Intercontinental dissemination and diversification of *Xanthomonas perforans* in tomato production

Sujan Timilsina*, Fernanda Iruegas-Bocardo*, Mustafa O. Jibrin*, Anuj Sharma, Aastha Subedi, Gerald V. Minsavage, Jose Huguet-Tapia, Jeannie Klein-Gordon, Pragya Adhikari, Tika B. Adhikari, Gabriella Cirvilleri, Laura Belen Tapia de la Barrera, Eduardo Bernal, Tom C. Creswell, Doan Thi Kieu Tien, Teresa A. Coutinho, Daniel S. Egel, Rubén Félix-Gastélum, David M. Francis, Misrak Kebede, Melanie Lewis Ivey, Frank J. Louws, Laixin Luo, Elizabeth T. Maynard, Sally A. Miller, Nguyen Thi Thu Nga, Ebrahim Osdaghi, Alice M. Quezado-Duval, Rebecca Roach, Francesca Rotondo, Gail E. Ruhl, Vou M. Shutt, Petcharat Thummabenjapone, Cheryl Trueman, Pamela D. Roberts, Jeffrey B. Jones†, Gary E. Vallad†, Erica M. Goss†

Affiliations

Sujan Timilsina, Fernanda Iruegas-Bocardo, Mustafa O. Jibrin, Anuj Sharma, Gerald V. Minsavage, Jose Huguet-Tapia, Aastha Subedi, Jeannie Klein-Gordon, Pamela D. Roberts, Jeffrey B. Jones, Gary E. Vallad, Erica M. Goss
Plant Pathology Department, University of Florida, Gainesville, Florida, USA.

Mustafa O. Jibrin
Department of Crop Protection, Ahmadu Bello University, Zaria, Nigeria.

Mustafa O. Jibrin, Pamela D. Roberts
Southwest Florida Research and Education Center, University of Florida, Immokalee, FL 34142, USA.

Tika B. Adhikari, Frank J. Louws
Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695

Gabriella Cirvilleri
Dipartimento di Agricoltura, Alimentazione e Ambiente, Sezione Patologia Vegetale, Via S. Sofia 100, 95123 Catania, Italy

Laura Belen Tapia de la Barrera
Centro de Investigacíon an Alimentación y Desarrollo (CIAD), Mexico

Eduardo Bernal, David M. Francis
Department of Horticulture and Crop Science, The Ohio State University, Wooster, OH 44691, USA.

Tom C. Creswell, Daniel S. Egel, Gail E. Ruhl
Botany and Plant Pathology Department, Purdue University, West Lafayette, IN 47907, USA.

Teresa A. Coutinho and Vou M. Shutt
Department of Microbiology and Plant Pathology, Centre for Microbial Ecology and Genomics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0002, South Africa.

Rubén Félix-Gastélum

Departamento de Ciencias Biológicas, Universidad Autónoma de Occidente, Unidad Los Mochis, Blvd. Macario Gaxiola y Carretera Internacional s/n Los Mochis, Sinaloa, CP 81223, México

Misrak Kebede
Biotechnology Department, Collage of Biological and Chemical Engineering, Addis Ababa Science and Technology University, Addis Ababa Ethiopia.

Pragya Adhikari, Frank J. Louws
Department of Horticultural Science, NC State University
Raleigh, NC 27695-7609

Laixin Luo
Department of Plant Pathology, China Agricultural University, Beijing, China P. R.

Elizabeth T. Maynard
Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907, USA.

Francesca Rotondo, Melanie Lewis Ivey, Sally A. Miller
Department of Plant Pathology, The Ohio State University, Wooster, OH 44691, USA.

Doan Thi Kieu Tien, Nguyen Thi Thu Nga
Department of Plant Protection, College of Agriculture, Can Tho University, Can Tho City, Vietnam

Ebrahim Osdaghi
Department of Plant Protection, College of Agriculture, University of Tehran, Karaj, 31587-77871, Iran.

Alice M. Quezado-Duval
Embrapa Hortalicas. Laboratorio de Fitopatologia. Rod. BR 060 Km 9, C.P. 218 Brasilia-DF, Brazil.

Rebecca Roach
Queensland Department of Agriculture and Fisheries, Brisbane, Queensland, Australia

Vou M. Shutt
Department of Plant Science and Biotechnology, University of Jos, Nigeria.
shutt_vou25@yahoo.com

Petcharat Thummabenjapone
Khon Kaen University, Plant Science and Agricultural Resources, Section Plant Pathology, Khon Kaen, Thailand

Cheryl Trueman
Department of Plant Agriculture, Ridgetown Campus, University of Guelph, Ridgetown, Ontario, Canada

Gary E. Vallad
Gulf Coast Research and Education Center, University of Florida, Wimauma, FL 33598, USA.

Erica M. Goss
Emerging Pathogen Institute, University of Florida, Gainesville, FL 32610, USA.

Running title: Intercontinental dissemination of *X. perforans*

***All first three authors contributed equally.**
† Corresponding authors

Abstract

Tomato is a high value crop that requires quality seed each season. Tomato production often relies on international breeding and production chains with a global network to deliver seeds to growers. *Xanthomonas perforans*, one of the causal agents of bacterial spot of tomato, is seedborne and has rapidly emerged and displaced other bacterial spot xanthomonads in tomato production regions around the world. The objectives of this work were to determine if different tomato production regions contain genetically distinct *X. perforans* populations, examine genetic relatedness of strains collected in tomato seed production areas in East Asia and other production regions, estimate the timing of *X. perforans* population expansion relative to its first report in 1991, and evaluate variation in type III effectors, which are critical pathogenicity and virulence factors. Our results showed extensive genetic diversity in the pathogen and highly similar strains in distant tomato production regions, which is consistent with global dissemination. Furthermore, we found genetically similar strains in tomato seed production and fruit production regions, as would be expected if the pathogen was being moved through tomato production chains from a common origin. We found waves of diversification of populations in the United States, where our sampling was more extensive, before and since its first detection, consistent with emergence of new *X. perforans* genotypes from diverse source populations. Variation in gene content, including effector genes, within and between core genome lineages suggests the potential for rapid changes to the accessory genome upon dissemination. Evolution of the *X. perforans* pangenome, including the acquisition and loss of type III secreted effectors, will continue to challenge sustainable management of tomato bacterial spot.

Introduction/Main

Tomato is a high value fruiting-vegetable crop, with more than 189 million metric tons of global production in 2021 [1]. 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 [2]. 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 [3].

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 [4-6]. 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 [7-14]. 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 perforans* [15,16] (or *X. euvesicatoria* pv. *perforans*, [17]) has emerged rapidly and become a major player on tomato [12,18-26]. *X. perforans* was first reported in 1991 in Florida, USA [15] and is now found in all tomato production areas of the world, including regions with no history of the disease [27]. A plausible hypothesis for new outbreaks of *X. perforans* is pathogen movement with seeds and planting materials [28].

The success of *X. perforans* as a pathogen, and its replacement of *X. euvesicatoria* on tomato, 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 [29-35]. Distinct genetic lineages of *X. perforans*, 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 [30,33,34,36,37]. Additional lineages of *X. perforans* have been found in Nigeria, Iran, Italy, and the Southwest Indian Ocean islands [27,31,38].

Xanthomonas perforans strains, like other xanthomonads, acquire nutrients through colonization of compatible hosts. The type III secretion system (T3SS) and type III secreted effector (T3SE) proteins are critical for suppression of host defenses and virulence by *X. perforans* [39]. Effector content varies among *Xanthomonas* species and distinct lineages of *X. perforans* have distinguishable effector content [8,31,33,40,41]. Early populations of *X. perforans* were limited to tomato [42], but now it has been found to cause bacterial spot disease of pepper [33,41,43]. 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*. Effector variation may cause differences in disease epidemiology in addition to host range [32,40,44]. For example, wildtype strains with the acquired effector XopJ2 showed three times faster spread in the field than isogenic mutant strains without the effector [34].

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) [45]. *X. perforans* 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 *X. perforans* 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 *X. perforans* 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 *X. perforans;* genetically similar strains in distant geographic regions, inclusive of seed production regions; bouts 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. perforans strains from seed and commercial fruit production areas

A total of 270 *X. perforans* genomes from 13 different countries – representing seed and fruit production – were used in this study (Table 1). *X. perforans* strains were differentiated from other tomato-pathogenic xanthomonads using a real-time qPCR assay that specifically amplifies the *hrcN* (*hrpB7*) gene in *X. perforans* [46] 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.

Genomic diversity in *X. perforans*

To examine genetic diversity in the core genome, we identified 887 genes that were present in all 270 *X. perforans* 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. Maximum likelihood phylogenetic analysis showed a distinct and especially diverged lineage of 11 strains from Nigeria and Thailand, that included a previously defined atypical strain – NI1 – from Nigeria [31]. Due to the genetic distance introduced by these 11 strains (Extended Data Fig. 1), they were not included in subsequent phylogenetic analyses. Phylogenetic analysis of core SNPs for 259 strains, followed by correction of branch lengths for recombination, showed many diversifying lineages of *X. perforans* (Figure 1). ClonalFrameML [47] 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. A neighbor-joining split network constructed with core SNPs output by ClonalFrameML indicated that there were still conflicting sites in the alignment, suggesting a potentially complex history of recombination in the diversification of *X. perforans* (Extended Data Figure 2A).

To summarize population structure based on core gene SNPs, we used hierBAPS [48], which assigned individual strains to 9 clusters using allele frequencies (Figure 1; Extended Data Table 1). This analysis excluded the 11 highly diverged strains from Nigeria and Thailand, which were placed in cluster 10. In some cases, clusters corresponded to phylogenetic lineages, including clusters 2, 3, 7, 8, and 9 (Figure 1). The remaining clusters were polyphyletic, often encompassing multiple diverged clades or individual strains. Presence-absence gene variation in the pangenome largely paralleled nucleotide variation in core genes (Extended Data Figure 2B). For instance, polyphyletic clusters 1, 4, and 5 also showed the most variation in gene content (Extended Data Figure 2C).

Global distribution of *X. perforans* populations

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 1C) 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 (Extended Data Table 2). 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, alignment fraction of 0.937). Likewise, strains collected in a seed production region in China had ANIb up to 99.99% compared to strains from Florida and North Carolina (alignment fraction of 0.996). 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 (Extended Data Table 2). Other strains were similar between countries in core genes only after correction for recombination.

Timing of *X. perforans* lineage emergence

We estimated the timing of diversification within cluster 1, which included the first described strains of *X. perforans*, collected in Florida tomato fields in 1991, as well as strains collected more recently in Ethiopia, Australia, South Africa, and North Carolina. Bacterial pathogens of tomato were intensively studied in Florida since the 1960s and *X. perforans* was not detected prior to 1991 [15,22]. Sampling year was significantly correlated with root-to-tip distance ($R^2 = 0.67$, $P < 1\times10^{-4}$). Dating of the most recent common ancestor (MRCA) of 33 cluster 1 strains using a strict clock estimated the root of the tree to 1978 (95% confidence interval: 1974–1983) (Figure 2). The strict clock model was the preferred model based on the deviance information criterion (DIC) of 387.9 versus relaxed model DIC of 487.4). As a check of the time signal in the data, setting all tip dates to 2006 resulted in a much higher DIC of 624.43, indicating poor model fit. The phylogenetic analysis of cluster 1 shows two very closely related clades of strains, represented by 91-118 versus Xp894 and XV0938, at the time of its first detection in Florida. The 91-118 lineage diversified in Florida and the other lineage became established in Africa and Australia.

Type III Effectors vary among geographic regions and core gene clusters

We detected 32 predicted Type III secreted effector genes in our collection of 270 strains (Extended Data Figure 3, Extended Data Table 3). The diversity in amino acid sequences of predicted effectors ranged widely from a single conserved allele to 8 or more alleles (Figure 3). None of the effectors were present and intact in 100% of our genomes, in

part due to our analysis of draft genomes. A notable case was XopD, which produced hits in every strain, but many strains showed a contig break within the gene. PCR and Sanger sequencing showed this to be an assembly issue due to the repeats within the gene. There were BLAST hits to previously described transcription activator-like (TAL) effectors in 65 strains. TAL effectors do not assemble in draft genomes due to their characteristic repeat sequences. Amplification of the repeat region of five strains produced band sizes consistent with the 13.5 repeats of AvrHah1 and 15.5 repeats of PthXp1 (Extended Data Table 3). Sanger sequencing of the entire TAL effector from strain 2P6S1 collected in Italy (NCBI accession number OQ588696) showed that it has the same repeat variable diresidues as PthXp1 reported in *X. perforans* strains from Alabama [33].

The tomato resistance gene, *Xv3*, targets effector XopAF (AvrXv3) [49], which 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* [50], 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 and is a fitness factor in tomato during the infection process [32], was present in less than half of strains examined (43%) and overwhelmingly in strains from the United States.

We defined the effector profile of each strain as the predicted presence or absence of each effector and its allelic state. 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. A lack of low frequency effectors characterized effector cluster A, containing 188 strains from 11 of 13 countries (Extended Data Figure 4). The remaining effector clusters were defined by the presence of one to three low frequency effectors (Extended Data Figure 5). While most effectors were found in multiple countries and continents (Extended Data Figure 4), populations in Brazil, Ethiopia, Nigeria, Thailand, South Africa, and the United States contained low frequency effectors that were not widely distributed.

Discussion

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 [51,52]. 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 [53]. *Xanthomonas* species can infest pepper and tomato seed [54], and *X. perforans* has been isolated from tomato seed [55,56], supporting the hypothesis that seeds can be a source of inoculum for bacterial spot outbreaks [57,58]. Thirty years after its first report, *X. perforans* has been identified in tomato production areas around the world [21]. Our results showed extensive genetic diversity in the pathogen and genetically similar strains in distant tomato production regions, which is consistent with continued global

346 dissemination. Furthermore, we found genetically similar strains in seed production and
347 fruit production regions on different continents, as would be expected if the pathogen was
348 being moved in shared production chains. We found multiple waves of diversification of
349 populations in the United States, where our sampling was more extensive, before and since
350 its first detection, consistent with introduction of new *X. perforans* genotypes from one or
351 more genetically diverse source population(s). Variation in gene content within and
352 between core genome lineages suggests the potential for rapid changes to the accessory
353 genome upon dissemination. The acquisition and loss of Type III secreted effectors will
354 continue to challenge sustainable management of tomato bacterial leaf spot [32,59].
355 Using our broad strain collection, we found *X. perforans* variants in seed production
356 regions in Asia that were previously reported in Australia, Italy, Nigeria, and the United
357 States [28,30,31,40]. Strains from Italy were nearly identical in core genes and very similar in
358 accessory genomes to strains collected from Thailand and Vietnam, both major seed
359 production regions. The atypical bacterial spot strains from Nigeria, recently designated as
360 race T5 [6], were genetically similar to strains from Thailand. A recently described variant of
361 *X. perforans* in Florida [cluster 3; [28,30]], which was also found in Australia [40], was similar in
362 the core genome to strains found in China; however, these strains showed divergence in the
363 pangenome, consistent with accessory genome evolution in emergent populations. Beyond
364 previously described variants, we found strains in Iran that were closely related to strains
365 from China; multiple instances of genetic similarity between strains from North America
366 and Ethiopia; and highly similar strains shared between USA and Brazil, USA and
367 Australia, and between Australia and Vietnam. Given the variation of *X. perforans* across
368 our sample, genetic similarity in core genes and gene content across continents is strong
369 evidence of international dissemination. Genetically similar strains of bacterial spot
370 pathogens *X. euvesicatoria* pv. *euvesicatoria* and *X. hortorum* pv. *gardneri* collected from
371 different continents similarly suggest intercontinental dissemination in tomato and pepper
372 seed [41]. Whole genome analysis of *X. hortorum* pv. *pelargonii* strains from a 2022 epidemic
373 of bacterial blight of geranium in the USA showed zero to seven chromosomal SNPs among
374 isolates of the emergent strain that distributed to multiple states in plant cuttings [60,61].
375 Other *X. perforans* genotypes indicated a more limited distribution. We did not find
376 core gene cluster 2 strains in the seed production regions sampled (China, Thailand,
377 Vietnam), while this lineage was highly represented in our USA sample [28,30,6237]. To date,
378 strains in this cluster have been found only in the southeastern and midwestern USA [33,36]
379 and Mexico. Seedling nurseries in the southeast USA produce tomato transplants for
380 growers in multiple states. Interstate movement of strains on seedlings is likely
381 responsible, at least in part, for disseminating genetically similar strains to different states
382 [28]. We previously reported extensive recombination with *X. euvesicatoria* pv. *euvesicatoria*
383 in cluster 2 strains [30] and this cluster had a diverse accessory genome, perhaps suggesting
384 that it has a larger geographic distribution than represented in our sample, which is biased
385 towards the USA.
386 To better understand the initial emergence of *X. perforans*, we examined the timing
387 of divergence of some of the first isolated strains of *X. perforans*, which were assigned to
388 core gene cluster 1. Strains from this cluster are positioned along the backbone of the
389 maximum likelihood tree (Fig. 1A), suggesting that early strains may best represent the
390 pre-dissemination population of *X. perforans*. The root of cluster 1 strains dates to the late

1970s, which coincides with a major expansion in tomato seed production in response to demand for hybrid tomato varieties [63,64]. The *X. perforans* strains we examined from the USA were assigned to core gene clusters 1, 2, 3, 4, 5, 6, and 8, suggesting repeated emergence of genetically distinct lineages in tomato production systems. The release of new varieties of plants that carry disease resistance genes can have dramatic effects on pathogen population structure due to selection to overcome host resistance [65,66], and we have previously reported on the loss of function of effector AvrXv3 (XopAF) in *X. perforans* populations, which has occurred across lineages [12,36,59]. We hypothesize that changes in prevalence and distribution of lineages are not exclusively due to selection on the pathogen population but rather are associated with the multinational structure of tomato breeding and seed production, in which parental lines and geographic locations of seed production change over time [67]. Our results indicate that *X. perforans* have diverged by mutation and acquisition of genetic material by horizontal gene transfer, which may be related to the geographic expansion of the population and exposure to regional communities of *Xanthomonas* and other plant-associated bacteria.

Type III secreted effectors are important members of both the core and accessory genomes given their role in pathogenicity and virulence. Approximately one-third of effector genes in our *X. perforans* sample showed allelic variation associated with core gene variation, indicating that some effectors evolve with the core genome and may reflect homologous recombination affecting the chromosome [30]. The impact of allelic variation in *X. perforans* effectors on pathogen fitness, if any, is unknown. Low frequency effectors were found across core gene clusters, suggesting acquisition of new effectors and their exchange among *X. perforans* lineages. For example, some strains in clusters 3, 4, and 5 from the United States, Canada, and Mexico carried the same alleles of low frequency effectors XopAQ and XopE3 as strains from Asia, Nigeria, and Italy. BLAST analysis suggested the geographically widespread presence of transcription activation-like (TAL) effectors in *X. perforans*. Both TAL effectors described in *X. perforans avrHah1* and *pthXp1*, are associated with increased disease severity on tomato [33,68]. Acquisition of Type III effectors could increase the fitness of *X. perforans* relative to other bacterial spot pathogens and cause more damaging disease outbreaks [32,34,59].

Examination of Type III effector content at a global scale puts variation previously observed in Florida into a larger context. XopAF was present in strains collected in the 1990s (cluster 1), but absent or non-functional in most strains from Florida, Indiana Ohio, and North Carolina, USA [8,10,12,36,59]. Here, we found that *xopAF* was intact in many strains from seed production areas, suggesting strong selection for loss of function in commercial fruit production. The effector XopJ4 is a possible resistance target [59], but Klein-Gordon et al. [8] reported that it was missing from 3.2% of Florida strains collected in 2017 and, here, we found that it was absent in one North Carolina and 20 Florida, USA strains. All strains collected outside the USA contained *xopJ4*, except for cluster 10 strains. *xopJ2*, a horizontally acquired effector that contributes to bacterial fitness in the field [32,34,59], is common in North America, particularly in cluster 2 and 6 strains, but absent or infrequently detected in *X. perforans* populations elsewhere. Type III secreted effectors comprise only a small part of the accessory genome of *X. perforans* but may be representative of divergence in the accessory genome. Changes to Type III effector

435 composition may be indicative of high-risk variants of *X. perforans* that could affect
436 epidemic dynamics and disease severity.
437 In summary, we found strong evidence for continued global dissemination of *X.*
438 *perforans*, consistent with the global nature of tomato breeding and hybrid tomato seed
439 production. We found notable diversity in the global population of *X. perforans*, including in
440 seed production regions, and multiple variants of *X. perforans* that are not yet globally
441 distributed. The genomic diversity of *X. perforans* in seed and fruit production regions
442 creates the opportunity for recombination among strains and subsequent dissemination of
443 high fitness variants of *X. perforans*.
444
445
446 **Materials and Methods**
447 Bacterial strains, genome sequencing, and assembly
448 *X. perforans* strains were collected from 13 different countries (Table 1; Extended
449 Data Table 1). Strains from the United States were collected from seven states between
450 1991 to 2016 and comprised 181 strains. The remaining 89 strains were collected from
451 Canada, Mexico, Brazil, Italy, Ethiopia, Nigeria, South Africa, Iran, China, Thailand,
452 Australia, and Vietnam. Strains from China, Thailand, and Vietnam were collected from
453 fields designated for production of tomato seed for the global market. Strains from Brazil
454 were obtained from both staked fresh-market and processing tomato commercial fields.
455 Strains from Italy were isolated from tomato pith in greenhouse tomato showing wilting
456 symptoms [69,70]. Strains from South Africa were collected from commercial seed lots. Strains
457 from Nigeria were obtained from fields cultivated for both subsistence and commercial
458 purposes. Strains from other countries were collected from fields designated for commercial
459 fruit production.
460 A total of 270 *X. perforans* genome sequences were used during this study. Draft and
461 whole genomes of several strains were generated and published previously [28,30,31,40,41,69,71].
462 The remaining strains were sequenced for this study using Illumina platforms, generating
463 2 x 250 or 2 x 150 base pair reads. Raw reads were assembled into contigs with Spades
464 v.3.11 with read error correction and "--careful" switch [72]. Reads were then aligned to the
465 assembled contigs using Bowtie 2 v. 2.3.3 [73] and inconsistencies identified by Bowtie2
466 removed using Pilon [74]. Contigs smaller than 500 bp and with less than 2.0 k-mer coverage
467 were filtered out. Assembled genomes were annotated using the IMG/JGI platform [75]. The
468 genomes generated for this study are available in NCBI BioProject PRJNA941448.
469
470 Core gene phylogeny
471 In a previous study, we defined a set of 1,356 'core genes' from 58 genomes of *X.*
472 *perforans* strains isolated from Florida [30]. The core genes were determined based on amino
473 acid sequence homology using GET_HOMOLOGUES software package [76]. We used the core
474 genes from a representative *X. perforans* genome, Xp91-118, as query to search the
475 remaining 269 genomes using local BLAST [77]. BLAST results were filtered using query
476 coverage and pairwise nucleotide sequence alignment thresholds of 70% each and the
477 sequence was checked for the presence of standard start and stop codons at either end of
478 the gene and gene was removed if both were not present. A total of 887 genes were found to
479 be intact in all 270 genomes. Genes were individually parsed and aligned using MAFFT [78]
480 and concatenated using sequence matrix [79]. The result was a circa 617 Kbp alignment,
481 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 [80]. 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 [47].

Population structure and global distribution

SNPs were extracted from core genes for hierarchical clustering based on Bayesian analysis of population structure (hierBAPS) algorithm [81], implemented in the 'rhierBAPS' R package v 1.0.1 [48,82]. For visualization, hierBAPS clusters were added to the phylogenetic tree generated from ClonalFrameML using the 'ggtree' package in R [83]. The treemap function in plotly [84] was used to show the relative distribution of clusters across geographic locations. R package 'ggplot2' was used to map hierBAPS clusters to countries [85].

A distance-based phylogenetic network was calculated using the neighbor-net method in Splitstree v.4.14.6 [86,87] from the ClonalFrameML output. The network was annotated in R using phangron [88] and ggplot2 packages.

The assembled genomes were used for calculating average nucleotide identity and pangenome analysis. Average nucleotide identity (ANIb) between strains was calculated using assembled genomes with Pyani version 0.2.10 [89]. The pangenome was estimated using Roary v3.12.0 [90] after annotation from Prokka v1.12 [91]. The gene presence absence matrix from Roary was used as input for generation of NMDS plots using the 'dplyr' and 'ggplot2' packages from tidyverse [85] and to generate gene accumulation curves for each cluster using package 'micropan' [92].

Bayesian analysis of *X. perforans* divergence times

Ancestral dates of divergence within *X. perforans* cluster 1 were inferred using the R package BactDating [93]. A multifasta alignment of core chromosomal single nucleotide polymorphisms (SNPs) between cluster 1 strains was generated using the Prok-SNPTree pipeline [94]. Briefly, the trimmed fastq reads were aligned to reference genome for strain 91-118 using bwa v0.7.17 [95] and the alignment file was sorted and converted to binary format with SAMtools v1.15 [96]. Strains with low coverage across more than 25% of the reference genome were removed. Variations from the reference were identified by HaplotypeCaller tool in GATK v4.1.9.0 [97]. All non-SNP and non-chromosomal variations were ignored; the remaining SNPs were then filtered as follows: depth ≥ 10, quality-normalized by depth ≥ 10, mapping quality ≥ 50, and variant allele fraction ≥ 0.1. Chromosome-level sequence alignment of all strains was generated from reference chromosome and high-quality SNPs using GATK FastaAlternateReferenceMaker. Gubbins v3.3.0 [98] was used to identify potentially recombinant regions in the alignment, and SNPs in these regions were filtered out before subsequent analyses. The nucleotide at non-recombinant core SNP positions for each strain was concatenated together to produce core SNP alignment. The alignment was used to infer a phylogenetic tree using the GTRCAT model in RAxML version 8.2.10 [99]. The inputs to the BactDating analysis were the maximum likelihood tree and dates of isolation assigned as dates of tips. Edge lengths were transformed from substitutions per site to total substitutions. Tree root was inferred using the 'initRoot' command. Strains that were outliers in the comparison of root-to-tip distance to collection year were removed from the alignment because these strains could contain undetected recombination or high variation in mutation rate. Dates of internal nodes were inferred using the 'bactdate' function on the re-rooted tree, using $1x10^6$ iterations. The analysis was run with strict and relaxed clock models and compared using DIC values. To confirm a temporal signal, the best model was

compared to an equivalent analysis with all tip dates forced equal. Nexus trees were exported and viewed in FigTree v. 1.4.4 [100].

Type III Secreted Effector Profile Analysis

A T3S effector database was curated using amino acid sequences of 66 *Xanthomonas* effectors [101] (Extended Data Table 4). When available, functional annotations were retrieved from NCBI and Pfam databases [102]. Orthologous sequences were identified with the software BLASTp [103,104], by querying the curated effectors database against the amino acid sequences of the annotated genomes of 270 *X. perforans* 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 [105]. 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 XopP-homolog (XopP-h).

Orthologous sequences from each effector were extracted from the annotated genomes, aligned with MAFFT [78], and allelic variants identified [106] 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' [82,107]. The results were displayed as a heatmap with the package 'gplots' and the function heatmap.2 [108].

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' [107]. 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' [109,110]. Based on the NMDS analysis, we assigned strains to effector clusters, which were plotted on a worldwide map with the packages 'ggplot2' and 'scatterpie' [110,111]. 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' [109,112-119]. Geographic coordinates (longitude, latitude) of countries and states (for USA) of collection were obtained with the R package 'googleway' [120] and the function 'mutate_geocode' from Google maps.

Putative TAL effectors were amplified using primers Hah1R-F (GACACAGGCCAACTTTTCAAGATTGC) and Hah1R-R (GATCAGGGCGAGATAACTGTGC), based on primers designed by A Strauß and T. Lahaye (personal communication), or PthXp1R-F (TTGGAGACAGGCCAACTTCT) and PthXp1R-R (GGATCAGGGCGAGATAACTG). Strains Xe95-2 and Xg444 containing the TAL effectors *avrBs3* and *avrHah1* were included as controls [121]. Strains produced bands consistent with *avrHah1* or an approximately 2 Kbp band in between typical band sizes for *avrBs3* and

579 *avrHah1*, representing *pthXp1*. From one strain with intermediate band size, 2P6S1, native
580 plasmid DNA was isolated using the alkaline lysis method [122]. *Eco*RI digested DNA of the
581 plasmid prep was ligated into vector pLAFR3 [123] restricted with the same enzyme for
582 transformation into *E. coli* DH5α. Clones containing the TAL effector were identified by
583 PCR and analyzed by restriction digest. One clone, designated as p7.1, contained an approx.
584 5 Kbp *Eco*RI fragment and was selected for further sequencing and phenotype testing. For
585 sequencing of the TAL repeat region, DNA of p7.1 was restricted with *Nsi*I and the internal
586 fragment was ligated into vector pBluescript restricted with *Pst*I. Additional pBluescript
587 subclones were made using *Bam*HI (~3 Kbp and ~1.1 Kbp) and *Bam*HI/*Eco*RI (~1 Kbp) in
588 order to cover the entire cloned region in p7.1. All clones were transformed into DH5α for
589 sequencing using vector primers T3 and T7.
590
591 **Acknowledgements**
592 We thank D. Ritchie and C. Mauney for providing reference strains and collecting disease
593 samples, respectively. Tina Simonton and Phyllis May provided assistance with strain
594 collection in Ontario. This research was supported by USDA NIFA award 2015-51181-
595 24312, Ontario Agri-Food Innovation Alliance, Ontario Tomato Research Institute, and the
596 Fresh Vegetable Growers of Ontario, the North Carolina Tomato Growers Association.
597 Plant diagnostic laboratories receive funding from USDA-NIFA through the National Plant
598 Diagnostic Network. FIB was supported by Conacyt Mexico. MK was supported by Swedish
599 International Development Cooperation Agency.
600
601 **Author Contributions**
602 ST, FIB, MOJ, PDR, GEV, JBJ, and EMG conceived the study. MOJ, GVM, PA, TBA, GC,
603 LBTdlB, EB, TCC, DTKT, TAC, DSE, RFG, DMF, MK, MLI, FJL, LL, ETM, SAM, NTTN, EO,
604 AMQD, RR, FR, GER, VMS, PT, CT, JBJ, and GEV collected bacterial strains. ST, MOJ,
605 GVM, and JKG prepared genomic libraries. ST, FIB, MOJ, JHT, A Sharma, A Subedi, and
606 EMG performed data analysis. ST, FIB, MOJ, A Sharma, A Subedi, GEV, JBJ, and EMG
607 wrote the manuscript. All authors edited the manuscript.
608
609 Sujan Timilsina*, Fernanda Iruegas-Bocardo*, Mustafa O. Jibrin*, Anuj Sharma, Aastha
610 Subedi, Gerald V. Minsavage, Jose Huguet-Tapia, Jeannie Klein-Gordon, Pragya Adhikari, Tika
611 B. Adhikari, Gabriella Cirvilleri, Laura Belen Tapia de la Barrera, Eduardo Bernal, Tom C.
612 Creswell, Doan Thi Kieu Tien, Teresa A. Coutinho, Daniel S. Egel, Rubén Félix-Gastélum,
613 David M. Francis, Misrak Kebede, Melanie Lewis Ivey, Frank J. Louws, Laixin Luo, Elizabeth
614 T. Maynard, Sally A. Miller, Nguyen Thi Thu Nga, Ebrahim Osdaghi, Alice M. Quezado-Duval,
615 Rebecca Roach, Francesca Rotondo, Gail E. Ruhl, Vou M. Shutt, Petcharat Thummabenjapone,
616 Cheryl Trueman, Pamela D. Roberts, Jeffrey B. Jones, Gary E. Vallad, Erica M. Goss
617
618
619 **Data summary**
620 The raw read files and genome assemblies for all strains sequenced in this study are
621 deposited in NCBI under bioproject PRJNA941448. The sources of genome assemblies
622 acquired from public databases are listed in Table 1.
623
624 **Figure Captions**
625
626 Figure 1. Phylogenetic relationships and genome-wide identity of *Xanthomonas perforans*
627 strains collected from tomato production regions worldwide. (A) Maximum likelihood

628 phylogenetic tree constructed with nucleotide sequences from 887 core genes, corrected for
629 recombination by ClonalFrameML, of 259 *X. perforans* strains. Tips are colored according
630 to clusters identified by hierBAPS. Cluster 10 strains were genetically distant and excluded
631 from the tree (see Extended Data Figure 1A). (B) Distribution of strains in each hierBAPS
632 cluster by country or state of collection. (C) Map showing distribution of clusters by country
633 of collection. Countries connected by lines show selected instances of genetically similar
634 strains collected in different countries. (D) Pairwise comparison of whole genome average
635 nucleotide identity of selected strains shows high identity between strains isolated from
636 different continents based on BLAST-based alignment with pyani v.0.2.10. For each
637 comparison, genome coverage is shown by grayscale in boxes, scale shown to the right.
638 Values for each comparison are for genomes in rows when compared to genomes in columns.
639 Additional data is shown in Extended Data Table 2.
640
641 Figure 2. Divergence of *Xanthomonas perforans* strains from 1990s collections in Florida
642 using well-supported SNPs. The time-calibrated phylogeny shows diversification of Florida
643 strains around the time of first isolation in 1991. Two strains from early collections, Xp894
644 and XV0938, diverged in the 1980s from a common ancestor of strains collected in Australia
645 and Africa in the following decades. All strains shown were assigned to cluster 1 based on
646 core genes. Estimated dates (year) are shown for major nodes with 95% confidence intervals
647 in brackets. Estimated nucleotide substitution rate was 1.19 changes per year [95%
648 confidence interval: 1.01, 1.36].
649
650
651 Figure 3. Variation in Type III secreted effectors in *Xanthomonas perforans*. Type III
652 secreted effectors are in columns and *X. perforans* strain in rows. Effector status is shown
653 by allele type: absence is indicated by allele type 0 (white), while the most frequent allele
654 observed when the effector is present is allele type 1 (purple), second most frequent is allele
655 type 2 (blue), and so on. Putative pseudogenized effectors are shown as allele 13 (gray). The
656 order of columns was determined by hierarchical clustering analysis, placing similarly
657 distributed effectors adjacent to each other. *X. perforans* strains (rows) are organized by
658 core gene cluster.
659
660 **Extended Data**
661
662 Extended Data Figure 1. Maximum likelihood phylogenetic tree of *Xanthomonas perforans* strains
663 based on aligned nucleotide sequences of 887 core genes. The tree was inferred using RAxML using
664 a GTRGAMMAI substitution model. The tree was arbitrarily rooted using the 11 genetically
665 diverged strains that make up core gene cluster 10.
666
667 Extended Data Figure 2. Recombination in core genes and variation in gene content in
668 *Xanthomonas perforans*. (A) Neighbor-net of 887 core genes from 259 *X. perforans* strains
669 (excludes the phylogenetically distant cluster 10 strains). Tips are colored by core gene cluster.
670 Reticulations in the network indicate conflicting phylogenetic relationships. (B) Visualization of
671 pangenome variation by non-metric multidimensional scaling of gene presence-absence for all 270
672 *X. perforans* strains by core gene cluster. Ellipses assume a multivariate t-distribution. (C)
673 Increase in gene count with increasing number of strains sampled. Clusters 1 and 2 were
674 represented by the most strains, but other clusters showed similar rates of increase in the
675 pangenome of the cluster.
676

Extended Data Figure 3. Frequency of Xop effectors found among the 270 *Xanthomonas perforans* strains examined. 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.

Extended Data Figure 4. Clustering of *Xanthomonas perforans* effector profiles by non-metric multidimensional scaling and distribution of resulting clusters among sampled regions. (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 (Supplementary Figure 6). (E) Distribution of strains by effector clusters among sampled countries.

Extended Data Figure 5. Variation in Type III secreted effector profiles in *Xanthomonas perforans* with strains (rows) ordered according to NMDS of effector profiles. Type III secreted effectors are in columns and *X. 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. Order of rows is based on NMDS clustering analysis of effector profiles (see Extended Data Figure 5A).

Extended Data Table 1. A. BAPS cluster assignment for *Xanthomonas perforans* strains along with NCBI BioProject for each genome and associated statistics for newly sequenced strains. B. Depth of coverage relevant to ANI comparisons in Fig. 1D and phylogenetic analysis in Figure 2.

Extended Data Table 2. Average nucleotide identity (ANIb) comparisons between strains with highly similar core gene sequences collected across continents.

Extended Data Table 3. Type III secreted effectors (Xop proteins) found in the 270 *X. perforans* assembled genomes based on BLAST searches using reference sequences in Extended Data Table 4. Last column shows the results of TAL effector BLAST search. "-" indicates no hit. Putative TAL was inferred in a small sample of strains based on band size from PCR amplification of repeat regions. Asterisk indicates pthXp1 gene confirmed by cloning and sequencing entire gene.

Extended Data Table 4. Type III secreted effector database used to query assembled genomes for effector genes.

References Cited

1 FAO. *FAOSTAT Database*, <http://www.fao.org/faostat/> (2022).

2 Modor Intelligence. *Tomato Seeds Market - Growth, Trends, COVID-10 Impact, and Forecasts (2022 - 2027)*, <Available at: https://www.mordorintelligence.com/industry-reports/tomato-seeds-market [Accessed August 31, 2022]> (2022).

3 Abrahamian, P., Sharma, A., Jones, J. B. & Vallad, G. E. Dynamics and spread of bacterial spot epidemics in tomato transplants grown for field production. *Plant Dis.* **105**, 566-575 (2021).

4 Osdaghi, E. *et al.* A centenary for bacterial spot of tomato and pepper. *Mol. Plant Pathol.* **22**, 1500 (2021).

5 Potnis, N. Harnessing eco-evolutionary dynamics of Xanthomonads on tomato and pepper to tackle new problems of an old disease. *Annu. Rev. Phytopathol.* **59**, 289-310 (2021).

6 Jibrin, M. O. *et al.* Bacterial spot of tomato and pepper in Africa: Diversity, emergence of T5 race, and management. *Frontiers in Microbiology* **13**, 835647, doi:10.3389/fmicb.2022.835647 (2022).

7 Abbasi, P. A., Khabbaz, S. E., Weselowski, B. & Zhang, L. Occurrence of copper-resistant strains and a shift in *Xanthomonas* spp. causing tomato bacterial spot in Ontario. *Canadian Journal of Microbiology* **61**, 753-761, doi:10.1139/cjm-2015-0228 %M 26308592 (2015).

8 Klein-Gordon, J. M. *et al.* Assessing changes and associations in the *Xanthomonas perforans* population across Florida commercial tomato fields via a statewide survey. *Phytopathology* **111**, 1029-1041, doi:10.1094/PHYTO-09-20-0402-R (2021).

9 Lewis Ivey, M. L., Strayer, A., Sidhu, J. K. & Minsavage, G. V. Bacterial leaf spot of tomato (*Solanum lycopersicum*) in Louisiana is caused by *Xanthomonas perforans*, tomato race 4. *Plant Dis.* **100**, 1233-1233, doi:10.1094/PDIS-12-15-1451-PDN (2016).

10 Egel, D. S. *et al.* Distribution and characterization of *Xanthomonas* strains causing bacterial spot of tomato in Indiana. *Plant Health Progress* **19**, 319-321, doi:10.1094/php-07-18-0041-br (2018).

11 Abrahamian, P., Jones, J. B. & Vallad, G. E. Efficacy of copper and copper alternatives for management of bacterial spot on tomato under transplant and field production. *Crop Protect.* **126**, 104919, doi:https://doi.org/10.1016/j.cropro.2019.104919 (2019).

12 Adhikari, P. *et al.* Phenotypic and genetic diversity of *Xanthomonas perforans* populations from tomato in North Carolina. *Phytopathology®* **109**, 1533-1543, doi:10.1094/phyto-01-19-0019-r (2019).

13 Rotondo, F. *et al.* Shifts in *Xanthomonas* spp. causing bacterial spot in processing tomato in the Midwest of the United States. *Canadian Journal of Plant Pathology*, 1-16, doi:10.1080/07060661.2022.2047788 (2022).

14 Khanal, S., Hind, S. R. & Babadoost, M. Occurrence of copper-resistant *Xanthomonas perforans* and *X. gardneri* in Illinois tomato fields. *Plant Health Progress* **21**, 338-344, doi:10.1094/PHP-06-20-0048-RS (2020).

15 Jones, J. B. *et al.* A third tomato race of *Xanthomonas campestris* pv. *vesicatoria*. *Plant Dis.* **79**, 395-398 (1995).

16 Jones, J. B., Lacy, G. H., Bouzar, H., Stall, R. E. & Schaad, N. W. Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. *Syst. Appl. Microbiol.* **27**, 755-762, doi:http://dx.doi.org/10.1078/0723202042369884 (2004).

17 Constantin, E. C. *et al.* Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathol.* **65**, 792-806, doi:10.1111/ppa.12461 (2016).

18 Roach, R., Mann, R., Gambley, C. G., Shivas, R. G. & Rodoni, B. Identification of *Xanthomonas* species associated with bacterial leaf spot of tomato, capsicum and chilli crops in eastern Australia. *Eur. J. Plant Pathol.* **150**, 595-608, doi:10.1007/s10658-017-1303-9 (2018).

19 Burlakoti, R. R., Hsu, C.-F., Chen, J.-R. & Wang, J.-F. Population dynamics of xanthomonads associated with bacterial spot of tomato and pepper during 27 years across Taiwan. *Plant Dis.* **102**, 1348-1356, doi:10.1094/PDIS-04-17-0465-RE (2018).

20 Kebede, M. *et al.* Molecular characterization of *Xanthomonas* strains responsible for bacterial spot of tomato in Ethiopia. *Eur. J. Plant Pathol.* **140**, 677-688, doi:10.1007/s10658-014-0497-3 (2014).

21 Potnis, N. *et al.* Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Mol. Plant Pathol.* **16**, 907-920, doi:10.1111/mpp.12244 (2015).

22 Timilsina, S. *et al.* Multilocus sequence analysis of xanthomonads causing bacterial spot of tomato and pepper reveals strains generated by recombination among species and recent global spread of *Xanthomonas gardneri*. *Appl. Environ. Microbiol.* **81**, 1520-1529 (2015).

23 Hamza, A. A. *et al.* Genetic and pathological diversity among *Xanthomonas* strains responsible for bacterial spot on tomato and pepper in the Southwest Indian Ocean region. *Plant Dis* **94**, 993-999, doi:10.1094/PDIS-94-8-0993 (2010).

24 Araújo, E. R., Costa, J. R., Ferreira, M. A. S. V. & Quezado-Duval, A. M. Widespread distribution of *Xanthomonas perforans* and limited presence of *X. gardneri* in Brazil. *Plant Pathol.* **66**, 159-168, doi:https://doi.org/10.1111/ppa.12543 (2017).

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

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

27 Osdaghi, E., Taghavi, S. M., Koebnik, R. & Lamichhane, J. R. Multilocus sequence analysis reveals a novel phylogroup of *Xanthomonas euvesicatoria* pv. *perforans* causing bacterial spot of tomato in Iran. *Plant Pathol.* **67**, 1601-1611, doi:https://doi.org/10.1111/ppa.12864 (2018).

28 Abrahamian, P. *et al.* Molecular epidemiology of *Xanthomonas perforans* outbreaks in tomato plants from transplant to field as determined by single-nucleotide polymorphism analysis. *Appl. Environ. Microbiol.* **85**, e01220-01219, doi:10.1128/aem.01220-19 (2019).

29 Hert, A. *et al.* Relative importance of bacteriocin-like genes in antagonism of *Xanthomonas perforans* tomato race 3 to *Xanthomonas euvesicatoria* tomato race 1 strains. *Appl. Environ. Microbiol.* **71**, 3581-3588 (2005).

30 Timilsina, S. *et al.* Multiple recombination events drive the current genetic structure of *Xanthomonas perforans* in Florida. *Frontiers in Microbiology* **10**, 448 (2019).

31 Jibrin, M. O. *et al.* Genomic inference of recombination-mediated evolution in *Xanthomonas euvesicatoria* and *X. perforans*. *Appl. Environ. Microbiol.* **84**, e00136-00118, doi:10.1128/aem.00136-18 (2018).
32 Abrahamian, P. *et al.* The type III effector AvrBsT enhances *Xanthomonas perforans* fitness in field-grown tomato. *Phytopathology* **108**, 1355-1362, doi:10.1094/PHYTO-02-18-0052-R (2018).
33 Newberry, E. A. *et al.* Independent evolution with the gene flux originating from multiple *Xanthomonas* species explains genomic heterogeneity in *Xanthomonas perforans*. *Appl Environ Microbiol* **85**, e00885-00819, doi:10.1128/aem.00885-19 (2019).
34 Sharma, A. *et al.* Need for speed: Bacterial effector XopJ2 is associated with increased dispersal velocity of *Xanthomonas perforans*. *Environ. Microbiol.* **23**, 5850–5865 (2021).
35 Klein-Gordon, J. *et al.* Strength in numbers: Density-dependent volatile-induced antimicrobial activity by *Xanthomonas perforans*. *Phytopathology* **113**, 160-169 (2022).
36 Bernal, E. *et al.* Migration drives the replacement of *Xanthomonas perforans* races in the absence of widely deployed resistance. *Frontiers in Microbiology* **13**, 826386, doi:10.3389/fmicb.2022.826386 (2022).
37 Klein-Gordon, J. M. *et al.* Whole genome sequences reveal the *Xanthomonas perforans* population is shaped by the tomato production system. *The ISME Journal* **16**, 591-601, doi:10.1038/s41396-021-01104-8 (2022).
38 Richard, D. *et al.* Complete genome sequences of six copper-resistant *Xanthomonas* strains causing bacterial spot of Solaneous plants, belonging to *X. gardneri*, *X. euvesicatoria*, and *X. vesicatoria*, using long-read technology. *Genome Announcements* **5**, e01693-01616, doi:10.1128/genomeA.01693-16 (2017).
39 Timilsina, S. *et al. Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology* **18**, 1-13, doi:10.1038/s41579-020-0361-8 (2020).
40 Roach, R. *et al.* Genomic sequence analysis reveals diversity of Australian *Xanthomonas* species associated with bacterial leaf spot of tomato, capsicum and chilli. *BMC Genomics* **20**, 310, doi:10.1186/s12864-019-5600-x (2019).
41 Schwartz, A. R. *et al.* Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Frontiers in Microbiology* **6**, doi:10.3389/fmicb.2015.00535 (2015).
42 Horvath, D. M. *et al.* Transgenic resistance confers effective field level control of bacterial spot disease in tomato. *PLoS ONE* **7**, e42036 (2012).
43 Subedi, A., Minsavage, J., Jones, J. B., Goss, E. & Roberts, P. Exploring diversity of bacterial spot associated *Xanthomonas* population of pepper in Southwest Florida. *Plant Dis.*, doi:10.1094/PDIS-10-22-2484-RE (2023).
44 Cesbron, S. *et al.* Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Frontiers in Plant Science* **6**, 1126 (2015).
45 McCann, H. C. Skirmish or war: the emergence of agricultural plant pathogens. *Curr. Opin. Plant Biol.* **56**, 147-152, doi:https://doi.org/10.1016/j.pbi.2020.06.003 (2020).
46 Strayer, A. L. *et al.* A multiplex real-time pcr assay differentiates four *Xanthomonas* species associated with bacterial spot of tomato. *Plant Dis.* **100**, 1660-1668, doi:10.1094/PDIS-09-15-1085-RE (2016).

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

48 Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Research* **3**, 93, doi:https://doi.org/10.12688/wellcomeopenres.14694.1 (2019).

49 Stall, R. E., Jones, J. B. & Minsavage, G. V. Durability of resistance in tomato and pepper to xanthomonads causing bacterial spot. *Annu. Rev. Phytopathol.* **47**, 265-284, doi:10.1146/annurev-phyto-080508-081752 (2009).

50 Sharlach, M. *et al.* Fine genetic mapping of *RXopJ4*, a bacterial spot disease resistance locus from *Solanum pennellii* LA716. *Theor. Appl. Genet.* **126**, 601-609, doi:10.1007/s00122-012-2004-6 (2013).

51 Gitaitis, R. & Walcott, R. The epidemiology and management of seedborne bacterial diseases. *Annu. Rev. Phytopathol.* **45**, 371-397, doi:10.1146/annurev.phyto.45.062806.094321 (2007).

52 Munkvold, G. P. Seed pathology progress in academia and industry. *Annu. Rev. Phytopathol.* **47**, 285-311, doi:10.1146/annurev-phyto-080508-081916 (2009).

53 Bank, W. *United States Seed; vegetable seed, of a kind used for sowing imports by country in 2019*, <https://wits.worldbank.org/trade/comtrade/en/country/USA/year/2019/tradeflow/Imports/partner/ALL/product/120991> (2019).

54 Jones, J. B., Pohronezny, K. L., Stall, R. E. & Jones, J. P. *Xanthomonas campestris* pv. *vesicatoria* in Florida on tomato crop residue, weeds, seeds, and voluteer tomato plants. *Phytopathology* **76**, 430-434 (1986).

55 Jones, J. *et al.* A third tomato race of *Xanthomonas campestris* pv. *vesicatoria*. *Plant Dis.* **79**, 395-398 (1995).

56 Jones, J. B. *et al.* Systematic analysis of xanthomonads (*Xanthomonas* spp.) associated with pepper and tomato lesions. *Int J Syst Evol Microbiol* **50 Pt 3**, 1211-1219, doi:10.1099/00207713-50-3-1211 (2000).

57 Dutta, B. *et al.* Role of Blossom Colonization in Pepper Seed Infestation by Xanthomonas euvesicatoria. *Phytopathology®* **104**, 232-239, doi:10.1094/PHYTO-05-13-0138-R (2013).

58 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.* **67**, 1168-1176, doi:10.1111/ppa.12839 (2018).

59 Timilsina, S. *et al.* Analysis of sequenced genomes of *Xanthomonas perforans* identifies candidate targets for resistance breeding in tomato. *Phytopathology* **106**, 1097-1104, doi:10.1094/PHYTO-03-16-0119-FI (2016).

60 Iruegas-Bocardo, F. *et al.* Whole genome sequencing-based tracing of a 2022 introduction and outbreak of *Xanthomonas hortorum* pv. *pelargonii*. *Phytopathology* **113**, 975-984, doi:10.1094/PHYTO-09-22-0321-R (2022).

61 Roman-Reyna, V. *et al.* Live tracking of a plant pathogen outbreak reveals rapid and successive, multidecade plasmid reduction. *mSystems* (2024).

62 Klein-Gordon, J. M. *et al.* Whole genome sequences reveal the Xanthomonas perforans population is shaped by the tomato production system. *The ISME journal* **16**, 591-601 (2022).

63 Gedgaew, C., Simaraks, S. & Rambo, A. T. Trends in hybrid tomato seed production under contract farming in Northeast Thailand. *Southeast Asian Studies* **6**, 339-333 (2017).
64 Rosset, P., Rice, R. & Watts, M. Thailand and the world tomato: globalization, new agricultural countries (NACs) and the agrarian question. *The International Journal of Sociology of Agriculture and Food* **8**, 71-94 (1999).
65 Quibod, I. L. *et al.* The Green Revolution shaped the population structure of the rice pathogen *Xanthomonas oryzae* pv. *oryzae*. *The ISME Journal* **14**, 492-505, doi:10.1038/s41396-019-0545-2 (2020).
66 McDonald, B. A. & Linde, C. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* **40**, 349-379 (2002).
67 Kloppenburg, J. R. *First the seed: The political economy of plant biotechnology, 1492-2000 (2nd edition)*. (University of Wisconsin Press, 2004).
68 Subedi, A. *et al.* Population genomics reveals an emerging lineage of *Xanthomonas perforans* on pepper. *Phytopathology*, doi:10.1094/PHYTO-04-23-0128-R (2023).
69 Torelli, E., Aiello, D., Polizzi, G., Firrao, G. & Cirvilleri, G. Draft genome of a *Xanthomonas perforans* strain associated with pith necrosis. *FEMS Microbiol. Lett.* **362**, doi:10.1093/femsle/fnv001 (2015).
70 Aiello, D. *et al.* A pith necrosis caused by *Xanthomonas perforans* on tomato plants. *Eur. J. Plant Pathol.* **137**, 29-41, doi:10.1007/s10658-013-0214-7 (2013).
71 Potnis, N. *et al.* Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genomics* **12**, 146 (2011).
72 Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and Its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).
73 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
74 Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963, doi:10.1371/journal.pone.0112963 (2014).
75 Markowitz, V. M. *et al.* IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115-D122 (2012).
76 Contreras-Moreira, B. & Vinuesa, P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* **79**, 7696-7701, doi:10.1128/AEM.02411-13 (2013).
77 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, doi:10.1016/s0022-2836(05)80360-2 (1990).
78 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
79 Vaidya, G., Lohman, D. J. & Meier, R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* **27**, 171-180, doi:https://doi.org/10.1111/j.1096-0031.2010.00329.x (2011).
80 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:https://doi.org/10.1093/bioinformatics/btu033 (2014).

81 Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M. & Corander, J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* **30**, 1224-1228, doi:10.1093/molbev/mst028 (2013).
82 R Core Team. R: A language and environment for statistical computing (R Foundation for Statistical Computing https://www.R-project.org/, Vienna, Austria, 2019).
83 Yu, G. Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics* **69**, e96, doi:https://doi.org/10.1002/cpbi.96 (2020).
84 Sievert, C. *Interactive web-based data visualization with R, plotly, and shiny*. (CRC Press, 2020).
85 Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686, doi:doi: 10.21105/joss.01686 (2019).
86 Huson, D. H. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68-73, doi:10.1093/bioinformatics/14.1.68 (1998).
87 Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254-267, doi:10.1093/molbev/msj030 (2006).
88 Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592-593 (2011).
89 Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G. & Toth, I. K. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* **8**, 12-24, doi:10.1039/C5AY02550H (2016).
90 Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691-3693, doi:10.1093/bioinformatics/btv421 (2015).
91 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).
92 Snipen, L. & Liland, K. H. micropan: an R-package for microbial pan-genomics. *BMC Bioinformatics* **16**, 79, doi:10.1186/s12859-015-0517-0 (2015).
93 Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R. & Wilson, D. J. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* **46**, e134-e134, doi:10.1093/nar/gky783 (2018).
94 Sharma, A. Prok-SNPTree: phylogenetic tree from next-gen sequencing of prokaryotes (v0.2b) (Zenodo, 2022).
95 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
96 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008, doi:10.1093/gigascience/giab008 (2021).
97 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498, doi:10.1038/ng.806 (2011).
98 Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15-e15 (2015).
99 Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinforma* **22**, doi:10.1093/bioinformatics/btl446 (2006).
100 Rambaut, A. FigTree 1.4.0 (http://tree.bio.ed.ac.uk/, 2012).
101 Costa, J. *et al.* A community-curated DokuWiki resource on diagnostics, diversity, pathogenicity and genetic control of xanthomonads. *Mol. Plant-Microbe Interact.*, doi:10.1094/MPMI-11-23-0184-FI (2023).

102 Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **38**, D211-D222, doi:10.1093/nar/gkp985 (2010).
103 Boratyn, G. M. *et al.* BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* **41**, W29-W33, doi:10.1093/nar/gkt282 (2013).
104 Altschul, S. F. BLAST Algorithm. *eLS*, doi:https://doi.org/10.1002/9780470015902.a0005253.pub2 (2014).
105 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010).
106 Takebayashi, N. uniqueHaplo.pl, http://raven.wrrb.uaf.edu/~ntakebay/teaching/programming/perl-scripts/uniqHaplo.pl, accessed 4-01-23 (2015).
107 Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.5-6. . https://CRAN.R-project.org/package=vegan. (2019).
108 Warnes, G. *et al.* gplots: Various R Programming Tools for Plotting Data. R package version 3.1.3, <https://CRAN.R-project.org/package=gplots> (2020).
109 Slowikowski, K. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.1, <https://CRAN.R-project.org/package=ggrepel> (2020).
110 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. 2 edn, (Springer-Verlag New York, 2016).
111 Yu, G. scatterpie: Scatter Pie Plot. R package version 0.1.8, <https://CRAN.R-project.org/package=scatterpie>. (2020).
112 Wilke, C. cowplot: Streamlined Plot Theme and Plot Annotations for 'gglot2'. R package version 1.0.0, <https://cran.r-project.org/package=cowplot> (2019).
113 Dunnington, D., Thorne, B. & Hernangómez, D. ggspatial: Spatial Data Framework for ggplot2. R package v. 1.1.1, <https://cran.r-project.org/package=ggspatial>. (2020).
114 Pebesma, E. Simple features for R: Standardized support for spatial vector data. *The R Journal* **10**, 439–446, doi:10.32614/RJ-2018-009 (2018).
115 Bivand, R. & Rundel, C. rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package v. 0.5-1, https://cran.r-project.org/package=rgeos (2019).
116 Elff, M. memisc: Management of Survey Data and Presentation of Analysis Results. R package version 0.99.27, <https://cran.r-project.org/package=memisc> (2020).
117 Venables, B. & Hornik, K. oz: Plot the Australian Coastline and States. R package version 1.0, <https://cran.r-project.org/package=oz> (2016).
118 Bivand, R. & Lewin-Koh, N. maptools: Tools for Handling Spatial Objects. R package version 1.0. https://CRAN.R-project.org/package=maptools (2020).
119 South, A. rnaturalearth: World Map Data from Natural Earth. R package version 0.1.0, <https://cran.r-project.org/package=rnaturalearth> (2017).
120 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).
121 Schornack, S., Minsavage, G. V., Stall, R. E., Jones, J. B. & Lahaye, T. Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *New Phytol.* **179**, 546-556 (2008).
122 Minsavage, G. V., Canteros, B. I. & Stall, R. E. Plasmid- mediated resistance to streptomycin in *Xanthomonas campestris* pv. *vesicatoria*. *Phytopathology* **80**, 719-723 (1990).

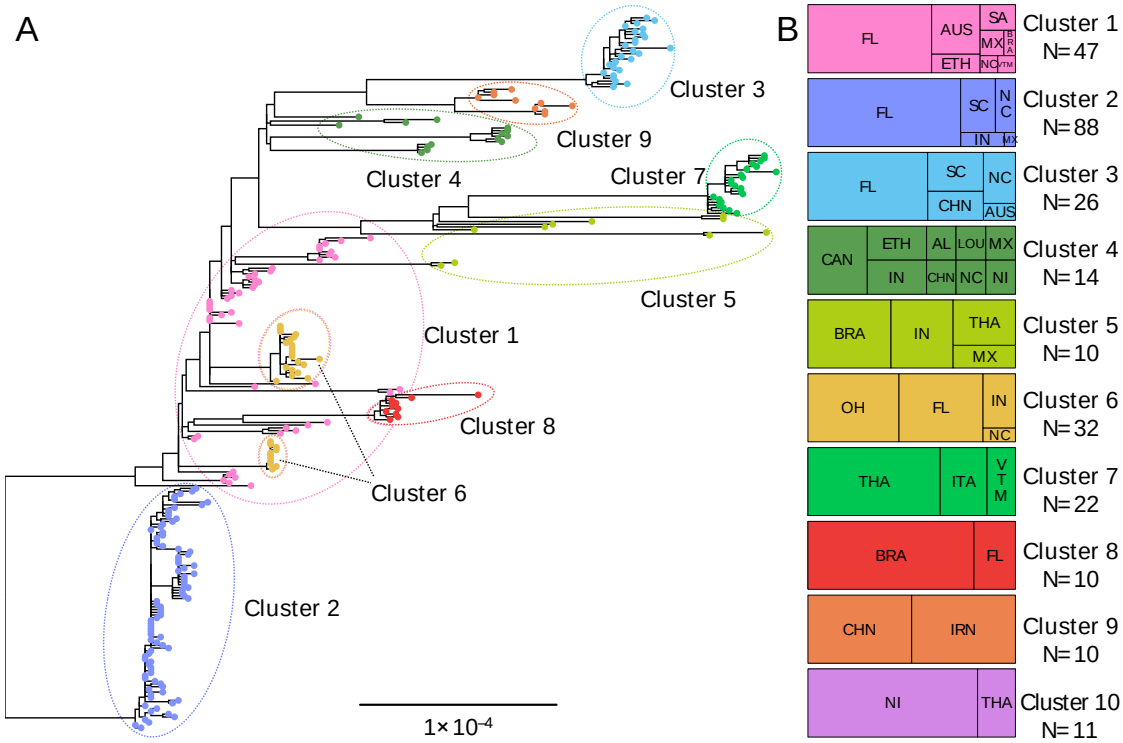
123 Staskawicz, B. J., Dahlbeck, D., Keen, N. & Napoli, C. Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*. *J. Bacteriol.* **169**, 5789-5794 (1987).

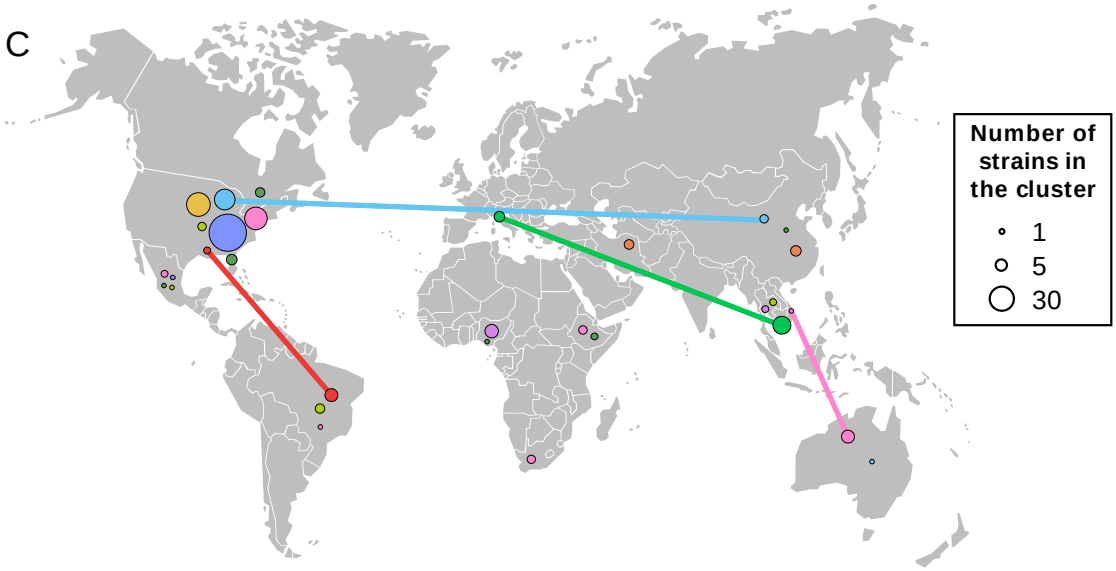
124 Jibrin, M. O. *et al.* First report of atypical *Xanthomonas euvesicatoria* strains causing bacterial spot of tomato in Nigeria. *Plant Dis.* **99**, 415, doi: https://doi.org/10.1094/PDIS-09-14-0952-PDN (2015).

1051 Table 1. *Xanthomonas* strains used in this study.

Country	Locality	Year	Strain (original name, if applicable)
Australia [40]	Queensland	2015	Aus3, Aus7, Aus14
		2016	Aus5, Aus10, Aus11
		2017	Aus1, Aus15, Aus16
Brazil [24]	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 [20]		2011	ETH5, ETH11, ETH21, ETH25, ETH33
Iran [26]		2013	K41, F210, F215, TOM801, TOM816
Italy [70]		2011	1P6S1, 2P4S1, 2P4S1D, 2P6S1, 1P4S1D
Mexico			Mexico-1, Mexico-3, Mexico-LT1, Mexico-LT3, Mexico-LT5
Nigeria [22,124]		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 [59]	1996	Xp1861
	Indiana [10]	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 [15,16,28,30,41,71]	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 [9]	2013	mli-2
	North Carolina [12]	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 [28]	2016	GEV2407, GEV2408, GEV2384, GEV2388, GEV2389, GEV2390, GEV2391, GEV2392, GEV2393, GEV2396, GEV2397, GEV2399, GEV2400, GEV2403, GEV2410, GEV2420
	Ohio [36]	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





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

Coverage
- 100%
- 98%
- 96%
- 94%
- 92%
- 90%

1052

1053 Figure 1. Phylogenetic relationships and average nucleotide identity of *Xanthomonas*
1054 *perforans* strains collected from tomato production regions worldwide. (A) Maximum
1055 likelihood phylogenetic tree constructed with nucleotide sequences from 887 core genes,
1056 corrected for recombination by ClonalFrameML, of 259 *X. perforans* strains. Tips are
1057 colored according to clusters identified by hierBAPS. Cluster 10 strains were genetically
1058 distant and excluded from the tree (see Extended Data Figure 1A). (B) Distribution of
1059 strains in each hierBAPS cluster by country or state of collection. (C) Map showing
1060 distribution of clusters by country of collection. Countries connected by lines show selected
1061 instances of genetically similar strains collected in different countries. (D) Pairwise
1062 comparison of whole genome average nucleotide identity of selected strains shows high
1063 identity between strains isolated from different continents based on BLAST-based
1064 alignment with pyani v.0.2.10. For each comparison, genome coverage is shown by
1065 grayscale in boxes, scale shown to the right. Values for each comparison are for genomes in
1066 rows when compared to genomes in columns. Additional data is shown in Extended Data
1067 Table 2.
1068

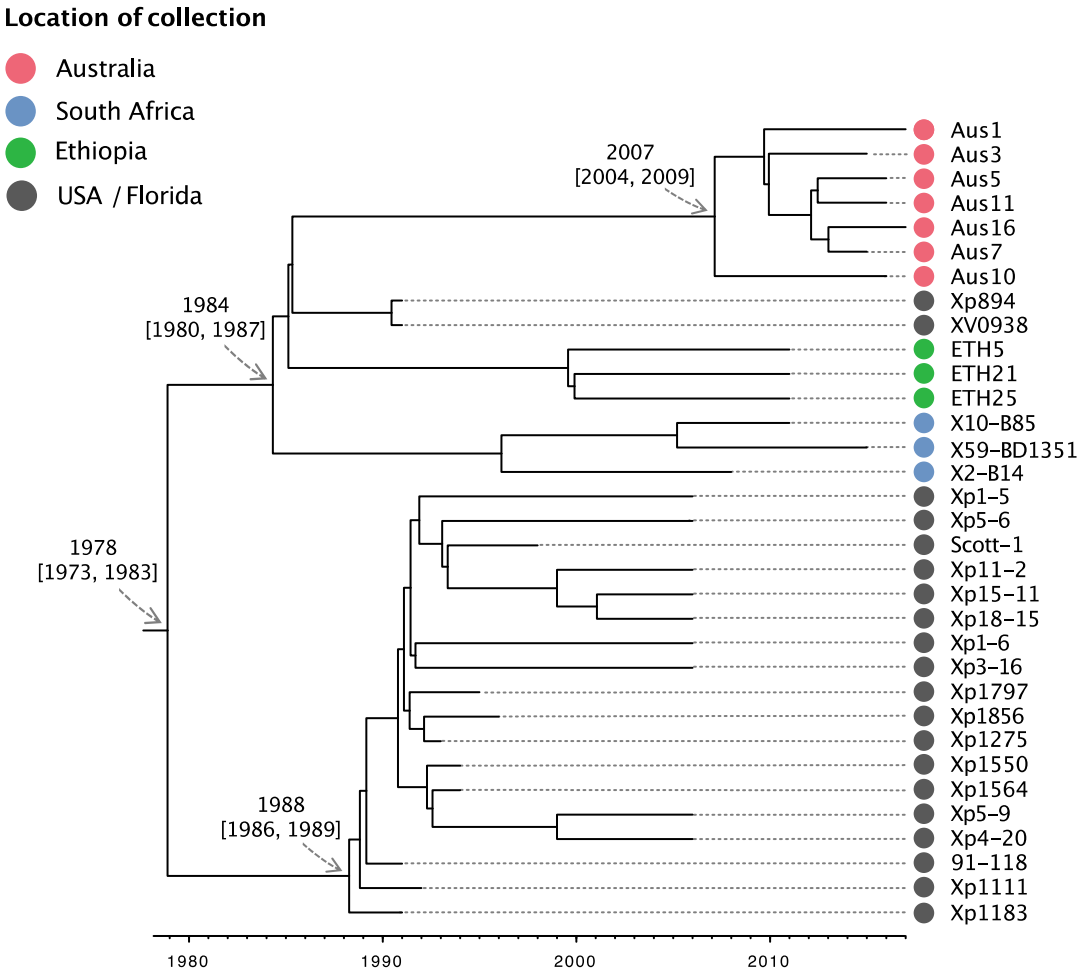


Figure 2. *Xanthomonas perforans* strains from 1990s collections in Florida show genetic variation in early population. The time-calibrated phylogeny using well-supported SNPs shows genetically distinct strains in Florida tomato fields in 1991 and subsequent diversification. The common ancestor of early Florida strains and later collections in Australia and Africa was dated to approximately 1978. Estimated dates (year) are shown for major nodes with 95% confidence intervals in brackets. All strains shown were assigned to cluster 1 based on core genes. Estimated nucleotide substitution rate was 1.19 changes per year [95% confidence interval: 1.01, 1.36].

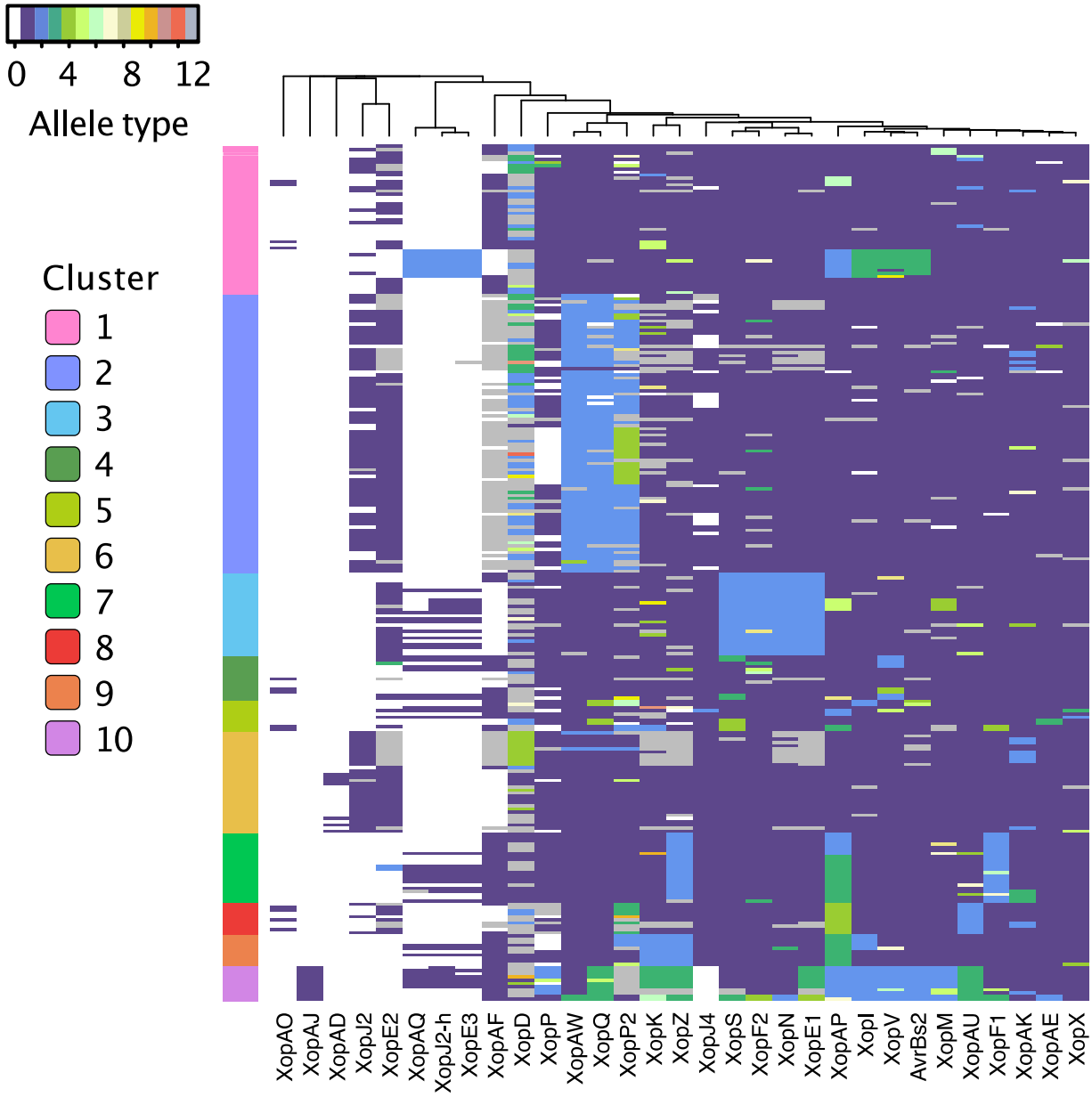
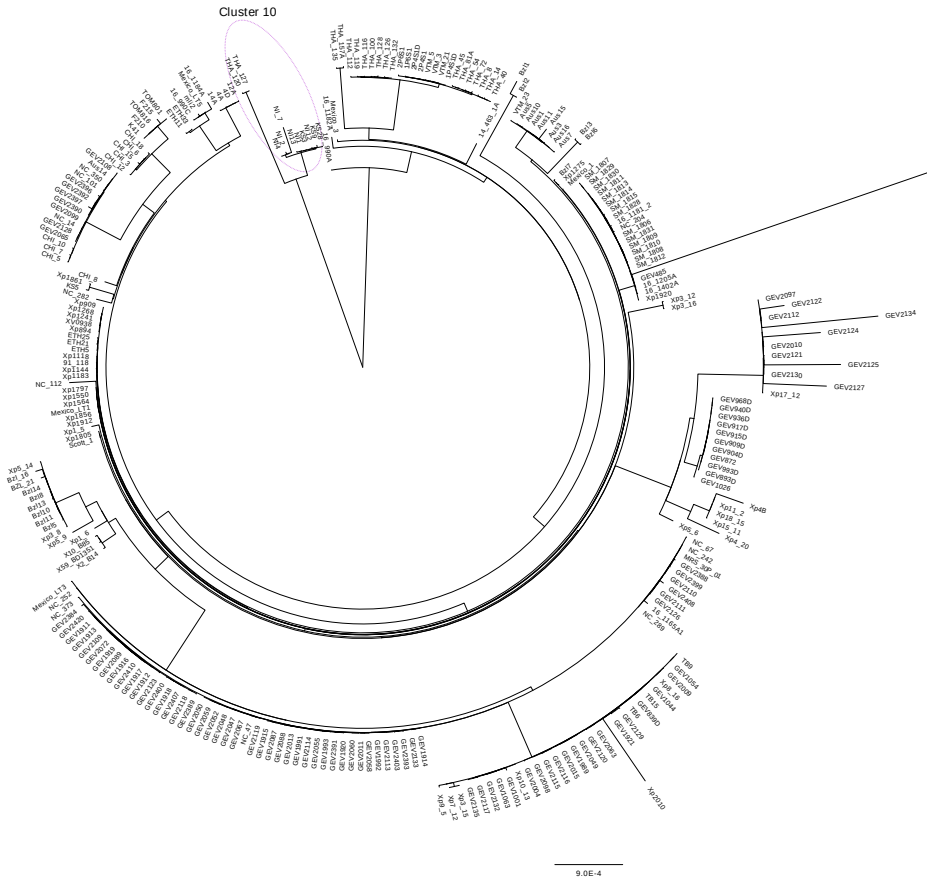
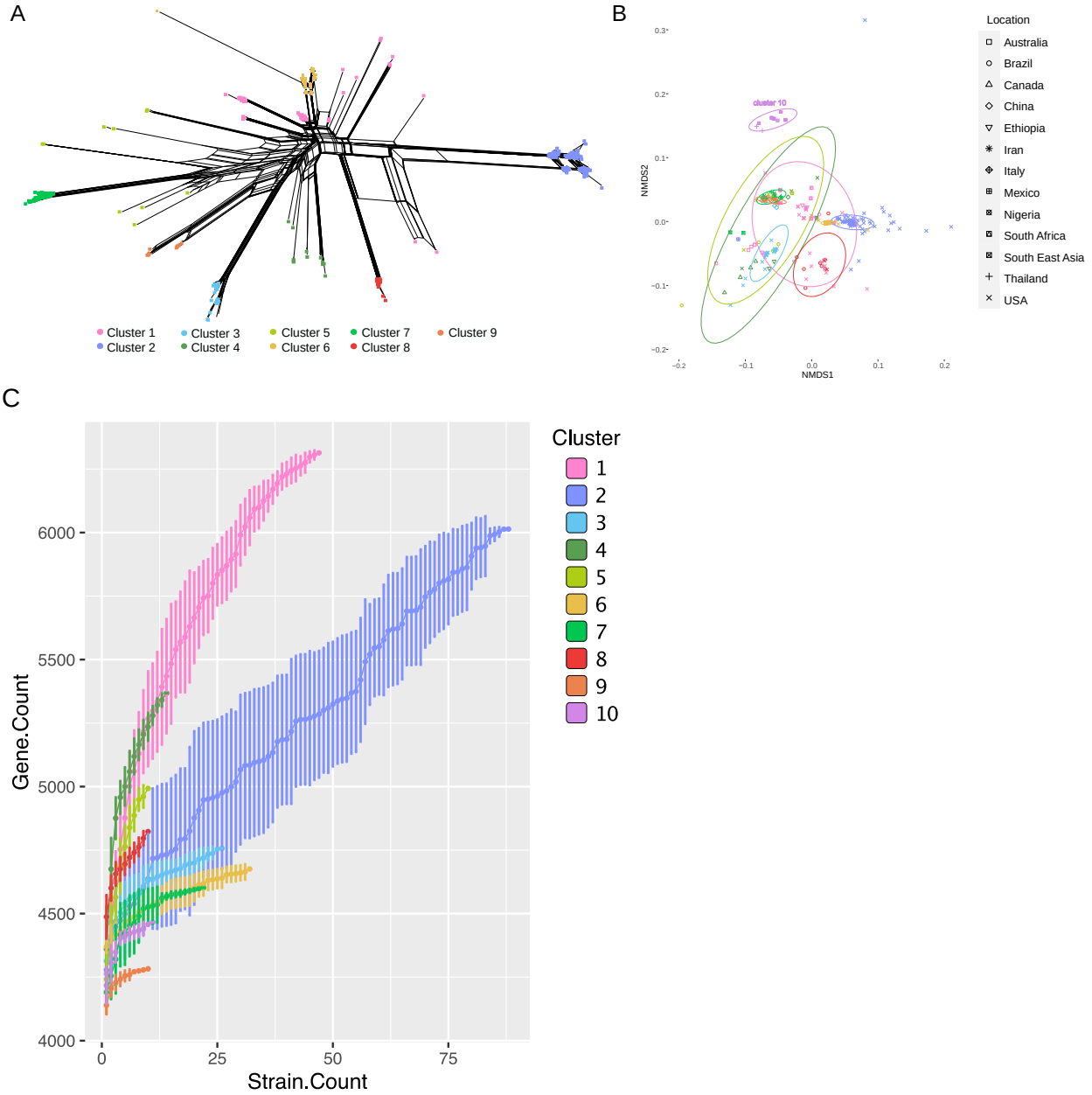


Figure 3. Type III secreted effectors in *Xanthomonas perforans* show variation in presence/absence and amino acid sequence. Type III secreted effectors are in columns and *X. perforans* strain in rows, organized by core gene cluster, which shows allelic variation in effectors is associated with core gene variation. 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.

Extended Data

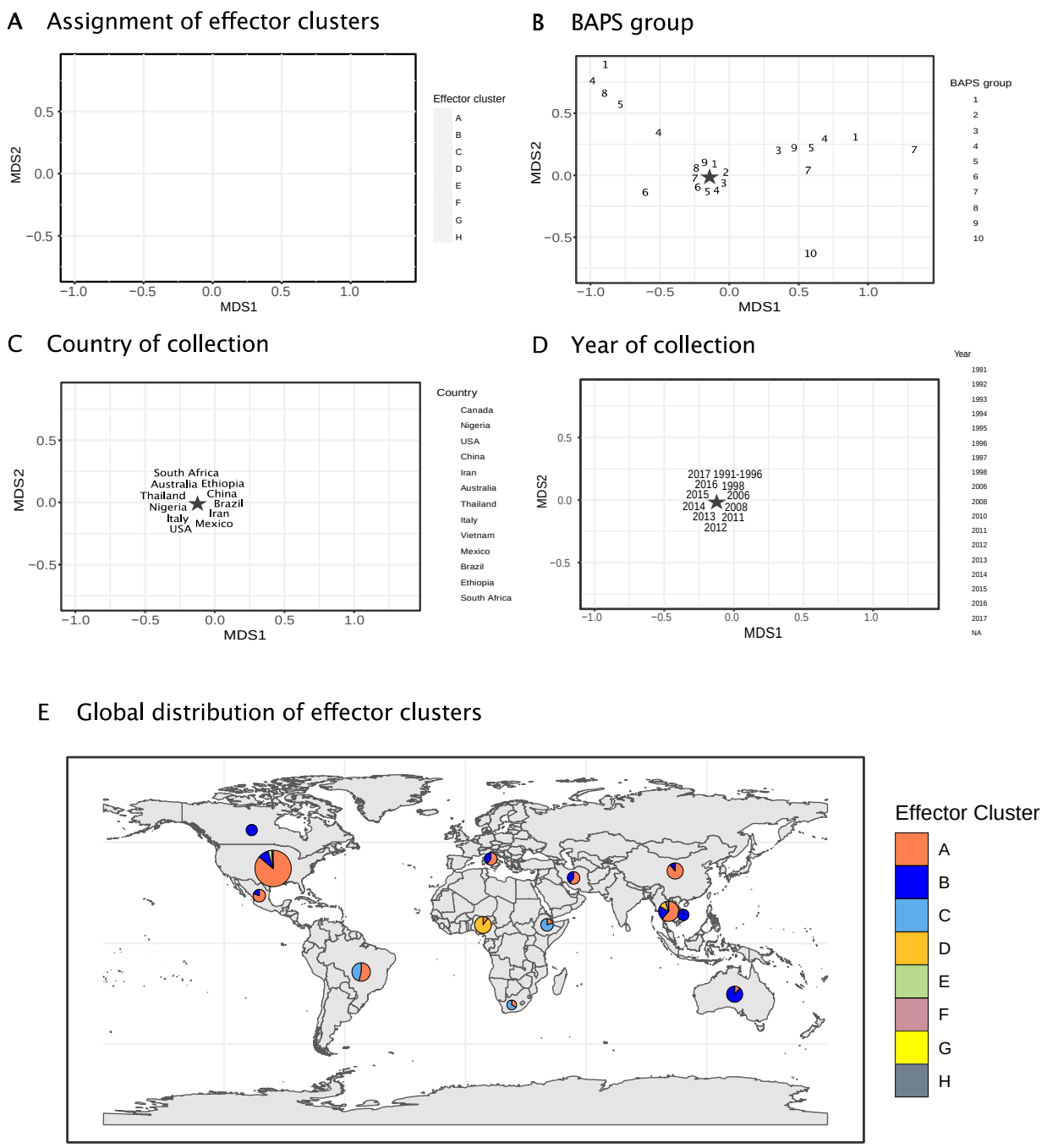


Extended Data Figure 1. Maximum likelihood phylogenetic tree of *Xanthomonas perforans* strains based on aligned nucleotide sequences of 887 core genes. The tree was inferred using RAxML using a GTRGAMMAI substitution model. The tree was arbitrarily rooted using the 11 genetically diverged strains that make up core gene cluster 10.

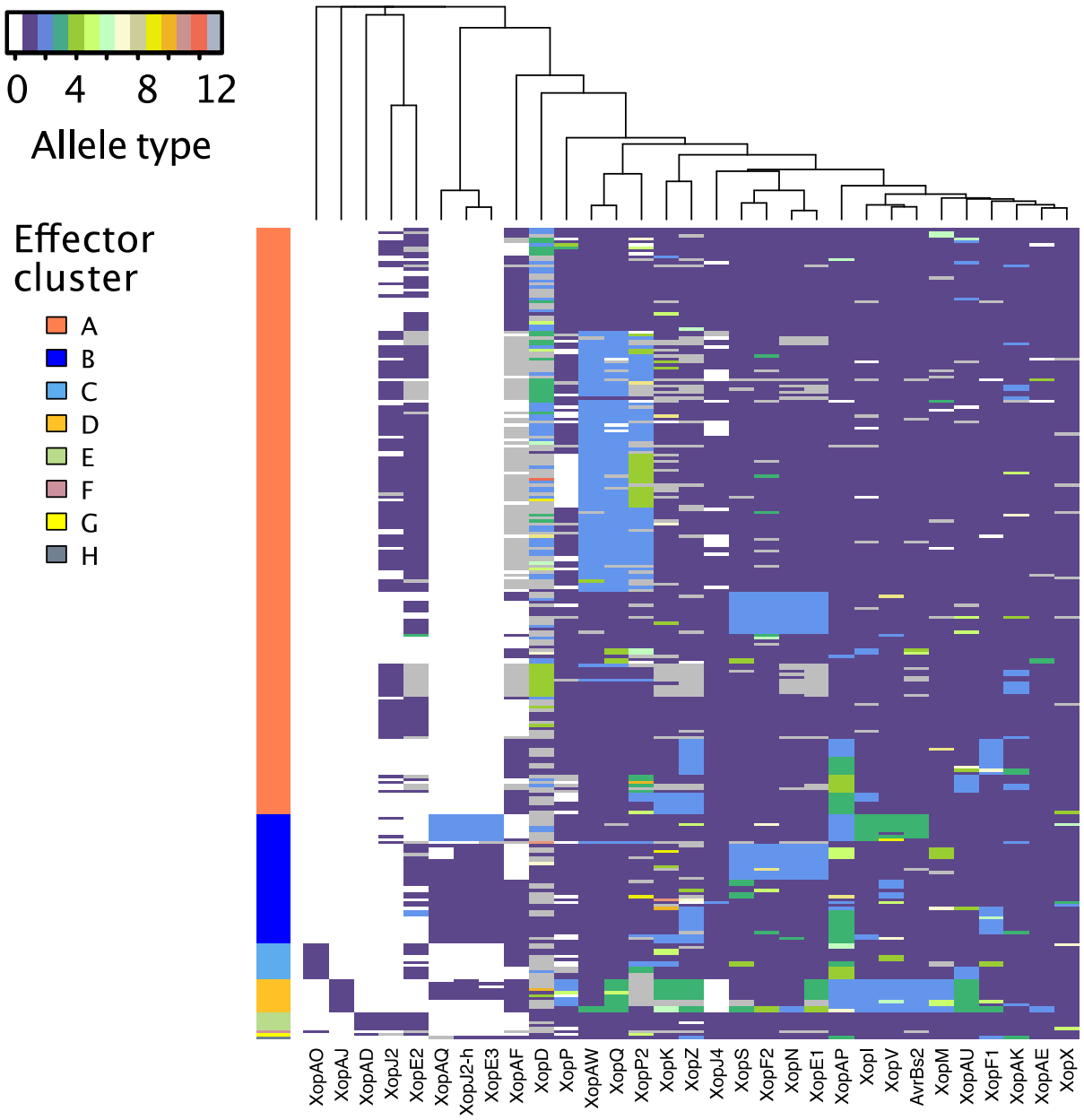


Extended Data Figure 2. Recombination in core genes and variation in gene content in *Xanthomonas perforans*. (A) Neighbor-net of 887 core genes from 259 *X. perforans* strains (excludes phylogenetically distant cluster 10 strains). Tips are colored by BAPS cluster. Reticulations in the network indicate conflicting phylogenetic relationships. (B) 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. (C) 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.

Extended Data Figure 3. Frequency of Xop effectors found among the 270 *Xanthomonas perforans* strains examined. 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.



Extended Data Figure 4. Clustering of *Xanthomonas perforans* effector profiles by non-metric multidimensional scaling and distribution of resulting clusters among sampled regions. (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 (Supplementary Figure 6). (E) Distribution of strains by effector clusters among sampled countries.



Extended Data Figure 5. Variation in Type III secreted effector profiles in *Xanthomonas perforans* with strains (rows) ordered according to NMDS of effector profiles. Type III secreted effectors are in columns and *X. 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. Order of rows is based on NMDS clustering analysis of effector profiles (see Supplementary Figure 5A).