Plant Disease • 2022 • 106:1558-1565 • https://doi.org/10.1094/PDIS-11-21-2498-SC

Short Communication

e-Xtra*

Whole-Genome Resources and Species-Level Taxonomic Validation of 89 Plant-Pathogenic *Xanthomonas* Strains Isolated from Various Host Plants

James T. Tambong,[1,2,†] Renlin Xu,[1] Diane Cuppels,[3] Julie Chapados,[1] Suzanne Gerdis,[1] Jackson Eyres,[4] Adam Koziol,[5] and Jeremy Dettman[1]

[1] Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada
[2] Department of Plant Science, University of Manitoba, Winnipeg, Manitoba, Canada
[3] London Research and Development Centre, Agriculture and Agri-Food Canada, London, Ontario, Canada (Retired)
[4] BICoE, Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada
[5] Canadian Food Inspection Agency, Ottawa, Ontario, Canada

Abstract

Bacterial spot disease caused by *Xanthomonas* spp. is a global threat to tomato and pepper plants. A recent classification of these pathogens indicated the need for a diverse dataset of whole-genome resources. We report whole-genome resources of 89 *Xanthomonas* strains isolated from Canada (n = 44), the United States (n = 29), Argentina (n = 4), Brazil (n = 3), Costa Rica (n = 3), New Zealand (n = 1), Australia (n = 1), Mexico (n = 1), Taiwan (n = 1), Thailand (n = 1), and unknown (n = 1). Of these strains, 48 were previously identified to species-level based on nongenome-based approaches while 41 strains were classified only at the genus level. The average coverage of the sequencing reads was 103×. The draft genome sizes ranged from 4.53 to 5.46 Mbp with a G + C content of 63.53 to 67.78% and comprised 4,233-5,178 protein-coding sequences. Using average nucleotide identity (ANI) and genome-based DNA-DNA hybridization (gDDH) values, the taxonomic classifications were validated for 38 of the 48 strains previously assigned to species level using other methods. Ten strains previously identified as *Xanthomonas campestris*, *X. axonopodis*, *X. vasicola*, and *X. arboricola* were incorrectly assigned, and new species-level delineations are proposed. Data from ANI, gDDH, and pangenome phylogeny of shared protein families were used to assign the 41 strains, previously identified only to genus level, into five distinct species: *X. euvesicatoria* (pv. *euvesicatoria* or pv. *perforans*), *X. hortorum* pv. *gardneri*, *X. vesicatoria*, *X. campestris*, and *X. arboricola*. These 89 whole-genome sequences of *Xanthomonas* strains, the majority (49.4%) of which are from Canada, could be useful resources in our understanding of the global population structure and evolution of these pathogens.

Keywords: average nucleotide identity, bacteria, bacterial spot disease, genome-based DNA-DNA hybridization, genome sequencing, pepper, prokaryotes, taxonomy, whole-genome sequencing, *Xanthomonas*, *X. euvesicatoria*, *X. gardneri*, *X. hortorum*, *X. perforans*, *X. vesicatoria*

The genus *Xanthomonas*, a member of the *Xanthomonadaceae* family (γ-*Proteobacteria*), encompasses a ubiquitous group of plant-pathogenic bacteria (Buttner and Bonas 2010). A distinct characteristic of this group of Gram-negative bacteria is the production of xanthan, an important ingredient in the food industry, and from which the taxonomic name of the genus is derived (da Silva et al. 2017). The genus has 38 validly published species or child taxa and 6 that are not validly published (https://lpsn.dsmz.de/genus/xanthomonas; accessed on October 16, 2021). Until recently, the taxonomic nomenclature of *Xanthomonas* spp. causing the bacterial spot disease on tomato and pepper has been unclear. In the early 1990s, the causal agent was taxonomically classified as *Xanthomonas campestris* pv. *vesicatoria* (Vancheva et al. 2021). The reclassifications of Jones et al. (2004), using DNA-DNA hybridization and metabolic profiles, resulted in the identification of four *Xanthomonas* spp. (*X. vesicatoria, X. euvesicatoria, X. perforans*, and *X. gardneri*) causing bacterial spot disease on tomato and pepper. A series of recent taxonomic studies (Constantin et al. 2016; Morinière et al. 2020; Timilsina et al. 2019; Vancheva et al. 2021) based on whole-genome sequences have reclassified the bacterial-spot-causing agents into three distinct *Xanthomonas* spp.: *X. vesicatoria; X. euvesicatoria* (*X. euvesicatoria* pv. *euvesicatoria* and *X. euvesicatoria* pv. *perforans*), and *X. hortorum* (pv. *gardneri*). This recent and more accurate taxonomic classification of the causal agents of bacterial spot disease on tomato and pepper underscores the significance of the availability of whole-genome resources. Although the availability of the genome resources is important, a more diverse geographic origin of the data is also crucial, given that the bacterial spot disease of tomato and pepper is a worldwide agricultural problem. The majority of the *Xanthomonas* genomes available, to date, are mainly from the United States, with a few from European Union countries. These represent a temporally and geographically biased dataset (Barak et al. 2016) from which to infer a global genome-based population structure and taxonomic perspective on the causal agents of bacterial spot disease on tomato and pepper. This might have partly contributed to the ever-changing taxonomic classification of these worldwide pathogens of tomato and pepper, important agricultural field and greenhouse crops. Here, we report high-quality draft whole-genome resources of 89 strains of *Xanthomonas* isolated from Canada (n = 42), the United States (n = 29), Argentina (n = 4), Brazil (n = 3), Australia (n = 2), Costa Rica (n = 3), New Zealand (n = 1), Mexico (n = 1), Taiwan (n = 1), and Thailand (n = 1). The origin of strain X40 isolated from cassava is unknown. In all, 80 of the 89 *Xanthomonas* strains were isolated from tomato and pepper while single isolates were obtained from nine other crop species (alfalfa, cowpea, cassava, cotton, begonia, sorghum, prunus, citrus, or barley). Of these 89 strains, a group of

[†]Corresponding author: J. T. Tambong; james.tambong@agr.gc.ca

Funding: Funding for genome sequencing and assembly of these bacterial strains was provided by Agriculture and Agri-Food Canada through the Biological Collections Data Mobilization Initiative project (BioMob; J-001564; Work Package 2); and comparative genomics analysis was funded through projects J-002272, J-001577, and J-000409.

*The *e*-Xtra logo stands for "electronic extra" and indicates that supplementary materials are published online.

The author(s) declare no conflict of interest.

Accepted for publication 28 January 2022.

48 were previously identified to species level based on phenotypic and nongenome-based methods while 41 strains were classified only at the genus level. The goal of this work was to contribute to the global whole-genome resources of the causal agents of bacterial spot disease on tomato and pepper; and, in particular, provide the scientific community with a "trove" of Canadian genome sequences for downstream analyses. The subobjectives were to (i) use genome-based average nucleotide identity (ANI) to validate species-level identification of 48 strains previously assigned using other approaches and (ii) identify, to species level, 41 strains that were previously classified only to the genus level using two Overall Genome Related Indices (OGRI) (Chun et al. 2018): ANI (Jain et al. 2018) and genome-based (digital) DNA-DNA hybridization (dDDH) (Meier-Kolthoff et al. 2013, 2021).

The Canadian strains of *Xanthomonas* were isolated between 1991 and 2006 from leaf and fruit tissues of tomato and pepper exhibiting typical bacterial spot symptoms in Ontario, Canada. The pathogens were isolated from diseased fresh material as previously described (Cuppels et al. 2006) using CKTM medium, a semiselective medium for the bacterial spot pathogen (Sijam et al. 1991). Pathogenicity was confirmed as previously described (Kuflu and Cuppels 1997). Purified strains were stored at −80°C in nutrient broth-yeast extract broth containing 15% glycerol (Cuppels et al. 1990). The bacterial isolates were cultured overnight in Luria-Bertani broth, and the genomic DNA of each strain was extracted using the DNeasy UltraClean Microbial kit (Qiagen), with minor modifications. Briefly, 600 µl of microbial culture was used to generate a cell pellet which was stored at −20°C until used; and 36 µl of RNase Cocktail (Invitrogen) was added to the cell suspension before processing. Homogenization was performed using a Tissue Lyser II (Qiagen) at a frequency of 20 Hz for 5 min, and the process was repeated once, with plates reoriented at 180° prior to incubation at 37°C for 30 min. The strains were characterized to the genus *Xanthomonas* using 16S ribosomal RNA PCR amplification and sequencing (Lane 1991; Tambong et al. 2017) followed by BLAST analysis (Altschul et al. 1990). Libraries were constructed using the Nextera DNA Flex prep kit (Illumina) following the manufacturer's instructions. The draft genome sequences were determined by paired-end sequencing using Illumina NextSeq technology at the Molecular Technologies Laboratory (Ottawa Research and Development Centre, Ottawa, Canada).

The quality of the reads was checked using FastQc (Andrews 2010) and trimmed with BBDUK (Bushnell 2014), if required, and de novo assemblies were performed using SKESA (Souvorov et al. 2018) as implemented in COWBAT version 0.5.0.7 pipeline (https://github.com/OLC-Bioinformatics/COWBAT). Scaffolds <500 bp long were discarded. The best assemblies, based on N_{50} and L_{50} values and nearness to the expected genome size, were obtained. The completeness and contamination rate of the assembled genomes were analyzed using the checkM algorithm (Parks et al. 2015) prior to annotation with RASTtk (Brettin et al. 2015) as implemented in the pipeline of the PathoSystems Resource Integration Center (PATRIC) (Davis et al. 2020; Wattam et al. 2014).

The validation of 48 *Xanthomonas* strains previously identified to the species level using nongenome-based methods was done using ANI at a 95 to 96% cut-off value (Goris et al. 2007) in comparison with 38 validly described *Xanthomona*s spp., including synonyms. In case of a nonconfirmatory result, gDDH values (cut-off value of 70%) were computed using Genome-Genome-Distance Calculator (DDGC) version 3.0 (Meier-Kolthoff et al. 2013, 2021) to ascertain the new species-level assignment. Species-level identification of 41 stains previously assigned only to the genus level was achieved with both ANI and gDDH at the cut-off thresholds indicated above. The CMG-Biotools pipeline (Vesth et al. 2013) was used to compute the pangenome (the entire set of protein families present from all genomes) and core genomes (protein families with representatives in all genomes) of the new population comprising the *Xanthomonas* sp. strains. In addition, a pangenome phylogenetic tree based on shared gene families was generated using the pancoreplot_tree command as implemented in the CMG-Biotools pipeline.

The mean number of 150-bp paired-end reads generated from the 48 strains previously identified to the species level was 3,290,825 (2,497,262 to 4,874,574) and represented an average estimated coverage of 93.2× within a range of 69× to 140×. (Table 1). The 48 assembled whole-genome sequences were found to be of good quality, with completeness of 99.6 to 100% and coarse and fine consistencies of 99.3 to 99.9 and 98.2 to 99.6%, respectively (Table 1). In all, 44 of the 48 strains showed a 0% contamination rate, 3 had a 0.4% rate, and 1 had a 3.2% rate, with all rates being below the acceptable level. The average size of these 48 draft whole-genome sequences is 5.13 Mbp, with a range of 4.53 to 5.46 Mbp (average maximum contig = 244,112 bp and N_{50} = 85,600 bp) (Table 1). The average G + C content of the draft genome sequences is 64.52% (63.53 to 67.78%) (Table 1). The assembled draft whole-genome sequences were annotated using the RAST tool kit (RASTtk) (Brettin et al. 2015) as implemented in the PATRIC pipeline, which identified 4,233 to 5,178 protein-coding sequences (Table 1). Species-level validation of the 48 strains using ANI values, at a cut-off threshold of 95 to 96% (Jain et al. 2018), confirmed the taxonomic classification of 38 strains to be accurate while 10 strains were incorrectly assigned. Three of the four strains previously identified as *X. campestris* were reclassified to either *X. phaseoli, X. cannabis*, or *X. arboricola* based on ANI values that are ≥95 to 96% (Table 1). These reclassifications were strongly supported by gDDH values of 82.3 and 92.2% when the genome sequences of strains X40 and Xcz 13 were compared with the type strains of *X. phaseoli* and *X. cannabis*, respectively (Table 1). The gDDH computed between strain Xcz5 (previously identified as *X. campestris*) and the type strain of *X. arboricola* was 68.0% (Table 1). This is less than the 70% cut-off threshold for species-level delineation but was the highest value of the 39 *Xanthomonas* type strains evaluated. Strain Xcz 5 had an ANI value of 87.0% and a gDDH value of 30.9% with the type strain of *X. campestris* (the previously assigned species), which are significantly less than their respective cut-off values of 95 to 96% and 70% required for species-level assignment. Strain Xcz5 cannot be assigned to "new species" status because it is required that both the ANI and gDDH values be below the thresholds of 95 and 70%, respectively (Kim et al. 2014; Osdaghi et al. 2020). Also, all five strains (X60, G55, X203, FB570, and JB1) previously classified as *X. axonopodis*, were correctly identified as either *X. euvesicatoria*, *X. fuscans*, *X. codiaei*, or *X. phaseoli* (Table 1) based on ANI and gDDH values that are greater than the recommended cut-off threshold values (Table 1). Two strains (X56 and Xc69), previously identified as *X. vasicola* and *X. arboricola*, respectively, were correctly reclassified as *X. hortorum* pv. *gardneri* based on ANI and gDDH values of, respectively, 98 and 80.8% for X56 and 98.3 and 85.1% for Xc69.

The whole-genome sequencing of 41 *Xanthomonas* strains that were previously classified only to the genus level yielded an average read count of 4,001,410 (2,356,947 to 5,736,201), with an average coverage of 114× (68× to 157×) (Table 2). The assembled draft genome sizes ranged from 4.78 to 5.31 Mb in contig counts ranging from 39 to 274 and N_{50} of 34,494 to 367,373 bp, with an average maximum largest length of 369,191 bp (100,664 to 765,731) (Table 2). The average G + C content is 64.5% (63.66 to 65.9%). Based on CheckM as implemented in PATRIC pipeline, all of the 41 draft whole-genome sequences were of good quality, with 0% percent contamination rate. The 41 whole-genome sequences were annotated using RASTkt, showing an average of 4,772 (4,309 to 5,033) protein-encoding genes, of which 61.4% (2,932) had functional assignment and 29.4% (mean = 1,839) are hypothetical protein-encoding genes (Table 2).

The 41 *Xanthomonas* sp. strains were classified into five distinct species based on ANI and gDDH values. The majority (n = 25) of the strains were accurately assigned to *X. euvesicatoria*, 7 strains were taxonomically affiliated with *X. hortorum*, 6 were identified as *X. vesicatoria*, 2 as *X. campestris*, and 1 as *X. arboricola* (Table 3). The 25 strains identified as *X. euvesicatoria* all showed ANI values of 98.5 to 99.9% and gDDH values of 85.4 to 99.99% with *X. euvesicatoria* pv. *euvesicatoria*, *X. euvesicatoria* pv. *perforans*, and *X. euvesicatoria* (Table 3). This truly suggests that these three pathovars belonged to the same species, as previously described (Constantin et al. 2016). Twenty strains of *X. euvesicatoria* that showed ANI and gDDH values of 99.8 to 99.9% and 99.1 to 99.4%, respectively, with the pathotype

Table 1. Basic genome statistics and validation of species-level identification using genome-based average nucleotide identity (ANI) values of previously identified *Xanthomonas* strains[a]

Previous species-level identification	Strain	Locality, country; host	Year of isolation	Total number of reads	Average coverage depth	Contig count	Contigs N_{50} (bp)	Contigs L_{50}	Largest contig	Total length (bp)	GC%	Total number of protein-coding sequences	Protein-encoding genes with functional assignment	Protein-encoding genes without functional assignment	Proposed species-level identification (ANI % [gDDH %])
Xanthomonas campestris* pv. *manihotis	**X40**	**n/a; Cassava**	**n/a**	**2,645,432**	**76**	**161**	**70,842**	**23**	**214,647**	**5,023,684**	**64.93**	**4,790**	**2,925**	**1,865**	***X. phaseoli* (98.1; [82.3])**
X. campestris pv. *armoraciae*	71	Minnesota, United States; radish	n/a	2,930,725	87	147	58,042	26	192,931	4,908,515	65.22	4,534	2,806	1,728	*X. campestris* (97.28)
X. campestris* pv. *zinniae	**Xcz 5**	**Ohio, United States; zinnia**	**2001**	**3,382,176**	**105**	**112**	**74,774**	**21**	**260,551**	**4,722,046**	**65.85**	**4,233**	**2,746**	**1,487**	***X. arboricola* (96.3; [68.0])**
X. campestris* pv. *zinniae	**Xcz13**	**Ohio, United States; zinnia**	**2001**	**3,530,214**	**109**	**123**	**72,657**	**22**	**262,244**	**4,706,474**	**65.75**	**4,237**	**2,741**	**1,496**	***X. cannabis* (99.17; [92.2])**
X. euvesicatoria	Xv 72	Florida, United States; pepper	n/a	4,874,574	131	129	101,621	19	254,049	5,458,903	64.5	5,178	3,065	2,113	*X. euvesicatoria* pv. *euvesicatoria* (99.9)
X. euvesicatoria	Xv 79	Florida, United States; pepper	n/a	4,778,562	129	117	114,481	17	254,049	5,439,628	64.51	5,162	3,056	2,106	*X. euvesicatoria* pv. *euvesicatoria* (99.9)
X. euvesicatoria	ATCC 11633	New Jersey, United States; pepper	n/a	2,976,684	82	125	101,995	16	389,113	5,316,656	64.47	5,029	3,006	2,023	*X. euvesicatoria* pv. *euvesicatoria* (100)
X. euvesicatoria	75-3	Florida, United States; tomato	1975	2,906,050	80	140	83,698	20	284,696	5,268,957	64.54	4,983	2,972	2,011	*X. euvesicatoria* pv. *euvesicatoria* (100)
X. euvesicatoria	85-16	Florida, United States; tomato	1985	2,806,300	78	153	76,453	24	160,817	5,251,155	64.57	4,984	2,965	2,019	*X. euvesicatoria* pv. *euvesicatoria* (99.9)
X. euvesicatoria	86-2	Florida, United States; tomato	1986	2,549,972	69	173	70,706	23	197,106	5,369,677	64.62	5,143	3,061	2,082	*X. euvesicatoria* pv. *euvesicatoria* (99.9)
X. euvesicatoria	86-22	Florida, United States; pepper	1986	2,789,146	77	148	86,089	22	203,485	5,292,853	64.61	5,037	3,011	2,026	*X. euvesicatoria* pv. *euvesicatoria* (99.9)
X. euvesicatoria	86-46	Florida, United States; pepper	1986	2,834,260	79	156	76,829	23	209,799	5,230,469	64.59	4,913	2,999	1,914	*X. euvesicatoria* pv. *euvesicatoria* (100)
X. euvesicatoria	87-21	Taiwan; pepper	1987	3,488,165	101	101	112,894	16	237,591	5,038,083	64.73	4,669	2,930	1,739	*X. euvesicatoria* pv. *euvesicatoria* (99.9)
X. euvesicatoria	87-47	Australia; pepper	1987	3,350,949	91	132	102,010	17	302,187	5,383,125	64.56	5,119	3,059	2,060	*X. euvesicatoria* pv. *euvesicatoria* (99.8)
X. euvesicatoria	89-10	Thailand; tomato	1989	2,885,326	80	127	113,408	18	237,608	5,230,924	64.59	4,929	2,984	1,945	*X. euvesicatoria* pv. *euvesicatoria* (99.8)
X. euvesicatoria	E3	Florida, United States; n/a	n/a	2,512,913	69	131	95,630	19	236,441	5,272,033	64.57	4,944	2,998	1,946	*X. euvesicatoria* pv. *euvesicatoria* (100)
X. perforans	DC00T15A	Harwich, ON, Canada; n/a	2000	3,776,167	112	49	205,101	8	791,193	4,961,873	64.94	4,564	2,873	1,691	*X. euvesicatoria* pv. *perforans* (99.8)
X. perforans	DC00T16A	Chatham, ON, Canada; n/a	2000	3,736,577	111	53	171,935	9	547,435	4,961,602	64.94	4,577	2,879	1,698	*X. euvesicatoria* pv. *perforans* (99.8)
X. perforans	Xv 1220	Mexico; tomato	n/a	2,672,447	79	62	128,849	13	364,637	4,941,273	64.94	4,557	2,865	1,692	*X. euvesicatoria* pv. *perforans* (100)
X. perforans	Xv 1484	Florida, United States; tomato	n/a	2,786,810	82	74	107,756	15	336,471	4,945,286	64.98	4,565	2,872	1,693	*X. euvesicatoria* pv. *perforans* (100)
X. perforans	97-2	Florida, United States; n/a	1997	2,666,058	78	63	157,482	12	404,446	5,029,231	64.96	4,642	2,910	1,732	*X. euvesicatoria* pv. *perforans* (100)
X. gardneri	DC98T7A	Dover, ON, Canada; tomato	1998	4,167,901	117	228	40,954	42	126,763	5,182,154	63.66	4,765	2,980	1,785	*X. hortorum pv. gardneri* (100)
X. gardneri	DC99T6A	Richards Farm, ON	1999	3,735,352	105	214	41,617	41	131,463	5,183,664	63.66	4,768	2,977	1,791	*X. hortorum pv. gardneri* (100)
X. gardneri	DC99T9A	Richards Farm, ON	1999	3,836,434	108	227	39,653	43	131,471	5,178,688	63.67	4,750	2,976	1,774	*X. hortorum pv. gardneri* (100)
X. gardneri	DC99T9B	Richards Farm, ON	1999	3,880,941	109	232	39,591	46	100,664	5,178,529	63.66	4,749	2,973	1,776	*X. hortorum pv. gardneri* (100)
X. gardneri	Xg101	Cost Rica; tomato	n/a	3,125,371	89	277	32,811	50	100,657	5,106,706	63.73	4,710	2,970	1,740	*X. hortorum pv. gardneri* (100)

(*Continued on next page*)

[a] Whole-genome sequences were generated using Illumina NextSeq platform and assembled using SKESA (Souvorov et al. 2018). Contigs < 500 bp were discarded. Based on CheckM algorithm (Parks et al. 2015) as implemented in the PATRIC pipeline (Davis et al. 2020; Wattam et al. 2014), the completeness of the genomes was 99.6 to 100% with coarse and fine consistencies of 99.3 to 99.9% and 98.2 to 99.6%, respectively; most genome sequences exhibited 0% rate of contamination while three had a 0.4% rate and one reported a 3.2% rate. Annotations and protein function assignments were implemented in PATRIC (Davis et al. 2020; Wattam et al. 2014). GC = G + C content. ANI values are shown in parentheses and genome-based DNA-DNA hybridization (gDDH) values are in square brackets (where required), with species delineation cut-off threshold values of 96.0 and 70%, respectively. Strains in bold have a different species-level taxonomic affiliation from the initial identification; n/a = not available.

Table 1. (*Continued from previous page*)

Previous species-level identification	Strain	Locality, country; host	Year of isolation	Total number of reads	Average coverage depth	Contig count	Contigs N_{50} (bp)	Contigs L_{50}	Largest contig	Total length (bp)	GC%	Total number of protein-coding sequences	Protein-encoding genes with functional assignment	Protein-encoding genes without functional assignment	Proposed species-level identification (ANI % [gDDH %])
X. gardneri	Xg 444	Cost Rica; tomato	n/a	3,106,559	85	266	36,723	46	180,418	5,336,796	63.53	4,988	3,028	1,960	*X. hortorum pv. gardneri* (100)
X. gardneri	Xg 451	Costa Rica; pepper	n/a	3,729,344	93	262	38,060	45	211,279	5,336,718	63.53	4,974	3,029	1,945	*X. hortorum pv. gardneri* (100)
X. gardneri	IBSBF1782	Brazil; tomato	1997	2,897,601	79	275	36,723	48	143,249	5,314,135	63.53	4,960	3,023	1,937	*X. hortorum pv. gardneri* (100)
X. gardneri	IBSBF1783	Brazil; tomato	n/a	2,497,262	72	276	33,908	49	112,887	5,039,300	63.72	4,634	2,937	1,697	*X. hortorum pv. gardneri* (100)
X. vesicatoria	BA 27-1	Florida, United States; tomato	n/a	4,808,631	135	90	146,239	11	429,576	5,194,179	64.15	4,910	2,863	2,047	*X. vesicatoria* (98.7)
X. vesicatoria	BA 29-1	Florida, United States; tomato	n/a	4,802,553	140	86	168,749	10	396,017	4,999,732	64.37	4,675	2,827	1,848	*X. vesicatoria* (98.8)
X. vesicatoria	Bv 5-3A	Argentina; pepper	n/a	3,951,784	110	59	182,341	11	440,883	5,282,410	64.09	4,945	2,905	2,040	*X. vesicatoria* (98.6)
X. vesicatoria	Bv 5-4A	Argentina; tomato	n/a	4,037,123	110	108	94,568	19	237,686	5,361,357	64.02	5,066	2,964	2,102	*X. vesicatoria* (98.6)
X. vesicatoria	Bv 5-4B	Argentina; tomato	n/a	2,849,473	82	92	103,388	18	262,581	5,067,310	64.33	4,705	2,853	1,852	*X. vesicatoria* (99.0)
X. vesicatoria	Bv 20-3A	Argentina; tomato	n/a	2,919,270	84	93	100,308	19	262,554	5,068,066	64.33	4,690	2,858	1,832	*X. vesicatoria* (100)
X. vesicatoria	ATCC 11551	Indiana, United States; tomato	n/a	2,924,230	84	90	95,073	18	262,441	5,081,946	64.27	4,720	2,848	1,872	*X. vesicatoria* (100)
X. vesicatoria	71-4	New Zealand; tomato	n/a	2,713,036	74	91	94,932	19	246,087	5,330,114	64.06	5,030	2,909	2,121	*X. vesicatoria* (100)
X. vesicatoria	MME	Florida, United States; n/a	n/a	2,854,124	83	92	88,549	20	245,369	5,028,534	64.35	4,661	2,843	1,818	*X. vesicatoria* (99.0)
X. axonopodis* pv. *alfalfae	**X60**	**n/a; Alfalfa**	**n/a**	**2,351,332**	**68**	**106**	**93,986**	**18**	**222,115**	**4,990,035**	**64.86**	**4,713**	**2,988**	**1,725**	***X. euvesicatoria pv. alfalfa* (98.5; [88.7])**
X. axonopodis* pv. *vignicola	**G55**	**n/a; Cowpea**	**n/a**	**3,001,914**	**85**	**123**	**50,848**	**33**	**230,737**	**5,055,235**	**64.71**	**4,836**	**2,987**	**1,849**	***X. fuscans* (98.5; [88.2])**
X. axonopodis* pv. *malvacearum	**X203**	**n/a; Cotton**	**n/a**	**3,367,486**	**93**	**199**	**94,618**	**17**	**155,324**	**5,285,520**	**66.03**	**4,593**	**2,865**	**1,728**	***X. codiaei* (97.43; [77.4])**
X. axonopodis* pv. *vasculorum	**FB570**	**n/a; sugarcane**	**n/a**	**3,148,106**	**87**	**115**	**113,488**	**16**	**257,569**	**5,300,250**	**64.54**	**5,033**	**3,013**	**2,020**	***X. euvesicatoria* pv. *euvesicatoria* (99.84; [99.2])**
X. axonopodis* pv. *begoniae	**JB1**	**n/a; Harrow-begonia**	**n/a**	**3,519,302**	**102**	**198**	**55,995**	**30**	**139,530**	**5,025,585**	**64.79**	**4,768**	**2,929**	**1,839**	***X. phaseoli* (96.6; [69.0])**
X. vasicola* pv. *holcicola	**X56**	**n/a; Sorghum**	**n/a**	**3,850,480**	**108**	**362**	**25,629**	**61**	**96,792**	**5,083,636**	**63.8**	**4,761**	**2,950**	**1,811**	***X. hortorum* pv. *gardneri* (98.0; [80.8])**
X. arboricola* pv. *pruni	**Xc 69**	**n/a; Prunus**	**n/a**	**3,307,661**	**93**	**299**	**29,350**	**60**	**66,612**	**5,155,497**	**63.65**	**4,797**	**2,948**	**1,849**	***X. hortorum* pv. *gardneri* (98.3; [85.1])**
X. fuscans subsp. *aurantifolii*	Xc70	Brazil; citrus	n/a	3,006,707	89	300	32,412	48	105,679	4,836,802	64.96	4,673	2,894	1,779	*X. fuscans* (98.8)
X. translucens pv. *translucens*	ATCC 19319	n/a; Barley	n/a	2,690,105	85	500	14,998	89	79,489	4,533,263	67.78	4,469	2,661	1,808	*X. translucens pv. translucens* (99.8)

of *X. euvesicatoria* pv. *euvesicatoria* were taxonomically assigned to this group. The remaining five strains were accurately assigned to *X. euvesicatoria* pv. *perforans* based on high ANI and gDDH values of 99.8 to 99.9% and 98.9 to 99.9%, respectively, compared with ANI of 98.5 to 98.6% (gDDH, 87.7 to 88.3%) and 98.8% (gDDH, 89.2 to 89.8%) recorded with *X. euvesicatoria* pv. *euvesicatoria* or pv. *alfalfa* (Table 3). These data suggest that 1 or 2% differences in ANI and gDDH values could be useful to taxonomically differentiate these pathovars at the genome level. A similar pattern but with greater margins was observed with seven strains that were taxonomically assigned to the recently reclassified *X. hortorum* (Morinière et al. 2020). Three of the *X. hortorum* pathovars (pvs. *hederae*, *cynarae*, and *gardneri*) showed ANI values of 95.9 to 96.0% (gDDH, 65.1 to 65.3%), 99.2% (gDDH, 93.4%), and 99.9% (gDDH, 99.8%), respectively (Table 3). The differences between the ANI and gDDH values of pv. *hederae* and the other two pathovars were about 3 and 34.2%, suggesting a potentially wider genetic distance. Based on ANI and gDDH values, these nine strains were assigned to the *X. hortorum* pv. *gardneri* group (Table 3). Also, six *Xanthomonas* sp. strains were identified as *X. vesicatoria* (ANI, 98.8 to 99.0%; gDDH, 89.4 to 91.0%), 2 strains were accurately assigned to *X. campestris* (ANI, 97.1 to 97.2%; gDDH, 75.9 to 76.0%), while strain DC06P2B was taxonomically affiliated with *X. arboricola* (Table 3). The species-level taxonomic classifications of the 41 *Xanthomonas* sp. strains were corroborated by pangenome phylogenetic tree based on shared gene families (Fig. 1). The tree inferred using the relative Manhattan distances clustered the strains into five distinct groups that corresponded to the five different *Xanthomona*s spp. identified using ANI and gDDH (Fig. 1). The tree branches were well supported by high bootstrap values. In total, 8,394 protein families constituted the pangenome of the 41 strains, with a core genome of 2,981 protein families (Supplementary Fig. S1). We generated and annotated 89 new whole-genome sequences of *Xanthomonas* strains. Using ANI and gDDH values and core gene phylogeny, the strains were accurately identified to the species level. These 89 whole-genome sequences of *Xanthomonas* strains, the majority of which are from Canada, could be useful resources required for a better understanding of global population structure as well as provide additional insight on the evolution of these important pathogens of tomato and pepper. Finally, accurate identification of pathogens is key to developing successful management strategies.

Data availability. The whole-genome sequences are deposited in DNA Data Bank of Japan/European Nucleotide Archive/GenBank under the accession numbers JAJITE000000000 to JAJIWN000000000.

Table 2. Strain code, origin, and basic genome statistics of 41 strains (37 from Canada, 3 from the United States and 1 from Australia) previously identified as *Xanthomonas* spp. isolated from pepper or tomato[a]

Strain code	Locality, country; host[b]	Year of isolation	Total number of reads	Average coverage depth	Number of contigs	Contigs N_{50} (bp)	Contigs L_{50}	Largest contig (bp)	Total length (bp)	GC%	Total number of protein-coding sequences	Protein-encoding genes with functional assignment	Protein-encoding genes without functional assignment
DC05T6	Smythe Farm	2005	5,736,201	169	39	367,373	5	765,731	4,983,679	64.97	4,561	2,887	1,674
DC01T1A	Knotek Farm, ON; tomato	2001	3,883,796	115	52	205,101	8	593,974	4,960,750	64.93	4,552	2,880	1,672
DC01T40A	Bradley; tomato	2001	4,899,371	144	49	209,972	7	663,495	4,962,251	64.94	4,572	2,884	1,688
DC01T37B	Delrue Farm; tomato	2001	4,345,501	128	46	211,153	7	675,480	4,961,791	64.94	4,564	2,880	1,684
DC01T3A	Demaiter Farm	2001	5,411,012	160	55	171,935	9	577,349	4,961,360	64.93	4,581	2,882	1,699
Xcv DC 96-3	Hank VanderPol Farm, Blenheim, ON; pepper	1996	3,962,385	114	101	135,247	12	386,839	5,099,772	64.69	4,760	2,952	1,808
Xcv DC 93-7	Dresden, ON; tomato	1993	4,125,061	116	86	151,232	12	445,571	5,187,864	64.7	4,847	2,953	1,894
Xcv DC 97 P1A	Hank VanderPol GH, Blenheim, ON; pepper	1997	2,758,366	77	125	123,182	15	325,746	5,259,187	64.57	4,959	2,974	1,985
Xcv DC 93-8	Dresden, ON; tomato	1993	3,685,117	104	83	157,660	11	385,077	5,193,949	64.69	4,872	2,954	1,918
Xcv DC 93-9	Dresden, ON; tomato	1996	4,157,109	117	94	148,294	12	401,869	5,192,676	64.69	4,875	2,958	1,917
Xcv DC 96-4	Hank VanderPol Farm, Blenheim, ON; pepper	1996	3,092,797	85	99	125,853	13	447,498	5,301,682	64.55	5,012	3,005	2,007
Xcv DC 96-5	Hank VanderPol Farm, Blenheim, ON; pepper	1996	3,646,691	105	95	163,207	12	386,865	5,102,336	64.69	4,760	2,951	1,809
Xcv DC 98 P2A	Labradel GH, ON; pepper	1998	4,333,524	119	99	151,194	12	386,869	5,300,503	64.55	5,005	3,003	2,002
Xcv DC 93-6	R. Dick Farm, Leamington, ON; pepper	1993	3,328,731	92	101	123,171	15	358,772	5,300,570	64.54	5,001	3,008	1,993
Xcv DC 96-1	Janssen Farm, ON; pepper	1996	3,947,942	109	91	157,637	12	400,250	5,301,432	64.55	5,008	3,003	2,005
DC99P1B1	Keitzer GH, Hartford, MI; pepper	1999	3,788,715	104	106	136,738	12	357,399	5,306,922	64.54	5,033	2,996	2,037
Ps-7	Harrow, ON; pepper	n/a	3,002,108	82	135	90,883	19	343,970	5,298,144	64.55	5,029	3,019	2,010
Xv157	Australia; pepper	n/a	2,724,076	76	116	99,908	16	291,444	5,230,725	64.59	4,927	2,978	1,949
Xcv DC 97 P3A	Jean LaPrise GH, Chatham, ON; pepper	1997	4,380,038	126	73	152,579	11	448,278	5,076,590	64.69	4,709	2,919	1,790
DC99P1A1	Keitzer GH, Hartford, MI; pepper	1999	4,033,992	111	111	113,398	14	357,178	5,303,540	64.54	5,025	2,990	2,035
Xcv DC 96-2	Janssen Farm, ON; pepper	1996	3,000,672	85	91	140,980	12	363,388	5,143,413	64.66	4,815	2,972	1,843
Xcv DC 97 P2A	Jean LaPrise GH, Chatham, ON; pepper	1997	2,990,660	86	96	113,434	15	257,574	5,075,487	64.69	4,738	2,937	1,801
71-21	Florida, United States; n/a	n/a	2,356,947	68	111	94,079	16	362,895	5,068,748	64.78	4,720	2,955	1,765
Xcv Sspep 92	Ontario; pepper	1992	3,852,726	109	80	153,729	10	448,269	5,195,177	64.59	4,858	2,923	1,935
Ps-1	Harrow Res Centre, ON; pepper	n/a	2,614,030	73	122	83,920	18	389,124	5,192,512	64.59	4,859	2,929	1,930
MSU 1138	Indiana F	1996	2,679,457	75	274	34,494	49	112,877	5,173,356	63.66	4,777	2,988	1,789
MSU 1140	Purdue 9502	1996	3,262,182	92	263	38,768	47	100,664	5,175,429	63.66	4,774	2,987	1,787
DC00T17A	Harwich, NE	2000	4,389,424	124	221	41,618	42	126,761	5,182,893	63.66	4,766	2,979	1,787
DC04T7	n/a	n/a	4,806,286	135	236	40,959	43	115,931	5,182,271	63.66	4,776	2,984	1,792
DC05T4	Wilson Farm	2005	4,788,328	135	230	40,134	41	126,761	5,181,516	63.66	4,782	2,981	1,801
DC05T3	Poppe, cv. K1480	2005	4,792,507	135	232	42,087	40	142,930	5,185,246	63.66	4,795	2,986	1,809
DC06T5	Ecoland Farms, Kent County	2006	5,488,650	155	230	41,617	42	116,192	5,184,452	63.66	4,780	2,983	1,797
DC01T36A	P. Caza Farm, ON; n/a	2001	4,928,291	142	65	170,864	9	577,375	5,095,928	64.21	4,713	2,847	1,866
Xcv DC 98 T1A	Labradel GH, ON; tomato	1998	4,800,407	138	61	196,877	10	470,290	5,096,029	64.22	4,677	2,842	1,835
Xcv DC 97 T16B	Litschko Farm, ON; tomato	1997	4,369,284	126	66	151,790	11	508,142	5,077,786	64.33	4,685	2,851	1,834
Xcv DC 93-5	Tiessen Farm, ON; tomato	1993	3,672,490	106	57	153,353	11	496,018	5,078,788	64.33	4,683	2,850	1,833
Xcv DC 97 T16A	Litschko Farm, ON; tomato	1997	5,436,426	157	64	148,188	14	349,312	5,079,180	64.32	4,690	2,852	1,838
Xcv DC 97 T18A	Piacek Farm, ON; tomato	1997	4,267,491	123	64	148,632	13	349,288	5,078,107	64.32	4,682	2,852	1,830
DC06T4A	Tomato plug plant	2006	3,738,740	111	185	49,518	34	139,893	4,926,387	65.27	4,628	2,868	1,760
Xcv DC 91-1	Leamington, ON; tomato	1991	3,963,471	118	103	90,603	18	253,342	4,908,184	65.36	4,494	2,811	1,683
DC06P2B	Chatham, ON; Pepper	2006	4,616,818	140	85	125,618	13	330,143	4,781,111	65.88	4,309	2,765	1,544

[a] Whole-genome sequences were generated using Illumina NextSeq platform and assembled using SKESA (Souvorov et al. 2018). Contigs < 500 bp were discarded. Based on the PATRIC (Davis et al. 2020; Wattam et al. 2014), the completeness of the genomes was 99.6 to 100%, with coarse and fine consistencies of 99.4 to 99.9% and 98.6 to 99.6, respectively; a 0% rate of contamination as determined using CheckM algorithm. Annotations and protein function assignments were implemented in PATRIC (Davis et al. 2020; Wattam et al. 2014). GC = G + C content. Strain Xv157 from Australia and strains DC 71-21, MSU 1138 and MSU 1140 obtained from the United States are the only non-Canadian isolates; n/a = not available.

[b] Locations include Ontario, Canada (ON) and Michigan (MI) and Nebraska (NE) in the United States.

Table 3. Species-level identification using genome-based average nucleotide identity (ANI) and digital DNA-DNA hybridization (gDDH; in parentheses) values of 41 strains isolated from Canadian farms and greenhouses between 1991 and 2006, previously identified as *Xanthomonas* spp.[a]

		ANI (gDDH) values (%)								
		Xanthomonas euvesicatoria			*X. hortorum*					
Strain code	**Genome-based species-level identification**	**pv. *euvesicatoria* LMG 27970**	**pv. *perforans* DSM 18975**	**pv. *alfalfa* GEV-Rose-07**	**pv. *hederae* CFBP 4925**	**pv. *gardneri* CFBP 8163**	**pv. *cynarae* CFBP 4188**	***X. vesicatoria* LMG 911**	***X. campestris* pv. *campestris* ATCC 33913**	***X. arboricola* CFBP 2528**
DC05T6	*X. euvesicatoria* pv. *perforans*	98.5 (87.7)	**99.9 (99.9)**	98.8 (89.8)	–	–	–	–	–	–
DC01T1A	*X. euvesicatoria* pv. *perforans*	98.6 (88.3)	**99.8 (98.9)**	98.8 (89.2)	–	–	–	–	–	–
DC01T40A	*X. euvesicatoria* pv. *perforans*	98.6 (88.3)	**99.8 (98.9)**	98.8 (89.2)	–	–	–	–	–	–
DC01T37B	*X. euvesicatoria* pv. *perforans*	98.6 (88.3)	**99.8 (98.9)**	98.8 (89.2)	–	–	–	–	–	–
DC01T3A	*X. euvesicatoria* pv. *perforans*	98.6 (88.3)	**99.8 (98.9)**	98.8 (89.2)	–	–	–	–	–	–
Xcv DC 96-3	*X. euvesicatoria* pv. *euvesicatoria*	**99.8 (99.3)**	98.5 (87.2)	98.5 (86.5)	–	–	–	–	–	–
Xcv DC 93-7	*X. euvesicatoria* pv. *euvesicatoria*	**99.8 (99.1)**	98.5 (87.3)	98.5 (87.1)	–	–	–	–	–	–
Xcv DC 97 P1A	*X. euvesicatoria* pv. *euvesicatoria*	**99.8 (99.2)**	98.5 (87.2)	98.5 (86.5)	–	–	–	–	–	–
Xcv DC 93−8	*X. euvesicatoria* pv. *euvesicatoria*	**99.8 (99.1)**	98.5 (87.2)	98.6 (87.1)	–	–	–	–	–	–
Xcv DC 93−9	*X. euvesicatoria* pv. *euvesicatoria*	**99.8 (99.1)**	98.5 (87.2)	98.5 (87.1)	–	–	–	–	–	–
Xcv DC 96−4	*X. euvesicatoria* pv. *euvesicatoria*	**99.8 (99.2)**	98.4 (86.5)	98.4 (85.4)	–	–	–	–	–	–
Xcv DC 96−5	*X. euvesicatoria* pv. *euvesicatoria*	**99.8 (99.3)**	98.5 (87.2)	98.5 (86.5)	–	–	–	–	–	–
Xcv DC 98 P2A	*X. euvesicatoria* pv. *euvesicatoria*	**99.8 (99.2)**	98.5 (86.5)	98.4 (85.5)	–	–	–	–	–	–
Xcv DC 93−6	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.2)**	98.4 (86.6)	98.4 (85.5)	–	–	–	–	–	–
Xcv DC 96−1	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.2)**	98.4 (86.5)	98.4 (85.5)	–	–	–	–	–	–
DC99P1B1	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.2)**	98.5 (87.1)	98.5 (86.2)	–	–	–	–	–	–
Ps−7	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.2)**	98.4 (86.6)	98.4 (85.5)	–	–	–	–	–	–
Xv157	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.3)**	98.5 (87.2)	98.5 (86.3)	–	–	–	–	–	–
DC99P1A1	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.2)**	98.5 (87.1)	98.5 (86.3)	–	–	–	–	–	–
Xcv DC 97 P3A	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.3)**	98.5 (86.7)	98.5 (86.3)	–	–	–	–	–	–
Xcv DC 96−2	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.2)**	98.4 (86.5)	98.5 (86.1)	–	–	–	–	–	–
Xcv DC 97 P2A	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.3)**	98.5 (87.1)	98.4 (86.4)	–	–	–	–	–	–
71−21	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.4)**	98.5 (86.8)	98.5 (86.5)	–	–	–	–	–	–
Xcv Sspep 92	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.4)**	98.5 (87.4)	98.6 (87.0)	–	–	–	–	–	–
Ps−1	*X. euvesicatoria* pv. *euvesicatoria*	**99.9 (99.4)**	98.5 (87.4)	98.5 (87.0)	–	–	–	–	–	–
MSU 1138	*X. hortorum* pv. *gardneri*	–	–	–	95.9 (65.3)	**99.9 (99.8)**	99.2 (93.4)	–	–	–
MSU 1140	*X. hortorum* pv. *gardneri*	–	–	–	96.0 (65.3)	**99.9 (99.8)**	99.2 (93.4)	–	–	–
DC00T17A	*X. hortorum* pv. *gardneri*	–	–	–	96.0 (65.1)	**99.9 (99.8)**	99.2 (93.4)	–	–	–
DC04T7	*X. hortorum* pv. *gardneri*	–	–	–	96.0 (65.1)	**99.9 (99.8)**	99.2 (93.4)	–	–	–
DC05T4	*X. hortorum* pv. *gardneri*	–	–	–	96.0 (65.1)	**99.9 (99.8)**	99.2 (93.5)	–	–	–
DC05T3	*X. hortorum* pv. *gardneri*	–	–	–	96.0 (65.1)	**99.9 (99.8)**	99.2 (93.4)	–	–	–

(*Continued on next page*)

[a] ANI values were computed using FastAni (Jain et al. 2018) while gDDH values were obtained using GGDC v3.0 (Meier-Kolthoff et al. 2021, 2013) with species-level cut-off values of 96 and 70%, respectively. Only values of type strains or pathotypes showing highest ANI and gDDH values are shown. For *X. euvesicatoria* or *X. hortorum*, where pathovars show ANI and gDDH values greater than the required threshold, the highest value (in bold) determines the prevailing pathovar.

Table 3. (*Continued from previous page*)

Strain code	Genome-based species-level identification	ANI (gDDH) values (%)								
		Xanthomonas euvesicatoria			*X. hortorum*					
		pv. *euvesicatoria* LMG 27970	pv. *perforans* DSM 18975	pv. *alfalfa* GEV-Rose-07	pv. *hederae* CFBP 4925	pv. *gardneri* CFBP 8163	pv. *cynarae* CFBP 4188	*X. vesicatoria* LMG 911	*X. campestris* pv. *campestris* ATCC 33913	*X. arboricola* CFBP 2528
DC06T5	*X. hortorum* pv. *gardneri*	–	–	–	95.9 (65.1)	**99.9 (99.8)**	99.2 (93.4)	–	–	–
DC01T36A	*X. vesicatoria*	–	–	–	–	–	–	98.8 (89.5)	–	–
Xcv DC 98 T1A	*X. vesicatoria*	–	–	–	–	–	–	98.8 (89.4)	–	–
Xcv DC 97 T16B	*X. vesicatoria*	–	–	–	–	–	–	99.0 (91.0)	–	–
Xcv DC 93-5	*X. vesicatoria*	–	–	–	–	–	–	99.01 (91.0)	–	–
Xcv DC 97 T16A	*X. vesicatoria*	–	–	–	–	–	–	99.0 (91.1)	–	–
Xcv DC 97 T18A	*X. vesicatoria*	–	–	–	–	–	–	99.0 (91.0)	–	–
DC06T4A	*X. campestris*	–	–	–	–	–	–	–	97.1 (76.1)	–
Xcv DC 91−1	*X. campestris*	–	–	–	–	–	–	–	97.2 (75.9)	–
DC06P2B	*X. arboricola*	–	–	–	–	–	–	–	–	96.2 (67.2)

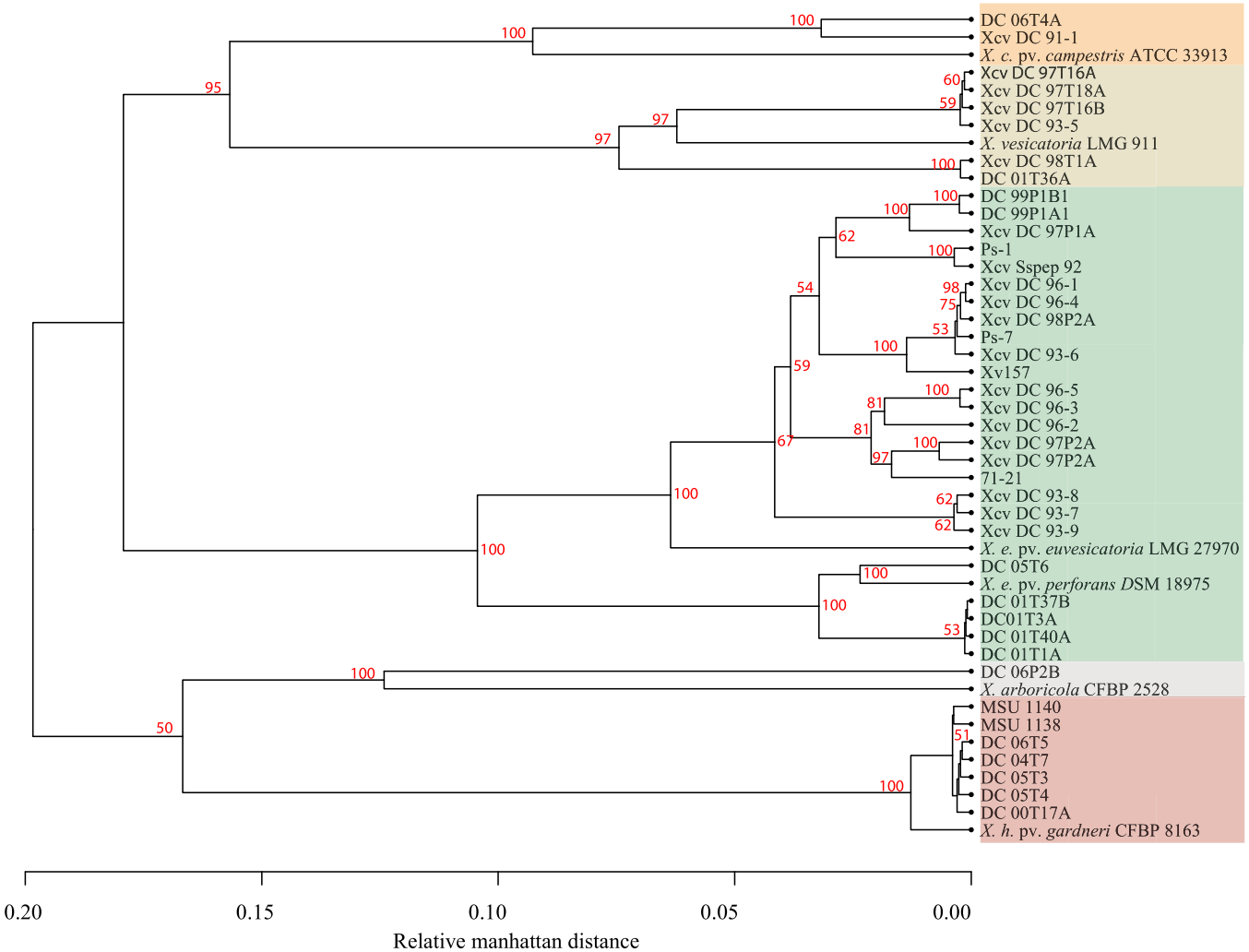

Fig. 1. Core-gene phylogenetic tree based on shared gene families shows five distinct clusters representing the different *Xanthomona*s spp. identified and the delineation supported by high bootstrap values. Bootstrap values <50% are not shown. The pangenome tree was created using the pancoreplot_tree command (Vesth et al. 2013) based on shared gene families.

The raw reads are deposited in the NCBI Sequence Read Archive under project number PRJNA779265.

Acknowledgments

We thank the staff of Molecular Technologies Laboratory, Ottawa Research and Development Centre, Ottawa, Canada (MTL-ORDC), for the Illumina NextSeq sequencing; the London Research and Development Centre (LRDC) for resources provided to D. Cuppels (retired) to collect these strains; and all of the summer students and technicians at ORDC and LRDC who worked on these projects.

Literature Cited

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.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Barak, J. D., Vancheva, T., Lefeuvre, P., Jones, J. B., Timilsina, S., Minsavage, G. V., Vallad, G. E., and Koebnik, R. 2016. Whole-genome sequences of *Xanthomonas euvesicatoria* strains clarify taxonomy and reveal a stepwise erosion of type 3 effectors. Front. Plant Sci. 7:1805.

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., 3rd, Stevens, R., Vonstein, V., Wattam, A. R., and Xia, F. 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5:8365.

Bushnell, B. 2014. BBTools software package. https://jgi.doe.gov/data-and-tools/software-tools/bbtools/

Buttner, D., and Bonas, U. 2010. Regulation and secretion of *Xanthomonas* virulence factors. FEMS Microbiol. Rev. 34:107-133.

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., Rooney, A. P., Yi, H., Xu, X. W., De Meyer, S., and Trujillo, M. E. 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int. J. Syst. Evol. Microbiol. 68:461-466.

Constantin, E. C., Cleenwerck, I., Maes, M., Baeyen, S., Van Malderghem, C., De Vos, P., and Cottyn, B. 2016. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. Plant Pathol. 65:792-806.

Cuppels, D. A., Louws, F. J., and Ainsworth, T. 2006. Development and evaluation of PCR-based diagnostic assays for the bacterial speck and bacterial spot pathogens of tomato. Plant Dis. 90:451-458.

Cuppels, D. A., Moore, R. A., and Morris, V. L. 1990. Construction and use of a nonradioactive DNA hybridization probe for detection of *Pseudomonas syringae* pv. *tomato* on tomato plants. Appl. Environ. Microbiol. 56:1743-1749.

da Silva, L. C. C., Targino, B. N., Furtado, M. M., Pinto, M. A. O., Rodarte, M. P., and Hungaro, H. M. 2017. Xanthan: Biotechnological production and applications. Pages 385-422 in: Microbial Production of Food Ingredients and Additives. A. M. Holban and A. M. Grumezescu, eds. Academic Press, London, U.K.

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E. M., Gabbard, J. L., Gerdes, S., Guard, A., Kenyon, R. W., Machi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E. K., Olsen, G. J., Olson, R. D., Overbeek, J. C., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomas, C., VanOeffelen, M., Vonstein, V., Warren, A. S., Xia, F., Xie, D., Yoo, H., and Stevens, R. 2020. The PATRIC Bioinformatics Resource Center: Expanding data and analysis capabilities. Nucleic Acids Res. 48:D606-D612.

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. Int. J. Syst. Evol. Microbiol. 57:81-91.

Jain, C., Rodriguez, R. L., 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.

Jones, J. B., Lacy, G. H., Bouzar, H., Stall, R. E., and Schaad, N. W. 2004. Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. Syst. Appl. Microbiol. 27:755-762.

Kim, M., Oh, H. S., Park, S. C., and Chun, J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of 662 prokaryotes. Int. J. Syst. Evol. Microbiol. 64:346-351.

Kuflu, K. M., and Cuppels, D. A. 1997. Development of a diagnostic DNA probe for xanthomonads causing bacterial spot of peppers and tomatoes. Appl. Environ. Microbiol. 63:4462-4470.

Lane, L. D. 1991. 16S/23S rRNA sequencing. Pages 115-175 in: Nucleic Acid Techniques in Bacterial Systematics. E. Stackebrandt and M. Goodfellow, eds. John Wiley & Sons, Ltd., Chichester, England.

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.

Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., and Goker, M. 2021. TYGS and LPSN: A database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res. 50:D801-D807.

Morinière, L., Burlet, A., Rosenthal, E. R., Nesme, X., Portier, P., Bull, C. T., Lavire, C., Fischer-Le Saux, M., and Bertolla, F. 2020. Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that *Xanthomonas cynarae* Trébaol 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.

Osdaghi, E., Rahimi, T., Taghavi, S. M., Ansari, M., Zarei, S., Portier, P., Briand, M., and Jacques, M. A. 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.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25:1043-1055.

Sijam, K., Chang, C. J., and Gitaitis, R. D. 1991. An agar medium for the isolation and identification of *Xanthomonas campestris* pv. *vesicatoria* from seed. Phytopathology 81:831-834.

Souvorov, A., Agarwala, R., and Lipman, D. J. 2018. SKESA: Strategic k-mer extension for scrupulous assemblies. Genome Biol. 19:153.

Tambong, J. T., Xu, R., and Bromfield, E. S. P. 2017. *Pseudomonas canadensis* sp. nov., a biological control agent isolated from a field plot under long-term mineral fertilization. Int. J. Syst. Evol. Microbiol. 67:889-895.

Timilsina, S., Kara, S., Jacques, M. A., Potnis, N., Minsavage, G. V., Vallad, G. E., Jones, J. B., and Fischer-Le Saux, M. 2019. Reclassification of *Xanthomonas gardneri* (ex Šutič 1957) Jones et al. 2006 as a later heterotypic synonym of *Xanthomonas cynarae* Trébaol et al. 2000 and description of *X. cynarae* pv. *cynarae* and *X. cynarae* pv. *gardneri* based on whole genome analyses. Int. J. Syst. Evol. Microbiol. 69:343-349.

Vancheva, T., Bogatzevska, N., Moncheva, P., Mitrev, S., Verniere, C., and Koebnik, R. 2021. Molecular epidemiology of *Xanthomonas euvesicatoria* strains from the Balkan peninsula revealed by a new multiple-locus variable-number tandem-repeat analysis scheme. Microorganisms 9:536.

Vesth, T., Lagesen, K., Acar, O., and Ussery, D. 2013. CMG-Biotools, a free workbench for basic comparative microbial genomics. PLoS One 8:e60120.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., Gillespie, J. J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E. K., Olson, R., Overbeek, R., Pusch, G. D., Shukla, M., Schulman, J., Stevens, R. L., Sullivan, D. E., Vonstein, V., Warren, A., Will, R., Wilson, M. J., Yoo, H. S., Zhang, C., Zhang, Y., and Sobral, B. W. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. 42:D581-D591.