Diversification of an emerging bacterial plant pathogen; insights from the global spread of *Xanthomonas euvesicatoria* pv. *perforans*

Short title: Intercontinental dissemination of *X. euvesicatoria* pv. *perforans*

Sujan Timilsina[1¶], Fernanda Iruegas-Bocardo[1¶], Mustafa O. Jibrin[1,2,3,#a¶], Anuj Sharma[1], Aastha Subedi[1], Amandeep Kaur[1], Gerald V. Minsavage[1], Jose Huguet-Tapia[1], Jeannie Klein-Gordon[1], Pragya Adhikari[12], Tika B. Adhikari[4], Gabriella Cirvilleri[5], Laura Belen Tapia de la Barrera[6], Eduardo Bernal[7], Tom C. Creswell[8], Doan Thi Kieu Tien[16], Teresa A. Coutinho[9], Daniel S. Egel[8], Rubén Félix-Gastélum[10], David M. Francis[7], Misrak Kebede[11], Melanie Lewis Ivey[15], Frank J. Louws[4,12], Laixin Luo[13], Elizabeth T. Maynard[14], Sally A. Miller[15], Nguyen Thi Thu Nga[16], Ebrahim Osdaghi[17], Alice M. Quezado-Duval[18], Rebecca Roach[19], Francesca Rotondo[15], Gail E. Ruhl[8], Vou M. Shutt[9,20], Petcharat Thummabenjapone[21], Cheryl Trueman[22], Pamela D. Roberts[1,2], Jeffrey B. Jones[1*], Gary E. Vallad[1,23*], Erica M. Goss[1,24*]

[1] Department of Plant Pathology, University of Florida, Gainesville, Florida, United States of America

[2] Department of Crop Protection, Ahmadu Bello University, Zaria, Nigeria

[3] Southwest Florida Research and Education Center, University of Florida, Immokalee, Florida, United States of America

[4] Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, North Carolina, United States of America

[5] Dipartimento di Agricoltura, Alimentazione e Ambiente, Sezione Patologia Vegetale, Catania, Italy

[6] Centro de Investigacíon an Alimentación y Desarrollo, Mexico

[7] Department of Horticulture and Crop Science, The Ohio State University, Wooster, Ohio, United States of America

[8] Botany and Plant Pathology Department, Purdue University, West Lafayette, Indiana, United States of America

[9] Department of Microbiology and Plant Pathology, Centre for Microbial Ecology and Genomics, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa

[10] Departamento de Ciencias Biológicas, Universidad Autónoma de Occidente, Unidad Los Mochis, Blvd. Macario Gaxiola y Carretera Internacional s/n Los Mochis, Sinaloa, México

[11] Biotechnology Department, Collage of Biological and Chemical Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

[12] Department of Horticultural Science, North Carolina State University
Raleigh, North Carolina, United States of America

[13] Department of Plant Pathology, China Agricultural University, Beijing, China

[14] Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, Indiana, United States of America

[15] Department of Plant Pathology, The Ohio State University, Wooster, Ohio, United States of America

[16] Department of Plant Protection, College of Agriculture, Can Tho University, Can Tho City, Vietnam

[17] Department of Plant Protection, College of Agriculture, University of Tehran, Karaj, Iran

[18] Embrapa Hortalicas. Laboratorio de Fitopatologia. Brasilia-DF, Brazil

[19] Queensland Department of Agriculture and Fisheries, Brisbane, Queensland, Australia

[20] Department of Plant Science and Biotechnology, University of Jos, Nigeria

[21] Khon Kaen University, Plant Science and Agricultural Resources, Section Plant Pathology, Khon Kaen, Thailand

[22] Department of Plant Agriculture, Ridgetown Campus, University of Guelph, Ridgetown, Ontario, Canada

[23] Gulf Coast Research and Education Center, University of Florida, Wimauma, Florida, United States of America

[24]Emerging Pathogen Institute, University of Florida, Gainesville, FL 32610, United States of America

[#a]Current Address: Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, Oklahoma, United States of America

[*] Corresponding authors

Email: emgoss@ufl.edu (EMG); gvallad@ufl.edu (GEV); jbjones@ufl.edu (JBJ)

[¶]These authors contributed equally to this work.

Abstract

Emerging and re-emerging plant diseases continue to present multifarious threats to global food security. Considerable recent efforts are therefore being channeled towards understanding the nature of pathogen emergence, their spread and evolution. *Xanthomonas euvesicatoria* pv. *perforans (Xep*), one of the causal agents of bacterial spot of tomato, rapidly emerged and displaced other bacterial spot xanthomonads in tomato production regions around the world. In less than three decades, it has become a dominant xanthomonad pathogen in tomato production systems across the world and presents a model for understanding diversification of recently emerged bacterial plant pathogens. Although *Xep* has been continuously monitored in Florida since its discovery, the global population structure and evolution at the genome-scale is yet to be fully explored. The objectives of this work were to determine genetic diversity globally to ascertain if different tomato production regions contain genetically distinct *Xep* populations, to examine genetic relatedness of strains collected in tomato seed production areas in East Asia and other production regions, and to evaluate variation in type III effectors, which are critical pathogenicity and virulence factors, in relationship to population structure. We used genome data from 270 strains from 13 countries for phylogenetic analysis and characterization of *Xop* effector gene diversity among strains. Our results showed notable genetic diversity in the pathogen. We found genetically similar strains in distant tomato production regions, including seed production regions, and diversification over the past 100 years, which is consistent with intercontinental dissemination of the pathogen in hybrid tomato production chains. Evolution of the *Xep* pangenome, including the acquisition and loss of type III secreted effectors, is apparent within and among phylogenetic lineages. The apparent long-distance movement of the pathogen, together with variants that may not yet be widely distributed, poses risks of emergence of new variants in tomato production.

Introduction/Main

Emerging and re-emerging plant diseases are a constant threat to global food security [1-3]. Bacterial plant pathogens cause some of the most intractable diseases of crops worldwide [4-7]. Novel emergence and re-emergence of bacterial diseases continue to be reported across the globe and is associated with an upsurge in efforts devoted to understanding the nature of pathogen emergence, spread, and evolution [8-17]. A bacterial plant pathogen that emerged in the last few decades and is of global epidemiological consequences is *Xanthomonas euvesicatoria* pv. *perforans,* one of the causal agents of bacterial spot of tomato [18].

Bacterial spot disease of tomato affects all aboveground plant parts including leaves, stems, flowers and fruit. Under optimal environmental conditions, fruit lesions and/or extensive defoliation can dramatically limit marketable yields and poses a continuous challenge to tomato production [19-21]. Once epidemics are initiated, growers have limited management tools and have relied heavily on copper-based bactericides. However, reliance on copper compounds has led to widespread copper tolerance [22-30]. Alternative bactericides are often costly, provide insufficient control when the weather favors rapid disease development, and rarely improve yields. While historically four taxa have caused this disease, *Xanthomonas euvesicatoria* pv. *perforans* (*Xep*) [31] (syn. *X. perforans* [32, 33]) has emerged rapidly and become a major player on tomato [18, 27, 34-41]. *Xep* was first reported in 1991 in Florida, USA [32] and is now found in all tomato production areas of the world, including regions with no history of the disease [42]. *Xep* has been isolated from tomato seed [32]; therefore, a plausible hypothesis for new outbreaks of *Xep* is pathogen movement with seeds and planting materials [43].

Tomato production is characterized by a high seed replacement rate (99.3%), meaning that growers require seeds each season, which in turn requires large-scale seed production [44]. Tomato hybrid seed production is concentrated in geographic areas where environmental conditions minimize seed contamination by pathogens and seed production costs are low. These seed production regions supply hybrid seeds globally for commercial production of tomato fruits for the fresh market or for processing into tomato products (e.g., sauce, paste, and diced tomatoes). The long-distance movement of seeds poses

a high risk for dissemination of seed borne pathogens to commercial tomato production areas. Seedlings are typically grown in transplant facilities and then transplanted into fields for the regional and international transplant markets, potentially amplifying and further disseminating seed-borne pathogens [45].

The success of *Xep* as a pathogen has been attributed to its production of bacteriocins against competing bacterial spot species, rapid genome evolution via recombination affecting the chromosome, and introduction of genes via horizontal gene transfer that contribute to fitness in tomato fields [46-52]. Distinct genetic lineages of *Xep*, each with unique patterns of allelic variation among core genes (genes present in all strains), were identified in fresh market and processing tomato production fields in the United States [47, 50, 51, 53, 54]. Additional lineages of *Xep* were found in Nigeria, Iran, Italy, and the Southwest Indian Ocean islands [42, 48, 55].

Xanthomonas perforans strains, like other xanthomonads, acquire nutrients through colonization of compatible hosts. The type III secretion system (T3SS) and type III effector (T3E) proteins are critical for suppression of host defenses and virulence by *Xep* [56]. Effector content varies among *Xanthomonas* species and distinct lineages of *Xep* have distinguishable effector content [23, 48, 50, 57, 58]. Strains of *Xep* isolated in the 1990s were limited to tomato [59], but now strains of *Xep* are causing bacterial spot disease of pepper [50, 58, 60]. Host range expansion was attributed, in part, to loss of effectors that act as avirulence factors in pepper and other genomic changes as a result of recombination with other *Xanthomonas* lineages, including pepper pathogenic *X. euvesicatoria* pv. *euvesicatoria*. Effector variation may cause differences in disease epidemiology in addition to host range [49, 57, 61]. For example, wildtype strains with the acquired effector XopJ2 showed three times faster spread in the field than isogenic mutant strains without the effector [51].

Emerging pathogens may show limited genetic variation if they experienced a bottleneck during the ecological and evolutionary processes that often precede emergence (e.g., host jump or introduction event) [62]. *Xep* appears genetically diverse but it is not known how this variation is structured across global tomato production regions. The first objective of this work was to determine if different tomato

production regions contain genetically distinct *Xep* populations. Second, we asked if there was evidence for long-distance pathogen dissemination, as would be indicated by genotypes shared among distant regions. Specifically, we obtained strains from tomato seed production areas in East Asia and determined if they resembled strains from fruit production fields elsewhere in the world, which would be expected if strains are being disseminated in seeds. Third, we estimated the timing of *Xep* population expansion relative to its first report in 1991. Finally, we evaluated T3SE content and allelic variation in the context of geography and core genome variation as a proxy for genetic variation in virulence. Overall, we found extensive genetic diversity within *Xep;* genetically similar strains in distant geographic regions, inclusive of seed production regions; evidence of diversification prior and subsequent to the first report of emergence; and lineage-specific T3SE repertoires. Together, these results illustrate the capacity for this pathogen to rapidly evolve and strongly support the potential for intra- and intercontinental movement of pathogens in tomato production systems.

Results

X. euvesicatoria pv. *perforans* strains from seed and commercial fruit production areas

A total of 270 *Xep* genomes from 13 different countries – representing seed and fruit production – were used in this study (Table 1). We generated new genome sequence data for 153 strains (S1 Table; NCBI BioProject PRJNA941448). *Xep* strains were differentiated from other tomato-pathogenic xanthomonads using a real-time qPCR assay that specifically amplifies the *hrcN* (*hrpB7*) gene in *Xep* [63] and inoculated on tomato cv. 'Bonny Best' to confirm pathogenicity. Strains from China, Thailand, and Vietnam were collected from seed production areas (n = 31) and all other strains (n = 239) were collected in commercial fruit production areas from Australia, Brazil, Canada, Ethiopia, Iran, Italy, Mexico, Nigeria, South Africa, and the United States. Within the US, strains were collected from seven different states in the Midwest and Southeast, including strains collected since 1991 from Florida.

186 **Table 1. *Xanthomonas euvesicatoria* pv. *perforans* strains used in this study.**

Country	Locality	Year	Strain (original name, if applicable)
Australia [57]	Queensland	2015	Aus3, Aus7, Aus14
		2016	Aus5, Aus10, Aus11
		2017	Aus1, Aus15, Aus16
Brazil [39]	São Paulo	2011	Bzl1 (2011-107), Bzl2 (2011-132)
	Goiás	2012	Bzl3 (2012-08)
	Goiás, São Paulo	2013	Bzl5 (2013-16), Bzl6 (2013-42)
	Goiás, Minas Gerais	2014	Bzl7 (2014-10), Bzl8 (2014-17)
	Minas Gerais	2015	Bzl10 (2015-53), Bzl11 (2015-56)
	Goiás	2016	Bzl13 (2016-08)
	Goiás	2017	Bzl14 (2017-21)
Canada	Ontario	2016	4A, 4D, 12A, 14A
China		2016	CHI-3, CHI-5, CHI-6, CHI-7, CHI-8, CHI-10, CHI-12, CHI-15, CHI-18
Ethiopia [36]		2011	ETH5, ETH11, ETH21, ETH25, ETH33
Iran [41]		2013	K41, F210, F215, TOM801, TOM816
Italy [64]		2011	1P6S1, 2P4S1, 2P4S1D, 2P6S1, 1P4S1D
Mexico			Mexico-1, Mexico-3, Mexico-LT1, Mexico-LT3, Mexico-LT5
Nigeria [37, 65]		2014	NI-1, NI-2, NI-4, NI-7, NI-12, NI-13
		2015	KS3, KS5, KS9, KS28
South Africa	Pretoria		X2-B14, X10-B85, X59-BD1351, X47-BD167

Vietnam			SEA-3, SEA-5, SEA-21, SEA-23
Thailand		2016	THA-8, THA-14, THA-40, THA-45, THA-54, THA-72, THA-81A, THA-100, THA-112, THA-116, THA-119, THA-120, THA-126, THA-127, THA-128, THA-132, THA-135, THA-157A
United States	Alabama [66]	1996	Xp1861
	Indiana [25]	2016	16-1165A1, 16-1181-2, 16-1182A, 16-1184A, 16-1187A, 16-1205A, 16-1402A, 16-974C, 16-990A, 16-990C
		2014	14-463-1A
	Florida [32, 33, 43, 47, 58, 67]	1991	XV0938, Xp91-118, Xp894, Xp909, Xp1183
		1992	Xp1118, Xp1144
		1993	Xp1241, Xp1268, Xp1275
		1994	Xp1550, Xp1564
		1995	Xp1797, Xp1805
		1996	Xp1856
		1997	Xp1912
		1998	Scott-1, Xp1920
		2006	Xp1-5, Xp1-6, Xp3-12, Xp3-15, Xp3-16, Xp3-8, Xp4-20, Xp5-14, Xp5-6, Xp5-9, Xp7-12, Xp8-16, Xp9-5, Xp10-13, Xp11-2, Xp15-11, Xp17-12, Xp18-15
		2007	Xp4B
		2010	Xp2010
		2011	GEV485

		2012	GEV839, GEV872, GEV893, GEV904, GEV909, GEV915, GEV917, GEV936, GEV940, GEV968, GEV993, GEV1001, GEV1026, GEV1044, GEV1054, GEV1063
		2013	TB6, TB9, TB15
		2015	GEV2047, GEV2048, GEV2049, GEV2050, GEV2052, GEV2055, GEV2058, GEV2059, GEV2060, GEV2063, GEV1989, GEV1991, GEV1992, GEV1993, GEV2004, GEV2009, GEV2010, GEV2011, GEV2013, GEV2015, GEV1911, GEV1912, GEV1913, GEV1914, GEV1915, GEV1916, GEV1917, GEV1918, GEV1919, GEV1920, GEV1921
		2016	GEV2065, GEV2067, GEV2072, GEV2087, GEV2088, GEV2089, GEV2097, GEV2098, GEV2099, GEV2108, GEV2109, GEV2110, GEV2111, GEV2112, GEV2113, GEV2114, GEV2115, GEV2116, GEV2117, GEV2118, GEV2119, GEV2120, GEV2121, GEV2122, GEV2123, GEV2124, GEV2125, GEV2126, GEV2127, GEV2128, GEV2129, GEV2130, GEV2132, GEV2133, GEV2134, GEV2135
	Louisiana [24]	2013	mli-2
	North Carolina [27]	2015	NC-14, NC-47, NC-67, NC-101, NC-112, NC-204
		2016	NC-242, NC-252, NC-282, NC-289, NC-350, NC-373, MRS-30P-011

	South Carolina [43]	2016	GEV2407, GEV2408, GEV2384, GEV2388, GEV2389, GEV2390, GEV2391, GEV2392, GEV2393, GEV2396, GEV2397, GEV2399, GEV2400, GEV2403, GEV2410, GEV2420
	Ohio [53]	2017	SM-1806, SM-1807, SM-1808, SM-1809, SM-1810, SM-1811, SM-1812, SM-1813, SM-1814, SM-1815, SM-1828, SM-1829, SM-1830, SM-1831

Genomic diversity in *X. euvesicatoria* pv. *perforans*

To examine genetic diversity in the core genome, we curated a set of 887 genes that were present in all 270 *Xep* genomes based on IMG/JGI gene annotation. The aligned sequence length of concatenated core genes was 617,855 bp, which contained 14,427 polymorphic sites after removing ambiguous nucleotides and any alignment gaps (S1 Data). Maximum likelihood phylogenetic analysis showed a distinct and especially diverged lineage of 11 strains from Nigeria and Thailand (S1 Figure), that included a previously defined atypical strain – NI1 – from Nigeria [48]. Grouping strains by state within the United States and country elsewhere produced an FST [68] of 0.66. The number of core gene SNPs by geographic location represented by more than one strain ranged from 15 to 5929 (S2 Table). Nucleotide diversity (average number of differences among sequences) ranged from 3 to 1287, with both extremes in nucleotide diversity coming from the Midwestern U.S., Ohio and Indiana respectively (S2 Table). Tajima's D [69] by geographic location ranged from –2.0 to 1.7 (S2 Table).

Phylogenetic analysis of core SNPs, followed by correction of branch lengths for recombination, showed diversifying lineages of *Xep* (Figure 1; S1 Figure). After excluding a particularly diverged lineage of 11 strains (S1 Figure), ClonalFrameML [70] estimated an overall ratio of recombination rate to mutation rate (R/theta) of 0.60, with recombination causing approximately seven times more base changes than mutation (delta = 231; nu = 0.05). There were an estimated 221 recombination events that

affected more than 96 Kbp in terminal branches and 494 recombination events detected in internal branches encompassing 190 Kbp.

Figure 1. Population structure of *Xanthomonas euvesicatoria* pv. *perforans* strains collected from tomato production regions. (A) Maximum likelihood phylogenetic tree of 259 *X. euvesicatoria* pv. *perforans* strains constructed with nucleotide sequences from 887 core genes, corrected for recombination by ClonalFrameML. Tips are colored according to clusters identified by hierBAPS. Cluster 10 strains (n=11) were genetically distant and excluded from the tree (see S1 Figure). Nucleotide alignment is available as S1 Data. (B) Distribution of 270 strains in each hierBAPS cluster by country or state of collection. Abbreviations are as follows: AUS – Australia; BRA – Brazil; CAN – Canada; CHN – China; ETH – Ethiopia; FL – Florida, USA; IN – Indiana, USA; IRN – Iran; ITA – Italy; LOU – Louisiana, USA; MX – Mexico; NC – North Carolina, USA; NI – Nigeria; AL – Alabama; OH – Ohio, USA; SA – South Africa; SC – South Carolina, USA; THA – Thailand; VTM – Vietnam. (C) Map showing distribution of clusters by country of collection. Countries connected by lines show selected instances of genetically similar strains collected in different countries. (D) Pairwise comparison of whole genome average nucleotide identity (ANIb) of selected strains shows high identity between strains isolated from different continents. For each comparison, genome coverage is shown by grayscale in boxes, scale shown to the right. Values for each comparison are for genomes in rows when compared to genomes in columns. See S3 Table for additional ANI output.

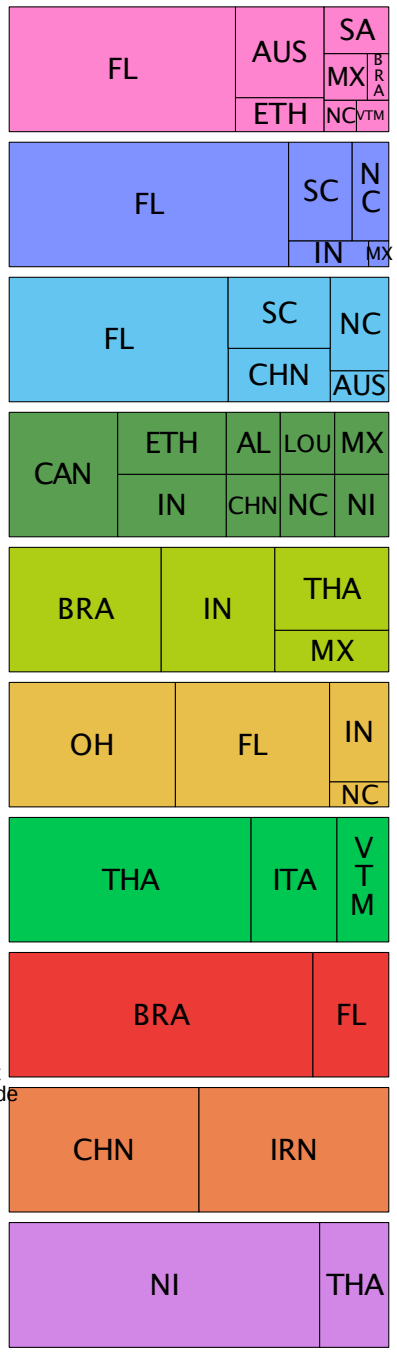
To summarize population structure based on core gene SNPs, we used hierBAPS [71], which assigned individual strains to 9 clusters using allele frequencies (Figure 1; S1 Table). This analysis excluded the 11 highly diverged strains from Nigeria and Thailand, which we designated as cluster 10. F_{ST} among clusters was 0.80. In some cases, clusters corresponded to phylogenetic lineages, including clusters 2, 3, 7, 8, and 9 (Figure 1). The remaining clusters were polyphyletic, encompassing multiple diverged clades or individual strains. Nucleotide diversity within clusters ranged from 13.8 to 675.9 and

A

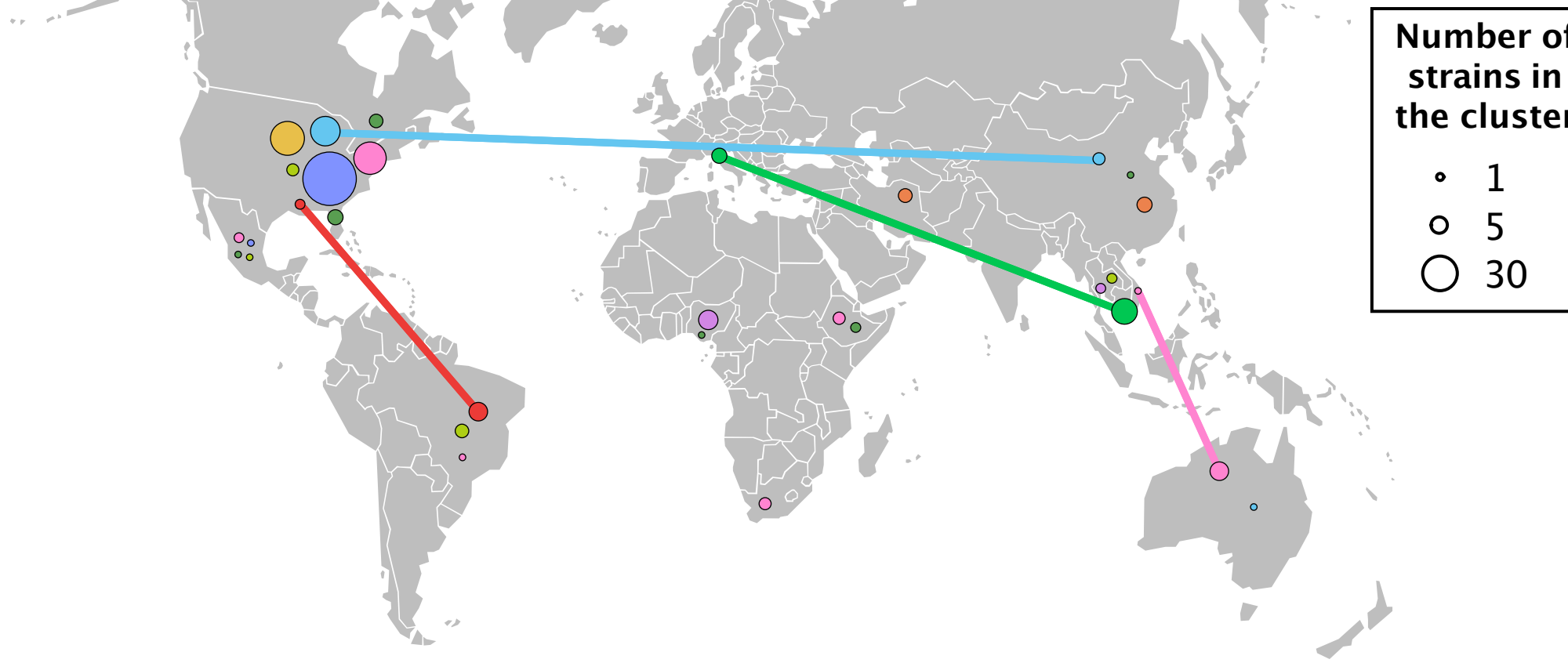
Cluster 3
Cluster 9
Cluster 4
Cluster 7
Cluster 5
Cluster 1
Cluster 8
Cluster 6
Cluster 2

1×10^{-4}

B

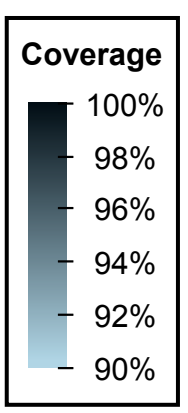


C



D

	VTM-5 (cluster 7)	THA-5 (cluster 7)	1P4S1D (cluster 7)	VTM-23 (cluster 1)	Aus3 (cluster 1)	CHI-5 (cluster 3)	NC350 (cluster 3)	GEV2099 (cluster 3)
VTM-5 (cluster 7)	100.00	99.94	99.95	99.65	99.66	99.69	99.55	99.54
THA-5 (cluster 7)	99.99	100.00	99.99	99.75	99.75	99.75	99.75	99.75
1P4S1D (cluster 7)	99.99	99.99	100.00	99.73	99.74	99.74	99.75	99.74
VTM-23 (cluster 1)	99.74	99.76	99.76	100.00	99.99	99.80	99.77	99.77
Aus3 (cluster 1)	99.74	99.76	99.76	99.98	100.00	99.81	99.77	99.77
CHI-5 (cluster 3)	99.74	99.75	99.75	99.78	99.79	100.00	99.99	99.98
NC350 (cluster 3)	99.57	99.69	99.68	99.69	99.70	99.92	100.00	99.98
GEV2099 (cluster 3)	99.57	99.71	99.70	99.70	99.70	99.93	99.99	100.00
	Vietnam	Thailand	Italy	Vietnam	Australia	China	USA	USA



Tajima's D from –2.6 to 1.0. Presence-absence gene variation in the pangenome largely paralleled the phylogenetic diversity of core genes in that polyphyletic clusters 1, 4, and 5 also showed the most variation in gene content (S2 Figure).

Geographic distribution of *X. euvesicatoria* pv. *perforans* core gene clusters

Cluster 1 encompasses genetically diverse strains from seven countries, including most of the strains from Australia, all four strains from South Africa, and one strain from Southeast Asia (Figure 1B). All USA strains assigned to cluster 1 were isolated in or before 2006 from Florida except for one strain from North Carolina. Cluster 2 contains 88 strains from the United States and one from Mexico, while Cluster 3 includes strains isolated from Florida, North and South Carolina, China, and Australia. Cluster 4 encompasses multiple lineages of strains from the United States, Canada, Ethiopia, China, and Nigeria. Cluster 5 is polyphyletic with diverged strains from three continents. Cluster 6 was isolated only within the United States from Florida, Indiana, North Carolina, and Ohio. Cluster 7 is a monophyletic group of strains from Southeast Asia and Italy. Cluster 8 is another monophyletic group found only in Brazil and Florida. Cluster 9 includes two clades of strains, one from China and the other from Iran and Nigeria. Cluster 10 comprises the atypical strains from Nigeria and similar strains from Thailand. Most countries contained strains from more than one core gene cluster (Figure 1C).

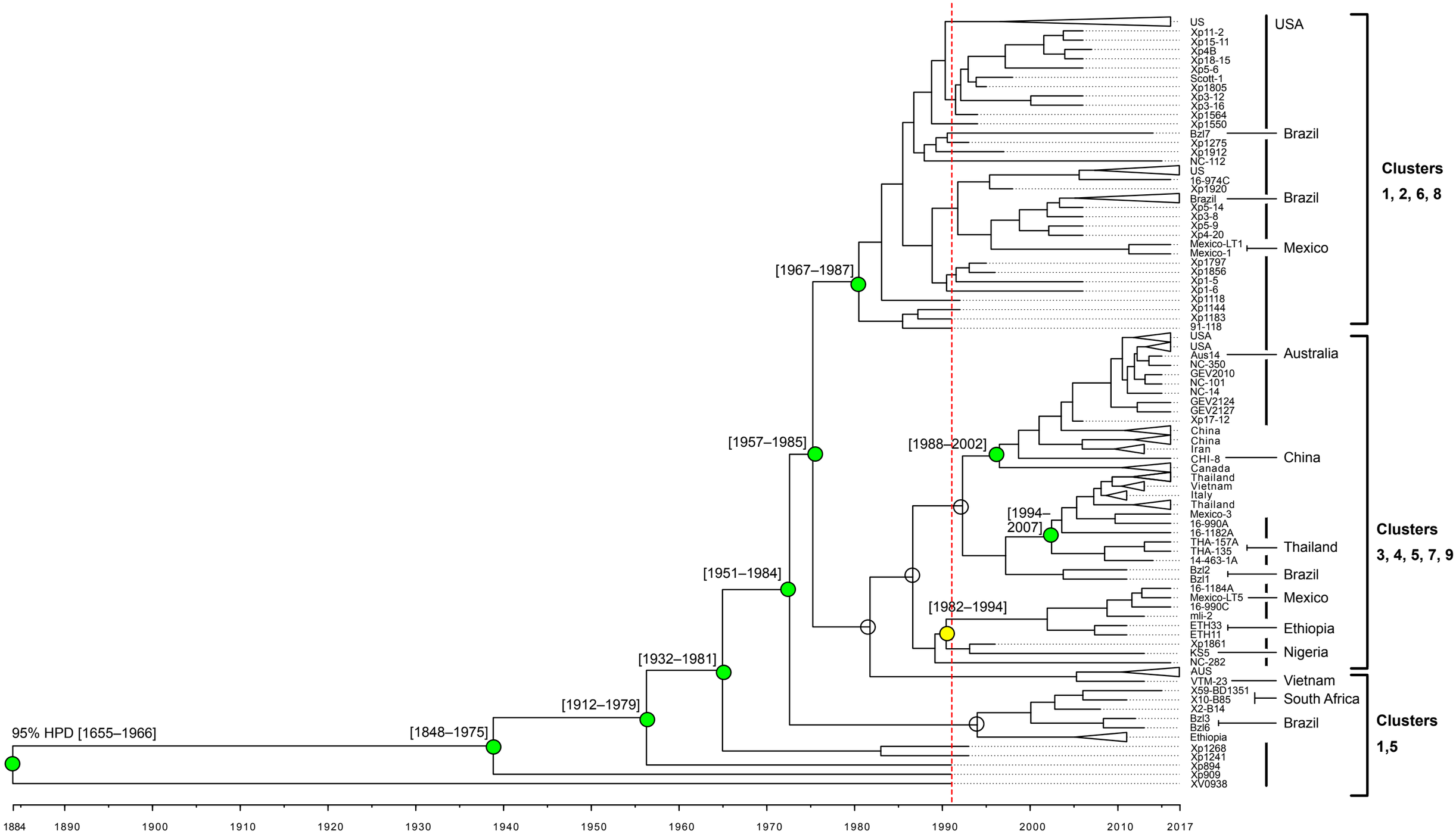
Clusters 1, 3, 4, 5, 7, 9, and 10 contain strains isolated from both seed production and commercial fruit production regions, whereas strains in clusters 2, 6, and 8 were only isolated from commercial fruit production regions. Some strains found on different continents were nearly identical in core gene sequences (Figure 1D) with very high average nucleotide identity. Strains in cluster 1 from Australia differed by 6 to 10 SNPs in more than 617 Kbp of core gene sequence from strain VTM-23 from Vietnam. Pairwise average nucleotide identity (ANIb) between VTM-23 to Aus3 was 99.99% with alignment fraction of 0.998 (Figure 1D). Strains from the USA had up to 99.87 ANIb with strains from Australia and Vietnam (S3 Table). A different strain from Vietnam, VTM-5 in cluster 7, had as few as four SNPs in the core genome when compared to strains from Italy and ANIb of 99.95% to Italian strain

1P4S1D (Figure 1D). Likewise, strains collected in a seed production region in China had ANIb up to 99.99% with strains from Florida and North Carolina. We also found similar strains between Brazil and USA, for example Bzl-10 (Minas Gerais) and Xp3-8 (Florida) had greater than 99.9% ANI (S3 Table). Other strains were similar between countries in core genes only after correction for recombination.

Timing of *X. euvesicatoria* pv. *perforans* lineage emergence

We used the years of strain collection to estimate the timing of diversification of our sample of *Xep*, excluding the cluster 10 strains. We inferred dated phylogenies using whole genome alignments with inferred recombinant sites removed by Gubbins [72]. Due to recombination with other *X. euvesicatoria* lineages, we did not include an outgroup (S1 Figure, part B; [48]). Sampling year was significantly correlated with root-to-tip distance ($R^2 = 0.20$ for the whole genome alignment, $P < 1\times10^{-4}$, S3 Figure). The root inferred by the BactDating R package [73] was placed between strains isolated in Florida in 1991. The most recent common ancestor (MRCA) of all strains was dated to 1884 (95% HPD: 1655–1966). Notably, strains that were isolated in the early 1990s, when *Xep* was first detected in U.S. tomato production [32, 37], represented multiple lineages (Figure 2). The MRCA of the clade representing core gene clusters 1, 2, 6, and 8 (including strains from USA, Brazil, and Mexico) was dated to 1980 (95% HPD: 1967–1987). A major clade, encompassing strains in clusters 3, 4, 5, 7, 9, which were collected in Africa, the Americas, Asia, Australia, and Europe, did not have a significant temporal signal across the clade. We repeated the analysis with BEAST, which inferred a different rooting. The tree inferred by BEAST placed the root between two strains isolated in 2011 from Brazil and all other strains (S4 Figure, part B). The MCRA of the BEAST tree was dated to 1868 (95% HPD: 1862–1919), which was similar to the root date estimated using BactDating (S4 Figure).

Figure 2. Dated phylogeny of 259 *X. euvesicatoria* pv. *perforans* strains. BactDating analysis estimated an approximately 130-year history for *Xep* strains in core gene clusters 1 through 9 (Figure 1). Red dotted line indicates the first documented isolations in 1991. Internal nodes were collapsed for clades containing

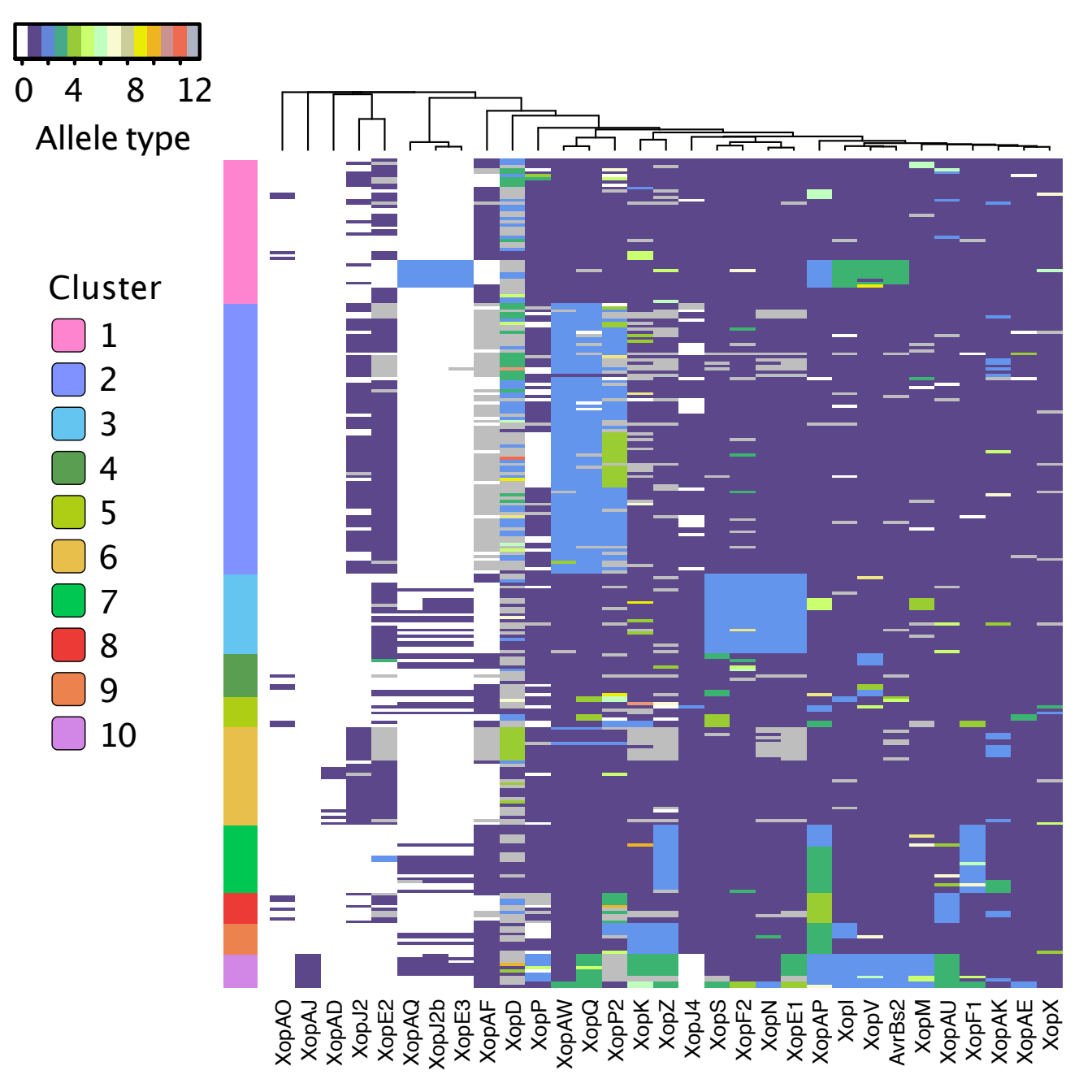

US
Xp11-2
Xp15-11
Xp4B
Xp18-15
Xp5-6
Scott-1
Xp1805
Xp3-12
Xp3-16
Xp1564
Xp1550
Bzl7
Xp1275
Xp1912
NC-112
US
16-974C
Xp1920
Brazil
Xp5-14
Xp3-8
Xp5-9
Xp4-20
Mexico-LT1
Mexico-1
Xp1797
Xp1856
Xp1-5
Xp1-6
Xp1118
Xp1144
Xp1183
91-118
USA
USA
Aus14
NC-350
GEV2010
NC-101
NC-14
GEV2124
GEV2127
Xp17-12
China
China
Iran
CHI-8
Canada
Thailand
Vietnam
Italy
Thailand
Mexico-3
16-990A
16-1182A
THA-157A
THA-135
14-463-1A
Bzl2
Bzl1
16-1184A
Mexico-LT5
16-990C
mli-2
ETH33
ETH11
Xp1861
KS5
NC-282
AUS
VTM-23
X59-BD1351
X10-B85
X2-B14
Bzl3
Bzl6
Ethiopia
Xp1268
Xp1241
Xp894
Xp909
XV0938
USA
Brazil
Brazil
Mexico
Australia
China
Thailand
Brazil
Mexico
Ethiopia
Nigeria
Vietnam
South Africa
Brazil
Clusters
1, 2, 6, 8
Clusters
3, 4, 5, 7, 9
Clusters
1,5
[1967–1987]
[1957–1985]
[1988–2002]
[1994–
2007]
[1951–1984]
[1982–1994]
[1932–1981]
[1912–1979]
[1848–1975]
95% HPD [1655–1966]
1884
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2017

283 strains from a single country with branch tips indicating country or strain (for full tree see S4 Figure).
284 Vertical line to the right of tip labels indicates strains from USA; other countries are labeled. Temporal
285 signal was assessed using Phylostems and results are shown for major nodes (for full results see S3
286 Figure). Empty circles indicate no significant temporal signal. Colored circles indicate nodes with
287 statistically significant temporal signal based on adjusted R^2 values: green – 0.13-0.19; yellow – 0.45. The
288 95% highest posterior density (95% HPD) of date estimates for major nodes with significant temporal
289 signals are shown in brackets.
290
291 Type III effector content
292 We detected 32 predicted type III effectors in our collection of 270 strains (S5 Figure, S4 Table).
293 The diversity in amino acid sequences of predicted effectors ranged widely from a single conserved allele
294 to 8 or more alleles per locus (Figure 3). None of the effectors were present and intact in 100% of our
295 genomes, in part due to our analysis of draft genomes. The following effector genes were present in more
296 than 95% of strains and can be considered "core effectors": *avrBs2*, *xopF1*, *xopF2*, *xopI*, *xopM*, *xopQ*,
297 *xopS*, *xopV*, *xopX*, *xopAE*, *xopAK*, *xopAP*, *xopAU*, and *xopAW*. The genes for *xopD*, *xopE1*, and *xopN*
298 were present in some form in all genomes but more than 5% of strains contained a contig break within the
299 gene. A closer examination of *xopD* by PCR and Sanger sequencing showed this to be an assembly issue
300 due to the repeats within the gene. Effectors at low frequency in our *Xep* strains (<25%) were *xopE3*,
301 *xopAD*, *xopAJ*, *xopAO*, and *xopAQ*. Transcription activator-like (TAL) effectors typically do not assemble
302 in draft genomes due to their characteristic repeat sequences, but there were BLAST hits to previously
303 described TAL effectors in 65 strains. We Sanger sequenced the TAL effector gene in strain 2P6S1
304 collected in Italy (NCBI accession number OQ588696), which confirmed that it had the same repeat
305 variable diresidues as PthXp1 reported in *Xep* strains from Alabama [50]. Previous phenotyping and
306 sequencing indicated that the strain isolated in Louisiana, USA origin has AvrHah1 [16].
307

Figure 3. Variation in type III effectors (Xop proteins) in *Xanthomonas perforans*. Type III effectors are in columns and 270 *X. euvesicatoria* pv. *perforans* strain in rows. Effector status is shown by allele type: absence is indicated by allele type 0 (white), while the most frequent allele observed when the effector is present is allele type 1 (purple), second most frequent is allele type 2 (blue), and so on. Putative pseudogenized effectors are shown as allele 13 (gray). The order of columns was determined by hierarchical clustering analysis, placing similarly distributed effectors adjacent to each other. Genomes showing BLAST hits to TAL effector(s) are indicated in Supporting Information Table S3 and not shown in heatmap. *X. euvesicatoria* pv. *perforans* strains (rows) are organized by core gene cluster.

The T3E effector XopAF (AvrXv3), which is targeted by the tomato resistance gene *Xv3* [74], was missing or pseudogenized in 64% of strains. Most strains examined from the United States did not have a complete copy of this gene, whereas it was intact in many strains collected in Asia and Africa. The gene for XopJ4 (AvrXv4), recognized by resistance gene *RXopJ4* from *S. pennellii* [75], was present in 88% of strains and absent in all cluster 10 strains and 19 of 88 cluster 2 strains. XopJ2 (AvrBsT), which elicits an HR in pepper but increases virulence in tomato [49], was present in less than half of strains examined (43%) and overwhelmingly in strains from the United States. A homolog of XopJ2, recently designated XopJ2b [76], was present in 50 strains, including two strains from Australia that carried both copies of XopJ2 (S4 Table).

We tested for evidence of positive selection in T3E by estimating synonymous and non-synonymous (dN/dS) substitution rates using a Bayesian approach for detecting pervasive selection (FUBAR, [77]) and maximum likelihood approach for detecting episodic selection (MEME, [78]). We found evidence of pervasive positive selection affecting at least one amino acid in AvrBs2, XopD, XopE1, XopF2, XopK, XopM, XopP and its paralog XopP2, XopQ, XopS, and XopAQ (S4 Table). We found evidence of episodic selection affecting at least one amino acid in XopF2, XopK, XopP, XopP2, XopQ, XopV, and XopAP (S4 Table).


0 4 8 12
Allele type
Cluster
1
2
3
4
5
6
7
8
9
10
XopAO
XopAJ
XopAD
XopJ2
XopE2
XopAQ
XopJ2b
XopE3
XopAF
XopD
XopP
XopAW
XopQ
XopP2
XopK
XopZ
XopJ4
XopS
XopF2
XopN
XopE1
XopAP
XopI
XopV
AvrBs2
XopM
XopAU
XopF1
XopAK
XopAE
XopX

We defined the effector profile of each strain as the predicted presence or absence of each effector and its allelic state, excluding TAL effector hits. Grouping effector profiles according to core gene cluster revealed that allelic variation of effectors often paralleled core genome variation (Figure 3). For example, specific alleles of effectors XopAW, XopQ, and XopP2 were mostly limited to strains in cluster 2. Cluster 3 strains carried unique alleles for effectors XopF2, XopS, XopN, and XopE1, and strains from cluster 7 shared unique alleles for effectors XopF1 and XopZ. Strains from highly diverged cluster 10 had rare alleles in many effectors, and it was the only cluster in which effector XopAJ was found (Figure 3). To visualize variation among strain effector profiles independent of core gene clusters, we transformed dissimilarities between profiles into distances represented in a two-dimensional plot and defined eight effector profile clusters (S6 Figure, part A). A lack of low frequency effectors characterized effector cluster A, containing 188 strains from 11 of 13 countries (S6 Figure). The remaining effector clusters were defined by the presence of one to three low frequency effectors (S7 Figure). While most effectors were found in multiple countries and continents (S6 Figure), populations in Brazil, Ethiopia, Nigeria, Thailand, South Africa, and the United States contained low frequency effectors that were not widely distributed.

Copper resistance genes

Xanthomonads, including *Xep*, have acquired genes conferring copper tolerance, likely in response to exposure to copper-based bactericides [30, 79-81]. In *Xep*, copper tolerance is conferred by an operon containing the copper resistance genes *copA* and *copB*, and regulator *copL* (*copLAB*) [80]. BLAST analysis showed that these genes were present in 73% of the genomes in our sample (S5 Table). Copper resistance genes are prevalent in the USA; only the genomes from strains isolated from Florida in the early 1990s and a strain from Louisiana lacked *copLAB*. The genes were also missing in the genomes of a few strains from Australia (1), Brazil (2), Ethiopia (2), Mexico (1), and Vietnam (2). In contrast, the genes were absent in all genomes of all strains from Nigeria, China, Iran, Italy, and Thailand.

Discussion

Emerging plant pathogens have the potential for global outbreaks, exacerbated by complex trade networks. Hybrid tomato production relies on international breeding and production chains with a global network to deliver seeds to growers. Global trade associated with vegetable seed production provides a pathway for global spread of pathogens, with quantities traded that challenge even strong phytosanitary measures [82, 83]. Over 100 countries import seeds of tomatoes and other vegetables; for example, 11.7 million kg of vegetable seed were imported to the USA in 2019, with China being the biggest supplier at 2.4 million kg [84]. *Xanthomonas* species can infest pepper and tomato seed [85], and *Xep* has been isolated from tomato seed [32, 86], supporting the hypothesis that seeds can be a source of inoculum for bacterial spot outbreaks [87, 88]. Thirty years after its first report, *Xep* has been identified in tomato production areas around the world [18]. Our results showed extensive genetic diversity in the pathogen, but also genetically similar strains in distant tomato production regions. Furthermore, we found genetically similar strains in seed production and fruit production regions on different continents, as would be expected if the pathogen was being moved in shared production chains. Dated phylogenies indicate multiple waves of diversification of the *Xep* population, before and since its first detection in 1991. Variation in gene content confirms that *Xep* acquired and lost type III effectors during its diversification, which will continue to challenge sustainable management of tomato bacterial leaf spot [49, 66].

Using our broad strain collection, we found *Xep* variants in seed production regions in Asia that were previously reported in Australia, Italy, Nigeria, and the United States [43, 47, 48, 57]. Strains from Italy were nearly identical in core genes and very similar in accessory genomes to strains collected from Thailand and Vietnam, both major seed production regions. The atypical bacterial spot strains from Nigeria, recently designated as race T5 [21], were genetically similar to strains from Thailand. A recently described variant of *Xep* in Florida [cluster 3; [43, 47]], which was also found in Australia [57], was similar in the core genome to strains found in China; however, these strains showed divergence in the pangenome, consistent with accessory genome evolution in emergent populations. Beyond previously

385 described variants, we found strains in Iran that were closely related to strains from China; multiple
386 instances of genetic similarity between strains from North America and Ethiopia; and highly similar
387 strains shared between USA and Brazil, USA and Australia, and between Australia and Vietnam. Given
388 the variation of *Xep* across our sample, genetic similarity in core genes and gene content across continents
389 is strong evidence of international dissemination. Genetically similar strains of bacterial spot pathogens *X.*
390 *euvesicatoria* pv. *euvesicatoria* and *X. hortorum* pv. *gardneri* collected from different continents similarly
391 suggest intercontinental dissemination in tomato and pepper seed [15, 58, 89]. Whole genome analysis of
392 *X. hortorum* pv. *pelargonii* strains from a 2022 epidemic of bacterial blight of geranium in the USA
393 showed zero to seven chromosomal SNPs among isolates of the emergent strain that was distributed to
394 multiple states in plant cuttings [9, 90].

395 Other *Xep* genotypes indicated a more limited distribution. We did not find core gene cluster 2
396 strains in the seed production regions sampled (China, Thailand, Vietnam), while this lineage was highly
397 represented in our USA sample. To date, strains in this cluster have been found only in the southeastern
398 and midwestern USA [43, 47, 50, 53, 54] and Mexico. Seedling nurseries in the southeast USA produce
399 tomato transplants for growers in multiple states. Interstate movement of strains on seedlings is likely
400 responsible, at least in part, for disseminating genetically similar strains to different states [43, 53]. We
401 previously reported extensive recombination with *X. euvesicatoria* pv. *euvesicatoria* in cluster 2 strains
402 [47] and this cluster had a diverse accessory genome, perhaps suggesting that it has a larger geographic
403 distribution than represented in our sample, which is biased towards the USA.

404 The *Xep* strains we examined from the USA were assigned to core gene clusters 1, 2, 3, 4, 5, 6,
405 and 8, representing several distinct genetic lineages. To better understand the initial emergence of *Xep*, we
406 used calibrated phylogenies to examine the timing of lineage divergence. International trade in F1 hybrid
407 tomato seed surged in the second half of the 20th century, after the first hybrid tomato cultivars were
408 released by 1940 [91, 92]. There was a 300-fold increase in hybrid tomato seeds exported from Asia
409 between 1962 and 1977 [93, 94] and subsequent rapid growth in tomato production. Our analyses
410 estimate the most recent common ancestor of our sample to ~150 years ago, while the major ancestral

411 lineages diverged during or after the early expansion in the hybrid seed trade. We hypothesize that the
412 emergence and geographic distribution of lineages may be associated with the multinational structure of
413 tomato breeding and seed production, in which parental lines and geographic locations of seed production
414 change over time [95].

415 The roles of T3E in pathogenicity and virulence make them important members of both the core
416 and accessory genomes. We found up to 16 putative core effector genes, most of which exhibited allelic
417 variation. The impact of allelic variation in *Xep* effectors on pathogen fitness, if any, is unknown. Low
418 frequency effectors were found across core gene clusters, suggesting acquisition of new effectors and
419 their exchange among *Xep* lineages. For example, some strains in clusters 3, 4, and 5 from the United
420 States, Canada, and Mexico carried the same alleles of low frequency effectors XopAQ and XopE3 as
421 strains from Asia, Nigeria, and Italy. BLAST analysis suggested the geographically widespread presence
422 of transcription activation-like (TAL) effectors in *Xep*. Both TAL effectors described in *Xep*, *avrHah1*
423 and *pthXp1*, are associated with increased disease severity on tomato [16, 50]. Acquisition of T3Es could
424 increase the fitness of *Xep* relative to other bacterial spot pathogens and cause more damaging disease
425 outbreaks [49, 51, 66].

426 The release of new plant varieties that carry disease resistance genes can have dramatic effects on
427 pathogen population structure due to selection to overcome host resistance [96-98], and we have
428 previously reported on the loss of function of effector AvrXv3 (XopAF) across lineages [27, 53, 66].
429 Examination of T3E content at a global scale puts variation previously observed in Florida into a larger
430 context. XopAF was present in strains collected in the 1990s (cluster 1), but absent or non-functional in
431 most strains from Florida, Indiana Ohio, and North Carolina, USA [23, 25, 27, 53, 66]. Here, we found
432 that *xopAF* was intact in many strains from seed production areas, suggesting strong selection for loss of
433 function in commercial fruit production. The effector XopJ4 is a possible resistance target [66], but Klein-
434 Gordon et al. [23] reported that it was missing from 3.2% of Florida strains collected in 2017 and, here,
435 we found that it was absent in one North Carolina and 20 Florida, USA strains. All strains collected
436 outside the USA contained *xopJ4*, except for cluster 10 strains. Another XopJ family member, *xopJ2*, is a

437 virulence factor in tomato [49, 51]. It is common in North America, particularly in cluster 2 and 6 strains,
438 but absent or infrequently detected in *Xep* populations elsewhere. An alternative form of this effector,
439 recently described as XopJ2b [76], is more common in strains from outside North America.
440 Bacterial spot is a destructive disease in areas where tomatoes are grown under humid conditions
441 and growers in the USA have relied heavily on copper bactericides to manage this disease. In response,
442 *Xep* strains have developed copper tolerance [99]. Most strains isolated from Florida in the 1990s lacked
443 the *copLAB* genes, but they are now common in strains collected in the USA. A recent study of Florida
444 strains found that these copper resistance genes are more frequently present on the chromosome than on a
445 plasmid, suggesting selection for vertical inheritance of copper tolerance [30]. In contrast, strains from
446 other countries lacked copper resistance genes, indicating little or no local selection for the acquisition of
447 *cop* genes.
448 In summary, we found strong evidence for intercontinental movement of *Xep*, consistent with the
449 international nature of tomato breeding and hybrid tomato seed production. We also found notable
450 diversity in our global sample of *Xep*, including in seed production regions, and multiple variants of *Xep*
451 that do not appear to be widely distributed. The genomic diversity of *Xep* in seed and fruit production
452 regions creates the opportunity for recombination among strains and subsequent dissemination of high
453 fitness variants of *Xep*.
454
455
456 **Materials and Methods**
457 Bacterial strains, genome sequencing, and assembly
458 *Xep* strains were collected from 13 different countries (Table 1; S1 Table). Strains from the
459 United States were collected from seven states between 1991 to 2016 and comprised 181 strains. The
460 remaining 89 strains were collected from Canada, Mexico, Brazil, Italy, Ethiopia, Nigeria, South Africa,
461 Iran, China, Thailand, Australia, and Vietnam. Strains from China, Thailand, and Vietnam were collected
462 from fields designated for production of tomato seed for the global market. Strains from Brazil were

463 obtained from both staked fresh-market and processing tomato commercial fields. Strains from Italy were
464 isolated from tomato pith in greenhouse tomato showing wilting symptoms [64, 100]. Strains from South
465 Africa were collected from commercial seed lots. Strains from Nigeria were obtained from fields
466 cultivated for both subsistence and commercial purposes. Strains from other countries were collected
467 from fields designated for commercial fruit production.

468 A total of 270 *Xep* genome sequences were used during this study (S1 Table). Draft and whole
469 genomes of 117 strains were generated and published previously [43, 47, 48, 57, 58, 67, 100]. The
470 remaining 153 strains were sequenced for this study using Illumina platforms. Genomic DNA was
471 extracted from single colony cultures grown for 24-hr in nutrient broth using the Wizard Genomic DNA
472 Purification Kit (Promega, Chicago, IL) following manufacturer instructions. Genomic libraries for
473 sequencing were prepared using the Nextera DNA library preparation kit from Illumina (Illumina, San
474 Diego, CA). Sequencing was performed at the Interdisciplinary Center for Biotechnology Research,
475 University of Florida, using an Illumina MiSeq to generate 250 bp paired end reads for each strain.
476 Additional genomic sequence data were generated for five strains for the ANI analysis (S1 Table, part B).
477 Genomic DNA was extracted using the above methods except that extracted genomic DNA was sent to
478 SeqCenter (Pittsburg, PA) for sequencing with Illumina NovaSeq 6000, producing 150 bp paired end
479 reads.

480 Raw reads were trimmed of adapters and paired with Trim Galore
481 (https://github.com/FelixKrueger/TrimGalore)[101], then assembled into contigs with Spades version
482 3.10.1 [102], with k-mers 21, 33, 55, 77, 99, and 127 with read error correction and "--careful" switch.
483 Reads were then aligned to the assembled contigs using Bowtie 2 v. 2.3.3 [103]. Inconsistencies were
484 identified and polished using Pilon [104]. Contigs smaller than 500 bp and with less than 2.0 k-mer
485 coverage were filtered out. Quality of genomes were assessed with CheckM [105]. Assembled genomes
486 were annotated using the IMG/JGI platform [106]. The genome data generated for this study are available
487 in NCBI BioProject PRJNA941448.
488

Core gene phylogeny

In a previous study, we defined a set of 1,356 'core genes' from 58 genomes of *Xep* strains isolated from Florida [47]. The core genes were determined based on amino acid sequence homology using GET_HOMOLOGUES software package [107]. We used the core genes from a representative *Xep* genome, Xp91-118, as query to search the remaining 269 genomes using local BLAST [108]. BLAST results were filtered using query coverage and pairwise nucleotide sequence alignment thresholds of 70% each and the sequence was checked for the presence of standard start and stop codons at either end of the gene and gene was removed if both were not present. A total of 887 genes were found to be intact in all 270 genomes. Genes were individually parsed and aligned using MAFFT [109] and concatenated using sequence matrix [110]. The result was a 617.854 Kbp alignment, hereafter referred to as core genes.

The concatenated core gene sequence was used to construct a maximum likelihood (ML) phylogenetic tree using RAxML v.8.2.12 [111]. General time reversible model with gamma distributed rates and invariant sites (GTRGAMMA) was used as the nucleotide substitution model. To account for recombination, the ML tree output from RAxML and concatenated core genome alignment were used as was input for ClonalFrameML v1.12 [70].

Population structure

SNPs were extracted from core genes for hierarchical clustering based on Bayesian analysis of population structure (hierBAPS) algorithm [112], implemented in the 'rhierBAPS' R package v 1.0.1 [71, 113]. For visualization, hierBAPS clusters were added to the phylogenetic tree generated from ClonalFrameML using the 'ggtree' package in R [114]. The treemap function in plotly [115] was used to show the relative distribution of clusters across geographic locations. R package 'ggplot2' was used to map hierBAPS clusters to countries [116]. The 'PopGenome' R package [117] was used to calculate FST, nucleotide diversity (pi), and Tajima's D statistic by geographic location and by hierBAPS cluster.

Assembled genomes were used for calculating average nucleotide identity and pangenome analysis. Average nucleotide identity (ANIb) between strains was calculated using assembled genomes

515 with Pyani version 0.2.10 [118]. The pangenome was estimated using Roary v3.12.0 [119] after
516 annotation from Prokka v1.12 [120]. The gene presence absence matrix from Roary (S4 Data) was used
517 as input for generation of NMDS plots using the 'dplyr' and 'ggplot2' packages from tidyverse [116] and
518 to generate gene accumulation curves for each cluster using package 'micropan' [121].
519
520 Bayesian analysis of *X. euvesicatoria* pv. *perforans* divergence times
521 A whole genome alignment was generated using split k-mer analysis version 2 (SKA2) [122] for
522 all 270 *Xep* strains plus outgroup *X. euvesicatoria* pv. *euvesicatoria* strain 85-10 (NCBI Accession
523 GCA_000009165.1; S2 Data). The alignment was reduced to variable sites only using Geneious 2023.2.1
524 (BioMatters Ltd.) A phylogenetic network was calculated from the resulting SNPs using the NeighborNet
525 2004 algorithm in SplitsTree5 [123, 124]. Phylogenetic conflict was indicated between the 259 strains,
526 cluster 10 strains, and outgroup (S1 Figure, part B). Removing the cluster 10 strains did not remove the
527 conflict between *Xep* and *Xee* outgroup. As a result, we limited our dating of the phylogeny of *Xep* to the
528 259 strains in BAPS clusters 1 through 9. We used Gubbins v. 2.4.1 [72] to remove putative recombinant
529 sites from whole genome alignments generated using SKA2 [122] and the complete genome of Xp91-118
530 as a reference (GCF_000192045.2). The resulting alignment was used to infer a phylogenetic tree using
531 the GTRGAMMI model in RAxML version 8.2.10 [125]. The temporal analysis was conducted with
532 BactDating v1.1.1 [73]. The inputs to the BactDating analysis were the maximum likelihood tree and
533 dates of isolation assigned as dates of tips. The rooting of the tree was estimated using the initRoot
534 function, which maximizes the correlation between tip date, the year the strain was collected, and root-to-
535 tip branch lengths. Dates of nodes were inferred using the bactdate function on the re-rooted tree using a
536 relaxed molecular clock with Markov chain Monte Carlo (MCMC) chains of 10^6 iterations. Phylostems
537 [126] was used to assess the temporal signals within internal clades for interpretation of node date
538 inferences.
539 We also used BEAST v. 1.10.4 (Suchard et al. 2018) to infer a dated phylogeny. The XML file
540 was manually edited to include the 'ascertained' flag in the alignment block (S3 Data). The HKY

nucleotide substitution model with empirical base frequencies and gamma distribution of site-specific rate heterogeneity was used with coalescent Bayesian skyline priors with an uncorrelated relaxed clock for Bayesian phylogenetic inference over MCMC chains of 200 million generations. Adequate mixing was assessed based on a minimum effective sample size of 200 for parameter estimates as calculated by Tracer v. 1.10.4. A maximum clade credibility tree was inferred from the posterior distribution of trees using TreeAnnotator v. 1.10.4, specifying a burn-in of 10% and the 'keep' option for node heights. Trees were visualized in iTOL version 6.9.1 [127].

Type III Effector Analysis

A T3E effector database was curated using amino acid sequences of 66 *Xanthomonas* effectors [128] (S6 Table). When available, functional annotations were retrieved from NCBI and Pfam databases [129]. Orthologous sequences were identified with the software BLASTp [130, 131], by querying the curated effectors database against the amino acid sequences of the annotated genomes of 270 *Xep* strains. Sequences (BLAST hits) were considered effector orthologs when at or above a threshold of 70 percent identity and 50 percent query coverage. When multiple sequences from the same strain had hits above the thresholds to a particular effector, a weighted calculation of the identity and coverage was used to select the best hit. Sequences with homology to multiple effectors and sequences with evidence of contig breaks were manually removed. Assignment of sequences as effector orthologues was confirmed by performing a clustering analysis of all sequences using the software USEARCH v. 11.0.667 and the algorithm HPC-CLUST [132]. For the duplicated effector XopP, we used a phylogenetic analysis of all sequences to distinguish likely orthologous alleles from the more genetically distant paralogous sequences, which were assigned to XopP2.

Orthologous sequences from each effector were extracted from the annotated genomes, aligned with MAFFT [109], and allelic variants identified [133] to generate a numeric matrix representing presence and allelic variant or absence. Hierarchical clustering analysis of effectors was performed by calculating a distance matrix with function 'dist' with the method 'manhattan', and the function 'hclust'

with the method 'complete' from the R package 'vegan' [113, 134]. The results were displayed as a heatmap with the package 'gplots' and the function heatmap.2 [135].

To investigate the presence of positive selection acting on the effector sequences, we used the software HyPhy (Hypothesis Testing using Phylogenies) implementing the methods FUBAR (Fast, Unconstrained Bayesian AprRoximation) and MEME (Mixed Effects Model of Evolution) [77, 78]. The Bayesian method FUBAR evaluates pervasive selection, assuming the same rates of synonymous and nonsynonymous substitution per site on all branches. The method MEME uses a maximum likelihood approach to evaluate episodic selection, i.e., selection only a subset of branches of the phylogeny. For each effector gene, a codon-aware alignment was generated with the software PRANK using the codon flag '-c' as settings [136]. RAxML [111] was used to infer a phylogenetic tree with the GTRGAMMA (gamma time-reversible) model of nucleotide substitution. The codon-aware alignment and phylogenetic tree were used as the input files for FUBAR and MEME.

To determine the relationship of the effector profiles with respect to core gene cluster, geographic and temporal distribution, we transformed the dissimilarities in the matrix of effector profiles into distances with non-metric multidimensional scaling (NMDS). We used the Bray-Curtis dissimilarity index, a robust index able to handle missing data that considers the presence and absence of effectors as equally informative, calculated with the package 'vegan' and the function 'metaMDS' [134]. We used a low number of dimensions (K=2) and set try=30 and trymax=500 for random starts to avoid the NMDS getting trapped in local optima. NMDS plots were created with the packages 'ggrepel' and 'ggplot2' [137, 138]. Based on the NMDS analysis, we assigned strains to effector clusters, which were plotted on a worldwide map with the packages 'ggplot2' and 'scatterpie' [138, 139]. The map was created in R with the packages 'cowplot', 'ggrepel', 'ggspatial', 'libwgeom', 'sf', 'rgeos', 'memisc', 'oz', 'maptools' and 'rnaturalearth' with the function 'ne_countries' [137, 140-147]. Geographic coordinates (longitude, latitude) of countries and states (for USA) of collection were obtained with the R package 'googleway' [148] and the function 'mutate_geocode' from Google maps.

592 To sequence the putative TAL effector from 2P6S1, native plasmid DNA was isolated using the
593 alkaline lysis method [149]. *Eco*RI digested DNA of the plasmid prep was ligated into vector pLAFR3
594 [150] restricted with the same enzyme for transformation into *E. coli* DH5α. Clones containing the TAL
595 effector were identified by PCR and analyzed by restriction digest. One clone, designated as p7.1,
596 contained an approx. 5 Kbp *Eco*RI fragment and was selected for Sanger sequencing and phenotype
597 testing. For Sanger sequencing of the TAL repeat region, DNA of p7.1 was restricted with *Nsi*I and the
598 internal fragment was ligated into vector pBluescript restricted with *Pst*I. Additional pBluescript
599 subclones were made using *Bam*HI (~3 Kbp and ~1.1 Kbp) and *Bam*HI/*Eco*RI (~1 Kbp) in order to cover
600 the entire cloned region in p7.1. All clones were transformed into DH5α for sequencing using vector
601 primers T3 and T7.
602 Copper resistance genes in assembled genomes were identified with BLASTn analysis using *copL*
603 (MBZ2440241.1), *copA* (MBZ2440240.1), and *copB* (MBZ2440239.1)
604 from *Xep* strain Xp2010 as reference sequences [30].
605

606 **Acknowledgements**

607 We thank D. Ritchie and C. Mauney for providing reference strains and collecting disease samples,
608 respectively. Tina Simonton and Phyllis May provided assistance with strain collection in Ontario. This
609 research was supported by USDA NIFA awards 2015-51181-24312, 2020-67013-31921, and 2022-
610 51181-38242, Ontario Agri-Food Innovation Alliance, Ontario Tomato Research Institute, and the Fresh
611 Vegetable Growers of Ontario, the North Carolina Tomato Growers Association. Plant diagnostic
612 laboratories receive funding from USDA-NIFA through the National Plant Diagnostic Network. FIB was
613 supported by Conacyt Mexico. MK was supported by Swedish International Development Cooperation
614 Agency.
615

616 **Author Contributions**

ST, FIB, MOJ, PDR, GEV, JBJ, and EMG conceived the study. MOJ, GVM, PA, TBA, GC, LBTdlB, EB, TCC, DTKT, TAC, DSE, RFG, DMF, MK, MLI, FJL, LL, ETM, SAM, NTTN, EO, AMQD, RR, FR, GER, VMS, PT, CT, JBJ, and GEV collected bacterial strains. ST, MOJ, GVM, and JKG prepared genomic libraries. ST, FIB, MOJ, JHT, A Sharma, A Subedi, AK, and EMG performed data analysis. ST, FIB, MOJ, A Sharma, A Subedi, GEV, JBJ, and EMG drafted the manuscript. All authors reviewed the manuscript.

Data summary

The raw read files and genome assemblies for all strains sequenced in this study are deposited in NCBI under BioProject PRJNA941448. The sources of genome assemblies acquired from public databases are listed in Table 1.

630 **Supporting Information**

631

632 **S1 Figure. Phylogenetic analysis of 270 *X. euvesicatoria* pv. *perforans* strains.** (A) Maximum likelihood
633 phylogenetic tree of *Xanthomonas euvesicatoria* pv. *perforans* strains based on aligned nucleotide sequences of
634 887 core genes. The tree was inferred using RAxML using a GTRGAMMAI substitution model. The tree was
635 rooted using the 11 genetically diverged strains that make up core gene cluster 10. (B) NeighborNet network
636 inferred using SNPs from aligned whole genome sequences, including *X. euvesicatoria* pv. *euvesicatoria* strain
637 85-10 (bolded) as an outgroup. Core gene cluster 10 strains are highlighted. Reticulations in the network indicate
638 conflicting phylogenetic relationships.

639

640 **S2 Figure. Accessory genome variation in *Xanthomonas euvesicatoria* pv. *perforans*.** (A) Visualization of
641 pangenome variation by non-metric multidimensional scaling of gene presence-absence for all 270 *X. perforans*
642 strains by BAPS cluster. Ellipses assume a multivariate t-distribution. (C) Increase in gene count with increasing
643 number of strains sampled. Clusters 1 and 2 were represented by the most strains, but other clusters showed
644 similar rates of increase in the pangenome of the cluster. Pangenome matrix used for analysis is available as S4
645 Data.

646

647 **S3 Figure. Temporal signal in phylogenetic tree of 259 *X. euvesicatoria* pv. *perforans* strains.** (A)
648 Correlation between sampling year and root-to-tip distance in maximum likelihood phylogenetic tree
649 inferred from alignment of whole genome sequences. Output was generated from BactDating R package.
650 (B) Temporal signal within the phylogenetic determined using Phylostems tool. Nodes with significant
651 temporal signals are indicated with colored circles. Adjusted R-squared values by color are: dark green 0–
652 0.2; light green 0.2–0.4; yellow 0.4–0.6; orange 0.6–0.8; red 0.8–1.

653

654 **S4 Figure. Dated phylogenies of 259 *X. euvesicatoria* pv. *perforans* strains.** (A) Dating using
655 BactDating relaxed clock analysis on RAxML-generated phylogeny. This is the tree shown in Figure 2,

shown here without collapsed nodes. (B) Dating of same dataset using BEAST with coalescent Bayesian skyline priors and an uncorrelated relaxed clock.

S5 Figure. Frequency of Xop effectors among 270 *Xanthomonas euvesicatoria* pv. *perforans* strains. The most common allele observed was assigned to allele type 1, second most frequent allele to allele type 2, and so on. Note that alleles classified as pseudogenes included contig breaks, which include assembly errors. For example, all strains appear to have *xopD*, but a repeat caused a contig break in the gene in nearly half of the genomes.

S6 Figure. Clustering of 270 *Xanthomonas perforans* effector profiles by non-metric multidimensional scaling and distribution of resulting clusters among geographic regions. Analysis did not include TAL effectors. (A) The most frequently observed group of effector profiles form cluster A. This cluster of 188 strains is represented as a star in plots B-C, as it is represented in most BAPS core gene clusters (B), most of the sampled tomato production regions (C), and in collections from 1991 to 2017 (D). Clusters were largely defined by low frequency effectors (S7 Figure). (E) Distribution of strains by effector clusters among sampled countries.

S7 Figure. Variation in type III effector profiles in 270 *Xanthomonas euvesicatoria* pv. *perforans* strains ordered according to NMDS of effector profiles. Analysis did not include TAL effectors. Type III effectors are in columns and *Xep* strains in rows. Effector status is shown by allele type: absence is indicated by allele type 0 (white), while the most frequent allele observed when the effector is present is allele type 1 (purple), second most frequent is allele type 2 (blue), and so on. Putative pseudogenized effectors are shown as allele 13 (gray). The order of columns was determined by hierarchical clustering analysis, placing similarly distributed effectors adjacent to each other. Order of rows is based on NMDS clustering analysis of effector profiles (see S6 Figure).

682 **S1 Table. Genome data and metadata for *X. euvesicatoria* pv. *perforans* strains.** (A) BAPS cluster
683 assignment for each strain and NCBI information for each genome. Genome assembly statistics are given
684 for newly sequenced strains. (B) Depth of coverage relevant to ANI comparisons in Fig. 1D.
685
686 **S2 Table. Genetic diversity statistics by geographic region and BAPS group.** (A) Statistics by country
687 and U.S. state. (B) Statistics by BAPS group.
688
689 **S3 Table. Average nucleotide identity (ANIb) comparisons between strains with highly similar core**
690 **gene sequences collected across continents.** (A) Proportion nucleotide identity. (B) Alignment fraction.
691
692 **S4 Table. Putative type III effectors (Xop proteins) found in 270 *X. euvesicatoria* pv. *perforans***
693 **assembled genomes.** (A) Summary for each locus. (B) Results by strain. Each different amino acid
694 sequence per gene was assigned a numerical allele type, such that the most common allele observed was
695 assigned to allele type 1. Potential pseudogenes are indicated with "pseudo" and absence indicated with
696 zero. Locus tags refer to JGI IMG annotations (https://img.jgi.doe.gov). Reference sequences used for
697 BLAST searches are given in S5 Table. The final column shows the result of BLAST searches for TAL
698 effectors.
699
700 **S5 Table. Presence or absence of copper genes (*copLAB*) in 270 *X. euvesicatoria* pv. *perforans***
701 **assembled genomes**. Symbols represent gene presence '+' or absence '-'. Contig break in gene is
702 indicated by (+).
703
704 **S6 Table. Type III effector database used to query assembled genomes for effector genes.**
705
706 **S1 Data. Nucleotide alignment of 887 core genes from 270 *X. euvesicatoria* pv. *perforans* strains.**
707 Alignment is 617,855 bp in FASTA format.

S2 Data. Nucleotide alignment of variable sites from whole genome alignment of 270 *X. euvesicatoria* pv. *perforans* strains and *X. euvesicatoria* pv. *euvesicatoria* strains 85-10.

S3 Data. XML file used for BEAST analysis.

S4 Data. Pangenome matrix for 270 *X. euvesicatoria* pv. *perforans* strains

716 **References Cited**

717

718 1. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of
719 pathogens and pests on major food crops. Nature Ecology & Evolution. 2019;3(3):430-9. doi:
720 10.1038/s41559-018-0793-y.
721 2. Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, et al. The
722 persistent threat of emerging plant disease pandemics to global food security. Proceedings of the National
723 Academy of Sciences. 2021;118(23):e2022239118. doi: doi:10.1073/pnas.2022239118.
724 3. Rizzo DM, Lichtveld M, Mazet JAK, Togami E, Miller SA. Plant health and its effects on food
725 safety and security in a One Health framework: four case studies. One Health Outlook. 2021;3(1):6. doi:
726 10.1186/s42522-021-00038-7.
727 4. Martins PMM, Merfa MV, Takita MA, De Souza AA. Persistence in Phytopathogenic Bacteria:
728 Do We Know Enough? Frontiers in Microbiology. 2018;9. doi: 10.3389/fmicb.2018.01099.
729 5. Jeger MJ, Fielder H, Beale T, Szyniszewska AM, Parnell S, Cunniffe NJ. What Can Be Learned
730 by a Synoptic Review of Plant Disease Epidemics and Outbreaks Published in 2021? Phytopathology®.
731 2023;113(7):1141-58. doi: 10.1094/phyto-02-23-0069-ia. PubMed PMID: 36935375.
732 6. Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, et al. Top 10 plant
733 pathogenic bacteria in molecular plant pathology. Mol Plant Pathol. 2012;13(6):614-29. doi:
734 https://doi.org/10.1111/j.1364-3703.2012.00804.x.
735 7. Pfeilmeier S, Caly DL, Malone JG. Bacterial pathogenesis of plants: future challenges from a
736 microbial perspective. Mol Plant Pathol. 2016;17(8):1298-313. doi: https://doi.org/10.1111/mpp.12427.
737 8. Morris CE, Sands DC, Vinatzer BA, Glaux C, Guilbaud C, Buffiere A, et al. The life history of
738 the plant pathogen *Pseudomonas syringae* is linked to the water cycle. The ISME journal. 2008;2(3):321-
739 34.

9. Roman-Reyna V, Sharma A, Toth H, Konkel Z, Omiotek N, Murthy S, et al. Live tracking of a plant pathogen outbreak reveals rapid and successive, multidecade plasmid reduction. mSystems. 2024;9. doi: https://doi.org/10.1128/msystems.00795-23.

10. Campos PE, Groot Crego C, Boyer K, Gaudeul M, Baider C, Richard D, et al. First historical genome of a crop bacterial pathogen from herbarium specimen: Insights into citrus canker emergence. PLoS Path. 2021;17(7):e1009714. doi: 10.1371/journal.ppat.1009714.

11. Xu J, Zhang Y, Li J, Teper D, Sun X, Jones D, et al. Phylogenomic analysis of 343 *Xanthomonas citri* pv. *citri* strains unravels introduction history and dispersal paths. PLoS Path. 2023;19(12):e1011876. doi: 10.1371/journal.ppat.1011876.

12. Weisberg AJ, Davis EW, Tabima J, Belcher MS, Miller M, Kuo C-H, et al. Unexpected conservation and global transmission of agrobacterial virulence plasmids. Science. 2020;368(6495):eaba5256. doi: 10.1126/science.aba5256.

13. Castillo AI, Chacón-Díaz C, Rodríguez-Murillo N, Coletta-Filho HD, Almeida RPP. Impacts of local population history and ecology on the evolution of a globally dispersed pathogen. BMC Genomics. 2020;21(1):369. doi: 10.1186/s12864-020-06778-6.

14. Castillo Andreina I, Bojanini I, Chen H, Kandel Prem P, De La Fuente L, Almeida Rodrigo PP. Allopatric Plant Pathogen Population Divergence following Disease Emergence. Appl Environ Microbiol. 2021;87(7):e02095-20. doi: 10.1128/AEM.02095-20.

15. Jibrin MO, Sharma A, Mavian CN, Timilsina S, Kaur A, Iruegas-Bocardo F, et al. Phylodynamic Insights into Global Emergence and Diversification of the Tomato Pathogen *Xanthomonas hortorum* pv. *gardneri*. Mol Plant-Microbe Interact. 2024:MPMI-04-24-0035-R. doi: 10.1094/MPMI-04-24-0035-R.

16. Subedi A, Barrera LBTdl, Lewis Ivey M, Egel D, Kebede M, Kara S, et al. Population genomics reveals an emerging lineage of *Xanthomonas perforans* on pepper. Phytopathology. 2023. doi: 10.1094/PHYTO-04-23-0128-R.

764 17. Dupas E, Durand K, Rieux A, Briand M, Pruvost O, Cunty A, et al. Suspicions of two bridgehead
765 invasions of Xylella fastidiosa subsp. multiplex in France. Communications Biology. 2023;6(1):103. doi:
766 10.1038/s42003-023-04499-6.
767 18. Potnis N, Timilsina S, Strayer A, Shantharaj D, Barak JD, Paret ML, et al. Bacterial spot of
768 tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a
769 worldwide challenge. Mol Plant Pathol. 2015;16(9):907-20. doi: 10.1111/mpp.12244.
770 19. Osdaghi E, Jones JB, Sharma A, Goss EM, Abrahamian P, Newberry EA, et al. A centenary for
771 bacterial spot of tomato and pepper. Mol Plant Pathol. 2021;22(12):1500.
772 20. Potnis N. Harnessing eco-evolutionary dynamics of Xanthomonads on tomato and pepper to
773 tackle new problems of an old disease. Annu Rev Phytopathol. 2021;59:289-310.
774 21. Jibrin MO, Timilsina S, Minsavage GV, Vallad GE, Roberts PD, Goss EM, et al. Bacterial spot
775 of tomato and pepper in Africa: Diversity, emergence of T5 race, and management. Frontiers in
776 Microbiology. 2022;13:835647. doi: 10.3389/fmicb.2022.835647.
777 22. Abbasi PA, Khabbaz SE, Weselowski B, Zhang L. Occurrence of copper-resistant strains and a
778 shift in *Xanthomonas* spp. causing tomato bacterial spot in Ontario. Canadian Journal of Microbiology.
779 2015;61(10):753-61. doi: 10.1139/cjm-2015-0228 %M 26308592.
780 23. Klein-Gordon JM, Xing Y, Garrett KA, Abrahamian P, Paret ML, Minsavage GV, et al.
781 Assessing changes and associations in the *Xanthomonas perforans* population across Florida commercial
782 tomato fields via a statewide survey. Phytopathology. 2021;111(6):1029-41. doi: 10.1094/PHYTO-09-20-
783 0402-R.
784 24. Lewis Ivey ML, Strayer A, Sidhu JK, Minsavage GV. Bacterial leaf spot of tomato (*Solanum*
785 *lycopersicum*) in Louisiana is caused by *Xanthomonas perforans*, tomato race 4. Plant Dis.
786 2016;100(6):1233-. doi: 10.1094/PDIS-12-15-1451-PDN.
787 25. Egel DS, Jones JB, Minsavage GV, Creswell T, Ruhl G, Maynard E, et al. Distribution and
788 characterization of *Xanthomonas* strains causing bacterial spot of tomato in Indiana. Plant Health
789 Progress. 2018;19(4):319-21. doi: 10.1094/php-07-18-0041-br.

790 26. Abrahamian P, Jones JB, Vallad GE. Efficacy of copper and copper alternatives for management
791 of bacterial spot on tomato under transplant and field production. Crop Protect. 2019;126:104919. doi:
792 https://doi.org/10.1016/j.cropro.2019.104919.
793 27. Adhikari P, Adhikari TB, Timilsina S, Meadows I, Jones JB, Panthee DR, et al. Phenotypic and
794 genetic diversity of *Xanthomonas perforans* populations from tomato in North Carolina. Phytopathology.
795 2019;109(9):1533-43. doi: 10.1094/phyto-01-19-0019-r. PubMed PMID: 31038016.
796 28. Rotondo F, Bernal E, Ma X, Ivey MLL, Sahin F, Francis DM, et al. Shifts in *Xanthomonas* spp.
797 causing bacterial spot in processing tomato in the Midwest of the United States. Canadian Journal of Plant
798 Pathology. 2022:1-16. doi: 10.1080/07060661.2022.2047788.
799 29. Khanal S, Hind SR, Babadoost M. Occurrence of copper-resistant *Xanthomonas perforans* and *X.*
800 *gardneri* in Illinois tomato fields. Plant Health Progress. 2020;21(4):338-44. doi: 10.1094/PHP-06-20-
801 0048-RS.
802 30. Bibi S, Weis K, Kaur A, Bhandari R, Goss E, Jones JB, et al. A Brief Evaluation of a Copper
803 Resistance Mobile Genetic Island in the Bacterial Leaf Spot Pathogen *Xanthomonas euvesicatoria* pv.
804 *perforans*. Phytopathology. 2023;113(8):1394-8. doi: 10.1094/PHYTO-02-23-0077-SC.
805 31. Constantin EC, Cleenwerck I, Maes M, Baeyen S, Van Malderghem C, De Vos P, et al. Genetic
806 characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic
807 revision of the *X. axonopodis* species complex. Plant Pathol. 2016;65(5):792-806. doi:
808 10.1111/ppa.12461.
809 32. Jones JB, Stall RE, Scott JW, Somodi GC, Bouzar H, Hodge NC. A third tomato race of
810 *Xanthomonas campestris* pv. *vesicatoria*. Plant Dis. 1995;79(4):395-8.
811 33. Jones JB, Lacy GH, Bouzar H, Stall RE, Schaad NW. Reclassification of the xanthomonads
812 associated with bacterial spot disease of tomato and pepper. Syst Appl Microbiol. 2004;27(6):755-62. doi:
813 http://dx.doi.org/10.1078/0723202042369884.

814 34. Roach R, Mann R, Gambley CG, Shivas RG, Rodoni B. Identification of *Xanthomonas* species
815 associated with bacterial leaf spot of tomato, capsicum and chilli crops in eastern Australia. Eur J Plant
816 Pathol. 2018;150(3):595-608. doi: 10.1007/s10658-017-1303-9.
817 35. Burlakoti RR, Hsu C-F, Chen J-R, Wang J-F. Population dynamics of xanthomonads associated
818 with bacterial spot of tomato and pepper during 27 years across Taiwan. Plant Dis. 2018;102(7):1348-56.
819 doi: 10.1094/PDIS-04-17-0465-RE.
820 36. Kebede M, Timilsina S, Ayalew A, Admassu B, Potnis N, Minsavage GV, et al. Molecular
821 characterization of *Xanthomonas* strains responsible for bacterial spot of tomato in Ethiopia. Eur J Plant
822 Pathol. 2014;140(4):677-88. doi: 10.1007/s10658-014-0497-3. PubMed PMID: WOS:000344533700006.
823 37. Timilsina S, Jibrin MO, Potnis N, Minsavage GV, Kebede M, Schwartz A, et al. Multilocus
824 sequence analysis of xanthomonads causing bacterial spot of tomato and pepper reveals strains generated
825 by recombination among species and recent global spread of *Xanthomonas gardneri*. Appl Environ
826 Microbiol. 2015;81:1520-9.
827 38. Hamza AA, Robène-Soustrade I, Jouen E, Gagnevin L, Lefeuvre P, Chiroleu F, et al. Genetic and
828 pathological diversity among *Xanthomonas* strains responsible for bacterial spot on tomato and pepper in
829 the Southwest Indian Ocean region. Plant Dis. 2010;94(8):993-9. doi: 10.1094/PDIS-94-8-0993. PubMed
830 PMID: 30743480.
831 39. Araújo ER, Costa JR, Ferreira MASV, Quezado-Duval AM. Widespread distribution of
832 *Xanthomonas perforans* and limited presence of *X. gardneri* in Brazil. Plant Pathol. 2017;66(1):159-68.
833 doi: https://doi.org/10.1111/ppa.12543.
834 40. Osdaghi E, Taghavi SM, Hamzehzarghani H, Lamichhane JR. Occurrence and characterization of
835 the bacterial spot pathogen *Xanthomonas euvesicatoria* on pepper in Iran. J Phytopathol.
836 2016;164(10):722-34. doi: https://doi.org/10.1111/jph.12493.
837 41. Osdaghi E, Taghavi SM, Hamzehzarghani H, Fazliarab A, Lamichhane JR. Monitoring the
838 occurrence of tomato bacterial spot and range of the causal agent *Xanthomonas perforans* in Iran. Plant
839 Pathol. 2017;66(6):990-1002. doi: https://doi.org/10.1111/ppa.12642.

840 42. Osdaghi E, Taghavi SM, Koebnik R, Lamichhane JR. Multilocus sequence analysis reveals a
841 novel phylogroup of *Xanthomonas euvesicatoria* pv. *perforans* causing bacterial spot of tomato in Iran.
842 Plant Pathol. 2018;67(7):1601-11. doi: https://doi.org/10.1111/ppa.12864.
843 43. Abrahamian P, Timilsina S, Minsavage GV, Potnis N, Jones JB, Goss EM, et al. Molecular
844 epidemiology of *Xanthomonas perforans* outbreaks in tomato plants from transplant to field as
845 determined by single-nucleotide polymorphism analysis. Appl Environ Microbiol. 2019;85(18):e01220-
846 19. doi: 10.1128/aem.01220-19.
847 44. Modor Intelligence. Tomato Seeds Market - Growth, Trends, COVID-10 Impact, and Forecasts
848 (2022 - 2027) 2022. Available from: Available at: https://www.mordorintelligence.com/industry-
849 reports/tomato-seeds-market [Accessed August 31, 2022].
850 45. Abrahamian P, Sharma A, Jones JB, Vallad GE. Dynamics and spread of bacterial spot epidemics
851 in tomato transplants grown for field production. Plant Dis. 2021;105(3):566-75.
852 46. Hert A, Roberts P, Momol M, Minsavage G, Tudor-Nelson S, Jones J. Relative importance of
853 bacteriocin-like genes in antagonism of *Xanthomonas perforans* tomato race 3 to *Xanthomonas*
854 *euvesicatoria* tomato race 1 strains. Appl Environ Microbiol. 2005;71(7):3581-8.
855 47. Timilsina S, Pereira-Martin JA, Minsavage G, Iruegas Bocardo F, Abrahamian P, Potnis N, et al.
856 Multiple recombination events drive the current genetic structure of *Xanthomonas perforans* in Florida.
857 Frontiers in Microbiology. 2019;10:448.
858 48. Jibrin MO, Potnis N, Timilsina S, Minsavage GV, Vallad GE, Roberts PD, et al. Genomic
859 inference of recombination-mediated evolution in *Xanthomonas euvesicatoria* and *X. perforans*. Appl
860 Environ Microbiol. 2018;84(13):e00136-18. Epub 2018/06/18. doi: 10.1128/aem.00136-18. PubMed
861 PMID: 29678917; PubMed Central PMCID: PMCPMC6007113.
862 49. Abrahamian P, Timilsina S, Minsavage GV, Kc S, Goss EM, Jones JB, et al. The type III effector
863 AvrBsT enhances *Xanthomonas perforans* fitness in field-grown tomato. Phytopathology.
864 2018;108(12):1355-62. doi: 10.1094/PHYTO-02-18-0052-R.

865 50. Newberry EA, Bhandari R, Minsavage GV, Timilsina S, Jibrin MO, Kemble J, et al. Independent
866 evolution with the gene flux originating from multiple *Xanthomonas* species explains genomic
867 heterogeneity in *Xanthomonas perforans*. Appl Environ Microbiol. 2019;85(20):e00885-19. Epub
868 2019/10/01. doi: 10.1128/aem.00885-19. PubMed PMID: 31375496; PubMed Central PMCID:
869 PMCPMC6805091.
870 51. Sharma A, Timilsina S, Abrahamian P, Minsavage GV, Colee J, Ojiambo PS, et al. Need for
871 speed: Bacterial effector XopJ2 is associated with increased dispersal velocity of *Xanthomonas perforans*.
872 Environ Microbiol. 2021;23(10):5850–65.
873 52. Klein-Gordon J, Cagmat JG, Minsavage GV, Meke LE, Vallad GE, Goss EM, et al. Strength in
874 numbers: Density-dependent volatile-induced antimicrobial activity by *Xanthomonas perforans*.
875 Phytopathology. 2022;113(2):160-9.
876 53. Bernal E, Rotondo F, Roman-Reyna V, Klass T, Timilsina S, Minsavage GV, et al. Migration
877 drives the replacement of *Xanthomonas perforans* races in the absence of widely deployed resistance.
878 Frontiers in Microbiology. 2022;13:826386. doi: 10.3389/fmicb.2022.826386.
879 54. Klein-Gordon JM, Timilsina S, Xing YR, Abrahamian P, Garrett KA, Jones JB, et al. Whole
880 genome sequences reveal the *Xanthomonas perforans* population is shaped by the tomato production
881 system. The ISME Journal. 2022;16(2):591-601. doi: 10.1038/s41396-021-01104-8. PubMed PMID:
882 WOS:000692961600002.
883 55. Richard D, Boyer C, Lefeuvre P, Canteros BI, Beni-Madhu S, Portier P, et al. Complete genome
884 sequences of six copper-resistant *Xanthomonas* strains causing bacterial spot of Solaneous plants,
885 belonging to *X. gardneri*, *X. euvesicatoria*, and *X. vesicatoria*, using long-read technology. Genome
886 Announcements. 2017;5:e01693-16. Epub 2017/02/23. doi: 10.1128/genomeA.01693-16. PubMed PMID:
887 28232425; PubMed Central PMCID: PMCPMC5323636.
888 56. Timilsina S, Potnis N, Newberry EA, Liyanapathiranage P, Iruegas-Bocardo F, White FF, et al.
889 *Xanthomonas* diversity, virulence and plant–pathogen interactions. Nature Reviews Microbiology.
890 2020;18(8):1-13. Epub 2020/04/30. doi: 10.1038/s41579-020-0361-8. PubMed PMID: 32346148.

891 57. Roach R, Mann R, Gambley CG, Chapman T, Shivas RG, Rodoni B. Genomic sequence analysis
892 reveals diversity of Australian *Xanthomonas* species associated with bacterial leaf spot of tomato,
893 capsicum and chilli. BMC Genomics. 2019;20(1):310. doi: 10.1186/s12864-019-5600-x.
894 58. Schwartz AR, Potnis N, Timilsina S, Wilson M, Patane J, Martins J, et al. Phylogenomics of
895 *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and
896 identifies determinants of host specificity. Frontiers in Microbiology. 2015;6. doi:
897 10.3389/fmicb.2015.00535.
898 59. Horvath DM, Stall RE, Jones JB, Pauly MH, Vallad GE, Dahlbeck D, et al. Transgenic resistance
899 confers effective field level control of bacterial spot disease in tomato. PLoS ONE. 2012;7(8):e42036.
900 60. Subedi A, Minsavage J, Jones JB, Goss E, Roberts P. Exploring diversity of bacterial spot
901 associated *Xanthomonas* population of pepper in Southwest Florida. Plant Dis. 2023. doi: 10.1094/PDIS-
902 10-22-2484-RE.
903 61. Cesbron S, Briand M, Essakhi S, Gironde S, Boureau T, Manceau C, et al. Comparative genomics
904 of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary
905 events linked to pathoadaptation. Frontiers in Plant Science. 2015;6:1126.
906 62. McCann HC. Skirmish or war: the emergence of agricultural plant pathogens. Curr Opin Plant
907 Biol. 2020;56:147-52. doi: https://doi.org/10.1016/j.pbi.2020.06.003.
908 63. Strayer AL, Jeyaprakash A, Minsavage GV, Timilsina S, Vallad GE, Jones JB, et al. A multiplex
909 real-time pcr assay differentiates four *Xanthomonas* species associated with bacterial spot of tomato. Plant
910 Dis. 2016;100(8):1660-8. doi: 10.1094/PDIS-09-15-1085-RE.
911 64. Aiello D, Scuderi G, Vitale A, Firrao G, Polizzi G, Cirvilleri G. A pith necrosis caused by
912 *Xanthomonas perforans* on tomato plants. Eur J Plant Pathol. 2013;137(1):29-41. doi: 10.1007/s10658-
913 013-0214-7.
914 65. Jibrin MO, Timilsina S, Potnis N, Minsavage GV, Shenge KC, Akpa AD, et al. First report of
915 atypical *Xanthomonas euvesicatoria* strains causing bacterial spot of tomato in Nigeria. Plant Dis.
916 2015; 99(3):415. doi: https://doi.org/10.1094/PDIS-09-14-0952-PDN.

66. Timilsina S, Abrahamian P, Potnis N, Minsavage GV, White FF, Staskawicz BJ, et al. Analysis of sequenced genomes of *Xanthomonas perforans* identifies candidate targets for resistance breeding in tomato. Phytopathology. 2016;106(10):1097-104. doi: 10.1094/PHYTO-03-16-0119-FI.

67. Potnis N, Krasileva K, Chow V, Almeida NF, Patil PB, Ryan RP, et al. Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. BMC Genomics. 2011;12(1):146.

68. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992;132.

69. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585-95.

70. Didelot X, Wilson DJ. ClonalFrameML: Efficient inference of recombination in whole bacterial genomes. PLoS Comp Biol. 2015;11(2):e1004041. doi: 10.1371/journal.pcbi.1004041.

71. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. Wellcome Open Research. 2019;3:93. doi: https://doi.org/10.12688/wellcomeopenres.14694.1.

72. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43(3):e15-e.

73. Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. Bayesian inference of ancestral dates on bacterial phylogenetic trees. Nucleic Acids Res. 2018;46(22):e134-e. doi: 10.1093/nar/gky783.

74. Stall RE, Jones JB, Minsavage GV. Durability of resistance in tomato and pepper to xanthomonads causing bacterial spot. Annu Rev Phytopathol. 2009;47(1):265-84. doi: 10.1146/annurev-phyto-080508-081752.

75. Sharlach M, Dahlbeck D, Liu L, Chiu J, Jiménez-Gómez JM, Kimura S, et al. Fine genetic mapping of *RXopJ4*, a bacterial spot disease resistance locus from *Solanum pennellii* LA716. Theor Appl Genet. 2013;126(3):601-9. doi: 10.1007/s00122-012-2004-6.

942 76. Sharma A, Iruegas-Bocardo F, Bibi S, Chen Y-C, Kim J-G, Abrahamian P, et al. Multiple
943 acquisitions of XopJ2 effectors in populations of Xanthomonas perforans. Mol Plant-Microbe Interact.
944 2024. doi: 10.1094/MPMI-05-24-0048-R.
945 77. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: A
946 Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Mol Biol Evol. 2013;30(5):1196-
947 205.
948 78. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting
949 Individual Sites Subject to Episodic Diversifying Selection. PLoS Genet. 2012;8(7):e1002764. doi:
950 10.1371/journal.pgen.1002764.
951 79. Behlau F, Canteros BI, Minsavage GV, Jones JB, Graham JH. Molecular characterization of
952 copper resistance genes from *Xanthomonas citri* subsp. *citri* and *Xanthomonas alfalfae* subsp.
953 *citrumelonis*. Appl Environ Microbiol. 2011;77(12):4089-96.
954 80. Behlau F, Gochez AM, Lugo AJ, Elibox W, Minsavage GV, Potnis N, et al. Characterization of a
955 unique copper resistance gene cluster in Xanthomonas campestris pv. campestris isolated in Trinidad,
956 West Indies. Eur J Plant Pathol. 2017;147(3):671-81. doi: 10.1007/s10658-016-1035-2.
957 81. Stall, R. E. L, D. C.,, Jones JB. Linkage of copper resistance and avirulence loci on a self-
958 transmissible plasmid in Xanthomonas campestris pv. vesicatoria. Phytopathology; 1986. p. 240-3.
959 82. Gitaitis R, Walcott R. The epidemiology and management of seedborne bacterial diseases. Annu
960 Rev Phytopathol. 2007;45(1):371-97. doi: 10.1146/annurev.phyto.45.062806.094321.
961 83. Munkvold GP. Seed pathology progress in academia and industry. Annu Rev Phytopathol.
962 2009;47(1):285-311. doi: 10.1146/annurev-phyto-080508-081916.
963 84. Bank W. United States Seed; vegetable seed, of a kind used for sowing imports by country in
964 2019
965 https://wits.worldbank.org/trade/comtrade/en/country/USA/year/2019/tradeflow/Imports/partner/ALL/pro
966 duct/1209912019 [2022-09-15]. Available from:

https://wits.worldbank.org/trade/comtrade/en/country/USA/year/2019/tradeflow/Imports/partner/ALL/product/120991.

85. Jones JB, Pohronezny KL, Stall RE, Jones JP. *Xanthomonas campestris* pv. *vesicatoria* in Florida on tomato crop residue, weeds, seeds, and voluteer tomato plants. Phytopathology. 1986;76(4):430-4.

86. Jones JB, Bouzar H, Stall RE, Almira EC, Roberts PD, Bowen BW, et al. Systematic analysis of xanthomonads (*Xanthomonas* spp.) associated with pepper and tomato lesions. Int J Syst Evol Microbiol. 2000;50 Pt 3:1211-9. doi: 10.1099/00207713-50-3-1211. PubMed PMID: 10843065.

87. Dutta B, Gitaitis R, Sanders H, Booth C, Smith S, Langston DB. Role of blossom colonization in pepper seed infestation by *Xanthomonas euvesicatoria*. Phytopathology. 2014;104(3):232-9. doi: 10.1094/phyto-05-13-0138-r. PubMed PMID: 24111576.

88. Giovanardi D, Biondi, E.,, Ignjatov M, Jevtić, R., Stefani E. Impact of bacterial spot outbreaks on the phytosanitary quality of tomato and pepper seeds. Plant Pathol. 2018;67:1168-76. doi: 10.1111/ppa.12839.

89. Parajuli A, Subedi A, Timilsina S, Minsavage GV, Kenyon L, Chen J-R, et al. Phenotypic and Genetic Diversity of Xanthomonads Isolated from Pepper (*Capsicum* spp.) in Taiwan from 1989 to 2019. Phytopathology. 2024;114(9):2033-44. doi: 10.1094/PHYTO-11-23-0449-R.

90. Iruegas-Bocardo F, Weisberg AJ, Riutta ER, Kilday K, Bonkowski JC, Creswell T, et al. Whole genome sequencing-based tracing of a 2022 introduction and outbreak of *Xanthomonas hortorum* pv. *pelargonii*. Phytopathology. 2022;113(6):975-84. doi: 10.1094/PHYTO-09-22-0321-R.

91. Emsweller SL. Fundamental Procedures in Breeding Crops. The Yearbook of Agriculture U.S. Department of Agriculture; 1961. p. 127-33.

92. Dorst J. Een en twintigste beschrijvende rassenlijst voor landbouwgewassen. Wageningen, Rijkscommissie voor de samenstelling van de rassenlijst voor landbouwgewassen; 1946. p. 221.

93. Gedgaew C, Simaraks S, Rambo AT. Trends in hybrid tomato seed production under contract farming in Northeast Thailand. Southeast Asian Studies. 2017;6(2):339-3.

992 94. Ltd." K-YNSPC, editor Production and marketing of F1 hybrid tomato seed in Taiwan. 1st
993 International Symposium on Tropical Tomato; 1978: AVRDC.
994 95. Kloppenburg JR. First the seed: The political economy of plant biotechnology, 1492-2000 (2nd
995 edition). Madison, WI: University of Wisconsin Press; 2004.
996 96. Quibod IL, Atieza-Grande G, Oreiro EG, Palmos D, Nguyen MH, Coronejo ST, et al. The Green
997 Revolution shaped the population structure of the rice pathogen *Xanthomonas oryzae* pv. *oryzae*. The
998 ISME Journal. 2020;14(2):492-505. doi: 10.1038/s41396-019-0545-2.
999 97. McDonald BA, Linde C. Pathogen population genetics, evolutionary potential, and durable
1000 resistance. Annu Rev Phytopathol. 2002;40:349-79.
1001 98. Gassmann W, Dahlbeck D, Chesnokova O, Minsavage Gerald V, Jones Jeffrey B, Staskawicz
1002 Brian J. Molecular Evolution of Virulence in Natural Field Strains of *Xanthomonas campestris* pv.
1003 *vesicatoria*. J Bacteriol. 2000;182(24):7053-9. doi: 10.1128/jb.182.24.7053-7059.2000.
1004 99. Klein J, Abrahamian P, Xing Y, Fulton J, Minsavage GV, Timilsina S, et al. It's complicated: The
1005 Xanthomonas perforans population across Florida commercial tomato fields. Phytopathology.
1006 2019;109(10):12-3. PubMed PMID: WOS:000492694200045.
1007 100. Torelli E, Aiello D, Polizzi G, Firrao G, Cirvilleri G. Draft genome of a *Xanthomonas perforans*
1008 strain associated with pith necrosis. FEMS Microbiol Lett. 2015;362(4). doi: 10.1093/femsle/fnv001.
1009 PubMed PMID: 25724775.
1010 101. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1011 EMBnetjournal; Vol 17, No 1: Next Generation Sequencing Data AnalysisDO - 1014806/ej171200. 2011.
1012 102. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new
1013 genome assembly algorithm and Its applications to single-cell sequencing. J Comput Biol.
1014 2012;19(5):455-77.
1015 103. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
1016 2012;9(4):357-9. doi: 10.1038/nmeth.1923.

104. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS ONE. 2014;9(11):e112963. doi: 10.1371/journal.pone.0112963.

105. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043-55.

106. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 2012;40(D1):D115-D22.

107. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol. 2013;79(24):7696-701. doi: 10.1128/AEM.02411-13.

108. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215. doi: 10.1016/s0022-2836(05)80360-2.

109. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. doi: 10.1093/molbev/mst010.

110. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 2011;27(2):171-80. doi: https://doi.org/10.1111/j.1096-0031.2010.00329.x.

111. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3. doi: https://doi.org/10.1093/bioinformatics/btu033.

112. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Mol Biol Evol. 2013;30(5):1224-8. doi: 10.1093/molbev/mst028.

113. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing https://www.R-project.org/; 2019.

1043 114. Yu G. Using ggtree to visualize data on tree-like structures. Current Protocols in Bioinformatics.
1044 2020;69(1):e96. doi: https://doi.org/10.1002/cpbi.96.
1045 115. Sievert C. Interactive web-based data visualization with R, plotly, and shiny. Florida: CRC Press;
1046 2020.
1047 116. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the
1048 Tidyverse. Journal of Open Source Software. 2019;4:1686. doi: doi: 10.21105/joss.01686.
1049 117. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: An Efficient Swiss
1050 Army Knife for Population Genomic Analyses in R. Mol Biol Evol. 2014;31(7):1929-36. doi:
1051 10.1093/molbev/msu136.
1052 118. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in
1053 diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods.
1054 2016;8(1):12-24. doi: 10.1039/C5AY02550H.
1055 119. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale
1056 prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3. Epub 2015/07/20. doi:
1057 10.1093/bioinformatics/btv421. PubMed PMID: 26198102; PubMed Central PMCID:
1058 PMCPMC4817141.
1059 120. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
1060 doi: 10.1093/bioinformatics/btu153.
1061 121. Snipen L, Liland KH. micropan: an R-package for microbial pan-genomics. BMC
1062 Bioinformatics. 2015;16:79. Epub 2015/03/12. doi: 10.1186/s12859-015-0517-0. PubMed PMID:
1063 25888166; PubMed Central PMCID: PMCPMC4375852.
1064 122. Harris SR. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology. bioRxiv.
1065 2018;453142. doi: https://doi.org/10.1101/453142.
1066 123. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics.
1067 1998;14(1):68-73. doi: 10.1093/bioinformatics/14.1.68.

124. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254-67. Epub 2005/10/14. doi: 10.1093/molbev/msj030. PubMed PMID: 16221896.

125. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinforma. 2006;22. doi: 10.1093/bioinformatics/btl446.

126. Doizy A, Prin A, Cornu G, Chiroleu F, Rieux A. Phylostems: a new graphical tool to investigate temporal signal of heterochronous sequences datasets. Bioinformatics Advances. 2023;3(1):vbad026. doi: 10.1093/bioadv/vbad026.

127. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52(W1):W78-W82. doi: 10.1093/nar/gkae268.

128. Costa J, Pothier JF, Bosis E, Boch J, Kölliker R, Koebnik R. A community-curated DokuWiki resource on diagnostics, diversity, pathogenicity and genetic control of xanthomonads. Mol Plant-Microbe Interact. 2023. doi: 10.1094/MPMI-11-23-0184-FI.

129. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, et al. The Pfam protein families database. Nucleic Acids Res. 2010;38(suppl_1):D211-D22. doi: 10.1093/nar/gkp985.

130. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: a more efficient report with usability improvements. Nucleic Acids Res. 2013;41(W1):W29-W33. doi: 10.1093/nar/gkt282.

131. Altschul SF. BLAST Algorithm. eLS. 2014. doi: https://doi.org/10.1002/9780470015902.a0005253.pub2.

132. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460-1.

133. Takebayashi N. uniqueHaplo.pl, http://raven.wrrb.uaf.edu/~ntakebay/teaching/programming/perl-scripts/uniqHaplo.pl, accessed 4-01-23. 2015.

134. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: Community Ecology Package. R package version 2.5-6. . https://CRAN.R-project.org/package=vegan. 2019.

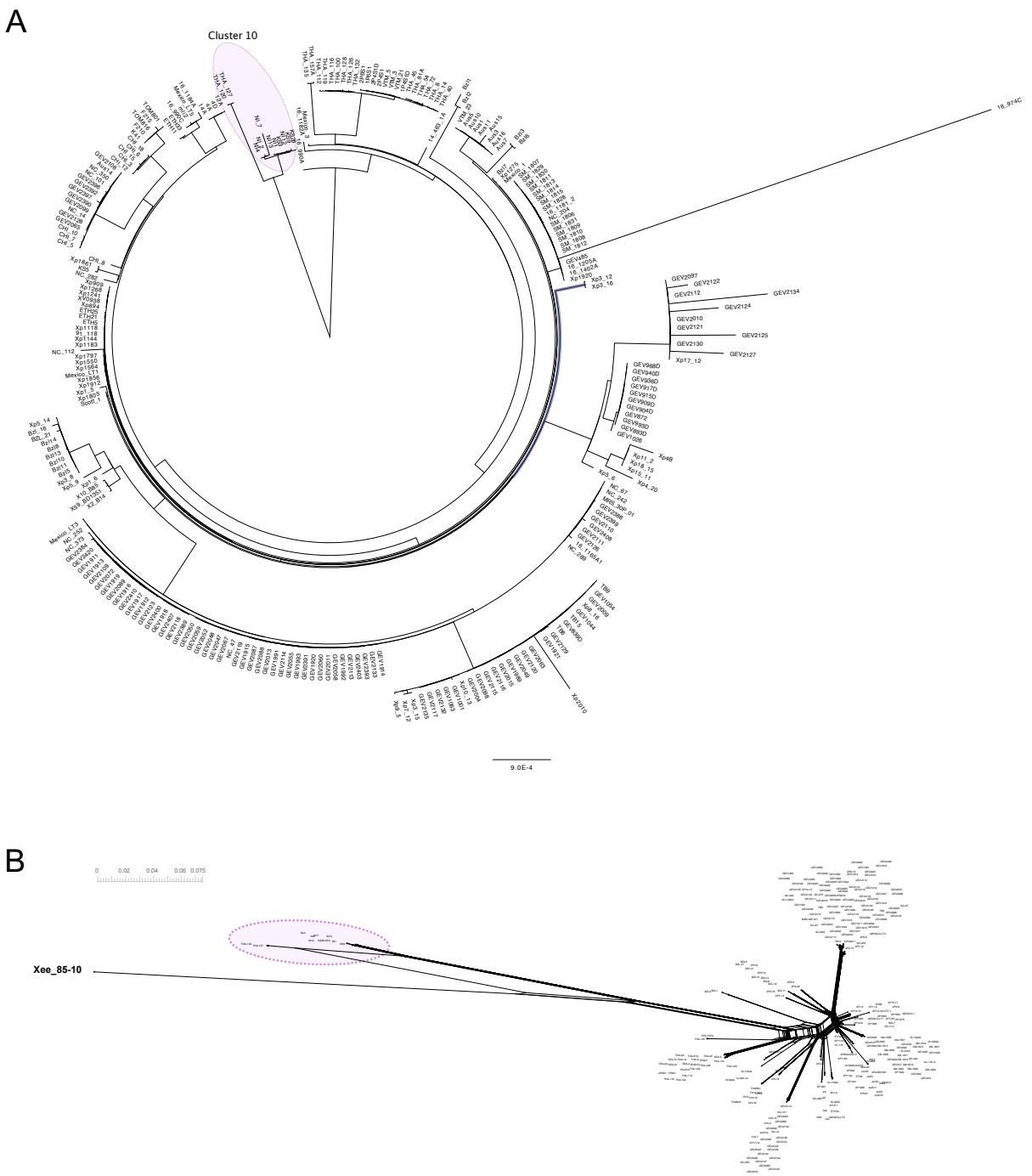
1094 135. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R
1095 Programming Tools for Plotting Data. R package version 3.1.3, <https://CRAN.R-
1096 project.org/package=gplots>. 2020.
1097 136. Löytynoja A. Phylogeny-aware alignment with PRANK. In: Russell DJ, editor. Multiple
1098 Sequence Alignment Methods. Totowa, NJ: Humana Press; 2014. p. 155-70.
1099 137. Slowikowski K. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R
1100 package version 0.9.1, <https://CRAN.R-project.org/package=ggrepel>. 2020.
1101 138. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2 ed: Springer-Verlag New York;
1102 2016.
1103 139. Yu G. scatterpie: Scatter Pie Plot. R package version 0.1.8,
1104 <https://CRAN.R-project.org/package=scatterpie>. 2020.
1105 140. Wilke C. cowplot: Streamlined Plot Theme and Plot Annotations for 'gglot2'. R package version
1106 1.0.0, <https://cran.r-project.org/package=cowplot>. 2019.
1107 141. Dunnington D, Thorne B, Hernangómez D. ggspatial: Spatial Data Framework for ggplot2. R
1108 package v. 1.1.1, <https://cran.r-project.org/package=ggspatial>. 2020.
1109 142. Pebesma E. Simple features for R: Standardized support for spatial vector data. The R Journal.
1110 2018;10(1):439–46. doi: 10.32614/RJ-2018-009.
1111 143. Bivand R, Rundel C. rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package v.
1112 0.5-1, https://cran.r-project.org/package=rgeos. 2019.
1113 144. Elff M. memisc: Management of Survey Data and Presentation of Analysis Results. R package
1114 version 0.99.27, <https://cran.r-project.org/package=memisc>. 2020.
1115 145. Venables B, Hornik K. oz: Plot the Australian Coastline and States. R package version 1.0,
1116 <https://cran.r-project.org/package=oz>. 2016.
1117 146. Bivand R, Lewin-Koh N. maptools: Tools for Handling Spatial Objects. R package version 1.0.
1118 https://CRAN.R-project.org/package=maptools. 2020.

147. South A. rnaturalearth: World Map Data from Natural Earth. R package version 0.1.0, <https://cran.r-project.org/package=rnaturalearth>. 2017.

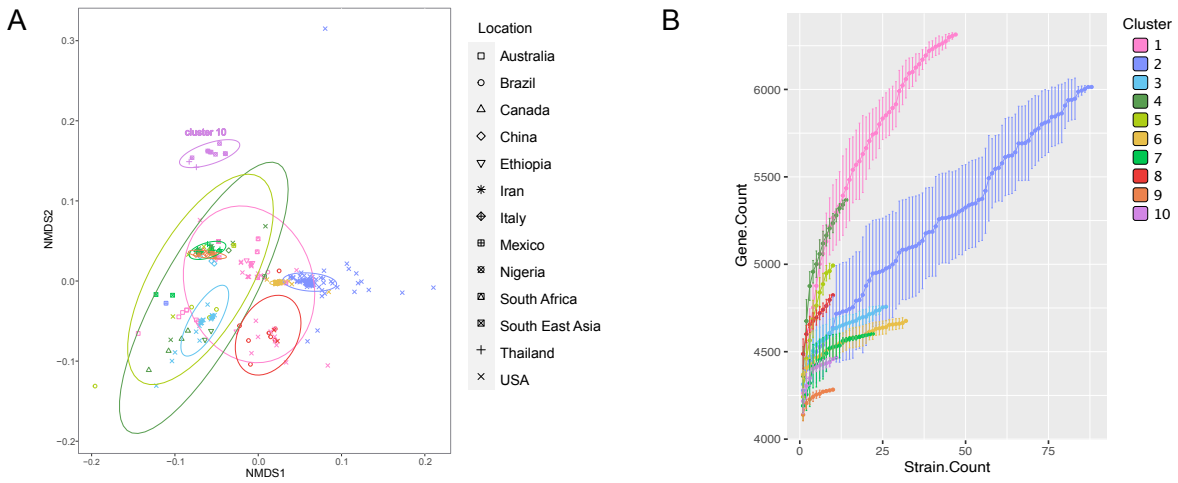
148. Cooley D. googleway: Accesses Google Maps APIs to Retrieve Data and Plot Maps. R package version 2.7.1. https://CRAN.R-project.org/package=googleway 2018.

149. Minsavage GV, Canteros BI, Stall RE. Plasmid- mediated resistance to streptomycin in *Xanthomonas campestris* pv. *vesicatoria*. Phytopathology. 1990;80:719-23.

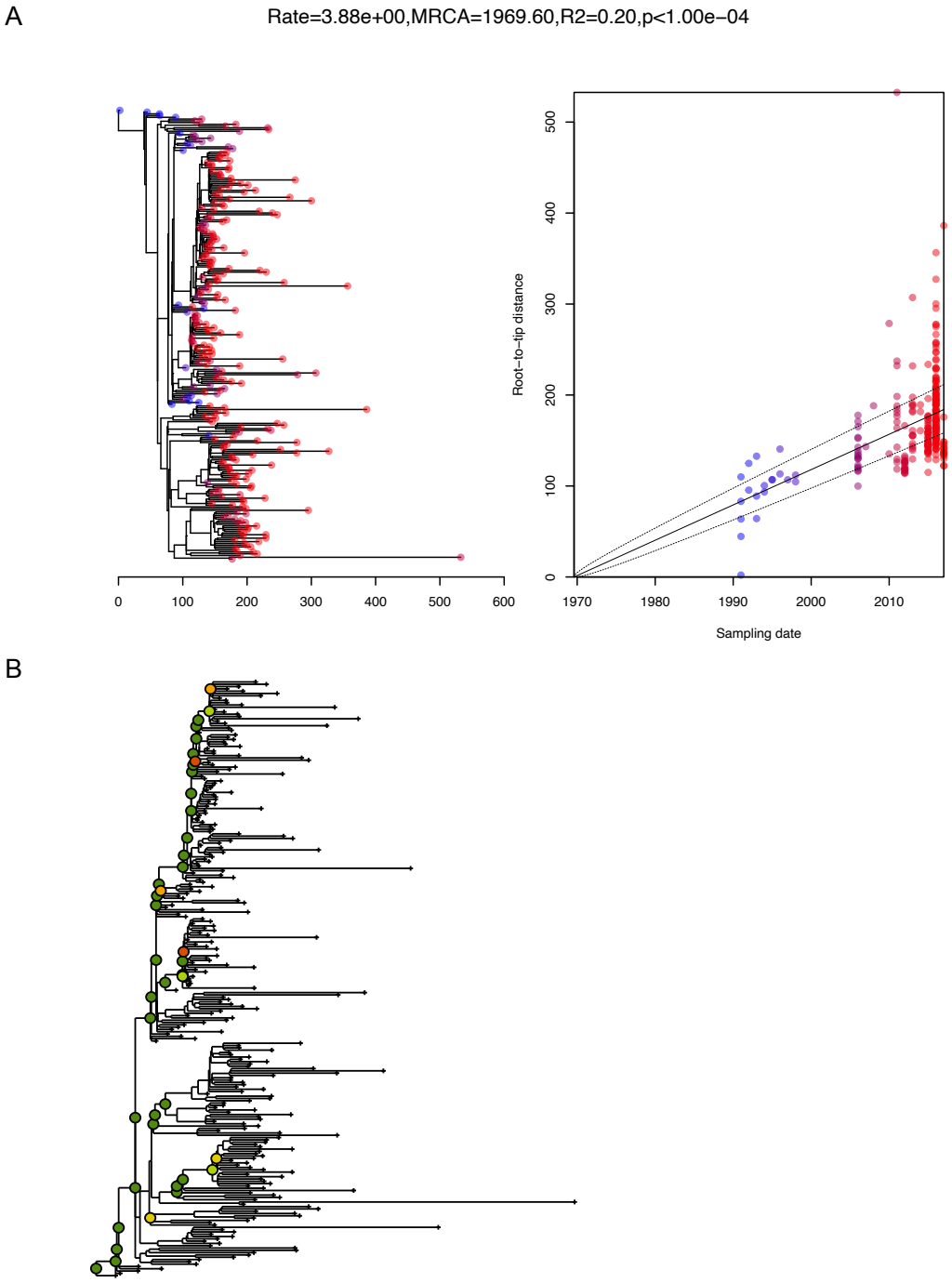
150. Staskawicz BJ, Dahlbeck D, Keen N, Napoli C. Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*. J Bacteriol. 1987;169:5789-94.



S1 Figure. Phylogenetic analysis of 270 *X. euvesicatoria* pv. *perforans* strains. (A) Maximum likelihood phylogenetic tree of *Xanthomonas euvesicatoria* pv. *perforans* strains based on aligned nucleotide sequences of 887 core genes. The tree was inferred using RAxML using a GTRGAMMAI substitution model. The tree was rooted using the 11 genetically diverged strains that make up core gene cluster 10. (B) Neighbor-net network inferred using SNPs from aligned whole genome sequences, including *X. euvesicatoria* pv. *euvesicatoria* strain 85-10 (bolded) as an outgroup. Core gene cluster 10 strains are highlighted. Reticulations in the network indicate conflicting phylogenetic relationships.



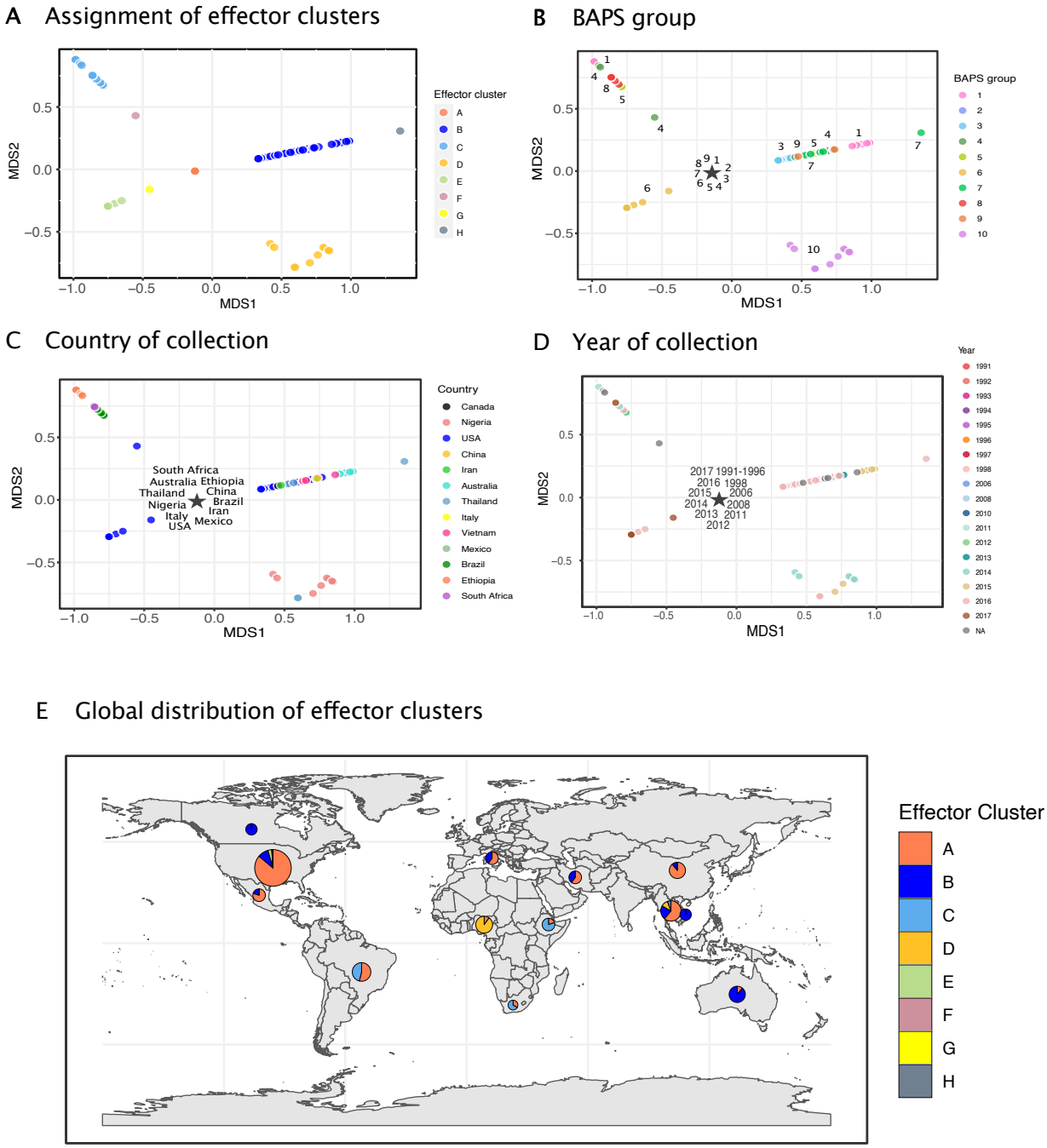
S2 Figure. Accessory genome variation in *Xanthomonas euvesicatoria* pv. *perforans*. (A) Visualization of pangenome variation by non-metric multidimensional scaling of gene presence-absence for all 270 *X. perforans* strains by BAPS cluster. Ellipses assume a multivariate t-distribution. (B) Increase in gene count with increasing number of strains sampled. Clusters 1 and 2 were represented by the most strains, but other clusters showed similar rates of increase in the pangenome of the cluster.



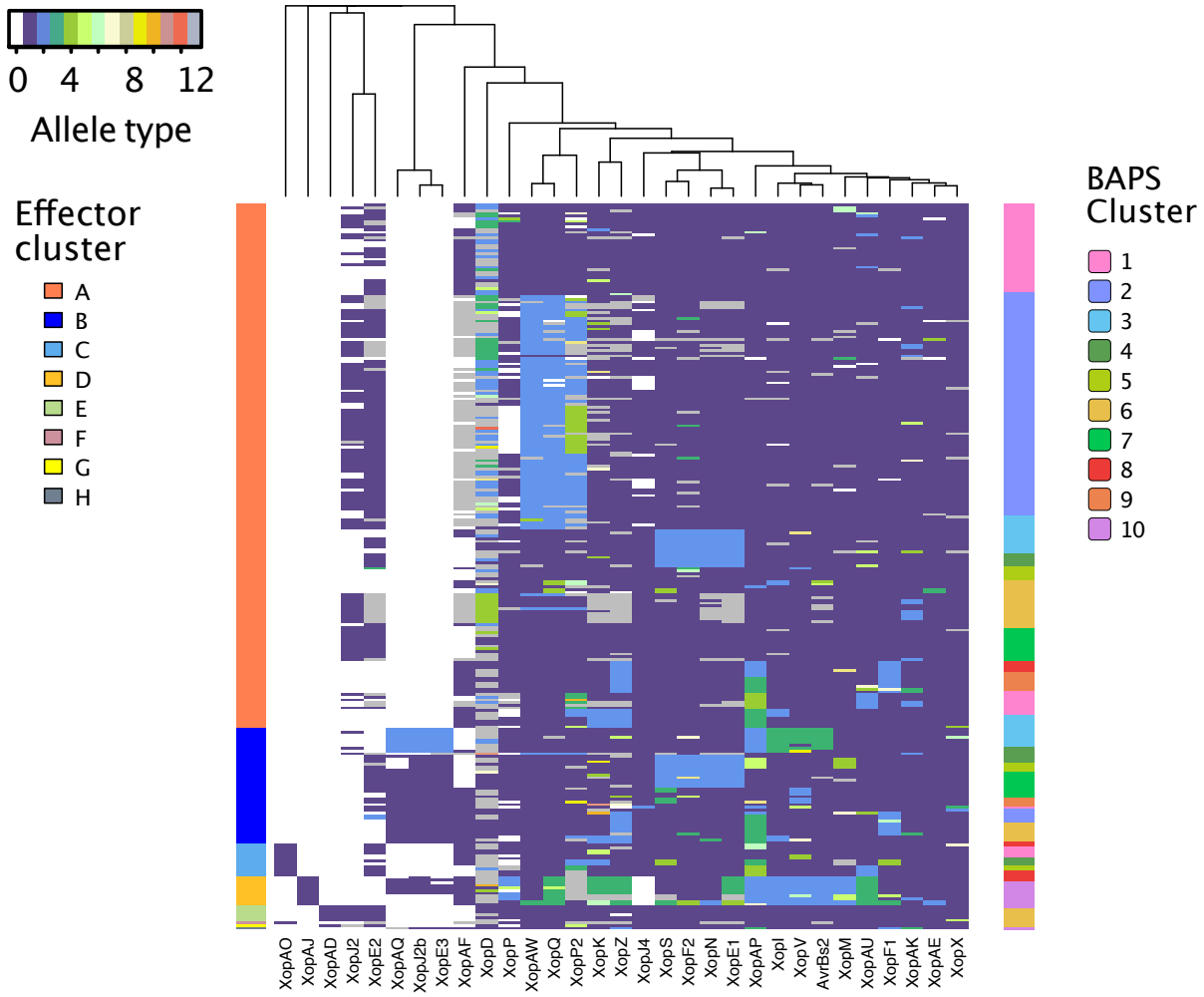
S3 Figure. Temporal signal in phylogenetic tree of 259 *X. euvesicatoria* pv. *perforans* strains. (A) Correlation between sampling year and root-to-tip distance in maximum likelihood phylogenetic tree inferred from alignment of whole genome sequences. Output was generated from BactDating R package. (B) Temporal signal within the phylogenetic determined using Phylostems tool. Tree is rooted as in Figure 2. Nodes with statistically significant temporal signals are indicated with colored circles. Adjusted R-squared values by color: dark green 0–0.2; light green 0.2–0.4; yellow 0.4–0.6; orange 0.6–0.8; red 0.8–1.

S4 Figure. Dated phylogenies of 259 *X. euvesicatoria* pv. *perforans* strains. (A) Dating using BactDating relaxed clock analysis on RAxML-generated phylogeny. This is the tree shown in Figure 2, shown here without collapsed nodes. (B) Dating of same dataset using BEAST with coalescent Bayesian skyline priors and an uncorrelated relaxed clock.

S5 Figure. Frequency of Xop effectors among 270 *Xanthomonas euvesicatoria* pv. *perforans* strains. The most common allele observed was assigned to allele type 1, second most frequent allele to allele type 2, and so on. Note that alleles classified as pseudogenes included contig breaks, which include assembly errors. For example, all strains appear to have *xopD*, but a repeat caused a contig break in the gene in nearly half of the genomes.



S6 Figure. Clustering of 270 *Xanthomonas perforans* effector profiles by non-metric multidimensional scaling and distribution of resulting clusters among geographic regions. Analysis did not include TAL effectors. (A) The most frequently observed group of effector profiles form cluster A. This cluster of 188 strains is represented as a star in plots B-C, as it is represented in most BAPS core gene clusters (B), most of the sampled tomato production regions (C), and in collections from 1991 to 2017 (D). Clusters were largely defined by low frequency effectors (S7 Figure). (E) Distribution of strains by effector clusters among sampled countries.



S7 Figure. Variation in Type III secreted effector profiles in 270 *Xanthomonas euvesicatoria* pv. *perforans* strains ordered according to NMDS of effector profiles. Analysis did not include TAL effectors. Type III secreted effectors are in columns and *Xep* strains in rows. Effector status is shown by allele type: absence is indicated by allele type 0 (white), while the most frequent allele observed when the effector is present is allele type 1 (purple), second most frequent is allele type 2 (blue), and so on. Putative pseudogenized effectors are shown as allele 13 (gray). The order of columns was determined by hierarchical clustering analysis, placing similarly distributed effectors adjacent to each other. Order of rows is based on NMDS clustering analysis of effector profiles (see S6 Figure, part A).