



## OPEN ACCESS

## EDITED BY

Michelle Teresa Hulin,  
The Sainsbury Laboratory, United Kingdom

## REVIEWED BY

David J. Studholme,  
University of Exeter, United Kingdom  
Manoj Choudhary,  
University of Florida, United States  
Ziyue Zeng,  
National Institute of Agricultural Botany  
(NIAB), United Kingdom

## \*CORRESPONDENCE

Tal Pupko

✉ talp@tauex.tau.ac.il

RECEIVED 31 January 2023

ACCEPTED 10 May 2023

PUBLISHED 02 June 2023

## CITATION

Wagner N, Ben-Meir D, Teper D and  
Pupko T (2023) Complete genome  
sequence of an Israeli isolate of  
*Xanthomonas hortorum* pv. pelargonii  
strain 305 and novel type III effectors  
identified in *Xanthomonas*.  
*Front. Plant Sci.* 14:1155341.  
doi: 10.3389/fpls.2023.1155341

## COPYRIGHT

© 2023 Wagner, Ben-Meir, Teper and  
Pupko. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Complete genome sequence of an Israeli isolate of *Xanthomonas hortorum* pv. pelargonii strain 305 and novel type III effectors identified in *Xanthomonas*

Naama Wagner<sup>1</sup>, Daniella Ben-Meir<sup>1</sup>, Doron Teper<sup>2</sup>  
and Tal Pupko<sup>1\*</sup>

<sup>1</sup>The Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel, <sup>2</sup>Department of Plant Pathology and Weed Research, Institute of Plant Protection Agricultural Research Organization (ARO), Volcani Institute, Rishon LeZion, Israel

*Xanthomonas hortorum* pv. pelargonii is the causative agent of bacterial blight in geranium ornamental plants, the most threatening bacterial disease of this plant worldwide. *Xanthomonas fragariae* is the causative agent of angular leaf spot in strawberries, where it poses a significant threat to the strawberry industry. Both pathogens rely on the type III secretion system and the translocation of effector proteins into the plant cells for their pathogenicity. Effectidor is a freely available web server we have previously developed for the prediction of type III effectors in bacterial genomes. Following a complete genome sequencing and assembly of an Israeli isolate of *Xanthomonas hortorum* pv. pelargonii - strain 305, we used Effectidor to predict effector encoding genes both in this newly sequenced genome, and in *X. fragariae* strain Fap21, and validated its predictions experimentally. Four and two genes in *X. hortorum* and *X. fragariae*, respectively, contained an active translocation signal that allowed the translocation of the reporter AvrBs2 that induced the hypersensitive response in pepper leaves, and are thus considered validated novel effectors. These newly validated effectors are XopBB, XopBC, XopBD, XopBE, XopBF, and XopBG.

## KEYWORDS

*Xanthomonas*, type-III secretion system, Effector proteins, type-III effectors, machine learning, Effectidor

## 1 Introduction

The *Xanthomonas* genus includes dozens of species divided to thousands of subspecies and strains with a wide range of lifestyles: from commensal, to opportunistic, to pathogenic. Among them are some of the major plant pathogens worldwide, affecting more than 400 plant species (Timilsina et al., 2020). These pathogens rely on the type III secretion system (T3SS) and type III effectors (T3Es) for their pathogenicity. The effectors

alter processes within the host cell for the benefit of the bacteria and thus promote disease in the plant (White et al., 2009; Ryan et al., 2011; An et al., 2019). Identification of the full effector repertoire encoded within the genome of a pathogenic bacterium is a prerequisite for detailed understanding of the molecular interactions between the pathogen and its host.

Discovering novel effectors is a challenging task, as effectors are highly diverse in their functionality, size, and structure. Moreover, the effector repertoire varies even among closely related strains (Jalan et al., 2013; Jiménez-Guerrero et al., 2020). The T3SS recognizes T3Es based on a secretion signal located in their N-terminus (Michiels and Cornelis, 1991; Sory and Cornelis, 1994). However, despite extensive efforts to characterize it, the secretion signal of T3Es is not characterized enough to allow accurate prediction of effectors as a sole feature (Wagner et al., 2022a). We have previously developed and applied machine-learning techniques to identify T3Es and type IVb effectors in various pathogenic bacteria (Burstein et al., 2009; Lifshitz et al., 2013; Lifshitz et al., 2014; Burstein et al., 2015; Burstein et al., 2016; Teper et al., 2016; Nissan et al., 2018; Jiménez-Guerrero et al., 2020; Ruano-Gallego et al., 2021). Following these efforts, we developed Effectidor: an automated machine-learning based web server for the prediction of T3Es (Wagner et al., 2022b). Effectidor combines dozens of different features to achieve accurate classification, e.g., sequence similarity to previously validated effectors, sequence similarity to host proteins, and atypical GC-content. Another feature is the sequence similarity to closely related bacteria without T3SS (putative effectors are expected not to have strong hits when searching against such genomes). Additional features that we consider are the amino acid composition, the genomic organization (effectors often reside close to each other in the genome), existence of known regulatory elements that are recognized by transcriptional regulators that regulate the T3SS and some of the T3Es, such as the plant-inducible promoter (PIP)-BOX (Cunnac et al., 2004; Koebnik et al., 2006), and a signal score reflecting the likelihood of the existence of a secretion signal in the 100 N-terminal amino-acids of the protein (Wagner et al., 2022a). Using these features, Effectidor trains a machine-learning classifier on the known effectors and non-effectors of the specific bacterial genome it analyzes, and outputs a prediction for all the other protein coding genes in the genome, reflecting their likelihood to encode an effector. Thus, we can pinpoint the T3Es candidates in the genome with no need for labor and cost intensive full-genomic screening. In this work we applied Effectidor to two *Xanthomonas* pathogens: *Xanthomonas fragariae*, the causative agent of angular leaf spot (ALS) in strawberries, and *X. hortorum* pv. *pelargonii*, the causative agent of bacterial blight in geranium.

The pathogen *X. fragariae* (Xfrg) was first reported in Minnesota, the United States in 1960 (Kennedy and King, 1962), and since then it has spread worldwide (Mazzucchi et al., 1973; McGechan and Fahy, 1976; Gubler et al., 2007; Matthews-Berry and Reed, 2009; Fernández-Pavía et al., 2014; Kamangar et al., 2017; Wu et al., 2020; Song et al., 2021). This bacterium is a quarantine pathogen in Europe and it is currently widely spread in North America, where it causes substantial loss in the strawberry nursery

industry (Puławska et al., 2020). In severe cases of the disease the crop production is significantly reduced either due to death of the plant or due to changes in the appearance of the fruit, which make them unmarketable. Yet, the most severe economic threat is to nurseries, where the bacteria spread easily. Currently there are neither resistant strawberry plants, nor effective treatments against the pathogen (Wang et al., 2018).

The bacterium *X. hortorum* pv. *pelargonii* (Xhp) is the causal agent of bacterial blight in geranium ornamental plants (also known by the name “pelargonium”). This is the most threatening bacterial disease of these plants worldwide (Barel et al., 2015; Balaž et al., 2016). The disease is widespread in various states of the USA, Europe, Australia and Israel, and may cause heavy economic losses. Warm and wet conditions favor infection and disease development. Normally, Xhp penetrates the plant *via* natural openings or wounds, and spreads systemically through the vascular system. Symptoms are characterized by wilting of the plant, localized water-soaked lesions that often become necrotic and rotted cuttings. All commercial cultivars of geranium are susceptible to Xhp (Zhang et al., 2009).

In this work, we aimed to discover new T3Es in these two pathogens. We first sequenced the genome of an Israeli isolate of Xhp, combining short and long reads to obtain high quality genome sequence. Next, we applied Effectidor (Wagner et al., 2022a; Wagner et al., 2022b; Wagner et al., 2022c) to predict T3Es in these two genomes. Our results suggested the existence of unknown T3Es in both genomes, i.e., putative T3Es without significant sequence similarity to previously identified effectors. We next experimentally validated the translocation of some of these putative T3Es in a T3SS mediated manner. We validated two and four novel T3Es in Xfrg and Xhp, respectively.

## 2 Materials and methods

### 2.1 Bacterial strains and plant material

The bacterium *X. hortorum* pv. *pelargonii* (Xhp) strain 305 was isolated from geranium plants in Israel and was a gift from Dr. Shulamit Manulis-Sasson from the Agricultural Research Organization (ARO), Volcani Center Israel (Barel et al., 2015). Genomic DNA of *X. fragariae* (Xfrg) Fap21 (BioSample SAMN05505397) (Henry and Leveau, 2016) was kindly provided by Dr. Joël Pothier (Zurich University of Applied Sciences). For the translocation assays, we used *X. euvesicatoria* (Xeu) *hrpG\**  $\Delta$ *avrBs2* (Roden et al., 2004). For cloning, we used NEB 5-alpha *Escherichia coli* that were obtained from New England Bio-Labs inc.

Strains of *E. coli* and *Xanthomonas* were grown in Luria-Bertani (LB), broth or agar, at 37°C and 28°C, respectively. The antibiotics used were spectinomycin (Sp; 100 µg/ml), kanamycin (Kan; 50 µg/ml) and gentamicin (Gm; 10 µg/ml). All antibiotics were from Sigma-Aldrich.

Pepper plants (*Capsicum annuum*) ECW20R (Kearney and Staskawicz, 1990) were grown in the greenhouse at 25°C and kept in long-day conditions (16 h light, 8 h dark).

## 2.2 Genome sequencing of *X. hortorum* pv. *pelargonii* 305

Genomic DNA of Xhp305 was isolated from 3 ml of overnight culture using Wizard<sup>®</sup> Genomic DNA Purification Kit – Promega. Microbial *De novo* sequencing was performed at Novogene Co., Ltd. using both PacBio (PacBio Sequel II) and Illumina (NovaSeq 6000) platforms. The shotgun genomic library for short-read sequencing and the library for long-read sequencing were prepared by the service provider, who also performed quality control. The Illumina sequencing yielded 17,914,078 paired-end reads of length 150 bp. The PacBio sequencing yielded, after trimming, 245,660 subreads, with mean length of 11,006 bp, N50 of 13,343 bp, for a total of 2,704 Mbp.

### 2.2.1 Genome assembly and annotation

For *de novo* assembly, the whole set of PacBio subreads was used as input for Canu v2.2 (Koren et al., 2017) with the following parameters: -pacbio-raw genomeSize = 5.6m. The average coverage was assessed by mapping corrected and trimmed reads obtained by Canu v2.2 against the assembly using BWA v0.7.17 (Li and Durbin, 2009) with default parameter values, calculating the alignment depth using SAMtools v1.3.3 (Li et al., 2009) with default parameter values, and the average depth per molecule using awk. We then used the draft genome and the corrected PacBio reads as input for Circlator (Hunt et al., 2015), together with BWA v0.7.17 (Li and Durbin, 2009), prodigal v2.6.3 (Hyatt et al., 2010), SAMtools v1.3.3 (Li et al., 2009), MUMmer v3.23 (Kurtz et al., 2004), and Canu v2.2 (Koren et al., 2017) to circularize the chromosome and plasmids, with the following parameters: circlator all -assembler canu. Following this step, we used the Illumina reads to polish the assembly using Pilon v1.22 (Walker et al., 2014), BWA v0.7.17 (Li and Durbin, 2009) and SAMtools v1.3.3 (Li et al., 2009) with the default parameter values and including -changes to keep track of the corrections done in the assembly. We applied three rounds of polishing using Pilon, until no further corrections were introduced in the fourth round. The average coverage of the Illumina reads was assessed in the same manner as assessed for the PacBio reads. For genome annotation, we used Prokka v1.13.3 (Seemann, 2014) with default parameter values.

## 2.3 Effectors prediction

Effectidor v1.04 (Wagner et al., 2022b) was used for T3Es predictions in each of the two genomes. The pipeline within Effectidor is divided into the following steps: (1) Defining the positive T3Es in the input genome either based on the input supplied by the user or based on homology to previously validated T3Es from various strains (this dataset can be viewed and downloaded from <https://effectidor.tau.ac.il/data.html>. For the analysis done in this work, version 1.04 was used). The homology criteria are E-values smaller than  $10^{-10}$  and at least 70% identical matches. If less than five effectors are identified based on this cutoff, the last criterion is reduced by 10 (i.e., 60% identical matches are required instead of 70%) until a minimum of 40% identical

matches; (2) Defining the negative set (i.e., non T3Es encoded in the input genome) based on homology to proteins of *E. coli* K12 MG1655 (accession GCF\_000005845.2); (3) Feature extraction. The features used in Effectidor vary based on the provided input. While the only mandatory input in Effectidor is a FASTA file containing all the ORFs records in the genome, additional inputs allow extraction of features outside the gene sequence alone. In our analysis we provided the following additional inputs, available in the advanced options of Effectidor: (3.1) GFF file, which holds information about the location of all the genes in the genome and allows Effectidor to extract genome organization features; (3.2) FASTA file of the full genome, which allows, together with the GFF file, to search for the PIP-box regulatory element in the promoters of the genes; (3.3) ZIP archive with FASTA files holding protein records of *Luteimonas* sp. MC1825 (accession GCF\_014764385), *Lysobacter capsica* 55 [accession GCF\_001442785 (de Bruijn et al., 2015)], *Pseudoxanthomonas suwonensis* J1 [accession GCF\_000972865 (Hou et al., 2015)], *Stenotrophomonas maltophilia* K279a [accession GCF\_000072485 (Crossman et al., 2008)], and *Xylella fastidiosa* 9a5c [accession GCF\_000006725 (Simpson et al., 2000; Marques et al., 2001)] that were used as input for the proteomes of closely related bacteria without T3SS. This input allows to run homology searches against these proteomes, and the results of these searches often serve as informative features for the machine-learning classifier, as T3Es are not expected to be found in these genomes, whereas many of the non-T3Es – are. In addition to these inputs, we conducted all searches including the optional feature that predicts the presence of the type 3 secretion signal in the protein sequence (Wagner et al., 2022a); (4) Training a machine learning classifier. Following the feature extraction step, several classifiers (i.e., Linear Discriminant Analysis, Naïve Bayes, Support Vector Machine, Logistic Regression, K Nearest Neighbors, and Random Forest) are trained on the labeled data (i.e., T3Es and non-T3Es defined in the first step). The labeled data are split into train and test sets, the classifiers are first trained in cross-validation on the training set (including feature selection) and are finally evaluated on the test set. The evaluation method used in Effectidor is the Area Under the Precision-Recall Curve (AUPRC). The temporary best classifier is defined as the one with the highest AUPRC on the test set. All classifiers are then evaluated according to the following criteria: (4.1) AUPRC measured on the test set is smaller than that achieved by the temporary best classifier by no more than 30%; (4.2) The range of the prediction scores of the genes in the training set is at least 0.75, to ensure that not all samples are classified as negative/positive; (4.3) The AUPRC measured on the train set and on the test set are compared, and the difference must be smaller than 0.25, to reduce chances of overfitting. The classifiers that meet all these criteria, are then merged to form a final voting classifier; (5) Applying the final classifier to identify potential novel T3Es in the genome. The final voting classifier is then applied to produce a score between 0 and 1 for each ORF in the genome. This score reflects the likelihood for this ORF to encode a T3E. Of note, in case all classifiers were dropped for not meeting some of the criteria mentioned in (4.3), the final classifier used to produce these predictions is a vote over all classifiers. In this case a message is

sent to the user of Effectidor. Genomes with a small number of known effectors are more susceptible to it.

## 2.4 Translocation assay

### 2.4.1 Plasmid construction

The plasmid pAvrBs2-HR (KanR) containing the Hypersensitive Response (HR) domain of *avrBs2* (amino acids 62–574), fused to an haemagglutinin (HA) tag (Teper et al., 2016), was used as vector for cloning and expression of candidate effector genes. The vector was linearized with XhoI and XbaI restriction enzymes (Thermo Fisher Scientific). The putative T3E genes of Xhp and Xfrg, including 24 bp upstream of their ATG start codon, were PCR amplified (Phusion Hot Start II High-Fidelity DNA Polymerase, Thermo Scientific) from genomic DNA of Xhp305 and XfrgFap21 using gene specific primers (Supplementary Table S1). In most cases the whole candidate gene was amplified, but in two cases, (PML25\_02815 and PML25\_02835) where the suspected gene was extremely long (> 3,000 bp), only the first ~600 bp were amplified. PCR products were purified and assembled (Gibson Assembly® Cloning Kit, NEB) into the linearized pAvrBs2-HR vector, upstream to the HR domain, according to the manufacturer directions. Assembly products were initially transformed into NEB 5-alpha competent *E. coli* according to the kit's instructions and grown on LB-Kan plates. The plasmids were then mobilized into Xeu *hrpG\*ΔavrBs2* that constitutively expresses the T3S apparatus and contains a mutation in the *avrBs2* gene, by using pRK2013 as a helper plasmid in triparental mating, as previously described (Figurski and Helinski, 1979). Conjugants were selected on LB-Kan-Gm plates. Presence of the recombinant plasmid was verified in each conjugated recipient by colony PCR using insert specific primers and by Sanger sequencing using the same primers.

### 2.4.2 Translocation

For translocation assays (Rodén et al., 2004), overnight bacterial cultures were suspended in 10 mM MgCl<sub>2</sub> at an optical density of 0.1 (at 600 nm) and infiltrated into the leaves of 7-week-old ECW20R (carrying the *Bs2* gene) pepper using a needleless syringe. Elicitation of HR was monitored at 36 h post-inoculation. For visualization of cell death, leaves were harvested and soaked for 24 h in a bleaching solution (40% ethanol, 40% chloroform, 10% acetic acid), and then transferred to a recovery solution (40% glycerol, 10% ethanol). For each translocation assay, three leaves of at least three pepper plants were infiltrated. Experiments were repeated three times with similar results.

## 3 Results

### 3.1 Genome assembly of *X. hortorum* pv. *pelargonii* strain 305

The genome assembly of Xhp305 was carried out using both long (PacBio Sequel II) and short (Illumina NovaSeq 6000) reads. Using the long reads of PacBio we could close the circular genome,

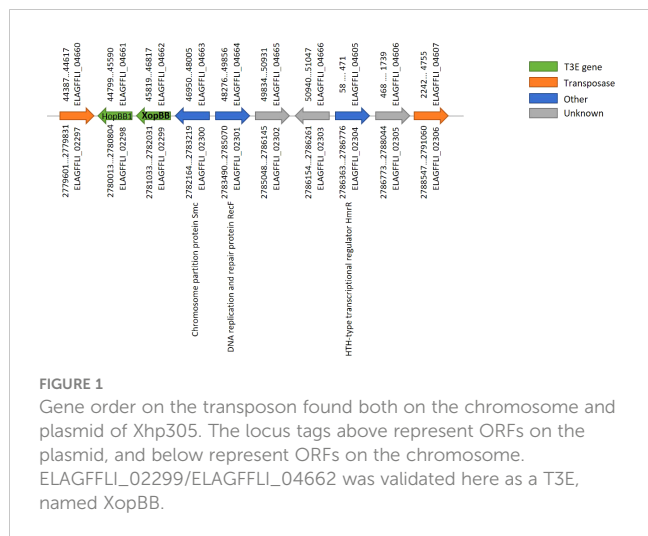
and the short Illumina reads were used to polish the assembly. The assembly resulted with three circular molecules: a chromosome of 5,216,813 bp, and two plasmids of 188,317 bp, and 51,091 bp, with average coverage of 36X, 35X, and 54X, respectively. The assembly polishing using the Illumina reads resulted with a few corrections in the chromosome and in the smaller plasmid. The average coverage of the assembly measured by mapping the Illumina reads was 454X, 467X, and 1,674X for the chromosome and plasmids, respectively. The coverage of the plasmids relative to the chromosome, measured both using the PacBio and the Illumina reads, suggests that the larger plasmid has a copy number of one, while the copy number of the smaller plasmid is either two or three, depending on the coverage of the PacBio versus Illumina, respectively. The average GC content of the chromosome is 0.64, while the average GC contents of the larger and smaller plasmids are 0.59 and 0.62, respectively. The chromosome and plasmids hold 4,357, 191, and 62 coding sequences, respectively. Genome assembly features are available in Table 1. The sequencing data and assembled genome were deposited to NCBI and can be found in BioProject PRJNA926924. The annotation available in NCBI was done using the internal PGAP annotation pipeline of NCBI. It differs from the annotation we obtained using Prokka, mainly in the prediction of translation start sites. The genomic features estimated here are based on the Prokka annotation. Our downstream analysis was done using Prokka annotation, and this annotation is available in the supplementary data.

#### 3.1.1 Detection of recent transposon duplication

Interestingly, an identical DNA segment of 11,818 bp, containing ten ORFs (see Supplementary Table S2), was found both on the chromosome and on the smaller plasmid. This segment is identical by DNA sequence based on the PacBio assembly, and no corrections were introduced within it using the Illumina reads mapping. Based on Prokka annotation, the first and last ORFs in this segment encode for two DNA transposases (IS3 family transposase ISMex7 and Tn3 family transposase ISPa43), which explains this duplication (termed transposon from now on). In

TABLE 1 Features of the assembled Xhp305 genome following sequencing with PacBio and Illumina, assembly with Canu, polish with Pilon, and annotation with Prokka.

Feature	Chromosome	plasmid1	plasmid2
Size (bp)	5,216,813	188,317	51,091
No. of circular contigs	1	1	1
No. of CDSs	4,357	191	62
G+C content (%)	63.8	58.8	62.1
Pac-bio average coverage	36X	35X	54X
Illumina average coverage	454X	467X	1,674X
No. of tRNA genes	54	1	0
No. of T3Es (+ No. of T3Es that were validated here)	35(+4)	0	1(+1)



order to verify the location of the two identical transposons (one chromosomal and one on the plasmid) we performed several PCR reactions on the total DNA prep we had previously sent for sequencing: Primers were designed to produce ~900 bp products covering the junction points between the chromosome/transposon or the plasmid/transposon (see primers table in the supplementary data), assuming that no product would be obtained in the plasmid/transposon combination if the transposon only existed on the chromosome and vice versa. Results of the PCR reaction (Figure S1) clearly show that the identical transposon can be found both on the chromosome and on the plasmid. The PCR products were Sanger-sequenced and found to be identical to the sequences of their respective source; chromosome or plasmid. The lack of point mutations between the two copies of the transposon indicates that the duplication event was very recent, on an evolutionary scale. A BLASTn search of the two transposases at the edges of this transposon, yielded identical hits to the chromosome of other Xhp strains. This suggests that the transposon was originally on the chromosome and was then duplicated to the plasmid of Xhp305. Among the ten ORFs found on the transposon we identified two T3Es, HopBB1 and XopBB. The latter was validated here (see

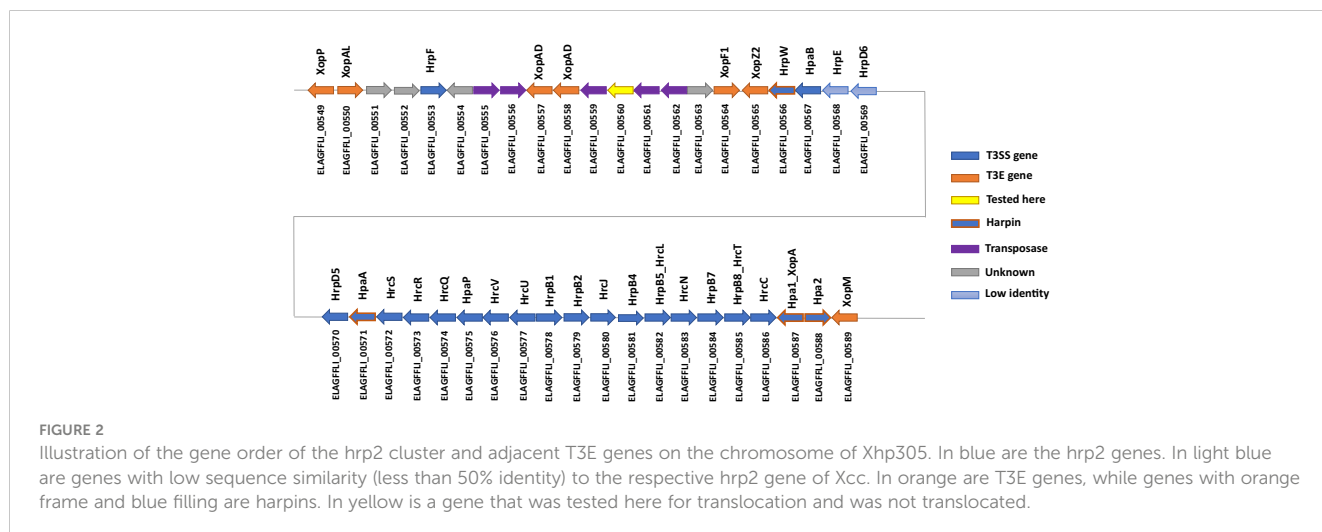
below). Other than the T3Es and the transposases, according to Prokka annotation, it also holds a chromosome partition protein SMC (Strunnikov, 2006), a DNA replication and repair protein, and a HTH-type transcriptional regulator HmrR, which stands for heavy metal-responsive regulator. The order and location in the genome of the ORFs on this segment is illustrated in Figure 1.

### 3.1.2 T3SS and T3Es genes on the genome

The Xhp305 genome possesses a full hrp2 class T3SS on the chromosome, but some of these genes are with a low percentage of identical matches to the respective *X. campestris* pv. *campestris* (Xcc) hrp2 genes that were used as reference. Specifically, ELAGFFLI\_00568 (PML25\_02850) has only 35% identity with Xcc HrpE (AAM40519), and ELAGFFLI\_00569 (PML25\_02855) has only 45% identity with Xcc HrpD6 (AAM40520). The genes order is the same as in the reference T3SS cluster from Xcc. Figure 2 shows the cluster of T3SS genes, adjacent T3E genes and harpins (Choi et al., 2013). Other than the two T3Es on the transposon, all the other T3Es found in this genome are on the chromosome.

### 3.1.3 Xhp305 genome compared to other *Xanthomonas hortorum* genomes

To compare the genome of Xhp305 with other *Xanthomonas hortorum* (Xh) genomes, we downloaded all the fully sequenced genomes of Xh available in NCBI on January 10<sup>th</sup> 2023, and used them, together with our genome of Xhp305, as well as Xfrg genome, as input for M1CR0B1AL1Z3R (Avram et al., 2019). M1CR0B1AL1Z3R is a web server for the analysis of large-scale microbial genomics data. We used default parameter values with the following changes: Maximal e-value cutoff of 10<sup>-4</sup>, Xfrg genome as an outgroup for the phylogeny reconstruction, and bootstrap over the species tree. The following genomes of Xh were analyzed: (1) Xh strain VT106 (accession GCF\_008728175); (2) Xh pv. vitians LM16734 [accession GCF\_014338485 (Morinière et al., 2021)]; (3) Xh pv. vitians strain CFBP498 (accession GCF\_903978195 (Dia et al., 2020)); (4) Xh strain Oregano108 (accession GCF\_026651895); (5) Xh strain jj2001 [accession GCF\_024339125]; (6) Xh pv. gardneri strain JS749-3 [accession

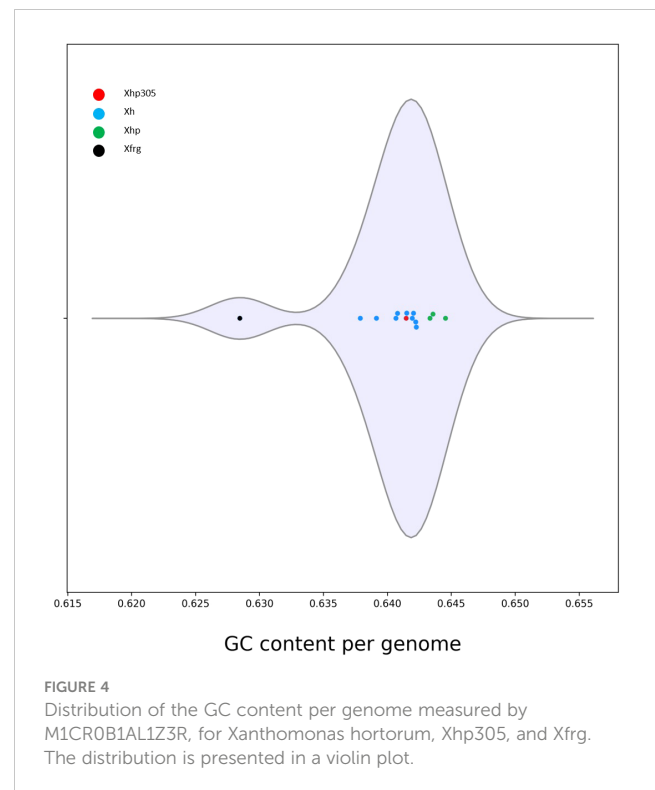
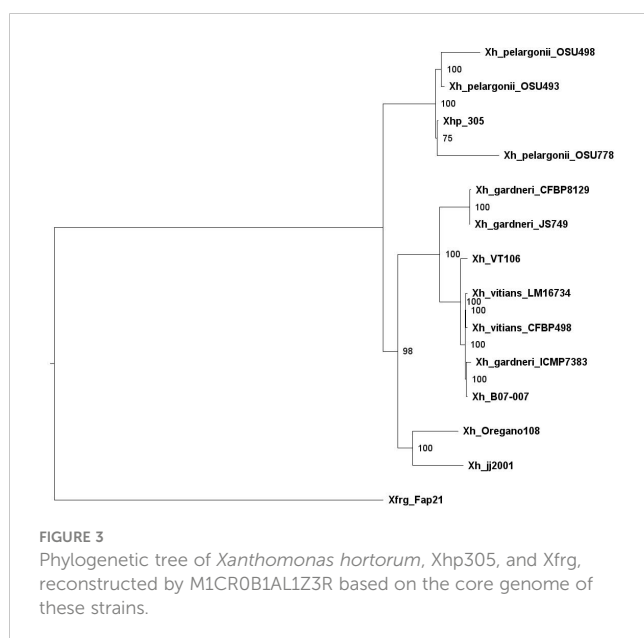


GCF\_001908755 (Richard et al., 2017)]; (7) *Xh* pv. *gardneri* strain ICMP7383 [accession GCF\_001908755 (Richard et al., 2017)]; (8) *Xh* pv. *gardneri* strain CFBP8129 [accession GCF\_903978225 (Dia et al., 2020)]; and (9) *Xh* strain B07-007 (accession GCF\_002285515). The following *Xhp*, other than *Xhp305*, were also included in the analysis: (1) *Xhp* strain OSU778 (accession GCA\_025452115); (2) *Xhp* strain OSU498 (accession GCA\_024498995); and (3) *Xhp* strain OSU493 (accession GCF\_024499015). Together with the strains studied here, *Xhp305* and *Xfrg*, a total of 14 strains were used in this analysis.

Using a minimum identity score of 80%, M1CR0B1AL1Z3R found 5,983 orthologous groups among these 14 genomes. Figure S2 summarizes the frequencies of the orthologous groups' sizes. As can be seen in this figure, of these 5,983 groups, 2,350 groups included genes shared by all genomes, thus defining the core genome. Using this core genome, M1CR0B1AL1Z3R reconstructed the phylogenetic tree (Figure 3). Based on this tree, we infer that the Israeli isolate, *Xhp305*, is evolutionary close to *Xhp* strain OSU778, isolated from a geranium leaf sample in the USA in 2012.

In addition to the phylogeny, the average GC-content of the ORFs in the genomes was evaluated and compared between all the genomes. While the GC content measured for the ORFs of *Xfrg* was lower than that of *X. hortorum* genomes, with an average of 0.628, the GC content measured for *Xhp* genomes was the highest, ranging between 0.643 and 0.645, and the GC content measured for the ORFs of *Xhp305* was slightly lower with an average of 0.641 (Figure 4).

A similar analysis using M1CR0B1AL1Z3R was conducted to compare the plasmids of *Xhp305* with the plasmids of other *X. hortorum* genomes. Apart from *Xhp305*, out of the 12 *X. hortorum* genomes (including three *Xhp* genomes and 9 *X. hortorum* genomes of other pathovars, as listed above), 11 had plasmids and thus were included in this analysis. The *X. hortorum* strain



jj2001 was the only genome without plasmids and was therefore excluded from this analysis. Each of the other genomes had between one and three plasmids. Interestingly, only 5% of the genes found on the larger plasmid of *Xhp* strain 305 had putative orthologous genes on other *pelargonii* strains. In fact, most of the genes on the larger plasmid had orthologs on plasmids of other *X. hortorum* variants (other than the *pelargonii* pathovar). Our observation suggests that this plasmid was acquired by *Xhp305* from another pathovar. In contrast, only 50% of the genes on the smaller plasmid had orthologs on plasmids of other *X. hortorum* genomes.

### 3.2 Effectidor predictions

Before running a machine-learning model, Effectidor searches for ORFs with significant sequence similarity to a database of previously validated T3Es from a large set of organisms (see Methods section 2.3). Another option is to provide a list of positives (i.e., known T3Es) as input, in addition or instead of the internal homology search. This list of positives, if supplied, should be in a FASTA format. For the analysis of *Xfrg* we chose to supply a list of positives instead of the internal search, as the built-in homology search resulted with only half of the known T3Es in this strain, due to high percentage of identity cutoff (70%) that led to missing some of the more distant homologs. The list of positives to consider was supplied by Dr. Doron Teper, accounting for sequence similarity to previously validated effectors from *Xanthomonas*, *Pseudomonas syringae*, *Ralstonia solanacearum*, *Acidovorax*, and *Pantoea* sp., sharing identity lower than 70%, as T3Es. While Effectidor can still build a classifier based on a partial effector list,

providing the full list of T3Es is preferable for better representation of the T3Es genomic organization, providing larger training set, and thus for more accurate predictions. For Xfrg this list of positives included 47 T3Es (Table 2). For Xhp305, Effectidor was executed without providing a list of positives, and in its internal homology search, it yielded 36 T3Es, which we consider as positive samples for training the machine-learning algorithm (Table 2).

In the next step within Effectidor, a machine-learning classifier was trained on the known T3Es found in the first step. Based on the trained classifiers, other than homology to effectors, among the ten most important features we used in these analyses were amino acid similarity to effectors vs. to non-effectors (Figures 5A2, 5B2), and score of the secretion signal in the N-terminal region (Wagner et al., 2022a) with 24 and 41 T3Es out of 36 and 47 in Xhp305 and Xfrg, respectively, with a signal score higher than 0.5 (Figures 5A1, 5B1). Homology to proteins of closely related bacteria without the T3SS was also important with only one and two T3Es in Xfrg and Xhp, respectively, which show high similarity to some of their proteins (Figures 5A3, 5B3). Specifically, the protein encoded by BER92\_03985 of Xfrg showed sequence similarity to OtsA of *Stenotrophomonas* and *Pseudoxanthomonas*, the protein encoded by ELAGFFLI\_03189 (PML25\_15775 + 141bp upstream) of Xhp showed sequence similarity to WQ53\_RS02800, glycoside hydrolase family 30 protein of *Pseudoxanthomonas*, and the protein encoded by ELAGFFLI\_01251 (PML25\_06220) of Xhp showed sequence similarity to OtsA of *Stenotrophomonas* from *Luteimonas* and *Pseudoxanthomonas*. The genomic organization of T3Es and specifically the distance to the closest known T3E on the genome was the next contributing feature, with 28 and 23 T3Es in Xfrg and Xhp, respectively, which were less than 15 ORFs away from a known effector on the genome (Figures 5A4, 5B4). In Xfrg the PIP-box was also important for the prediction, with 14 of the T3Es with a complete PIP-box (Figure 5B5), while in Xhp the GC-content was more important than the PIP-box for the predictions (Figure 5A5). Figure 5 shows the distribution of these features' values among T3Es and non-T3Es in Xhp305 and Xfrg.

Following the training step, the trained classifier was applied to the remaining genes in the given genome, to yield prediction scores, reflecting the likelihood of each of the genes to encode a T3E. This step yielded several high-scoring predictions in each genome (Tables 3, 4). The features of these highly ranked genes reveal that most of them have a high secretion signal score predicted for the N-terminal region of the protein sequence. Many of them have a perfect or nearly perfect PIP-box in their promoter. In addition, some of them show sequence similarity to known T3Es, or reside in proximity to other T3Es on the genome (Tables 3, 4). Of note, the sequence similarity to known effector, when present, was not high enough for these putative T3Es to be considered positive in the previous step, i.e., the identity percentage was less than 70%.

Out of the above predictions, several candidates were chosen for experimental validation. Candidates were selected based on predictions rank, lack of significant sequence similarity to known effectors in *Xanthomonas*, and minimal length (peptides smaller than 75 amino acids were ignored). In addition, two proteins that were identified as T3Es based on homology to previously validated

T3Es were used as positive controls. Of note, additional putative T3Es that we did not validate exist (see discussion).

### 3.3 Four Xhp and two Xfrg predicted T3E proteins are translocated into plant cells via the T3SS

To examine the translocation of predicted effectors, we utilized a reporter system based on the delivery of a truncated form of the Xeu T3E AvrBs2 (amino acids 62–574) into susceptible plant cells. AvrBs2<sub>62–574</sub> lacks a translocation signal, but is sufficient to elicit HR in plants expressing the *Bs2* resistance gene (Rodén et al., 2004). The deleted translocation signal is supplied (or not) by the cloned candidate effector. The conjugant strains we obtained were tested for elicitation of HR in the pepper line ECW20R, which encodes a functional *Bs2*-resistance gene.

Our results show that the following candidates induced the HR 36 h post infection (Figure 6): Xhp conjugants: ELAGFFLI\_04662 (conjugant of PML25\_23150+PML25\_23155)/ELAGFFLI\_02299 (conjugant of PML25\_11405+PML25\_11410), identical genes on the chromosome and plasmid, respectively, named hereafter XopBB; ELAGFFLI\_03194 (PML25\_15800 + 60bp upstream), named hereafter XopBC; ELAGFFLI\_00506 (PML25\_02555 + 252bp upstream), named hereafter XopBD; and ELAGFFLI\_01101 (PML25\_05510 - 249bp upstream), named hereafter XopBE. Xfrg conjugants: BER92\_21920, named XopBF; and BER92\_22150, named XopBG.

Of note, the protein encoded by the gene PML25\_02815 of Xhp contains 15 conserved SKW repeats, previously described in the effector XopAD of Xeu (Teper et al., 2016). It tested negative in the translocation assay, see discussion.

No HR was observed in leaf areas inoculated with Xeu strains expressing the other tested constructs (Tables 3, 4). Xeu expressing ELAGFFLI\_00550 (XopAL) and ELAGFFLI\_00565 (XopZ2), of the latter we cloned only the first 200 N-terminal amino acids, were tested as positive controls and also induced HR on pepper leaves (Figure 6). The parent strain Xeu *hrpG*\*  $\Delta$ *avrBs2* expressing the AvrBs2<sub>62–574</sub>::HA fusion (“empty” vector) was tested on the same pepper leaves, as negative control. As expected, this strain did not cause HR. All in all, four out of six Xhp genes and two out of six Xfrg genes we tested encode proteins that elicited HR response in the pepper line ECW20R, which encodes a functional *Bs2* resistance gene (Figure 6). Thus, all six can be defined as novel T3Es.

### 3.4 Presence of the newly discovered T3E genes in other strains

We next conducted sequence similarity searches to identify the taxonomic distribution of the newly identified T3Es.

The 332 amino acids long XopBB protein from Xhp305 is encoded by two identical genes, on the chromosome and on the smaller plasmid. We searched for homology of XopBB to a list of previously validated T3Es from *Xanthomonas*, *Pseudomonas*

TABLE 2 Known effectors found in (A) Xfrg and (B) Xhp305.

A - Xfrg						B - Xhp			
Locus	Effector family	Score	Locus	Effector family	Score	Locus	NCBI locus	Effector family	Score
BER92_23180	XopP	0.99	BER92_05605	XopAV	0.93	<b>ELAGFFLI_00550</b>	<b>PML25_02765</b>	<b>XopAL</b>	<b>0.99</b>
BER92_17495	XopX	0.99	BER92_12855	BapC	0.90	ELAGFFLI_00564	PML25_02830	XopF1	0.99
BER92_23185	XopP	0.99	BER92_12860	Hpa1	0.85	ELAGFFLI_04661	PML25_23145 <sup>+43</sup>	HopBB1	0.99
BER92_17490	XopF	0.99	BER92_17220	HopBL	0.84	ELAGFFLI_02298	PML25_11400 <sup>+82</sup>	HopBB1	0.99
BER92_23190	XopP	0.99	BER92_23165	HopBL	0.83	ELAGFFLI_00094	PML25_00480	XopAI	0.99
BER92_21950	HopBL	0.99	BER92_02515	XopV	0.82	ELAGFFLI_00549	PML25_02760	XopP	0.99
BER92_06295	XopAD	0.99	BER92_12375	HopBL	0.81	<b>ELAGFFLI_00565</b>	<b>PML25_02835</b>	<b>XopZ2</b>	<b>0.99</b>
BER92_18835	XopP	0.99	BER92_12940	XopA	0.81	ELAGFFLI_00650	PML25_03270	XopX	0.99
BER92_04220	XopL	0.99	BER92_23140	HopBD	0.81	ELAGFFLI_00589	PML25_02955 <sup>+84</sup>	XopM	0.99
BER92_02460	XopAD	0.99	BER92_03985	RipTPS	0.80	ELAGFFLI_03853	PML25_19105 <sup>+36</sup>	XopE2	0.99
BER92_01070	HopAS1	0.99	BER92_07165	XopM	0.80	ELAGFFLI_03195	PML25_15805 <sup>+16</sup>	XopQ	0.99
BER92_00360	HopBL	0.99				ELAGFFLI_03642	PML25_18040 <sup>tr</sup>	HopAK1	0.98
BER92_03355	XopF	0.99				ELAGFFLI_00566	PML25_02840 <sup>-32</sup>	HrpW	0.98
BER92_23155	XopC	0.99				ELAGFFLI_01371	PML25_06830 <sup>-41</sup>	XopD	0.98
BER92_18830	XopP	0.99				ELAGFFLI_00093	PML25_00475	XopAA	0.98
BER92_12975	XopF	0.99				ELAGFFLI_04247	PML25_21105 <sup>+221</sup>	XopAD	0.98
BER92_23085	XopC	0.99				ELAGFFLI_00302	PML25_01490	XopN	0.98
BER92_15380	XopAE	0.99				ELAGFFLI_00558	PML25_02805 <sup>tr</sup>	XopAD	0.98
BER92_11120	XopAG	0.99				ELAGFFLI_00557	PML25_02805 <sup>tr</sup>	XopAD	0.98
BER92_17775	XopN	0.99				ELAGFFLI_03644	PML25_18045	HopAK1	0.98
BER92_12970	XopZ	0.99				ELAGFFLI_03193	PML25_15795	XopAY	0.98
BER92_05755	XopC	0.99				ELAGFFLI_03199	PML25_15820	XopG	0.97
BER92_16595	XopR	0.99				ELAGFFLI_01370	PML25_06820 <sup>+64</sup>	XopK	0.97
BER92_15420	XopK	0.99				ELAGFFLI_03178	PML25_15720 <sup>+475</sup>	XopAM	0.97
BER92_22830	XopAD	0.99				ELAGFFLI_00588	PML25_02950 <sup>+38</sup>	Hpa2	0.97
BER92_05600	XopQ	0.99				ELAGFFLI_00411	PML25_02035	XopB	0.97
BER92_22580	XopE	0.99				ELAGFFLI_00055	PML25_00280	AvrBs2	0.97
BER92_15425	XopAU	0.99				ELAGFFLI_03189	PML25_15775 <sup>+47</sup>	Srfj	0.96
BER92_02195	XopAD	0.99				ELAGFFLI_00571	PML25_02865	HpaA	0.96
BER92_18860	XopB	0.98				ELAGFFLI_01251	PML25_06220	RipTPS	0.95
BER92_15435	XopAF	0.98				ELAGFFLI_01852	-	XopAQ	0.89
BER92_14315	XopAM	0.98				ELAGFFLI_00587	PML25_02945 <sup>+42</sup>	XopA	0.75
BER92_22445	HopBL	0.98				ELAGFFLI_00017	PML25_00090 <sup>-23</sup>	AvrRxv	0.71

(Continued)



TABLE 2 Continued

A - Xfrg			B - Xhp						
Locus	Effector family	Score	Locus	Effector family	Score	Locus	NCBI locus	Effector family	Score
BER92_01845	AvrBs2	0.97				ELAGFFLI_00798	PML25_04005 <sup>+6</sup>	PthG	0.70
BER92_21340	HopBL	0.97				ELAGFFLI_02490	PML25_12355	CigR	0.43
BER92_22450	HopBL	0.96				ELAGFFLI_03643	PML25_18040 <sup>tr</sup>	HopAK1	0.39

In bold are positive controls used for the translocation assay. "Score" is the prediction score given by Effectidor, reflecting the likelihood of this gene to encode a T3E. For Xhp305 genes, both Prokka and PGAP (NCBI) annotation locus tags are mentioned. The difference between the two annotations is in superscript, such that (-) and (+) mean the protein is shorter and longer than Prokka annotation, respectively. All these differences are in the N-terminus. Irregular differences ("tr" in superscript): ELAGFFLI\_03642 respective PGAP annotation is with a start codon upstream to Prokka annotation and is a pseudogene with a frameshift; ELAGFFLI\_00558 and ELAGFFLI\_00557 are both part of the respective PGAP annotation; ELAGFFLI\_01852 is not found in PGAP annotation; ELAGFFLI\_03643 is part of its respective PGAP annotation.

*syringae*, *Ralstonia solanacearum*, *Acidovorax*, and *Pantoea* sp, supplied by Dr. Doron Teper (available in the supplementary data), using BLASTp. The best hit was to APS58\_0178 of *Acidovorax citrulli* M6 that we have previously reported as a putative T3E based on sequence similarity to HopF2 (Jiménez-Guerrero et al., 2020). The alignment between XopBB and APS58\_0178 shared 50% identical matches on 70% coverage. In a regular BLASTp search, closer putative homologs were found in *X. hortorum*, *X. campestris*, and *X. hydrangea* (Table 5). All these inferred homologs are annotated as hypothetical proteins.

The Xhp305 T3E XopBC shares some sequence similarity with XopAV (Teper et al., 2016) and XopAY (Yang et al., 2015); The validated XopAV from Xeu is a protein of 165 amino acids. In contrast, XopBC is 249 amino acids long. The pairwise alignment between these two proteins is only between the 49 most N-terminal amino acids of XopBC, and a region near the C-terminus of XopAV, where they share only 50% identity. In contrast, XopBC shares 52% identity over 92% coverage with XopAY, which is

encoded by a gene adjacent to the gene encoding for XopBC. We therefore hypothesize that XopBC and XopAY are two paralogs. XopBC was annotated as XopAV in the PGAP annotation of NCBI. Nevertheless, since the sequence similarity observed between XopBC and the validated XopAV is between the N-terminal region of XopBC, which is expected to hold the translocation signal, and the C-terminal region of XopAV, which is expected to hold the active part of the effector, and since both of these effectors were found to encode an active translocation signal which enabled them to be translocated into pepper leaves in the translocation assay, we hypothesize that these are two different effectors, and that XopBC is a newly identified effector. A BLAST search using XopBC as query reveals the presence of putative homologs of this protein in *X. hortorum*, *X. campestris*, *X. arboricola*, *X. hydrangea*, and *X. codiae* (Table 5). These proteins are annotated as XopAV, based on a sequence similarity similar to the one we found between XopBC and XopAV. These proteins share a higher sequence similarity with XopBC than with the validated XopAV from Xeu, and with the

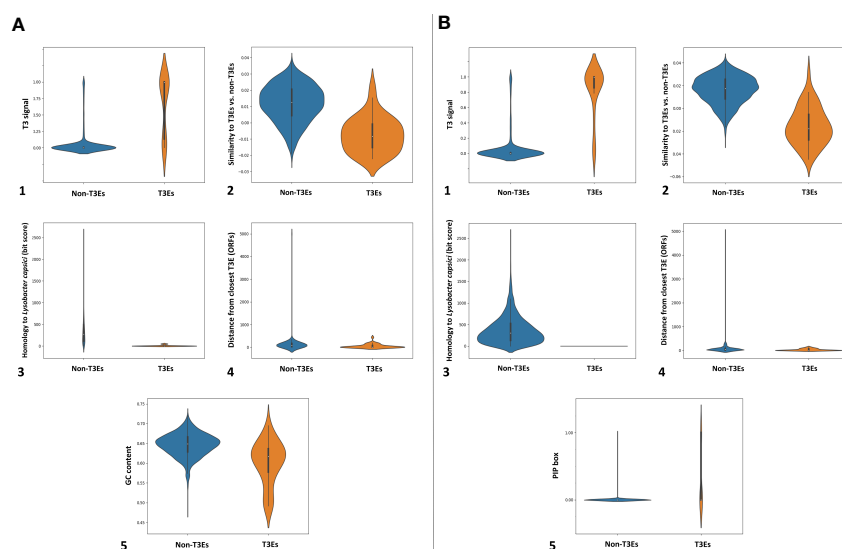


FIGURE 5

Distribution of informative feature values for T3Es and non-T3Es in Xhp305 (A) and Xfrg (B), as analyzed by Effectidor. The distributions are presented in violin plots.

TABLE 3 Top T3Es predictions and translocation assay results in Xfrg, with informative features values.

Locus	Score	Protein length	PIP-box	Secretion signal	Distance to closest T3E	Bit score to T3Es (T3E name)	Translocated
BER92_11960	0.83	556	V	0.475	80	214 (NopAC)	not tested
BER92_12965	0.77	371	-	0.997	1	407 (HrpW)	not tested
BER92_02770	0.72	465	V	0.017	50	179 (NopAC)	not tested
BER92_22860	0.61	134	V	0.995	111	0	not tested
BER92_22075	0.58	282	V	0.996	272	0	no
<b>BER92_21920 (XopBF)</b>	<b>0.57</b>	<b>163</b>	-	<b>0.990</b>	<b>18</b>	<b>81 (XopAV)</b>	<b>yes</b>
BER92_12945	0.57	311	V	0.065	1	140 (HrpW)	not tested
BER92_12955	0.56	90	-	0.781	3	0	not tested
BER92_22365	0.56	101	-	0.997	48	0	no
BER92_19605	0.56	142	V	0.929	1	0	not tested
BER92_21675	0.55	96	-	0.998	6	0	not tested
BER92_17025	0.54	192	-	0.981	2	0	not tested
<b>BER92_22150 (XopBG)</b>	<b>0.52</b>	<b>158</b>	<b>V</b>	<b>0.858</b>	<b>7</b>	<b>0</b>	<b>yes</b>
BER92_18820	0.52	165	V	0.973	2	0	not tested
BER92_18825	0.52	188	-	0.042	1	0	not tested
BER92_02200	0.52	157	-	0.002	1	83 (XopAD)	no
:	:	:	:	:	:	:	:
BER92_19140	0.40	338	-	0.991	51	0	no

PIP-box: "V"-perfect PIP-box is found in the gene promoter; "-"-lack of PIP-box in the gene promoter. Score: a score given by the classifier, reflecting the likelihood of the gene to encode a T3E. Secretion signal: a score reflecting the likelihood of an existence of a type 3 secretion signal within the protein sequence. Distance to closest T3E: number of genes on the genome separating between the given gene and a known T3E. Bit score to T3Es: alignment score of the best sequence similarity found to a known T3E within Effector's T3Es database. In bold are the newly identified translocated T3Es.

above results we suggest that they should not be annotated as XopAV, but rather as XopBC.

XopBD of Xhp305 did not have hits to any of the validated T3Es from *Xanthomonas*, *Pseudomonas syringae*, *Ralstonia*, *Acidovorax*, and *Pantoea* sp. It has some sequence similarity (60% identity over 69% coverage) to a protein from *Xanthomonas campestris* pv. raphani 756C, a pathogen of the plant model organism *Arabidopsis thaliana*, which was previously suggested as a T3E candidate XopAT, based on the presence of a PIP box and a -10 box-like sequence upstream of the coding sequence, low GC content, and eukaryotic motifs (Bogdanove et al., 2011). It has not been validated yet, and its function is unknown. ORFs with higher sequence similarity to XopBD were found in *X. hortorum*, *X. arboricola*, *X. cucurbitae*, *X. codiae*, and *X. campestris* (Table 5). All these putative homologs are annotated as hypothetical proteins.

XopBE of Xhp305 shows distant sequence similarity (47% identity) to XopC1 (Noél et al., 2003). Putative closer homologs were found in *X. hortorum* and Xfrg (Table 5). The proteins found in *X. hortorum* are annotated as hypothetical protein whereas in Xfrg they are annotated as hydrolase-like protein or hypothetical protein.

Multiple sequence alignments (MSAs), of the abovementioned T3Es and their homologs from *Xanthomonas*, produced by Clustal

Omega (Sievers and Higgins, 2014; Madeira et al., 2022) using default parameter values, are available in the supplementary data. All the homologs listed in Table 5 and used to produce the MSAs share at least 70% identity and 70% coverage with the respective T3E.

Of the six effector candidates tested in Xfrg, two tested positive in the translocation assay: BER92\_21920 (XopBF) and BER92\_22150 (XopBG).

XopBF is a hypothetical protein with unknown function. It shares some sequence similarity (identity of ~40%) with the proteins homologous to our newly validated XopBC ("XopAV", see above) of several strains of *Xanthomonas*, among which are: *X. hortorum* (WP\_159087131, WP\_176339450, WP\_180336534, WP\_168958006, WP\_152025508, WP\_268212485), *X. campestris* (WP\_228439322, WP\_169705357, WP\_273676157), *X. arboricola* (WP\_212583737, WP\_080591365, WP\_104562523), *X. oryzae* (WP\_019303846, WP\_113343065, WP\_075244353, WP\_044757351, WP\_027704160, WP\_047339610, WP\_041183112, WP\_240113023, WP\_029217345, WP\_113221815, WP\_113335989, WP\_113000154, WP\_069963882), *X. codiae* (WP\_104539725), and *X. prunicola* (WP\_101363523). Proteins with higher sequence similarity are

TABLE 4 Top T3Es predictions and translocation assay results in Xhp.

Locus	Score	Protein length	PIP-box	Secretion signal	Distance to closest T3E	Bit score to T3Es (T3E name)	Translocated
<b>ELAGFFLI_02299</b> <b>ELAGFFLI_04662 (PML25_11405+ PML25_11410/ PML25_23150+ PML25_23155) (XopBB)</b>	<b>0.98</b>	<b>332</b>	<b>V</b>	<b>0.999</b>	<b>1</b>	<b>238 (APS58_0178)</b>	<b>yes</b>
ELAGFFLI_00330 (PML25_01630)	0.96	372	V	1.000	28	259 (XopR)	not tested
<b>ELAGFFLI_03194 (PML25_15800<sup>20</sup>) (XopBC)</b>	<b>0.94</b>	<b>249</b>	<b>V</b>	<b>1.000</b>	<b>1</b>	<b>201 (XopAV)</b>	<b>yes</b>
ELAGFFLI_00560 (PML25_02815)	0.94	990	-	0.041	2	755 (XopAD)	no
<b>ELAGFFLI_00506 (PML25_02555<sup>84</sup>) (XopBD)</b>	<b>0.91</b>	<b>304</b>	<b>V</b>	<b>0.975</b>	<b>43</b>	<b>225 (XopAT)</b>	<b>yes</b>
ELAGFFLI_01763 (PML25_08750)	0.85	556	V	0.762	88	214 (NopAC)	no
ELAGFFLI_00820 (PML25_04110)	0.70	468	V	0.006	22	181 (NopAC)	not tested
<b>ELAGFFLI_01101 (PML25_05510<sup>83</sup>) (XopBE)</b>	<b>0.64</b>	<b>337</b>	<b>-</b>	<b>0.644</b>	<b>149</b>	<b>294 (XopC)</b>	<b>yes</b>
ELAGFFLI_04382 (PML25_21815 <sup>-3</sup> )	0.57	525	mm	1.000	47	188 (AvrB4-1)	not tested
ELAGFFLI_01276 (PML25_06350)	0.54	251	mm	0.677	25	183 (XopH)	not tested
ELAGFFLI_00092 (PML25_00470 <sup>91</sup> )	0.52	156	-	0.155	1	164 (APS58_0178)	not tested

PIP-box: "V"-perfect PIP-box is found in the gene promoter; "-"-lack of PIP-box in the gene promoter; "mm"-PIP-box with one mismatch is found in the gene promoter. Score: a score given by the classifier, reflecting the likelihood of the gene to encode a T3E. Secretion signal: a score reflecting the likelihood of an existence of a type 3 secretion signal within the protein sequence. Distance to closest T3E: number of genes on the genome separating between the given gene and a known T3E. Bit score to T3Es: alignment score of the best sequence similarity found to a known T3E within Effector's T3Es database. In bold are the newly identified translocated T3Es. Both Prokka and PGAP (NCBI) annotation locus tags are mentioned. The difference between the two annotations is in superscript, such that (-) and (+) mean the protein is shorter and longer than Prokka annotation, respectively. All these differences are in the N-terminus. Irregular differences between the two annotations: XopBB, encoded by ELAGFFLI\_02299 and ELAGFFLI\_04662 on the chromosome and plasmid, respectively, was broken into two ORFs in PGAP annotation, PML25\_11405+ PML25\_11410/ PML25\_23150+ PML25\_23155, such that the first ORF in each pair is annotated as a pseudogene lacking a C-terminus. In bold are the newly identified translocated T3Es.

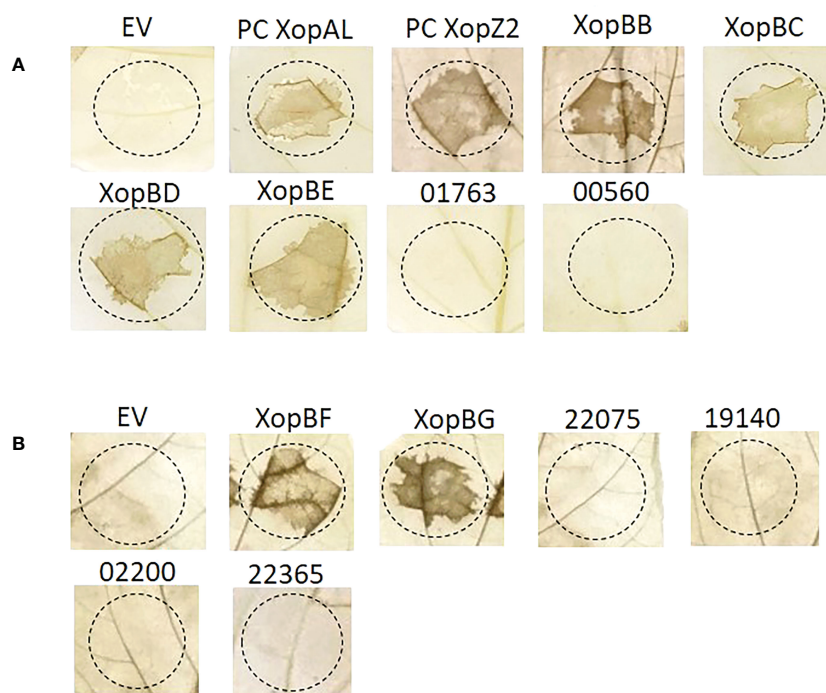


FIGURE 6

Translocation assay for predicted effectors in Xhp (A) and Xfrg (B). Xeu hrpG\* $\Delta$ avrBs2 bacteria were introduced with the indicated putative effectors of Xhp305 and Xfrg, fused to the HR domain of AvrBs2. Overnight cultures were infiltrated into leaves of pepper ECW20R var, which carries the Bs2 resistant gene. Leaves were harvested 48 h later, bleached in a bleaching solution and photographed. EV, empty vector; PC, positive control.

TABLE 5 Presence of the newly discovered T3Es in *Xanthomonas*.

T3E	<i>X. cucurbitae</i>	<i>X. hydrangeae</i>	<i>X. arboricola</i>	<i>X. campestris</i>	<i>X. codiae</i>	<i>X. hortorum</i>	<i>X. fragariae</i>
XopBB	-	WP_232372966	-	WP_256267875 WP_227971037 WP_221284749 WP_228434360 WP_274340180 WP_274508026 WP_164493567 WP_116891014	-	WP_180313544 WP_268215251 WP_268215250 WP_168959150 WP_273664738	-
XopBC	-	WP_210762087 WP_275415019 WP_275415023	WP_080591365 WP_212583737 WP_104562523	WP_273676157 WP_228439322 WP_169705357	WP_104539725	WP_268212485 WP_176339450 WP_168958006 WP_159087131 WP_152025508	-
XopBD	WP_158251449 WP_274396588	-	WP_126750625 WP_153064930	WP_227971169 WP_228878756 WP_139328362 WP_146011555 WP_274340431	WP_146091904	WP_168959089 WP_233366577 WP_251762182	-
XopBE	-	-	-	-	-	WP_168960092 WP_211317318	WP_269122454 WP_208587355 WP_197493698
XopBF	-	-	-	-	-	-	WP_134656590 WP_269125207 WP_159088880
XopBG	-	-	-	-	-	-	WP_159087733 WP_159088353

annotated as hypothetical proteins and are restricted to Xfrg (Table 5).

Finally, the gene product of BER92\_22150, XopBG, has various putative homologs, all restricted to sub-strains of Xfrg (Table 5). These are all hypothetical proteins with unknown function. Surprisingly, no homologs were detected in other *Xanthomonas* strains. The only other putative homology found, to some degree (coverage of 99% and identity of 45%), is a hypothetical protein from *Xylophilus ampelinus* (WP\_146228602), a grapevines pathogen that encodes a T3SS. This pathogen was previously termed *Xanthomonas ampelina*. It should be noted that this newly identified effector emphasizes the power of Effectidor in discovering novel T3Es without any sequence similarity to known effectors.

## 4 Discussion

The goal of this work was to identify and validate novel T3Es using the Effectidor web server. We selected two pathogens, Xhp and Xfrg, for which we had DNA samples. These species are not extensively studied, and we thus hypothesized there may be unknown effectors within them. To this end, we first sequenced an Israeli isolate of Xhp – strain 305 and assembled its genome. Analysis of the obtained genome revealed a recently duplicated transposon between the chromosome and the smaller of the two plasmids. Moreover, one of the newly validated T3Es was found on this transposon, both on the chromosome and on the plasmid.

We next applied Effectidor on our newly assembled Xhp305 genome, as well as on Xfrg Fap21 genome, to find putative novel T3Es within them. We tested six candidates in each of these genomes. In Xfrg and Xhp we showed that two of the six and four of six candidates, respectively, were translocated and elicited HR on pepper leaves. Interestingly, one of the two T3Es we validated in Xfrg (XopBG encoded by BER92\_22150) was found to be unique to Xfrg and showed no sequence similarity to any of the previously identified T3Es. As XopBG is restricted to Xfrg, it is possible that it plays a significant role in Xfrg pathogenicity and host specificity. It would be interesting to further study its structure and molecular function within its native host.

Effectidor combines dozens of features for the learning and prediction, none of which is capable to fully differentiate between effectors and non-effectors by itself. Among these features are sequence similarity to known T3Es, amino acid composition, proximity to effectors on the genome, existence of regulatory elements such as the PIP-box in the promoter, and prediction of the secretion signal in the N-terminal region. While effectors tend to cluster together on the genome in pathogenicity islands (Marcelletti and Scortichini, 2015), with ~60% of the T3Es residing in proximity of up to 15 ORFs from another T3E, there are also non-effectors in proximity to known T3Es. Thus, predictions based on proximity alone will miss some T3Es and will yield many false positives. Similarly, the PIP-box was found to be the regulatory motif to which HrpG/HrpX transcription regulators bind to regulate the expression of the T3SS and effector genes in *Xanthomonas* (Koebnik et al., 2006), yet we found a PIP-box in the promoters

of only 38% of the T3Es, while it was found also in 9% of the non-T3Es. The translocation signal prediction is an informative feature, with 78% of the T3Es with a score higher than 0.5, but some non-T3Es also have a score higher than 0.5. Thus, prediction based on this feature alone will lead to ~20% precision, which is far from optimal. By combining these and additional features in a machine-learning classification algorithm, Effectidor predicts effectors, in a way that could not be achieved by using any of the features separately.

We validated putative effectors using a truncated *avrBs2* reporter gene, which has a functional HR domain but lacks a translocation signal. Candidate genes were cloned upstream to the truncated *avrBs2* domain, assuming that genuine effectors would supply the translocation signal, and thus elicit HR on pepper leaves. Nevertheless, not all T3Es have a strong enough translocation signal and some require the assistance of chaperones for translocation. Furthermore, expressing the candidate on a plasmid, in a strain other than its original strain could mean that these were not the optimal conditions for the effector to be translocated. Thus, a negative result in this assay does not necessarily rule out the possibility that these candidates act as T3Es in the original pathogen they were isolated from. This may be the case with gene ELAGFFLI\_00560 of Xhp, which tested negative in the translocation assay. This protein contains 15 conserved SKW repeats, previously described in the effector XopAD of Xeu (Teper et al., 2016). Its predicted secretion signal score based on the annotated ORF was only 0.041. Since the prediction of ORFs occasionally suffers from mis-annotation of the start codon, we cloned this gene from an alternative start codon, 48 bp upstream to the predicted start codon. The predicted secretion signal score of this alternative N-terminus was 0.83. Nevertheless, it was not translocated in our system. This again raises the possibility that ELAGFFLI\_00560 requires assistance of a chaperone for translocation, which was absent in the Xeu system under the given conditions.

In this work we tested six candidates of each of the two pathogens, but according to the predictions of Effectidor, additional putative T3Es exist. Candidates for validation in this work were chosen based on prediction score and features such as lack of significant sequence similarity to previously validated T3Es, yet additional ORFs follow this rule. In Tables 3 and 4 are listed putative T3Es that were not tested. These candidates include BER92\_11960, BER92\_12965, BER92\_02770, BER92\_22860, BER92\_12945, BER92\_12955, BER92\_19605, BER92\_21675, BER92\_17025, BER92\_18820, and BER92\_18825 in Xfrg Fap21, and ELAGFFLI\_00330 (putative XopR), ELAGFFLI\_00820, ELAGFFLI\_04382 (putative XopAH/AvrB), ELAGFFLI\_01276 (putative hopD2, based on Prokka annotation), and ELAGFFLI\_00092 in Xhp305.

Identifying the T3E repertoire of a bacterial pathogen is a first step towards understanding the pathogen-host interaction at the molecular level. Open questions for further research include: (1) Validating the additional putative T3Es identified by Effectidor in both Xfrg21 and Xhp305; (2) Understanding how the T3Es are regulated within the bacteria; (3) Understanding their secretion signal; (4) Finding whether their translocation depends on specific chaperones; (5) Determining the order of their translocation into the host; (6) Finding their functions within the host cell, which include

discovering their interaction with host molecules and among themselves. We hope that computational tools, including machine-learning, in the future, can help accelerate discoveries towards such a detailed understanding of the molecular pathways involved in the pathogenicity.

## Data availability statement

The sequencing data and assembled genome were deposited to NCBI and can be found in BioProject PRJNA926924.

## Author contributions

NW, DBM, DT and TP conceived the project. DBM prepared Xhp DNA for sequencing and performed all cloning and expression of effector candidates. DBM and DT executed translocation assays. NW assembled Xhp305 genome and performed all computational analysis. NW and DBM wrote the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

Israel Science Foundation (ISF) [2818/21 to TP]. NW was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel Aviv University.

## Acknowledgments

Israel Science Foundation (ISF) [2818/21 to T.P.]; NW was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel Aviv University; TP's research is supported in part by the Edouard Seroussi Chair for Protein Nanobiotechnology, Tel Aviv University. We thank Dr. Shulamit Manulis and Dr. Joël Pothier who provided us with the Xhp305 bacteria and the Xfrg DNA, respectively. TP would like to thank Prof. Jeff Chang from Oregon State University for hosting him during a sabbatical and for numerous discussions on plant-pathogen interactions.

## 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.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1155341/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

## References

- An, S. Q., Potnis, N., Dow, M., Vorhölter, F. J., He, Y. Q., Becker, A., et al. (2019). Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. *FEMS Microbiol. Rev.* 44, 1–32. doi: 10.1093/femsre/fuz024
- Avram, O., Rapoport, D., Portugez, S., and Pupko, T. (2019). M1CR0B1AL1Z3R—a user-friendly web server for the analysis of large-scale microbial genomics data. *Nucleic Acids Res.* 47, W88–W92. doi: 10.1093/NAR/GKZ423
- Balaž, J., Ivanović, Ž., Davidović, A., Iličić, R., Janse, J., and Popović, T. (2016). Characterization of *Xanthomonas hortorum* pv. pelargonii isolated from geranium in Serbia. *Plant Dis.* 100, 164–170. doi: 10.1094/PDIS-03-15-0295-RE
- Barel, V., Chalupowicz, L., Barash, I., Sharabani, G., Reuven, M., Dror, O., et al. (2015). Virulence and in planta movement of *Xanthomonas hortorum* pv. pelargonii are affected by the diffusible signal factor (DSF)-dependent quorum sensing system. *Mol. Plant Pathol.* 16, 710–723. doi: 10.1111/MPP.12230/SUPPINFO
- Bogdanove, A. J., Koebnik, R., Lu, H., Furutani, A., Angiuoli, S. V., Patil, P. B., et al. (2011). Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic xanthomonas spp. *J. Bacteriol.* 193, 5450–5464. doi: 10.1128/JB.05262-11
- Burstein, D., Amaro, F., Zusman, T., Lifshitz, Z., Cohen, O., Gilbert, J. A., et al. (2016). Genomic analysis of 38 *Legionella* species identifies large and diverse effector repertoires. *Nat. Genet.* 48, 167–175. doi: 10.1038/ng.3481
- Burstein, D., Satanower, S., Simovitch, M., Belnik, Y., Zehavi, M., Yerushalmi, G., et al. (2015). Novel type III effectors in *Pseudomonas aeruginosa*. *MBio* 6, e00161-15. doi: 10.1128/mBio.00161-15
- Burstein, D., Zusman, T., Degtyar, E., Viner, R., Segal, G., and Pupko, T. (2009). Genome-scale identification of *Legionella pneumophila* effectors using a machine learning approach. *PLoS Pathog.* 5, e10000508. doi: 10.1371/journal.ppat.1000508
- Choi, M. S., Kim, W., Lee, C., and Oh, C. S. (2013). Harpins, multifunctional proteins secreted by gram-negative plant-pathogenic bacteria. *Mol. Plant Microbe Interact.* 26, 1115–1122. doi: 10.1094/MPMI-02-13-0050-CR
- Crossman, L. C., Gould, V. C., Dow, J. M., Vernikos, G. S., Okazaki, A., Sebahia, M., et al. (2008). The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol.* 9, R74.1–R74.13. doi: 10.1186/GB-2008-9-4-R74
- Cunnac, S., Boucher, C., and Genin, S. (2004). Characterization of the cis-acting regulatory element controlling HrpB-mediated activation of the type III secretion system and effector genes in *Ralstonia solanacearum*. *J. Bacteriol.* 186, 2309–2318. doi: 10.1128/JB.186.8.2309-2318.2004
- de Bruijn, I., Cheng, X., de Jager, V., Expósito, R. G., Watrous, J., Patel, N., et al. (2015). Comparative genomics and metabolic profiling of the genus *Lysobacter*. *BMC Genomics* 16, 991. doi: 10.1186/S12864-015-2191-Z
- Dia, N. C., Rezzonico, F., Smits, T. H. M., and Pothier, J. F. (2020). Complete or high-quality draft genome sequences of six *Xanthomonas hortorum* strains sequenced with short- and long-read technologies. *Microbiol. Resour. Announc.* 9, e00828-20. doi: 10.1128/MRA.00828-20
- Fernández-Pavía, S. P., Rodríguez-Alvarado, G., Garay-Serrano, E., and Cárdenas-Navarro, R. (2014). First report of *Xanthomonas fragariae* causing angular leaf spot on strawberry plants in México. *Plant Dis.* 98, 682. doi: 10.1094/PDIS-07-13-0691-PDN
- Figurski, D. H., and Helinski, D. R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Natl. Acad. Sci. U. S. A.* 76, 1648–1652. doi: 10.1073/PNAS.76.4.1648
- Gubler, W. D., Feliciano, A. J., Bordas, A. C., Civerolo, E. C., Melvin, J. A., and Welch, N. C. (2007). First report of blossom blight of strawberry caused by *Xanthomonas fragariae* and *Cladosporium cladosporioides* in California. *Plant Dis.* 83, 400. doi: 10.1094/PDIS.1999.83.4.400A
- Henry, P. M., and Leveau, J. H. J. (2016). Finished genome sequences of *Xanthomonas fragariae*, the cause of bacterial angular leaf spot of strawberry. *Genome Announc.* 4, e01271–e01216. doi: 10.1128/GENOMEA.01271-16
- Hou, L., Jiang, J., Xu, Z., Zhou, Y., and Leung, F. C. C. (2015). Complete genome sequence of *Pseudoxanthomonas suwonensis* strain J1, a cellulose-degrading bacterium isolated from leaf- and wood-enriched soil. *Genome Announc.* 3, e00614-15. doi: 10.1128/GENOMEA.00614-15
- Hunt, M., De Silva, N., Otto, T. D., Parkhill, J., Keane, J. A., and Harris, S. R. (2015). Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* 16, 1–10. doi: 10.1186/S13059-015-0849-0/FIGURES/3
- 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 Bioinf.* 11, 119. doi: 10.1186/1471-2105-11-119
- 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
- Jiménez-Guerrero, I., Pérez-Montaño, F., Da Silva, G. M., Wagner, N., Shkedy, D., Zhao, M., et al. (2020). Show me your secret(ed) weapons: a multifaceted approach reveals a wide arsenal of type III-secreted effectors in the cucurbit pathogenic bacterium *Acidovorax citrulli* and novel effectors in the *Acidovorax* genus. *Mol. Plant Pathol.* 21, 17–37. doi: 10.1111/mpp.12877
- Kamangar, S. B., Van Vaerenbergh, J., Kamangar, S., and Maes, M. (2017). First report of angular leaf spot on strawberry caused by *Xanthomonas fragariae* in Iran. *Plant Dis.* 101, 1031. doi: 10.1094/PDIS-11-16-1659-PDN
- Kearney, B., and Staskawicz, B. J. (1990). Widespread distribution and fitness contribution of *Xanthomonas campestris* avirulence gene *avrBs2*. *Nature* 346, 385–386. doi: 10.1038/346385A0
- Kennedy, B. W., and King, T. H. (1962). Angular leafspot of strawberry caused by *xanthomonas fragariae* sp. nov. *Phytopathology* 52, 873–875.
- Koebnik, R., Krüger, A., Thieme, F., Urban, A., and Bonas, U. (2006). Specific binding of the *Xanthomonas campestris* pv. vesicatoria AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *J. Bacteriol.* 188, 7652–7660. doi: 10.1128/JB.00795-06
- 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
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5, 1–9. doi: 10.1186/GB-2004-5-2-R12/FIGURES/3
- 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
- Lifshitz, Z., Burstein, D., Peeri, M., Zusman, T., Schwartz, K., Shuman, H. A., et al. (2013). Computational modeling and experimental validation of the *Legionella* and *Coxiella* virulence-related type-IVB secretion signal. *Proc. Natl. Acad. Sci. U. S. A.* 110, E707–E715. doi: 10.1073/pnas.1215278110
- Lifshitz, Z., Burstein, D., Schwartz, K., Shuman, H. A., Pupko, T., and Segal, G. (2014). Identification of novel *Coxiella burnetii* lcm/Dot effectors and genetic analysis of their involvement in modulating a mitogen-activated protein kinase pathway. *Infect. Immun.* 82, 3740–3752. doi: 10.1128/IAI.01729-14
- Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., et al. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50, W276–W279. doi: 10.1093/NAR/GKAC240

### SUPPLEMENTARY FIGURE 2

Orthologous genes group size distribution, among the 13 *X. hortorum* genomes and the Xfrg genome, as found by M1CR0B1AL1Z3R. Group of size k means the ortholog was found in k of the 14 genomes. This figure is an output of M1CR0B1AL1Z3R.

- Marcelletti, S., and Scortichini, M. (2015). Comparative genomic analyses of multiple *Pseudomonas* strains infecting *Corylus avellana* trees reveal the occurrence of two genetic clusters with both common and distinctive virulence and fitness traits. *PLoS One* 10, e0131112. doi: 10.1371/journal.pone.0131112
- Marques, M. V., Da Silva, A. M., and Gomes, S. L. (2001). Genetic organization of plasmid pXF51 from the plant pathogen *Xylella fastidiosa*. *Plasmid* 45, 184–199. doi: 10.1006/PLAS.2000.1514
- Matthews-Berry, S. S., and Reed, P. J. (2009). Eradication of the first outbreak of *Xanthomonas fragariae* in the united kingdom. *EPPO Bull.* 39, 171–174. doi: 10.1111/J.1365-2338.2009.02284.X
- Mazzucchi, U., Alberghina, A., and Dalli, A. (1973). Occurrence of *Xanthomonas fragariae* Kennedy et king in Italy. *Phytopathol. Z.* 76, 367–370. doi: 10.1111/j.1439-0434.1973.tb02680.x
- McGeachan, J. K., and Fahy, P. C. (1976). Angular leaf spot of strawberry, *Xanthomonas fragariae*: first record of its occurrence in Australia, and attempts to eradicate the disease. *Aust. Plant Pathol. Soc Newsl.* 5, 57–59. doi: 10.1071/APP9760057
- Michiels, T., and Cornelis, G. R. (1991). Secretion of hybrid proteins by the *Yersinia* yop export system. *J. Bacteriol.* 173, 1677–1685. doi: 10.1128/jb.173.5.1677-1685.1991
- Morinière, L., Lecomte, S., Gueguen, E., and Bertolla, F. (2021). *In vitro* exploration of the *Xanthomonas hortorum* pv. vitians genome using transposon insertion sequencing and comparative genomics to discriminate between core and contextual essential genes. *Microb. Genomics* 7, 546. doi: 10.1099/MGEN.0.000546
- Nissan, G., Gershovits, M., Morozov, M., Chalupowicz, L., Sessa, G., Manulis-Sasson, S., et al. (2018). Revealing the inventory of type III effectors in *Pantoea agglomerans* gall-forming pathogens using draft genome sequences and a machine-learning approach. *Mol. Plant Pathol.* 19, 381–392. doi: 10.1111/mpp.12528
- Noël, L., Thieme, F., Gäubler, J., Büttner, D., and Bonas, U. (2003). XopC and XopJ, two novel type III effector proteins from *Xanthomonas campestris* pv. vesicatoria. *J. Bacteriol.* 185, 7092–7102. doi: 10.1128/JB.185.24.7092-7102.2003
- Puławska, J., Warabieda, W., Pothier, J. F., Gétaz, M., and van der Wolf, J. M. (2020). Transcriptome analysis of *Xanthomonas fragariae* in strawberry leaves. *Sci. Rep.* 10, 1–10. doi: 10.1038/S41598-020-77612-Y
- Richard, D., Boyer, C., Lefeuvre, P., Canteros, B. I., Beni-Madhu, S., Portier, P., et al. (2017). Complete genome sequences of six copper-resistant *Xanthomonas* strains causing bacterial spot of solanaceous plants, belonging to *X. gardneri*, *X. euvesicatoria*, and *X. vesicatoria*, using long-read technology. *Genome Announc.* 5, e01693-16. doi: 10.1128/GENOMEA.01693-16
- Roden, J. A., Belt, B., Ross, J. B., Tachibana, T., Vargas, J., and Mudgett, M. B. (2004). A genetic screen to isolate type III effectors translocated into pepper cells during *Xanthomonas* infection. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16624–16629. doi: 10.1073/pnas.0407383101
- Ruano-Gallego, D., Sanchez-Garrido, J., Kozik, Z., Núñez-Berruero, E., Cepeda-Molero, M., Mullineaux-Sanders, C., et al. (2021). Type III secretion system effectors form robust and flexible intracellular virulence networks. *Sci. (80- ).* 371, eabc9531. doi: 10.1126/science.abc9531
- Ryan, R. P., Vorhölter, F. J., Potnis, N., Jones, J. B., Van Sluys, M. A., Bogdanove, A. J., et al. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* 9, 344–355. doi: 10.1038/NRMICRO2558
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Sievers, F., and Higgins, D. G. (2014). Clustal omega. *Curr. Protoc. Bioinforma.* 48, 3.13.1–3.13.16. doi: 10.1002/0471250953.BI0313S48
- Simpson, A. J. G., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M., Alvarenga, R., et al. (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. the *Xylella fastidiosa* consortium of the organization for nucleotide sequencing and analysis. *Nature* 406, 151–157. doi: 10.1038/35018003
- Song, Z., Yang, C., Zeng, R., Gao, S., Cheng, W., Gao, P., et al. (2021). First report of strawberry crown rot caused by *Xanthomonas fragariae* in China. *Plant Dis.* 105, 2711. doi: 10.1094/PDIS-03-21-0574-PDN
- Sory, M.-P., and Cornelis, G. R. (1994). Translocation of a hybrid YopE-adenylate cyclase from *Yersinia enterocolitica* into HeLa cells. *Mol. Microbiol.* 14, 583–594. doi: 10.1111/j.1365-2958.1994.tb02191.x
- Strunnikov, A. V. (2006). SMC complexes in bacterial chromosome condensation and segregation. *Plasmid* 55, 135–144. doi: 10.1016/j.PLASMID.2005.08.004
- Teper, D., Burstein, D., Salomon, D., Gershovitz, M., Pupko, T., and Sessa, G. (2016). Identification of novel *Xanthomonas euvesicatoria* type III effector proteins by a machine-learning approach. *Mol. Plant Pathol.* 17, 398–411. doi: 10.1111/mpp.12288
- Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathirana, 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
- Wagner, N., Albuquerque, M., Ecker, N., Dotan, E., Zerah, B., Pena, M. M., et al. (2022a). Natural language processing approach to model the secretion signal of type III effectors. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1024405
- Wagner, N., Avram, O., Gold-Binshok, D., Zerah, B., Teper, D., and Pupko, T. (2022b). Effector: an automated machine-learning-based web server for the prediction of type-III secretion system effectors. *Bioinformatics* 38, 2341–2343. doi: 10.1093/bioinformatics/btac087
- Wagner, N., Teper, D., and Pupko, T. (2022c). “Predicting type III effector proteins using the effector web server,” in *Bacterial virulence*. Ed. O. Gal-Mor (New York, New York, USA: Springer US), 25–36. doi: 10.1007/978-1-0716-1971-1\_3
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakhikumar, 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, H., McTavish, C., and Turechek, W. W. (2018). Colonization and movement of *Xanthomonas fragariae* in strawberry tissues. *Phytopathology* 108, 681–690. doi: 10.1094/PHYTO-10-17-0356-R/ASSET/IMAGES/LARGE/PHYTO-10-17-0356-R\_F4.JPG
- 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, H. Y., Lai, Q. J., Wu, Y. M., Chung, C. L., Chung, P. C., and Lin, N. C. (2020). First report of *Xanthomonas fragariae* causing angular leaf spot on strawberry (*fragaria x ananassa*) in Taiwan. *Plant Dis.* 105, 1187. doi: 10.1094/PDIS-07-20-1631-PDN
- Yang, L., Su, H., Yang, F., Jian, H., Zhou, M., Jiang, W., et al. (2015). Identification of a new type III effector XC3176 in *Xanthomonas campestris* pv. *campestris*. *Wei Sheng Wu Xue Bao* 55, 1264–1272. Available at: <https://europepmc.org/article/med/26939454>
- Zhang, S., Sairam, R. V., Grefer, D., Feasel, J., Ferencak, M., and Goldman, S. L. (2009). Resistance to *Xanthomonas campestris* pv. *pelargonii* in geranium and diagnosis of the bacterial blight using polymerase chain reaction. *Taylor Fr.* 42, 1109–1117. doi: 10.1080/03235400701622378