472

Wang, L., Rinaldi, F. C., Singh, P., Doyle, E. L., Dubrow, Z. E., Tran, T. T., et al. (2017). TAL effectors drive transcription bidirectionally in plants. *Mol. Plant* 10, 285–296. doi: 10.1016/j.molp.2016.12.002

White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Mol. Plant Pathol.* 10, 749–766. doi: 10.1111/j.1364-3703.2009.00590.x

Wichmann, F., Vorhölter, F.-J., Hersemann, L., Widmer, F., Blom, J., Niehaus, K., et al. (2013). The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol. Plant Pathol.* 14, 576–588. doi: 10.1111/mpp.12030

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 March 2022
doi: 10.3389/fmicb.2022.826386

Migration Drives the Replacement of *Xanthomonas perforans* Races in the Absence of Widely Deployed Resistance

*Eduardo Bernal[1], Francesca Rotondo[2], Veronica Roman-Reyna[3,4], Taylor Klass[3,4], Sujan Timilsina[5], Gerald V. Minsavage[5], Fernanda Iruegas-Bocardo[5], Erica M. Goss[5,6], Jeffrey B. Jones[5], Jonathan M. Jacobs[3,4], Sally A. Miller[2] and David M. Francis[1]**

[1] Department of Horticulture and Crop Science, College of Food, Agricultural, and Environmental Sciences, The Ohio State University, Wooster, OH, United States, [2] Department of Plant Pathology, College of Food, Agricultural, and Environmental Sciences, The Ohio State University, Wooster, OH, United States, [3] Department of Plant Pathology, College of Food, Agricultural, and Environmental Sciences, The Ohio State University, Columbus, OH, United States, [4] Infectious Diseases Institute, The Ohio State University, Columbus, OH, United States, [5] Department of Plant Pathology, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, United States, [6] Emerging Pathogens Institute, University of Florida, Gainesville, FL, United States

OPEN ACCESS

Edited by:
Jesus L. Romalde,
University of Santiago de Compostela, Spain

Reviewed by:
Margarita Gomila,
University of the Balearic Islands, Spain
David J. Studholme,
University of Exeter, United Kingdom

***Correspondence:**
David M. Francis
francis.77@osu.edu

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 30 November 2021
Accepted: 09 February 2022
Published: 18 March 2022

Citation:
Bernal E, Rotondo F, Roman-Reyna V, Klass T, Timilsina S, Minsavage GV, Iruegas-Bocardo F, Goss EM, Jones JB, Jacobs JM, Miller SA and Francis DM (2022) Migration Drives the Replacement of Xanthomonas perforans Races in the Absence of Widely Deployed Resistance. Front. Microbiol. 13:826386. doi: 10.3389/fmicb.2022.826386

Changes in *Xanthomonas* race and species composition causing bacterial spot of tomato have occurred throughout the world and are often associated with epidemics. Knowledge of bacterial population structure is key for resistance discovery and deployment. We surveyed *Xanthomonas* spp. composition from processing tomato fields in the Midwestern United States over a 4-year period between 2017 and 2020, compared these to strains collected previously, and found that *X. perforans* is currently the most prevalent species. We characterized 564 *X. perforans* isolates for sequence variation in *avrXv3* to distinguish between race T3 and T4 and validated race designation using hypersensitive response (HR) assays for 106 isolates. Race T4 accounted for over 95% of *X. perforans* isolates collected in the Midwest between 2017 and 2020. Whole genome sequencing, Average Nucleotide Identity (ANI) analysis, core genome alignment and single nucleotide polymorphism (SNP) detection relative to a reference strain, and phylogenomic analysis suggest that the majority of Midwestern *X. perforans* strains collected between 2017 and 2020 were nearly identical, with greater than 99.99% ANI to *X. perforans* isolates collected from Collier County, Florida in 2012. These isolates shared a common SNP variant resulting an a premature stop codon in *avrXv3.* One sequenced isolate was identified with a deletion of *avrXv3* and shared 99.99% ANI with a strain collected in Collier Co., Florida in 2006. A population shift to *X. perforans* T4 occurred in the absence of widely deployed resistance, with only 7% of tomato varieties tested having the resistant allele at the *Xv3*/*Rx-4* locus. The persistence of nearly identical strains over multiple years suggests that migration led to the establishment of an endemic population. Our findings validate a genomics-based framework to track shifts in *X. perforans* populations due to migration, mutation, drift, or selection based on comparisons to 146 genomes.

Keywords: bacterial spot, *X. perforans*, tomato, whole genome sequencing, overwintering, clonal populations, migration, phylogenomic analysis

INTRODUCTION

Xanthomonas spp. causing the disease bacterial spot of tomato (BST) are not restrained by geographical location, and pathogen populations continue to change over time. In recent years *X. perforans* has dominated many tomato growing regions of the United States. More specifically, *X. perforans* race T4 (T4) was first observed in 1998, and since 2006 it has been dominant in Florida (Horvath et al., 2012; Timilsina et al., 2016). In the Midwestern United States multiple species of *Xanthomonas* have persisted, with the abundance changing over time (Sahin, 1997; Ma et al., 2011; Bernal, 2020). Between 1994 and 1996 collections in Ohio of xanthomonads causing BST showed that 51% of strains were characterized as race T1 (now recognized as *X. euvesicatoria*), 38% as T2 (*X. vesicatoria*), and 11% as T3 (*X. perforans*) (T3). Strains collected between 1994 and 1996 resulted in the first report of the presence of T3 in Ohio (Sahin, 1997). Collections between 2010 and 2013 showed a shift in the *Xanthomonas* species composition in Ohio and Michigan. Seventy percent of isolates were *X. hortorum* pv. *gardneri* (no race designation), 19% were *X. perforans,* and 11% were *X. euvesicatoria* (Ma et al., 2011; Ma, 2015). Between 2016 and 2017, collections in Indiana and North Carolina determined that *X. perforans* race T4 has become the dominant species followed by *X. hortorum* pv. *gardneri* and *X. euvesicatoria* (Egel et al., 2018; Adhikari et al., 2019).

The T3 and T4 race designation for *X. perforans* is based on the presence of a hypersensitive response (HR) when the pathogen protein AVRXV3 is recognized by the tomato locus *Xv3/Rx4* which corresponds to Solyc11g069020 at 53,579,795–53,583,729 bp on chromosome 11 (Robbins et al., 2009; Wang et al., 2011; Pei et al., 2012; Zhang et al., 2021). The resistant allele for *Xv3/Rx4* is found in multiple tomato accessions, including H7981, PI 126932, and PI 128216 (Scott et al., 1995, 1996; Pei et al., 2012; Zhang et al., 2021). If the tomato accession carrying *Xv3/Rx4* reacts with a HR after syringe inoculation with *X. perforans,* the strain is designated T3. If the strain fails to elicit an HR on plants with *Xv3/Rx4* it is designated T4 or T2. The T4 designation also implies an HR on tomatoes carrying the RXopJ4 locus from LA0716 (Sharlach et al., 2013). Several variants have been identified in *avrXv3* resulting in a loss of HR (Timilsina et al., 2016). These variants include a single nucleotide polymorphism (SNP) leading to an early stop codon, an insertion/deletion (indel) upstream of the coding sequence creating a "pseudogene/indel," contig breaks in the coding sequence causing a truncation of the protein, or complete loss of the gene (Timilsina et al., 2016). Identification of *X. perforans* race T4 can therefore be characterized through plant-based assays using HR or sequence characterization of the bacterial effector.

In addition to effector mediated phenotypes, multi locus sequence alignment (MLSA) and whole genome sequencing have been used as tools to estimate phylogeny of *Xanthomonas* spp. (Timilsina et al., 2015, 2019). Whole genome sequencing has become a standard approach to measure the genomic similarity within and between species and to elucidate the predominant evolutionary forces governing population shifts of plant pathogens. Average nucleotide identity (ANI) uses a Basic Local Alignment Search Technique (BLAST) to scan 1 Kb fragments shared between contigs (or complete genomes). Estimates of the nucleotide identity from the individual shared fragments are then used to calculate the ANI across the genome (Goris et al., 2007; Figueras et al., 2014). Genomic variation can provide insight into pathogen dissemination across global geographic regions or within local agricultural production systems (Vinatzer et al., 2014; Monteil et al., 2016; Klein-Gordon et al., 2020, 2021; Perez-Quintero et al., 2020).

We collected *Xanthomonas* spp. from processing tomato fields in multiple counties in Ohio and Indiana between 2017 and 2020. Fruits were sampled at different stages of maturity, representing infection events that could occur 5–7 weeks apart. We sampled tomato varieties from various seed producers and fields of multiple growers established from seedlings grown in different transplant houses. Hypersensitive response assays, PCR amplification, and sequence characterization of *avrXv3*, *avrXv4*, and *avrBsT* were used for *X. perforans* race identification. Whole genome sequence data were used to assess the relationship between 38 *X. perforans* strains collected in the Midwest and 108 strains isolated from other geographical regions. The specific objectives of this study were to gain insight into the forces shaping pathogen populations by (i) describing the race composition of *X. perforans* in the Midwestern United States; (ii) identifying variants present in the *avrXv3* gene; and (iii) estimating the genomic relatedeness between *X. perforans* strains from Midwestern states and strains found in other geographical locations using phylogenomic approaches; and (iv) evaluating commercial tomato varieties for the presence of the *Xv3/Rx4* locus to determine the extent of resistance deployment in the region.

MATERIALS AND METHODS

Sampling Tomatoes With Bacterial Spot Lesions

Sampling of *Xanthomonas* populations was performed in tomato fields in 2017, 2018, 2019, and 2020. For the first 3 years of collection a hierarchical strategy was employed. Field locations were organized according to state, counties, and regions within state that correspond to specific processing plants. Agricultural staff for major Midwestern tomato processors were contacted for field locations and field maps. Fruits with lesions were collected from fields in Ohio (Putnam, Hancock, Wood, Ottawa, Erie, Sandusky, and Seneca) and Indiana (Madison and Tipton) counties. These counties were divided into four sub-regions (North Central, Northwest, West Central OH and Central, IN). Growers in each sub-region sourced transplants from the same regional greenhouse grower(s) and sent harvested fruit to the same regional processing plants. In each field, fruit were sampled on a transect, with multiple green and red fruit collected per variety and per field. Fruit were processed in the Vegetable Pathology Lab, Department of Plant Pathology at The Ohio State University, Wooster, OH, United States using standard procedures previously described (Ma et al., 2011). A single isolate was maintained per lesion and given a numerical strain

designation. In 2020, we isolated a few strains for the purpose of whole genome sequencing, described further below.

Identification via BOX-PCR and qPCR

BOX-PCR and Real-Time PCR (qPCR) were used to discriminate between *X. perforans*, *X. euvesicatoria,* and *X. hortorum* pv. *gardneri*. BOX-PCR (BOX A 1R primer) was used to compare DNA banding profiles of the palindromic intergenic repetitive BOX A subunits conserved in bacterial species (Louws et al., 1994; Ma et al., 2011). All isolates were screened via a multiplex qPCR amplification of *hrpB7* as described in Strayer et al. (2016) to compare both strain discrimination methods and as a more robust species identification. For this study, we focused on further characterization of strains identified as *X. perforans*. The species level survey is described in a previous study (Bernal, 2020).

Plant Material

Tomato line Ohio 88119 (Berry et al., 1995) and its near-isogenic sibling OH813A, having *Xv3/Rx4* (Bernal et al., 2020; Bernal and Francis, 2021) were used as susceptible and resistant controls to distinguish between *X. perforans* race T3 and T4 strains in all experiments. In addition, we tested isolates for HR against LA0716 as a control for *RXopJ4* interaction with the bacterial effector *avrXv4*. Tomato plants were grown in 1-gallon pots in a greenhouse managed by the Department of Horticulture and Crop Science at The Ohio State University, Wooster, OH, United States. The greenhouse temperature fluctuated between 22 and 27°C. The relative humidity was maintained between 50 and 70%.

Fresh-market and processing tomato varieties commonly used in the Midwestern United States were tested for the presence of the candidate gene for the *Xv3/Rx4* resistance locus. These include fresh-market varieties Camaro, Camp, FL91, FL47, Red Bounty, Southern Ripe, Grand Marshall, Bounty, Mariana, Southern Ripe and Tasti-Lee. Processing varieties were from OSU (Ohio and OX), Red Gold (GEM), Heinz Seed (H), Harris-Moran Seed (HM), Seminis Vegetable Seeds (Peto), Rispins (R), and Tomato Solutions (TSH) and included Ohio 8611, OX325, GEM 111, GEM183, GEM818, H1015, H101A, H1301, H1648, H1765, H1766, H2206, H3406, H3602, H5108, H7222, H8504, H9144, H9364, and H9706. H9996, H9997, HM1823, HM1892, HM3887, HM5900 HM8901, HM9903, Peto 696, R522695, TSH04, TSH05, TSH12, TSH18, TSH20, and TSH40. We used the marker PCC12 to indicate the presence of the *Xv3/Rx4* locus as described below (Pei et al., 2012).

Inoculum Preparation and Hypersensitive Response Evaluation

Bacterial suspensions were prepared from 48 h-old culture plates initiated from single colonies and maintained at 28°C in darkness on yeast dextrose calcium carbonate (YDC) agar. Bacterial suspensions were prepared in sterile deionized water, and the optical density was adjusted to 0.15 optical density (OD) at an absorbance of 600 nm ($\sim 10^7$ CFU/ml). Hypersensitive response (HR) leaf assays were performed within 2 h following the preparation of the suspension. The HR phenotype was evaluated to identify the race of 106 *X. perforans* strains and two positive controls, SM761 (T3) and Scott-1 (T4). This subset of strains comprised isolates from 2012 (16), 2013 (4), 2017 (29), 2018 (42), 2019 (15).

For evaluation of HR, inoculations were performed by leaf infiltration of 8–10-week-old potted plants of OH88119, OH813A, and LA0716 in the Williams Hall greenhouse, Wooster, OH. The plants were placed in a completely randomized design throughout the greenhouse. Bacterial suspensions were infiltrated into the underside of fully expanded leaflets using a 1-ml syringe (Robbins et al., 2009). The area of infiltration was 1–2 cm in diameter. Leaflets were inoculated between 10:00 AM-12:00 PM, and the HR phenotype was visually inspected and recorded daily for 5 days post inoculation (dpi). Five independent inoculation experiments were conducted between 2017 and 2019. Approximately 25 strains were evaluated at each time, with the addition of race T3 and T4 controls (Sahin, 1997; Timilsina et al., 2016). Three plants for each tomato genotype were syringe-inoculated with each strain to observe repeatability of HR symptoms. Inoculated leaflets were given a score of 1 = HR phenotype (confluent necrosis), 2 = susceptible response (necrosis expansion beyond inoculation area), 3 = no symptom development or unclear HR. In most cases, HR symptoms were most noticeable after 72 h. Strains that displayed the same phenotypic response in all biological reps ($n = 3$) within genotype were scored. Strains that displayed differences among the three replicates were re-tested.

Bacterial and Plant DNA Isolation

DNA from bacterial isolates was extracted using the boiling method (Abrahamian et al., 2018) modified by the addition of a chloroform extraction. Single colony glycerol stocks were streaked and grown on nutrient broth yeast (NBY) medium for 48–72 h. Bacteria from a single colony was transferred into a 1.5 ml microcentrifuge tube and re-suspended in 400 ml of sterile water. Tubes were boiled at 100°C for 15 min in a dry-bath (Thermo Fisher Scientific, Waltham, MA, United States), and then immediately placed on ice for 5 min followed by centrifugation at 10,000 rpm for 5 min. An aliquot of 100 μL of isolated DNA was mixed with 300 μL of Tris EDTA buffer [10 mM Tris, 0.1 mM EDTA], and 400 μL of chloroform isoamyl alcohol (24:1) followed by centrifugation at 5,000 × *g* for 15 min. About 150 μL of the aqueous phase was transferred to a 96-well round bottom microtiter plate (Corning Inc., Kennebunk, ME, United States) containing 15 μL of sodium acetate (3M, at pH 5.3) and 330 μL of 70% ethanol followed by centrifugation at 5,000 × *g* for 10 min for DNA precipitation. Plates were air-dried and 100 μL of ddH_20 was added. For DNA isolation of plant material we used methods described in Bernal et al. (2020).

Molecular Marker Analysis of *X. perforans* Races T3 and T4

Primers were designed to amplify *avrXv3* based on an alignment of wild-type and known variant alleles using the Geneious Prime software (Kearse et al., 2012). The reverse primer, avrXv3R, 5′-TGAGCGAGAGCTACTATCGCCTCC-3′

was previously described (Timilsina et al., 2016). The forward primer, avrXv3ef1, 5′-GGAAGCTTGGATTAAAGGGG-3′ was developed for this study such that the entire gene and known variants would be amplified. These primers flank *avrXv3* with a predicted amplicon of 358 bp, which spans a known SNP, and a 30 base indel upstream of the gene (Timilsina et al., 2019). All amplifications included controls, T3, T4, *X. euvesicatoria* and *X. hortorum* pv. *gardneri*. The *avrXv3* gene was PCR-amplified in 564 *X. perforans* strains to determine the allele associated with the HR phenotype. Polymerase Chain Reaction (PCR) was conducted in 20 μL reactions consisting of 2 μL of Buffer A [10 mM Tris–HCL, 50 mM KCl, 1.5 mM $MgCl_2$,], 0.8 μL of 1.25 mM dNTP, 0.2 μL of 10 μM forward and reverse primer, 12.4 μL of ddH_20, 0.4 units of *Taq* DNA polymerase, and 4 μL of DNA template using the following cycling conditions: (1) 94°C for 3 min, (2) 94°C for 1 min, (3) 63°C for 30 s, (4) 72°C for 1 min, (5) 72°C for 5 min. Steps 2–4 were repeated 35 times prior to the final annealing step. Amplicons were evaluated based on restriction digestion and/or polymorphism detected based on agarose gel electrophoresis. The presence of the known SNP205nt was evaluated using the restriction enzyme *Alu*I (New England Biolabs Incorporated, Ipswich, MA, United States) in a cleaved amplified polymorphic (CAP) detection approach. PCR products were digested by combining 16 μl of PCR product with 1.0 μL of 1X CutSmart® Buffer (New England Biolabs Incorporated, Ipswich, MA, United States), 0.4 μL of *Alu*I (10,000 units/ml), and 2.6 μL of ddH_20 for 3 h. Amplicons were separated on a 2% agarose gel at 180V for 1 h and 30 min and detected using ethidium bromide fluorescence.

Whole Genome Sequencing of *X. perforans* Strains

In total, 38 *X. perforans* strains from the Midwestern United States were selected for whole genome sequencing. The criteria for selecting a subset from the 564 strains for whole genome sequencing was based on (1) year of isolation, (2) location, (3) allele at the *avrXv3* locus, and (4) HR response (**Supplementary Table 1**). Strain SM1852 was included as a representative of those with a failure to amplify using the *avrXv3* primers. Isolates were selected from clean single colony cultures stored in glycerol for DNA extraction. A total of 14 unique strains were sequenced using an Illumina MiSeq System (Illumina Inc.) by University of Florida. Reads were assembled using SPAdes v.3.11 (Bankevich et al., 2012). A detailed pipeline for assembly and filtering process of genomes was described previously (Timilsina et al., 2019). The remaining 24 unique *X. perforans* strains were sequenced using an Illumina iSeq 100 System with the NextEra DNA Flex Library Prep protocol. One strain, SM1852-18 was sequenced three times over two sequence runs. The same glycerol stock for SM1852-18 was used in all cases. In the first run (R1) an independent single colony was used as a source of DNA. We consider this to be a biological replicate. In the second run, a different colony was selected, and the same DNA was used with two different sets of adapters (R2 and R3). We consider these to be technical replicates. This process of sequencing the same DNA was conducted as a control to determine error rates. FastQC (Andrews, 2010) was used for quality control of the raw sequence data. Reads were assembled using SPAdes v.3.14.1 (Bankevich et al., 2012). Reads and assemblies are available through the National Center for Biotechnology Information (NCBI) under BioProject accession PRJNA795842. These assemblies were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAG).

Characterization of *avrXv3* Alleles and Presence of *avrBsT* and *avrXv4*

A total of 108 sequences of strains available in public databases were narrowed to 24. The criteria for selecting the subset of 24 from the 108 publicly available sequences was (1) location, (2) year of isolation, (3) allele at the *avrXv3* locus, and (4) previously defined *X. perforans* clade or group (Schwartz et al., 2015; Newberry et al., 2019; Timilsina et al., 2019). Passport data that emphasize these criteria are included in **Supplementary Table 1**. For the analysis presented here, the genomes of 64 *X. perforans* strains, 38 strains collected between 2017 and 2020 in the Midwest, two biological and technical replicates, and 24 sequences of bacterial strains available in public databases were utilized to characterize alleles of *avrXv3*, *avrBsT*, and *avrXv4*. A 1.5 kb region including the *avrXv3* gene from *X. perforans* strain CFBP7293 (NZ_MOLQ01000001.1) was extracted and searched in a custom database containing the 64 *X. perforans* genome sequences formatted as FASTA files. The FASTA sequences were then formatted as a Basic Local Alignment Search Tool (BLAST) database and searched using BLASTN with default settings in the Geneious Prime software (Altschul et al., 1990; Kearse et al., 2012). The 1.5 kb region from all strains was aligned and mapped to CFBP 7293 using Multiple Sequence Comparison by Log-Expectation (MUSCLE) version 3.8.425 (Edgar, 2004). To further characterize novel *avrXv3* alleles we utilized Burrows-Wheeler aligner (BWA), BWA-MEM algorithm, v.0.7.17 (Li, 2013) to map the raw reads to *avrXv3* using the reference T3 strain Xp91-118. SAMtools v1.13 (Li et al., 2009) was utilized for post-processing of alignments. The Integrative Genomics Viewer (IGV) v2.8.13 (Robinson et al., 2011) was utilized to visualize mapped reads of sequenced strains using the "show read pairs" option. To determine the presence of *avrBsT* and *avrXv4* we extracted the gene sequences from *X. perforans* strain TB6 contig15 (NZ_JZWA01000015.1) and *X. perforans* strain NC373 (MK583029.1), respectively. This sequence was used for BLAST and MUSCLE alignment as described for *avrXv3*.

Average Nucleotide Identity Pairwise Analysis

We performed an average nucleotide identity (ANI) pairwise analyses to determine whole genome sequence variation between 24 reference strains, 38 *X. perforans* strains sequenced from our 2017 to 2020 collection, and two controls. ANI for *X. perforans* strains was completed using the enviomics toolbox ANI calculator from the Kostas lab (Rodriguez-R and Konstantinidis, 2016). The calculator estimates the ANI between genomes by comparing 1 kb intervals (Goris et al., 2007). The approach uses reciprocal best matches as a way to measure genetic relatedness.

The calculator uses the BLASTN algorithm with X = 150 (where X is the drop-off value for gapped alignment), q = 21 (where q is the penalty for nucleotide mismatch) and F = F (where F is the filter for repeated sequences) (Goris et al., 2007). The pairwise ANI analysis quantifies genome wide identity based on SNPs and gaps between strains.

Core Gene Analysis

A core-gene phylogenetic tree for *39 X. perforans* strains was created to assess the diversity within strains. Five housekeeping genes *lepA*, *gyrB*, *lacF*, *gapA*, and *gltA* were identified in these 39 strains and concatenated (Timilsina et al., 2015). The sequences of housekeeping gene accessions were extracted from NCBI as follows: *lepA* (KM492599.1), *gyrB* (KM492331.1), *lacF* (KM492465.1), *gapA* (KM492063.1), and *gltA* (KM492197.1). Sequences identified with the Basic Local Alignment Search Tool, BLASTN, in the Geneious Prime software (Altschul et al., 1990; Kearse et al., 2012). The phylogenetic tree was developed using the Geneious tree builder (Kearse et al., 2012). Standard parameters were set using a global alignment with free end gaps, a cost matrix of 65% similarity, Tamura-Nei genetics distance model, and a neighbor-joining tree method. For the pair-wise alignment option for building the distance matrix a gap open penalty and gap extension penalty of 12 and 3 were used, respectively.

Phylogenomic Analysis

Phylogenomic relatedness between *X. perforans* strains was assessed using a whole-genome SNP genotyping alignment method. The Harvest suit package, Parsnp v.1.2 (Treangen et al., 2014) utilizes whole-genome alignment, read mapping, and k-mer analyses for SNP identification. The Parsnp output produces a core genome alignment, SNP variant calls, gap analysis, and a phylogenomic tree (Treangen et al., 2014). Parsnp was used to align all 64 *X. perforans* genomes (including strain replicates) to Xp91-118 reference T3 strain. The sequence for Xp91-118 is the most complete genome for *X. perforans* and is extensively used as a reference (Abrahamian et al., 2019). Alignments were implemented using the log-expectation function in MUSCLE with trees rooted to Xp91-118 based on maximum-likelihood nearest-neighbor interchanges as implemented in FastTree2 (Price et al., 2010). Gingr v1.3, a graphic user interface from the Harvest suit package, was used to visualize the phylogenomic tree. The newick formatted text file was extracted from Gingr v.1.3 and the Interactive Tree of Life (iTOL) v5 (Letunic and Bork, 2021) online tool was used to display and annotate the phylogenomic tree. The Parsnp variant call output consisted of 23,203 SNPs that were used to determine experimental error and to assess polymorphism between and among strains.

Experimental error was determined by comparing SNPs between the SM1852-18 control strain sequenced three times (SM1852-R1, SM1852-R2, and SM1852-R3). The SNP comparison for the three controls was used to set a baseline for error that is a composite of sequencing error and assembly error but will not reflect systematic errors in assembly. This error rate was estimated based on the summation of all SNPs between technical replicates divided by the size of the reference genome (4,898,349 bp), Xp91-118 used in the alignment. Strains displaying a SNP polymorphism rate lower than or within the range detected for technical resequencing of control strain replicates were regarded as identical within the resolution of our methods.

Decay of Linkage Disequilibrium

Analysis of LD was performed using the R package "sommer" (Covarrubias-Pazaran, 2016, 2018). The matrix of 23,203 SNPs was extracted from the Parsnp variant call output and transformed such that accession sequences were row names and SNP (locus) were column headings. The SNP genotyping matrix consisted of "0" for reference genome and "1" for alternative allele calls. The locus name and position were extracted and combined with a vector for linkage group = 1, and this matrix provided the "map" file for analysis using the LD.decay function.

RESULTS

Changes in *X. perforans* Race Distribution

The species makeup of *Xanthomonas* strains collected between 2017 and 2020 in the Midwest has changed relative to previous reports (Sahin, 1997; Ma et al., 2011; Ma, 2015). *X. perforans* T4 now appears to be the predominant species and race based on hypersensitive response assays, and sequence analyses of *avrXv3* (Bernal, 2020; Rotondo et al., 2022). In 2012 *X. perforans* race composition was approximately 40% T3 and 60% race T4. Since 2017, race T4 has accounted for over 93% of isolated *X. perforans* strains (**Figure 1**). The tomato genotype OH813A displayed a clear HR when inoculated with a race T3 control strain, SM761, and water-soaked necrosis when inoculated with the T4 control strain, Scott-1. In all assays OH88119 was susceptible to both control strains and no HR was observed (**Figure 2A**). LA0716 displayed HR when inoculated with control strains. Out of 106 strains tested >95% were identified as race T4 based on HR response to tomatoes carrying *Xv3*/*Rx4* and *RXoJ4*.

We screened 564 *X. perforans* strains using PCR amplification and identified a known SNP variant within *avrXv3* with a high frequency (94%). This SNP was previously described at the 205 nucleotide position in *avrXv3* (Klein-Gordon et al., 2021), and was only found in isolates collected between 2017 and 2020 in the Midwest. No amplification of *avrXv3* was observed for *X. euvesicatoria* and *X. hortorum* pv. *gardneri* controls as expected, though faint bands that we ascribe to non-specific amplification were observed (**Figure 2B**). Several strains collected between 2010 and 2013 contained a previously described T4-Contig variant (Klein-Gordon et al., 2021). For thirty seven *X. perforans* strains (∼6.5%) *avrXv3* was not amplified. One of these strains, SM1852-18, was used for whole genome sequence and we detected a ∼4 Kb deletion which extended beyond. The coding sequence of *avrXv3* (see section "Characterization of *X. perforans* Strain SM1852-18 Lacking *avrXv3* and *avrBsT*" below). We did not detect the presence of novel insertion deletion variants or mutations within the coding sequence of

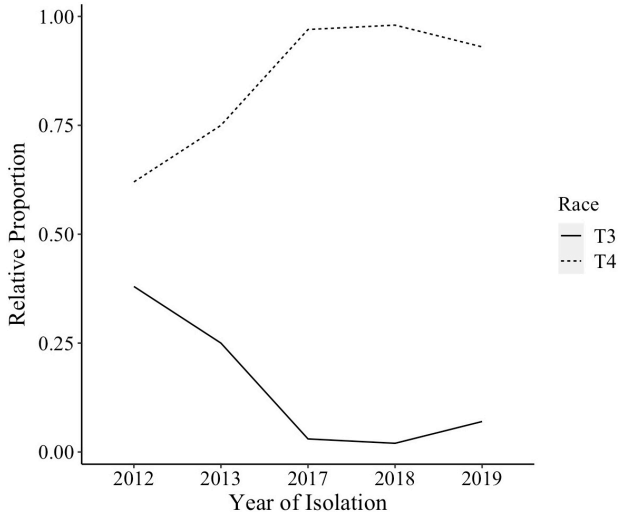


FIGURE 1 | *Xanthomonas perforans* race composition for strains isolated from tomatoes between 2012 and 2019. *X. perforans* strains isolated in 2012, 2013, 2017, 2018, and 2019 were screened using hypersensitive response (HR) assay to determine their race. By 2017, 93% of all *X. perforans* strains were categorized as race T4 based on hypersensitive response and PCR amplification of *avrXv3*.

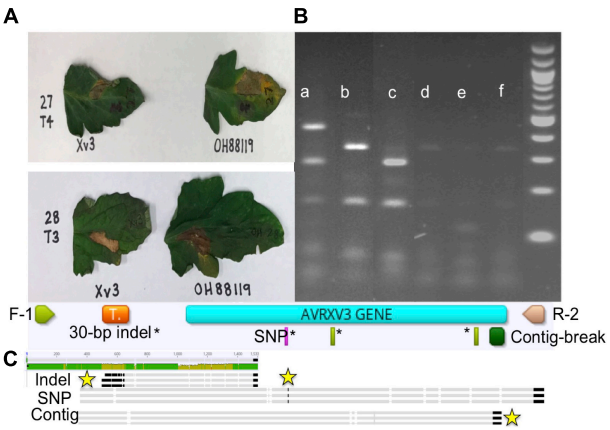


FIGURE 2 | Hypersensitive response (HR), genotype banding patterns for avrXv3 mutations, and graphical presentation of known AvrXv3 mutations. **(A)** Xv3 OH813A with Xv3/Rx4 inoculation elicits an HR in the presence of *Xanthomonas perforans* race T3 (lower panel) and not T4 (upper panel). OH88119 was utilized as a control to show susceptibility to both races. **(B)** Shows PCR amplification and banding patterns of *X. perforans* race T4 mutations and race T3. (a) *X. perforans* race T4 (Scott-1) digest band size 496, 280, and 164 bp, (b) *X. perforans* race T3 (SM761) digest band sizes 332 and 164 bp, (c) T4-SNP (1828-17) digest band sizes 280 and 164 bp (d,e) *X. euvesicatoria* (110C and 767), (f) *X. hortorum* pv. *gardneri* (SM775-12) show weak bands due to non-specific amplification as those *Xanthomonas* species do not possess *avrXv3*. A 1 kb ladder (far-right) was used as a size standard, each band represents 100 bp. **(C)** F-1 and R-2 refers to the flanking primers used to amplify avrXv3. The orange box represents the location of a previously known pseudogene. The pink rectangle highlights the location of a known SNP mutation, which also corresponds to a restriction site for *Alu*I. The green rectangles* represent other cut sites for *Alu*I. The large dark green box represents the location of a known mutation creating a contig break. Sequence alignment of *X. perforans* strains with different mutations is shown at the end. The yellow stars mark the location of the mutations.

avrXv3 (**Figure 2C**). The presence of the previously characterized SNP205nt variant and a lack of other variation was verified using Sanger sequencing of amplicons (data not shown). There was a strong association between the HR phenotype and the PCR-based genotypic screening of *avrXv3,* with 101 of 106 strains showing complete agreement. Five strains were phenotypically classified as T3 but displayed the SNP205nt variant in the *avrXv3* locus.

Tomato Germplasm

Forty-three commercial varieties were tested for the presence for the *Xv3/Rx4* allele conferring HR to *X. perforans* race T3. Heinz varieties H2206, H5108, and H9997 and the Harris-Moran variety HM6900 had the *Xv3/Rx4* allele associated with HR in the presence of *avrXv3*. H5108, an early season hybrid, accounts for less than 7% of the processing tomato acreage in Ohio and Indiana. Varieties H2206 and H9997 are considered experimental and are not widely used. None of the fresh-market varieties tested possessed the *Xv3/Rx4* allele associated with HR.

Whole Genome Sequence Analyses of *X. perforans* Strains

Sequence reads and assemblies are available through NCBI Bioproject PRJNA795842. Sequences generated through iSeq ranged from 6x to 10x coverage while sequences generated on the MiSeq ranged from 13x to 17x. Biological and technical replication was used to estimate an experimental error as determined by resequencing and reassembly prior to phylogenomic analysis (**Table 1**). This error estimate quantifies the number of single-nucleotide differences between replications and accounts for sequence and assembly error but will not account for systematic errors due to methodologies nor the potential for mutations between single colony isolates. We refer to experimental error in this context. A total of 23,203 SNPs were extracted from Parsnp to characterize the genomic relatedness between *X. perforans* strains. SNPs between specific *X. perforans* strains were calculated relative to the whole genome size (4,898,349 bp) of *X. perforans* strain 91–118 (**Table 1**). The allele mismatch between technical reps, SM1852-R1 (run 1, rep 1), R2 (run 2, rep 2), and R3 (run 2, rep 3) was between 23 and 67 SNPs. Comparison of SM1852-R2 vs. SM1852-R3 showed a difference of 50 SNP polymorphisms with an allele mismatch of 1.02×10^{-5} relative to the whole genome (**Table 1**).

After establishing an error based on resequencing and reassembly, we characterized the allele mismatch between reference strains and sequenced isolates collected from Midwest between 2017 and 2020. The sequences for isolates GEV893 and GEV909 from Collier County, FL displayed an allele mismatch of 1.84×10^{-6}. Comparisons between GEV893 and strains from the Midwest displayed a similar, or lower, allele mismatch relative to the SM1852-R1, R2, and R3 controls (**Table 1**). These mismatch estimates fall within our technical error and are consistent with expected mutation rates of genes per generation

TABLE 1 | Pairwise analysis of polymorphism between technical replicates, biological replicates, and strains.

Comparison 1				Comparison 2							
Isolate[a]	**Xp Group[b]**	**Location**	**Year**	**Isolate**	**Xp Group**	**Location**	**Year**	**SNP[c] no.**	**SNP[d] mismatch**	**GAP[e] analyses**	**CDS[f] diff.**
SM1852-R1	Group 1B (SC2)	OH	2018	SM1852-R2	Group 1B (SC2)	OH	2018	67	1.37E-05	0.79%	58
SM1852-R1	Group 1B (SC2)	OH	2018	SM1852-R3	Group 1B (SC2)	OH	2018	23	4.70E-06	0.15%	72
SM1852-R2	Group 1B (SC2)	OH	2018	SM1852-R3	Group 1B (SC2)	OH	2018	50	1.02E-05	0.45%	14
Xp17-12	Group 1B (SC2)	FL	2006	SM1852-R1	Group 1B (SC2)	OH	2018	28	5.72E-06	0.41%	92
GEV893	Group 1A (SC1)	FL	2012	GEV909	Group 1A (SC1)	FL	2012	9	1.84E-06	0.20%	29
GEV893	Group 1A (SC1)	FL	2012	SM1806	Group 1A (SC1)	OH	2017	9	1.84E-06	0.16%	76
GEV893	Group 1A (SC1)	FL	2012	SM1856	Group 1A (SC1)	OH	2018	46	9.39E-06	0.14%	77
GEV893	Group 1A (SC1)	FL	2012	SM234	Group 1A (SC1)	OH	2019	60	1.22E-05	0.12%	86
GEV893	Group 1A (SC1)	FL	2012	SM176	Group 1A (SC1)	OH	2020	64	1.31E-05	0.16%	80
SM1843	Group 1A (SC1)	OH	2017	SM1011	Group 2 (SC3)	OH	2013	5341	1.09E-03	3.01%	318
SM1807	Group 1A (SC1)	IN	2017	SM1809	Group 1A (SC1)	OH	2017	46	9.39E-06	0.08%	8
SM1807	Group 1A (SC1)	IN	2017	GEV893	Group 1A (SC1)	FL	2017	9	1.84E-06	0.20%	75
SM1809	Group 1A (SC1)	OH	2017	GEV893	Group 1A (SC1)	FL	2017	53	1.08E-05	0.10%	79
SM1807	Group 1A (SC1)	IN	2017	SM779	Group 2 (SC3)	OH	2012	5279	1.08E-03	3.01%	581
SM1809	Group 1A (SC1)	OH	2017	SM779	Group 2 (SC3)	OH	2012	5323	1.09E-03	3.03%	577
GEV893	Group 1A (SC1)	FL	2012	SM779	Group 2 (SC3)	OH	2012	5280	1.08E-03	2.95%	656

[a] *Isolate names are referenced in* ***Supplementary Table 1****. SM1852-R1 is a single colony isolate sequenced on the first run, SM1852-R2 and SM1852-R3 are from the same DNA source using separate sets of adapters during library preparation prior to sequencing.*

[b] *Xp Group refers to phylogenetic clustering described previously (Schwartz et al., 2015; Newberry et al., 2019; Timilsina et al., 2019).*

[c] *SNP no., the number of single nucleotide polymorphisms (SNPs) between two isolates.*

[d] *SNP mismatch, number of SNPs relative to whole genome size of Xp 91-118 (4,898,349 bp).*

[e] *GAP analyses, percent of total unaligned sequences displaying gaps in pairwise comparisons with ANI.*

[f] *CDS diff., pairwise comparison of the number of coding sequences derived from the NCBI Prokaryotic Genome Annotation Pipeline (PGAG).*

(Balin and Cascalho, 2010; Mhedbi-Hajri et al., 2013). These results provide evidence that these strains may be clonal based on their near identity.

Information extracted from the annotation of sequences provided comparative estimates of the number of coding sequences (CDS) while ANI provided pairwise SNP and alignment gap information (**Table 1**). Technical reps SM1852-18-R2, and SM1852-18-R3 have a difference of 14 CDS though the same DNA was used for both runs. SM1852-18 biological reps, R1 and R3, show a difference of 72 CDS. When comparing annotations of individuals collected between 2017 and 2020, we observe CDS differences lower than noted for our technical reps. For example, SM1807-17 and SM1809-17 have a difference of 8 CDS even though these were isolated in Indiana and Ohio, respectively. The ANI analysis provides another approach to assessing the potential for loss or gain of sequences based on open gap analysis. In comparing technical replicates R2 and R3 open gaps account for 0.45% of the genome. In comparing R1 and R2, open gaps accounting for 0.79% of the genome were found while comparisons of R1 and R3 reveal gaps accounting for 0.15%. Comparing strains known to differ in gene content, SM1807 with the SNP205nt variant of AvrXv3 and SM779 (contig), revealed gaps accounting for 3% of the genome. A comparison of SM1809 and GEV893 identified gaps accounting for only 0.1% of the genome. These results suggest gene-loss or addition are not accounting for large amounts of variation between genomes especially when comparing strains with near-identity based on SNP analysis.

Phylogenomic Whole Genome Analyses of *X. perforans* Strains

A preliminary comparison of phylogenetic analysis was conducted using three measures of genome variation. Comparison of five core genes, genome wide ANI, and core genome alignment and SNP detection relative to a reference strain Xp91-118 displayed similar structure (**Supplementary Figure 1**). Within the Group 1A (SC1) clade, 16 group identically between methods. However, the core-gene tree contains an additional 7 strains which do not have the SNP205nt allele. These strains are BRIP62383, BRIP2389, BRIP62404, 91-118, Xp5-6, 4P1S2, and CFBP7293. The T4-Contig strains [Group 1B (SC2)] group identically between methods. ANI and whole-core-genome approaches offered higher resolution, likely due to the increase in sequence comparisons. Therefore a phylogenomic approach based on core genome alignment and pairwise ANI was utilized to evaluate diversity for a larger set of *X. perforans* sequenced strains.

To determine the genomic diversity across the 38 strains sequenced in this study, a subset of 24 publicly available genomes showing variation across *avrXv3* were selected for reference. These 24 genomes were selected from a larger data set, showing unique sequence variation in *avrXv3* previously described (Timilsina et al., 2016; Newberry et al., 2019). In addition, genomes were selected across time and space based on collection of the original strains to further investigate potential geographical stratification of *X. perforans* isolates. Strains sequenced from 2010, 2011, 2012, and 2013 were classified as T3-like or T4. The designation T3-like refers to strains that have a complete *avrXv3*, but have not been evaluated for HR. T3-like strains SM1013-13 and SM587-11 were more genetically similar to the reference strain Xp91-118 race T3 (**Figure 3**). Sequenced strains SM176-10, SM779-12, and SM1011-13 from 2010, 2012, and 2013, respectively, were collected when *X. hortorum* pv. *gardneri* was the predominant species in the Midwest (Ma et al., 2011; Ma, 2015) and contained a T4-Contig variant. This sequence variant creates a break in the coding sequence of *avrXv3*, similar to other strains from Florida. These strains clustered under Group 2 (**Figure 3**), previously described (Schwartz et al., 2015; Timilsina et al., 2019).

The majority of *X. perforans* strains isolated from the Midwest between 2017 and 2020 clustered within the same clade as Florida strains GEV893 and GEV909, isolated in 2012 (**Figure 3**). This genetic cluster has previously been described as Group 1A, or sequence cluster 1 (SC1) (Newberry et al., 2019; Timilsina et al., 2019). ANI was calculated between strains and clusters within the phylogeny to further quantify relationships. For example, the ANI between SM1852-R2 and SM1852-R3 was 100% SD 0.02 while SM1852-R1 and SM1852-R2 was 100% SD 0.11%. In contrast, Group 1A sequence cluster SC1 strain SM1843-17 and Group 2 strain SM1011-13 were 99.78% SD 0.83% identical based on ANI (**Figure 3**). Comparison of whole genome sequence for GEV893 and isolates from the Midwest collected between 2017 and 2020, including SM194-20 collected from a volunteer plant in a wheat field, displayed an ANI between 99.99% and 100.00% (**Figure 3**). These strains all had the SNP205nt *avrXv3* variant and the effector *avrBsT*. Phylogenetic comparison of GEV893 with strains from the Midwest isolated between 2017 and 2020 suggests that Group 1A cluster SC1 predominates and is likely clonal. Further evidence of clonality is suggested by a lack of LD decay across the entire genome (data not shown).

One strain, SM1852-18 did not cluster in Group 1A. All three technical replicates of strain SM1852-18-R1, R2, and R3 clustered with the strain Xp17-12 previously described (Schwartz et al., 2015; Abrahamian et al., 2019). Strain Xp17-12 was collected in Florida in 2006 and was nearly identical to SM1852-18-R2 based on ANI (100% SD 0.08%). The strains SM1852-18 and Xp17-12 do not contain *avrXv3* or *avrBsT* (described further below).

Characterization of *X. perforans* Strain SM1852-18 Lacking *avrXv3* and *avrBsT*

SM1852-18 did not induce a hypersensitive response on OH813A and failed to amplify *avrXv3* in PCR, consistent with a deletion of *avrXv3*. Whole-genome sequence analyses of *X. perforans* strain SM1852-18 confirmed the absence of effectors *avrXv3* and *avrBsT*. A ∼4 Kb region encompassing *avrXv3* from Xp91-118 showed no reads mapped from SM1852-18 (**Figure 4**). The presence/absence of avrBsT was based on BLASTN searches and SM185218 lacked this effector as well (data not shown). All other sequenced strains classified as T4-SNP or T4-Contig contained both the effectors *avrBsT* and *avrXv4*. The presence of *avrBsT* was expected, as the effector is common

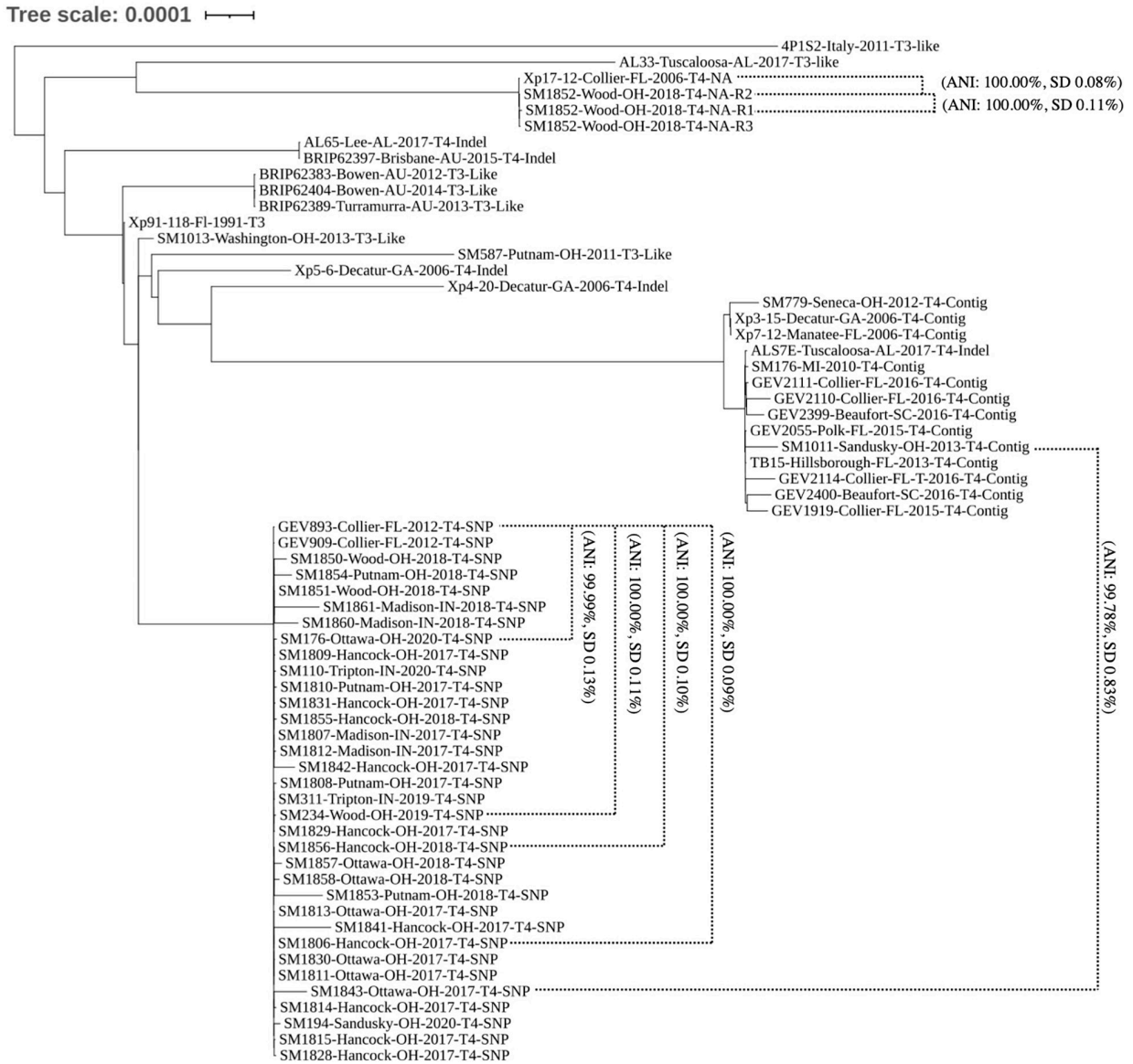


FIGURE 3 | Phylogenomic tree of *X. perforans* isolates geographically distributed worldwide. Phylogenomic tree of 64 *Xanthomonas perforans* strains based on whole genome sequencing. Strains are denoted by the name, location of isolation, year, race-type, and known mutation in *avrXv3*. NA refers to the absence of *avrXv3.* Two-way-Average Nucleotide Identity values and standard deviations are denoted for specific comparisons between and within clades. The majority of strains isolated and sequenced in 2017–2020 are genetically identical. These strains group under the same clade as strains GEV909, GEV893 isolated from Florida in 2012. These strains all have the same SNP mutation in *avrXv3*.

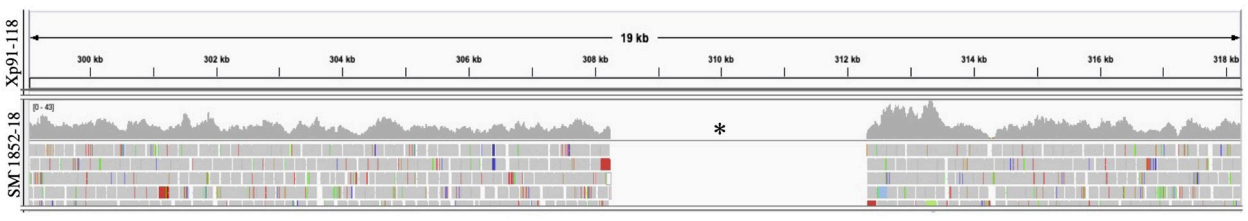


FIGURE 4 | Mapping of sequence reads to *avrXv3* from Xp91-118. Burrows Wheeler Alignment was utilized to map reads *from X. perforans* isolate SM1852-18 to Xp91-118. Integrative Genome Viewer was utilized to visualize mapped reads. No reads (denoted with a black asterisk) from SM1852-18 mapped to Xp91-118.

in both Group 1a/SC1 and Group 2/SC3, previously described (Newberry et al., 2019).

DISCUSSION

Our results validate a genomics-based framework to track shifts in *X. perforans* populations due to migration, mutation, drift, or selection. We accessed 108 *X. perforans* whole-genome data sets from public sources and distilled these into 24 reference strains representing diversity within the species, major clades (Timilsina et al., 2019), date of collection, and geographic origin of the isolate. We evaluated 106 *X. perforans* strains using HR to distinguish between race T3 or T4 and we utilized PCR to amplify *avrXv3* in 564 Midwestern isolates and controls strains. Our results demonstrate that *X. perforans* race T4 is currently the predominant species and race in the Midwestern United States. These results parallel those reported for Florida, Indiana, and North Carolina (Timilsina et al., 2016; Egel et al., 2018; Adhikari et al., 2019; Klein-Gordon et al., 2020).

We observed a strong correlation between HR response and the PCR amplification assay for the *avrXv3* effector. The error rate of <5% could have resulted from a misclassification of HR in OH813A, despite repeated tests. However, it is also naïve to assume that loss-of-function SNPs and insertion/deletions are the sole reason for loss of HR. It is possible that regulatory mechanisms affecting expression and or modification of the *avrXv3* effector might also account for this observed variability associated with genetic characterization and HR phenotype.

We utilized whole-genome phylogenomic analyses to investigate sequence variation between the 24 reference strains and 38 newly sequenced genomes from the Midwestern United States. Based on the characterization of strains collected between 2017 and 2020, we determined that the most abundant clone (37 of 38 sequenced strains) belongs to the Group 1A/SC1 clade of *X. perforans,* contains a known SNP variant in *avrXv3* (SNP205nt), and both *avrBsT* and *avrXv4.* The strains collected between 2017 and 2020 differ from strains collected in the same region between 2010 and 2013 and show sequence identity, within the bounds of experimental error, with a strain collected in Florida previously. Although these analyses do not provide insight into how the Group 1A/SC1 clone was introduced, transplant shipment from the South, seed introduction, or weather patterns all offer possible routes. We also identified a strain, SM1852-18, lacking *avrXv3* and *avrBsT*. This strain and other isolates that failed to amplify *avrXv3* account for less than 1.2% of the current Midwestern *X. perforans* population. Phylogenetic analysis place SM1852-18 in the previously described Group 1B/SC2 clade. ANI demonstrated near identity to Xp17-12, isolated in Florida in 2006, suggesting a migration mechanism for even rare constituents of the population. This strain accounts for nearly 16% of 280 sequenced from Florida in recent years (Klein-Gordon et al., 2021). The sequence differences detected within strains classified as Group 1A/SC1 are consistent with known mutation rates for Xanthomonas which range from 10^{-6} to 10^{-5} base changes per generation (Mhedbi-Hajri et al., 2013) and are within the range of experimental error based on technical replication. The lack of sequence polymorphism between *X. perforans* strains collected across space and time, the lack of LD decay, and the phylogenetic clustering suggests an established clonal population. Clonal populations are defined by having two distinct features, high linkage disequilibrium (LD) and strong phylogenic signal (Tibayrenc and Ayala, 2012) consistent with the *X. perforans* strains analyzed here and suggests that the strains we characterized are descended from a single common ancestor that has been maintained for multiple years.

It is unlikely that Group 1A/SC1 is reintroduced each year on seed. Hybrid tomato seed tends to be produced in India, Thailand, and China with production locations varying by producer and current costs. It is unlikely that the diversity of producers, varieties, and seed lots sampled would all carry a single clone. If continuous migration through seed exchange was occurring every year, or if strains were moving with weather annually, we would expect to find greater variation in Midwestern states as new species and genotypes are introduced annually. The occurrence of the T4-SNP clone across varieties, regions, growers would seem to rule out an epidemic originating from contaminated seed and reestablishing each year. Xanthomonads surviving in transplant facilities could also explain the maintenance of clonal populations. More recent *X. perforans* population studies across Florida commercial tomato fields have shown that strains could be associated with tomato production systems and transplant facilities (Klein-Gordon et al., 2020, 2021). It is possible that the T4-SNP clone was introduced into the Midwestern United States via seed or transplants and is now surviving in transplant greenhouses, in fields, or on alternate hosts.

We identified a *X. perforans* T4-SNP Group 1A/SC1 isolate from a volunteer tomato plant found in a wheat field in 2020. An early description of bacterial spot suggested that *Xanthomonas* spp. were capable of overwintering in Indiana (Gardner and Kendrick, 1921). It was reported that volunteer tomato plants with bacterial spot symptoms were observed in a corn field, even though no tomato fields were planted nearby. The disease was present the year before, suggesting that the bacteria overwintered in debris or seed from the field (Gardner and Kendrick, 1921). In a different study it was shown that *X. vesicatoria* contributed 0.3% of the field soil microbial population in winter and spring sampling in Indiana, and was attributed to continuous inoculum from standing stalks of dead tomato plants (Peterson, 1963). Other studies have shown and/or suggested that xanthomonads can overwinter on wheat, weeds, and plant debris (Diachun and Valleau, 1946; Leben, 1981). Identification of T4-SNP Group 1A/SC1 isolates on volunteer tomato plants provides evidence of a possible overwintering of *X. perforans* in the Midwest.

Mutation in *avrXv3* does not appear to be driving population shifts in the absence of widespread resistance. The SNP205nt variant and the strain lacking *avrXv3* were first described in Florida strains in 2012 and 2006, respectively. Of 564 isolates analyzed, only two previously described *avrXv3* variants were identified between 2017 and 2020. Previous research demonstrated that some Ohio *X. perforans* race T3 strains

characterized lack the effector *avrBsT* present in T4 strains (Timilsina et al., 2016). The effector *avrBsT* has been shown to provide a fitness advantage in the field compared to a knock-out recombinant (Abrahamian et al., 2018; Sharma et al., 2021). Therefore, the persistence of T4 may be related to fitness associated with a specific effector. We determined that all *X. perforans* T4-SNP strains had *avrBsT* with the exception of SM1852-18, similar to Xp17-12 isolated in Florida in 2006 (Timilsina et al., 2016).

Our results show that the race shift observed in the Midwestern United States is due to the introduction or migration of a strain of *X. perforans* possessing a previously described SNP205nt mutation in *avrXv3*. Given the low rate of *Rx-4*/*Xv3* deployment, it is unlikely that mutation in *avrXv3* was selected in response to the plant-pathogen arms race nor is resistance gene pressure responsible for maintaining this change. We cannot rule out multiple introductions, although it seems unlikely that the same clone would be introduced multiple years from multiple seed sources. The evidence based on amplification of *avrXv3*, whole genome sequencing, and analysis of isolates collected over several years suggests that the population of *X. perforans* T4 is largely descended from and is genetically identical to a single common ancestor that has been introduced and maintained for multiple years.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found online as cited in the **Supplementary Material**. Sequences and annotation are available at NCBI under BioProject accession PRJNA795842.

AUTHOR CONTRIBUTIONS

DF developed the experimental design, sampled tomatoes with bacterial spot lesions throughout the Midwestern United States, and assisted with data analysis. EB established the experimental design, performed the bacterial inoculations, genotyping, phylogenomics, and strain characterization, and analyzed all data. FR and SAM isolated and genotyped the bacterial strains at the species level. FR cultured and prepared bacterial strains for inoculations. VR-R, TK, ST, GM, FI-B, JMJ, JBJ, and EG conducted library preparation and whole genome sequencing of strains isolated over the 4-year period, and provided guidance on tools and resources to conduct data analyses. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the USDA Agriculture and Food Research Initiative (AFRI) Specialty Crop Research Initiative (SCRI) award 2015-51181-24312, and the Ohio Department of Agriculture Specialty Crops Block Grant (AGR-SCG-19-03) to JMJ and SAM, and USDA NIFA FACT (2021-67021-34343).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.826386/full#supplementary-material

Supplementary Figure 1 | Comparison of Multilocus Sequence analysis (MLSA) Tree, Average Nucleotide Identity (ANI) Tree, and Parsnp Tree. MLSA, ANI, and Parsnp trees, containing 39 *X. perforans* strains, were developed to compare and identify known groups and clades. **(A)** Concatenation of five house-keeping genes, *lepA, gyrB, lacF, gapA,* and *gltA,* was used to build the MLSA phylogenetic tree. **(B)** Whole genome sequence was used for the ANI phylogenomic tree. **(C)** Whole Genome sequence was used for Parsnp phylogenomic tree. Strains are denoted by the name, location of isolation, year, race-type, and known mutation in *avrXv3*.

REFERENCES

Abrahamian, P., Timilsina, S., Minsavage, G. V., Goss, E. M., and Jones, J. B. (2018). The Type III Effector AvrBsT Enhances *Xanthomonas perforans* Fitness in Field-Grown Tomato. *Phytopathology* 108, 1355–1362. doi: 10.1094/PHYTO-02-18-0052-R

Abrahamian, P., Timilsina, S., Minsavage, G. V., Potnis, N., Jones, J. B., Goss, E. M., et al. (2019). Molecular Epidemiology of *Xanthomonas perforans* Outbreaks in Tomato Plants from Transplant to Field as Determined by Single-Nucleotide Polymorphism Analysis. *Appl. Env. Microb.* 85, e1220–e1219. doi: 10.1128/AEM.01220-19

Adhikari, P., Adhikari, T. B., Timilsina, S., Meadows, I., Jones, J. B., Panthee, D. R., et al. (2019). Phenotypic and Genetic Diversity of *Xanthomonas perforans* Populations from Tomato in North Carolina. *Phytopathology* 109, 1533–1543. doi: 10.1094/PHYTO-01-19-0019-R

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mole. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Balin, S. J., and Cascalho, M. (2010). The rate of mutation of a single gene. *Nucleic Acids Res.* 38, 1575–1582. doi: 10.1093/nar/gkp1119

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bernal, E. (2020). *Development of tomato (S. lycopersicum) lines with resistance to Xanthomonas spp. and use of genetic resources to characterize infection and diversity in pathogen populations. [Doctoral dissertation] Ohio State University]. OhioLINK*. Columbus: Electronic Theses and Dissertations Center.

Bernal, E., and Francis, D. M. (2021). Processing Tomato Germplasm with Improved Resistance to Bacterial Spot. *HortScience* 56, 519–520. doi: 10.21273/hortsci15616-20

Bernal, E., Liabeuf, D., and Francis, D. M. (2020). Evaluating Quantitative Trait Locus Resistance in Tomato to Multiple *Xanthomonas* spp. *Plant Dis.* 104, 423–429. doi: 10.1094/PDIS-03-19-0669-RE

Berry, S. Z., Aldrich, T. S., Wiese, K. L., and Bash, W. D. (1995). 'Ohio OX38' Hybrid Processing Tomato. *HortScience* 30:159. doi: 10.21273/hortsci.30.1.159

Covarrubias-Pazaran, G. (2016). Genome assisted prediction of quantitative traits using the R package sommer. *PLoS One* 11, 1–15. doi: 10.1371/journal.pone.0156744

Covarrubias-Pazaran, G. (2018). Software update: moving the R package sommer to multivariate mixed models for genome-assisted prediction. *bioRxiv* doi: 10.1101/354639

Diachun, S., and Valleau, W. (1946). Growth and overwintering of *Xanthomonas vesicatoria* in association with wheat roots. *Phytopathology* 36, 277–280.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Egel, D. S., Jones, J. B., Minsavage, G. V., Creswell, T., Ruhl, G., Maynard, E., et al. (2018). Distribution and Characterization of Xanthomonas Strains Causing Bacterial Spot of Tomato in Indiana. *Plant Health Prog.* 19, 319–321. doi: 10.1094/PHP-07-18-0041-BR

Figueras, M. J., Beaz-Hidalgo, R., Hossain, M. J., and Liles, M. R. (2014). Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc.* 2, e927–e914. doi: 10.1128/genomeA.00927-14

Gardner, M. W., and Kendrick, J. B. (1921). *Tomato bacterial spot and seed disinfection*. Carmel: Purdue University Agricultural Experiment Station.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Internat. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Horvath, D. M., Stall, R. E., Jones, J. B., Pauly, M. H., Vallad, G. E., Dahlbeck, D., et al. (2012). Transgenic resistance confers effective field level control of bacterial spot disease in tomato. *PLoS One* 7:8. doi: 10.1371/journal.pone.0042036

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Klein-Gordon, J. M., Timilsina, S., Xing, Y., Abrahamian, P., Garrett, K. A., Jones, J. B., et al. (2021). Whole genome sequences reveal the Xanthomonas perforans population is shaped by the tomato production system. *ISME J.* 2021:8. doi: 10.1038/s41396-021-01104-8

Klein-Gordon, J. M., Xing, Y., Garrett, K. A., Abrahamian, P., Paret, M. L., Minsavage, G. V., et al. (2020). Assessing Changes and Associations in the *Xanthomonas perforans* Population Across Florida Commercial Tomato Fields Via a Statewide Survey. *Phytopathology* 2020:0402. doi: 10.1094/phyto-09-20-0402-r

Leben, C. (1981). How plant-pathogenic bacteria survive. *Plant Dis.* 65, 633–637.

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics* 2013:1303.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Louws, F. J., Fulbright, D. W., Stephens, C. T., and de Bruijn, F. J. (1994). Specific genomic fingerprints of phytopathogenic Xanthomonas and *Pseudomonas* pathovars and strains generated with repetitive sequences and PCR. *Appl. Environ. Microbiol.* 60:2286. doi: 10.1128/aem.60.7.2286-2295.1994

Ma, X. (2015). *Characterization and Management of Bacterial Leaf Spot of Processing Tomato in Ohio. [Doctoral dissertation, Ohio State University]*. Columbus: OhioLINK Electronic Theses and Dissertations Center.

Ma, X., Lewis Ivey, M. L., and Miller, S. A. (2011). First Report of Xanthomonas gardneri Causing Bacterial Spot of Tomato in Ohio and Michigan. *Plant Dis.* 95:1584. doi: 10.1094/PDIS-05-11-0448

Mhedbi-Hajri, N., Hajri, A., Boureau, T., Darrasse, A., Durand, K., Brin, C., et al. (2013). Evolutionary History of the Plant Pathogenic Bacterium Xanthomonas axonopodis. *PLoS One* 8:e58474. doi: 10.1371/journal.pone.0058474

Monteil, C. L., Yahara, K., Studholme, D. J., Mageiros, L., Méric, G., Swingle, B., et al. (2016). Population-genomic insights into emergence, crop adaptation and dissemination of *Pseudomonas* syringae pathogens. *Microb. Gen.* 2, e000089–e000089. doi: 10.1099/mgen.0.000089

Newberry, E. A., Bhandari, R., Minsavage, G. V., Timilsina, S., Jibrin, M. O., Kemble, J., et al. (2019). Independent Evolution with the Gene Flux Originating from Multiple *Xanthomonas Species* Explains Genomic Heterogeneity in *Xanthomonas perforans*. *Appl. Environ. Microbiol.* 85, e885–e819. doi: 10.1128/AEM.00885-19

Pei, C., Wang, H., Zhang, J., Wang, Y., Francis, D. M., and Yang, W. (2012). Fine mapping and analysis of a candidate gene in tomato accession PI128216 conferring hypersensitive resistance to bacterial spot race T3. *Theor. Appl. Genet.* 124, 533–542. doi: 10.1007/s00122-011-1726-1

Perez-Quintero, A. L., Ortiz-Castro, M., Lang, J. M., Rieux, A., Wu, G., Liu, S., et al. (2020). Genomic Acquisitions in Emerging Populations of Xanthomonas vasicola pv. vasculorum Infecting Corn in the United States and Argentina. *Phytopathology* 110, 1161–1173. doi: 10.1094/PHYTO-03-19-0077-R

Peterson, G. H. (1963). Survival of Xanthomionas vesicatoria in soil & diseased tomato plants. *Phytopathology* 53, 765–767.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Robbins, M. D., Darrigues, A., Sim, S. C., Masud, M. A., and Francis, D. M. (2009). Characterization of hypersensitive resistance to bacterial spot race T3 (*Xanthomonas perforans*) from tomato accession PI 128216. *Phytopathology* 99, 1037–1044. doi: 10.1094/PHYTO-99-9-1037

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* 4:e1900v1.

Rotondo, F., Bernal, E., Ma, X. Lewis Ivey, M. L., Sahin, F., Francis, D. M., et al. (2022). Shifts in Xanthomonas spp. causing bacterial spot in processing tomato in the U.S. Midwest. *Can. J. Plant Pathol.* (in press).

Sahin, F. (1997). *Detection, identification and characterization of strains of Xanthomonas campestris pv. vesicatoria by traditional and molecular methods, and resistance in Capsicum species to Xanthomonas campestris pv. vesicatoria pepper race 6*. 9731706 Ph.D. Columbus: The Ohio State University.

Schwartz, A. R., Potnis, N., Timilsina, S., Wilson, M., PatanÈ, J., Martins, J. Jr., et al. (2015). Phylogenomics of Xanthomonas field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front. Microb.* 2015:6. doi: 10.3389/fmicb.2015.00535

Scott, J. W., Jones, J. B., Somodi, G. C., and Stall, R. E. (1995). Screening Tomato Accessions for Resistance to *Xanthomonas campestris pv. vesicatoria*, Race T3. *HortScience* 30:579. doi: 10.21273/hortsci.30.3.579

Scott, J., Stall, R., Jones, J., and Somodi, G. (1996). A single gene controls the hypersensitive response of Hawaii 7981 to race 3 (T3) of the bacterial spot pathogen. *Tomato Genet. Coop.* 46:23.

Sharlach, M., Dahlbeck, D., Liu, L., Chiu, J., Jiménez-Gómez, J. M., Kimura, S., et al. (2013). Fine genetic mapping of *RXopJ4*, a bacterial spot disease resistance locus from Solanum pennellii LA716. *Theor. Appl. Genet.* 126, 601–609. doi: 10.1007/s00122-012-2004-6

Sharma, A., Timilsina, S., Abrahamian, P., Minsavage, G. V., Colee, J., Ojiambo, P. S., et al. (2021). Need for speed: bacterial effector XopJ2 is associated with increased dispersal velocity of Xanthomonas perforans. *Env. Microb.* 2021:15541. doi: 10.1111/1462-2920.15541

Strayer, A. L., Jeyaprakash, A., Minsavage, G. V., Timilsina, S., Vallad, G. E., Jones, J. B., et al. (2016). A Multiplex Real-Time PCR Assay Differentiates Four Xanthomonas Species Associated with Bacterial Spot of Tomato. *Plant Dis.* 100:8. doi: 10.1094/PDIS-09-15-1085-RE

Tibayrenc, M., and Ayala, F. J. (2012). Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. *Proc. Natl. Acad. Sci.* 109:E3305. doi: 10.1073/pnas.1212452109

Timilsina, S., Abrahamian, P., Potnis, N., Minsavage, G. V., White, F. F., Staskawicz, B. J., et al. (2016). Analysis of Sequenced Genomes of *Xanthomonas perforans* Identifies Candidate Targets for Resistance Breeding in Tomato. *Phytopathology* 106, 1097–1104. doi: 10.1094/PHYTO-03-16-0119-FI

Timilsina, S., Jibrin, M. O., Potnis, N., Minsavage, G. V., Kebede, M., Schwartz, A., et al. (2015). Multilocus sequence analysis of xanthomonads causing bacterial spot of tomato and pepper plants reveals strains generated by recombination among species and recent global spread of Xanthomonas gardneri. *Appl. Env. Microb.* 81, 1520–1529. doi: 10.1128/AEM.03000-14

Timilsina, S., Pereira-Martin, J. A., Minsavage, G. V., Iruegas-Bocardo, F., Abrahamian, P., Potnis, N., et al. (2019). Multiple Recombination Events Drive

the Current Genetic Structure of *Xanthomonas perforans* in Florida. *Front. Microb.* 2019:10. doi: 10.3389/fmicb.2019.00448

Treangen, T. J., Ondov, B. D., Koren, S., and Phillippy, A. M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15:524. doi: 10.1186/s13059-014-0524-x

Vinatzer, B. A., Monteil, C. L., and Clarke, C. R. (2014). Harnessing Population Genomics to Understand How Bacterial Pathogens Emerge, Adapt to Crop Hosts, and Disseminate. *Ann. Rev. Phytopathol.* 52, 19–43. doi: 10.1146/annurev-phyto-102313-045907

Wang, H., Hutton, S. F., Robbins, M. D., Sim, S. C., Scott, J. W., Yang, W., et al. (2011). Molecular mapping of hypersensitive resistance from tomato 'Hawaii 7981' to Xanthomonas perforans race T3. *Phytopathology* 101, 1217–1223. doi: 10.1094/PHYTO-12-10-0345

Zhang, X., Li, N., Liu, X., Wang, J., Zhang, Y., Liu, D., et al. (2021). Tomato protein Rx4 mediates the hypersensitive response to Xanthomonas euvesicatoria pv. perforans race T3. *Plant J.* 105, 1630–1644. doi: 10.1111/tpj.15138

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fmicb.2022.840308

Phylogenetic Distribution and Evolution of Type VI Secretion System in the Genus *Xanthomonas*

Prabha Liyanapathiranage[1†], Naama Wagner[2†], Oren Avram[2], Tal Pupko[2] and Neha Potnis[1]*

[1] Department of Entomology and Plant Pathology, Auburn University, Auburn, AL, United States, [2] The Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

The type VI secretion system (T6SS) present in many Gram-negative bacteria is a contact-dependent apparatus that can directly deliver secreted effectors or toxins into diverse neighboring cellular targets including both prokaryotic and eukaryotic organisms. Recent reverse genetics studies with T6 core gene loci have indicated the importance of functional T6SS toward overall competitive fitness in various pathogenic *Xanthomonas* spp. To understand the contribution of T6SS toward ecology and evolution of *Xanthomonas* spp., we explored the distribution of the three distinguishable T6SS clusters, i3*, i3***, and i4, in approximately 1,740 *Xanthomonas* genomes, along with their conservation, genetic organization, and their evolutionary patterns in this genus. Screening genomes for core genes of each T6 cluster indicated that 40% of the sequenced strains possess two T6 clusters, with combinations of i3*** and i3* or i3*** and i4. A few strains of *Xanthomonas citri*, *Xanthomonas phaseoli*, and *Xanthomonas cissicola* were the exception, possessing a unique combination of i3* and i4. The findings also indicated clade-specific distribution of T6SS clusters. Phylogenetic analysis demonstrated that T6SS clusters i3* and i3*** were probably acquired by the ancestor of the genus *Xanthomonas*, followed by gain or loss of individual clusters upon diversification into subsequent clades. T6 i4 cluster has been acquired in recent independent events by group 2 xanthomonads followed by its spread via horizontal dissemination across distinct clades across groups 1 and 2 xanthomonads. We also noted reshuffling of the entire core T6 loci, as well as T6SS spike complex components, *hcp* and *vgrG*, among different species. Our findings indicate that gain or loss events of specific T6SS clusters across *Xanthomonas* phylogeny have not been random.

Keywords: T6SS, *Xanthomonas*, evolution, non-pathogenic, phylogenetics

OPEN ACCESS

Edited by:
Chih-Horng Kuo, Institute of Plant and Microbial Biology, Academia Sinica, Taiwan

Reviewed by:
Alexandra J. Weisberg, Oregon State University, United States
Yong-Qiang He, Guangxi University, China

***Correspondence:**
Neha Potnis
nzp0024@auburn.edu

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *21 December 2021*
Accepted: *10 February 2022*
Published: *14 April 2022*

Citation:
Liyanapathiranage P, Wagner N, Avram O, Pupko T and Potnis N (2022) Phylogenetic Distribution and Evolution of Type VI Secretion System in the Genus Xanthomonas. Front. Microbiol. 13:840308. doi: 10.3389/fmicb.2022.840308

INTRODUCTION

The genus *Xanthomonas* belongs to the gamma subdivision of the phylum Proteobacteria (Jun et al., 2010). It consists of more than 35 species, some of which cause economically important diseases in more than 400 host plants (Timilsina et al., 2020). Even though a wide host range is observed with *Xanthomonas* spp., individual strains can be tightly restricted to a particular host or host group. Hence, the *Xanthomonas* spp. are further differentiated into pathovars based on this

characteristic (White et al., 2009; An et al., 2020). To date, *Xanthomonas* spp. have been isolated from host plants belonging to 124 monocotyledonous and 268 dicotyledonous species (White et al., 2009; Kyrova et al., 2020). In terms of host tissue specificity, *Xanthomonas* spp. can either invade the vascular tissues or survive in the intercellular spaces of the mesophyll parenchyma tissue (An et al., 2020). To successfully adapt to the hostile host environment, *Xanthomonas* spp. usually depend on several factors that are important for their invasion. These factors are responsible for various functions, including adherence to the plant surface, invasion to the intercellular space of the host tissue, manipulation of plant cellular processes, acquisition of nutrients, counteracting plant defense responses, and competing with other microbes that occupy the same niche (Büttner and Bonas, 2010). Identification of the adaptation factors that contribute to host–pathogen interactions is crucial in implementing future management strategies.

To date, more than 1,700 *Xanthomonas* genomes are publicly available in the National Center for Biotechnology Information (NCBI) database. With the availability of such a large amount of genomic data, comparative genomic studies enabled the identification of many potential virulence and fitness strategies contributing to the adaptation to a variety of monocot/dicot hosts and the establishment of successful vascular or non-vascular modes of infection (Jacques et al., 2016; An et al., 2020). These strategies include various secretion systems, by which pathogenic bacteria translocate proteins/DNA directly into the host cell or into the extracellular milieu.

In Gram-negative bacteria, 10 different secretion systems have been identified, but only six of them (I–VI) have been described in the genus *Xanthomonas* (Meuskens et al., 2019; Alvarez-Martinez et al., 2020; Palmer et al., 2020). Of these six systems, types II, III, and IV have been extensively studied in the genus *Xanthomonas*. In contrast, the role, distribution, and evolutionary history of the most recently identified type VI secretion system (T6SS) has been described only in few studies (Bayer-Santos et al., 2018; Bansal et al., 2019, 2020; Bayer-Santos et al., 2019; Ceseti et al., 2019; Alvarez-Martinez et al., 2020; Choi et al., 2020; Zhu et al., 2020; Montenegro Benavides et al., 2021). Most of the aforementioned studies used only representative *Xanthomonas* spp., to derive conclusions regarding the distribution of the T6SS clusters in the whole genus *Xanthomonas*. Yet, the evolutionary history of multiple T6SS clusters, their distribution across the genus, and their association with particular pathogen lifestyle all remain unclear.

T6SS is present in approximately 25% of Gram-negative bacteria, mainly within the phylum Proteobacteria and the classes α-, β-, and γ-proteobacteria, including pathogenic, beneficial, and commensal bacteria, indicating possible functions of this system unrelated to pathogenicity (Boyer et al., 2009; Durand et al., 2014). T6SS is composed of 15 to 23 different proteins, including 13 core structural proteins that assemble into an injectisome-like structure (Shyntum et al., 2014). Core components are named as Tss (for type VI secretion) A-M and are usually encoded in the same gene cluster (Zoued et al., 2014). In addition to the core genes, most T6 clusters also include Tag (type VI–associated genes), which encode accessory proteins, suggested to be important for the transcriptional or posttranscriptional regulation and assembly of the T6SS. Finally, T6SS includes T6 effectors and their associated immunity proteins (Shalom et al., 2007; Boyer et al., 2009; Silverman et al., 2011; Coulthurst, 2013). This Tss nomenclature is now widely adopted by many, while common names, such as Hcp, VgrG, and ClpV, are also in use.

T6SS is comprised of three major complexes: the transmembrane complex, the baseplate, and the tail. Together, they form a larger machine that extends from the bacterial cytoplasm, across the bacterial cell wall and into the target cell (Silverman et al., 2012; Coulthurst, 2013; Durand et al., 2015; Nguyen et al., 2018; and Hernandez et al., 2020). The transmembrane complex made from TssM, TssJ, and TssL acts as a docking site for the cytoplasmic baseplate-like structure made from TssE, TssF, TssG, and TssK. It also serves as an adaptor between the transmembrane complex and the tail. The internal tube of the tail component is made of Hcp tubes that are assembled into the central VgrG–PAAR spike and is enclosed by a contractile sheath composed of TssB and TssC. Secretion of effector proteins through the T6SS is a one-step process, independent of the Sec or Tat secretion machinery, which is mediated through contraction of the sheath to propel the Hcp–VgrG–PAAR complex toward the target cell and deliver a cocktail of effectors in each contraction. Unlike other secretion systems, after each firing event, the T6 tube disassembles, and Hcp is lost to the extracellular milieu. AAA+ClpV (TssH) ATPase is recruited to the contracted sheath to disassemble the complex and to recycle the sheath components for the next round of firing. Apart from being core components of the T6SS, Hcp (hemolysin coregulated protein) and VgrG (valine-glycine repeat protein G) have also been identified as specialized effectors important in antibacterial activity in several pathogen systems (Lien and Lai, 2017; Ma et al., 2017). Effectors that get delivered via the T6SS are categorized as either "cargo" or "specialized." Cargo effectors interact with the Hcp–VgrG–PAAR complex non-covalently, whereas specialized effectors are covalently linked to these proteins. T6 effectors and their antitoxins or immunity proteins may be encoded either within the T6 gene cluster, which encodes the core genes, or scattered in other parts of the bacterial genome (Hernandez et al., 2020).

Depending on the 13 core proteins, T6 clusters have been separated into four subtypes (T6SS^{i-iv}). The majority of the proteobacteria have T6 clusters belonging to the subtype T6SS^{i}, and because of the diversity observed in this subgroup, it has been further subdivided into five phylogenetic clades (i1–i5) (Boyer et al., 2009). Bacteria belonging to the family Xanthomonadales harbor three such clades: i1, i3, and i4, whereas bacteria within genus *Xanthomonas* harbor only the i3 and i4 clades (Bayer-Santos et al., 2019; Alvarez-Martinez et al., 2020). Clade i3 was further subdivided into distinct subclades: i3*, i3**, and i3*** based on the phylogenetic analysis conducted using tssC gene (Bayer-Santos et al., 2019). Species within the genus *Xanthomonas* harbor subclades i3* (e.g., *Xanthomonas citri*), i3*** (e.g., *Xanthomonas oryzae*), or both i3* and i3*** (e.g., *Xanthomonas perforans*), but not i3**.

The phylogenetic-based division to clades and subclades is also reflected in the set of genes included within the T6SS gene cluster

and its genome organization. For example, all *Xanthomonas* species within the i3 clade lack *tssJ*, a core component in the outer membrane complex of T6 machinery (Alvarez-Martinez et al., 2020). In addition, i3* and i3*** differ in their genetic architecture, for example, i3* cluster in *Xanthomonas euvesicatoria* 85-10 is approximately 40 kb, whereas the i3*** cluster is approximately 20 kb. This difference in cluster size is due to the presence of different genes encoding hypothetical proteins and transcriptional regulators in each cluster (Boyer et al., 2009; Abendroth et al., 2017; Alvarez-Martinez et al., 2020).

Similar to observations with other bacteria, some *Xanthomonas* spp. encode multiple T6 copies: either both i3* and i3*** or one copy from each subclade together with the i4 cluster (Bingle et al., 2008; Alvarez-Martinez et al., 2020). These multiple copies of T6 gene clusters are believed to be acquired by independent horizontal gene transfer events, suggesting different roles played by the T6 clusters in different bacteria (Bingle et al., 2008; Blondel et al., 2009; Boyer et al., 2009). In *X. oryzae* pv. oryzae, one T6 cluster is distantly related to the *phaseoli* clade, whereas the second cluster forms a monophyletic clade with i3*** cluster of *X. euvesicatoria* and *X. perforans*, implicating a recent common origin (Montenegro Benavides et al., 2021).

The goal of this study is to harness a large set of publicly available *Xanthomonas* genomes spanning the entire genus *Xanthomonas* with the genomes of multiple important pathogenic and non-pathogenic or commensal species and study the conservation, genetic organization, and evolutionary patterns of multiple T6SS clusters. We hypothesize that the acquisition of the multiple T6SS clusters occurred as independent events during the adaptation of *Xanthomonas* species to specific hosts, as they diverged into various clades. Specific cases of loss or gain events happened in certain clades as an adaptation to vascular or non-vascular lifestyles. We examined the phylogenetic distribution of T6SS clusters and systematically assessed the acquisition and loss events among different species along with the reshuffling of entire loci or individual genes, to obtain insights into the significance of T6SS in host adaptation of xanthomonads.

MATERIALS AND METHODS

Xanthomonas Genomes

All *Xanthomonas* RefSeq genomes available in NCBI on January 16, 2022, were downloaded. This included a total of 1,740 genomes of 39 *Xanthomonas* species (see full list of species and accession numbers in **Supplementary Table 1**). For each of these genomes, protein translation of open reading frames was extracted using Prodigal (Hyatt et al., 2010).

Initial Dataset of T6 Genes

To find the core genes of T6-i3*, T6-i3***, and T6-i4 clusters in each of the *Xanthomonas* genomes, we first established datasets of core genes of each of the three clusters. The core genes of T6-i3* and T6-i3*** were taken from *X. euvesicatoria* str. 85-10 (accession NC_007508.1), whereas the core genes of T6-i4 were taken from *X. oryzae* pv. oryzae PXO99A (accession NC_010717.2). The core gene loci of all three clusters included TssA, TssB, TssF, TssG, TssK, TssL (T6-i4 included both TssL1 and TssL2), and TssM. T6-i3* and T6-i3*** also included TssC and TssD. T6-i3*** and T6-i4 also included TssE. T6-i3* also included TssH, T6-i3*** also included ImpE-like protein TagF, and T6-i4 also included TssJ and ClpV.

Identification of T6SS Clusters Across All 1,740 *Xanthomonas* Genomes

The core genes of T6-i3*, T6-i3***, and T6-i4 clusters were searched for in each of the *Xanthomonas* genomes using BLASTx (Altschul et al., 1990). We used the core gene sequences described above as query and the protein sequences of each of the genomes as databases. The commands used were **makeblastdb -in {dataset} -dbtype prot -out {dataset_db}** to create the BLAST database, and **blastx -db {dataset_db} -query {query} -outfmt 6 -out {out}** to run BLASTx. "dataset" was the protein FASTA file of a certain *Xanthomonas* genome, and "query" was a DNA FASTA file containing the core genes of a certain T6 system. It was repeated for each genome and system combination. For each of the core genes in the query, the best BLAST hit in the genome was considered, if it had an E-value lower than 10^{-10} and identity of at least 50% to the core gene, and a coverage of at least 30% (i.e. that the length of the aligned region is at least 30% of the query gene length). In case a certain gene in the genome was found to be a hit matching our criteria for genes of more than one T6 system, for example, TssA of T6-i3* and TssA of T6-i3***, the hit was saved only for the gene of the system it had the highest identity to. ClpV shares domains with ClpB, so to avoid classifying ClpB genes as ClpV; we blasted ClpB versus the genomes, and protein sequences that showed a higher sequence similarity to ClpB than to ClpV were discarded. Then, for each genome, the presence of each T6 system was evaluated. If all the core genes of a specific system were found in the genome, we considered this system to be full in this genome. Otherwise, if it had at least one core gene, we considered it partial. Gene Neighborhood tool was used to identify gene synteny and cluster organization (Chen et al., 2021).

Construction of T6 Core Gene Phylogenies

Following the identification of core genes, as described previously, the corresponding protein sequences of each gene were aligned using MAFFT (v.7.471) with the L-ins-I option (Katoh et al., 2002). Phylogenetic trees under the maximum likelihood (ML) criterion were reconstructed using the IQ-TREE multicore version 2.1.2 COVID edition (Minh et al., 2020). The "–m MFP" function was used for the model selection, whereas "–merge rclusterf" was used for partition finding. Branch support was determined using 1,000 bootstraps and 1,000 SH-aLRT bootstrap replicates (-B 1000, -alrt 1000). Nodes with $\geq$80% support with the SH-aLRT and $>$80% with normal bootstrap supports were considered well supported (Davis et al., 2021). Trees were visualized using FigTree v1.4.4 (Rambaut et al., 2018).

SplitsTree Analysis for Homologous Recombination Visualization

To identify and visualize possible conflicting signals, which would suggest possible recombination events and evolutionary relationships within sequence data, phylogenetic networks were inferred using SplitsTree4 (version 4.17.0) (Huson and Bryant, 2006). Concatenated T6 core genes from each cluster were used to generate SplitsTree files, for a selected subset of genomes using the automated multilocus sequence analysis pipeline (automlsa2) (Davis et al., 2021). Resultant splits.nex files were visualized using the SplitsTree4 (version 4.17.0), with the rooted equal angle algorithm (Gambette and Huson, 2008). Recombination events were identified by the branches that form parallelograms (Joseph and Forsythe, 2012).

Multilocus Sequence Analysis

To reconstruct the species phylogeny, we conducted a multilocus sequence analysis (MLSA) using 12 housekeeping genes: *atpD* (ATP synthase β chain), *dnaK* (70-kDa heat shock protein), *efp* (elongation factor P), *fusA* (translation elongation factor aEF-2), *fyuA* (transmembrane protein; Ton-B–dependent transporter), *gapA* (glyceraldehyde 3-phosphate dehydrogenase), *glnA* (glutamine synthetase I), *gltA* (citrate (Si)-synthase gene), *gyrB* (DNA gyrase β subunit), *lacF* (sugar ABC transporter permease), *lepA* (GTP-binding protein), and *rpoD* (RNA polymerase sigma 70 factor). These genes were previously used for MLSA (Young et al., 2008; Fischer-Le Saux et al., 2015; Timilsina et al., 2015). We created a multiple sequence alignment (MSA) for each of these 12 genes using MAFFT (Katoh et al., 2002; Katoh and Standley, 2013) and then concatenated these MSAs. The concatenated MSA was used as input for RaxML (Stamatakis, 2014) to create the phylogenetic tree.

Inferring Evolutionary Gain/Loss Events of T6SS Clusters Across *Xanthomonas* Genus Phylogeny

We used GLOOME to analyze the gain and loss probabilities for three T6 clusters during evolution of the entire *Xanthomonas* genus (Cohen et al., 2010). GLOOME analyzes phyletic patterns based on presence and absence of individual core genes of the cluster and accurately infers branch-specific and site-specific gain and loss events. We generated the input file of phyletic pattern based on the BLASTx results described above, in which "1" indicates presence and "0" indicates absence. The MLSA tree was used as input tree. Cladogram was drawn with FigTree v1.4.4 (Rambaut et al., 2018). Gain and loss events were added manually based on the GLOOME output files.

Construction of Core Genome Phylogeny

For the reconstruction of the core genome phylogeny Roary was used (Page et al., 2015). All the genomes downloaded from the NCBI GenBank in the fna format were first annotated using the Prokka (v1.14.5) software tool (Seemann, 2014). After annotating the genome to determine the location and attributes of the genes contained in them, the resultant gff file in the GFF3 format was used as input file to perform the pan-genome analysis using Roary. Roary extracted the coding sequences from the input GFF3 files, converted them into protein sequences, removed partial clusters, and created preclusters. Next Roary performed BLASTP with 95% sequence identity, and sequences were clustered with Markov cluster algorithm. As the next step, Roary generated clusters and inferred orthologs. Finally, Roary identified core genes and generated a concatenated core gene alignment that was used to generate maximum likelihood phylogeny using RAxML (v 8.2.12) (Stamatakis, 2014).

RESULTS

Distribution of T6SS Clusters in the Genus *Xanthomonas*

In this study, we analyzed 1,740 complete genome sequences, belonging to 39 *Xanthomonas* spp. retrieved from the NCBI RefSeq with a large proportion of them from species, *X. oryzae* (396 genomes), *X. citri* (195 genomes), *X. phaseoli* (97 genomes), *X. perforans* (151 genomes), and *X. arboricola* (136 genomes). This collection of genomes also included commensal *Xanthomonas* spp., isolated from environment (rain) as well as isolated from diverse plant species (Potnis et al., unpublished). Based on core gene sequence identity of T6SS clusters, presence of complete T6 i3*** was identified in 832 genomes; T6 i3* in 710 genomes and 497 genomes had the complete T6 i4 cluster (**Table 1**). Presence of only a single T6 cluster, T6 i3* was identified in 23% of *Xanthomonas genomes*, the majority belonging to *X. citri*, *X. phaseoli*, *X. vesicatoria*, *X. axonopodis*, and few strains belonging to *X. arboricola*, *X. campestris*, *X. pisi*, *X. dyei*, and *X. cucurbitae*, whereas that of T6SS i3*** alone in *X. vasicola* and that of T6 i4 alone in *X. oryzae*, *X. translucens*, *X. fragarie*, couple of *X. hortorum* strains and *X. bromi*. Some *X. oryzae* strains contain both T6 i3*** and T6 i4. The T6SS makeup in *X. citri* is highly variable, with majority of strains containing T6 i3* alone, 14 strains with T6 i3* and T6 i4, whereas a couple of them lack T6 clusters. *X. phaseoli* also shows variability in genetic makeup of T6 clusters, with 40 strains with T6 i3* alone, 52 strains with T6 i3* and T6 i3***, and 2 strains containing T6 i3* and T6 i4. Eighty-nine percent of the *X. oryzae* strains possess both T6 i3*** and T6 i4, and 2.77% contain T6 i4 alone. *X. euvesicatoria* and *X. perforans* contain both T6 i3* and T6 i3*** clusters. *Xanthomonas albilineans*, *Xanthomonas sacchari*, *X. cannabis*, and majority of *X. arboricola* strains lack T6 cluster. *Xanthomonas maliensis* strains contain all three T6 clusters. Approximately 40% of the currently classified *X. campestris* strains contain either i3* or i3*** complete clusters, with 13% of them containing both i3* and i3*** clusters (**Table 1**).

Among early-branching strains, also known as group 1 xanthomonads, *X. translucens* and few strains belonging to unclassified species contain T6 clusters. Of 65 sequenced *X. translucens* strains, 44 strains contain intact T6 i3*** cluster, and 42 strains contain intact i4 cluster, with 39 strains containing both i3*** and i4 clusters (**Table 1**). While *Xanthomonas hyancinthi* genomes available in RefSeq do not contain any intact T6 cluster, *X. hyacinthi* DSM 19077 contains intact T6 i3***

TABLE 1 | Analysis of T6 cluster distribution in genus *Xanthomonas*.

Xanthomonas spp.	Total no. of genomes	i3***	i3*	i4	i3*** only	i3 only	i4 only	i3*** and i3*	i3*** and i4	i3* and i4	i3*** and i3* and i4
Xanthomonas campestris	125	23 (6)	44 (71)	0 (0)	6	27	0	17	0	0	0
Xanthomonas cassavae	2	0 (2)	2 (0)	0 (0)	0	2	0	0	0	0	0
Xanthomonas cannabis	4	0 (0)	0 (0)	0 (0)	0	0	0	0	0	0	0
Xanthomonas vasicola	100	98 (2)	0 (100)	0 (0)	98	0	0	0	0	0	0
Xanthomonas citri	195	0 (49)	194 (0)	14 (0)	0	180	0	0	0	14	0
Xanthomonas oryzae	396	359 (37)	0 (396)	364 (32)	6	0	11	0	353	0	0
Xanthomonas phaseoli	97	52 (3)	94 (3)	2 (0)	0	40	0	52	0	2	0
Xanthomonas floridensis	1	0 (0)	1 (0)	0 (0)	0	1	0	0	0	0	0
Xanthomonas nasturtii	2	0 (0)	0 (2)	0 (0)	0	0	0	0	0	0	0
Xanthomonas vesicatoria	30	0 (0)	30 (0)	0 (0)	0	30	0	0	0	0	0
Xanthomonas cissicola	2	0 (0)	2 (0)	2 (0)	0	0	0	0	0	2	0
Xanthomonas hortorum	74	0 (0)	0 (0)	5 (0)	0	0	5	0	0	0	0
Xanthomonas prunicola	3	0 (0)	0 (3)	0 (0)	0	0	0	0	0	0	0
Xanthomonas codiaei	2	1 (1)	2 (0)	0 (0)	0	1	0	1	0	0	0
Xanthomonas dyei	4	1 (2)	2 (2)	0 (0)	0	1	0	1	0	0	0
Xanthomonas melonis	6	0 (4)	6 (0)	0 (0)	0	6	0	0	0	0	0
Xanthomonas pisi	2	0 (2)	0 (2)	0 (0)	0	0	0	0	0	0	0
Xanthomonas populi	1	0 (0)	0 (0)	0 (1)	0	0	0	0	0	0	0
Xanthomonas sacchari	5	0 (0)	0 (0)	0 (0)	0	0	0	0	0	0	0
Xanthomonas maliensis	2	2 (0)	2 (0)	2 (0)	0	0	0	0	0	0	2
Xanthomonas sontii	5	0 (0)	0 (0)	0 (0)	0	0	0	0	0	0	0
Xanthomonas hyacinthi	2	0 (0)	0 (0)	0 (0)	0	0	0	0	0	0	0
Xanthomonas cucurbitae	2	0 (2)	0 (2)	0 (0)	0	0	0	0	0	0	0
Xanthomonas albilineans	17	0 (0)	0 (0)	0 (0)	0	0	0	0	0	0	0
Xanthomonas axonopodis	83	7 (4)	72 (11)	0 (1)	6	71	0	1	0	0	0
Xanthomonas theicola	2	0 (0)	0 (0)	0 (0)	0	0	0	0	0	0	0
Xanthomonas translucens	65	44 (4)	0 (48)	42 (3)	5	0	3	0	39	0	0
Xanthomonas perforans	151	148 (3)	142 (9)	0 (0)	9	3	0	139	0	0	0
Xanthomonas bromi	2	0 (0)	0 (0)	1 (1)	0	0	1	0	0	0	0
Xanthomonas fragariae	63	0 (0)	0 (0)	62 (0)	0	0	62	0	0	0	0
Xanthomonas euroxanthea	6	0 (0)	0 (0)	0 (0)	0	0	0	0	0	0	0
Xanthomonas hydrangea	3	0 (0)	0 (0)	0 (0)	0	0	0	0	0	0	0
Xanthomonas arboricola	136	1 (0)	7 (0)	0 (0)	0	6	0	1	0	0	0
Xanthomonas euvesicatoria	98	95 (1)	93 (5)	0 (0)	4	2	0	91	0	0	0
Xanthomonas sp.	47	1 (1)	17 (0)	1 (0)	0	15	0	1	0	1	0
Xanthomonas alfalfae	1	0 (1)	0 (1)	0 (0)	0	0	0	0	0	0	0
Xanthomonas massiliensis	2	0 (0)	0 (0)	2 (0)	0	0	2	0	0	0	0
Xanthomonas retroflexus	1	0 (0)	0 (0)	0 (0)	0	0	0	0	0	0	0

Numbers of strains with partial T6 clusters are shown inside the parenthesis.

cluster, with core genes of this cluster showing 100% nucleotide sequence identity to those from *Luteibacter rhizovicinus* DSM 16549 strain. This strain did not get included in **Table 1** as it lacks RefSeq record. *L. rhizovicinus* is a gamma proteobacterium in the order *Xanthomonadales* isolated from the rhizosphere of barley (*Hordeum vulgare*) (Johansen et al., 2005). This observation indicates the possible independent acquisitions of the T6 clusters by individual strains within the early-branching clade from species belonging to closely related genera within *Xanthomonadaceae* family.

The size of each T6 cluster was also compared in this study. Most of the species with complete i3* cluster were 49–60 kbp whereas the i3* cluster in *X. cucurbitae* CFBP 2542 was approximately 70 kbp (**Supplementary Figure 1**). Shorter i3*** clusters with 23–28 kbps in size were identified in the analyzed genomes, whereas determining the size of the i4 cluster was not possible because of the presence of the i4 cluster in two loci.

Evolutionary Patterns Among Core Genes of T6 Clusters in *Xanthomonas* spp.

Phylogeny based on concatenated sequences of 12 housekeeping genes (MLSA) was compared with that of T6 i3***, i3*, and i4 core gene phylogenies including conserved effector genes, *vgrG* and *hcp* (**Figure 1**). Incongruencies among MLSA phylogeny

and that of T6 i3* were noted for the phylogenetic placements of clades belonging to species *X. perforans* and *X. euvesicatoria*. *X. euvesicatoria* is nestled between two *X. perforans* clades, whereas *X. euvesicatoria*–related pathovars (those pathogenic on other hosts such as alfalfa, and citrus) form a separate clade based on the MLSA phylogeny. T6 i3* phylogeny with *hcp* and *vgrG* sequences reshuffled some *X. perforans* strains among the two clades. Incongruencies between MLSA phylogeny and T6 i4 cluster phylogeny with *hcp* and *vgrG* sequences were observed in case of *X. oryzae*, *X. translucens*, and *X. fragariae* species. Each of these species forms a distinct clade in the MLSA phylogeny. However, based on T6 i4 phylogeny, *X. oryzae* was highly diverse, with strains dispersed throughout the phylogeny. Some *X. translucens* and *X. fragarie* strains carry T6 i4 loci that are closely related to *X. oryzae*. *X. translucens* strain ART-Xtg2 clustered with *X. bromi* based on T6 i4 phylogeny (**Figures 1A,B**). No apparent incongruencies were observed when i3*** phylogeny was compared with that of the MLSA phylogeny.

Hcp and VgrG are the components of T6SS spike complex and play a role in secretion of effectors by covalently linking to the effectors. To investigate if *hcp* and *vgrG* contribute to the incongruencies observed among the above phylogeny comparisons, these two gene sequences were removed from the concatenated T6 core gene query file, and the phylogenies of T6SS core genes without *hcp* and *vgrG* sequences were reconstructed for the T6 clusters (**Figure 1C**). Phylogeny of core T6SS i3*** genes without *hcp* and *vgrG* showed *X. euvesicatoria*–related pathovars that form a separate clade in MLSA phylogeny, now clustering together with pepper pathogenic *X. euvesicatoria* strains. Likewise, some *X. perforans* strains that formed a distinct clade nested between *X. euvesicatoria* and *X. euvesicatoria*–related pathovars in MLSA phylogeny from the other *X. perforans* clade were now clustering together in phylogeny based on T6SS i3*** genes without *hcp* and *vgrG*. A distantly related species, *X. maliensis*, was observed closely related to *X. euvesicatoria* species complex based on T6 i3*** phylogeny with or without *hcp* and *vgrG* genes. Interestingly, comparison of i3* phylogeny without *hcp* and *vgrG* to that with *hcp* and *vgrG* showed change in the placement of some *X. perforans* strains in separate clades. The incongruencies observed with placement of *X. oryzae*, and *X. fragarie* observed in T6 i4 phylogeny with *vgrG* genes in comparison to MLSA phylogeny were resolved in the phylogeny constructed without *vgrG* genes (**Figure 1**). *X. oryzae* was observed as highly diverse species in phylogeny with *vgrG* and formed one distinct clade in T6 i4 phylogeny without *vgrG*, indicating a high degree of variation in *vgrG* genes in this species in addition to a gene flow with other species such as *X. translucens* and *X. fragarie*. Multiple copies of *vgrG* are presently dispersed throughout the genome of *X. oryzae* (**Supplementary Figure 2**). Group 1 species, *X. translucens*, was found to cluster with group 2 species, *X. oryzae* based on core gene T6 i4 phylogeny without *vgrG* genes, indicating a shared evolutionary path for T6 i4 cluster in these two species. A strain of *X. translucens*, ART-Xtg2, was an exception, which is more closely related to *X. bromi*, a group 2 species, based on T6 i4 core gene phylogeny with or without *vgrG* genes, suggesting shared evolutionary patterns for T6 i4 cluster (**Figure 1**).

Given our observation of incongruencies among MLSA phylogeny and T6SS core gene phylogeny without *hcp, vgrG* in case of clusters i3* and i3***, we hypothesized that there is genetic exchange of T6 core genes among different strains belonging to different species. We constructed phylogenetic networks using SplitsTree focusing on representatives from individual species. SplitsTree can demonstrate conflicting signals for phylogenetic placements as reticulation. Upon examining the split decomposition tree generated for T6 i3* without *hcp* and *vgrG*, we observed reticulation at the base of the branches connecting unusual recombinant strains thought to be hybrid between *X. perforans* and *X. euvesicatoria* (NI1, NI38), strains belonging to *X. euvesicatoria* sister species (CFBP 6369, F1, LMG12749, LMG 495, CFBP 3836, CFBP3371, DAR26930), commensal *Xanthomonas* strains (CFBP7921, CFBP7922), some strains belonging to *X. phaseoli* (A1962, LMG12749), *X. axonopodis* (LMG26789), and *X. euvesicatoria* 85-10 and *X. perforans* 91-118 (**Figure 2A**). Another evident reticulated event was identified to involve *X. citri* (LMG859, *Bagalkot,* LMG7439, 12609, NCPPB3660, NCPPB1495, NCPPB1056), *X. axonopodis* (LMG753, LMG9049, LMG9050, LMG558), and *X. campestris* strains (LMG954, LMG940, LMG939, LMG9044), indicating flow of genetic information between them with respect to T6 i3* core genes. In a split decomposition tree inferred for T6 i3*** without *hcp* and *vgrG*, a highly reticulated network encompassing above identified *X. perforans*, *X. euvesicatoria*, *X. euvesicatoria*–related pathovars, and some strains of *X. phaseoli* was observed (**Figure 2B**).

Estimating the Acquisition or Loss of T6SS Clusters in the Genus *Xanthomonas* Phylogeny

Gene gain and loss evolutionary dynamics among bacterial species can be inferred by contrasting the reference species phylogeny to that of individual genes (Kettler et al., 2007; Cohen and Pupko, 2011; Iranzo et al., 2019). We next analyzed the evolutionary origin and distribution of multiple T6SS clusters using such an approach, as implemented in GLOOME (Cohen et al., 2010). For the reference species tree, we used a phylogenetic tree generated based on a concatenated MSA of 12 housekeeping gene sequences: *atpD*, *dnaK*, *efP*, *fusA*, *fyuA*, *gapA*, *glnA*, *gltA*, *gyrB*, *lacF*, *lepA*, and *rpoD*. A profile of presence and absence of core genes of individual T6 clusters was used as input, and the output is a probabilistic mapping of events over the reference tree (**Figure 3**). We additionally added clade information based on core genome phylogeny presented by Merda et al. (2017) (**Supplementary Figure 4**). Our results clearly present a pattern of multiple gain and loss events for all clusters (**Figure 3** and **Supplementary Figure 4**).

The i3* cluster is present in the majority of group 2 xanthomonads, and in few *X. translucens* strains belonging to group 1. GLOOME predicted a strong signal for multiple loss events of the i3* cluster in both groups 1 and 2 *Xanthomonas* spp. For example, in group 1, the bacteria in which this cluster

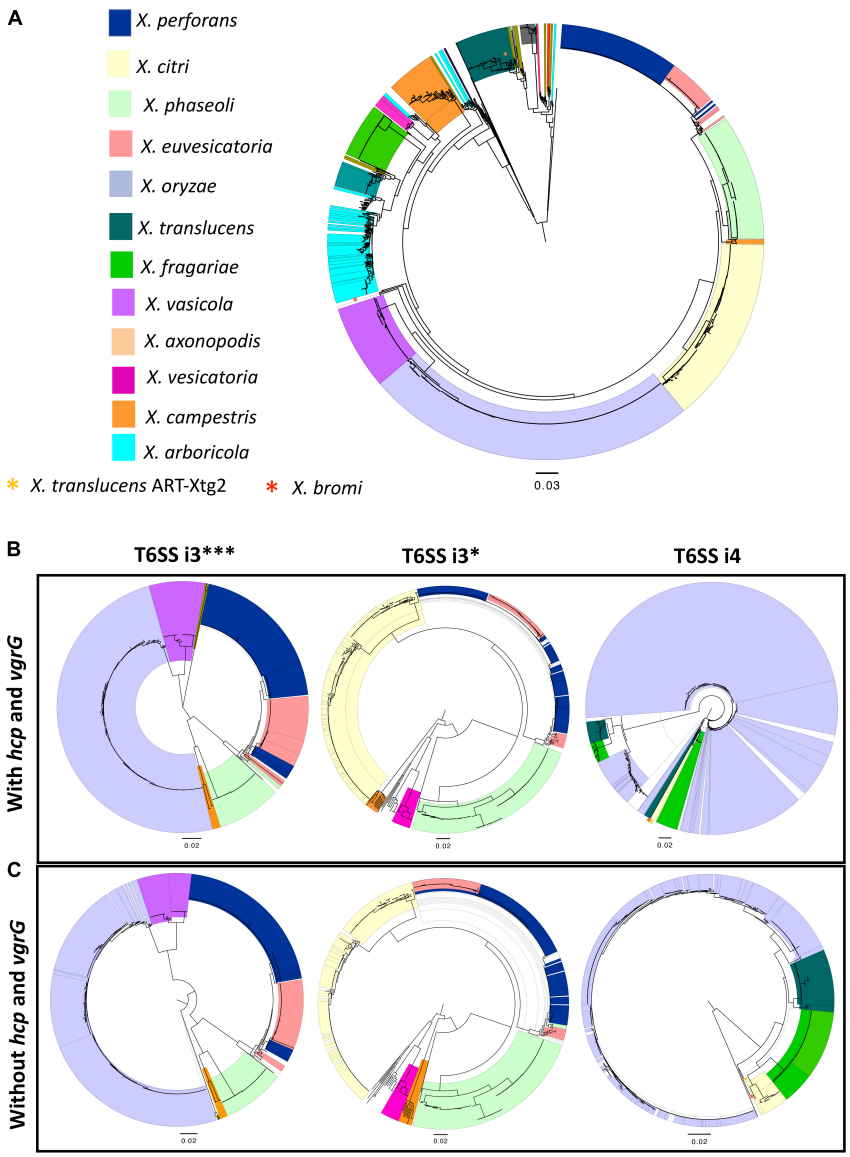


FIGURE 1 | Distribution of T6 clusters in the genus *Xanthomonas*. **(A)** Relationship among 1,577 *Xanthomonas* genomes, multilocus sequence analysis (MLSA) phylogeny of *Xanthomonas* spp. based on the concatenation of partial sequences of housekeeping genes atpD, dnaK, efP, fusA, fyuA, gapA, glnA, gltA, gyrB, lacF, lepA, and rpoD. **(B)** T6 core gene phylogeny of *Xanthomonas* spp. based on the concatenation of T6 core genes *tssA*, *tssB*, *tssC*, *tssD (hcp)*, *tssE*, *tssF*, *tssG*, *tssH*, *tssI (vgrG)*, *tssJ* (only in i4 cluster), *tssK*, *tssL*, and *tssM*. **(C)** T6SS core gene phylogeny of i3***, i3*, and i4 clusters in the absence of *hcp* and *vgrG* core genes. **(B,C)** T6 core gene phylogenies were generated using the automated multilocus sequence analysis pipeline (automlsa2). Midpoint rooted phylogenetic trees in maximum likelihood (ML) criterion were reconstructed using the IQ-TREE and branch support was determined using 1,000 bootstraps and 1,000 SH-aLRT bootstrap replicates.

is present are nested among multiple strains in which the cluster is absent. Thus, a likely scenario, is that the T6 i3* cluster was acquired by lateral gene transfer event by the ancestor of the *X. translucens* clade, after its divergence from the clade that includes *X. albilineans*, *Xanthomonas sontii*, and *X. sacchari* (**Figure 3**). The i3* cluster is present in many but not all group 2 xanthomonads. The phyletic pattern analysis suggests that it experienced multiple independent loss events, for example, in the lineage leading to the clade that includes *X. cannabis* and also in the lineage leading to a clade that includes *X. arboricola* 3307 (**Supplementary Figure 4**). The presence of the i3* cluster in *X.* sp. leaf 131, a commensal *Xanthomonas*, suggests a recent independent acquisition of the cluster in that strain (**Figure 3**). *X. oryzae*, *X. vasicola*, *X. prunicola*, *X. bromi*, and *Xanthomonas nasturtii* have recently lost T6 i3*, whereas some strains belonging to these species have retained a partial cluster. These species belong to clade B based on Merda et al., 2017 core genome phylogeny (**Supplementary Figure 4**). In addition, several strains of *X. phaseoli*, *X. axonopodis*, and *X. euvesicatoria* have experienced recent partial loss of this cluster (**Figure 3**).

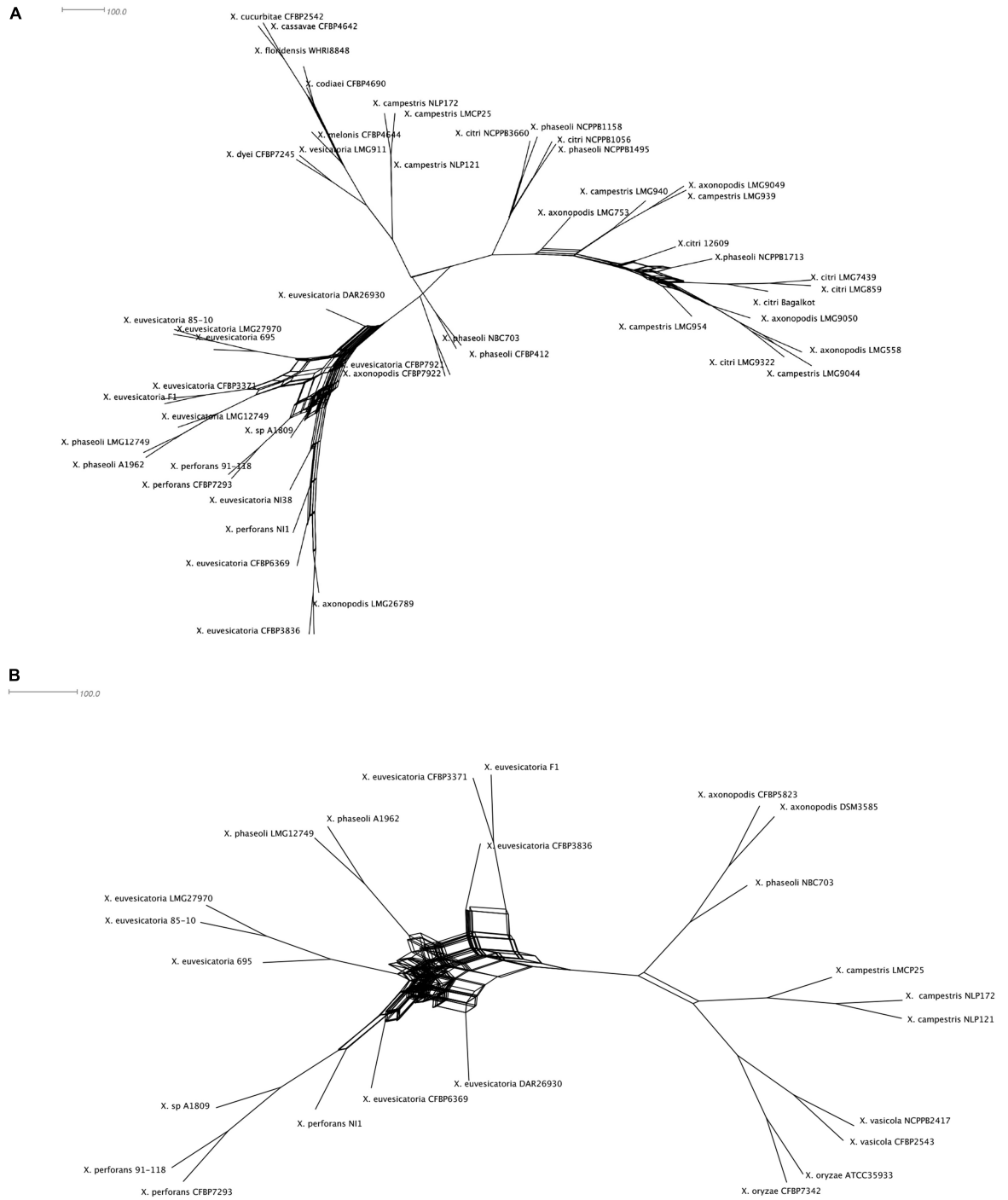


FIGURE 2 | Split decomposition tree of the selected *Xanthomonas* spp. demonstrating incongruencies between the MLSA-based phylogeny and **(A)** T6 i3* without *hcp* and *vgrG* and **(B)** T6 i3*** without *hcp* and *vgrG*. The figure was drawn to scale using Splitstree4. The formation of parallel lines indicates conflicting phylogeny or possible recombination events.

Our analysis suggests that the T6 i4 cluster was independently acquired in *X. oryzae*, *X. translucens*, *X. fragarie*, and in some strains of *X. hortorum*, *X. citri*, *X. phaseoli*, and *X. maliensis*. These multiple gain events are also supported by a SplitsTree analysis that indicated a gene flow of core genes of T6 i4 cluster among *X. oryzae* and *X. translucens*, as well as among *X. citri* and *X. hortorum*, suggesting more recent spread of this cluster with repeated events of horizontal dissemination into distinct clades across *Xanthomonas* phylogeny (**Figure 2**).

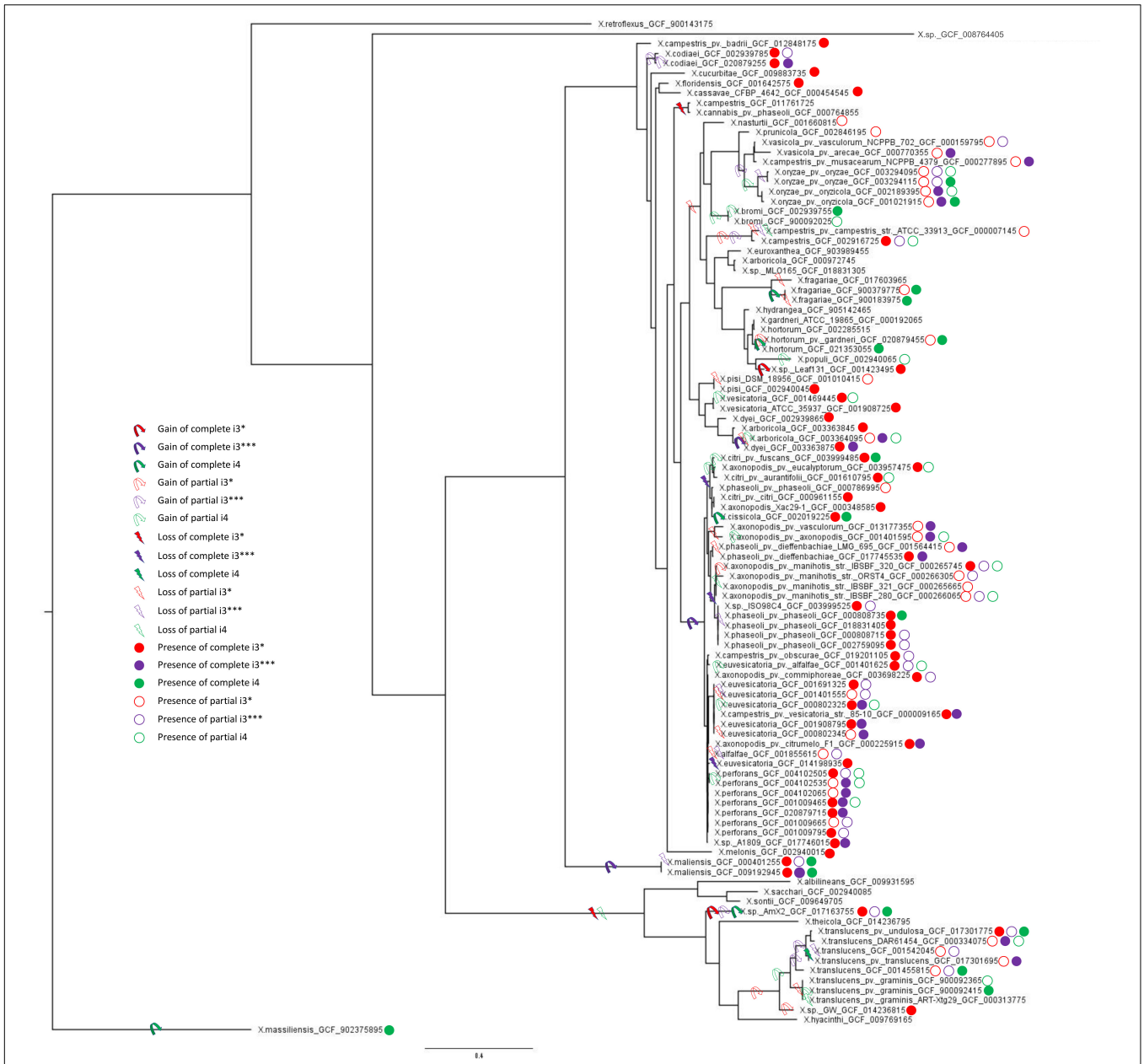


FIGURE 3 | Multilocus sequence analysis phylogenetic tree and T6SS clusters presence/absence and gain/loss prediction for the genus *Xanthomonas*. Gain and loss of individual T6SS genes were mapped onto the MLSA phylogeny of 1,740 *Xanthomonas* spp. From each species, one genome with each phyletic pattern was included; that is, if two or more genomes from the same species had identical presence/absence profile of T6SS genes, only a single representative was included in the analysis. The tree was rooted using *X. massiliensis* as outgroup. The presence of complete and partial T6 clusters (i3*, i3***, and i4) in all the strains was taken from **Table 1**. Ancestral gain and loss events were mapped onto the tree based on the output of GLOOME. For all the predicted events, the probability was higher than 0.5.

T6 i3*** cluster was likely gained by the ancestor of *X. axonopodis* complex, containing species *euvesicatoria*, *citri*, *phaseoli*, and *perforans*, and was subsequently lost by some species within this clade, as indicated by the presence of partial clusters in strains of these species. An independent gain is inferred to have occurred in the lineage leading to species *X. campestris* (clade D, as shown in **Supplementary Figure 4**), a conclusion that is corroborated by the SplitsTree analysis that indicated a gene flow of core genes of this cluster among *X. axonopodis* species complex (clade B) and *X. campestris*. The presence of the i3*** cluster in *X. maliensis* sharing higher similarity to that of *X. euvesicatoria* compared with *X. campestris* might suggest a recent independent acquisition of this cluster into *X. maliensis*. Another independent acquisition of this cluster was also identified in *X. oryzae* pv. oryzicola strains. Presence of intact T6 i3*** cluster in some *X. translucens* strains and

partial in some *X. translucens* strains might indicate its presence in the ancestor of group 1. It is likely that the availability of more genomes of species belonging to group 1 might be able to resolve the phylogenetic inference of evolutionary patterns of the three clusters. Nevertheless, this analysis highlights that all three types of clusters experienced numerous transfer and loss events along the evolution of the analyzed *Xanthomonas* genomes.

Contribution of the T6SS Toward the Evolutionary History of the *Xanthomonas axonopodis* Complex

Presence of more than one T6SS cluster was detected in *X. axonopodis* complex (i3* and i3***) and in *X. oryzae* (i3*** and i4). Strong evidence of acquisition of both i3* and i3*** by the ancestors of *X. axonopodis* species complex was obtained in aforementioned GLOOME analysis. We also observed genetic exchange of core T6 genes among different species of *X. axonopodis* in the SplitsTree analysis. Evolutionary history of *X. axonopodis* complex has been recently well studied, where it has been proposed that the first step of generalist diversification into the *X. axonopodis* subgroups was followed by the second step of ecology-driven specialization that favored the emergence of novel pathovars (Mhedbi-Hajri et al., 2013). Given our prior work that suggested the potential role of core gene *tssM* of cluster T6 i3*, in overall epiphytic fitness in *X. perforans* and during initial stages of disease development (Liyanapathiranage et al., 2021), we hypothesized that during diversification into *X. axonopodis* subgroups, a combination of specific T6SS clusters in individual species provided significant advantage during the process of niche adaptation onto specific hosts. We closely examined the patterns of the evolutionary history of the i3* and i3*** T6 clusters in the *X. axonopodis* complex subspecies groups and the lifestyle of these pathogens. When the distribution of the T6 clusters was examined according to the subspecies group they belong to, it was observed that strains belonging to the same subgroup have a similar distribution of the T6 clusters except for *X. phaseoli* pv. dieffenbachiae LMG 695 in subgroup 9.4. Strains belonging to subgroup 9.2 and *X. phaseoli* pv. dieffenbachiae LMG 695 in subgroup 9.4 possess both complete clusters, whereas strains belonging to subgroups 9.4, 9.5, and 9.6 have complete i3* clusters. Subgroup 9.3 contains only a single complete i3*** cluster, along with an incomplete i3* cluster with one to three core genes (**Table 2**). Interestingly, 9.3 is the only subgroup that has strains pathogenic only on monocot plants. *X. axonopodis* pv. axonopodis belonging to subgroup 9.3 are pathogenic on *Axonopus scoparius*, whereas *X. axonopodis* pv. vasculorum are pathogenic on *Saccharum officinarum*. All the other subgroups consist of strains that are pathogenic on dicot plants, except for *X. axonopodis* pv. allii (9.2) that is pathogenic on *Allium* sp.

To determine if the T6 clusters follow the same evolutionary history as the species belonging to the *X. axonopodis* complex, a phylogeny developed based on the T6 core genes in i3*/i3*** clusters was compared with the phylogenies generated based on the core genome of strains with either i3* or i3*** core genes encoded in their genomes. Incongruencies were observed when comparing core genome phylogeny and phylogeny based on i3* or i3*** core genes (**Figure 4**). Core genome phylogeny clustered subgroups 9.2 and 9.4 in a single clade, whereas i3*** phylogeny clustered subgroups 9.4 and 9.3 in one clade, and i3* phylogeny indicates an early divergence of the subgroup 9.4. Clustering of subgroups 9.5 and 9.6 shows the same arrangement when core genome phylogeny

TABLE 2 | Distribution of the i3* and i3*** clusters in the *X. axonopodis* complex.

Subgroups	*X. axonopodis* pathovars/strains	i3***	i3*
9.1	*X. axonopodis* pv. begoniae strain CFBP2524		
	X. axonopodis pv. spondiae strain CFBP2623		
9.4	*X. phaseoli* pv. dieffenbachiae LMG 695	x	x
	X. axonopodis pv. manihotis str. CFBP1851		x
	X. phaseoli pv. phaseoli CFBP6546 GL1		x
9.2	*X. axonopodis* pv. citrumelo F1	x	x
	X. axonopodis pv. allii CFBP6369	x	x
	X. alfalfae subsp. alfalfae CFBP3836	x	x
	X. euvesicatoria 85-10	x	x
9.3	*X. axonopodis* pv. axonopodis LMG 982	x	only *tssA*
	X. axonopodis pv. vasculorum NCPPB900	x	only *tssA, tssB* and *tssC*
9.5	*X. citri* CFBP3369		x
	X. citri pv. glycines CFBP2526		x
	X. citri pv. malvacearum XcmN1003		x
	X. citri pv. mangiferaeindicae CFBP1716		x
9.6	*X. citri* pv. anacardii CFBP2913	Only *tssB*, *tssC* partial genes	x
	X. citri pv. phaseoli var. fuscans CFBP6165	Only *tssB*, *tssC* partial genes	x
	X. citri pv. fuscans CFBP6991 *phaseoli* GL2	Only *tssB*, *tssC* partial genes	x
	X. citri pv. fuscans pv. phaseoli GL3 CFBP6996	Only *tssB*, *tssC* partial genes	x
	X. citri pv. vignicola CFBP7112	Only *tssB*, *tssC* partial genes	x
	X. citri pv. aurantifolii ICPB 10535	Only *tssB*, *tssC* partial genes	x

"x" indicates the presence of all the core genes.

was compared with the i3* phylogeny. These differences in the phylogeny hint at a possible evolution pattern in T6 clusters independent of its core genome evolution, thus further suggesting that T6 might have played a major role in host adaptation.

T6SS Presence and Lifestyle of the Pathogen

We further assessed the distribution of T6 clusters according to vascular and non-vascular mode of colonization as well as preferred dicot versus monocot host. Most of the dicotyledonous species infecting *Xanthomonas* contain T6 i3*, whereas most of the monocotyledonous host plants infecting *Xanthomonas* contain T6 i3*** cluster (**Table 3**). Absence of i3* cluster was noted in the genomes of species infecting monocots. Furthermore, the T6 cluster distribution was analyzed depending on their lifestyles of vascular versus non-vascular mode. We observed presence of i4 cluster and its association with the ability of the majority of those strains to colonize vascular tissues. Notable examples of the presence of i4 cluster and demonstrated vascular colonization are mentioned below. Presence of i4 cluster was noted in only some strains of species, *X. citri*, more specifically, *X. citri* pv. fuscans and *X. citri* pv. glycines, of which *X. citri* pv. fuscans has been demonstrated to colonize vascular tissues (**Table 3**). While all *X. citri* strains including vascular colonizers contain intact i3* cluster, loss of i3* was seen in *X. oryzae* strains. Presence of intact i4 cluster was observed in 393 strains out of 396 sequenced genomes of *X. oryzae*. Vascular mode of infection has been demonstrated in pathovar oryzae, whereas pathovar oryzicola colonizes non-vascular tissues during infection. A partial i4 cluster was observed in some oryzicola strains. Another vascular pathogen, *X. fragariae*, encodes an intact i4 cluster, similar to *X. oryzae*, with absence of any other intact T6 cluster (**Table 3**). Other known vascular colonizers, *X. phaseoli* pv. manihotis and pv. dieffenbachiae, were found to lack complete i4 cluster, although a partial cluster was noted in some pathovar dieffenbachiae strains. Two *X. phaseoli* pv. phaseoli strains contain intact i4 clusters, strains of which are known to colonize both vascular and mesophyll tissues. *X. campestris*, *X. arboricola* pv. pruni, and *X. hortorum* pv. vitians can be considered as exception to the observation of association of intact i4 cluster with vascular mode of pathogenesis (**Table 3**). Only one sequenced strain of pathovar vitians contains intact i4 cluster. One strain of *X. hortorum* pv. gardneri was found to contain intact i4 cluster. While *X. hortorum* pv. gardneri is not known to colonize vascular tissues, Bernal et al. (2021) recently showed colonization of hydathodes upon inoculation of tomato plants with *X. hortorum* pv. gardneri. *X. translucens* strains show diversity in their preferences to colonize vascular and non-vascular tissues. However, our above observation of association of intact i4 cluster with vascular pathogenesis does not hold true with *X. translucens* species complex, which contains strains known to colonize vascular as well as non-vascular tissues. Some *X. translucens* strains despite being vascular lack T6 cluster (**Table 3**).

T6SS in Non-pathogenic and Environmental Strains

Next, we expanded our analysis of distribution of T6SS clusters to non-pathogenic or environmental isolates. This could provide added clarification to its role in *Xanthomonas* strains regardless of their pathogenicity. Thirteen strains collected from rainwater and 16 non-pathogenic *Xanthomonas* collected from tomato, beans, citrus, pepper, and rice seed or phyllosphere were included in a core genome phylogenetic analysis (**Supplementary Table 2**). Among the non-pathogenic *Xanthomonas* species, *X. maliensis* strains, early-branching strain *X. translucens* F5, and *X. floridensis* WHRI 8848 contain i3* cluster encoded in their genome. Interestingly, *X. maliensis* contains all three clusters, i3*, i3***, and i4 clusters, whereas other non-pathogenic rice leaf–associated *Xanthomonas* species such as *X. sontii* and *X. sacchari* lack any T6 clusters. Similarly, non-pathogenic *X. floridensis* WHRI 8848 isolated from watercress has a complete i3* cluster, whereas watercress pathogenic *X. nasturtii* WHRI 8853 only has a single i3* core gene with flanking regions similar to the *X. floridensis*, suggesting recent loss (**Supplementary Figure 1**). A commensal *Xanthomonas* strain F5 isolated from pepper belonging to group 1 contains complete i3* cluster, with core genes showing 95% to 100% identity to the genes from non-pathogenic *Xanthomonas* strains SI, SS, and GW isolated from perennial ryegrass seed (**Supplementary Table 3**). These strains showed antagonistic activity against key fungal pathogens. Li et al. (2020) hypothesized the involvement of unique secondary metabolites in this antagonistic activity. Our results suggest that T6SS may also play a role in this antagonism to some extent.

Among the *Xanthomonas* strains collected from the rainwater, two *X. arboricola* strains (4461 and 3793) possess a T6 i3* cluster (**Supplementary Figure 4**). Other *X. arboricola* strains including both known pathogens and commensals contain a partial *tssB* gene along with potential regulatory genes encoding serine/threonine protein kinase and RNA polymerase sigma factor and conserved synteny for i3* cluster, indicating recent loss.

DISCUSSION

In this study, we conducted a comprehensive analysis to study the distribution and the evolutionary patterns of multiple T6 clusters in the 1,740 genomes spanning the entire genus *Xanthomonas*. Complete T6 clusters, partial clusters, and even single T6 genes belonging to i3*, i3***, and i4 subgroups of T6 secretion system are dispersed throughout the genus. Our data indicate that T6SS cluster genes have been inherited both vertically and horizontally throughout the genus phylogeny. Like other genera, *Xanthomonas* species contain multiple copies of T6 clusters, with presence of either i3* + i3*** or i3* + i4 combinations being frequently observed. *X. maliensis*, associated with healthy rice seeds or leaves in Mali (Triplett et al., 2015), was the only species identified carrying all three clusters, i3*, i3***, and i4. Being an early branching species within group 2 xanthomonads, conservation of all three clusters in this species and the genomic context is important to infer evolutionary scenarios explaining

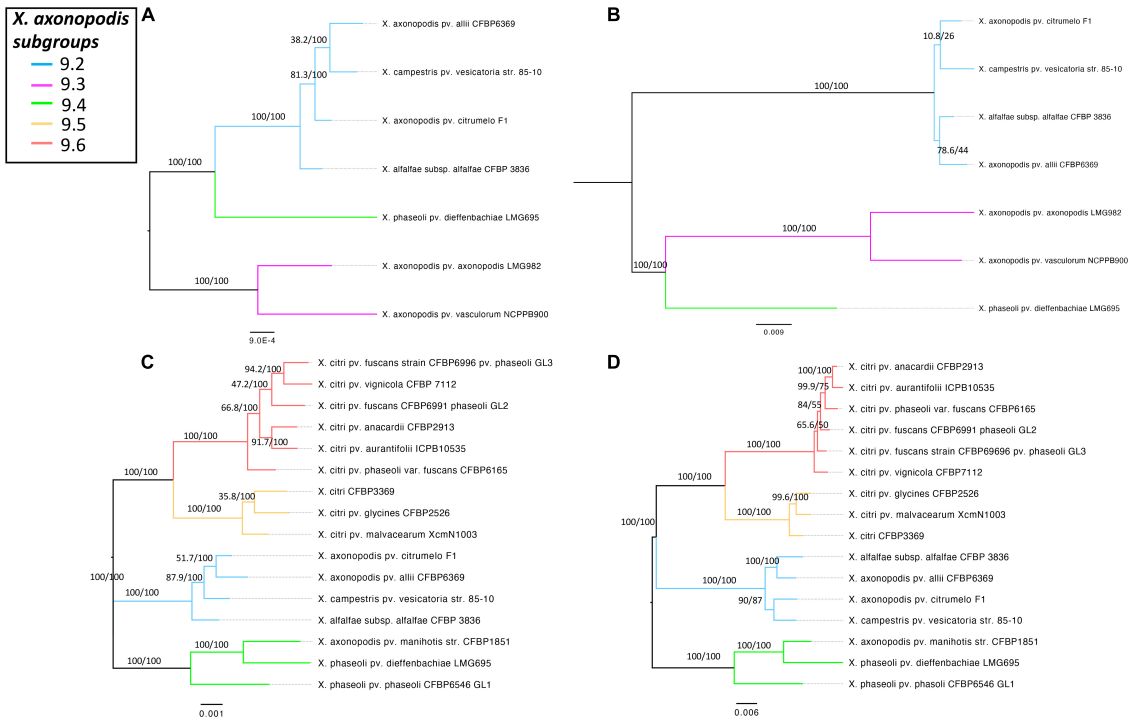


FIGURE 4 | Comparison of the core genome phylogeny of *X. axonopodis* subspecies with T6 i3***/i3* core genome phylogeny to T6 i3*/i3*** core gene phylogeny. **(A)** Core genome phylogeny for the strains with i3***. **(B)** T6SS i3*** phylogeny. **(C)** Core genome phylogeny for the strains with i3*. **(D)** T6SS i3* phylogeny. Core genome phylogenies were developed based on the core alignment of nucleotide sequences using the Roary pipeline (Page et al., 2015), whereas T6 phylogenies were generated using the automated multilocus sequence analysis pipeline (automlsa2). Midpoint rooted phylogenetic tree in maximum likelihood (ML) criterion was developed using the IQ-TREE, and branch support was determined using 1,000 bootstraps and 1,000 SH-aLRT bootstrap replicates.

gain/loss events of the three types of T6 clusters. Simultaneous maintenance of all three clusters in the species is surprising, given the energy investment in these large clusters. Sequencing multiple strains of this species might help address whether presence of all three clusters may represent random flicker or a sign of positive selection for acquisition in the evolutionary history. If the latter is true, this species would represent an ideal candidate for studying differential regulation of each cluster and how the environment influences this regulation.

The majority of the species belonging to *X. axonopodis* complex possess both i3* and i3*** clusters, with exception of *X. citri* that lacks cluster i3***. The loss of i3*, observed in clade A strains and strains of *X. oryzae*, and *X. vasicola* from clade B, was accompanied by gain of i4 cluster. Some strains belonging to *X. citri* and *X. phaseoli*, which possess both i3 and i4 clusters in 1% to 5% of the strains are an exception to the rule of two i3 clusters in *X. axonopodis* complex and to the observation of replacement of i3 by i4. The identification of two *X. phaseoli* strains with the i4 cluster was an unusual observation as the *X. phaseoli* strains were previously only known to harbor i3* and i3*** clusters (Montenegro Benavides et al., 2021). This highlights the importance of the availability of genomes of multiple hundreds of strains of each species for understanding intraspecific diversity. It is unclear whether the loss of i3* and subsequent gain of i4 cluster in species such as *X. oryzae*, and *X. vasicola*, was in response to selection pressure from the environment including resident microflora.

The role of i3* has been suggested not only in competition against resident flora, more specifically, resistance to predation by the bacterivorous amoeba, *Dictyostelium discoideum*, but also in adaptation to the phyllosphere non-vascular environment (Bayer-Santos et al., 2018; Liyanapathiranage et al., 2021). Loss of i3* in lineages containing *X. oryzae* and *X. vasicola*, both vascular pathogens, might be in response to the vascular lifestyle. In the rice pathogen *X. oryzae* pv. oryzicola, T6SS i4 has been identified for its role in the interbacterial competition, whereas in *X. oryzae* pv. oryzae, it was identified for its role in virulence (Choi et al., 2020; Zhu et al., 2020; Montenegro Benavides et al., 2021).

As it has been identified that presence of all the core T6SS genes is required for the functionality of the T6SS (Boyer et al., 2009), it is intriguing to identify some strains that lack single T6 genes and raise the questions about the possible function of that apparently inactive T6 cluster. In our analysis, we observed the T6SS locus lacking *clpV* gene in some of the *X. oryzae* strains, and interestingly, in previous studies on the distribution of the T6 clusters in *Campylobacter jejuni*, T6SS locus lacking the *clpV (tssH)* gene has been noted (Bönemann et al., 2009; Bleumink-Pluym et al., 2013). Even though ClpV is required for the recycling of the TssB/TssC tubular sheath, in *Vibrio cholerae*, a *clpV* mutant was still effective in killing *Escherichia coli*, thus indicating the possibility of function of a *clpV*-related ClpB family

of ATPase encoded from elsewhere compensating for the lack of *clpV* in the T6SS (Pietrosiuk et al., 2011; Zheng et al., 2011).

The distribution of T6SS among *Xanthomonas* species does not correlate with the host preference. Four *Xanthomonas* spp. that cause bacterial leaf spot in tomato and pepper show the diverse distribution of T6 clusters. *X. euvesicatoria* and *X. perforans* have both i3* and i3*** clusters; whereas *X. vesicatoria* has i3*cluster, *X. hortorum pv.* gardneri does not have T6 clusters encoded in its genome, with the exception of a single strain carrying i4 cluster (Potnis et al., 2011). Even the three species that have the i3* cluster show differences in cluster organization (**Supplementary Figure 3**), with different set of putative effectors. Another example could be of rice-associated strains. An extensive diversity in the T6 cluster distribution was observed in the rice-associated xanthomonads, ranging from absence of T6 clusters in non-pathogenic strains, *X. sontii* (Bansal et al., 2019) and *X. sacchari* R1 (Fang et al., 2015), to presence of i4 alone or i3*** and i4 clusters in pathogenic *X. oryzae* and *X. oryzicola*. Another such example is of watercress-associated strains. A non-pathogenic strain, *X. floridensis* WHRI 8848 isolated from the diseased leaves of watercress (*Nasturtium officinale*) (Vicente et al., 2017), contains intact i3* cluster, but a pathogenic strain, *X. nasturtii* WHRI 8853, lacks a functional cluster and encodes only a single partial T6SS gene (*tssA*), indicating a recent loss of the cluster with similar flanking regions.

The distribution of T6 clusters does not depend on host of isolation or pathogenic status. We observed presence of intact T6 clusters in non-pathogenic or commensal strains, even those recovered from rainwater. This observation highlights the role of T6 clusters and associated effectors in conferring fitness during host colonization and also possibly outside the host environment, in conferring traits such as nutrient acquisition, or bacterial communication. While host environment has been suggested to

TABLE 3 | Distribution of the T6 clusters according to the lifestyle of *Xanthomonas* spp.

***Xanthomonas* spp.**	**Vascular(V)/ non-vascular (NV)**	**Reference for tissue/host specificity**	**Host of isolation monocot (M)/dicot (D)**	**T6 clusters**		
				i3* **	**i3***	**i4**
X. floridensis	NV	Vicente et al., 2017	D		x	
X. nasturtii	V	Vicente et al., 2017	D		x(partial)	
X. vesicatoria	NV	Potnis et al., 2015	D		x	
X. citri pv. citri	NV	Lee et al., 2009	D		x	
X. citri pv. malvacearum	NV, V	Mijatovic et al., 2021	D		x	
X. citri pv. fuscans	V, NV	Darrasse et al., 2018	D		x	x*
X. perforans	NV	Potnis et al., 2015	D	x	x	
X. euvesicatoria	NV	Potnis et al., 2015	D	x	x	
X. fragariae	V	Wang et al., 2018	D			x
X. cucurbitae	NV	Vieira et al., 2021	D		x	
X. phaseoli pv. phaseoli	NV,V	Vieira et al., 2021	D		x	x*
X. phaseoli pv. manihotis	V	Zarate-Chaves et al., 2021	D		x	
X. phaseoli pv. dieffenbachiae LMG695	V	Constantin et al., 2017	D	x*	x	
X. cassavae	NV	Zarate-Chaves et al., 2021	D		x	
X. codiaei	V	Rockey et al., 2015	D	x*	x	
X. hortorum pv. vitians	V	Barak et al., 2002	D			x**
X. hortorum pv. gardneri	NV	Bernal et al., 2021	D			x**
X. campestris pv. campestris	V	Bogdanove et al., 2011	D			
X. campestris pv. raphani	NV	Bogdanove et al., 2011	D			
X. arboricola pv. pruni	V	du Plessis, 1984	D			
X. oryzae	V	Bogdanove et al., 2011	M	x		x
X. oryzae pv. oryzicola	NV	Bogdanove et al., 2011	M	x		x#
X. vasicola pv. vasculorum	V	Vieira et al., 2021	M	x		
X. campestris pv. musacearum	V	Vieira et al., 2021	M	x		
X. translucens pv. translucens DSM18974	V	Gluck-Thaler et al., 2020	M	x		
X. translucens pv. arrhenatheri UPB455	V	Gluck-Thaler et al., 2020	M			
X. translucens pv. graminis ART-Xtg29	V	Gluck-Thaler et al., 2020	M			
X. translucens pv. graminis CNC2-P4	V	Gluck-Thaler et al., 2020	M			
X. translucens pv. translucens UPB787	V	Gluck-Thaler et al., 2020	M	x		x
X. translucens pv. cerealis NCPPB1943	V	Gluck-Thaler et al., 2020	M	x		x
X. translucens pv. translucens BLSB3	NV	Gluck-Thaler et al., 2020	M	x		x
X. translucens pv. undulosa NARK-1	NV	Gluck-Thaler et al., 2020	M	x		x
X. translucens DAR61454	NV	Gluck-Thaler et al., 2020	M	x		
X. sacchari	NV	Gluck-Thaler et al., 2020	M			

"x" indicates the presence of all the core genes.
"" indicates that not all strains carry this cluster.*
"#" indicates that some strains carry partial cluster.
*"**" only one sequenced strain so far reported to carry this cluster.*

contribute to induction of T6 clusters in pathogenic bacteria, commensal strains would be an ideal background to understand how T6SS is regulated in absence of network of other virulence factors, although constitutive expression of T6SS cannot be ruled out as observed in *V. cholerae* non-pathogenic strains (Lopez et al., 2020).

Based on the probability of gain/loss events during *Xanthomonas* evolution, as well as information gathered by analysis of flanking regions of T6 clusters and identities of core genes, the data obtained in this study indicate that T6SS was already present in the ancestors of the major xanthomonad lineages. When we mapped the presence of various T6 clusters onto the MLSA tree, we inferred numerous gain events via horizontal gene transfer as well as many loss events. We hypothesize that such evolutionary dynamics is driven by selective pressure related to adaptation to specific environments. In case of T6 i3* cluster, a conserved flanking region across group 2 xanthomonads indicate vertical acquisition of T6 clusters by a common ancestor in a single acquisition event. However, T6 i3* cluster experienced subsequent loss in several lineages, including those in *X. arboricola*, *X. oryzae*, and *X. hortorum*, and has left signatures within genomes, such as partial cluster, or presence of tRNA indicating a rearrangement event. This loss was also followed by gain of T6 i4 cluster in some vascular pathogenic species such as *X. oryzae* and *X. fragarie*. T6 i4 gain events were identified as independent events across the phylogeny. Recent gain of the i4 cluster via horizontal transfer events in multiple species was further confirmed by the fact that flanking regions of the i4 cluster show extreme heterogeneity.

Comparing T6 cluster core gene phylogenies to that of housekeeping genes suggested gene flow of core genes among different species of xanthomonads. This is in agreement with previous findings of T6SS gene clusters as recombination hotspots in *X. euvesicatoria* species complex (Newberry et al., 2019), which suggests reshuffling of alleles of core genes of T6 clusters. In addition to core genes, we also observed reticulations in phylogenetic network, indicating frequent exchange of Hcp and VgrG among *Xanthomonas* species. Such events of reshuffling of VgrG and Hcp have been observed as a strong force in diversification of effector-immunity proteins in many genera, such as *Agrobacterium* and *Vibrio* (Salomon et al., 2015; Thomas et al., 2017; Wu et al., 2021). Although it is true that poorly resolved phylogenies or phylogenies of closely related strains with low branch support could appear as reticulations in the network, reticulations observed specifically for distantly related species suggest that recombination could be a highly likely explanation in those cases.

Species belonging to the *X. axonopodis* complex are pathogenic on a wide range of crops (Vauterin et al., 1995) and are classified into six groups named 9.1 to 9.6 within *X. axonopodis* (Rademaker et al., 2005; Young et al., 2008; Hajri et al., 2009; Bui Thi Ngoc et al., 2010; Mhedbi-Hajri et al., 2013). According to the core genome phylogeny and i3* or i3*** core genes phylogeny in this current study, we observe the same pattern in the way each strain categorizes into the subgroupings as explained in the previous literature, suggesting acquisition of these clusters by the ancestor of *X. axonopodis*. Mhedbi-Hajri et al. (2013) speculated that the first step of generalist diversification was followed by a second step of ecology-driven specialization that probably occurred in the past two centuries and favored the emergence of novel pathovars. This second event of host specialization has occurred likely through agricultural intensification and expansion that facilitated genetic exchanges of virulence-associated genes. All of these pathovars are restricted to one of the six subgroups and most of them formed monophyletic (e.g., citri, mangiferaeindicae, malvacearum, begonia, and manihotis) clusters, whereas some strains were polyphyletic (e.g., dieffenbachiae, glycines, and phaseoli) (Hajri et al., 2009; Mhedbi-Hajri et al., 2013). Some strains that belong to the same pathovar can be found in phylogenetically distant clusters. For example, *X. axonopodis* strains that infect legumes can be found in subgroups 9.2, 9.4, and 9.6; mango *Mangifera indica*–infecting strains were found in subgroups 9.5 and 9.6, and citrus-infecting strains were identified in clusters 9.2, 9.5, and 9.6. Similarly, *X. axonopodis* strains that are pathogenic on phylogenetically distant hosts are clustered in the same subgroups, as seen with subgroup 9.5, where citrus-, mango-, legume-, and cotton-infecting *X. axonopodis* strains are grouped together. An independent rapid evolution in the subgroup 9.3 compared with the other subgroups might also be reflected in its loss of i3* cluster during niche adaptation. *X. phaseoli* pv. dieffenbachiae is polyphyletic and found in subgroups 9.6 (A- *Dieffenbachia* sp.), 9.2 (B- *Philodendron* sp.) and 9.4 (C- *Anthurium* sp.) according to the *rpoD* sequences analysis and Rep-PCR, this grouping is also dependent on the host of isolation (Hajri et al., 2009; Mhedbi-Hajri et al., 2013). Our observation of heterogeneity of T6 cluster distribution in the subgroup 9.4 suggests preferences in specific T6 clusters during niche adaptation of different host species. This phylogenetic analysis and T6SS cluster distributions suggest that T6SS cluster conservation and associated effector alleles might have played a role during host specialization in the *X. axonopodis* strains in addition to the already suggested factors such as methyl-accepting chemotaxis proteins and type III effectors (Hajri et al., 2009; Mhedbi-Hajri et al., 2013).

In conclusion, our findings point to the scenario in which the different T6 clusters are often gained and lost during the evolution of *Xanthomonas* species, with some species encoding two clusters, whereas others, not a single one. We hypothesize that the majority of gain and loss events are not random, but rather reflect adaptations and specialization to specific environments. The factors involved in niche adaptation have been of interest in case of plant pathogenic bacteria, and a new paradigm has suggested involvement of multiple genetic determinants underlying this process with small to major effect genes during the process of adaptation of a pathogen to a host type or to a particular tissue (Bogdanove et al., 2011; Huang et al., 2015; Jacques et al., 2016). Here, we explored association of T6 clusters and preference of *Xanthomona*s species toward dicot or monocot hosts and vascular or non-vascular mode of infection. As not all species have been studied for their breadth of host range or tissue colonization, the associations evaluated here are based on current literature. We noted the absence of i3* cluster in monocot-infecting pathogens. The gain/loss patterns

also indicated loss of i3* followed by gain of i4 cluster in vascular pathogenic species in group 2 xanthomonads, such as *oryzae* and *fragarie*. But this observation was not true in case of *X. translucens* species complex and *X. campestris* pv. campestris, *X. vasicola*. In brief, there does not seem to be association of specific T6 cluster for a lifestyle of *Xanthomonas* when evaluating the entire genus as a whole. However, functional characterization using T6SS-defective mutants of *Xanthomonas* has suggested importance of T6SS in overall fitness during host colonization, as well as its role in competing with the resident flora. As pathogen fitness criteria vary depending on the niche that pathogen experiences, involvement of T6 cluster/s and its role in regulatory network may vary depending on the host environment. Thus, it is likely that associations that we have observed with gain/loss patterns for a specific T6 cluster may be valid for specific clades that experience similar host environments or that harness similar virulence factor repertoires to coordinate the process of colonization. It is also possible that secreted effectors or toxins may be more important in specialization to the host environment rather than association of a specific T6 cluster/s.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

NP conceived the project. PL, NW, and NP conducted the comparative genome analyses and SplitsTree analysis, generated the core T6 gene and core genome phylogenies, and wrote the manuscript. OA conducted the Gloome analysis. OA, NP, NW, and TP interpreted the Gloome analysis results. NW, OA, and TP edited the manuscript. All authors approved final manuscript.

FUNDING

This research was supported in part with funds from the NSF-CAREER award to NP, award, IOS-1942956 from the Plant Biotic Interactions Program of the National Science Foundation (NSF). OA and NW were supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. OA was also supported by a fellowship from the Dalia and Eli Hurvitz Foundation LTD. This work was made possible in part by a grant of high-performance computing resources and technical support from the Alabama Supercomputer Authority.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.840308/full#supplementary-material

REFERENCES

Abendroth, U., Adlung, N., Otto, A., Grüneisen, B., Becher, D., and Bonas, U. (2017). Identification of new protein-coding genes with a potential role in the virulence of the plant pathogen *Xanthomonas euvesicatoria*. *BMC Genomics* 18:625. doi: 10.1186/s12864-017-4041-7

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Alvarez-Martinez, C. E., Sgro, G. G., Araujo, G. G., Paiva, M. R. N., Matsuyama, B. Y., Guzzo, C. R., et al. (2020). Secrete or perish: the role of secretion systems in Xanthomonas biology. *Comput. Struct. Biotechnol. J.* 19, 279–302. doi: 10.1016/j.csbj.2020.12.020

An, S.-Q., Potnis, N., Dow, M., Vorhölter, F.-J., He, Y.-Q., Becker, A., et al. (2020). Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen Xanthomonas. *FEMS Microbiol. Rev.* 44, 1–32. doi: 10.1093/femsre/fuz024

Bansal, K., Kumar, S., and Patil, P. B. (2020). Phylogenomic insights into diversity and evolution of nonpathogenic Xanthomonas strains associated with citrus. *mSphere* 5, e00087–20. doi: 10.1128/mSphere.00087-20

Bansal, K., Midha, S., Kumar, S., Kaur, A., Sonti, R. V., and Patil, P. B. (2019). Ecological and evolutionary insights into pathogenic and non-pathogenic rice associated Xanthomonas. *bioRxiv* [Preprint]. doi: 10.1101/453373

Barak, J. D., Koike, S. T., and Gilbertson, R. L. (2002). Movement of Xanthomonas campestris pv. vitians in the stems of lettuce and seed contamination. *Plant Pathol.* 51, 506–512.

Bayer-Santos, E., Ceseti, L., de, M., Farah, C. S., and Alvarez-Martinez, C. E. (2019). Distribution, function and regulation of type 6 secretion systems of Xanthomonadales. *Front. Microbiol.* 10:1635. doi: 10.3389/fmicb.2019.01635

Bayer-Santos, E., Lima, L. D. P., Ceseti, L. M., Ratagami, C. Y., de Santana, E. S., da Silva, A. M., et al. (2018). *Xanthomonas citri* T6SS mediates resistance to Dictyostelium predation and is regulated by an ECF σ factor and cognate Ser/Thr kinase. *Environ. Microbiol.* 20, 1562–1575. doi: 10.1111/1462-2920.14085

Bernal, E., Deblais, L., Rajashekara, G., and Francis, D. M. (2021). Bioluminescent Xanthomonas hortorum pv. gardneri as a tool to quantify bacteria in planta, screen germplasm, and identify infection routes on leaf surfaces. *Front. Plant Sci.* 12:667351. doi: 10.3389/fpls.2021.667351

Bingle, L. E., Bailey, C. M., and Pallen, M. J. (2008). Type VI secretion: a beginner's guide. *Curr. Opin. Microbiol.* 11, 3–8. doi: 10.1016/j.mib.2008.01.006

Bleumink-Pluym, N. M. C., van Alphen, L. B., Bouwman, L. I., Wösten, M. M., and van Putten, J. P. (2013). Identification of a functional type VI secretion system in campylobacter jejuni conferring capsule polysaccharide sensitive cytotoxicity. *PLoS Pathog.* 9:e1003393. doi: 10.1371/journal.ppat.1003393

Blondel, C. J., Jiménez, J. C., Contreras, I., and Santiviago, C. A. (2009). Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. *BMC Genomics* 10:354. doi: 10.1186/1471-2164-10-354

Bogdanove, A. J., Koebnik, R., Lu, H., Furutani, A., White, F. F., Salzberg, S. L., et al. (2011). Two new complete genome sequences offer insights into host and tissue specificity of plant pathogenic Xanthomonas spp. *J. Bacteriol.* 193, 5450–5464. doi: 10.1128/JB.05262-11

Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H., and Mogk, A. (2009). Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J.* 28, 315–325. doi: 10.1038/emboj.2008.269

Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y., and Attree, I. (2009). Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics* 10:104. doi: 10.1186/1471-2164-10-104

Bui Thi Ngoc, L., Vernière, C., Jouen, E., Ah-You, N., Lefeuvre, P., Chiroleu, F., et al. (2010). Amplified fragment length polymorphism and multilocus sequence analysis-based genotypic relatedness among pathogenic variants of

Xanthomonas citri pv. citri and Xanthomonas campestris pv. bilvae. *Int. J. Syst. Evol. Microbiol.* 60, 515–525. doi: 10.1099/ijs.0.009514-0

Büttner, D., and Bonas, U. (2010). Regulation and secretion of Xanthomonas virulence factors. *FEMS Microbiol. Rev.* 34, 107–133. doi: 10.1111/j.1574-6976.2009.00192.x

Ceseti, L. M., de Santana, E. S., Ratagami, C. Y., Barreiros, Y., Lima, L. D. P., Dunger, G., et al. (2019). The *Xanthomonas citri* pv. citri Type VI secretion system is induced during epiphytic colonization of citrus. *Curr. Microbiol.* 76, 1105–1111. doi: 10.1007/s00284-019-01735-3

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2021). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* 49, D751–D763. doi: 10.1093/nar/gkaa939

Choi, Y., Kim, N., Mannaa, M., Kim, H., Park, J., Jung, H., et al. (2020). Characterization of Type VI secretion system in Xanthomonas oryzae pv. oryzae and its role in virulence to rice. *Plant Pathol. J.* 36, 289–296. doi: 10.5423/PPJ.NT.02.2020.0026

Cohen, O., and Pupko, T. (2011). Inference of gain and loss events from phyletic patterns using stochastic mapping and maximum parsimony a simulation study. *Genome Biol. Evol.* 3, 1265–1275. doi: 10.1093/gbe/evr101

Cohen, O., Ashkenazy, H., Belinky, F., Huchon, D., and Pupko, T. (2010). GLOOME: gain loss mapping engine. *Bioinformatics* 26, 2914–2915. doi: 10.1093/bioinformatics/btq549

Constantin, E. C., Haegeman, A., Van Vaerenbergh, J., Baeyen, S., Van Malderghem, C., Maes, M., et al. (2017). Pathogenicity and virulence gene content of *Xanthomonas* strains infecting Araceae, formerly known as *Xanthomonas axonopodis* pv. *dieffenbachiae*. *Plant Pathol.* 66, 1539–1554. doi: 10.1111/ppa.12694

Coulthurst, S. J. (2013). The Type VI secretion system – a widespread and versatile cell targeting system. *Res. Microbiol.* 164, 640–654. doi: 10.1016/j.resmic.2013.03.017

Darrasse, A., Barret, M., Cesbron, S., Compant, S., and Jacques, M.-A. (2018). Niches and routes of transmission of *Xanthomonas citri* pv. fuscans to bean seeds. *Plant Soil* 422, 115–128.

Davis, E. W., Okrent, R. A., Manning, V. A., and Trippe, K. M. (2021). Unexpected distribution of the 4-formylaminooxyvinylglycine (FVG) biosynthetic pathway in *Pseudomonas* and beyond. *PLoS One* 16:e0247348. doi: 10.1371/journal.pone.0247348

du Plessis, H. J. (1984). Scanning electron microscopy of Xanthomonas campestris pv. pruni in plum petioles and buds. *J. Phytopathol.* 109, 277–284.

Durand, E., Cambillau, C., Cascales, E., and Journet, L. (2014). VgrG, Tae, Tle, and beyond: the versatile arsenal of Type VI secretion effectors. *Trends Microbiol.* 22, 498–507. doi: 10.1016/j.tim.2014.06.004

Durand, E., Nguyen, V. S., Zoued, A., Logger, L., Péhau-Arnaudet, G., Aschtgen, M.-S., et al. (2015). Biogenesis and structure of a type VI secretion membrane core complex. *Nature* 523, 555–560. doi: 10.1038/nature14667

Fang, Y., Lin, H., Wu, L., Ren, D., Ye, W., Dong, G., et al. (2015). Genome sequence of Xanthomonas sacchari R1, a biocontrol bacterium isolated from the rice seed. *J. Biotechnol.* 206, 77–78. doi: 10.1016/j.jbiotec.2015.04.014

Fischer-Le Saux, M., Bonneau, S., Essakhi, S., Manceau, C., and Jacques, M.-A. (2015). Aggressive emerging pathovars of *Xanthomonas arboricola* represent widespread epidemic clones distinct from poorly pathogenic strains, as revealed by multilocus sequence typing. *Appl. Environ. Microbiol.* 81, 4651–4668. doi: 10.1128/AEM.00050-15

Gambette, P., and Huson, D. H. (2008). Improved layout of phylogenetic networks. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 5, 472–479. doi: 10.1109/tcbb.2007.1046

Gluck-Thaler, E., Cerutti, A., Perez-Quintero, A. L., Butchacas, J., Roman-Reyna, V., Madhavan, V. N., et al. (2020). Repeated gain and loss of a single gene modulates the evolution of vascular plant pathogen lifestyles. *Sci. Adv.* 6:eabc4516. doi: 10.1126/sciadv.abc4516

Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., et al. (2009). A «Repertoire for repertoire» hypothesis: repertoires of type three effectors are candidate determinants of host specificity in Xanthomonas. *PLoS One* 4:e6632. doi: 10.1371/journal.pone.0006632

Hernandez, R. E., Gallegos-Monterrosa, R., and Coulthurst, S. J. (2020). Type VI secretion system effector proteins: effective weapons for bacterial competitiveness. *Cell. Microbiol.* 22:e13241. doi: 10.1111/cmi.13241

Huang, C.-L., Pu, P.-H., Huang, H.-J., Sung, H.-M., Liaw, H.-J., Chen, Y. M., et al. (2015). Ecological genomics in Xanthomonas: the nature of genetic adaptation with homologous recombination and host shifts. *BMC Genomics* 16:188. doi: 10.1186/s12864-015-1369-8

Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Iranzo, J., Wolf, Y. I., Koonin, E. V., and Sela, I. (2019). Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. *Nat. Commun.* 10:5376. doi: 10.1038/s41467-019-13429-2

Jacques, M.-A., Arlat, M., Boulanger, A., Boureau, T., Carrere, S., Cesbron, S., et al. (2016). Uisng ecology, physeiology, and genomics to understand host specificity in Xanthomonas. *Annu. Rev. Phytopathol.* 54, 163–187. doi: 10.1146/annurev-phyto-080615-100147

Johansen, J. E., Binnerup, S. J., Kroer, N., and Mølbak, L. (2005). Luteibacter rhizovicinus gen. nov., sp. nov., a yellow-pigmented gammaproteobacterium isolated from the rhizosphere of barley (Hordeum vulgare L.). *Int. J. Syst. Evol. Microbiol.* 55, 2285–2291. doi: 10.1099/ijs.0.63497-0

Joseph, S., and Forsythe, S. (2012). Insights into the emergent bacterial pathogen cronobacter spp., generated by multilocus sequence typing and analysis. *Front. Microbiol.* 3:397. doi: 10.3389/fmicb.2012.00397

Jun, S.-R., Sims, G. E., Wu, G. A., and Kim, S.-H. (2010). Whole-proteome phylogeny of prokaryotes by feature frequency profiles: an alignment-free method with optimal feature resolution. *Proc. Natl. Acad. Sci. U. S. A.* 107, 133–138. doi: 10.1073/pnas.0913033107

Katoh, M., Miyata, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., et al. (2007). Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. *PLoS Genet.* 3:e231. doi: 10.1371/journal.pgen.0030231

Kyrova, E. I., Dzhalilov, F. S., and Ignatov, A. N. (2020). The role of epiphytic populations in pathogenesis of the genus Xanthomonas bacteria. *Bio Web Conf.* 23:03010. doi: 10.1051/bioconf/20202303010

Lee, I.-J., Kim, K.-W., Hyun, J. W., Lee, Y. H., and Park, E.-W. (2009). Comparative ultrastructure of nonwounded mexican lime and yuzu leaves infected with the citrus canker bacterium *Xanthomonas citri* pv. citri. *Microscopy Res. Technique* 72, 507–516. doi: 10.1002/jemt.20707

Li, T., Mann, R., Sawbridge, T., Kaur, J., Auer, D., and Spangenberg, G. (2020). Novel Xanthomonas species from the perennial ryegrass seed microbiome - Assessing the bioprotection activity of non-pathogenic relatives of pathogens. *Front. Microbiol.* 11:1991. doi: 10.3389/fmicb.2020.01991

Lien, Y.-W., and Lai, E.-M. (2017). Type VI secretion effectors: methodologies and biology. *Front. Cell Infect. Microbiol.* 7:254. doi: 10.3389/fcimb.2017.00254

Liyanapathiranage, P., Jones, J. B., and Potnis, N. (2021). A mutation of a single core gene, tssM, of Type VI secretion system of *Xanthomonas perforans* influences virulence, epiphytic survival and transmission during pathogenesis on tomato. *Phytopathology*. doi: 10.1094/PHYTO-02-21-0069-R [Epub ahead of print].

Lopez, J., Ly, P. M., and Feldman, M. F. (2020). The tip of the VgrG spike is essential to functional type VI secretion system assembly in *Acinetobacter baumannii*. *mBio* 11, e02761–19. doi: 10.1128/mBio.02761-19

Ma, J., Pan, Z., Huang, J., Sun, M., Lu, C., and Yao, H. (2017). The Hcp proteins fused with diverse extended-toxin domains represent a novel pattern of antibacterial effectors in type VI secretion systems. *Virulence* 8, 1189–1202. doi: 10.1080/21505594.2017.1279374

Merda, D., Briand, M., Bosis, E., Rousseau, C., Portier, P., Barret, M., et al. (2017). Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in Xanthomonas plant pathogens. *Mol. Ecol.* 26, 5939–5952. doi: 10.1111/mec.14343

Meuskens, I., Saragliadis, A., Leo, J. C., and Linke, D. (2019). Type V secretion systems: an overview of passenger domain functions. *Front. Microbiol.* 10:1163. doi: 10.3389/fmicb.2019.01163

Mhedbi-Hajri, N., Hajri, A., Boureau, T., Darrasse, A., Durand, K., Brin, C., et al. (2013). Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PLoS One* 8:e58474. doi: 10.1371/journal.pone.0058474

Mijatovic, J., Severns, P. M., Kemerait, R. C., Walcott, R. R., and Kvitko, B. (2021). Patterns of seed-to-seedling transmission of *Xanthomonas citri* pv. malvacearum, the causal agent of cotton bacterial blight. *Phytopathology* 111, 2176–2184. doi: 10.1094/PHYTO-02-21-0057-R

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534.

Montenegro Benavides, N. A., Alvarez, B. A., Arrieta-Ortiz, M. L., Rodriguez-R, L. M., Botero, D., Tabima, J. F., et al. (2021). The type VI secretion system of *Xanthomonas phaseoli* pv. manihotis is involved in virulence and in vitro motility. *BMC Microbiol.* 21:14. doi: 10.1186/s12866-020-02066-1

Newberry, E. A., Bhandari, R., Minsavage, G. V., Timilsina, S., Jibrin, M. O., Kemble, J., et al. (2019). Independent evolution with the gene flux originating from multiple xanthomonas species explains genomic heterogeneity in Xanthomonas *perforans*. *Appl. Environ. Microbiol.* 85, e00885–19. doi: 10.1128/AEM.00885-19

Nguyen, V. S., Douzi, B., Durand, E., Roussel, A., Cascales, E., and Cambillau, C. (2018). Towards a complete structural deciphering of Type VI secretion system. *Curr. Opin. Struct. Biol.* 49, 77–84. doi: 10.1016/j.sbi.2018.01.007

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Palmer, T., Finney, A. J., Saha, C. K., Atkinson, G. C., and Sargent, F. (2020). A holin/peptidoglycan hydrolase-dependent protein secretion system. *Mol. Microbiol.* 115, 345–355. doi: 10.1111/mmi.14599

Pietrosiuk, A., Lenherr, E. D., Falk, S., Bönemann, G., Kopp, J., Zentgraf, H., et al. (2011). Molecular basis for the unique role of the AAA+ chaperone ClpV in type VI protein secretion. *J. Biol. Chem.* 286, 30010–30021. doi: 10.1074/jbc.M111.253377

Potnis, N., Krasileva, K., Chow, V., Almeida, N. F., Patil, P. B., Ryan, R. P., et al. (2011). Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genomics* 12:146. doi: 10.1186/1471-2164-12-146

Potnis, N., Timilsina, S., Strayer, A., Shantharaj, D., Barak, J. D., Paret, M. L., et al. (2015). Bacterial spot of tomato and pepper: diverse Xanthomonas species with a wide variety of virulence factors posing a worldwide challenge. *Mol. Plant Pathol.* 16, 907–920. doi: 10.1111/mpp.12244

Rademaker, J. L. W., Louws, F. J., Schultz, M. H., Rossbach, U., Vauterin, L., Swings, J., et al. (2005). A comprehensive species to strain taxonomic framework for Xanthomonas. *Phytopathology* 95, 1098–1111. doi: 10.1094/PHYTO-95-1098

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rockey, W., Potnis, N., Hong, J. C., Timilsina, S., Vallad, G. E., Jones, J. B., et al. (2015). Multilocus sequence analysis reveals genetic diversity in Xanthomonads associated with poinsettia production. *Plant Dis.* 99, 874–882. doi: 10.1094/PDIS-08-14-0867-RE

Salomon, D., Klimko, J. A., Trudgian, D. C., Kinch, L. N., Grishin, N. V., Mirzaei, H., et al. (2015). Type VI secretion system toxins horizontally shared between marine bacteria. *PLoS Pathog.* 11:e1005128. doi: 10.1371/journal.ppat.1005128

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shalom, G., Shaw, J. G., and Thomas, M. S. (2007). In vivo expression technology identifies a type VI secretion system locus in Burkholderia pseudomallei that is induced upon invasion of macrophages. *Microbiology* 153, 2689–2699. doi: 10.1099/mic.0.2007/006585-0

Shyntum, D., Venter, S., Moleleki, L., Toth, I., and Coutinho, T. (2014). Comparative genomics of type VI secretion systems in strains of Pantoea ananatis from different environments. *BMC Genomics* 15:163. doi: 10.1186/1471-2164-15-163

Silverman, J. M., Austin, L. S., Hsu, F., Hicks, K. G., Hood, R. D., and Mougous, J. D. (2011). Separate inputs modulate phosphorylation-dependent and -independent type VI secretion activation: posttranslational regulation of type VI secretion. *Mol. Microbiol.* 82, 1277–1290. doi: 10.1111/j.1365-2958.2011.07889.x

Silverman, J. M., Brunet, Y. R., Cascales, E., and Mougous, J. D. (2012). Structure and Regulation of the Type VI Secretion System. *Ann. Rev. Microbiol.* 66, 453–472. doi: 10.1146/annurev-micro-121809-151619

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Thomas, J., Watve, S. S., Ratcliff, W. C., and Hammer, B. K. (2017). Horizontal gene transfer of functional type VI killing genes by natural transformation. *mBio* 8, e00654–17. doi: 10.1128/mBio.00654-17

Timilsina, S., Jibrin, M. O., Potnis, N., Minsavage, G. V., Kebede, M., Schwartz, A., et al. (2015). Multilocus sequence analysis of xanthomonads causing bacterial spot of tomato and pepper plants reveals strains generated by recombination among species and recent global spread of Xanthomonas gardneri. *Appl. Environ. Microbiol.* 81, 1520–1529. doi: 10.1128/AEM.03000-14

Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F. F., et al. (2020). Xanthomonas diversity, virulence and plant–pathogen interactions. *Nat. Rev. Microbiol.* 18, 415–427. doi: 10.1038/s41579-020-0361-8

Triplett, L. R., Verdier, V., Campillo, T., Van Malderghem, C., Cleenwerck, I., Maes, M., et al. (2015). Characterization of a novel clade of Xanthomonas isolated from rice leaves in Mali and proposal of Xanthomonas maliensis sp. nov. *Antonie van Leeuwenhoek* 107, 869–881. doi: 10.1007/s10482-015-0379-5

Vauterin, L., Hoste, B., Kersters, K., and Swings, J. Y. (1995). Reclassification of Xanthomonas. *Int. J. Syst. Evol. Microbiol.* 45, 472–489. doi: 10.1099/00207713-45-3-472

Vicente, J. G., Rothwell, S., Holub, E. B., and Studholme, D. J. (2017). Pathogenic, phenotypic and molecular characterisation of Xanthomonas nasturtii sp. nov. and Xanthomonas floridensis sp. nov., new species of Xanthomonas associated with watercress production in Florida. *Int. J. Syst. Evol. Microbiol.* 67, 3645–3654. doi: 10.1099/ijsem.0.002189

Vieira, P. S., Bonfim, I. M., Araujo, E. A., Melo, R. R., Lima, A. R., Fessel, M. R., et al. (2021). Xyloglucan processing machinery in Xanthomonas pathogens and its role in the transcriptional activation of virulence factors. *Nat. Commun.* 12:4049. doi: 10.1038/s41467-021-24277-4

Wang, H., McTavish, C., and Turecheck, W. W. (2018). Colonization and movement of Xanthomonas fragariae in strawberry tissues. *Phytopathology* 108, 681–690. doi: 10.1094/PHYTO-10-17-0356-R

White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. (2009). The type III effectors of Xanthomonas. *Mol. Plant Pathol.* 10, 749–766. doi: 10.1111/j.1364-3703.2009.00590.x

Wu, C. F., Weisberg, A. J., Davis, E. II., Chou, L., Khan, S., Lai, E.-M., et al. (2021). Diversification of the type VI secretion system in Agrobacteria. *mBio* 12:e0192721. doi: 10.1128/mBio.01927-21

Young, J. M., Park, D.-C., Shearman, H. M., and Fargier, E. (2008). A multilocus sequence analysis of the genus Xanthomonas. *Syst. Appl. Microbiol.* 31, 366–377. doi: 10.1016/j.syapm.2008.06.004

Zarate-Chaves, C. A., de la Cruz, D. G., Verdier, V., Lopez, C. E., and Bernal, A. (2021). Cassava diseases caused by *Xanthomonas phaseoli* pv. manihotis and *Xanthomonas cassavae*. *Mol. Plant Pathol.* 22, 1520–1537. doi: 10.1111/mpp.13094

Zheng, J., Ho, B., and Mekalanos, J. J. (2011). Genetic analysis of anti-amoebae and anti-bacterial activities of the type VI secretion system

in *Vibrio cholerae*. *PLoS One* 6:e23876. doi: 10.1371/journal.pone.0023876

Zhu, P.-C., Li, Y.-M., Yang, X., Zou, H.-F., Zhu, X.-L., Niu, X.-N., et al. (2020). Type VI secretion system is not required for virulence on rice but for inter-bacterial competition in Xanthomonas oryzae pv. oryzicola. *Res. Microbiol.* 171, 64–73. doi: 10.1016/j.resmic.2019.10.004

Zoued, A., Brunet, Y. R., Durand, E., Aschtgen, M.-S., Logger, L., Douzi, B., et al. (2014). Architecture and assembly of the Type VI secretion system. *Biochim. Biophys. Acta Mol. Cell Res.* 1843, 1664–1673.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 April 2022
doi: 10.3389/fmicb.2022.840311

Comparative Genomic Analysis of the Lettuce Bacterial Leaf Spot Pathogen, *Xanthomonas hortorum* pv. *vitians*, to Investigate Race Specificity

Emma Rosenthal[1†], *Neha Potnis*[2†] *and Carolee T. Bull*[1*†]

[1] Department of Plant Pathology and Environmental Microbiology, Pennsylvania State University, University Park, PA, United States, [2] Department of Entomology and Plant Pathology, Auburn University, Auburn, AL, United States

OPEN ACCESS

Edited by:
Chih-Horng Kuo,
Academia Sinica, Taiwan

Reviewed by:
Joana G. Vicente,
Fera Science Ltd., United Kingdom
Nemanja Kuzmanovic,
Julius Kühn-Institut, Germany

***Correspondence:**
Carolee T. Bull
caroleebull@psu.edu

†ORCID:
Emma Rosenthal
orcid.org/0000-0002-6397-9696
Neha Potnis
orcid.org/0000-0002-5418-5894
Carolee T. Bull
orcid.org/0000-0002-2592-407X

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 21 December 2021
Accepted: 09 February 2022
Published: 18 April 2022

Citation:
Rosenthal E, Potnis N and Bull CT (2022) Comparative Genomic Analysis of the Lettuce Bacterial Leaf Spot Pathogen, Xanthomonas hortorum pv. vitians, to Investigate Race Specificity. Front. Microbiol. 13:840311. doi: 10.3389/fmicb.2022.840311

Bacterial leaf spot (BLS) of lettuce caused by *Xanthomonas hortorum* pv. *vitians* (*Xhv*) was first described over 100 years ago and remains a significant threat to lettuce cultivation today. This study investigated the genetic relatedness of the *Xhv* strains and the possible genetic sources of this race-specific pathogenicity. Whole genome sequences of eighteen *Xhv* strains representing the three races, along with eight related *Xanthomonas* strains, were included in the analysis. A maximum likelihood phylogeny based on concatenated whole genome SNPs confirmed previous results describing two major lineages of *Xhv* strains. Gene clusters encoding secretion systems, secondary metabolites, and bacteriocins were assessed to identify putative virulence factors that distinguish the *Xhv* races. Genome sequences were mined for effector genes, which have been shown to be involved in race specificity in other systems. Two effectors identified in this study, *xopAQ* and the novel variant *xopAF2*, were revealed as possible mediators of a gene-for-gene interaction between *Xhv* race 1 and 3 strains and wild lettuce *Lactuca serriola* ARM-09-161-10-1. Transposase sequence identified downstream of *xopAF2* and prophage sequence found nearby within *Xhv* race 1 and 3 insertion sequences suggest that this gene may have been acquired through phage-mediated gene transfer. No other factors were identified from these analyses that distinguish the *Xhv* races.

Keywords: bacterial plant pathogens, plant-microbe interactions, comparative genomics, *Xanthomonas*, effectors

INTRODUCTION

Bacterial leaf spot (BLS) of lettuce is caused by *Xanthomonas hortorum* pv. *vitians* (*Xhv*) Morinière et al. (2020) (formally *X. campestris* pv. *vitians*), which was first described in 1918 following an outbreak in South Carolina lettuce fields (Brown, 1918). BLS is a sporadic, yet significant threat to lettuce production worldwide, and several subsequent outbreaks in the last century have

prompted research to better understand this plant-pathogen interaction. Investigation into *Xhv* strain diversity has been vital to the development and application of effective strategies for the management of BLS. In 2015, Bull et al. (2015) demonstrated that some, but not all, of the 120 *Xhv* strains evaluated resulted in a hypersensitive response (HR) to limit the spread of infection upon inoculation into *Lactuca sativa* cvs. Little Gem, Pavane, and La Brillante, and that these strains did not cause HR in the other lettuce cultivars that were tested. This subset of *Xhv* strains was designated race 1 to correspond to the HR elicited in Little Gem, Pavane, and La Brillante and the hypothesized interaction with the resistance gene (R-gene) *Xar1* mapped to chromosome 2 in these cultivars (Hayes et al., 2014). Another identified source of resistance to *Xhv* race 1 strains isolated from Florida lettuce fields was the *Xcv1* R-gene from *L. serriola* PI358001-1 (Wang et al., 2016). The presence of these R-genes led to the hypothesis that *Xhv* race 1 strains engage in a gene-for-gene interaction with these lettuce cultivars. Additional races of the pathogen were designated among those *Xhv* strains that did not result in HR in the *Xar1*- or *Xcv1*-enconding cultivars, but instead triggered HR in either *L. serriola* PI491114 (designated race 2) or *L. serriola* ARM-09-161-10-1 (designated race 3). Recent HR screening also suggests that *Xhv* race 1 strains might also induce HR in ARM-09-161-10-1 (Rosenthal et al., unpublished).

Bull et al. (2016) found that the diversity they observed in the resistance phenotypes of the lettuce cultivars corresponded to the genetic diversity of the pathogen. They demonstrated the presence of five distinct sequetypes among 120 *Xhv* strains using a multi-locus sequence analysis (MLSA) scheme originally designed for distinguishing *Xanthomonas* spp. (Maiden et al., 1998; Young et al., 2008). The results of this study corroborated previous work on *Xhv* diversity (Sahin et al., 2003; Fayette et al., 2016), and they found that the strains that induce HR upon injection into the *Xar1*- or *Xcv1*-encoding cultivars all belonged to sequetypes B, D, or E. The strains which triggered HR in PI491114 all belonged to sequetype A, and those that triggered HR in ARM-09-161-10-1 all belonged to sequetype C. These results supported the hypothesis that genotypic differences underly the differences in resistance phenotype. The relationship between sequetype and race also supported the finding that *Xhv* populations are predominantly clonal, though with some sequence variation due to recombination (Fayette et al., 2016).

Studies of plant-pathogen interactions have suggested a constant cycle in which plants and their pathogens evolve to outperform each other, in defense strategy and pathogenicity, respectively (Dangl and Jones, 2001; Zipfel and Felix, 2005). Pathogen presence can trigger basal plant defenses to prevent further colonization, which can be then interrupted by the delivery of bacterial effector proteins through the type three-secretion system. In turn, effector activity can be thwarted by plant host recognition of those effectors and activation of resistance (R-)genes, such as those that activate HR to kill off invaded tissues and limit the spread of infection. The presence of the R-genes *Xar1* and *Xcv1* in lettuce cultivars capable of HR exclusively to *Xhv* race 1 strains (Hayes et al., 2014; Wang et al., 2016) suggested the possibility that these strains might produce an effector that is recognized by those R-genes. Other gene-for-gene interactions might also be at play for *Xhv* races 2 and 3 and their respective HR-inducible lettuce cultivars, though R-genes have not yet been demonstrated for PI491114 and ARM-09-161-10-1.

In other bacterial plant pathogens, the proximity of many type III effectors (T3Es) to mobile genetic elements suggests that they were acquired or lost through horizontal gene transfer with other bacteria in their local microbiome (McCann and Guttman, 2008; Merda et al., 2017). Following effector gene acquisition or loss, the selective environment created by the host's defenses may have caused the new genotype to be maintained and to be host-specific (Sarkar et al., 2006; Hajri et al., 2009) or even race-specific. Such a relationship between race-specific genotype and a tomato cultivar has already been demonstrated for *X. euvesicatoria* pv. *perforans* (Astua-Monge et al., 2000).

Other gene products, such as secondary metabolites and bacteriocins, have been shown to play a role in virulence. The *Xanthomonas* pigment Xanthomonadin and siderophore Xanthoferrin contribute to bacterial fitness by protecting bacterial cells from UV damage and sequestering iron in low-iron conditions, respectively; functional knockouts of these gene clusters resulted in reduced virulence according to Rajagopal et al. (1997) and Pandey et al. (2017). A variety of bacteriocins identified in *Xanthomonas* spp. enhance bacterial competitiveness through antimicrobial activity (Holtsmark et al., 2008). Although a role in host-specific pathogenicity for these components has not yet been demonstrated, it is possible that race-specific inclusion of these genes may be used to further distinguish the *Xhv* races.

This study was designed to investigate variation in effector repertoires between the *Xhv* races as a possible genetic source for race-specific HR elicitation. Several important virulence factors were also studied to look for genotypic differences that could be used to distinguish the *Xhv* races. Eighteen geographically diverse isolates of *Xhv* and seven related plant pathogenic *Xanthomonas* strains were sequenced and assembled. A genome assembly for *X. hortorum* pv. *carotae* M081 was also retrieved and included for comparison. Genome statistics and predicted features, including plasmids and secretions systems, were determined for each strain. A maximum likelihood phylogeny based on SNP data corroborated the taxonomic relationships predicted by MLSA (Bull et al., 2016; Fayette et al., 2016). Mining the genomes for effector sequences and other potential virulence factors revealed insights into the race-specificity of the BLS pathogen. A preliminary report of this data has been given (Rosenthal et al., 2018).

METHODS

Bacterial Strains and Culturing Methods

All strains used in this study and their sources are shown in **Table 1**. Eighteen strains of *Xhv* were selected as representatives of the three pathogenic races designated in previous studies: race 1 (12 strains), race 2 (4 strains), and race 3 (2 strains).

The type and pathotype strains of *X. hortorum* pv. *hederae, X. hortorum* pv. *taraxaci, X. hortorum* pv. *pelargonii, X. hortorum* pv. *gardneri*, *and X. hortorum* pv. *cynarae* were included for comparison to the *Xhv* strains, as well as one *Xanthomonas hortorum* strain from radicchio. *X. campestris* pv. *coriandri* was included as an outgroup. Bacteria were routinely cultured on nutrient agar (NA) and in nutrient broth (NB), both at room temperature (20–28°C).

DNA Extraction and Whole Genome Sequencing

DNA extraction was performed using the Qiagen DNeasy UltraClean Microbial Kit (cat. no. 12224; Valencia, CA, United States) according to the manufacturer's instructions, with the following change: five replicates were completed for each strain and after eluting each first replicate, that elute was used to elute the next replicate, and so on. This resulted in >2.0 μg of total DNA per strain, as measured using the Thermo Fischer Qubit Fluorometer 3.0 and the Invitrogen dsDNA Broad Range Assay Kit (cat nos. Q33216 and Q32850; Waltham, MA, United States). Library preparation and 250 × 250 paired-end whole genome sequencing was performed at the Pennsylvania State University Genomics Facility using the TruSeq DNA PCR Free Library Construction Kit and the Illumina MiSeq System (cat. nos. FC-121-3001 and SY-410-1003; San Diego, CA, United States).

Quality Control and Whole Genome and Plasmid Assembly

Trimmomatic PE v.0.39 (Bolger et al., 2014) was used to remove Illumina adapters and poor-quality sequence, defined as four consecutive bases with phred scores lower than 20 and DNA fragments shorter than 50 bases. Sequence quality reports were generated using FastQC v0.11.9 (Andrews, 2010) before and after using Trimmomatic PE to verify an improvement in sequence quality which would improve the accuracy of the genome assembly. This procedure was completed in a loop using trimming.sh, spades.sh, and spades_loop.txt (**Supplementary Data 1–3**).

Trimmed reads were assembled *de novo* using SPAdes v3.12.0 (Bankevich et al., 2012) and Quast v5.0.0 (Gurevich et al., 2013) was used to provide descriptive statistics about the assemblies (**Table 2**). Plasmid counts were predicted using the Recycler pipeline (Rozov et al., 2017) and plasmid sequences were predicted using plasmidSPAdes (Antipov et al., 2019); using both programs for plasmid analysis has been recommended by the creators of plasmidSPAdes due to the differing predictive strengths of each tool. Raw reads and whole genome sequence assemblies are available at BioProject PRJNA790934. Completeness of the whole genome assemblies was assessed using BUSCO v5.2.2 (lineage = xanthomonadales_odb10 and mode = genome; Manni et al., 2021) to compare these assemblies against several publicly available genome assemblies for related *Xanthomonas* strains obtained from NCBI's genome database, including *X. campestris* pv. *campestris* (ATCC 33913^{T}), *X. populi* (CFBP 1817^{T}), *X. arboricola* pv. *juglandis* (CFBP 2528^{T}), *X. hortorum* pv. *pelargonii* (CFBP 2533^{PT}), *X. hortorum* pv. *taraxaci* (CFBP 410^{PT}), *X. hortorum* pv. *cynarae* (CFBP 4188^{PT}), *X. hortorum* pv. *carotae* (CFBP 7900), *X. hortorum* pv. *gardneri* (CFBP 8163^{PT}), *X. hortorum* pv. *vitians* (CFBP 8686^{PT}), *X. campestris* pv. *coriandri* (ICMP 5725^{PT}), *X. fragariae* (PD 885^{T}), and *X. hortorum* pv. *hederae* (WHRI 7744^{T}).

Variant-Based Phylogeny Using Whole Genome Sequences

To investigate the genetic relatedness of the *Xhv* strains and related strains used in this study, a phylogenetic tree was constructed based on a core alignment of single nucleotide polymorphisms (SNPs). The whole genome assembly of *Xanthomonas hortorum* pv. *carotae* (*Xhc*) M081 was downloaded from NCBI (GenBank assembly accession GCA_000505565.1) to be used for comparison. The snippy-multi pipeline v4.3.6 (Seemann, 2015) was used to map trimmed Illumina sequencing reads to the *Xhc* M081 reference and to identify and align a set of core SNPs. In the resulting core alignment, all non-[AGTCN-] characters were replaced using snippy-clean_full_aln and then loci with high levels of base substitution, indicating possible recombination, were removed using run_gubbins.py v2.4.1 (Croucher et al., 2015). SNPs were extracted from the filtered multi-fasta alignment using SNP-sites with the option "-c" to only output AGTC into a final alignment file. This file was imported into CLC Genomics Workbench and General Time Reversible with rate variation (four categories) and estimated topology was selected as the appropriate substitution model following model testing with the neighbor-joining clustering method. This model was used to produce a maximum-likelihood phylogeny with 1,000 bootstraps and branch lengths that represent the expected number of nucleotide substitutions per sequence site.

Effector Gene Mining

BLAST databases were constructed from each whole genome assembly scaffold file or plasmid scaffold file. A FASTA file containing known bacterial effector protein sequences, including those previously published for plant pathogenic *Xanthomonas* and *Pseudomonas* species, was provided (White et al., 2009). Translated nucleotide BLAST was used to align each entry of this effector sequence catalog to the whole genome sequence databases. Matches were filtered to include those with greater than 60% identity, greater than 40% query coverage per high scoring sequence pair, and less than 0.0001 expect value. Other potential matches were found using a filter of greater than 45% identity, greater than 40% qcovhsp, and less than 0.00001 e-value; this allowed for inclusion of some other effectors which may be present but with moderate sequence variation compared to those in our database. The same searches were performed using the predicted plasmid sequences to determine which effectors may be carried on plasmids. The filtered matches constitute the hypothesized effector repertoires of that strain. Mining for these effector repertoires was completed using the script effectorgene_mining.sh (**Supplementary Data 4**).

TABLE 1 | *Xanthomonas* strains included in this study.

Organism	Strain designations	Other strain IDs	Geographic origin	Host of isolation (common name)	Race*	Source or citation
X. hortorum pv. *vitians*	BP5172	*Xav* 98-37 2/01	Salinas, CA, United States	*L. sativa* (lettuce)	1	J. Barak
	BS0339	Salinas 2/01	Salinas, CA, United States	*L. sativa* (lettuce)	1	J. Barak
	BS0340	*Xav* 98-23 2/01	Salinas, CA, United States	*L. sativa* (lettuce)	1	J. Barak
	BS0347	*Xcv* 5/01	Salinas, CA, United States	*L. sativa* (lettuce)	1	J. Barak
	BP5176	*Xcv* 5/01	Salinas, CA, United States	*L. sativa* (lettuce)	1	J. Barak
	BP5177	"Edge A"	Colorado, United States	*L. sativa* (lettuce)	1	S. Koike
	BP5179	"Daniel Rom"	Salinas, CA, United States	*L. sativa* (lettuce)	1	S. Koike
	BP5182	"Moreno Let"	Santa Maria, CA, United States	*L. sativa* (lettuce)	1	S. Koike
	NCPPB 4058	N/A	United Kingdom	*L. sativa* (lettuce)	1	H. Stanford
	CFBP 8686^{PT}	LMG 938^{PT}, NCPPB 2248^{PT}, MR20213^{PT}	Zimbabwe	*L. sativa* (lettuce)	1	Vauterin et al., 1995; Morinière et al., 2020
	BP5191	VT111	Canada	*L. sativa* (lettuce)	1	V. Toussaint
	BP5192	*Xcv*-2	CA, United States	*L. sativa* (lettuce)	1	C. T. Bull
	ICMP 1408	PDDCC 1408	Ithaca, NY, United States	*L. sativa* (lettuce)	2	W. H. Burkholder
	ICMP 4165	LMG 7508, PDDCC 4165	New Zealand	*L. sativa* (lettuce)	2	H. J. Boesewinkel
	BS3127	VT106	Canada	*L. sativa* (lettuce)	2	V. Toussaint
	BP5194	917	Ohio	*L. sativa* (lettuce)	2	Sahin et al., 2003
	BS2861	"Christy BuLet 2"	King City, CA, United States	*L. sativa* (lettuce)	3	S. Koike, Rianda
	BP5181	"Christy BuLet 3"	King City, CA, United States	*L. sativa* (lettuce)	3	S. Koike, Rianda
X. hortorum from radicchio	BP5178	N/A	Salinas, CA, United States	*Cichorium intybus* (radicchio)		Zacaroni et al., 2012
X. hortorum pv. *hederae*	CFBP 4925^{T}	ICMP 453^{T}, NCPPB 939^{T}, LMG 733^{T}	United States	*Hedera helix* (English ivy)		Arnaud, 1920; Dye, 1978; Vauterin et al., 1995
X. hortorum pv. *taraxaci*	CFBP 410^{PT}	ATCC 19318^{PT}, NCPPB 940^{PT}, LMG 870^{PT}	Ithaca, NY, United States	*Taraxacum kok-saghyz* (Russian dandelion)		Niederhauser, 1943; Dye, 1978; Vauterin et al., 1995
X. hortorum pv. *pelargonii*	CFBP 2533^{PT}	ICMP 4321^{PT}, LMG 7314^{PT}, NCPPB 2985^{PT}	Auckland, New Zealand	*Pelargonium peltatum L. L'Hér.* (pelargonium)		Brown, 1923; Dye, 1978
X. hortorum pv. *gardneri*	CFBP 8163^{PT}	LMG 962^{PT}, ATCC19865^{PT}, NCPPB 881^{PT}, PDCC 1620^{PT}	Yugoslavia	*Lycopersicon esculentum* Mill. (tomato)		(ex Šutić 1957) Jones and Dangl, 2006
X. hortorum pv. *cynarae*	CFBP 4188^{PT}	ICMP 16775^{PT}	France	*Cynara scolymus* (artichoke)		Trébaol et al., 2000.
X. hortorum pv. *carotae*	CFBP 7900	M081	Hungary	*Daucus carota* L. var. sativus (carrot)		Kendrick, 1934; Dye, 1978; Vauterin et al., 1995
X. campestris pv. *coriandri*	CFBP 8452^{PT}	LMG 687^{PT}, ATCC 17996^{PT}, ICMP 5725^{PT}, NCPPB 1758^{PT}, PDDCC 5725^{PT}	India	*Coriandrum sativum* (coriander)		Srinivasan et al., 1961; Dye, 1978; Vauterin et al., 1995

Type and pathotype strains are marked T and PT, respectively, in the strain designation column.
**There are currently 10 Xhv race 2 strains and four Xhv race 3 strains described.*

TABLE 2 | Genome characteristics for *Xanthomonas hortorum* pv. *vitians* strains and related type and pathotype strains.

	Strain	Genome size (Mb)	N50	GC%	Contig no.	CDSs no.	CDSs Length (Mb)	% Genome CDS	Plasmid no.	ISs compared to BS3127
X. hortorum pv. *vitians* race 1	BP5172	5.1	90636	63.7	128	4590	4.6	91.1	2	97
	BS0339	5.2	45509	63.6	223	4747	4.7	90.8	1	93
	BS0340	5.4	82765	63.4	324	4653	4.7	86.7	0	105
	BS0347	5.1	114952	63.7	114	4646	4.7	91.0	0	91
	BP5176	5.1	40310	63.7	264	4644	4.7	91.4	3	72
	BP5177	5.2	89202	63.7	138	4674	4.7	90.8	1	104
	BP5179	5.2	26996	63.7	364	4702	4.7	91.0	4	87
	BP5182	5.2	55827	63.7	179	4712	4.7	90.9	2	94
	NCPPB 4058	5.3	80295	63.6	130	4793	4.8	90.6	0	95
	CFBP 8686^{PT}	5.1	53428	63.8	183	4591	4.6	91.4	2	88
	BP5191	5.3	49126	63.5	220	4848	4.8	90.8	1	64
	BP5192	5.1	106975	63.7	130	4665	4.7	91.0	0	100
X. hortorum pv. *vitians* race 2	ICMP 1408	5.2	107536	63.7	109	4720	4.7	91.3	0	0
	ICMP 4165	5.3	129151	63.6	107	4865	4.9	91.2	1	0
	BS3127	5.1	106407	63.7	98	4647	4.7	91.6	0	0
	BP5194	5.2	33155	63.6	271	4725	4.7	91.5	0	0
X. hortorum pv. *vitians* race 3	BS2861	5.2	118085	63.7	120	4690	4.7	90.9	0	no data
	BP5181	5.2	98381	63.7	128	4684	4.7	91.0	1	111
X. hortorum from radicchio	BP5178	5.2	58212	63.7	199	4655	4.7	91.1	1	no data
X. hortorum pv. *hederae*	CFBP 4925^{T}	5.4	43552	63.8	270	4916	5.0	91.7	1	no data
X. hortorum pv. *taraxaci*	CFBP 410^{PT}	5.0	53166	63.9	173	4535	4.6	92.5	1	no data
X. hortorum pv. *pelargonii*	CFBP 2533^{PT}	5.2	128949	63.8	83	4788	4.9	92.9	1	no data
X. hortorum pv. *gardneri*	CFBP 8163^{PT}	5.2	112448	63.7	95	4707	4.7	91.5	1	no data
X. hortorum pv. *cynarae*	CFBP 4188^{PT}	5.4	129044	63.4	335	4633	4.7	87.3	1	no data
X. campestris pv. *coriandri*	CFBP 8452^{PT}	5.1	422303	64.9	38	4294	4.4	87.1	0	no data

Gene Alignments and Trees

Nucleotide sequences for two genes of interest, *xopAQ* and *xopAF*, were extracted from our assembly files using the program samtools v1.9 faidx and the gene positions found from the effector gene mining procedure (Li et al., 2009). These sequences were used as input for the ORFfinder program available online at https://www.ncbi.nlm.nih.gov/orffinder/, and the longest ORFs were selected for protein alignment using MEGA11: Molecular Evolutionary Genetics Analysis (Tamura et al., 2021), along with the original protein sequences published for the homologs XopAQ, HopAQ1, Rip6, and Rip11 (Guttman et al., 2002; Mukaihara et al., 2010; Potnis et al., 2011); or XopAF (AvrXv3) and HopAF1 (Astua-Monge et al., 2000; Petnicki-Ocwieja et al., 2002). Overhanging sequences were trimmed. Maximum-likelihood phylogenies were generated from these alignments to show the relatedness of the extracted genes. The Jones-Taylor-Thornton (JTT) amino acid substitution model, an assumption of uniform substitution rates among sites, and the nearest-neighbor-interchange (NNI) method were used to generate the phylogenies. The phylogenies were tested using 1,000 bootstraps and branch lengths represented the number of amino acid substitutions per site.

Mobile Genetic Elements

Mobile genetic elements were assessed for their effector content, especially for *xopAQ* and *xopAF* presence, to find evidence of recent acquisition by a common ancestor of *Xhv* race 1 and 3 strains. Large insertion sequences in *Xhv* race 1 and 3 strains were identified by aligning their whole genome sequencing reads to the assembled sequence for *Xhv* race 2 strain BS3127, identifying the sequence regions present in their genomes that were not present in BS3127, and aligning the regions flanking the putative insertion to their whole genome assemblies to find the complete

insertion sequences. These steps were completed following the MGEfinder pipeline (Durrant et al., 2020). *Xhv* race 3 strain BS2861 was excluded from this analysis because its inclusion caused the program to fail. The identified insertion sequences were then mined for effector genes using the method described above, as well as for transposase sequences included in the ISEScan package (Xie and Tang, 2017) using the same method and phage elements using PHASTER's online platform (Arndt et al., 2016) with the default settings.

xopAF and *xopAQ* Gene Regions in *Xhv* Race 2 Strains

Whether there were other known genes present at the sites in *Xhv* race 2 strains where *xopAF* and *xopAQ* were absent was analyzed. Sequences flanking the *xopAF* and *xopAQ* genes from BP5172 (*Xhv* race 1) were extracted using samtools faidx, including 1,000 nucleotides on either side. *Xhv* race 2 strains were then parsed for these sequences using BLAST to identify the same regions of their genome sequence. Alignments of these regions were completed for a subset of *Xhv* race 1 and 2 strains, and both the *Xhv* race 3 strains. These alignments revealed the sequence within the *Xhv* race 2 genomes that were present instead of the *xopAF* and *xopAQ* genes. These sequences were copied into NCBI's ORFfinder to look for coding sequence and NCBI's Conserved Domain Search to determine the possible function of that coding sequence. One protein-coding gene of interest which was present in all *Xhv* race 2 strains instead of *xopAF* was used as a query in a BLAST search of *Xhv* race 1 and 3 genomes to determine whether this gene was specific to *Xhv* race 2 strains or present in all *Xhv* strains tested.

Secretion System Gene Mining

Genes associated with or encoding secretion system components were compiled into a database by downloading their sequences from representative species from NCBI (Pena et al., in preparation). Nucleotide BLAST using the script SSgene_mining.sh (**Supplementary Data 5**) revealed genes from our database which aligned to the whole genome sequence assemblies. Plasmid sequences were also queried for secretion system genes. Matches were filtered to retain only those that had 65% identity and 65% query coverage per high scoring sequence pair or higher and an e-value less than 0.00001. To confirm these results, additional global alignments of extracted secretion system gene clusters were completed in CLC Genomics Workbench, and aligned genes were again filtered for 65% identity. Reported results represent the sum of these two methods.

Secondary Metabolite/Bacteriocin Gene Mining

Secondary metabolite gene clusters were predicted for each strain using the antiSMASH v5.0 online server using the "relaxed strictness" parameter and the "KnownClusterBlast" feature (Blin et al., 2019). Individual genes within these clusters were manually filtered to include only hits with greater than 60% identity and 40% query coverage.

To verify the antiSMASH results, each whole genome assembly was also uploaded to the BAGEL4 webserver and queried against a database of known bacteriocin genes and motifs using default settings (van Heel et al., 2018). Areas of interest that were scored as significant by the program (where the sum of weight factors is greater than 175) are reported here along with their identified class.

RESULTS AND DISCUSSION

Comparative Genomic Analysis

Eighteen assembled genomes representing the three pathogenic races of *Xhv*, as well as those of seven related *Xanthomonas* strains, are available at BioProject PRJNA790934. The bioinformatics pipeline used in this study is shown in **Figure 1**. Genome-associated statistics are reported in **Table 2**. The average *Xhv* genome size was 5.2 Mbp and in this study the number of contigs ranged from 98 for the *Xhv* race 2 strain BS3127 to 364 for the *Xhv* race 1 strain BP5179, with an average of 180 contigs per strain. The N50 for *Xhv* strains varied from 27,000 to 129,000 bp. Whole genome annotation showed that each *Xhv* strain contained about 4,700 coding sequences (CDS), with an average total length of CDS per genome being 4.7 Mb and accounting for 91% of the total genome size. The number of plasmids predicted for each *Xhv* strain varied from zero to four, with no pattern by race observed.

Compared to the *Xhv* strains, the related *Xanthomonas* strains had approximately the same genome sizes, fewer gaps in the sequence with an average of 170 contigs per strain, and similar variation in N50. Annotations for these reference strains revealed similar statistics for coding sequences, with 4,650 CDSs at an average length of 4.7 Mb and representing 91% of the total genome size. A single plasmid is predicted for each of the reference strains except for the pathotype of *X. campestris* pv. *coriandri*, which did not have any predicted plasmid sequence. BUSCO analysis for measuring the completeness of all of these sequence assemblies showed that they are complete and similar to other published *Xanthomonas* genomes (**Figure 2**).

Phylogenetic Analysis

The maximum likelihood phylogeny using whole genome SNP data showed a distinct clade to which all the *Xhv* strains belong (**Figure 3**). Within this larger clade there were three groups that corresponded to the three *Xhv* races, originally distinguished by their differing host reactions on various lettuce plant introduction lines (Bull et al., 2016). *Xhv* race 1 and 2 appear to have diverged from a common ancestor that itself diverged from strains of *X. hortorum* pv. *gardneri* and *X. hortorum* pv. *cynarae*, which are pathogenic on tomato and artichoke, respectively. The *Xhv* race 3 strains appeared to be more recently evolved, stemming from *Xhv* race 1 strains. This corroborated the finding of Fayette et al. (2016) in which *Xhv* race 3 strains made up a single MLST. The relationship between the *Xhv* strains and the other strains of *Xanthomonas hortorum* corroborated results presented in Morinière et al. (2020), which were based on a

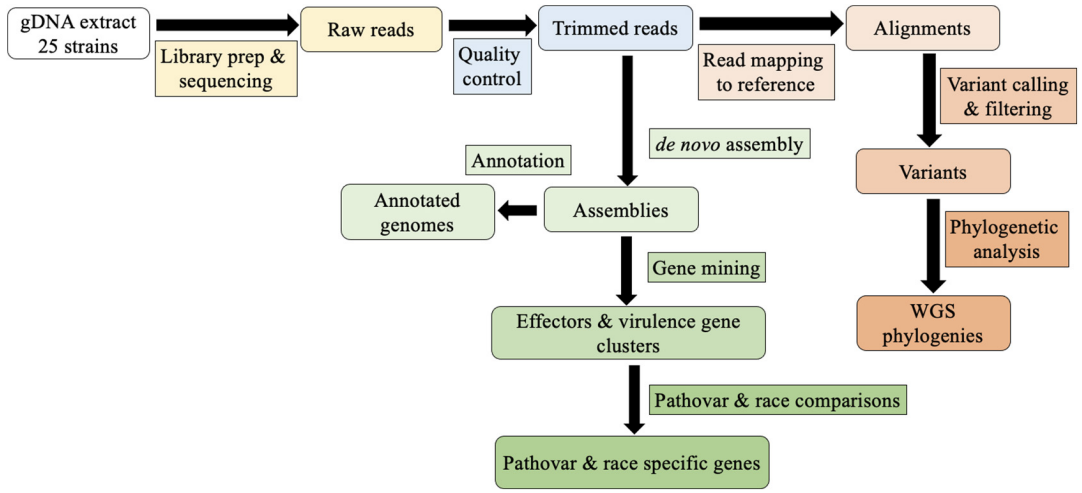


FIGURE 1 | Bioinformatics workflow for comparing whole genome sequences.

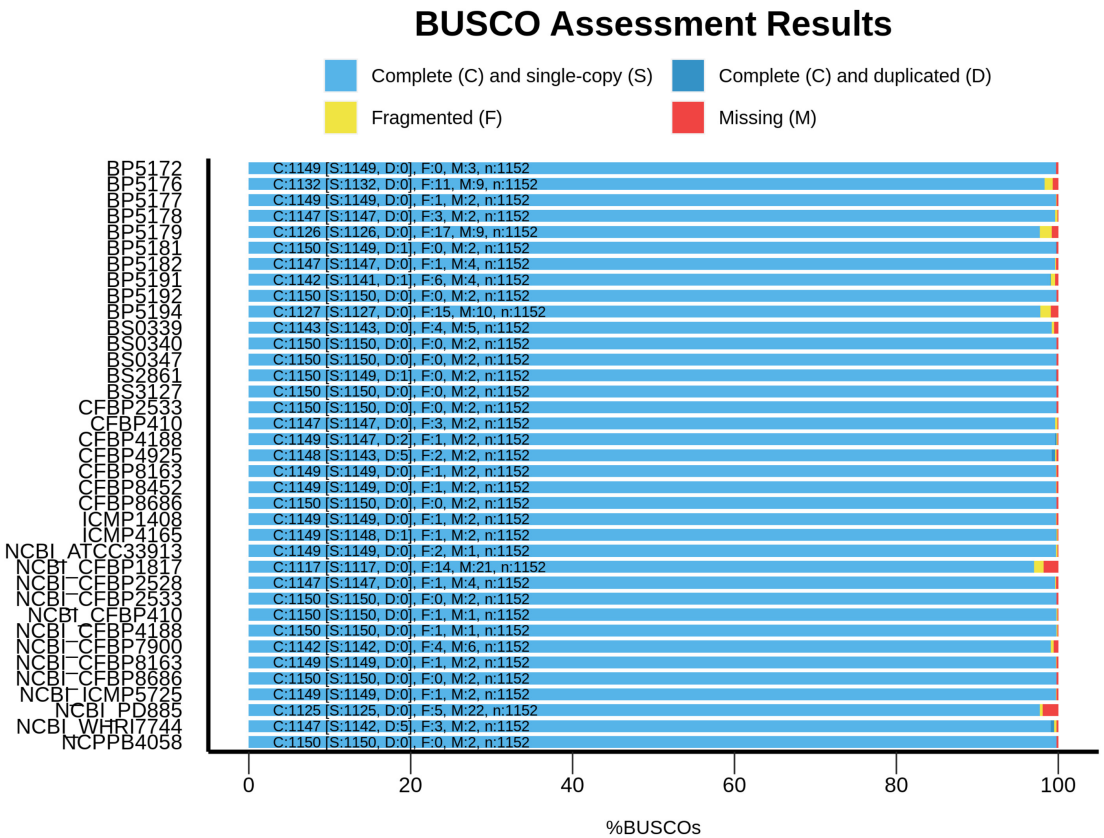


FIGURE 2 | Genome assembly completeness assessment for *Xhv* and related strains. The completeness of the genome assemblies produced in this study was evaluated using the benchmarking universal single copy orthologs method (BUSCO; Manni et al., 2021). The proportions of complete and single-copy orthologs in these assemblies were compared to those of previously published *Xanthomonas* assemblies retrieved from NCBI's genome database, including *X. campestris* pv. *campestris* (ATCC 33913^{T}; GenBank assembly accession GCA_000007145.1), *X. populi* (CFBP 1817^{T}; GCA_002940065.1), *X. arboricola* pv. *juglandis* (CFBP 2528^{T}; GCA_001013475.1), *X. hortorum* pv. *pelargonii* (CFBP 2533^{PT}; GCA_012922215.1), *X. hortorum* pv. *taraxaci* (CFBP 410^{PT}; GCA_012922225.1), *X. hortorum* pv. *cynarae* (CFBP 4188^{PT}; GCA_002939985.1), *X. hortorum* pv. *carotae* (CFBP 7900; GCA_000505565.1), *X. hortorum* pv. *gardneri* (CFBP 8163^{PT}; GCA_012922265.1), *X. hortorum* pv. *vitians* (CFBP 8686^{PT}; GCA_012922135.1), *X. campestris* pv. *coriandri* (ICMP 5725^{PT}; GCA_019201305.1), *X. fragariae* (PD 885^{T}; GCA_900183975.1), and *X. hortorum* pv. *hederae* (WHRI 7744^{T}; GCA_003064105.1).

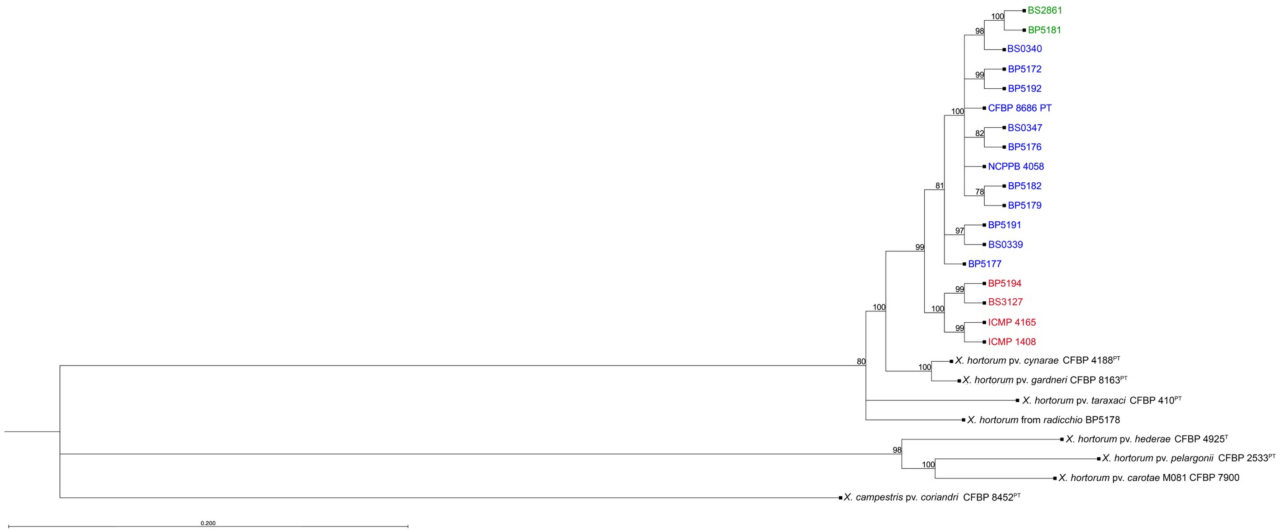


FIGURE 3 | Phylogenetic tree constructed from whole genome SNP data. Maximum-likelihood phylogeny of *Xanthomonas hortorum* pv. *vitians* strains of race 1 (blue), race 2 (red), and race 3 (green) and related *Xanthomonas* type and pathotype strains (black) used in this study. SNP data was extracted from trimmed Illumina sequence reads and core aligned using the snippy-multi pipeline (Seemann, 2015) and the tree was built from sequence alignment using CLC Genomics Workbench. This phylogeny was constructed using the neighbor-joining clustering method and the general time-reversible substitution method with rate variation (4 categories) and estimated topology. Bootstrap values are shown above branches and branch length represents the expected number of nucleotide substitutions per site. Branches shorter than 0.0100 are shown as having a length of 0.0100.

predicted core proteome rather than the whole genome SNP data presented here.

Effector Repertoire Variations

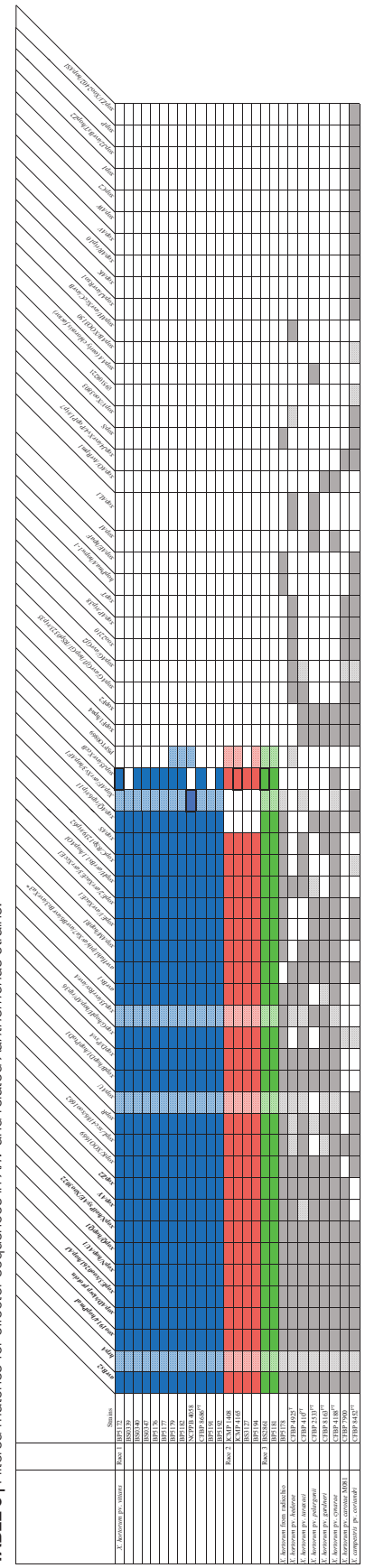
The focus of this study was to identify race-specific avirulence genes among the *Xhv* strains that could explain the race-specific HR phenotypes observed upon inoculation into various lettuce-cultivars. Of the total sixty-seven effector homologs predicted, fifteen were identified in all *X. hortorum* strain sequences tested here and constitute the core effector repertoire of the species (**Table 3**). Regarding the *Xhv* strains, thirty-four of the predicted effector homologs were present in all *Xhv* strains tested, representing the core repertoire of the pathovar. Homologs for the transcription-activator like effector (TALE) *avrHah1* were identified in all *Xhv* strains tested here, as well as in the *X. hortorum* pv. *hederae*, *X. hortorum* pv. *taraxaci*, *X. hortorum* pv. *gardneri*, *X. hortorum* pv. *cynarae*, *and X. campestris* pv. *coriandri* sequences evaluated. This class of effectors is significant for its ability to enter the host cell nucleus and alter transcription in such a way that makes the host susceptible to disease (Bogdanove et al., 2010). Although race-specific variation in *avrHah1* sequence was not observed in our strains, this could be a limitation in the ability of our short-read sequencing method to resolve the repeat sequences characteristic of TALEs. Further research is needed using a long-read sequencing platform that can demonstrate whether there are race-specific differences as well as pathovar-specific differences in the *avrHah1* gene.

Two additional effectors, *xopAF* and *xopAQ*, were predicted in the genomes of all *Xhv* strains belonging to races 1 and 3 but were absent from those of *Xhv* race 2 strains. The presence of *xopAF* was also predicted in the genomes of *X. hortorum* pv. *taraxaci*, *X. hortorum* pv. *cynarae*, and *X. campestris* pv. *coriandri* but not in the genomes of *X. hortorum* pv. *pelargonii*, *X. hortorum* pv. *gardneri*, *X. hortorum* pv. *carotae* M081, or the *X. hortorum* strain from radicchio. Further, the presence of *xopAQ* was predicted in the genomes of *X. hortorum* from radicchio, *X. hortorum* pv. *pelargonii*, *X. hortorum* pv. *cynarae*, and *X. campestris* pv. *coriandri* but not of *X. hortorum* pv. *hederae*, *X. hortorum* pv. *taraxaci,* or *X. hortorum* pv. *carotae* M081.

The gene *xopAF* was first identified as *avrXv3* in *X. euvesicatoria* pv. *perforans* strains of tomato race 3 (T3) and was demonstrated to be responsible for race-specific HR elicitation in tomato cultivar Hawaii 7981 (Astua-Monge et al., 2000). Acting intracellularly, *xopAF* causes the upregulation of defense related genes (Balaji et al., 2007). Since its original identification, homologs have been identified in strains of *X. citri* pv. *citri* (*Xcc*) isolated only from Mexican lime and *X. vasicola* pv. *vasculorum* (*Xvv*) isolated only from sugarcane. Experiments with *xopAF* mutants showed that the gene was not responsible for limiting the host range of *Xcc*, but did contribute to virulence (Jalan et al., 2013). Cultivar-specificity for *Xvv* strains does not appear to have been tested.

The other gene, *xopAQ*, was first identified as *rip6/rip11* in a screen for *Ralstonia solanacearum* proteins that are injected into host plant cells (Mukaihara et al., 2010). The *Xanthomonas* homolog *xopAQ* was first described in *X. hortorum* pv. *gardneri* and hypothesized to be part of a group of pathotype-specific genes responsible for its aggressive disease compatibility on tomato and pepper (Potnis et al., 2011). Homologs have since been found in *X. arboricola*, *X. citri*, and *X. euvesicatoria* and have

TABLE 3 | Filtered matches for effector sequences in *Xhv* and related *Xanthomonas* strains.



Bolded effectors are those predicted for all X. hortorum strains. Those effectors with an asterisk indicate that they are TALEs. Solid color indicates a match of >60% identity, >40% qcovhsp, and >0.00001 e-value; striped color lowers the percent identity filter to >45%; and no fill corresponds to absence of that effector gene. Blue color corresponds to Xhv race 1 hits, red to Xhv race 2 hits, green to Xhv race 3 hits, and gray to related Xanthomonas reference strain hits. Thick box borders indicate gene presence verified through targeted gene amplification and visualization by gel electrophoresis.

a proposed function in lipid modification that may be involved in host recognition and hypersensitive response (Thieme et al., 2007; Barak et al., 2016).

Five of the effectors predicted for all *Xhv* strains were found within their plasmid sequences: *avrBs1*, *avrHah1*, *xopE1*, *xopE2*, and *xopH* (**Supplementary Data 6**). Several other effector genes were predicted among various *Xhv* strains, including *avrBs2*, *xopL*, *xopJ5*, *xopAQ*, *xopB*, *xopC*, *xopD*, *xopG*, and *xopZ2*, but with no pattern that could underly the race-specific differences in HR elicitation in lettuce cultivars. Each of the reference strains had a different predicted repertoire of plasmid-borne effectors, and there were no plasmid-borne effectors that were predicted for all *X. hortorum* pathovars. The genes *avrBs1*, *avrHah1*, *xopH*, *xopAQ, avrBsT* (predicted only in *X. campestris* pv. *coriandri*), and *xopAO* (predicted only in *X. hortorum* pv. *gardneri* and *X. hortorum* pv. *cynarae*) are known to be plasmid-borne in other *Xanthomonas* species (Swanson et al., 1988; Minsavage et al., 1990; Schornack et al., 2008; Potnis et al., 2011; Newberry et al., 2019; Roach et al., 2019), and so were likely acquired through plasmid transfer in the strains tested here. All other predicted plasmid-borne effectors (see **Supplementary Data 6** for the complete list) are not yet demonstrated to be plasmid-borne in other species but may still have been acquired through plasmid transfer to a strain ancestral to *X. hortorum* pv. *vitians*. Further study with completed *Xhv* genomes produced using long-read sequencing may be able to address these questions regarding the acquisition of these effectors. All effectors not predicted on plasmid sequence here are assumed to be chromosomal genes.

From these data we hypothesize that the differences in effector repertoire composition between *Xhv* races 1 and 3, and *Xhv* race 2, is responsible for the race-specific elicitation of HR in lettuce. It is possible that *xopAQ* and/or *xopAF* interact with an R-gene in ARM-09-161-10-1 to trigger resistance to *Xhv* race 1 and 3 strains, and that the absence of these genes in *Xhv* race 2 strains allows them to go undetected. Further work is needed to confirm the avirulence function of these *xopAQ* and *xopAF* homologs in ARM-09-161-10-1, and that the resistance in these cultivars is governed by a single dominant R gene.

XopAQ Homolog and New XopAF2 Variant

Alignment of the XopAQ protein sequences revealed that they were present among X. *hortorum* strains (**Table 3**), it was identical to XopAQ protein sequence from *X. hortorum* pv. *gardneri*. In the gene tree, all XopAQ protein sequences from *X. hortorum* strains clustered together and were separated from the HopAQ1 protein sequences from *Pst* DC3000 (**Figure 4A**).

The XopAF protein sequence alignment showed that all *Xhv* race 1 and 3 homologs were identical except for that of NCPPB 4058, which had significant variation. Phylogenetic analysis revealed that all XopAF protein sequences from *X. hortorum* strains clustered together and were more closely related to the HopAF1 protein sequences from *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000 than the XopAF/AvrXv3 from *Xanthomonas euvesicatoria* pv. *perforans* (**Figure 4B**). It is

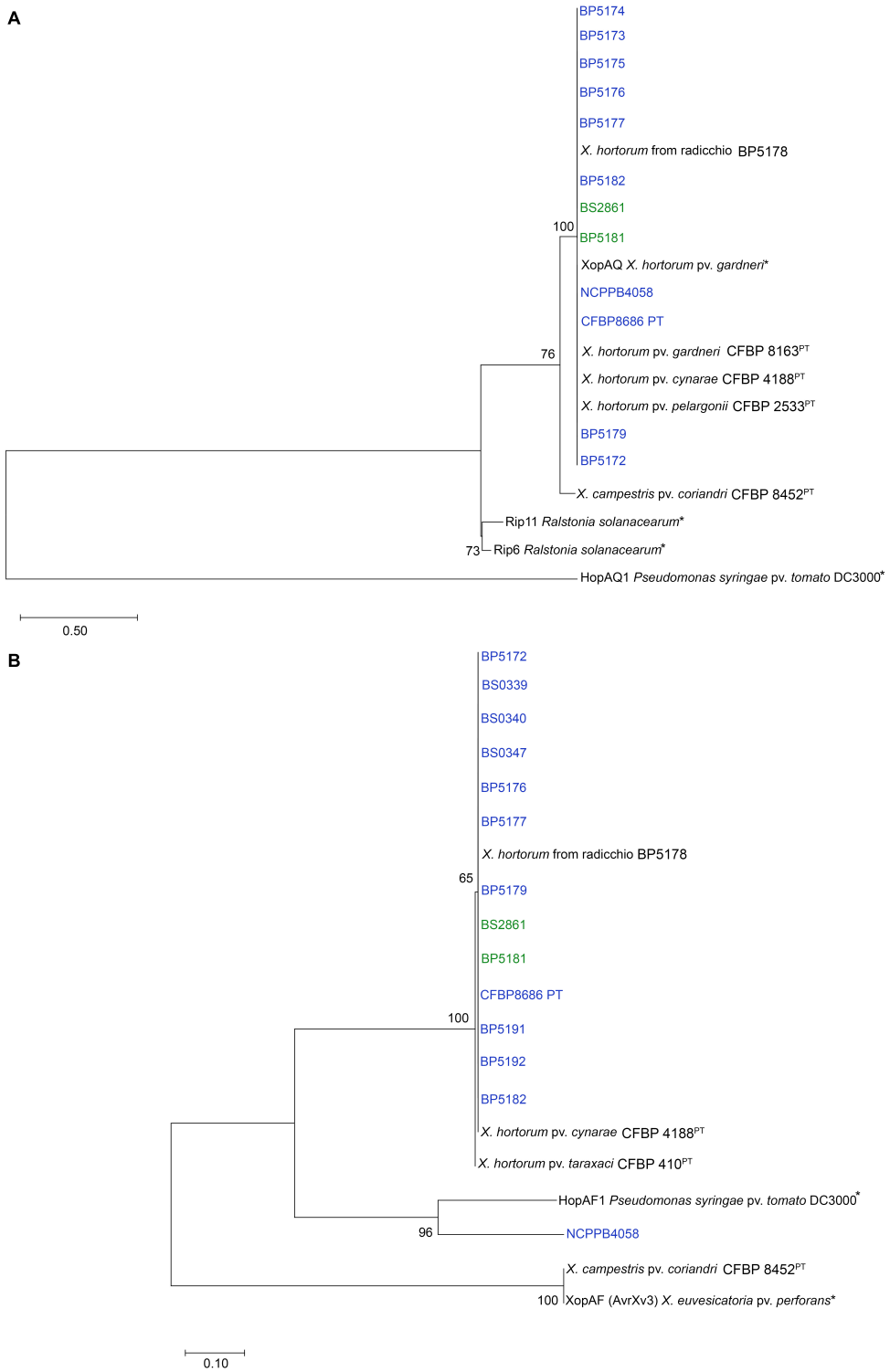


FIGURE 4 | Phylogenetic trees constructed from **(A)** XopAQ amino acid alignment and **(B)** XopAF amino acid alignment. Maximum-likelihood phylogenies were constructed for the two genes present in *Xhv* race 1 and race 3 strains (blue and green) but missing from *Xhv* race 2 strains. Nucleotide sequences for these genes were extracted from the whole genome sequence assemblies and converted to amino acid sequences using OrthoFinder. Alignments and trimming were completed using MEGA11, as well as the phylogeny building using the Jones-Taylor-Thornton (JTT) substitution model, uniform substitution rates among sites, and nearest-neighbor-interchange (NNI) method for tree inference. Bootstrap values are shown beside branches and branch length represents the expected number of amino acid substitutions per site. Asterisks indicate published sequences used as references for proteins HopAF1 and XopAF (Astua-Monge et al., 2000; Petnicki-Ocwieja et al., 2002) and XopAQ, HopAQ1, Rip6, and Rip11 (Guttman et al., 2002; Mukaihara et al., 2010; Potnis et al., 2011).

possible this effector gene was originally acquired by the ancestor of *X. hortorum* complex from a *Pseudomonas* species. Here, we designate this new effector variant XopAF2 for *X. hortorum* and propose to distinguish it from protein XopAF1 identified in *X. euvesicatoria*.

Determining the regions flanking *xopAQ* and *xopAF2* in *Xhv* race 1 and 3 strains allowed for the identification of those same regions in *Xhv* race 2 strains, and alignments revealed the gene sequence present in place of those effector genes. Where *xopAQ* would have been present, the *Xhv* race 2 strains vary in the sequence at this position: conserved domain search revealed that ICMP 1408 encoded for a DNA breaking-rejoining enzyme, while ICMP 4165 encoded for an IS3 family transposase and BS3127 did not encode any conserved domain. In contrast, where *xopAF* would have been present, alignments showed that the *Xhv* race 2 strains all had the same sequence which was predicted to encode a glycosyl hydrolase family 3 C-terminal domain. Mining the *Xhv* race 1 and 3 genomes showed that this sequence could be found elsewhere in those genomes, and so its presence alone is not likely a distinguishing feature of a particular *Xhv* race.

In comparison to *Xhv* race 2 strain BS3127, the number of insertion sequences present in *Xhv* race 1 and 3 strains ranged from 64 to 111, with an average of 92 (**Table 2**). Searching these insertion sequences for effector genes yielded no matches using the same effector gene mining method used for the whole genome sequences. Searches for the transposase genes included in the ISEScan package reveals a negative sense transposase sequence on the same node as *xopAF* and 953 bp downstream. This transposase sequence belongs to insertion sequence family IS5, and the sequence itself is 371 bp in length. Further, searching the *Xhv* race 1 and 3 insertion sequences for phage elements using PHASTER reveals two intact prophage regions. These results indicate that the effector gene *xopAF* may have been acquired by a strain ancestral to *Xhv* race 1 and 3 strains through phage-mediated gene transfer. As expected, no insertion sequences were found in other *Xhv* race 2 strains when compared to BS3127, which supports the assertion that these strains are more related to one another than to the *Xhv* race 1 and 3 strains.

Secondary Metabolites

Though the main interest in this study was to identify effector repertoires as possible genetic sources of host-specificity, several other virulence factors were investigated for genetic differences among the strains tested that could be useful for characterization of the *Xhv* races. The secondary metabolites Xanthomonadin and Xanthoferrin are encoded by gene clusters that are conserved among *Xanthomonas* spp. Xanthomonadin was first described in *X. oryzae* pv. *oryzae* (*Xoo*) and is a yellow pigment that protects bacteria from the increase in lipid-degrading free radicals caused by UV exposure (Rajagopal et al., 1997; Ayala et al., 2014). Xanthoferrin is a siderophore that enhances bacterial survival in low-iron conditions, such as the plant epidermis, due to its iron search and uptake function. Knock-out mutants for these metabolites resulted in decreased virulence for *X. campestris* pv. *campestris*, the causal agent of black rot on cabbage. For Xanthomonadin, lipid peroxidation is higher in knock-out mutants than in WT strains (Rajagopal et al., 1997). For Xanthoferrin, the knock-out results in smaller leaf lesions and less migratory capacity through the tissue (Pandey et al., 2017).

Secondary metabolite gene cluster mining of the whole genome sequences using antiSMASH revealed the possible presence of two clusters (**Table 4**). Nine to eleven genes of the 14-gene cluster encoding Xanthomonadin I, as described for *Xathomonas oryzae* pv. *oryzae* (*Xoo*), were found only in the genome sequences of *Xhv* race 1 strains BP5172 and BS0340, and those of the reference strains of *X. hortorum* from radicchio and *X. hortorum* pv. *hederae*. If functional, the production of Xanthomonadin I would afford these strains a fitness advantage in their ability to resist UV damage. The entire 7-gene cluster for Xanthoferrin production, as described for *Xoo*, was found in all genome sequences tested except for those of *Xhv* race 1 strain BP5179, the *X. hortorum* pv. *hederae* pathotype, and the *X. hortorum* pv. *taraxaci* pathotype, which were missing one gene, *XOO1360*. The strains that do produce Xanthoferrin would be better equipped to survive in low-iron conditions and may be able to spread further through host tissues than strains that do not produce Xanthoferrin. Further research is necessary to demonstrate the production and activity of these metabolites empirically.

Bacteriocins

Bacteriocins are also potential virulence factors. Three classes of bacteriocins have been proposed: class I which are heat-stable and ribosomally produced, post-translationally modified peptides (RiPPs); class II which are heat-stable and unmodified; and class III which are thermolabile and unmodified. Xanthomonin I/II, microcin, and rhodanodin are all class I bacteriocins called lasso peptides, so named for their 3-D looped structure (Hegemann et al., 2013). Sactipeptides and lanthipeptides also belong to class I, and all class I peptides have a variety of functions and may aid in bacterial fitness *via* antibacterial or antiviral activity (Maksimov et al., 2012). Zoocin A is a class III bacteriocin known as a bacteriolysin, which also serves as an antimicrobial peptide but differs from the others in its ability to degrade target bacterial cell walls (Simmonds et al., 1996).

Searches for bacteriocin gene clusters using antiSMASH revealed the presence of one bacteriocin among the strains tested (**Table 4**), and a parallel analysis using BAGEL4 produced more detailed as well as some contradictory results (**Table 5**). The entire 4-gene cluster encoding Xanthomonin I/II, as previously described for *Xanthomonas hortorum* pv. *gardneri* (ATCC 19865), was found in all strains except the *X. hortorum* pv. *hederae* pathotype which was missing half of the cluster, and *Xanthomonas campestris* pv. *coriandri*, which was missing the entire cluster. This might suggest that all strains except *X. hortorum* pv. *hederae* have an advantage in bacterial fitness due to their production of Xanthomonin I/II. However, searches with BAGEL4 revealed that five of the *Xhv* race 1, two of the *Xhv* race 2, and both *Xhv* race 3 strains encoded for the Xanthomonin I cluster, and the remaining seven *Xhv* race 1 and two of the *Xhv* race 2 strains encoded for the Xanthomonin II cluster. Among the non-*Xhv* strains, *X. hortorum* from radicchio, *X. hortorum* pv. *hederae*, *X. hortorum* pv. *taraxaci*, *X. hortorum* pv. *pelargonii*,

TABLE 4 | Predicted secondary metabolite gene clusters.

	Type	Aryl polyene	Lasso peptide	Siderophore
	Strains	**Xanthomonadin I**	**Xanthomonin I/II**	**Xanthoferrin**
X. hortorum pv. *vitians* race 1	BP5172	11/14	4/4	7/7
	BS0339	0/14	4/4	7/7
	BS0340	10/14	4/4	7/7
	BS0347	0/14	4/4	7/7
	BP5176	0/14	4/4	7/7 (split contigs)
	BP5177	0/14	4/4	7/7
	BP5179	0/14	4/4	6/7
	BP5182	0/14	4/4	7/7
	NCPPB 4058	0/14	4/4	7/7
	CFBP 8686^{PT}	0/14	4/4	7/7
	BP5191	0/14	4/4	7/7
	BP5192	0/14	4/4	7/7
X. hortorum pv. *vitians* race 2	ICMP 1408	0/14	4/4	7/7
	ICMP 4165	0/14	4/4	7/7
	BS3127	0/14	4/4	7/7
	BP5194	0/14	4/4	7/7
X. hortorum pv. *vitians* race 3	BS2861	0/14	4/4	7/7
	BP5181	0/14	4/4	7/7
X. hortorum from radicchio	BP5178	10/14	4/4	7/7
X. hortorum pv. *hederae*	CFBP 4925^{T}	10/14	2/4	6/7
X. hortorum pv. *taraxaci*	CFBP 410^{PT}	9/14	4/4	6/7
X. hortorum pv. *pelargonii*	CFBP 2533^{PT}	10/14	4/4	7/7
X. hortorum pv. *gardneri*	CFBP 8163^{PT}	11/14	4/4	7/7
X. hortorum pv. *cynarae*	CFBP 4188^{PT}	0/14	4/4	7/7
X. hortorum pv. *carotae*	CFBP 7900	0/14	4/4	7/7
X. campestris pv. *coriandri*	CFBP 8452^{PT}	10/14	0/14	7/7

Solid color indicates predicted gene presence, no fill indicates predicted gene absence. Blue color corresponds to Xhv race 1 hits, red to Xhv race 2 hits, green to Xhv race 3 hits, and gray to related Xanthomonas reference strain hits. Fractions indicate the observed number of genes over the expected number of genes for that cluster.

and *X. hortorum* pv. *cynarae* all encoded for the Xanthomonin I cluster; *X. hortorum* pv. *gardneri* and *X. hortorum* pv. *carotae* M081 encoded for the Xanthomonin II cluster. *X. campestris* pv. *coriandri* did not encode for either Xanthomonin cluster. These results suggest that all strains tested except *X. campestris* pv. *coriandri* might lack the fitness advantage afforded by Xanthomonin I/II production, however further study would be necessary to determine whether Xanthomonin I or II is expressed by these strains and confer a fitness advantage.

Several other bacteriocins were found to be encoded in the strains tested using BAGEL4. A cluster encoding for microcin production was found in 5 of the 12 *Xhv* race 1 strains, all the *Xhv* race 2 and 3 strains, and all the related *Xanthomonas* strains except for *X. hortorum* pv. *hederae* and *X. hortorum* pv. *taraxaci*. *X. hortorum* pv. *hederae*^{PT} and *X. hortorum* pv. *carotae* M081 were the only strains encoding for rhodandoin and lanthipeptide, respectively. Nine *Xhv* race 1 strains, all the *Xhv* race 2 and 3 strains, *X. hortorum* from radicchio, *X. hortorum* pv. *hederae*, and *X. hortorum* pv. *taraxaci* encoded for sactipeptide. For each strain in which they are produced, these class I bacteriocins likely enhance that strain's ability to compete for resources through antibacterial and antiviral activity. All strains tested encoded for zoocin A production, a type of class III bacteriocin known as a bacteriolysin. The presence of this cluster in all strains suggests that they have a fitness advantage due to the ability to degrade the cell walls of competitor bacteria.

Type II Secretion System

The type II, type III, type IV, and type VI bacterial secretion systems are important delivery systems for pathogenicity and virulence factors, and so the whole genome sequence data was also used to assess the possible presence and composition of these secretions system gene clusters. Type II secretion systems (T2SSs) enhance the virulence of plant pathogens by releasing cell wall degrading enzymes and toxins (Jha et al., 2005). The secretion complex is composed of an inner membrane platform, a pseudopilin extending from that platform into the periplasm, and a secretin forming a pore in the outer membrane. Its substrates are transported into the periplasm by the Sec or Tat pathways and into the path of the type II pseudopilin, which then pushes the substrate out through the outer membrane pore. There are two variations of the type II secretion system, known as Xps and Xcs; although only the former contributes to virulence, the latter has some homology that allows for it to complement the former. Both variations can be present in the same strain, or the Xps system

TABLE 5 | Predicted bacteriocin genes.

	Class:	I	I	I	I	I	I	II	III
		Lasso peptide	Lasso peptide	Lasso peptide	Lasso peptide	Sactipeptides	Lanthipeptide	Leaderless	Bacteriolysin
		Xanthomonin I	Xanthomonin II	Microcin	Rhodandoin	Sactipeptides	Lanthipeptide	Enterocin	Zoocin A
	Strains								
X. hortorum pv. *vitians* race 1	BP5172								
	BS0339								
	BS0340								
	BS0347								
	BP5176								
	BP5177								
	BP5179								
	BP5182								
	NCPPB 4058								
	CFBP 8686^{PT}								
	BP5191								
	BP5192								
X. *hortorum* pv. *vitians* race 2	ICMP 1408								
	ICMP 4165								
	BS3127								
	BP5194								
X. *hortorum* pv. *vitians* race 3	BS2861								
	BP5181								
X. hortorum from radicchio	BP5178								
X. hortorum pv. *hederae*	CFBP 4925^{T}								
X. hortorum pv. *taraxaci*	CFBP 410^{PT}								
X. hortorum pv. *pelargonii*	CFBP 2533^{PT}								
X. hortorum pv. *gardneri*	CFBP 8163^{PT}								
X. hortorum pv. *cynarae*	CFBP 4188^{PT}								
X. hortorum pv. *carotae*	CFBP 7900								
X. campestris pv. *coriandri*	CFBP 8452^{PT}								

Solid color indicates predicted gene presence, no fill indicates predicted gene absence. Blue color corresponds to Xhv race 1 hits, red to Xhv race 2 hits, green to Xhv race 3 hits, and gray to related Xanthomonas reference strain hits.

may be present alone, and they are both encoded in gene clusters (*xpsE-N*, *xpsD*; *xcsC-N*, respectively).

T2SS gene cluster mining of each of the 25 whole genome sequences included in this study revealed that all encoded for the complete Xcs-T2SS (**Figure 5A**). However, the *xcs* gene cluster has not been implicated in the virulence of a plant pathogen to date, except in its ability to complement the *xpsD-H* genes of the Xps system, and so it is unlikely to play a role in virulence for the strains in included in this study. Regarding the *xps* gene cluster, *X. hortorum* pv. *taraxaci* and *X. hortorum* pv. *gardneri* genomes encoded for its entirety, and all the other genomes encoded for the entire cluster minus the *xpsD* gene (**Figure 5A**). One exception was the *Xhv* race 1 strain BP5179 genome sequence, which lacked *xpsI* and parts of *xpsJ* and *xpsD*—all of which may be explained by gaps between nodes at their expected positions in the sequence. The incomplete or absent *xpsD* genes may be complemented by the presence of *xcsD* due to significant homology. No T2SS genes were identified within predicted plasmid sequences, suggesting that these genes are chromosomal. Overall, these results suggest that all strains studied here contained a functional Xps-T2SS that may contribute to virulence through the delivery of toxins and cell-wall degrading enzymes, though empirical study is necessary to confirm this hypothesis.

Type III Secretion System

Type III secretion systems (T3SSs) play a major role in the virulence and pathogenicity of bacterial plant pathogens (Büttner, 2016). The transmembrane complex includes a needle-like structure that extends into the extracellular space and can pierce the host plant cell wall to directly deliver effector proteins into the host cell cytoplasm. Once inside the host cell, the effectors can halt the basal immune response, tag R-proteins for degradation, disrupt protective phytohormones, or enter the nucleus and alter gene transcription—all of these are strategies to enhance virulence or enable pathogenicity. A cluster of approximately 20 genes encodes the T3SS (Rossier et al., 1999). The effector genes that they translocate are usually dispersed throughout the genome and near to mobile genetic elements, which facilitates their transfer and potential for gain or loss (Darrasse et al., 2013; Cesbron et al., 2015).

Twenty-four genes contributing to the T3SS structure and function were queried against the 25 *Xanthomonas* strains included in this study, and matches are demonstrated in **Figure 5B**. The genome sequences of the *X. hortorum* pv. *vitians* strains, *X. hortorum* from radicchio, *X. hortorum* pv. *taraxaci*, *X. hortorum* pv. *gardneri*, and *X. hortorum* pv. *cynarae* all encoded for the complete hrp gene cluster and likely have functional systems (Hueck, 1998). The *X. campestris* pv. *coriandri* genome sequence tested here lacked *hpa1* and *hrpW*. The *hpa1* gene product transports effectors across the T3SS apparatus in *X. oryzae* pv. *oryzae* (Wang et al., 2018) and the *hrpW* gene is a type III secreted protein that resembles both a harpin and pectate lyase (Charkowski et al., 1998). Either of these two genes may contribute to virulence, but they are not required for a functional T3SS; therefore, this *X. campestris* pv. *coriandri* strain likely has a functional T3SS. On the other hand, the genome sequences of *X. hortorum* pv. *hederae* and *X. hortorum* pv. *pelargonii* both lacked *hrpE*, and that of *X. hortorum* pv. *pelargonii* also lacked *hrpD6*. The *hrpE* gene encoded the type 3 pilus in *X. vesicatoria* (Weber et al., 2005) and the *hrpD6* gene was responsible for regulating the expression of other T3SS structural genes and type three secreted effectors in *X. oryzae* pv. *oryzicola* (Li et al., 2011); without these genes the T3SS is likely not functional in these strains. Further research would be necessary to demonstrate the translocation activity of these secretion systems empirically, especially for *X. hortorum* pv. *hederae* and *X. hortorum* pv. *pelargonii*, which both encode T3Es but seem to lack key T3SS structural gene sequences in their genomes. No T3SS genes were found to be encoded by predicted plasmid sequences, indicating that they are chromosomal.

Type IV Secretion System

There are two classes of type IV secretion systems (T4SSs); class A has 12 core genes (*virB1-11* and *virD4*; Tzfira and Citovsky, 2006) and class B has 5 core genes (*Dot*/*IcmCDFG*[*H*/*K*]) and up to 27 total genes (Hilbi et al., 2017; Qiu and Luo, 2017). Class A substrates have antimicrobial activity to reduce competition for resources; Class B substrates interfere with plant host cell signaling. Class B has not yet been found in *Xanthomonas* spp. The database used here for type IV secretion system gene mining only included the 10 class A core genes for which DNA sequences are available.

All genome sequences tested here encoded for incomplete T4SSs of class A (**Figure 5C**). The *Xhv* genome sequences and those of *X. hortorum* pv. *hederae*, *X. hortorum* pv. *pelargonii*, and *X. hortorum* pv. *taraxaci*, all encoded for the genes *virB3*, *virB4*, *virB8*, *virB9*, *virB10*, *virB11*, and *virD4*, and those of two *Xhv* race 1 strains also encoded for *virB6* (BP5177 and NCPPB 4058). For three of the *Xhv* race 2 strains (ICMP 4165, BS3127, and BP5194), *virD4* was encoded for on predicted plasmid sequence and all other predicted T4SS genes were predicted on chromosomal sequence (**Supplementary Data 6**). In other *Xanthomonas* species, T4SS the genes *virB1*, *virB2*, *virB4*, *virB9*, *virB10*, *virD2*, and *virD4* have been shown to be plasmid-borne, and among those all but *virD2* can be found duplicated in a larger T4SS gene cluster located on the chromosome (Potnis et al., 2011). The genomic sequence of *X. hortorum* pv. *gardneri* encoded for nearly the same T4SS composition as the *Xhv* strains but was missing *virB11* and had the additional *virB6* gene. The genome sequence of *X. hortorum* isolated from radicchio encoded for only *virD4*. *X. campestris* pv. *coriandri* genome sequence encoded for only *virB4, virB8*, *virB9*, and *virD4*. All chromosomal and plasmid sequences assessed here lacked genes that encoded the core complex and pilus of this system (Sgro et al., 2019). Without core structural genes, it is likely that none of these strains have functional T4SSs. More research is necessary to confirm this hypothesis.

Type VI Secretion System

The type VI secretion system (T6SS) in plant pathogenic bacteria has antimicrobial ability; they are used to directly

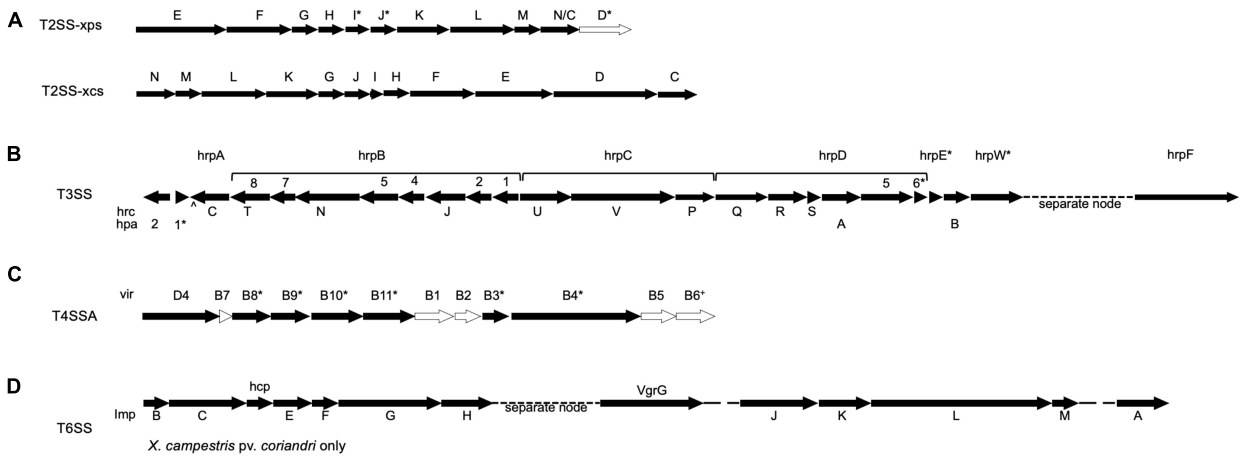


FIGURE 5 | Predicted Secretion system gene clusters for *Xhv* and related *Xanthomonas* strains. Presence of each labeled gene in the majority of *Xhv* and related strains is indicated by a black arrow, and absence from all strains is indicated by a white arrow. Asterisks, carets, and crosses indicate exceptions listed here. **(A)** T2SSs xps and xcs—exceptions include *X. hortorum* pv. *taraxaci* and *X. hortorum* pv. *gardneri*, for which the entire xps gene cluster is predicted, and *Xhv* race 1 strain BP5179, for which the genes *xpsI*, *xpsJ*, and *xpsD* were truncated; **(B)** T3SS—exceptions include *X. hortorum* pv. *hederae*, which had a large insertion between *hpa1* and *hrcC* and lacked *hrpE*, *X. campestris* pv. *coriandri*, which lacked *hpa1* and *hrpW*, and *X. hortorum* pv. *pelargonii*, which lacked *hrpE* and *hrpD6*; **(C)** T4SSA—exceptions include *Xhv* race 1 strains BS0541 and BS3044, which encoded the additional gene *virB6*, *X. hortorum* pv. *gardneri*, which encoded the additional *virB6* gene and lacked *virB11*, *X. hortorum* from radicchio, which only encoded for *virD4*, and *X. campestris* pv. *coriandri*, which only encoded for *virB4*, *virB8*, *virB9*, and *virD4*; and **(D)** T6SS—which was only predicted for the *X. campestris* pv. *coriandri* strain and not found in any other strain tested.

inject toxins into competitor bacteria cells to reduce competition for resources (Bernal et al., 2018). The contribution of T6SS in plant host interactions, however, has not yet been investigated (Liyanapathiranage et al., 2021). A cluster of thirteen type six secretion genes, *tssA-tssM*, encode the system and were formerly referred to as imp (impaired in nitrogen fixation) genes. The injection structure they form resembles a bacteriophage and toxin delivery relies on its direct contact with bacterial cell walls. The genes *tssD* and *tssI*, also known as *hcp* and *vgrG*, respectively, are thought to be both structural components and substrates of this system (Russell et al., 2014).

The only whole genome sequence included in this study that encoded for a T6SS was *X. campestris* pv. *coriandri* (**Figure 5D**). The gene cluster was split between two nodes; node 6 contained genes *tssB*, *tssC*, *tssD*, *tssE-H* and node 15 contained genes *tssI-M* and *tssA*. None of these genes were found among predicted plasmid sequences, suggesting that they are all encoded chromosomally. The presence of all core genes suggests that the T6SS is functional in this strain of *X. campestris* pv. *coriandri* and may contribute to its fitness against competitor bacteria.

Secretion system mining revealed that the *X. hortorum* strains, including all *Xhv* strains tested here, likely have functional type II and III secretion systems and do not have type IV or type VI secretion systems. *X. campestris* pv. *coriandri* also likely has functional type II, III, and VI secretion systems but likely lacks the type IV secretion system. The presence of type II and type III secretion system gene clusters suggests that these strains can deliver type II secreted toxins and enzymes and type III secreted effectors that may enhance virulence. In *X. campestris* pv. *coriandri*, the type VI secretion system may offer a fitness advantage in the ability to secrete type VI effectors that target competitor bacteria.

CONCLUSION

This study revealed insights into the race structure of *X. hortorum* pv. *vitians*. Phylogenetic analysis showed that strains of *Xhv* races 1 and 3 were more closely related to each other than to strains of *Xhv* race 2. HR induction in ARM-09-161-10-1 upon inoculation with *Xhv* race 1 and 3 strains may have been due to a gene-for-gene interaction, in which a race-specific effector is recognized by a resistance protein expressed by this cultivar. Such a race-specific interaction has been demonstrated for tomato cultivars and three races of *X. euvesicatoria* pv. *perforans* (Astua-Monge et al., 2000). The one TALE identified among the genome sequences tested, *avrHah1*, did not show race-specific variation here, but long-read sequencing to resolve the repeat regions may yet reveal such variation. Two other putative effectors identified in this study are possible candidate genes responsible for race-specificity because they were present only in *Xhv* races 1 and 3 genome sequences and not in *Xhv* race 2 genome sequences. These genes are a *xopAQ* homolog and the novel effector variant *xopAF2*. Antisense transposase sequence downstream from *xopAF2* and the prophage sequence identified among nearby *Xhv* race 1 and 3 insertion sequences suggest that an ancestor of *Xhv* race 1 and 3 strains may have acquired *xopAF2* in a phage-mediated gene transfer. Our ongoing research seeks to demonstrate whether these effectors are responsible for HR induction in ARM-09-161-10-1. Additionally, more study is necessary to identify other possible gene products involved in the incompatible interactions of *Xhv* races 1 and 2 with *Xar1*-containing lettuce cultivars,

such as Little Gem, and PI491114, respectively. Furthermore, the close relatedness of the *Xhv* race 1 and 3 strains may make it difficult to identify genotypic differences that explain their different disease phenotypes on lettuce cultivar Little Gem, as none were identified in this study.

This study also provided several insights into *Xanthomonas* genetic diversity. The genome statistics for our assemblies are consistent with what is expected for *Xanthomonas* spp. Secondary metabolite and bacteriocin production varied by strain, and no pattern was found that would be useful for characterizing the *Xhv* races. All strains tested here encoded for type II and type III secretion systems, and *X. campestris* pv. *coriandri* also encoded for a type VI secretion system. These systems all appear to be encoded chromosomally, as none of the genes were identified among predicted plasmid sequences.

Understanding the genetic variation among strains of *Xhv* is crucial for developing effective disease management strategies, especially the breeding of lettuce cultivars with durable resistance. Chemical control is most effective when applied prophylactically, but the sporadic nature of the disease can render this approach unnecessary and costly in years that the pathogen does not appear (Bull et al., 2007). A more efficient strategy is the use of resistant cultivars, especially those with multiple genetic sources of resistance against the different races of *Xhv* (Sandoya et al., 2019). Lettuce cultivars resistant to *Xhv* race 1 strains have already been found (Hayes et al., 2014; Wang et al., 2016) and additional germplasm is being sought (Sandoya et al., submitted manuscript). These resistant cultivars encode R-genes *Xar1* and *Xcv1*, which have been mapped to lettuce chromosome two, but the precise sequence has not yet been determined. The presence of these R-genes suggests a possible gene-for-gene interaction between these cultivars and *Xhv* race 1, and the race-specific interactions of *Xhv* race 2 and 3 with lettuce cultivars PI491114 and ARM-09-161-10-1, respectively, may also be due to gene-for-gene interactions with other yet unidentified R-genes. With knowledge of the specific genes interacting with lettuce hosts to induce HR, research could proceed for the precise identification of R-gene interaction partners and subsequent engineering of lettuce cultivars expressing those R-genes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

ER co-authored grants funding research, conducted the analyses, and wrote the manuscript. NP provided bioinformatic and phytobacterial expertise to direct the research. CB co-wrote grants funding research, designed research hypotheses, mentored Ph.D. candidate ER, and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by NSF GRFP Grant No. 2018265841 and USDA SCMP Grant No. 134853. This work was supported, in part, by the USDA National Institute of Food and Federal Appropriations.

ACKNOWLEDGMENTS

We thank Istvan Albert, Aswathy Sebastian, and Juan Francisco Iturralde Martinez for sharing their expertise in bioinformatics and genomics. We also thank Laura Ramos Sepulveda for her early contributions to this work and Samuel Martins for his unwavering support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.840311/full#supplementary-material

REFERENCES

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed January 22, 2019).

Antipov, D., Raiko, M., Lapidus, A., and Pevzner, P. A. (2019). Plasmid detection and assembly in genomic and metagenomic data sets. *Genome Res.* 29, 961–968. doi: 10.1101/gr.241299.118

Arnaud, M. G. (1920). Une maladie bactcriennc du lierre (I fe-dera helix L.). *C. R. Lebdomadaires Seances Acad. Sci.* 171, 121–122.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Astua-Monge, G., Minsavage, G. V., Stall, R. E., Davis, M. J., Bonas, U., and Jones, J. B. (2000). Resistance of tomato and pepper to T3 strains of *Xanthomonas campestris* pv. vesicatoria is specified by a plant-inducible avirulence gene. *Mol. Plant Microbe Interact.* 13, 911–921. doi: 10.1094/MPMI.2000.13.9.911

Ayala, A., Muñoz, M. F., and Argüelles, S. (2014). Lipid peroxidation: production, metabolism, and signaling mechanisms of malondialdehyde and 4-hydroxy-2-nonenal. *Oxid. Med. Cell. Longev.* 2014:360438. doi: 10.1155/2014/360438

Balaji, V., Gibly, A., Debbie, P., and Sessa, G. (2007). Transcriptional analysis of the tomato resistance response triggered by recognition of the *Xanthomonas* type III effector AvrXv3. *Funct. Integr. Genom.* 7, 305–316. doi: 10.1007/s10142-007-0050-y

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Barak, J. D., Vancheva, T., Lefeuvre, P., Jones, J. B., Timilsina, S., Minsavage, G. V., et al. (2016). Whole-genome sequences of *Xanthomonas euvesicatoria* strains clarify taxonomy and reveal a stepwise erosion of type 3 effectors. *Front. Plant Sci.* 7:1805. doi: 10.3389/fpls.2016.01805

Bernal, P., Llamas, M. A., and Filloux, A. (2018). Type VI secretion systems in plant-associated bacteria. *Environ. Microbiol.* 20, 1–15. doi: 10.1111/1462-2920.13956

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87. doi: 10.1093/nar/gkz310

Bogdanove, A. J., Schornack, S., and Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* 13, 394–401. doi: 10.1016/j.pbi.2010.04.010

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brown, N. A. (1918). Some bacterial diseases of lettuce. *J. Agric. Res.* 13, 367–388.

Brown, N. A. (1923). Bacterial leaf spot of geranium in the eastern United States. *J. Agric. Res.* 23, 361–372.

Bull, C., Trent, M., and Hayes, R. (2016). Three races of *Xanthomonas campestris* pv. vitians causing bacterial leaf spot on lettuce identified. *Phytopathol. S* 4:100.

Bull, C. T., Gebben, S. J., Goldman, P. H., Trent, M. A., and Hayes, R. J. (2015). Host genotype and hypersensitive reaction influence population levels of *Xanthomonas campestris* pv. vitians in lettuce. *Phytopathology* 105, 316–324. doi: 10.1094/PHYTO-07-14-0185-R

Bull, C. T., Goldman, P. H., Hayes, R., Madden, L. V., Koike, S. T., and Ryder, E. (2007). Genetic diversity of lettuce for resistance to bacterial leaf spot caused by *Xanthomonas campestris* pv. vitians. *Plant Health Prog.* 8:15. doi: 10.1094/PHP-2007-0917-02-RS

Büttner, D. (2016). Behind the lines-actions of bacterial type III effector proteins in plant cells. *FEMS Microbiol. Reviews* 40, 894–937. doi: 10.1093/femsre/fuw026

Cesbron, S., Briand, M., Essakhi, S., Gironde, S., Boureau, T., Manceau, C., et al. (2015). Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Front. Plant Sci.* 6:1126. doi: 10.3389/fpls.2015.01126

Charkowski, A. O., Alfano, J. R., Preston, G., Yuan, J., He, S. Y., and Collmer, A. (1998). The *Pseudomonas syringae* pv. tomato HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate. *J. Bacteriol.* 180, 5211–5217. doi: 10.1128/JB.180.19.5211-5217.1998

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., et al. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 43:e15. doi: 10.1093/nar/gku1196

Dangl, J. L., and Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161

Darrasse, A., Carrère, S., Barbe, V., Boureau, T., Arrieta-Ortiz, M. L., Bonneau, S., et al. (2013). Genome sequence of *Xanthomonas fuscans* subsp. fuscans strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* 14:761. doi: 10.1186/1471-2164-14-761

Durrant, M. G., Li, M. M., Siranosian, B. A., Montgomery, S. B., and Bhatt, A. S. (2020). A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. *Cell Host Microbe* 27, 140.e9–153.e9. doi: 10.1016/j.chom.2019.10.022

Dye, D. W. (1978). Genus IX *Xanthomonas*. In a proposed nomenclature and classification for plant pathogenic bacteria. *N. Z. J. Agric. Res.* 21, 153–177. doi: 10.1080/00288233.1978.10427397

Fayette, J., Raid, R., Roberts, P. D., Jones, J. B., Pernezny, K., Bull, C. T., et al. (2016). Multilocus sequence typing of strains of bacterial spot of lettuce collected in the United States. *Phytopathology* 106, 1262–1269. doi: 10.1094/PHYTO-11-15-0302-R

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Guttman, D. S., Vinatzer, B. A., Sarkar, S. F., Ranall, M. V., Kettler, G., and Greenberg, J. T. (2002). A functional screen for the type III (Hrp) secretome of the plant pathogen *Pseudomonas* syringae. *Science* 295, 1722–1726. doi: 10.1126/science.295.5560.1722

Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., et al. (2009). A "repertoire for repertoire" hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS One* 4:e6632. doi: 10.1371/journal.pone.0006632

Hayes, R. J., Trent, M. A., Truco, M. J., Antonise, R., Michelmore, R. W., and Bull, C. T. (2014). The inheritance of resistance to bacterial leaf spot of lettuce caused by *Xanthomonas campestris* pv. vitians in three lettuce cultivars. *Hortic. Res.* 66:14066. doi: 10.1038/hortres.2014.66

Hegemann, J. D., Zimmermann, M., Zhu, S., Klug, D., and Marahiel, M. A. (2013). Lasso peptides from *Proteobacteria*: genome mining employing heterologous expression and mass spectrometry. *Biopolymers* 100, 527–542. doi: 10.1002/bip.22326

Hilbi, H., Nagai, H., Kubori, T., and Roy, C. R. (2017). Subversion of host membrane dynamics by the legionella Dot/Icm type IV secretion system. *Curr. Top. Microbiol. Immunol.* 413, 221–242. doi: 10.1007/978-3-319-75241-9_9

Holtsmark, I., Eijsink, V. G., and Brurberg, M. B. (2008). Bacteriocins from plant pathogenic bacteria. *FEMS Microbiol. Lett.* 280, 1–7. doi: 10.1111/j.1574-6968.2007.01010.x

Hueck, C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* 62, 379–433. doi: 10.1128/MMBR.62.2.379-433.1998

Jalan, N., Kumar, D., Andrade, M. O., Yu, F., Jones, J. B., Graham, J. H., et al. (2013). Comparative genomic and transcriptome analyses of pathotypes of *Xanthomonas citri* subsp. citri provide insights into mechanisms of bacterial virulence and host range. *BMC Genomics* 14:551. doi: 10.1186/1471-2164-14-551

Jha, G., Rajeshwari, R., and Sonti, R. V. (2005). Bacterial type two secretion system secreted proteins: double-edged swords for plant pathogens. *Mol. Plant Microbe Interact.* 18, 891–898. doi: 10.1094/MPMI-18-0891

Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286

Kendrick, J. B. (1934). Bacterial blight of carrot. *J. Agric. Res.* 49, 493–510.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, Y. R., Zou, H. S., Che, Y. Z., Cui, Y. P., Guo, W., Zou, L. F., et al. (2011). A novel regulatory role of HrpD6 in regulating hrp-hrc-hpa genes in *Xanthomonas oryzae* pv. oryzicola. *Mol. Plant Microbe Interact.* 24, 1086–1101. doi: 10.1094/MPMI-09-10-0205

Liyanapathiranage, P., Jones, J. B., and Potnis, N. (2021). A mutation of a single core gene, tssM, of Type VI secretion system of *Xanthomonas perforans* influences virulence, epiphytic survival and transmission during pathogenesis on tomato. *Phytopathology* [Epub ahead of print]. doi: 10.1094/PHYTO-02-21-0069-R

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140

Maksimov, M. O., Pan, S. J., and James Link, A. (2012). Lasso peptides: structure, function, biosynthesis, and engineering. *Nat. Product Rep.* 29, 996–1006. doi: 10.1039/c2np20070h

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199

McCann, H. C., and Guttman, D. S. (2008). Evolution of the type III secretion system and its effectors in plant-microbe interactions. *New Phytol.* 177, 33–47. doi: 10.1111/j.1469-8137.2007.02293.x

Merda, D., Briand, M., Bosis, E., Rousseau, C., Portier, P., Barret, M., et al. (2017). Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in *Xanthomonas* plant pathogens. *Mol. Ecol.* 26, 5939–5952. doi: 10.1111/mec.14343

Minsavage, G. V., Dahlbeck, D., Whalen, M. C., Kearney, B., Bonas, U., Staskawicz, B. J., et al. (1990). Gene-for-gene relationships specifying disease resistance in *Xanthomonas campestris* pv. vesicatoria-Pepper interactions. *Mol. Plant Microbe Interact.* 3, 41–47. doi: 10.1094/MPMI-3-041

Morinière, L., Burlet, A., Rosenthal, E. R., Nesme, X., Portier, P., Bull, C. T., et al. (2020). Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that *Xanthomonas cynara* Trébol et al. 2000 emend. Timilsina et al. 2019 is a later heterotypic synonym of *Xanthomonas hortorum* Vauterin et al. (1995). *Syst. Appl. Microbiol.* 43:126087. doi: 10.1016/j.syapm.2020.126087

Mukaihara, T., Tamura, N., and Iwabuchi, M. (2010). Genome-wide identification of a large repertoire of Ralstonia solanacearum type III effector proteins by a new functional screen. *Mol. Plant Microbe Interact.* 23, 251–262. doi: 10.1094/MPMI-23-3-0251

Newberry, E. A., Bhandari, R., Minsavage, G. V., Timilsina, S., Jibrin, M. O., Kemble, J., et al. (2019). Independent evolution with the gene flux originating from multiple *Xanthomonas* species explains genomic heterogeneity in *Xanthomonas perforans*. *Appl. Environ. Microbiol.* 85:e00885-19. doi: 10.1128/AEM.00885-19

Niederhauser, J. S. (1943). A bacterial leaf spot and blight of the Russian da∼delion. *Phytopathology* 33, 959–961.

Pandey, P., Irulappan, V., Bagavathiannan, M. V., and Senthil-Kumar, M. (2017). Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Front. Plant Sci.* 8:537. doi: 10.3389/fpls.2017.00537

Petnicki-Ocwieja, T., Schneider, D. J., Tam, V. C., Chancey, S. T., Shan, L., Jamir, Y., et al. (2002). Genomewide identification of proteins secreted by the Hrp type III protein secretion system of *Pseudomonas syringae* pv. tomato DC3000. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7652–7657. doi: 10.1073/pnas.112183899

Potnis, N., Krasileva, K., Chow, V., Almeida, N. F., Patil, P. B., Ryan, R. P., et al. (2011). Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genom.* 12:146. doi: 10.1186/1471-2164-12-146

Qiu, J., and Luo, Z. Q. (2017). Legionella and Coxiella effectors: strength in diversity and activity. *Nat. Rev. Microbiol.* 15, 591–605. doi: 10.1038/nrmicro.2017.67

Rajagopal, L., Sundari, C. S., Balasubramanian, D., and Sonti, R. V. (1997). The bacterial pigment xanthomonadin offers protection against photodamage. *FEBS Lett.* 415, 125–128. doi: 10.1016/s0014-5793(97)01109-5

Roach, R., Mann, R., Gambley, C. G., Chapman, T., Shivas, R. G., and Rodoni, B. (2019). Genomic sequence analysis reveals diversity of Australian *Xanthomonas* species associated with bacterial leaf spot of tomato, capsicum and chilli. *BMC Genom.* 20:310. doi: 10.1186/s12864-019-5600-x

Rosenthal, E. R., Sebastian, A., Potnis, N., Albert, I., and Bull, C. T. (2018). *Investigating Effector Diversity as a Source of Race-Specific Pathogenicity for the Lettuce Bacterial Leaf Spot Pathogen*. Boston, MA: The American Phytopathological Society.

Rossier, O., Wengelnik, K., Hahn, K., and Bonas, U. (1999). The *Xanthomonas* Hrp type III system secretes proteins from plant and mammalian bacterial pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9368–9373. doi: 10.1073/pnas.96.16.9368

Rozov, R., Brown Kav, A., Bogumil, D., Shterzer, N., Halperin, E., Mizrahi, I., et al. (2017). Recycler: an algorithm for detecting plasmids from de novo assembly graphs. *Bioinformatics* 33, 475–482. doi: 10.1093/bioinformatics/btw651

Russell, A. B., Peterson, S. B., and Mougous, J. D. (2014). Type VI secretion system effectors: poisons with a purpose. *Nat. Rev. Microbiol.* 12, 137–148. doi: 10.1038/nrmicro3185

Sahin, F., Abbasi, P. A., Lewis Ivey, M. L., Zhang, J., and Miller, S. A. (2003). Diversity among strains of *Xanthomonas campestris* pv. vitians from lettuce. *Phytopathology* 93, 64–70. doi: 10.1094/PHYTO.2003.93.1.64

Sandoya, G. V., Maisonneuve, B., Truco, M. J., Bull, C. T., Simko, I., Trent, M., et al. (2019). Genetic analysis of resistance to bacterial leaf spot in the heirloom lettuce cultivar Reine des Glaces. *Mol. Breed.* 39:160. doi: 10.1007/s11032-019-1072-6

Sarkar, S. F., Gordon, J. S., Martin, G. B., and Guttman, D. S. (2006). Comparative genomics of host-specific virulence in *Pseudomonas syringae*. *Genetics* 174, 1041–1056. doi: 10.1534/genetics.106.060996

Schornack, S., Minsavage, G. V., Stall, R. E., Jones, J. B., and Lahaye, T. (2008). Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *New Phytol.* 179, 546–556. doi: 10.1111/j.1469-8137.2008.02487.x

Seemann, T. (2015). *Snippy: Fast Bacterial Variant Calling from NGS Reads*. Available online at: https://github.com/tseemann/snippy (accessed November 23, 2020).

Sgro, G. G., Oka, G. U., Souza, D. P., Cenens, W., Bayer-Santos, E., Matsuyama, B. Y., et al. (2019). Bacteria-killing type IV secretion systems. *Front. Microbiol.* 10:1078. doi: 10.3389/fmicb.2019.01078

Srinivasan, M. C., Patel, M. K., and Thirumalachar, M. J. (1961). A bacterial blight disease of coriander. *Proc. Indian Acad. Sci. U.S.A.* 53B, 298–301.

Simmonds, R. S., Pearson, L., Kennedy, R. C., and Tagg, J. R. (1996). Mode of action of a lysostaphin-like bacteriolytic agent produced by *Streptococcus zooepidemicus* 4881. *Appl. Environ. Microbiol.* 62, 4536–4541. doi: 10.1128/aem.62.12.4536-4541.1996

Swanson, J., Kearney, B., Dahlbeck, D., and Staskawicz, B. J. (1988). Cloned avirulence gene of *Xanthomonas campestris* pv. vesicatoria complements spontaneous race-change mutants. *Mol. Plant Microbe Interact.* 1, 5–9.

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Thieme, F., Szczesny, R., Urban, A., Kirchner, O., Hause, G., and Bonas, U. (2007). New type III effectors from *Xanthomonas campestris* pv. vesicatoria trigger plant reactions dependent on a conserved N-myristoylation motif. *Mol. Plant Microbe Interact.* 20, 1250–1261. doi: 10.1094/MPMI-20-10-1250

Trébaol, G., Gardan, L., Manceau, C., Tanguy, J. L., Tirilly, Y., and Boury, S. (2000). Genomic and phenotypic characterization of *Xanthomonas cynarae* sp. nov., a new species that causes bacterial bract spot of artichoke (*Cynara scolymus* L.). *Int. J. Syst. Evol. Microbiol.* 50, 1471–1478. doi: 10.1099/00207713-50-4-1471

Tzfira, T., and Citovsky, V. (2006). Agrobacterium-mediated genetic transformation of plants: biology and biotechnology. *Curr. Opin. Biotechnol.* 17, 147–154. doi: 10.1016/j.copbio.2006.01.009

van Heel, A. J., de Jong, A., Song, C., Viel, J. H., Kok, J., and Kuipers, O. P. (2018). BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* 46, W278–W281. doi: 10.1093/nar/gky383

Vauterin, L., Hoste, B., Kersters, K., and Swings, J. (1995). Reclassification of *Xanthomonas*. *Int. J. Syst. Biol.* 45, 472–489. doi: 10.1099/00207713-45-3-472

Wang, X., Zhang, L., Ji, H., Mo, X., Li, P., Wang, J., et al. (2018). Hpa1 is a type III translocator in *Xanthomonas oryzae* pv. oryzae. *BMC Microbiol.* 18:105. doi: 10.1186/s12866-018-1251-3

Wang, Y., Lu, H., and Hu, J. (2016). Molecular mapping of high resistance to bacterial leaf spot in lettuce PI 358001-1. *Phytopathology* 106, 1319–1325. doi: 10.1094/PHYTO-09-15-0238-R

Weber, E., Ojanen-Reuhs, T., Huguet, E., Hause, G., Romantschuk, M., Korhonen, T. K., et al. (2005). The type III-dependent Hrp pilus is required for productive interaction of *Xanthomonas campestris* pv. vesicatoria with pepper host plants. *J. Bacteriol.* 187, 2458–2468. doi: 10.1128/JB.187.7.2458-2468.2005

White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Mol. Plant Pathol.* 10, 749–766. doi: 10.1111/j.1364-3703.2009.00590.x

Xie, Z., and Tang, H. (2017). ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33, 3340–3347. doi: 10.1093/bioinformatics/btx433

Young, J. M., Park, D. C., Shearman, H. M., and Fargier, E. (2008). A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 31, 366–377. doi: 10.1016/j.syapm.2008.06.004

Zacaroni, A. B., Koike, S. T., de Souza, R. M., and Bull, C. T. (2012). Bacterial leaf spot of radicchio (*Cichorium intybus*) is caused by *Xanthomonas hortorum*. *Plant Dis.* 96:1820. doi: 10.1094/PDIS-07-12-0672-PDN

Zipfel, C., and Felix, G. (2005). Plants and animals: a different taste for microbes? *Curr. Opin. Plant Biol.* 8, 353–360. doi: 10.1016/j.pbi.2005.05.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 May 2022
doi: 10.3389/fmicb.2022.804681

Comparative Genomics of *Prunus*-Associated Members of the *Pseudomonas syringae* Species Complex Reveals Traits Supporting Co-evolution and Host Adaptation

Michela Ruinelli[1], *Jochen Blom*[2], *Theo H. M. Smits*[1] *and Joël F. Pothier*[1]*

[1]*Environmental Genomics and Systems Biology Research Group, Institute for Natural Resources Sciences, Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland,* [2]*Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany*

Members of the *Pseudomonas syringae* species complex cause symptoms that are ranging from leaf spots to cankers on a multitude of plant species, including some of the genus *Prunus*. To date, a total of two species of the *P. syringae* species complex and six different pathovars have been associated with diseases on *Prunus* spp., which were shown to belong to different phylogenetic units (phylogroups, PG) based on sequence similarity of housekeeping genes or whole genomes, suggesting that virulence to *Prunus* spp. may be the result of convergent pathoadaptation. In this study, a comparative genomics approach was used to determine genes significantly associated with strains isolated from *Prunus* spp. across a phylogeny of 97 strains belonging to the *P. syringae* species complex. Our study revealed the presence of a set of orthologous proteins which were significantly associated with strains isolated from *Prunus* spp. than in strains isolated from other hosts or from non-agricultural environments. Among them, the type III effector HopAY predicted to encode for a C58 cysteine protease was found to be highly associated with strains isolated from *Prunus* spp. and revealed patterns supporting co-evolution and host adaptation.

Keywords: *Pseudomonas syringae* species complex, comparative genomics, pathogenicity, co-evolution, host adaptation

OPEN ACCESS

Edited by:
Neha Potnis,
Auburn University, United States

Reviewed by:
Mustafa Ojonuba Jibrin,
Wenatchee Tree Fruit Research and Extension Center, United States
Vittoria Catara,
University of Catania, Italy

***Correspondence:**
Joël F. Pothier
joel.pothier@zhaw.ch

Specialty section:
This article was submitted to Microbe and Virus Interactions With Plants, a section of the journal Frontiers in Microbiology

Received: *29 October 2021*
Accepted: *07 March 2022*
Published: *03 May 2022*

Citation:
Ruinelli M, Blom J, Smits THM and Pothier JF (2022) Comparative Genomics of Prunus-Associated Members of the Pseudomonas syringae Species Complex Reveals Traits Supporting Co-evolution and Host Adaptation. Front. Microbiol. 13:804681. doi: 10.3389/fmicb.2022.804681

INTRODUCTION

Members of the *Pseudomonas syringae* species complex are responsible for the development of plant disease-causing blights, spots, specks, galls, and cankers on a wide range of economically important plant species including both herbaceous and woody hosts. Strains belonging to the *P. syringae* species complex have also been isolated from non-agricultural habitats, and therefore, their persistence and transmission is probably linked to the water cycle (Morris et al., 2008). Despite the economic and ecological importance of this bacterium, the taxonomy and nomenclature of strains belonging to the *P. syringae* species complex is quite confusing and remains largely unsettled (Palleroni, 2005; Gomila et al., 2017). Within the *P. syringae* species complex, at

least nine independent species have been determined based on phenotypical and molecular characteristics while more than 60 pathovars have been defined based on the host range (Dye et al., 1980; Palleroni, 2005; Young, 2010). DNA–DNA hybridization experiments among strains belonging to 48 different pathovars of *P. syringae* revealed the existence of nine different genomospecies (Gardan et al., 1999), which were later reflected by the so-called phylogroups (PG) obtained based on sequence similarity of housekeeping genes (Sarkar and Guttman, 2004; Hwang et al., 2005; Sarkar et al., 2006; Parkinson et al., 2011). With the inclusion of strains isolated from non-agricultural environments, a total of 13 PG were defined (Berge et al., 2014).

Many studies have been performed in the last decades with the intent to investigate and determine factors related to pathogenicity of *P. syringae* strains. The presence of the *hypersensitive reaction* and *pathogenicity* (*hrp*)/*hypersensitive reaction* and *conserved* (*hrc*) cluster was shown to be essential for pathogenicity of *P. syringae* pv. phaseolicola on bean and for triggering hypersensitive response (HR) on non-host plants, such as tobacco and tomato (Lindgren et al., 1986, 1988; Bogdanove et al., 1996). A homologous region with similar function was found also in other plant pathogens (Beer et al., 1991; Bonas et al., 1991; Bogdanove et al., 1996) and was later shown to encode for a type III secretion system (T3SS) with homology to the virulence protein secretion system (Yop) of animal-pathogenic *Yersinia* spp. (Gough et al., 1992). In *P. syringae*, the T3SS encodes for a protein apparatus which is responsible for the delivery of virulence-related factors, so-called type III effectors (T3E), into the plant cell (Wei et al., 2000). T3E generally act by promoting pathogenicity or by suppressing host immune defense but constitute a double-edge sword since T3E can also be recognized by specific plant resistance proteins which in turn trigger host immune system (Mackey et al., 2002; Shao et al., 2003; Xiang et al., 2008). However, many T3E have been shown to be functionally redundant thus decreasing the selective pressure on the host to evolve resistance proteins against single T3E (Kvitko et al., 2009). This observation suggested that the compatible interaction between *P. syringae* and its host is defined by the totality of its T3E repertoire (Lindeberg et al., 2012).

With the advent of affordable next-generation sequencing technologies, many complete and draft genome sequences of strains belonging to the *P. syringae* species complex have become available. Comparative genomics studies within different pathovars of the *P. syringae* species complex also revealed that adaptation to woody hosts was reflected by the presence of genes involved in the degradation of woody plant species-related compounds like the pentose sugar xylose and aromatic compounds, such as toluene and catechol (Green et al., 2010; Bartoli et al., 2015; Caballo-Ponce et al., 2016; Nowell et al., 2016; Hulin et al., 2020). Many studies have focused on the determination of the T3E repertoire of strains isolated from different hosts (Lindeberg et al., 2012; Ruinelli et al., 2019) and it is only recently that a few of them reported the convergent acquisition of T3E in strains adapted to the same host (Hulin et al., 2018; Newberry et al., 2019; Moreno-Pérez et al., 2020) or that strain differences in T3E alleles could be linked to host specificity (Zembek et al., 2018; Jayaraman et al., 2020). These findings underline the importance of whole genome-based comparisons to investigate factors involved in the host–pathogen interactions, which indeed are more complex than initially thought.

The plant genus *Prunus* includes economically important stone fruit trees, such as sweet cherry (*Prunus avium*), sour cherry (*Prunus cerasus*), and peach (*Prunus persica*), which in 2018 accounted for 11.6% of the total fruit orchard area in Europe (Eurostat, 2018). Even more important for the European market are almond trees (*Prunus amygdalus*) which in 2018 occupied as single species 22.6% of the total area dedicated to growing fruits (Eurostat, 2018). Bacterial canker on *Prunus* spp. caused by members of the *P. syringae* species complex affects all aboveground organs of the tree causing heavy yield reduction (up to 75%) and can lead to death of the whole tree, especially in young orchards (Crosse, 1966; Spotts et al., 2010; Hulin et al., 2020). Typical symptoms visible on trunks and branches include sunken, dark brown dieback, and cankers, which are sometimes accompanied by gummy leaks (Puławska et al., 2017). Blossom wilting and browning is mainly visible on highly susceptible varieties and constitute an important source of secondary infection. In addition, necrotic spots can be observed on leaves and on fruits which then lose their commercial values (Puławska et al., 2017). Within the *P. syringae* species complex, three different PG contain two *Pseudomonas* species and six *P. syringae* pathovars, which were found in association with diseases on *Prunus* spp.

Bacterial canker of sweet and sour cherry is mainly caused by strains belonging to *P. syringae* pv. morsprunorum race 1 and *P. syringae* pv. morsprunorum race 2 (Crosse, 1959; Crosse and Garrett, 1963; Freigoun and Crosse, 1975; Ruinelli et al., 2019). Despite being classified as races of the same pathovar, phylogenetic analysis based on sequence similarity of four housekeeping genes or of core genome of 2,085 coding sequences revealed that strains of the *P. syringae* pv. morsprunorum race 1 belong to PG3, whereas strains of the *P. syringae* pv. morsprunorum race 2 cluster within PG1 (Nowell et al., 2016; Ruinelli et al., 2019), underlying the need for clarification of the nomenclature of members of the *P. syringae* species complex.

Bacterial dieback of peach is caused by *P. syringae* pv. persicae (PG1; Young, 1987) which is also causes disease on nectarine and is weakly pathogenic to plum but not causing disease on apricot and cherry (Young, 1987). Due to its limited distribution in Europe, *P. syringae* pv. persicae was classified as quarantine organism from the European and Mediterranean Plant Protection Organization (EPPO, 2005) and as recommended regulated non-quarantine pest in the EU plant health regulation in force since December 2019 (Picard et al., 2018). Strains belonging to the *P. syringae* pv. avii (PG1) were isolated from wild cherry trees (*Prunus avium*) affected by bacterial canker in France and were shown to be only weakly pathogenic to peach, plum, and apricot (Ménard et al., 2003). *Pseudomonas amygdali* and *P. syringae* pv. cerasicola, both belonging to PG3, are the causal agents of the bacterial hyperplastic canker of almond (*P. amygdalus*;

Psallidas, 1997) and bacterial gall of ornamental cherry (*Prunus* × *yedoensis*; Kamiunten et al., 2000), respectively. A few years ago, a new species belonging to PG2, namely, *Pseudomonas cerasi* (Kałużna et al., 2016), was found to be responsible for the development of bacterial canker on cherry trees in Poland and more recently on pear tree in South Korea (Choi et al., 2020).

In addition, symptoms of bacterial canker on *Prunus* spp. are also caused by strains of *P. syringae* pv. syringae belonging to PG2 (Crosse and Garrett, 1966). However, in contrast to all above-mentioned pathovars which have been specifically found in association with plant species belonging to the genus *Prunus*, strains of *P. syringae* pv. syringae display a broader host range and are responsible for diseases on many other woody and herbaceous hosts (Cazorla et al., 1998; Garibaldi et al., 2007; Zhou et al., 2012; Popović et al., 2015; Ivanović et al., 2017).

In this study, a comparative genomics approach was used to investigate factors potentially involved in the adaptation of *P. syringae* to plant species belonging to the *Prunus* genus. Our study revealed the presence of a set of orthologous proteins, which were significantly more present in strains isolated from *Prunus* spp. than in strains isolated from other hosts or environments. Among them, the T3E HopAY, potentially encoding for a C58 cysteine protease was found to be highly associated with strains isolated from *Prunus* spp. and revealed patterns supporting co-evolution and host adaptation.

MATERIALS AND METHODS

Phylogenomics

For comparative genomics purpose, the whole genomes data of 97 strains belonging to the *P. syringae* species complex, together with one *P. fluorescens* (strain Pf0-1) and one *P. putida* (strain KT2440) were used (**Table 1**). A total of 20 genomes were complete and 79 were draft. The selected set of *P. syringae* genomes consisted of strains isolated from plants ($n = 81$) as well as strains isolated from non-agricultural environments ($n = 15$) and represents 11 of the 13 PG defined by Berge et al. (2014). Plant-associated strains were isolated from over 30 different plant species comprising *Prunus* spp. ($n = 20$), *Actinidia chinensis* ($n = 4$), *Solanum lycopersicum* ($n = 4$), *Corylus avellana* ($n = 5$), *Cucumis* spp. ($n = 5$), *Aesculus hippocastanum* ($n = 3$), *Triticum aestivum* ($n = 3$), *Hordeum vulgare* ($n = 3$), *Phaseolus vulgaris* ($n = 2$), *Olea europaea* ($n = 2$), *Glycine max* ($n = 2$), *Nicotiana* sp. ($n = 2$), *Pyrus* sp. ($n = 2$), and other herbaceous and woody hosts ($n = 22$). Non-annotated genomes retrieved from the NCBI database were annotated using a command line annotation pipeline based on HMMer against an EDGAR based database of *Pseudomonas* ortholog groups followed by reference genome annotation and a comparison to the Swiss-Prot and RefSeq databases for genes that had no high-quality hit in previous steps (Linke et al., 2011).

The core genome phylogenetic relationships were obtained using EDGAR 2.2 (Blom et al., 2016) as previously described (Ruinelli et al., 2019).

Comparative Genomics, Gene Sets Calculation, and Identification of *Prunus*-Associated Genes

Based on the core genome phylogeny, four subsets of genomes were defined (subsets A–D; **Table 1**) to be used in comparative genomics. Within each of the subsets, the sets of orthologous proteins present in *Prunus*-associated strains but absent in their phylogenetically closely related non-*Prunus*-associated strains were determined using EDGAR 2.2 (Blom et al., 2016). The protein sequences ($n = 1{,}058$) resulting for each of the subsets (subsets A–D; **Figure 1**) were used as large query against each other using standalone BLAST v.2.2.29+ (Camacho et al., 2009). All BLASTP hits having identity and coverage higher or equal to 70% were considered as ortholog and displayed in a Venn diagram. Orthologous proteins shared among each combination of subsets ($n = 52$) were checked for orthologs in the whole set of genomes ($n = 97$) using EDGAR 2.2.

Using the core genome phylogeny as a reference, associations were identified between the presence/absence of each orthologous protein in the analyzed genomes ($n = 99$) and the discrete binary trait designated "*Prunus* spp. isolate" or "other host/environment isolate" using BayesTraits v.3.0.5 (Pagel, 1994; Barker and Pagel, 2005; Pagel and Meade, 2006). The goodness of fit of the dependent versus independent model was compared with a likelihood ratio (LR) test by using a Perl script to run both models (available from https://github.com/reubwn/bayestraits-wrapper; Nowell et al., 2016). The LR test was conducted for the 52 genes that occurred in either greater than six or fewer than 92 strains, resulting in a total of 49 LRs. A null LR distribution model was constructed by randomly permuting a total of 100 times either the gene occurrence data for each of the 52 tested genes, the binary trait designation or both variables, in each case calculating a new LR statistic (Nowell et al., 2016). The null distribution was then used to derive the *p*-value thresholds. The proteins considered to be significantly more present in *Prunus*-associated strains were also used as online TBLASTN queries against nucleotide databases from 13 additional *Pseudomonas* species closely related to *P. syringae* (Mulet et al., 2010; Lalucat et al., 2020; **Supplementary Figure 4**). The presence of an ortholog in any of these 13 closely related *Pseudomonas* species was then reported when at least one TBLASTN hit having identity and coverage higher or equal to 70% was detected.

HopAY and HopAR Ortholog Retrieval and Phylogenetic Analysis

The bidirectional best hits protein orthology criteria used in EDGAR 2.2 in the previous step is mostly designed to determine the presence of a complete and probably functional ortholog protein among different genomes. However, in order, to investigate the evolution of a gene within different strains it is also important to differentiate between absence or inactivation of that gene. For this purpose, the *hopAY* reference sequence (GenBank accession number CP000059.1; locus tag: PSPPH_A0129) was derived from the T3E database (PPI, 2010) and used as online

TABLE 1 | List of strains used for this study.

Strain[1]	Code	Origin[2]	Host	GenBank accession[3]	Reference[4]	Genome subset
P. syringae pv. avii CFBP 3846^{P}	Pavii CFBP 3846	FR, 1991	*Prunus avium*	LT963402-LT963407	Ruinelli et al., 2019	A
P. syringae pv. persicae CFBP 1573^{P}	Ppe CFBP 1573	FR, 1974	*Prunus persica*	ODAL01	Ruinelli et al., 2019	A
P. syringae pv. persicae NCPPB 2254	Ppe NCPPB 2254	FR, 1969	*P. persica*	ODAM01	Ruinelli et al., 2019	A
P. syringae pv. persicae NCPPB 2254*	Ppe NCPPB 2254*	FR, 1969	*P. persica*	LAZV01	Zhao et al., 2015	–
P. amygdali pv. morsprunorum race 2 HRI W 5261	Pmp2 HRIW5261	UK, 1990	*P. avium*	LIIA01	Nowell et al., 2016	B
P. syringae pv. morsprunorum race 2 CFBP 3800	Pmp2 CFBP 3800	UK, N.D.	*Prunus cerasus*	OLMQ01	Ruinelli et al., 2019	B
P. syringae pv. morsprunorum race 2 CFBP 6411	Pmp2 CFBP 6411	UK, 1995	*P. avium*	LT963408	Ruinelli et al., 2019	B
P. amygdali pv. morsprunorum race 2 MAFF 302280^{P}	Pmp2 MAFF 302280	US, N.D.	*Prunus domestica*	AEAE01	Baltrus et al., 2011	B
P. cerasi PL58^{T}	P. cerasi PL58	PL, 2007	*P. cerasus*	LT222313-LT222319	Kałużna et al., 2016	D
P. cerasi PL963	P. cerasi PL963	PL, 2007	*P. avium*	LT963395-LT963400	Ruinelli et al., 2019	D
P. syringae pv. syringae 2339	Psy 2339	HU, 1984	*P. avium*	LIHU01	Nowell et al., 2016	–
P. syringae pv. syringae CFBP 2118	Psy CFBP 2118	FR, 1979	*P. cerasus*	LT962481	Ruinelli et al., 2019	–
P. syringae pv. syringae CFBP 4215	Psy CFBP 4215	FR. 1997	*P. avium*	LT962480	Ruinelli et al., 2019	–
P. amygdali CFBP 3205^{T}	P. amygdali CFBP 3205	GR, 1967	*Prunus amygdalus*	JYHB01	Bartoli et al., 2015	–
P. amygdali pv. morsprunorum race 1 2341	Pmp1 2341	HU, 1988	*P. cerasus*	LIIB01	Nowell et al., 2016	C
P. amygdali pv. morsprunorum FTRS U7805*	Pmp FTRSU7805*	JP, 1978	*Prunus mume*	LGLQ01	N.A.	–
P. amygdali pv. morsprunorum race 1 HRI W 5269	Pmp1 HRIW5269	UK, 1990	*P. cerasus*	LIHZ01	Nowell et al., 2016	C
P. syringae pv. morsprunorum race 1 CFBP 2116	Pmp1 CFBP 2116	FR, 1974	*P. cerasus*	LT985192-LT985195; OLMD01	Ruinelli et al., 2019	C
P. syringae pv. morsprunorum race 1 CFBP 3840	Pmp1 CFBP 3840	FR, 1996	*P. avium*	LT963409-LT963413	Ruinelli et al., 2019	C
P. syringae pv. cerasicola CFBP 6109^{P}	Pscer CFBP 6109	JP, 1995	*Prunus* × *yedoensis*	LT963391-LT963394	Ruinelli et al., 2019	C
P. syringae pv. cerasicola CFBP 6110	Pscer CFBP 6110	JP, 1995	*Prunus* × *yedoensis*	LT985210-LT985212; OLMP01	Ruinelli et al., 2019	C
P. syringae pv. actinidiae biovar 3 ICMP 18884	Psa3 ICMP 18884	NZ, 2010	*Actinidia chinensis*	CP011972-CP011973	Templeton et al., 2015	B
P. syringae pv. actinidiae biovar 2 ICMP 19073	Psa2 ICMP 19073	KR, 1998	*A. chinensis*	AOJR01	McCann et al., 2013	B
P. syringae pv. actinidiae biovar 1 ICMP 9617^{P}	Psa1 ICMP 9617	JP, 1984	*A. chinensis*	CM002753-CM002754	McCann et al., 2013	B
P. syringae pv. actinidifoliorum ICMP 18883	Pfm ICMP 18883	NZ, 2010	*A. chinensis*	AOKH01	McCann et al., 2013	B
P. syringae pv. tomato DC3000	Pto DC3000	UK, 1960	*Solanum lycopersicum*	AE016853-AE016855	Buell et al., 2003	A
P. syringae pv. tomato NCPPB 1108	Pto NCPPB 1108	UK, 1960	*S. lycopersicum*	ADGA01	N.A.	A
P. syringae pv. tomato NYS-T1	Pto NYS-T1	US, 2009	*S. lycopersicum*	JRRA01	Jones et al., 2015	A
P. syringae pv. tomato T1	Pto T1	1986	*S. lycopersicum*	ABSM01	Almeida et al., 2009	A
P. avellanae BPIC 631^{T}	Pav BPIC631	GR, 1976	*Corylus avellana*	AKBS01	O'Brien et al., 2012	B

(Continued)

TABLE 1 | Continued

Strain[1]	Code	Origin[2]	Host	GenBank accession[3]	Reference[4]	Genome subset
P. avellanae CRAFRUEC1	Pav CRAFRUEC1	IT, 2003	*C. avellana*	ATLL01	Scortichini et al., 2013	B
P. avellanae PaVt10	Pav PaVt10	IT, 2010	*C. avellana*	JYHC01	Bartoli et al., 2015	B
P. syringae pv. avellanae ISPAVE013	Psav ISPAVE013	IT, 1992	*C. avellana*	AKCJ01	O'Brien et al., 2012	–
P. syringae pv. avellanae ISPAVE037	Psav ISPAVE037	IT, 1992	*C. avellana*	AKCK01	O'Brien et al., 2012	–
P. amygdali pv. lachrymans MAFF 302278	Pla M302278	US, 1935	*Cucumis sativus*	AEAM01	Baltrus et al., 2011	A
P. syringae CC440	CC440	FR, 2002	*Cucumis melo*	AVEC02	Baltrus et al., 2014	–
P. syringae CC457	CC457	FR, 2003	*C. melo*	AVEB02	Baltrus et al., 2014	–
P. syringae CC94	CC94	FR, 1997	*C. melo*	AVEA02	Baltrus et al., 2014	–
P. amygdali pv. lachrymans MAFF 301315	Pla MAFF 301315	JP, 1975	*C. sativus*	AEAF01	Baltrus et al., 2011	C
P. amygdali pv. aesculi 2250	Pae 2250	United Kingdom	*Aesculus hippocastanus*	ACXT01	Green et al., 2010	–
P. amygdali pv. aesculi 0893_23	Pae 0893_23	IN, 1969	*A. hippocastanus*	AEAD01	Baltrus et al., 2011	–
P. amygdali pv. aesculi NCPPB 3681[P]	Pae NCPPB 3681	IN, 1980	*A. hippocastanus*	ACXS01	Green et al., 2010	–
P. syringae pv. atrofaciens DSM 50255	Paf DSM 50255	CA, 1942	*Triticum aestivum*	AWUI01	N.A.	–
P. syringae pv. syringae B64	Psy B64	N.D.	*T. aestivum*	ANZF01	Dudnik and Dudler, 2013b	–
P. syringae pv. syringae SM	Psy SM	United States	*T. aestivum*	APWT01	Dudnik and Dudler, 2013a	–
P. syringae BRIP39023	BRIP39023	AU, 1971	*Hordeum vulgare*	AMZX01	Gardiner et al., 2013	D
P. syringae BRIP34876	BRIP34876	AU, 1971	*H. vulgare*	AMXK01	Gardiner et al., 2013	–
P. syringae BRIP34881	BRIP34881	AU, 1971	*H. vulgare*	AMXL01	Gardiner et al., 2013	–
P. syringae pv. syringae B728a	Psy B728a	US, 1987	*Phaseolus vulgaris*	CP000075	Feil et al., 2005	–
P. syringae pv. phaseolicola 1448a	Pph 1448a	ET, 1985	*P. vulgaris*	CP000058-CP000060	Joardar et al., 2005	C
P. savastanoi pv. savastanoi DAPP-PG722	Psv DAPP-PG722	IT, 2007	*Olea europaea*	JOJV01	Moretti et al., 2014	–
P. savastanoi pv. savastanoi PseNe107	Psv PseNe107	NP, 2007	*O. europaea*	JYHF01	Bartoli et al., 2015	–
P. syringae CC1458	CC1458	US, 2005	*Dodecantheon pulchellum*	AVEN02	Baltrus et al., 2014	–
P. syringae CC1466	CC1466	US, 2005	*D. pulchellum*	AVEM02	Baltrus et al., 2014	–
P. savastanoi pv. glycinea B076	Pgy B076	2007	*Glycine max*	AEGG01	Qi et al., 2011	C
P. savastanoi pv. glycinea str. race 4	Pgy r4	1977	*G. max*	AEGH01	Qi et al., 2011	C
P. amygdali pv. tabaci 6605	Pta 6605	JP	*Nicotiana* sp.	AJXI01	N.A.	C
P. amygdali pv. tabaci ATCC 11528	Pta ATCC 11528	US, 1905	*Nicotiana* sp.	AEAP01	Baltrus et al., 2011	C
P. syringae pv. syringae A2	Psy A2	N.D.	*Pyrus calleryana*	LGKU01	N.A.	–
P. syringae pv. syringae B301D-R	Psy B301D-R	UK, 1969	*Pyrus communis* L.	JALJ01	Dudnik and Dudler, 2014	–
P. syringae CC1630	CC1630	US, 2007	*Onobrychis* sp.	AVED02	Baltrus et al., 2014	–
P. syringae pv. maculicola CFBP 1657[P]	Pma CFBP1657	NZ, 1965	*Brassica oleracea*	JYHH01	Bartoli et al., 2015	A
P. syringae pv. theae ICMP 3923	Pth ICMP 3923	JP, 1970	*Camellia sinensis*	LJRU01	N.A.	–
P. syringae pv. viburni ICMP 3963[P]	Pvi ICMP 3963	US, N.d.	*Viburnum* sp.	LJRR01	N.A.	–
P. syringae pv. papulans ICMP 4048[P]	Ppp ICMP 4048	CN, 1973	*Malus* × *domestica*	LJRB01	N.A.	D
P. syringae UMAF0158	UMAF0158	ES, 1993	*Mangifera* sp.	CP005970-CP005971	Martínez-García et al., 2015	D
P. syringae pv. panici LMG 2367[P]	Ppa LMG 2367	US, 1963	*Panicum* sp.	ALAC01	Liu et al., 2012	–
P. syringae pv. syringae DSM 10604[T]	Psy DSM 10604	UK, 1950	*Syringa vulgaris*	JALK01	N.A.	–
P. syringae pv. syringae HS191	Psy HS191	AU, 1979	*Panicum miliaceum*	CP006256-CP006257	Ravindran et al., 2015	–

(Continued)

TABLE 1 | Continued

Strain[1]	Code	Origin[2]	Host	GenBank accession[3]	Reference[4]	Genome subset
P. syringae pv. syringae 642	Psy 642	US, 2007	Unidentified weed	ADGB01	Clarke et al., 2010	–
P. syringae pv. syringae 1212	Psy 1212	United Kingdom	*Pisum sativum*	AVCR02	Baltrus et al., 2014	–
P. amygdali pv. dendropanacis CFBP 3226^{P}	Pde CFBP 3226	JP, 1979	*Dendropanax trifidus*	JYHG01	Bartoli et al., 2015	–
P. meliae CFBP 3225^{T}	P.meliae CFBP 3225	JP, 1974	*Melia azedarach*	JYHE01	Bartoli et al., 2015	–
P. coronafaciens pv. atropurpurea ICMP 4457^{P}	Par ICMP 4457	JP, 1967	*Lolium multiflorum*	LJPS01	N.A.	–
P. syringae CC1513	CC1513	FR, 2006	*Hutchinsia alpina*	AVEL02	Baltrus et al., 2014	–
P. syringae CC1629	CC1629	US, 2007	*Avena sativa*	AVEE02	Baltrus et al., 2014	–
P. cannabina ICMP 2823	Pcb ICMP 2823	HU, 1957	*Cannabis sativa*	LJPX01	N.A.	–
P. cannabina pv. alisalensis ES4326	Pal ES4326	US, 1965	*Raphanus sativus*	AEAK01	Baltrus et al., 2011	–
P. syringae pv. helianthi ICMP 4531^{P}	Phe ICMP4531	MX, 1972	*Helianthus annuus*	LJQM01	N.A.	–
P. syringae pv. tagetis ICMP 4091^{P}	Ptg ICMP4091	ZW, 1972	*Tagetes erecta*	LJRM01	N.A.	–
P. viridiflava TA043	Pvir TA043	FR, 2007	*Primula officinalis*	AVDV01	Baltrus et al., 2014	–
P. viridiflava UASWS0038	Pvir UASWS0038	CH, 2007	*Rhododendron* sp.	AMQP01	Lefort et al., 2013	–
P. syringae CC1416	CC1416	US, 2004	Epilithon	AVEP02	Baltrus et al., 2014	–
P. syringae CC1544	CC1544	FR, 2006	Lake water	AVEI02	Baltrus et al., 2014	–
P. syringae CC1559	CC1559	FR, 2006	Snow	AVEG02	Baltrus et al., 2014	–
P. syringae USA007	USA007	US, 2007	Stream water	AVDY02	Baltrus et al., 2014	–
P. syringae CC1543	CC1543	FR, 2006	Lake water	AVEJ02	Baltrus et al., 2014	–
P. syringae UB0390	UB0390	FR,2007	River water	JPQV01	N.A.	–
P. syringae UB303	UB303	FR, 2006	Lake water	AVDZ02	Baltrus et al., 2014	–
P. syringae USA011	USA011	US, 2007	Stream water	AVDX02	Baltrus et al., 2014	–
P. viridiflava CC1582	Pvir CC1582	FR, 2006	Epilithon	AVDW01	Baltrus et al., 2014	–
P. syringae CC1417	CC1417	US, 2004	Epilithon	AVEO02	Baltrus et al., 2014	–
P. syringae CC1524	CC1524	FR, 2006	Stream water	AVEK02	Baltrus et al., 2014	–
P. syringae CC1583	CC1583	FR, 2006	Epilithon	AVEF02	Baltrus et al., 2014	–
P. syringae CC1557	CC1557	FR, 2006	Snow	CP007014-CP007015	N.A.	–
P. syringae GAW0119	GAW0119	FR, 2010	Irrigation canal	JPQU01	N.A.	–
P. syringae CEB003	CEB003	FR, 2010	Stream water	JPQT01	N.A.	–
P. fluorescens Pf0-1	Pfl Pf0-1	US, 1987	Soil	CP000094	Silby et al., 2009	–
P. putida KT2440	Ppu KT2440	N.D.	Soil	AE015451	Nelson et al., 2002	–

[1]Superscript following strain names indicate ^{T} the type strain of a species and ^{P} the pathotype strain for a pathovar.
Superscript asterisk following strain name indicates strains with a suspected misnaming based on core genome phylogeny. Culture collections providing strains are abbreviated in the strain names as ATCC (American Type Culture Collection, Manassas, Virginia, United States), CFBP (Collection Française de Bactéries associées aux Plantes, FR), DSM (German Collection of Microorganisms and Cell Cultures, DE), ICMP (International Collection of Microorganisms from Plants, NZ), LMG (Bacteria collection of the Laboratory for Microbiology of the Faculty of Sciences of the Ghent University, BE), NCPPB (National Collection of Plant Pathogenic Bacteria, United Kingdom), and MAFF (NIAS Genebank of the Ministry of Agriculture, Forestry and Fisheries, JP).
[2]N.D.: not determined.
[3]For Whole Genome Shotgun (WGS) sequences, accession numbers are provided as four letters prefixes and two digits for the version number of the data set.
[4]N.A.: not applicable.

BLASTN query against all genomes selected for comparative genomics ($n = 97$). The resulting nucleotide sequence was translated using the ExPASy translate tool (ExPASy) and the longest open reading frame corresponding to the reference HopAY sequence (GenBank accession number: AAZ37994.1) was used for alignment. Deviations between the BLASTN hit and the identified protein were investigated in comparison to the reference *hopAY* gene for the possibility of pseudogenization due to frameshift or insertion of a stop codon in the correct reading frame. DNA and amino acid sequences were aligned using ClustalW, while MEGA 6.0 was used to generate neighbor-joining (NJ) phylogeny using the Jones–Taylor–Thornton model with the gamma parameter set at 2.25 and bootstrap values after 1,000 repeats as suggested elsewhere (Lindeberg et al., 2005). A similar method was used twith *hopAR* (GenBank accession number AJ870974.1 positions 17,471–18,274) and *hopAU* (GenBank accession number LT963409.1; locus tag: CFBP3840_01698).

Comparison of the *Arabidopsis* PBS Resistance Protein Among Different Plant Species

The T3E HopAR1 (formerly AvrPphB) from *P. syringae* pv. phaseolicola belongs to the same family of C58 protease as HopAY and has been shown to proteolytically cleave the serine/threonine protein kinase PBS1 in *Arabidopsis*. The amino acid sequence of PBS1 from *Prunus persica* (GenBank accession number XP_007225732) was used to perform a TBLASTN search in the Transcriptome Shotgun Assembly Sequence (TSA) and Protein NCBI databases of the

plants associated with strains possessing a full-length or truncated HopAY.

The transcribed mRNA sequences retrieved from the TSA were translated using the ExPASy translate tool and the obtained amino acid sequences were aligned to the PBS1 amino acid sequence retrieved from the NCBI protein database of 22 additional plant species (**Supplementary Table 1**) using ClustalW on the MEGA 6.0 software. To clarify the phylogenetic relationships among the PBS1 proteins of different plants, a maximum likelihood phylogeny was reconstructed using the Jones–Taylor–Thornton model with the gamma parameter set at 2.25 and bootstrap values of 1,000.

Data Availability

The data sets analyzed for this study are available at the NCBI GenBank/DDJ/EMBL database under the accession detailed in **Table 1**.

RESULTS

Phylogenomics

In order to clarify the exact phylogenetic position of the *Prunus*-associated strains in the data set within the *P. syringae* species complex and to define suitable strains and subgroups for comparative genomics (**Table 1**), a core genome-based phylogeny was generated for the selected set of genomes using EDGAR 2.2 (Blom et al., 2016). The obtained tree was generated based on the concatenated and aligned amino acid sequences of 1,344 genes consisting of a total length of 536,722 amino acids (**Figure 1**).

The main clustering obtained from the core genome phylogeny reflects the PG previously defined by Multi Locus Sequence Analysis (MLSA; Sarkar and Guttman, 2004; Hwang et al., 2005; Sarkar et al., 2006) and single locus phylogeny (Parkinson et al., 2011; Berge et al., 2014). However, our analysis revealed that two genomes obtained from the Whole Genome Shotgun (WGS) NCBI database which were supposed to represent strains isolated from *Prunus* spp. did not cluster as expected based on previous work (Parkinson et al., 2011). Indeed, the sequence with the GenBank WGS accession prefix LAZV01 which is supposed to represent *P. syringae* pv. persicae strain NCPPB 2254 and should cluster close to *P. syringae* pv. avii (Parkinson et al., 2011) was found to be clustering really close to the complete genome of *P. syringae* pv. tomato DC3000 and quite distant from the two other *P. syringae* pv. persicae genomes generated previously (Ruinelli et al., 2019). Additionally, the sequence with accession number LGLQ01 which was deposited in the NCBI database as *P. amygdali* pv. morsprunorum strain FTRSU7805 clustered closer to *P. syringae* pv. cerasicola and *Pseudomonas meliae* than to other strains of *P. syringae* pv. morsprunorum race 1. This observation was supported by the calculation of the average nucleotide identity (ANI) values among the suspected strains, their observed phylogenetically closely related strains and their supposed closely related strains (**Supplementary Figure 1**). Considering these facts, the sequences with the WGS accession prefixes LAZV01 and LGLQ01 were not included in further comparative genomics analysis.

Correlation Between Genes Presence and *Prunus* spp. Association

The number of orthologous proteins present in *Prunus*-associated strains but absent in non-*Prunus*-associated strains retrieved for each of the compared genome subsets (subsets A–D; **Table 1**; **Figure 1**) ranged from 41 (PG3, genome subset D) to 758 (PG2a, genome subset C; **Figure 2**). This considerable difference could be because *Prunus*-associated strains within PG3 belonged to different pathovars and species (*P. syringae* pv. cerasicola, *P. syringae* pv. morsprunorum race 1 and *P. amygdali*), whereas within PG2a only strains of *P. cerasi* have been described to date as being associated with *Prunus* diseases. Among *P. syringae* pv. avii and *P. syringae* pv. persicae (PG1a, genome subset A), a relatively high number of orthologous proteins were retrieved ($n = 249$), whereas only 70 orthologous proteins were found within strains of the genome subset B (PG1b; **Figure 2**).

To verify which proteins were shared among *Prunus* spp. associated strains belonging to different PG, the proteins retrieved for each genome subset ($n = 1{,}058$) were compared for orthology using BLASTP and the results obtained for each possible combination represented in a Venn diagram (**Figure 2**). A total of 52 proteins were found to be shared at least between two genome subsets but no protein was found to be shared among all *Prunus*-associated members of the *P. syringae* species complex. Each protein was checked for distribution across all the initially selected set of genomes ($n = 97$). None of the analyzed proteins ($n = 52$) was found exclusively in *Prunus*-associated strains but 19 of them were found to be significantly more abundant in *Prunus*-associated strains than in non-*Prunus*-associated strains (likelihood ratio statistic exceeding the $p \leq 0.05$ threshold of 5.36; **Figure 1**; **Table 2**; **Supplementary Figure 2**). Out of these, only proteins present in at least 60% of the *Prunus* spp. isolated strains were finally considered, giving a total of 13 proteins (**Figure 1**; **Table 2**). Strains isolated from *Prunus* spp. belonging to PG1a, PG1b, PG2a, and PG3 possessed a similar distribution profile with exception of the *P. syringae* pv. morsprunorum race 2 strain CFBP 6411 (PG1b) and strains from PG2d which were more divergent (**Figure 1**).

A third of the analyzed proteins were hypothetical proteins ($n = 4$) and also a third were located potentially on plasmids ($n = 4$) when complete genomes were available (**Table 1**; **Supplementary Figure 3**). However, three known virulence factors were found to be significantly more present in *Prunus* spp. associated members of the *P. syringae* species complex, namely, three T3E (HopAY, HopAU, and HopBB; **Figure 1**). These three known virulence factors were only reported in the species *P. syringae* during ortholog analysis within 13 additional *Pseudomonas* species closely related to the *P. syringae* species complex (**Supplementary Figure 4**).

The T3E HopAY was the protein with the highest LR statistic and the most abundant in *Prunus* spp. associated strains (89%) if compared to all other considered proteins ($n = 12$) and it was found only in 19% of strains isolated from other hosts

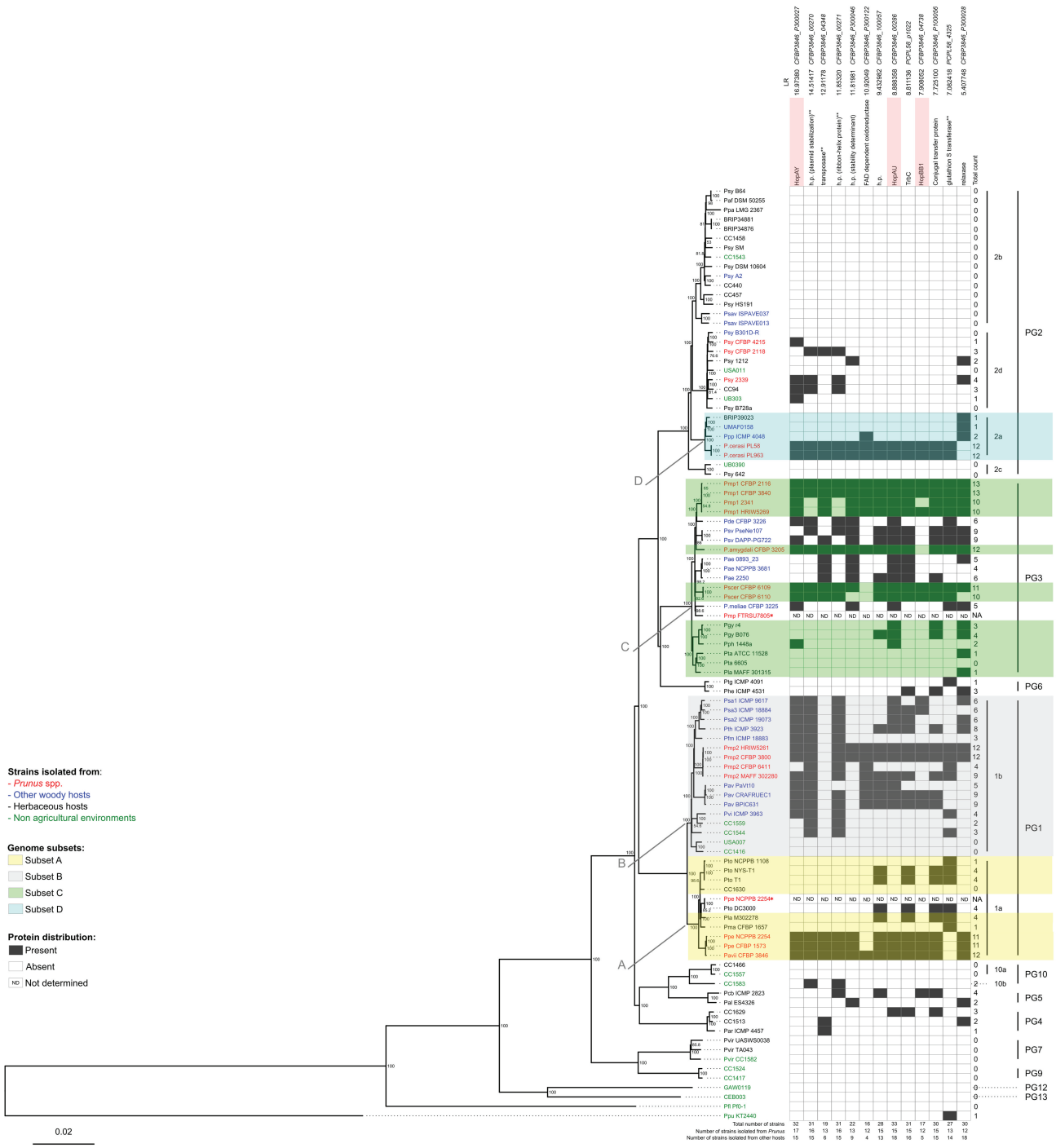


FIGURE 1 | Neighbor-joining (NJ) phylogeny based on the core genome of the *Pseudomonas syringae* species complex and distribution profile of 13 proteins considered to be significantly associated with *Prunus* spp. among these strains. A set of 99 genomes of strains belonging to the *P. syringae* species complex as well as one *Pseudomonas fluorescens* and one *Pseudomonas putida* strains were used for this analysis (**Table 1**). The tree was built using EDGAR 2.2 (Blom et al., 2016) out of a core genome of 1,344 CDS giving a total alignment of 5,36,722 amino acids per genome. Percent bootstrap (bt) support values calculated for 500 reiterations are indicated near nodes. Only bt values over 51 are displayed. The strain names refer to the code field from **Table 1**. Phylogroups (PG) and clades are indicated on the right. Strains isolated from *Prunus* spp. are indicated in red, from other woody hosts in blue, from herbaceous hosts in black, and from non-agricultural environments in green. Strain names followed by an asterisk (*) indicate strains which were excluded from further comparative analyses due to a presumed misnaming of the genome. Genome subsets (A-D) used to determine the correlation between gene presence and *Prunus* spp. association are indicated with color highlights. Arrows indicate the node at which *Prunus*-associated strains are diverging from the non-*Prunus*-associated strains within the same genome subset. Protein orthologs were retrieved out of these 97 genomes using EDGAR 2.2 (Blom et al., 2016). Black squares indicate presence of the protein based on

(Continued)

FIGURE 1 | the orthology criteria of EDGAR 2.2. Proteins highlighted in pink are involved in virulence based on their annotation. Protein descriptions followed by two asterisks (**) indicate that orthologs were also found using online TBLASTN analysis against 13 additional *Pseudomonas* species closely related to the *P. syringae* species complex as reported in **Supplementary Figure 4**. The proteins are ordered by decreasing significance of the likelihood ratio (LR) statistic when exceeding the $p \leq 0.05$ threshold of 5.36. This order is not indicative of any physical proximity. Locus tags and LR statistic are reported over each considered protein; h.p.: hypothetical protein; NA: not applicable; and ND: not determined.

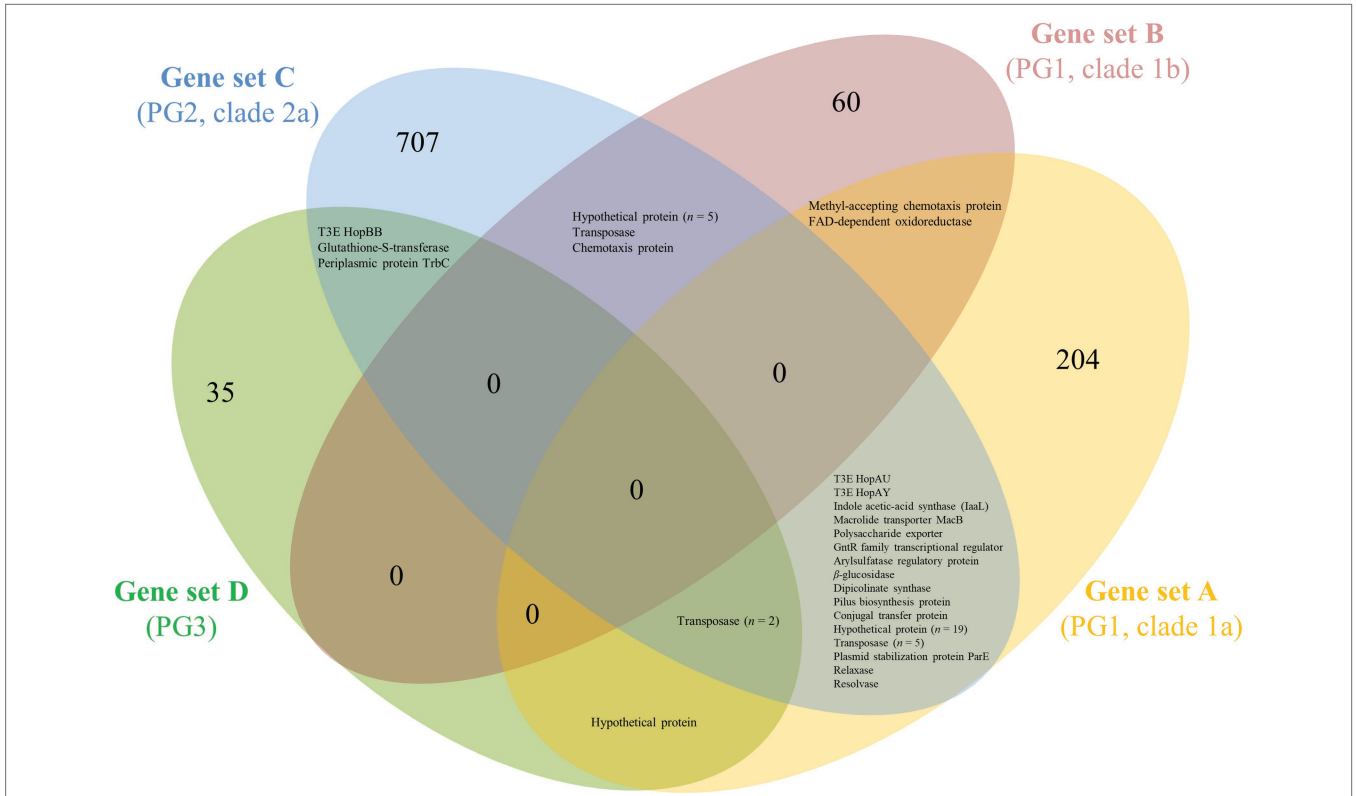


FIGURE 2 | Venn diagram showing groups of ortholog proteins within *Prunus*-associated *Pseudomonas syringae* strains belonging to different PG but absent from phylogenetically closely related non-*Prunus*-associated strains. Subsets of genomes to be compared were defined based on core genome phylogeny within PG1a (subset A), PG1b (subset B), PG2a (subset C), and PG3 (subset D; see **Figure 1**). For each subset of genomes, the core genome of the *Prunus*-associated strains was calculated, and at the same time, all ortholog proteins found in non-*Prunus*-associated strains within the same subset were discarded using EDGAR 2.2 (Blom et al., 2016). The protein sequences resulting for each of the subsets (gene sets A–D) were used as BLASTP query against each other using standalone BLAST v.2.2.29+ (Camacho et al., 2009) and considered as ortholog if identity and coverage were higher or equal to 60%.

TABLE 2 | Number of genes significantly more present in *Prunus* spp. isolated strains.

p-value	LR value	Number of genes			Proportion (%)	
		Expected[1]	Observed	Retained[2]	Tested[3]	Flexible[4]
0.05	5.36	3	19	13	38.78	1.80
0.01	6.40	<1	17	12	34.69	1.61
0.001	9.08	<1	11	7	22.45	1.04
0.0001	12.62	<1	3	3	6.12	0.28
0.00001	12.63	<1	3	3	6.12	0.28

[1]*Expected number of Type I (false-positive) errors under the null model.*
[2]*Retained based on the criteria present in 60% of the Prunus isolated strains.*
[3]*Proportion of the 49 tested genes (three genes skipped based on the criteria occurring in either greater than six or fewer than 92 strains).*
[4]*Proportion of the total flexible genome of subsets A-D (1,058 genes).*

or from non-agricultural environment (**Figure 1**). Within strains of the PG2 ($n = 31$) only six strains harbored HopAY of which four were isolated from *Prunus* spp. (**Figure 1**).

A similar distribution was observed for the T3E HopAU, which was present in 80% of *Prunus*-associated strains and 23% of strains isolated from other hosts. Out of the 32 strains

possessing HopAY, 27 also possessed HopAU (**Figure 1**; **Supplementary Figure 5**). The T3E HopBB was present in only 6% of non-*Prunus* isolated strains but its abundance was also lower in strains isolated from *Prunus* (63%, **Figure 3**).

HopAY is predicted to belong to the same class of C58 peptidases like the well-characterized T3E HopAR (formerly AvrPphB). HopBB has been shown to interact with regulators of the jasmonic acid hormone signaling pathway in *Arabidopsis* (Yang et al., 2017), whereas HopAU was recently shown to activate plant immunity by interacting with a calcium-sensing receptor in *Nicotiana benthamiana* and in kiwifruit (Zhang et al., 2022).

Sequence Comparison of HopAY and HopAR, a Very Well-Characterized C58 Cysteine Protease in *Pseudomonas syringae*

The T3E HopAY showed the strongest level association with *Prunus* isolated strains in respect to all other genes ($n = 13$) analyzed in this study (**Figure 1**) and belongs to the C58 cysteine proteases family. Another well-studied and characterized T3E encoding for a C58 peptidase is HopAR which was initially identified in *P. syringae* pv. phaseolicola as being responsible for elicitation of HR in bean (Jenner et al., 1991; Puri et al., 1997). Orthologs of *hopAR* were retrieved from 10 out of 19 strains isolated from *Prunus* spp. and in 13 strains isolated from other hosts (**Supplementary Figure 5**). Around 15 strains, including nine strains isolated from *Prunus* spp., possessed both *hopAY* and *hopAR* orthologs. The target of HopAR in *Arabidopsis* is the serine/threonine protein kinase AVRPPHB SUSCEPTIBLE 1 (PBS1) and the ability of HopAR to cleave PBS1 is related to the presence in PBS1 of the Glycine (G241)-Aspartate (D242)-Lysine (K243) motif which is also found at the autocleavage site of HopAR (Shao et al., 2003). Mutations in the amino acids G241, D242, and K243 of PBS1 in *Arabidopsis* reduced the proteolytic activity of HopAR by 90, 75, and 15%, respectively (Shao et al., 2003). The cleavage of PBS1 by HopAR induces a conformational change of PBS1 causing the exposition of a particular motif (SEMPH) which is sensed by the resistance protein RESISTANCE TO PSEUDOMONAS SYRINGAE 5 (RPS5) in *Arabidopsis*, leading to HR (Ade et al., 2007; Qi et al., 2012, 2014). In addition, the determination of the crystal structure of HopAR1 revealed the presence of a catalytic triad composed by a cysteine (C98), histidine (H212), and aspartate (D227) which has been shown to be essential for catalysis (Zhu et al., 2004). As already noticed by Zumaquero et al. (2010), the amino acid sequence similarity between HopAY and HopAR is very limited (68% query coverage and 27% identity; **Supplementary Figure 6**). Nevertheless, motifs corresponding to the catalytic triad were identified also on HopAY and localized at C156, H265, and D280 using the HopAY reference present in the T3E database (PPI, 2010; NCBI locus tag: PSPPH_A0129), whereas no motif corresponding to the cleavage site of HopAR (GDK) was found in the HopAY sequence (**Figure 3B**). Secondary structure prediction revealed a conserved pattern of α-helices and β-sheets between HopAR and HopAY as well as other members of the C58 proteases (Zhu et al., 2004). Alignment of the PBS1 protein sequence from different plant species ($n = 31$) revealed that the protein kinase PBS1 is quite conserved among different plant families (Qi et al., 2014). However, we noticed that members of the *Prunus* spp. ($n = 4$) possess an EDK motif instead of the GDK motif essential for HopAR cleavage in PBS1, which was in contrast conserved in all other plant species included in the comparison ($n = 27$; **Figure 3A**). The alignment of the PBS1 sequence of *P. persica* with HopAY revealed that the same EDK motif was found also within the N-terminal half of HopAY (E76, D77, and K78) followed by a stretch of four amino acids with the same physical properties (**Figure 3B**). In addition, all PBS1 sequences analyzed in this study with exception of PBS1 of *Arabidopsis thaliana* and *Capsella rubella* were also lacking the SEMPH motif, which was shown to be essential for RPS5 mediated resistance in *Arabidopsis* (Qi et al., 2014).

Sequence Comparison of HopAY Among Different Members of the *Pseudomonas syringae* Species Complex

In order to determine the evolutionary relationships of *hopAY* within different strains of the *P. syringae* species complex, a BLASTN search was performed using the *hopAY* sequence of *P. syringae* pv. phaseolicola 1448a (PPI; NCBI locus tag: PSPPH_A0129) against the set of genomes selected for comparative genomics ($n = 97$; **Table 1**). The BLASTN analysis revealed the presence of 43 *hopAY* sequences in a total of 36 strains. In addition to the strains retrieved by the protein-based ortholog search ($n = 32$; **Figure 1**), a *hopAY* ortholog was present in the horse chestnut-associated *P. amygdali* pv. aesculi strains 2250, 0893_23, and NCPPB 3681 as well as in the apple tree pathogen *P. syringae* pv. papulans ICMP 4048. With exception of strain HRIW5269, all other *P. syringae* pv. morsprunorum race 1 strains analyzed in this study ($n = 3$) were possessing more than one copy of *hopAY*. In the genomes of *P. syringae* pv. avii strain CFBP 3846, *P. amygdali* CFBP 3205, and *P. syringae* pv. dendropanacis CFBP 3226, two copies of *hopAY* were found as well.

Sequence analysis revealed that the retrieved *hopAY* sequences ($n = 43$) could be divided into five major groups based on the insertion–deletion (indel) scheme affecting this gene (**Figure 4**). The indel group 1 ($n = 25$) consisted of sequences with no insertions or deletions if compared to the reference *hopAY* sequence available in the *hop* database and were mostly retrieved from genomes of strains isolated from *Prunus* spp. ($n = 16$). Sequences belonging to the indel group 2 ($n = 3$) were affected by a probable transposase insertion leading to a 41-bp deletion at the 5′ end (**Figure 4**) and were retrieved only from *Pseudomonas avellanae* strains. Indel groups 3, 4, and 5 displayed an additive indel profile. In fact, the indel group 3 ($n = 2$) displayed a 4-bp deletion at position 66–70 which was shared also from groups 4 ($n = 6$) and 5 ($n = 2$). A 1-bp deletion located at position 737 was also present in sequences of groups 4 and 5, whereas group 5 was additionally having a 12-bp deletion at position 149–160. Sequences of the indel group 4 were

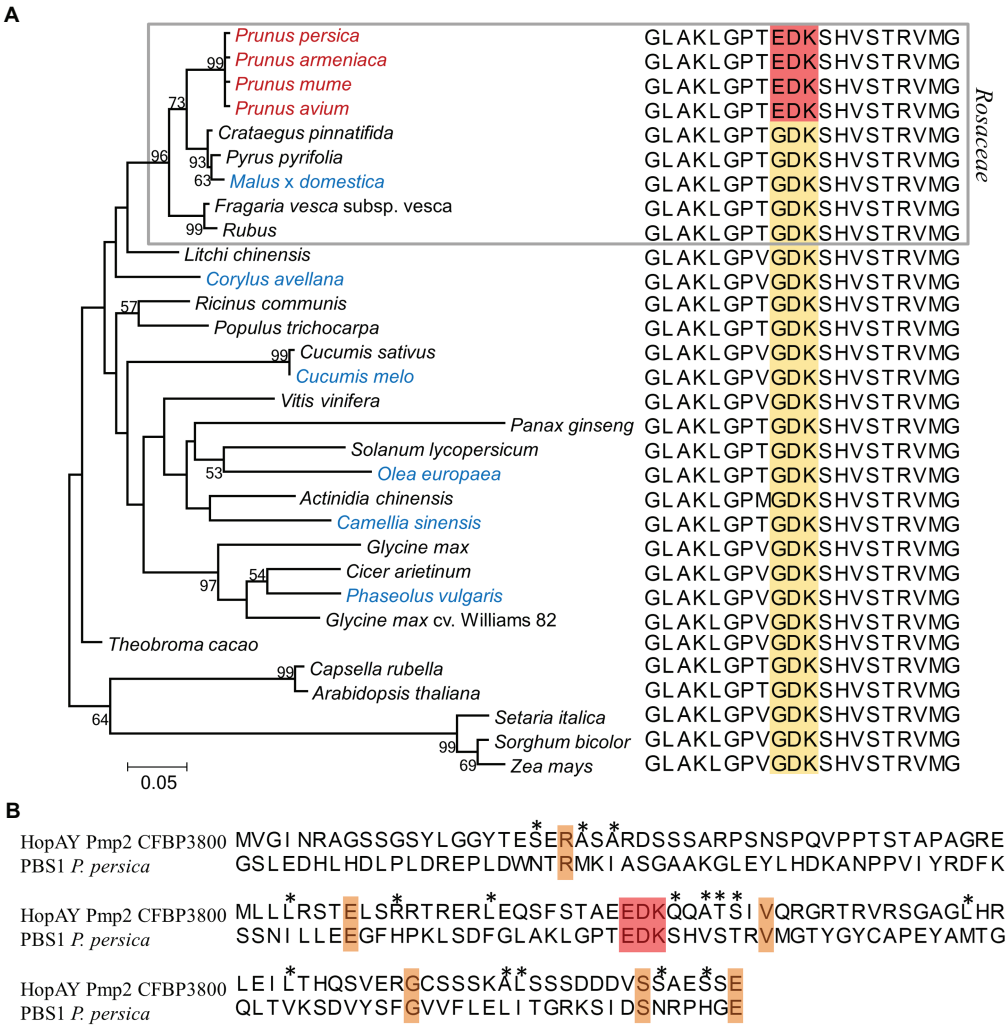


FIGURE 3 | Comparison of the serine/threonine protein kinase PBS1 among different plant species **(A)** and relation to the HopAY sequence **(B)**. **(A)** Maximum likelihood phylogeny of the PBS1 protein among different plant species and relative PBS1 sequence stretch corresponding to the amino acids 233–252 of the *Arabidopsis thaliana* PBS1 sequence (NCBI Acc. Nr. NP_196820) containing the HopAR1 cleavage site GDK (brown block). The corresponding EDK region in the PBS1 sequence of the *Prunus* spp. is highlighted in red. In the phylogeny *Prunus* spp. members are reported in red, whereas plants other than *Prunus* spp. from which *Pseudomonas syringae* strains possessing the HopAY gene were isolated are indicated in blue. The gray block indicates species belonging to the *Rosaceae* family. The evolutionary distances were computed using the Jones–Taylor–Thornton model matrix-based with a gamma distribution (shape parameter = 2.25). Percent bootstrap (bt) support values calculated for 1,000 reiterations are indicated near node. Only bt values over 51 are displayed. All ambiguous positions were removed for each sequence pair giving a total of 598 positions. Alignments were obtained using ClustalW and phylogenetic analysis was performed with MEGA 6.0. **(B)** Alignment of HopAY and PBS1 around the EDK region. The first 133 amino acids (aa) of the HopAY sequence of *Pseudomonas syringae* pv. morsprunorum race 2 CFBP 3800 (Pmp2 CFBP3800) were aligned to the aa 170–302 of the *Prunus persica* PBS1 sequence (NCBI Acc. Nr. XP_007225732). The EDK motif is highlighted in red, identical residues are highlighted in orange, whereas residues sharing similar side chain properties at a specific position are indicated by asterisks.

retrieved only from strains of the PG3 and mostly isolated from *Prunus* spp. ($n = 5$), namely, *P. syringae* pv. cerasicola and *P. syringae* pv. morsprunorum race 1. Based on the complete genomes previously sequenced using PacBio (Ruinelli et al., 2019), it was possible to determine that all *hopAY* of the indel group 4 were located on the chromosome, whereas the *hopAY* of indel group 1 were located on both chromosome and plasmids. Sequences of the indel group 5 were retrieved from two *P. syringae* pv. aesculi strains isolated in Europe, whereas the *P. syringae* pv. aesculi isolated in India displayed an additional resolvase insertion within *hopAY* (**Figure 4**). In addition to the above-described groups, four sequences displayed unique indel profiles varying from transposase insertions (Psa ICMP 18884) to 1-bp deletions (CC94; **Figure 4**).

Alignment of HopAY sequences retrieved from the BLASTN search ($n = 43$) revealed that sequences belonging to the previously described indel groups 4 and 5 as well as four of five additional sequences with unique indel profiles (**Figure 4**) were missing both H256 and D280 due to the introduction of a premature stop codon (**Figure 4**). On the other side, the transposase insertion

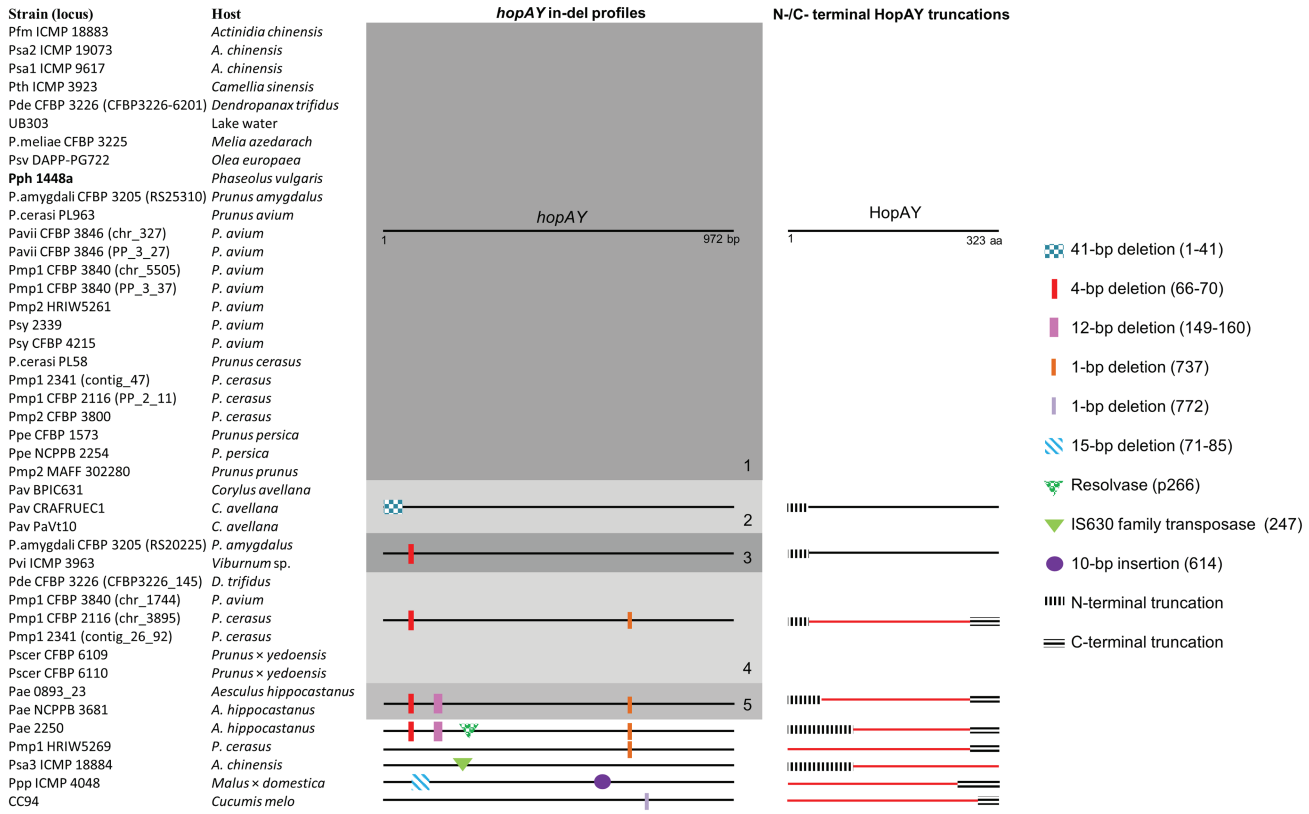


FIGURE 4 | Insertion–deletion profiles of *hopAY* retrieved by BLASTN and corresponding HopAY truncation scheme. Five major groups (1–5) were defined based on conserved indel mutations. Numbers in bracket indicate the position of the insertion or deletion based on the reference *hopAY* sequence from *Pseudomonas syringae* pv. *phaseolicola* 1448a (indicated in bold) available in the Hop database (PPI). HopAY proteins with a truncation affecting the catalytic domains characteristic for C58 cysteine protease and thus considered as non-functional are represented in red. The strain names used refer to the code field from **Table 1**. Figure is not to scale.

within *hopAY* of *P. syringae* pv. actinidiae ICMP 18884 led to a N-terminal truncation deleting the C156 motif. In addition, the HopAY from *P. amygdali* CFBP 3205 belonging to the indel group 1 possessed a tyrosine instead of the expected H256. With exception of both *P. syringae* pv. cerasicola strains, the other three strains isolated from *Prunus* belonging to the indel group 4 thus possess an inactivated HopAY and at least another copy of *hopAY* encoding a full-length protein. The N-terminal truncations observed in the *P. avellanae* strains (indel group 2) and in the sequences of the indel group 3 did not affect the catalytic triad of HopAY and thus it was not possible to determine if the derived protein would be functional or not (**Figure 4**).

Phylogeny of HopAY

The NJ phylogeny obtained from the 43 retrieved HopAY sequences did not reflect the phylogeny obtained from the core genome of the 36 strains possessing a *hopAY* ortholog (**Figures 5A,B**). In particular, the HopAY sequence of *Prunus*-associated strains belonging to PG1a, PG1b, PG2a, and PG3 cluster closer to each other than to strains isolated from other hosts belonging to the same PG.

For example, HopAY sequences from *P. syringae* pv. morsprunorum race 2 strains belonging to clade PG1b cluster closer to *Prunus* isolated strains of PG1a than to strains of the PG1b, namely, *P. syringae* pv. actinidiae and *P. avellanae* (**Figure 5B**). In addition, protein sequences from strains of the PG2d form a monophyletic cluster, which is distantly related to strains of the PG2a clade (**Figure 5B**). Within PG2a, the proteins from *Prunus*-associated *P. cerasi* strains are more closely related to the full-length protein of the PG3 strain *P. amygdali* CFBP 3205, also isolated from *Prunus*, than to that of *P. syringae* pv. papulans strain ICMP 4048 (PG2a). HopAY sequences belonging to the indel groups 4 and 5, which were all retrieved from members of the PG3, form a clearly separated cluster together with the proteins from two strains with unique indel profile, being *P. syringae* pv. morsprunorum race 1 HRIW5269 and *P. syringae* pv. aesculi strain 2250 (**Figure 5B**).

DISCUSSION

The development of effective measures to control plant diseases would be facilitated by a founded knowledge on the pathogen biology as well as on mechanisms involved in the plant–pathogen interactions. Diseases caused by members of the *P. syringae*

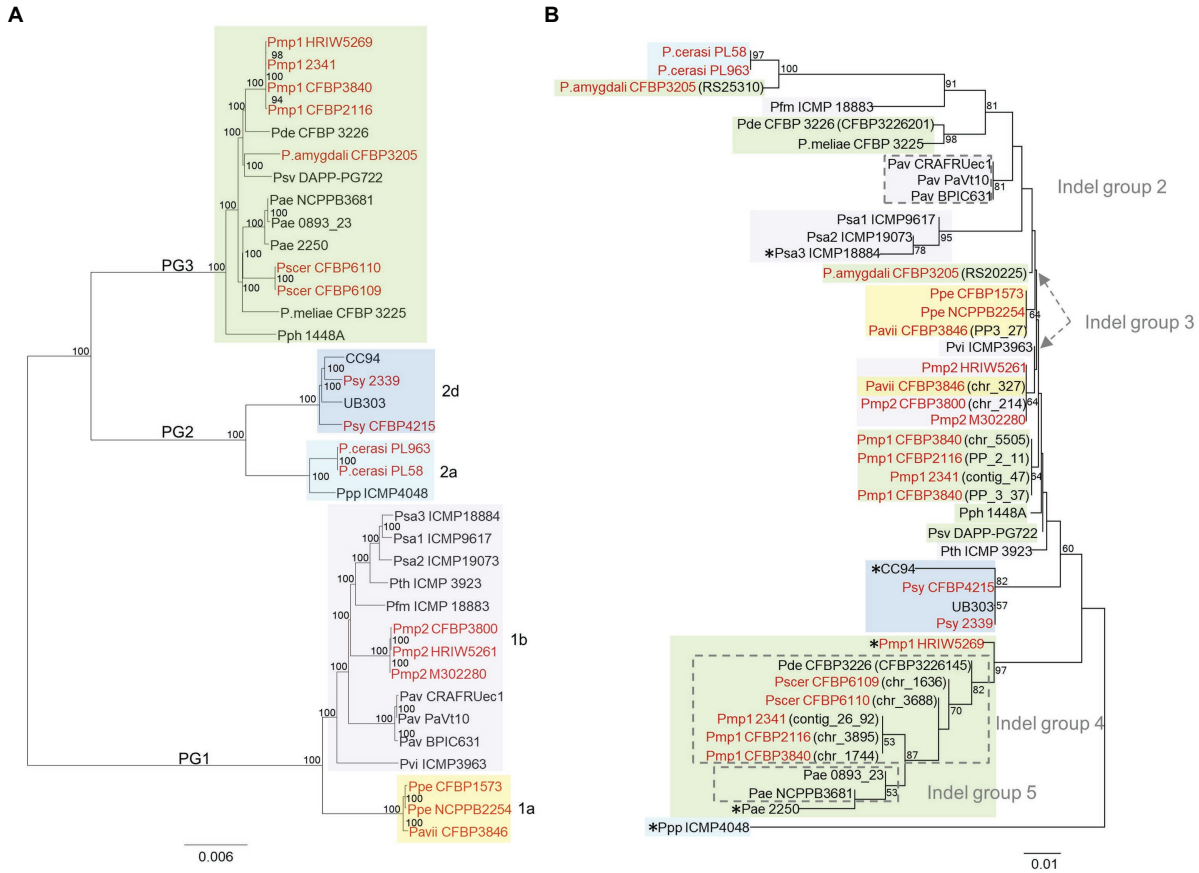


FIGURE 5 | Comparison of the neighbor-joining phylogeny based on the core genome **(A)** and on HopAY **(B)**. **(A)** The core genome of the 36 strains possessing an *hopAY* ortholog based on the BLASTN search was determined using EDGAR 2.2 (Blom et al., 2016) out of a core genome of 2,511 CDS giving a total alignment of 872,675 amino acids per genome. The strain names used refer to the code field from **Table 1**. Phylogroups and clades are indicated on the left and on the right, respectively. **(B)** The phylogeny of the 43 retrieved HopAY sequences was computed using the Jones–Taylor–Thornton model matrix-based with a gamma distribution (shape parameter = 2.25). Percent bootstrap (bt) support values calculated for 1,000 reiterations are indicated near nodes. Only bt values over 51 are displayed. All ambiguous positions were removed for each sequence pair giving a total of 323 positions in the final data set. Evolutionary analyses were conducted in MEGA 6.0. Indel groups (gray dashed lines boxes) refer to **Figure 4**. Asterisks (*) indicate strains with unique indel profiles. If nothing stated, sequences belong to indel group 1. For strains possessing multiple copies of HopAY, the locus tag is indicated in brackets. Strains isolated from *Prunus* spp. are highlighted in red. PG and clades are indicated with the same color code as used on the left in panel **(A)**.

species complex on species belonging to the *Prunus* genus are responsible for relevant yield losses, particularly in young orchards (Puławska et al., 2017). To date, a total of two species and six pathovars belonging to three different PG of the *P. syringae* species complex have been found in association with diseases of species within the *Prunus* genus. Despite their economic importance, not much is known about the evolution and adaptation strategies of members belonging to the *P. syringae* species complex toward *Prunus* spp. In this study, a whole-genome comparison approach was used aiming to identify genetic traits shared among these phylogenetically distantly related pathovars and species that could give insights into the evolutionary aspects related to the adaptation toward *Prunus* spp. hosts.

From the core genome-based phylogeny obtained in this study, it was evident that the pathoadaptation toward *Prunus* spp. is not the result of a single evolutionary event but have evolved independently at least three times in the evolutionary history of the *P. syringae* species complex. This convergent pathoadaptation in distantly related strains leading to virulence on the same host is not unique for the *P. syringae—Prunus* spp. pathosystem within the *P. syringae* species complex (Morris et al., 2019). In fact, phylogenetically distantly related members of the *P. syringae* species complex were also found to have converged onto hazelnut (Wang et al., 2007).

Wang et al. (2007) explained the occurrence of convergent pathoadaptation toward a specific host is not only by the independent acquisition of genes necessary for a successful association but also by the specific loss or inactivation of genes resulting in the same host range limitation. The predominant evolutionary force driving such events in the *P. syringae* species complex is horizontal gene transfer (HGT), which allows the transfer (gain or loss) of genes between closely and distantly related strains within relatively short evolutionary periods (Nowell et al., 2014). Based on HGT, genes having a selective

advantage can be easily accumulated leading to new pathovars or lineages which can adapt to new ecological niches and hosts.

The comparative genomic analysis performed in this study revealed a strong correlation between the presence of the T3E HopAY and the association of members of the *P. syringae* species complex with hosts belonging to the genus *Prunus*. The gene *hopAY* was claimed to be significantly associated with the woody host niche (Nowell et al., 2016), something that we also noticed if considering the *hopAY* orthologs. However, our analysis highlighted the importance of considering not only the gene sequence but also the protein sequence to correctly interpret T3E profiles. A few studies recently took this also into consideration and showed that T3E alleles were linked to host specificity (Zembek et al., 2018; Jayaraman et al., 2020).

Unlike many T3E which have no known function, HopAY is a putative member of the C58 cysteine protease family which is characterized by the presence of an invariant catalytic triad composed by a Cysteine (C), a Histidine (H), and an Aspartate (D) which are essential for catalysis (Shao et al., 2002). Based on that knowledge, it was possible to determine that half of the *hopAY* sequences retrieved based on DNA orthology were encoding for proteins missing at least one of those essential amino acids (C/H/D) and thus would not be functional. Inactivated HopAY were found also in *Prunus*-associated strains but most of them were shown to possess an additional *hopAY* encoding for a full-length protein possessing the C/H/D catalytic triad, often located on a plasmid. The evolutionary dynamics observed within the retrieved HopAY sequences suggests that this protein may be of selective disadvantage on certain hosts and therefore mutated at higher rate than other T3E, like already observed for other T3E families (Baltrus et al., 2011). The phylogeny obtained based on HopAY did not reflect the core genome-based phylogeny, thus excluding a vertical pattern of inheritance and further support the importance of HGT as adaptive force in the evolution of the *P. syringae* species complex. In addition, it revealed that the HopAY sequence present in many *Prunus* spp. associated strains belonging to PG1a, PG1b, and PG3 was nearly identical, supporting the theory of convergent pathoadaptation of these strains.

HopAR (former AvrPphB), another T3E of the C58 cysteine protease family, was subject of many molecular studies in the last decades. These studies revealed that HopAR targets the protein kinase PBS1 in *Arabidopsis* due to the presence of a particular recognition motif (GDK) which was also found in the sequence of HopAR (Shao et al., 2003). Cleavage of PBS1 by HopAR could result in increased virulence or lead to resistance in *Arabidopsis* plants lacking or possessing the resistance protein RPS5, respectively (Ade et al., 2007). The PBS1 protein is quite conserved among different plant species representing a good target for T3Es. In contrast to all other considered plant species, including other members of the *Rosaceae* family, the PBS1 sequence found in *Prunus* spp. lacked the GDK motif necessary for HopAR cleavage and possessed instead an EDK motif which was found also in the N-terminal half of HopAY. The N-terminal part of members of the C58 cysteine protease family is known to be involved in substrate specificity as shown for HopAR and for the DKM motif of Y4zC, a putative T3E of *Rhizobium* (Zhu et al., 2004). Based on this observation, we speculate that HopAY could act in a similar way as HopAR but specifically evolved to cleave the PBS1 ortholog of *Prunus* spp., thus explaining why HopAY is significantly associated with strains adapted to this group of hosts. This hypothesis is supported by the fact that strains isolated from other hosts, such as *Corylus avellana* and *Aesculus hippocastanum*, both harboring a GDK motif in the PBS1 sequence, possessed a truncated or non-functional HopAY, respectively. In addition, Zumaquero et al. (2010) showed that knocking out HopAY does not affect pathogenicity of *P. syringae* pv. phaseolicola 1448a on bean, whose PBS1 protein also possesses a GDK motif. Of course, it could also be hypothesized that the PBS1 protein in *Prunus* has evolved to be cleaved by HopAY to trigger resistance by action of a third unknown resistance protein (similarly to RPS5). However, pathogenicity tests using wild-type strains revealed no direct correlation with presence or absence of *hopAY* (Ruinelli et al., 2019). Therefore, this suggests that *Prunus* spp. does not possess a recognition system for HopAY. At the same time, based on that data, HopAY does not seem to be the determinant factor for pathogenicity but it could still play a role interfering with plant immune response. In order to confirm this hypothesis, additional experiments are needed to show that HopAY is a functional protease able to cleave PBS1 from *Prunus* spp. but the comparative genomic analysis conducted here already provided evidence for sequence correlation between HopAY and its putative cognate target in *Prunus* spp.

This study identifies traits supporting the adaptation between members of the *P. syringae* species complex with species belonging to the *Prunus* genus. It also revealed that most of the mutations affecting *hopAY* were short insertions or deletions that would not be detected by regular PCR and gel electrophoresis, a method that was often used to determine T3E profiles of *P. syringae* and other plant pathogens before the advent of next-generation sequencing technologies (Escalon et al., 2013; Ferrante and Scortichini, 2015). Besides highlighting the biases linked to DNA-based T3E profiling, this study also underlines the importance of integrating host genomic data to correctly interpret the relevance of genomic traits found in the pathogen.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

TS and JP conceptualized the study with the assistance of MR. MR designed the methodology and carried out the experiments. MR and JP analyzed the data with the assistance from JB and TS, contributed to the data visualization, prepared the original draft with assistance from JB and TS for review and editing, and curated the data. JB and JP helped with software. All the authors revised the final version of the manuscript,

while JP acted as the corresponding author. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Swiss Secretariat for Education, Research and Innovation, grant number SBFI C12.0099, within the European research network COST Action FA1104 "Sustainable production of high-quality cherries for the European market" and in part by the European Union Seventh Framework (FP7/2007–2013) under the grant agreement no. 613678 (DROPSA). TS and JP were supported by the Department of Life Sciences and Facility Management of ZHAW. The EDGAR platform is financially supported by the BMF grant FKZ 031A533 within the de.NBI network. Open access funding provided by University Library, Zurich University of Applied Sciences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.804681/full#supplementary-material

REFERENCES

Ade, J., DeYoung, B. J., Golstein, C., and Innes, R. W. (2007). Indirect activation of a plant nucleotide binding site–leucine-rich repeat protein by a bacterial protease. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2531–2536. doi: 10.1073/pnas.0608779104

Almeida, N. F., Yan, S., Lindeberg, M., Studholme, D. J., Schneider, D. J., Condon, B., et al. (2009). A draft genome sequence of *Pseudomonas syringae* pv. tomato T1 reveals a type III effector repertoire significantly divergent from that of *Pseudomonas syringae* pv. tomato DC3000. *Mol. Plant-Microbe Interact.* 22, 52–62. doi: 10.1094/mpmi-22-1-0052

Baltrus, D. A., Nishimura, M. T., Romanchuk, A., Chang, J. H., Mukhtar, M. S., Cherkis, K., et al. (2011). Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathog.* 7:e1002132. doi: 10.1371/journal.ppat.1002132

Baltrus, D. A., Yourstone, S., Lind, A., Guilbaud, C., Sands, D. C., Jones, C. D., et al. (2014). Draft genome sequences of a phylogenetically diverse suite of *Pseudomonas syringae* strains from multiple source populations. *Genome Announc.* 2, e01195–e01213. doi: 10.1128/genomeA.01195-13

Barker, D., and Pagel, M. (2005). Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. *PLoS Comput. Biol.* 1:e3. doi: 10.1371/journal.pcbi.0010003

Bartoli, C., Lamichhane, J. R., Berge, O., Guilbaud, C., Varvaro, L., Balestra, G. M., et al. (2015). A framework to gauge the epidemic potential of plant pathogens in environmental reservoirs: the example of kiwifruit canker. *Mol. Plant Pathol.* 16, 137–149. doi: 10.1111/mpp.12167

Beer, S.V., Bauer, D.W., Jiang, X.H., Laby, R.J., Sneath, B.J., Wei, Z.-M., et al. (1991). "The *hrp* gene cluster of *Erwinia amylovora*," in Advances in Molecular Genetics of Plant-Microbe Interactions, *Vol. 1. Proceedings of the 5th International Symposium on the Molecular Genetics of Plant-Microbe Interactions, Interlaken, Switzerland.* (eds.) H. Hennecke and D.P.S. Verma. September 9–14, 1990. Dordrecht: Springer Netherlands, 53–60.

Berge, O., Monteil, C. L., Bartoli, C., Chandeysson, C., Guilbaud, C., Sands, D. C., et al. (2014). A user's guide to a data base of the diversity of *Pseudomonas syringae* and its application to classifying strains in this phylogenetic complex. *PLoS One* 9:e105547. doi: 10.1371/journal.pone.0105547

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., et al. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* 44, W22–W28. doi: 10.1093/nar/gkw255

Bogdanove, A. J., Beer, S. V., Bonas, U., Boucher, C. A., Collmer, A., Coplin, D. L., et al. (1996). Unified nomenclature for broadly conserved *hrp* genes of phytopathogenic bacteria. *Mol. Microbiol.* 20, 681–683. doi: 10.1046/j.1365-2958.1996.5731077.x

Bonas, U., Schulte, R., Fenselau, F., Minsavage, G. V., Staskawicz, B. J., and Stall, R. E. (1991). Isolation of a gene cluster from *Xanthomonas campestris* pv. vesicatoria that determines pathogenicity and the hypersensitive response on pepper and tomato. *Mol. Plant-Microbe Interact.* 4, 81–88. doi: 10.1094/MPMI-4-081

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., et al. (2003). The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10181–10186. doi: 10.1073/pnas.1731982100

Caballo-Ponce, E., van Dillewijn, P., Wittich, R. M., and Ramos, C. (2016). WHOP, a genomic region associated with woody hosts in the *Pseudomonas syringae* complex contributes to the virulence and fitness of *Pseudomonas savastanoi* pv. savastanoi in olive plants. *Mol. Plant-Microbe Interact.* 30, 113–126. doi: 10.1094/MPMI-11-16-0233-R

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

Cazorla, F. M., Torés, J. A., Olalla, L., Pérez-García, A., Farré, J. M., and de Vicente, A. (1998). Bacterial apical necrosis of mango in southern Spain: a disease caused by *Pseudomonas syringae* pv. syringae. *Phytopathology* 88, 614–620. doi: 10.1094/PHYTO.1998.88.7.614

Choi, O., Kang, B., Lee, Y., Kim, S., Oh, J., Kim, H., et al. (2020). Bacterial shoot blight caused by *Pseudomonas cerasi*, a new pathogen of pear tree. *Aust. Plant Dis. Notes* 15:24. doi: 10.1007/s13314-020-00393-w

Clarke, C. R., Cai, R., Studholme, D. J., Guttman, D. S., and Vinatzer, B. A. (2010). *Pseudomonas syringae* strains naturally lacking the classical *P. syringae hrp/hrc* locus are common leaf colonizers equipped with an atypical type III secretion system. *Mol. Plant-Microbe Interact.* 23, 198–210. doi: 10.1094/mpmi-23-2-0198

Crosse, J. E. (1959). Bacterial canker of stone-fruits: IV. Investigation of a method for measuring the inoculum potential of cherry trees. *Ann. Appl. Biol.* 47, 306–317. doi: 10.1111/j.1744-7348.1959.tb02546.x

Crosse, J. (1966). Epidemiological relations of the pseudomonad pathogens of deciduous fruit trees. *Annu. Rev. Phytopathol.* 4, 291–310. doi: 10.1146/annurev.py.04.090166.001451

Crosse, J. E., and Garrett, C. M. E. (1963). Studies on the bacteriophagy of *Pseudomonas morsprunorum*, *Ps. syringae* and related organisms. *J. Appl. Microbiol.* 26, 159–177. doi: 10.1111/j.1365-2672.1963.tb04764.x

Crosse, J. E., and Garrett, C. M. E. (1966). Bacterial canker of stone-fruits: infection experiments with *Pseudomonas morsprunorum* and *P. syringae*. *Ann. Appl. Biol.* 58, 31–41. doi: 10.1111/j.1744-7348.1966.tb05068.x

Dudnik, A., and Dudler, R. (2013a). High-quality draft genome sequence of *Pseudomonas syringae* pv. syringae strain SM, isolated from wheat. *Genome Announc.* 1, e00610–e00613. doi: 10.1128/genomeA.00610-13

Dudnik, A., and Dudler, R. (2013b). Non contiguous-finished genome sequence of *Pseudomonas syringae* pathovar *syringae* strain B64 isolated from wheat. *Stand. Genomic Sci.* 8, 420–429. doi: 10.4056/sigs.3997732

Dudnik, A., and Dudler, R. (2014). Genome and transcriptome sequences of *Pseudomonas syringae* pv. syringae B301D-R. *Genome Announc.* 2, e00306–e003314. doi: 10.1128/genomeA.00306-14

Dye, D., Bradbury, J., Goto, M., Hayward, A., Lelliott, R., and Schroth, M. N. (1980). International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. *Rev. Plant Pathol.* 59, 153–168. doi: 10.1079/cabireviews/19801362818

EPPO (2005). PM 7/43 (1) *Pseudomonas syringae* pv. persicae. *EPPO Bull.* 35, 285–287. doi: 10.1111/j.1365-2338.2005.00823.x

Escalon, A., Javegny, S., Vernière, C., Noël, L. D., Vital, K., Poussier, S., et al. (2013). Variations in type III effector repertoires, pathological phenotypes and host range of *Xanthomonas citri* pv. citri pathotypes. *Mol. Plant Pathol.* 14, 483–496. doi: 10.1111/mpp.12019

Eurostat (2018). Database—Eurostat [Online]. Available at: https://ec.europa.eu/eurostat/databrowser/view/APRO_CPSH1 (Accessed January 13, 2022).

Feil, H., Feil, W. S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., et al. (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. syringae B728a and pv. tomato DC3000. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11064–11069. doi: 10.1073/pnas.0504930102

Ferrante, P., and Scortichini, M. (2015). Redefining the global populations of *Pseudomonas syringae* pv. actinidiae based on pathogenic, molecular and phenotypic characteristics. *Plant Pathol.* 64, 51–62. doi: 10.1111/ppa.12236

Freigoun, S. O., and Crosse, J. E. (1975). Host relations and distribution of a physiological and pathological variant of *Pseudomonas morsprunorum*. *Ann. Appl. Biol.* 81, 317–330. doi: 10.1111/j.1744-7348.1975.tb01647.x

Gardan, L., Shafik, H., Belouin, S., Broch, R., Grimont, F., and Grimont, P. A. D. (1999). DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959). *Int. J. Syst. Evol. Microbiol.* 49, 469–478. doi: 10.1099/00207713-49-2-469

Gardiner, D. M., Stiller, J., Covarelli, L., Lindeberg, M., Shivas, R. G., and Manners, J. M. (2013). Genome sequences of *Pseudomonas* spp. isolated from cereal crops. *Genome Announc.* 1, e00209–e00213. doi: 10.1128/genomeA.00209-13

Garibaldi, A., Minuto, A., Scortichini, M., and Gullino, M. L. (2007). First report of syringae leaf spot caused by *Pseudomonas syringae* pv. syringae on tomato in Italy. *Plant Dis.* 91:1518. doi: 10.1094/PDIS-91-11-1518B

Gomila, M., Busquets, A., Mulet, M., García-Valdés, E., and Lalucat, J. (2017). Clarification of taxonomic status within the *Pseudomonas syringae* species group based on a phylogenomic analysis. *Front. Microbiol.* 8:2422. doi: 10.3389/fmicb.2017.02422

Gough, C. L., Genin, S., Zischek, C., and Boucher, C. A. (1992). *Hrp* genes of *Pseudomonas solanacearum* are homologous to pathogenicity determinants of animal pathogenic bacteria and are conserved among plant pathogenic bacteria. *Mol. Plant-Microbe Interact.* 5, 384–389. doi: 10.1094/MPMI-5-384

Green, S., Studholme, D. J., Laue, B. E., Dorati, F., Lovell, H., Arnold, D., et al. (2010). Comparative genome analysis provides insights into the evolution and adaptation of *Pseudomonas syringae* pv. aesculi on *Aesculus hippocastanum*. *PLoS One* 5:e10224. doi: 10.1371/journal.pone.0010224

Hulin, M. T., Armitage, A. D., Vicente, J. G., Holub, E. B., Baxter, L., Bates, H. J., et al. (2018). Comparative genomics of *Pseudomonas syringae* reveals convergent gene gain and loss associated with specialization onto cherry (*Prunus avium*). *New Phytol.* 219, 672–696. doi: 10.1111/nph.15182

Hulin, M. T., Jackson, R. W., Harrison, R. J., and Mansfield, J. W. (2020). Cherry picking by pseudomonads: after a century of research on canker, genomics provides insights into the evolution of pathogenicity towards stone fruits. *Plant Pathol.* 69, 962–978. doi: 10.1111/ppa.13189

Hwang, M. S. H., Morgan, R. L., Sarkar, S. F., Wang, P. W., and Guttman, D. S. (2005). Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. *Appl. Environ. Microbiol.* 71, 5182–5191. doi: 10.1128/aem.71.9.5182-5191.2005

Ivanović, Ž., Perović, T., Popović, T., Blagojević, J., Trkulja, N., and Hrnčić, S. (2017). Characterization of *Pseudomonas syringae* pv. syringae, causal agent of citrus blast of mandarin in Montenegro. *Plant Pathol. J.* 33, 21–33. doi: 10.5423/PPJ.OA.08.2016.0161

Jayaraman, J., Yoon, M., Applegate, E. R., Stroud, E. A., and Templeton, M. D. (2020). AvrE1 and HopR1 from *Pseudomonas syringae* pv. actinidiae are additively required for full virulence on kiwifruit. *Mol. Plant Pathol.* 21, 1467–1480. doi: 10.1111/mpp.12989

Jenner, C., Hitchin, E., Mansfield, J., Walters, K., Betteridge, P., Teverson, D., et al. (1991). Gene-for-gene interactions between *Pseudomonas syringae* pv. phaseolicola and *Phaseolus*. *Mol. Plant-Microbe Interact.* 4, 553–562. doi: 10.1094/MPMI-4-553

Joardar, V., Lindeberg, M., Jackson, R. W., Selengut, J., Dodson, R., Brinkac, L. M., et al. (2005). Whole-genome sequence analysis of *Pseudomonas syringae* pv. phaseolicola 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *J. Bacteriol.* 187, 6488–6498. doi: 10.1128/JB.187.18.6488-6498.2005

Jones, L. A., Saha, S., Collmer, A., Smart, C. D., and Lindeberg, M. (2015). Genome-assisted development of a diagnostic protocol for distinguishing high virulence *Pseudomonas syringae* pv. tomato strains. *Plant Dis.* 99, 527–534. doi: 10.1094/pdis-08-14-0833-re

Kałużna, M., Willems, A., Pothier, J. F., Ruinelli, M., Sobiczewski, P., and Puławska, J. (2016). *Pseudomonas cerasi* sp. nov. (non griffin, 1911) isolated from diseased tissue of cherry. *Syst. Appl. Microbiol.* 39, 370–377. doi: 10.1016/j.syapm.2016.05.005

Kamiunten, H., Nakao, T., and Oshida, S. (2000). *Pseudomonas syringae* pv. cerasicola, pv. nov., the causal agent of bacterial gall of cherry tree. *J. Gen. Plant Pathol.* 66, 219–224. doi: 10.1007/PL00012949

Kvitko, B. H., Park, D. H., Velásquez, A. C., Wei, C.-F., Russell, A. B., Martin, G. B., et al. (2009). Deletions in the repertoire of *Pseudomonas syringae* pv. tomato DC3000 type III secretion effector genes reveal functional overlap among effectors. *PLoS Pathog.* 5:e1000388. doi: 10.1371/journal.ppat.1000388

Lalucat, J., Mulet, M., Gomila, M., and García-Valdés, E. (2020). Genomics in bacterial taxonomy: impact on the genus *Pseudomonas*. *Gene* 11:139. doi: 10.3390/genes11020139

Lefort, F., Calmin, G., Crovadore, J., Osteras, M., and Farinelli, L. (2013). Whole-genome shotgun sequence of *Pseudomonas viridiflava*, a bacterium species pathogenic to *Ararabidopsis thaliana*. *Genome Announc.* 1, e00116–e00212. doi: 10.1128/genomeA.00116-12

Lindeberg, M., Cunnac, S., and Collmer, A. (2012). *Pseudomonas syringae* type III effector repertoires: last words in endless arguments. *Trends Microbiol.* 20, 199–208. doi: 10.1016/j.tim.2012.01.003

Lindeberg, M., Stavrinides, J., Chang, J. H., Alfano, J. R., Collmer, A., Dangl, J. L., et al. (2005). Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen *Pseudomonas syringae*. *Mol. Plant-Microbe Interact.* 18, 275–282. doi: 10.1094/MPMI-18-0275

Lindgren, P. B., Panopoulos, N. J., Staskawicz, B. J., and Dahlbeck, D. (1988). Genes required for pathogenicity and hypersensitivity are conserved and interchangeable among pathovars of *Pseudomonas syringae*. *Mol. Gen. Genet.* 211, 499–506. doi: 10.1007/bf00425707

Lindgren, P. B., Peet, R. C., and Panopoulos, N. J. (1986). Gene cluster of *Pseudomonas syringae* pv. "phaseolicola" controls pathogenicity of bean plants and hypersensitivity of nonhost plants. *J. Bacteriol.* 168, 512–522. doi: 10.1128/jb.168.2.512-522.1986

Linke, B., Giegerich, R., and Goesmann, A. (2011). Conveyor: a workflow engine for bioinformatic analyses. *Bioinformatics* 27, 903–911. doi: 10.1093/bioinformatics/btr040

Liu, H., Qiu, H., Zhao, W., Cui, Z., Ibrahim, M., Jin, G., et al. (2012). Genome sequence of the plant pathogen *Pseudomonas syringae* pv. panici LMG 2367. *J. Bacteriol.* 194, 5693–5694. doi: 10.1128/jb.01267-12

Mackey, D., Holt Iii, B. F., Wiig, A., and Dangl, J. L. (2002). RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* 108, 743–754. doi: 10.1016/S0092-8674(02)00661-X

Martínez-García, P. M., Rodríguez-Palenzuela, P., Arrebola, E., Carrión, V. J., Gutiérrez-Barranquero, J. A., Pérez-García, A., et al. (2015). Bioinformatics analysis of the complete genome sequence of the mango tree pathogen *Pseudomonas syringae* pv. syringae UMAF0158 reveals traits relevant to virulence and epiphytic lifestyle. *PLoS One* 10:e0136101. doi: 10.1371/journal.pone.0136101

McCann, H. C., Rikkerink, E. H. A., Bertels, F., Fiers, M., Lu, A., Rees-George, J., et al. (2013). Genomic analysis of the kiwifruit pathogen *Pseudomonas syringae* pv. actinidiae provides insight into the origins of an emergent plant disease. *PLoS Pathog.* 9:e1003503. doi: 10.1371/journal.ppat.1003503

Ménard, M., Sutra, L., Luisetti, J., Prunier, J. P., and Gardan, L. (2003). *Pseudomonas syringae* pv. avii (pv. nov.), the causal agent of bacterial canker of wild cherries (*Prunus avium*) in France. *Eur. J. Plant Pathol.* 109, 565–576. doi: 10.1023/A:1024786201793

Moreno-Pérez, A., Pintado, A., Murillo, J., Caballo-Ponce, E., Tegli, S., Moretti, C., et al. (2020). Host range determinants of *Pseudomonas savastanoi* pathovars of woody hosts revealed by comparative genomics and cross-pathogenicity tests. *Front. Plant Sci.* 11:973. doi: 10.3389/fpls.2020.00973

Moretti, C., Cortese, C., Passos da Silva, D., Venturi, V., Ramos, C., Firrao, G., et al. (2014). Draft genome sequence of *Pseudomonas savastanoi* pv. savastanoi strain DAPP-PG 722, isolated in Italy from an olive plant affected by knot disease. *Genome Announc.* 2, e00864–e00914. doi: 10.1128/genomeA.00864-14

Morris, C. E., Lamichhane, J. R., Nikolić, I., Stanković, S., and Moury, B. (2019). The overlapping continuum of host range among strains in the *Pseudomonas syringae* complex. *Phytopathol. Res.* 1:4. doi: 10.1186/s42483-018-0010-6

Morris, C. E., Sands, D. C., Vinatzer, B. A., Glaux, C., Guilbaud, C., Buffière, A., et al. (2008). The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. *ISME J.* 2, 321–334. doi: 10.1038/ismej.2007.113

Mulet, M., Lalucat, J., and García-Valdés, E. (2010). DNA sequence-based analysis of the *Pseudomonas* species. *Environ. Microbiol.* 12, 1513–1530. doi: 10.1111/j.1462-2920.2010.02181.x

Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V. A., et al. (2002). Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* 4, 799–808. doi: 10.1046/j.1462-2920.2002.00366.x

Newberry, E. A., Ebrahim, M., Timilsina, S., Zlatković, N., Obradović, A., Bull, C. T., et al. (2019). Inference of convergent gene acquisition among *Pseudomonas syringae* strains isolated from watermelon, cantaloupe, and squash. *Front. Microbiol.* 10:270. doi: 10.3389/fmicb.2019.00270

Nowell, R. W., Green, S., Laue, B. E., and Sharp, P. M. (2014). The extent of genome flux and its role in the differentiation of bacterial lineages. *Genome Biol. Evol.* 6, 1514–1529. doi: 10.1093/gbe/evu123

Nowell, R. W., Laue, B. E., Sharp, P. M., and Green, S. (2016). Comparative genomics reveals genes significantly associated with woody hosts in the plant pathogen *Pseudomonas syringae*. *Mol. Plant Pathol.* 17, 1409–1424. doi: 10.1111/mpp.12423

O'Brien, H. E., Thakur, S., Gong, Y., Fung, P., Zhang, J., Yuan, L., et al. (2012). Extensive remodeling of the *Pseudomonas syringae* pv. avellanae type III secretome associated with two independent host shifts onto hazelnut. *BMC Microbiol.* 12:141. doi: 10.1186/1471-2180-12-141

Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 255, 37–45. doi: 10.1098/rspb.1994.0006

Pagel, M., and Meade, A. (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* 167, 808–825. doi: 10.1086/503444

Palleroni, N. J. (2005). "Genus I. *Pseudomonas*" in *Bergey's Manual of Systematic Bacteriology. The Proteobacteria, Part C: The Gammaproteobacteria*. eds. D. J. Brenner, N. R. Krieg, J. T. Staley and G. M. Garrity (Dordrecht: Springer, Netherlands), 323–379.

Parkinson, N., Bryant, R., Bew, J., and Elphinstone, J. (2011). Rapid phylogenetic identification of members of the *Pseudomonas syringae* species complex using the *rpoD* locus. *Plant Pathol.* 60, 338–344. doi: 10.1111/j.1365-3059.2010.02366.x

Picard, C., Afonso, T., Benko-Beloglavec, A., Karadjova, O., Matthews-Berry, S., Paunovic, S. A., et al. (2018). Recommended regulated non-quarantine pests (RNQPs), associated thresholds and risk management measures in the European and Mediterranean region. *EPPO Bull.* 48, 552–568. doi: 10.1111/epp.12500

Popović, T., Ivanović, Ž., Trkulja, N., Milosavljević, A., and Ignjatov, M. (2015). First report of *Pseudomonas syringae* pv. syringae on pea (*Pisum sativum*) in Serbia. *Plant Dis.* 99:724. doi: 10.1094/PDIS-11-14-1212-PDN

PPI (2010). *Pseudomonas syringae* genome resources home page [Online]. Available at: www.pseudomonas-syringae.org (Accessed January 23, 2017).

Psallidas, P. G. (1997). Hyperplastic canker—a perennial disease of almond caused by *Pseudomonas amygdali*. *EPPO Bull.* 27, 511–517. doi: 10.1111/j.1365-2338.1997.tb00675.x

Puławska, J., Gétaz, M., Kałużna, M., Kuzmanović, N., Obradović, A., Pothier, J. F., et al. (2017). "Bacterial diseases" in *Cherries: Botany, Production and Uses*. eds. J. Quero-García, A. Iezzoni, J. Puławska and G. Lang (Wallingford, United Kingdom: CAB International), 365–385.

Puri, N., Jenner, C., Bennett, M., Stewart, R., Mansfield, J., Lyons, N., et al. (1997). Expression of *avrPphB*, an avirulence gene from *Pseudomonas syringae* pv. phaseolicola, and the delivery of signals causing the hypersensitive reaction in bean. *Mol. Plant-Microbe Interact.* 10, 247–256. doi: 10.1094/MPMI.1997.10.2.247

Qi, D., DeYoung, B. J., and Innes, R. W. (2012). Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein. *Plant Physiol.* 158, 1819–1832. doi: 10.1104/pp.112.194035

Qi, D., Dubiella, U., Kim, S. H., Sloss, D. I., Dowen, R. H., Dixon, J. E., et al. (2014). Recognition of the protein kinase AVRPPHB SUSCEPTIBLE1 by the disease resistance protein RESISTANCE TO PSEUDOMONAS SYRINGAE5 is dependent on S-acylation and an exposed loop in AVRPPHB SUSCEPTIBLE1. *Plant Physiol.* 164, 340–351. doi: 10.1104/pp.113.227686

Qi, M., Wang, D., Bradley, C. A., and Zhao, Y. (2011). Genome sequence analyses of *Pseudomonas savastanoi* pv. *glycinea* and subtractive hybridization-based comparative genomics with nine pseudomonads. *PLoS One* 6:e16451. doi: 10.1371/journal.pone.0016451

Ravindran, A., Jalan, N., Yuan, J. S., Wang, N., and Gross, D. C. (2015). Comparative genomics of *Pseudomonas syringae* pv. syringae strains B301D and HS191 and insights into intrapathovar traits associated with plant pathogenesis. *Microbiology* 4, 553–573. doi: 10.1002/mbo3.261

Ruinelli, M., Blom, J., Smits, T. H. M., and Pothier, J. F. (2019). Comparative genomics and pathogenicity potential of members of the *Pseudomonas syringae* species complex on *Prunus* spp. *BMC Genomics* 20:172. doi: 10.1186/s12864-019-5555-y

Sarkar, S. F., Gordon, J. S., Martin, G. B., and Guttman, D. S. (2006). Comparative genomics of host-specific virulence in *Pseudomonas syringae*. *Genetics* 174, 1041–1056. doi: 10.1534/genetics.106.060996

Sarkar, S. F., and Guttman, D. S. (2004). Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Appl. Environ. Microbiol.* 70, 1999–2012. doi: 10.1128/aem.70.4.1999-2012.2004

Scortichini, M., Marcelletti, S., Ferrante, P., and Firrao, G. (2013). A Genomic redefinition of *Pseudomonas avellanae* species. *PLoS One* 8:e75794. doi: 10.1371/journal.pone.0075794

Shao, F., Golstein, C., Ade, J., Stoutemyer, M., Dixon, J. E., and Innes, R. W. (2003). Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. *Science* 301, 1230–1233. doi: 10.1126/science.1085671

Shao, F., Merritt, P. M., Bao, Z., Innes, R. W., and Dixon, J. E. (2002). A *Yersinia* effector and a *Pseudomonas* avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell* 109, 575–588. doi: 10.1016/S0092-8674(02)00766-3

Silby, M. W., Cerdeño-Tárraga, A. M., Vernikos, G. S., Giddens, S. R., Jackson, R. W., Preston, G. M., et al. (2009). Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biol.* 10:R51. doi: 10.1186/gb-2009-10-5-r51

Spotts, R. A., Wallis, K. M., Serdani, M., and Azarenko, A. N. (2010). Bacterial canker of sweet cherry in Oregon—Infection of horticultural and natural wounds, and resistance of cultivar and rootstock combinations. *Plant Dis.* 94, 345–350. doi: 10.1094/pdis-94-3-0345

Templeton, M. D., Warren, B. A., Andersen, M. T., Rikkerink, E. H., and Fineran, P. C. (2015). Complete DNA sequence of *Pseudomonas syringae* pv. actinidiae, the causal agent of kiwifruit canker disease. *Genome Announc.* 3, e01054–e01115. doi: 10.1128/genomeA.01054-15

Wang, P. W., Morgan, R. L., Scortichini, M., and Guttman, D. S. (2007). Convergent evolution of phytopathogenic pseudomonads onto hazelnut. *Microbiology* 153, 2067–2073. doi: 10.1099/mic.0.2006/001545-0

Wei, W., Plovanich-Jones, A., Deng, W., Jin, Q., Collmer, A., Huang, H., et al. (2000). The gene coding for the Hrp pilus structural protein is required for type III secretion of Hrp and Avr proteins in *Pseudomonas syringae* pv. tomato. *Proc. Natl. Acad. Sci. U. S. A.* 97, 2247–2252. doi: 10.1073/pnas.040570097

Xiang, T., Zong, N., Zou, Y., Wu, Y., Zhang, J., Xing, W., et al. (2008). *Pseudomonas syringae* effector AvrPto blocks innate immunity by targeting receptor kinases. *Curr. Biol.* 18, 74–80. doi: 10.1016/j.cub.2007.12.020

Yang, L., Teixeira, P. J., Biswas, S., Finkel, O. M., He, Y., Salas-Gonzalez, I., et al. (2017). *Pseudomonas syringae* type III effector HopBB1 promotes host transcriptional repressor degradation to regulate phytohormone responses and virulence. *Cell Host Microbe* 21, 156–168. doi: 10.1016/j.chom.2017.01.003

Young, J. M. (1987). New plant disease record in New Zealand: *Pseudomonas syringae* pv. persicae from nectarine, peach, and Japanese plum. *N. Z. J. Agric. Res.* 30, 235–247. doi: 10.1080/00288233.1987.10430502

Young, J. M. (2010). Taxonomy of *Pseudomonas syringae*. *J. Plant Pathol.* 92, S1.5–S1.14. doi: 10.4454/jpp.v92i1sup.2501

Zembek, P., Danilecka, A., Hoser, R., Eschen-Lippold, L., Benicka, M., Grech-Baran, M., et al. (2018). Two strategies of *Pseudomonas syringae* to avoid recognition of the HopQ1 effector in *Nicotiana* species. *Front. Plant Sci.* 9:978. doi: 10.3389/fpls.2018.00978

Zhang, J., Zhou, M., Liu, W., Nie, J., and Huang, L. (2022). *Pseudomonas syringae* pv. actinidiae effector HopAU1 interacts with calcium-sensing

receptor to activate plant immunity. *Int. J. Mol. Sci.* 23:508. doi: 10.3390/ijms23010508

Zhao, W., Jiang, H., Tian, Q., and Hu, J. (2015). Draft genome sequence of *Pseudomonas syringae* pv. persicae NCPPB 2254. *Genome Announc.* 3, e00555–e00565. doi: 10.1128/genomeA.00555-15

Zhou, L. H., Han, Y., Ji, G. H., Wang, Z. S., and Liu, F. (2012). First report of bacterial leaf spot disease caused by *Pseudomonas syringae* pv. syringae on *Panax notoginseng*. *Plant Dis.* 97:685. doi: 10.1094/PDIS-11-12-1047-PDN

Zhu, M., Shao, F., Innes, R. W., Dixon, J. E., and Xu, Z. (2004). The crystal structure of *Pseudomonas* avirulence protein AvrPphB: A papain-like fold with a distinct substrate-binding site. *Proc. Natl. Acad. Sci. U. S. A.* 101, 302–307. doi: 10.1073/pnas.2036536100

Zumaquero, A., Macho, A. P., Rufián, J. S., and Beuzón, C. R. (2010). Analysis of the role of the type III effector inventory of *Pseudomonas syringae* pv. phaseolicola 1448a in interaction with the plant. *J. Bacteriol.* 192, 4474–4488. doi: 10.1128/JB.00260-10

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership