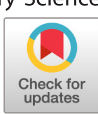
RESEARCH ARTICLE
Ecological and Evolutionary Science



An Introduced Crop Plant Is Driving Diversification of the Virulent Bacterial Pathogen *Erwinia tracheiphila*

Lori R. Shapiro,[a,b,c] Joseph N. Paulson,[d] Brian J. Arnold,[e] Erin D. Scully,[f] Olga Zhaxybayeva,[g,h] Naomi E. Pierce,[b] Jorge Rocha,[a,i] Vanja Klepac-Ceraj,[j] Kristina Holton,[k] Roberto Kolter[a]

[a]Department of Microbiology and Immunology, Harvard Medical School, Boston, Massachusetts, USA
[b]Department of Organismal and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA
[c]Department of Applied Ecology, North Carolina State University, Raleigh, North Carolina, USA
[d]Department of Biostatistics, Product Development, Genentech Inc., San Francisco, California, USA
[e]Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA
[f]Stored Product Insect and Engineering Research Unit, USDA-ARS Center for Grain and Animal Health Research, Manhattan, Kansas, USA
[g]Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire, USA
[h]Department of Computer Science, Dartmouth College, Hanover, New Hampshire, USA
[i]CIDEA Consortium Conacyt-Centro de Investigación en Alimentación y Desarrollo, Hermosillo, Mexico
[j]Department of Biological Sciences, Wellesley College, Wellesley, Massachusetts, USA
[k]Department of Biostatistics, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

ABSTRACT *Erwinia tracheiphila* is the causal agent of bacterial wilt of cucurbits, an economically important phytopathogen affecting few cultivated Cucurbitaceae host plant species in temperate eastern North America. However, essentially nothing is known about *E. tracheiphila* population structure or genetic diversity. To address this shortcoming, a representative collection of 88 *E. tracheiphila* isolates was gathered from throughout its geographic range, and their genomes were sequenced. Phylogenomic analysis revealed three genetic clusters with distinct *hrp*T3SS virulence gene repertoires, host plant association patterns, and geographic distributions. Low genetic heterogeneity within each cluster suggests a recent population bottleneck followed by population expansion. We showed that in the field and greenhouse, cucumber (*Cucumis sativus*), which was introduced to North America by early Spanish conquistadors, is the most susceptible host plant species and the only species susceptible to isolates from all three lineages. The establishment of large agricultural populations of highly susceptible *C. sativus* in temperate eastern North America may have facilitated the original emergence of *E. tracheiphila* into cucurbit agroecosystems, and this introduced plant species may now be acting as a highly susceptible reservoir host. Our findings have broad implications for agricultural sustainability by drawing attention to how worldwide crop plant movement, agricultural intensification, and locally unique environments may affect the emergence, evolution, and epidemic persistence of virulent microbial pathogens.

IMPORTANCE *Erwinia tracheiphila* is a virulent phytopathogen that infects two genera of cucurbit crop plants, *Cucurbita* spp. (pumpkin and squash) and *Cucumis* spp. (muskmelon and cucumber). One of the unusual ecological traits of this pathogen is that it is limited to temperate eastern North America. Here, we complete the first large-scale sequencing of an *E. tracheiphila* isolate collection. From phylogenomic, comparative genomic, and empirical analyses, we find that introduced *Cucumis* spp. crop plants are driving the diversification of *E. tracheiphila* into multiple lineages. Together, the results from this study show that locally unique biotic (plant population)

Received 14 June 2018 **Accepted** 17 August 2018 **Published** 2 October 2018

Citation Shapiro LR, Paulson JN, Arnold BJ, Scully ED, Zhaxybayeva O, Pierce NE, Rocha J, Klepac-Ceraj V, Holton K, Kolter R. 2018. An introduced crop plant is driving diversification of the virulent bacterial pathogen *Erwinia tracheiphila*. mBio 9:e01307-18. https://doi.org/10.1128/mBio.01307-18.

Editor David S. Guttman, University of Toronto

Copyright © 2018 Shapiro et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Lori R. Shapiro, Lori.R.Shapiro@gmail.com, or Roberto Kolter, roberto_kolter@hms.harvard.edu.

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

and abiotic (climate) conditions can drive the evolutionary trajectories of locally endemic pathogens in unexpected ways.

KEYWORDS *Erwinia*, agriculture, cucurbit, disease ecology, host jump, monoculture, pathogen emergence

Complex interactions between human behavior, demographic change, the local environment, and microbial evolution underlie the emergence and transmission of the pathogenic microorganisms that have shaped human history. Many pathogens first emerged into human populations during the Neolithic Revolution, when the widespread adoption of agricultural technologies led small, isolated hunter-gatherer groups to settle into larger, denser civilizations. These agriculture-driven demographic changes facilitated the emergence and evolution of some virulent microbial pathogens that specialized on humans as hosts (1–3). These newly emerged, human-specialized pathogens remained geographically restricted until global trade and human migration inadvertently introduced these pathogens to novel geographic areas (4, 5). Despite modern advances in medicine and public health, complex local ecological conditions—such as exponential human population growth, rapid urbanization, human-livestock and human-wild animal contact, and microbial evolution—are continuing to drive local emergence of novel human pathogens (6, 7). Devising strategies to predict pathogen emergence and to control newly emerged pathogens remains one of the most pressing public health concerns and, deservedly, is attracting intense international research efforts (8).

Less recognized is that similarly complex anthropogenic and ecological factors are likely driving the emergence of microbial pathogens into cultivated crop plant populations. Humans are continually creating new ecological niches by transforming complex ecological habitats into simplified agroecosystems (1, 9–11). Since the Neolithic Revolution, and accelerating with global trade, the geographic range of many crop plant species has expanded from the limited geographic region of origin (where the wild progenitors evolved with the endemic biotic communities for millions of years) to worldwide cultivation (12, 13). This creates landscapes of crop plants with distinct biogeographic histories suddenly being cultivated in close proximity to each other and at times to wild, undomesticated progenitors. These mosaic landscapes facilitate continual encounters of locally endemic insects and microbes with high-density populations of genetically similar native and introduced crop plant species (14–18). This increases the probability that a novel virulent pathogen will be generated through mobile DNA invasion and subsequently encounter a large, genetically homogeneous host population into which it can emerge and then rapidly spread.

Erwinia tracheiphila Smith (*Enterobacteriaceae*), the etiological agent of bacterial wilt of cucurbits, is one plant pathogen with genomic changes consistent with a recent emergence into a novel host plant population (19, 20). *E. tracheiphila* is a highly virulent phytopathogen known to affect only two genera of Cucurbitaceae crop plants—*Cucumis* spp. (cucumber and muskmelon) and *Cucurbita* spp. (pumpkin, squash, and yellow-flowered gourds). *E. tracheiphila* induces characteristic wilt symptoms by blocking xylem sap flow (Fig. 1), causing infected plants to die within 2 to 3 weeks after the first symptoms appear. Curiously, losses due to *E. tracheiphila* are reported from only a very limited geographic range in temperate midwestern and northeastern North America (20–30). This conspicuously contrasts with the worldwide distribution of susceptible cucurbit host plant species (31–34). *Cucurbita* spp. are native to the New World, and undomesticated *Cucurbita* populations naturally occur from subtropical South America through the southeastern United States (35–37). Wild *Cucumis* species are native to the Eurasian, Australian, and African tropics and subtropics (33). *Cucumis* spp. did not occur in eastern North America until Spanish colonists introduced cultivated varieties in the early 1500s (33, 38). *E. tracheiphila* causes the most severe losses in introduced *Cucumis* crop plants and less severe losses in native *Cucurbita* crops (25, 29, 39, 40). *E. tracheiphila* is obligately insect transmitted by two species of highly

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

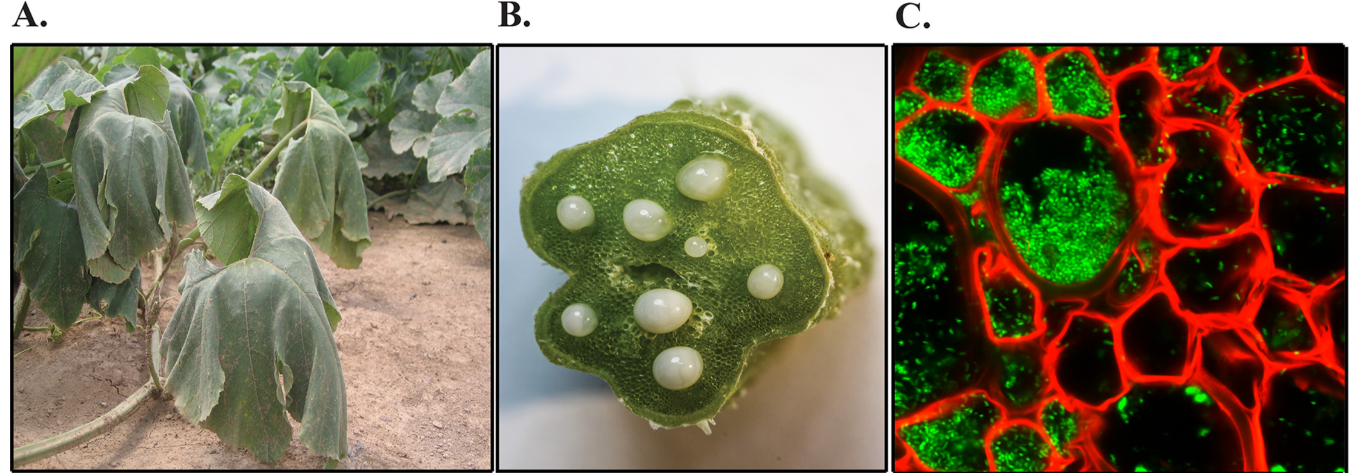


FIG 1 *Erwinia tracheiphila* infection at the macroscopic and microscopic levels. (A) A vine of a field-infected *Cucurbita pepo* plant shows characteristic systemic wilting symptoms. (B) *E. tracheiphila* can be seen oozing from multiple blocked xylem vessels in a cross-section of a symptomatic cucumber stem. (C) *In planta* confocal microscopy image of *E. tracheiphila* (green) blocking the xylem (red) of a wilting squash plant.

specialized leaf beetles that are endemic to North America: the striped (*Acalymma vittatum*) and spotted (*Diabrotica undecimpunctata*) cucumber beetles (Coleoptera: Chrysomelidae: Luperini: Galerucinae: Diabroticina). *E. tracheiphila* transmission can occur when frass from infective beetles contacts recent leaf wounds or floral nectaries (25, 41–45). Direct losses from leaf beetle herbivory and *E. tracheiphila* infection and indirect costs to control populations of the beetle vector amount to many millions of dollars annually (29).

Despite its economic burden, nothing is known about the population structure of *E. tracheiphila*, the genetic basis of virulence against the two cucurbit genera that *E. tracheiphila* infects, or why *E. tracheiphila* only occurs in such a restricted geographic range. To address this knowledge gap, we collected and sequenced an 88-isolate collection sampled from all susceptible host plants across the entire geographic range where *E. tracheiphila* is known to occur. Via analysis of the genomes of these isolates, we evaluated *E. tracheiphila* genetic diversity in relation to its plant host range and geographic distribution. We then tested for interactions between the abiotic environment (temperature) and host plant species on *E. tracheiphila* virulence. We find that these isolates group into three distinct clusters that differ in host plant associations, geographic ranges, and horizontally acquired virulence gene repertoires. Low genetic heterogeneity and an excess of rare alleles within each lineage are consistent with a recent bottleneck and expansion into a susceptible host population. In controlled inoculation experiments, *E. tracheiphila* is more virulent at temperate than subtropical summer temperatures. Further, we find that cucumber—a crop plant recently introduced into eastern North America—is the most susceptible to *E. tracheiphila* overall and the only plant species susceptible to infection by isolates from all three lineages. From this, we infer that both genetic factors (i.e., horizontal acquisition of virulence genes) and ecological factors (i.e., foreign crop plant introductions and low genetic diversity in agricultural populations) may have driven the recent emergence and epidemic persistence of *E. tracheiphila* into cucurbit agricultural populations in temperate eastern North America.

RESULTS

***Erwinia tracheiphila* is comprised of three phylogenetic lineages that have different plant host and geographic ranges.** Of 88 isolates, 68 were recovered from introduced *Cucumis* crop plants (cucumber and muskmelon) and only 20 were recovered from native *Cucurbita* crop plants (squash and pumpkin) (Table 1). A phylogenetic network analysis, which can account for and visualize phylogenetic conflict due to

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

TABLE 1 Summary of the host associations of the sequenced *Erwinia tracheiphila* isolates[a]

Cluster	No. of isolates	Host plant species of isolation	Host plant status in the Americas
Et-melo	27	Muskmelon (*Cucumis melo*)	Introduced
Et-melo	28	Cucumber (*Cucumis sativus*)	Introduced
Et-C1	14	Squash and pumpkin (*Cucurbita* spp.)	Native
Et-C1	12	Cucumber (*Cucumis sativus*)	Introduced
Et-C2	6	Squash and pumpkin (*Cucurbita* spp.)	Native
Et-C2	1	Cucumber (*Cucumis sativus*)	Introduced

[a]See Table S1 for the detailed metadata for each individual isolate.

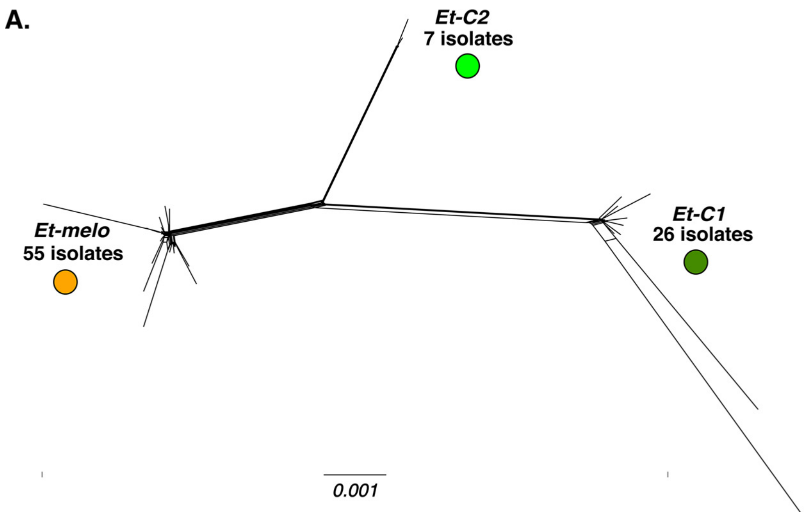
recombination and gene flow (46, 47), revealed that *E. tracheiphila* is comprised of three distinct, coexisting phylogenetic clusters, designated *Et-C1*, *Et-C2*, and *Et-melo* (Fig. 2A; also see Fig. S1 in the supplemental material and see the text file at https://figshare.com/projects/Recent_emergence_of_a_virulent_phytopathogen/35108). Faint reticulations along the long branches connecting *Et-C1* and *Et-melo* suggest some limited gene flow between these two groups. *Et-C2* is on a nonreticulating long branch and shows no evidence of gene flow with either *Et-C1* or *Et-melo* (Fig. 2A). We refer to these three distinct groups as phylogenetic "clusters" instead of "pathovars," as "pathovar" assignments are often inconsistent with phylogenetic group (48–51).

The three clusters are present at different frequencies, over different geographic ranges, and have distinct host plant association patterns (Fig. 2B and Table 1; see also Table S1). The most frequently recovered *E. tracheiphila* isolates (55 isolates) belong to the *Et-melo* cluster and were collected exclusively from cucumber and muskmelon. *Et-melo* also has the largest geographic distribution, encompassing the known range of *E. tracheiphila* throughout the midwestern and northeastern United States (Fig. 2B). The 26 *Et-C1* isolates were recovered from both introduced cucumber and native squash plants collected in the Mid-Atlantic and Northeast (Table 1). Of the 7 *Et-C2* isolates, six were recovered from squash and one from cucumber (Table 1), and all *Et-C2* isolates were found in the northeastern United States (Fig. 2B). Isolates from all three clusters were found in field-infected cucumber plants, while muskmelon was infected only by the *Et-melo* isolates, and squash was infected only by the *Et-C1* and *Et-C2* isolates (Table 1). All three lineages are geographically restricted to temperate eastern North America (Fig. 2B). This is further north than where wild, undomesticated *Cucurbita* spp. naturally occur in the American tropics and subtropics (31, 35, 52).

All three *Erwinia tracheiphila* lineages have low genetic diversity. To investigate the recent population history of *E. tracheiphila*, genetic diversity was measured with the Watterson estimator (θ_W) and Tajima's D. These were calculated separately within each phylogenetic cluster and within each collection period (collection period one from 2008-2010, and collection period two in 2015). The core genes shared by all isolates within each lineage were assigned as putatively functional (Intact), or mobile DNA/putatively pseudogenized (Pseudogenized + Repetitive) using published, manually curated gene annotations from the BuffGH reference genome (formerly PSU-1) (20, 30). There is low within-cluster nucleotide diversity (θ_W) in all three lineages (Table 2) despite clear between-cluster genetic divergence (Fig. 2A), which is consistent with small effective population sizes. *Et-C2*, which was observed only in the 2015 collection, has the fewest segregating sites, is represented by the fewest isolates in the smallest geographic range, and has isolates with the shortest branch lengths (Fig. 2A), which together suggest that *Et-C2* may be the most recently emerged lineage (Tables 1 and 2; Fig. 2B). For both *Et-C1* and *Et-melo*, θ_W increased over the 7-year period, although diversity increased 7-fold faster in *Et-C1* than *Et-melo*. The low overall heterogeneity within each *E. tracheiphila* cluster is compatible with recent emergence from a small founder population and recent divergence into distinct genetic clusters.

In addition to the density of polymorphic sites (θ_W), the allele frequencies at these sites also contain information about recent population history. Tajima's D, which measures the degree to which the allele frequency spectrum is compatible with that of

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest



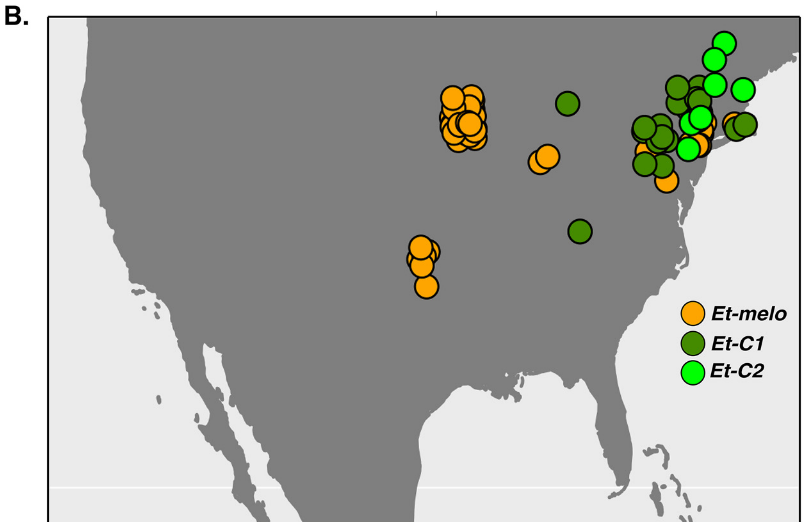


FIG 2 Three genetically distinct lineages of *Erwinia tracheiphila* and their geographic distributions. (A) The phylogenetic network of 88 *Erwinia tracheiphila* isolates. The network was reconstructed from concatenated alignments of the core gene families identified with OrthoMCL in all 88 *E. tracheiphila* genomes. Three distinct clusters separated by long branches are named *Et-melo*, *Et-C1*, and *Et-C2* based on the host plant that they were found to infect (Table 1): isolates from clusters *Et-C1* and *Et-C2* were found only on squash and cucumber plants, while strains from cluster *Et-melo* were found only on muskmelon and cucumber. Host plant, year of isolation, location, and assembly metadata for each isolate are listed in Table S1. Scale bar shows number of substitutions per site. Figure S1 shows individual isolate identifiers for each isolate in the network. (B) Geographic distribution of the three clusters. Each of the 88 isolates is plotted as a single circle on the map according to its collection site and colored according to the genetic cluster to which it belongs (see panel A). The isolate-specific locations and year of collection are listed in Table S1.

a neutral population of constant size, is negative for all three clusters (Table 2). This reflects an excess of rare variants and suggests that these three lineages are experiencing an ongoing population expansion after a bottleneck. The excess of rare alleles is consistent with the hypothesis that these three relatively monomorphic lineages are rapidly spreading within genetically homogenous host plant populations that are susceptible to infection by pathogen variants with the same virulence alleles. All three lineages show evidence of limited within-lineage recombination, although the large number of repetitive regions likely makes recombination estimates inexact (Table 2). While estimated rates of homologous recombination are relatively low for all three clusters, this process may be contributing to lack of within-cluster phylogenetic structure (Fig. 2A).

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

TABLE 2 Nucleotide diversity of Intact versus Pseudogenized + Repetitive genes within the three *Erwinia tracheiphila* lineages collected during the two collection periods, from 2008 to 2010, and in 2015

			No. of segregating sites (S)		Tajima's D		Watterson estimator (θ_W)	
Cluster	r/m[b]	Gene type	2008–2010	2015	2008–2010	2015	2008–2010	2015
Et-melo	0.317	Pseudogenized + Repetitive	655	2,568	−2.64	−2.12	6.24×10^{-4}	3.61×10^{-3}
		Intact	1,444	5,346	−2.44	−2.15	2.15×10^{-4}	1.18×10^{-3}
Et-C1	0.682	Pseudogenized + Repetitive	87	3,485	−1.56	−2.17	9.03×10^{-5}	4.05×10^{-3}
		Intact	264	10,026	−1.81	−2.17	4.94×10^{-5}	2.06×10^{-3}
Et-C2	0.229	Pseudogenized + Repetitive	NC[a]	1,070	NC	−0.73	NC	1.38×10^{-3}
		Intact	NC	1,512	NC	−0.64	NC	3.79×10^{-4}

[a]NC, not collected, i.e., not found in that collection period.
[b]Estimated recombination-to-mutation-rate ratio within a cluster.

Estimation of the *Erwinia tracheiphila* core genome, pangenome, and functional repertoire. The entire *E. tracheiphila* pangenome of the 88 strains sequenced here, encompassing all core, accessory, and unique genes, is 10,598 gene families (Fig. 3A). The pangenomes of geographically widespread microbes with environmental reservoirs such as *Prochlorococcus* or *Escherichia coli* have almost an order of magnitude more gene clusters (53, 54). The relatively small *E. tracheiphila* pangenome size of ~10,600 genes is compatible with the hypotheses that *E. tracheiphila* is a host-restricted pathogen that recently emerged from a population bottleneck and/or is predominantly circulating in low-diversity agricultural host plant populations.

Of the 4,032 gene families present in at least 95% of sequenced genomes, 2,907 (72%) can be assigned to a functional category of the Clusters of Orthologous Groups (COG) database (55). These "core" gene families are enriched in almost all COG categories associated with cellular processes and metabolism (Fig. 3B and Table S1). This finding is consistent with these gene families being essential for survival and therefore ubiquitous in all isolates in the population. Only 699 out of 3,720 (18.8%) genes found in fewer than 5% of *E. tracheiphila* sequenced genomes are assignable to a COG functional category. This set of gene families that are "rare" in the population are enriched in only "Mobilome" (X), suggesting that most rare genes are accessory genes or mobile DNA and are not involved in cellular, metabolic, or information processes (Fig. 3B and Table S2).

***Erwinia tracheiphila* clusters vary by *hrp*T3SS effector content.** Many Gram-negative bacterial phytopathogens use a hypersensitive response and pathogenicity type III secretion system (*hrp*T3SS) to translocate effector proteins directly into the host plant cell. In the plant cell cytoplasm, T3SS effectors may reveal the presence of a pathogen and initiate a cascade of antipathogen defenses, often mediated through salicylic acid (56). Alternatively, effectors may promote pathogen virulence by suppressing induced plant defense responses. *E. tracheiphila* contains an *hrp*T3SS locus, and *E. tracheiphila* suppresses salicylic acid production in a wild gourd host (*Cucurbita pepo* subsp. *texana*) (20, 44), suggesting that *E. tracheiphila* may use effectors for suppressing plant-induced defenses during disease development.

We found that the 88 *E. tracheiphila* isolates collectively carry at least 23 *hrp*T3SS effector genes (Fig. 4 and Fig. S2). Because differences in T3SS effector repertoire can drive host plant specificity, we also examined the distribution of effector genes between the three *E. tracheiphila* clusters. Cluster *Et-melo* has one unique effector gene, Eop3, which is homologous to the Eop3 gene in *Erwinia amylovora* (57), the uncharacterized *Pseudomonas syringae* pv. actinidiae effector HopBN1 (16), and the *P. syringae* effector HopX1 (58). Two other effector genes, NleD and AvrRpm1, are unique to the *Et-C1* cluster. In the BuffGH reference genome, NleD is present in six copies, including in an intact phage region (20, 30). The *E. tracheiphila* NleD genes have 99% amino acid identity to an NleD gene in an active phage region in the emerging mouse pathogen *Citrobacter rodentium* (59) (Fig. S2). The functional significance for *E. tracheiphila* having

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

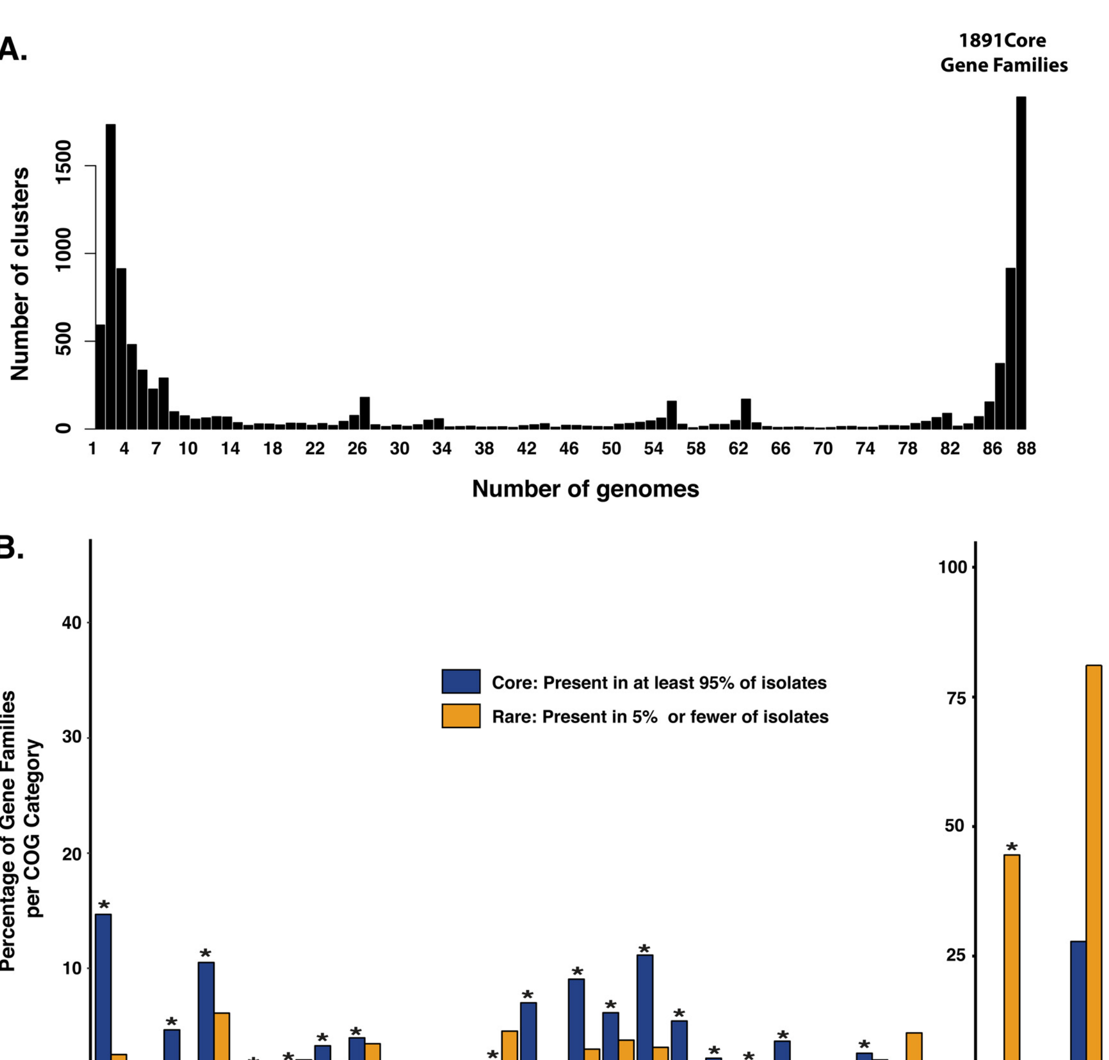


FIG 3 The *Erwinia tracheiphila* pangenome and its functional annotations. (A) Distribution of detected gene families among core and rare pangenome. The number of gene families (*y* axis) is plotted as a histogram, with counts of the number of sequenced *E. tracheiphila* genomes that contain them (*x* axis). Of 10,598 gene families in the species, Micropan identifies 1,891 gene families present in all 88 genomes. There are 4,032 gene families present in at least 95% (84 out of 88 genomes), which are referred to as core genes; 4,890 gene families are found at intermediate frequency (between 6 and 84 genomes); and 3,720 gene families are present in less than 5% (6 genomes) and are referred to as rare genes. (B) Distribution of predicted functions in the core and rare gene families of *E. tracheiphila*. The core and rare gene families are grouped into COG categories (*x* axis), which are annotated by their one-letter abbreviations (see Table S1 for notations). The *y* axis shows the percentage of the gene families within each COG category. The bar to the far right shows the overall percentage of the core and rare gene families that were not represented in COG. 'Mobile' (X) and the number of genes not assigned to a COG are shown with a 100% *y* axis, while the other categories are shown with a *y* axis scaled to 40%. Asterisks designate the functional categories that are significantly overrepresented compared to the distribution of all genes in that category (Fisher's exact test, $P < 0.05$; Table S2). The percentages of rare and core genes not in COG (far right) are shown for scale but were not included in the statistical tests.

six NleD copies—if there is functional significance—is not yet known. There are no effector genes that are unique only to the *Et-C2* cluster, but a gene for effector HopAM1 is present in *Et-C2* and *Et-C1* isolates, and a gene for HopAF1 is present in *Et-C2* and *Et-melo* isolates (Fig. 4). In *P. syringae*, HopAM1 manipulates abscisic acid-mediated responses and water availability via stomatal closure (60), but how it affects the virulence phenotype for *E. tracheiphila* is unknown. In *P. syringae*, HopAF1 inhibits pathogen-associated molecular pattern (PAMP)-mediated increases in ethylene pro-

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

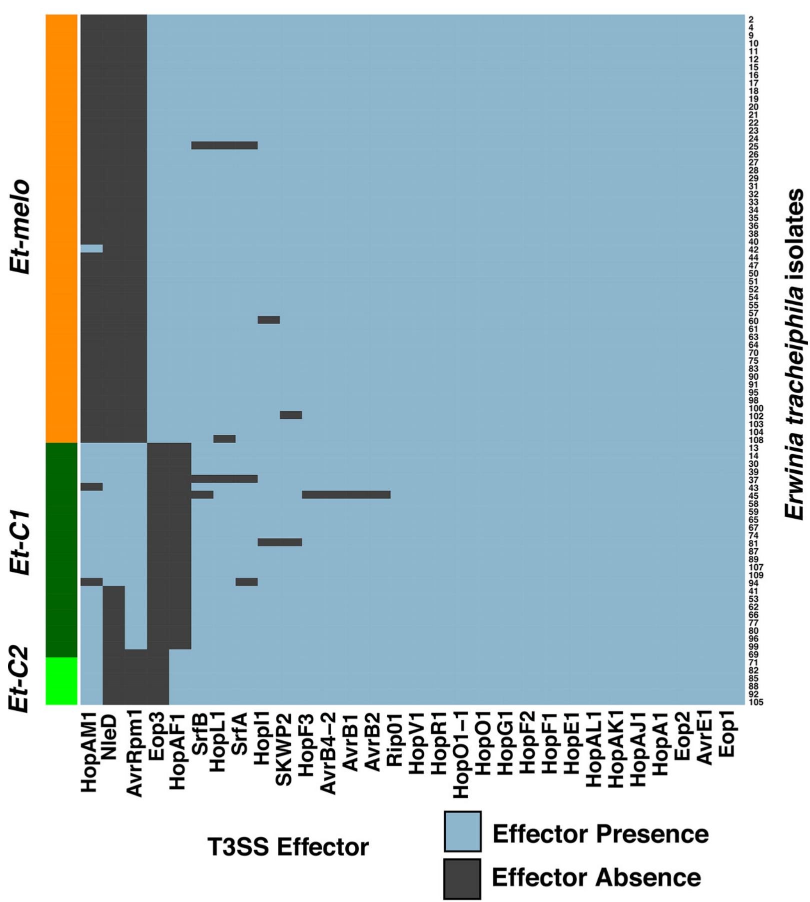


FIG 4 Distribution of *hrp*T3SS effector genes across the genomes of 88 *Erwinia tracheiphila* isolates. Each individual sequenced isolate is represented by a row, and the rows are grouped by phylogenetic cluster (*y* axis). The 23 effector genes found in the *Erwinia tracheiphila* pangenome are arranged on the *x* axis. Each cell in the matrix is color coded by the presence (blue) or absence (dark gray) of a *hrp*T3SS effector gene in a genome of an individual isolate. The effector presence/absence matrix with isolate names is included in Fig. S2. Phylogenetic trees for the five cluster-specific effectors (HopAM1, NleD, AvrRpm1, Eop3, and HopAF1) are shown in Fig. S2.

duction, and homologs are widely distributed in many bacterial phytopathogens (61). All five cluster-specific effectors (HopAM1, NleD, AvrRpm1, Eop3, and HopAF1) are physically located far from the *hrp*T3SS locus, and their evolutionary histories are all consistent with horizontal acquisition (Fig. S2). Phytopathogen effectors are often determinants of host range, and the horizontal acquisition of these five effectors may underlie the split of *E. tracheiphila* into phylogenetic clusters with distinct virulence phenotypes and host plant association patterns.

Cucumber is the only host plant susceptible to all *Erwinia tracheiphila* lineages. Controlled cross-inoculation experiments were used to test whether the patterns of lineage-specific host plant associations observed in the field were due to random sampling patterns or were reflective of genetic differences. In the greenhouse, three isolates from *Et-melo*, three isolates from *Et-C1*, and one isolate from *Et-C2* were all cross-inoculated into 2-week-old seedlings of squash, cucumber, and muskmelon. Isolates from *Et-melo* killed all experimental cucumber and muskmelon plants (Fig. 5). In squash, *Et-melo* isolates induced localized wilt symptoms, but all squash plants inoculated with *Et-melo* recovered (Fig. 5). Isolates from *Et-C1* and *Et-C2* were highly virulent against cucumber, killing 98% of experimental cucumber plants, but less virulent against squash and muskmelon (Fig. 5 and Table 3). The attenuation of *Et-C1* and *Et-C2* virulence on muskmelon compared to *Et-melo* in the greenhouse is likely ecologically important, as none of these strains have yet been isolated from field-

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

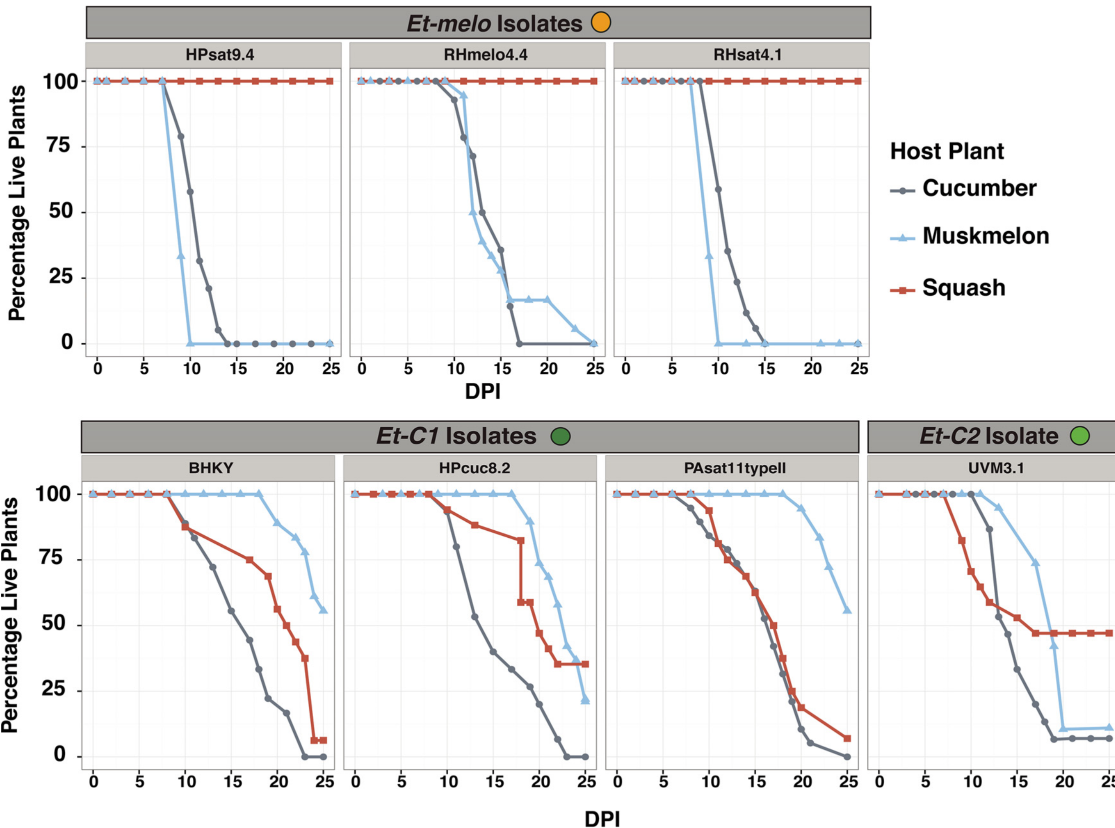


FIG 5 *In planta* virulence of isolates from different clusters in muskmelon, cucumber, and squash. Individual panels show the change in the percentage of live plants (*y* axis) over 25 days postinoculation (DPI; *x* axis) in controlled greenhouse cross-inoculation experiments. The name of each tested isolate is shown inside a light gray bar, and the isolates are grouped according to the phylogenetic cluster to which they belong (Fig. 2A).

infected muskmelon (Table 1). Squash showed variable susceptibility to isolates from *Et-C1* and *Et-C2*, which is consistent with previous reports that this genus is moderately resistant to *E. tracheiphila* (Fig. 5 and Table 3) (25). In summary, cucumber is the most susceptible of the three host plant species and is the only host plant susceptible to infection by isolates from all three *E. tracheiphila* clusters in both the field (Table 1 and Fig. 2A) and greenhouse (Fig. 5).

Subtropical temperatures inhibit *Erwinia tracheiphila in vitro* growth and *in vivo* virulence. We tested the effects of temperature on *in vitro* growth and *in vivo* virulence to determine whether the temperatures in temperate eastern North America, the only region in the world where *E. tracheiphila* is known (see "Confirmation of

TABLE 3 Summary statistics describing wilt disease progression in cucumber (*Cucumis sativus*), muskmelon (*Cucumis melo*), and squash plants (*Cucurbita pepo*) inoculated with isolates from different phylogenetic clusters

		Total no. of plants			Avg no. of days until:	
Phylogenetic origin of inoculating strains	Host plant	Inoculated	Dead[a]	% dead plants[a]	First wilt symptoms	Plant death
Et-melo	Muskmelon	54	54	100	4.2	10.98
	Cucumber	50	50	100	6.5	11.96
	Squash	46	0	0	12	None died
Et-C1 + *Et-C2*	Muskmelon	74	49	66.2	5.6	21.2
	Cucumber	66	67	98	7.4	15.9
	Squash	66	50	75.8	6.5	16.8

[a]At the end of the 25-day observation period.

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

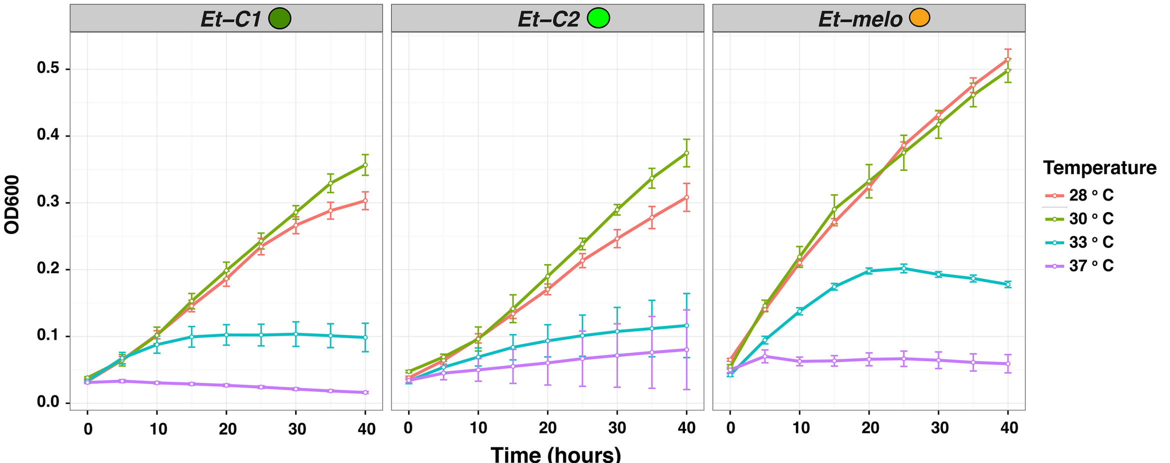


FIG 6 Effects of temperature on *in vitro* growth of *Erwinia tracheiphila*. Individual panels show *in vitro* growth for 7 isolates from *Et-C1*, 2 isolates from *Et-C2*, and 4 isolates from *Et-melo* grown at four different temperatures. Bacterial growth was assessed via optical density (OD_{600}; *y* axis), which was measured hourly over 40 h, and displayed in intervals of 10 h on the *x* axis. Each individual curve shows the average values of all tested isolates in a corresponding cluster. Error bars, standard error of the mean.

restricted *Erwinia tracheiphila* geographic range" in Materials and Methods), are more favorable for *E. tracheiphila* growth than subtropical temperatures. For isolates from all three clusters, we find that the final concentration indicated as optical density at 600 nm (OD_{600}) after 40 h of *in vitro* growth is suppressed at warmer 33°C and 37°C incubation temperatures, compared to incubation at cooler temperatures of 28°C or 30°C ($P \leq 0.001$) (Fig. 6 and Table 4).

To test the effects of temperature on *in vivo* virulence, we isolated an *E. tracheiphila* strain from a field-infected cucumber and a second strain from a field-infected squash. Each isolate was then inoculated into the host species in which it was found. Half of the plants were incubated at average July temperatures measured in Massachusetts (27°C day/18°C night) to represent the temperature in the northeastern United States. This is the region where *E. tracheiphila* is an annual epidemic, all three *E. tracheiphila* lineages were found, and cultivated squash, cucumber, and muskmelon are present only due to human agriculture. The other half of the inoculated plants were incubated at average July temperatures measured in Texas (33°C day/23°C night) to represent the subtropical southwestern United States, where the wild squash progenitor (*Cucurbita pepo* subsp. *texana*) is native but *E. tracheiphila* has never been reported (35). At "southwestern U.S." temperatures, only three inoculated squash plants developed localized symptoms in the inoculated leaf, and these three plants recovered. At cooler "northeastern U.S." temperatures, half of the squash plants developed localized wilt symptoms, but only six of these plants developed systemic disease and died within the 25-day experiment (Fig. 7 and Table 5). In contrast to squash, at southwestern U.S. temperatures, 34 out of 36 cucumber plants died by 25 days postinfection (DPI). At cooler northeastern U.S. temperatures, all 36 cucumber plants died by 19 DPI. Moreover, plant death at temperate northeastern U.S. temperatures occurred significantly faster in cucumber (mean of 13.8 days) than in squash (mean of 18.3 days). In summary, cucumber is

TABLE 4 Two-way analysis of variance that tested the effects of temperature, phylogenetic group, and their interaction on *E. tracheiphila in vitro* growth

Source of variation	Degrees of freedom	Sum of squares	Mean squares	*F* value	Pr(>*F*)
Temp	3	2.773	0.9244	113.372	<2e−16
Cluster	2	0.725	0.3625	44.451	<2e−16
Temp × cluster	6	0.214	0.0356	4.371	0.000263
Residuals	456	3.718	0.0082		

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

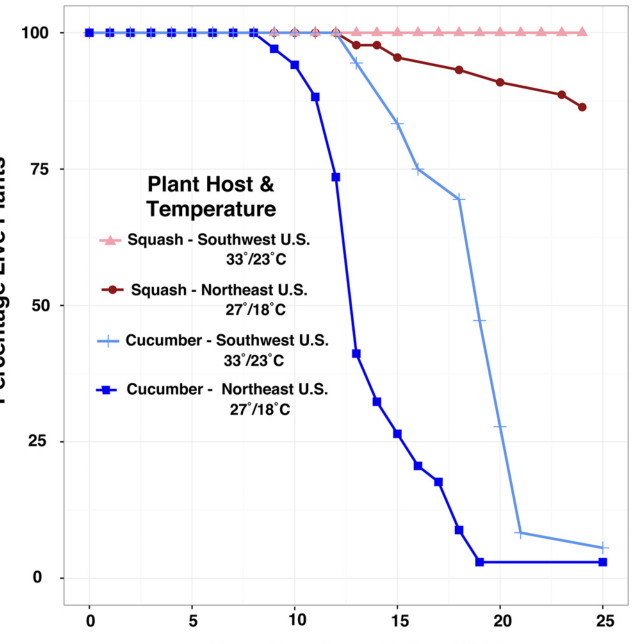


FIG 7 Effects of temperature on *in planta* virulence. Change in percentage of live squash and cucumber seedlings (*y* axis) was tracked at two different temperatures over time (*x* axis). The time is shown in days since inoculation of plants with *Erwinia tracheiphila* (day 0). Incubation was at either temperate (dark red for squash and dark blue for cucumber) or subtropical (light red for squash and light blue for cucumber) temperatures.

TABLE 5 Effect of average summer subtropical temperatures normal in the *Cucurbita pepo* native range (measured in Texas) versus average summer temperatures in temperate northeastern United States (measured in Massachusetts) on *Erwinia tracheiphila in vivo* virulence in cucumber and squash plants

		No. of plants				Mean no. of days until:	
Geographic location	Plant	Total inoculated	Showing wilt symptoms	Died	% that died	First wilt symptom	Death
Southwestern United States (Texas)	Cucumber	36	35	34	94	10.8	18.6
	Squash	44	3	0	0	19	None died
Temperate eastern North America (Massachusetts)	Cucumber	34	34	34	100	7.8	13.8
	Squash	44	22	6	13.6	9.3	18.3

significantly more susceptible than squash at both tested temperatures. Cooler temperatures (normal in the Northern introduced range) are required for *E. tracheiphila* virulence in squash and also increase virulence of *E. tracheiphila* against cucumber.

DISCUSSION

In our comprehensive study of *Erwinia tracheiphila* genomic diversity, host plant association patterns, and demographic history, we found that *E. tracheiphila* is comprised of three distinct, homogeneous phylogenetic lineages that have an excess of rare genetic variants. From this, we infer that these three clusters were recently founded by small populations and are currently experiencing rapid population expansions to fill new agroecological niches (3, 62–64, 134, 135). These inferences about *E. tracheiphila* demographic history correlate with recent anthropogenic changes to cucurbit agroecosystems in eastern North America. The recent introduction of all cucurbit crop plants into temperate eastern North America, one of the world's most agriculturally intensive regions, likely created a novel ecological niche (33, 65, 66). Cucumber is the most susceptible plant species in the greenhouse and field and the only plant species highly susceptible to infection by isolates from all three *E. tracheiphila* lineages. The high susceptibility of cucumber to isolates from all three clusters

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

in both the field and greenhouse suggests that cucumbers could be functioning ecologically as a highly susceptible reservoir host. This presents the possibility that *E. tracheiphila* (which was already present in the midwestern United States by 1900 [67, 68]) could not have emerged or persisted as an annual epidemic without the human-mediated introduction of cultivated *Cucumis* spp. into temperate North America in the early 1500s (33).

E. tracheiphila has among the most dramatic structural genomic changes—including gene decay through pseudogenization, mobile element invasion and proliferation, and horizontal gene acquisitions—of any bacterial pathogen (20). These structural changes are consistent with a recent evolutionary transition from a progenitor with multiple environmental reservoirs and diverse metabolic capabilities to a pathogen with a narrow, host-specialized ecological niche. However, the species identity, geographic origin, and host relationships of the direct *E. tracheiphila* progenitor are all unknown, limiting our ability to investigate the evolutionary transition from the *E. tracheiphila* direct progenitor—presumably a plant commensal or weak pathogen—to a virulent pathogen (62, 69, 70). The genomic evidence of the recent transition of *E. tracheiphila* to a virulent, host-restricted pathogen (20) highlights the continuing risk of nonpathogenic environmental microbes acquiring virulence genes via continual and naturally occurring mobile DNA invasion (71). Virulent pathogens are unlikely to persist in ecologically intact habitats with higher plant species diversity and higher diversity of pathogen resistance (R) genes (72–74). When pathogens evolve or acquire novel virulence genes, this acts as a selective pressure on host plant populations and causes a rise in frequency of plant resistance genes. However, repeatedly planting the same crop plant varieties in agricultural populations interferes with this coevolutionary dynamic by preventing a rise in frequency of effective host plant resistance alleles. Identifying cultivars or wild crop relatives with resistance genes, and crossing them into cultivated crop populations, is one method favored by plant breeders. However, the probability of success from this approach for controlling *E. tracheiphila* is likely to be low. Cucumber is the best characterized of all cucurbit crops, and this species was found to contain among the lowest genetic heterogeneity of any vegetable crop, with an estimated effective population size of only 500 individuals at the time of domestication (75, 76). The *E. tracheiphila*-cucurbit association is evolutionarily novel (20), suggesting that genetic resistance to *E. tracheiphila* may not exist in any undomesticated cucurbit populations. Even if the genetic basis of host resistance is identified in wild relatives or rare cultivars of cucumber, squash, or melon and successfully introduced into agricultural populations, *E. tracheiphila* is amenable to invasion by mobile DNA, including acquisition of virulence effector genes (20). This could function to quickly overcome potential host plant genetic resistance, especially if the same resistance gene(s) is broadly deployed in large, homogeneous crop plant populations (77, 78). This potential to rapidly generate novel variants from a recombining source population(s), together with the ability to horizontally acquire virulence effectors, will be important to consider when attempting to design durable resistance strategies for agricultural systems (20, 79).

Many—perhaps most—of the economically damaging plant pathogens and insect pests have emerged after the Neolithic Revolution (11, 16, 63, 64, 80–85). Yet, little effort has been put toward using ecological principles to plan genetic, physiological, and/or structural complexity into agricultural systems to mitigate susceptibility to outbreaks of insect pests or microbial pathogens (10). We hypothesize that the kind of local pathogen (or insect pest) emergence such as what has happened with *E. tracheiphila* is more common than currently understood. Further, we predict that these local emergence events can in some cases be followed by rapid dissemination through genetically homogeneous agricultural populations. Given the potential of such infections to threaten globalized crop populations, including staple crops that are vital for local and global food security, we urgently need to develop approaches for building sustainable agroecosystems that are rooted in ecological and evolutionary principles.

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

MATERIALS AND METHODS

Study system. Wild species in the gourd family, Cucurbitaceae, occur in tropical and subtropical regions worldwide, and cultivars from this family are among the world's most widely grown fruit and vegetable crops (34, 86). Like many Cucurbitaceae, *Cucurbita* spp. and *Cucumis* spp. produce a class of secondary metabolites called cucurbitacins (87–89). Cucurbitacins are among the most bitter and toxic compounds ever characterized and function as highly effective herbivory deterrents for almost all insect and mammalian herbivores, including humans (89–92). The exceptions are a few genera of highly coevolved Luperini leaf beetles (Coleoptera: Chrysomelidae), and for these beetles, cucurbitacins function as arrestants and feeding stimulants (90, 93, 94). *Acalymma* is a strictly New World genus of highly specialized leaf beetles that has coevolved in Mesoamerica with *Cucurbita.* In natural settings, *Acalymma* spp. are obligately dependent on *Cucurbita* plants in all life stages (95–97). *E. tracheiphila* has no known environmental reservoirs and persists only within infected *Cucurbita* or *Cucumis* host plants or the digestive tracts of the highly specialized beetle vectors. Beetle vectors are the only documented winter reservoirs of *E. tracheiphila* (45, 89, 98). The Eastern striped cucumber beetle (*Acalymma vittatum*) is the only *Acalymma* species that has received substantial research attention because of its status as an important agricultural pest and plant pathogen vector in eastern North America (97). *A. vittatum*, which is the predominant insect vector of *E. tracheiphila*, occurs only in northeastern and midwestern North America. It is likely that *A. vittatum* only recently emerged into this geographic area following the domestication and range expansion of *Cucurbita* for agriculture, as was recently shown for the obligate pollinator of *Cucurbita* in eastern North America (66, 99). In the Old World, *Aulocophora* species (Coleoptera: Chrysomelidae: Luperini) are obligate cucurbit specialists, although natural history information is almost completely absent for almost all species (100, 101).

Confirmation of restricted *Erwinia tracheiphila* geographic range. Losses from *E. tracheiphila* are an annual epidemic in temperate eastern North America (22, 25, 26, 29, 41, 87, 98, 102–106). No losses from *E. tracheiphila* have been reported anywhere else in the world. To evaluate whether the reported geographic restriction of *E. tracheiphila* to temperate eastern North America is a reflection of its actual geographic range or an artifact of this pathogen not being recognized outside this range, one of us (L.R.S.) undertook extensive scouting expeditions of wild and cultivated *Cucurbita*, *Cucumis*, *Luffa*, and *Lagenaria* populations in diverse areas of the world, including the entire southern United States from California to South Carolina; on the west coast of Mexico from Jalisco to Oaxaca; in Europe; and in Southeast Asia. There is one report of *E. tracheiphila* in New Mexico (107), but this isolate was said to be from a cultivated watermelon (which is not susceptible) and this isolate is not archived, nor do gene sequences from it exist, and we must therefore at this time consider it a single erroneous report.

No *E. tracheiphila* symptoms were observed in undomesticated populations of *Cucurbita digitata* in California and Arizona or in undomesticated or domesticated *Cucurbita* spp. or *Cucumis* spp. in California, Arizona, New Mexico, Texas, Louisiana, Mississippi, Alabama, Georgia, South Carolina, or Missouri. In Mexico, *E. tracheiphila* was not found in wild or cultivated cucurbits in the Mexican states of Jalisco, Guerrero, Michoacán, Oaxaca, Guanajuato, or Querétaro. Nor was *E. tracheiphila* observed in any cucurbits in commercial or academic farms in Thailand, Philippines, or Vietnam. In Europe, *E. tracheiphila* was never observed in cucumber or squash plants in Spain or Germany. These observations are consistent with the lack of reports of *E. tracheiphila* outside temperate northeastern and midwestern North America. *E. tracheiphila* has never been shown to survive outside a few agricultural species of cucurbit hosts and beetle vectors (45, 98). Therefore, the isolates collected in this study (Fig. 2A and B; see also Table S1 in the supplemental material) are hypothesized to cover the entire plant host and geographic range where *Erwinia tracheiphila* exists.

Collecting single isolates of *Erwinia tracheiphila*. Single *E. tracheiphila* isolates were obtained from symptomatic squash (*Cucurbita pepo*), muskmelon (*Cucumis melo*), and cucumber (*Cucumis sativus*) plants in agroecosystems from across the entire geographic range where economic losses from *E. tracheiphila* are reported (Fig. 2B; Table S1). In the field, infected plants were visually identified by characteristic wilting symptoms (Fig. 1A). All wilting, symptomatic plants in a given field were gathered to avoid collection bias. Symptomatic vines from infected plants were removed with a sterile knife, immediately placed in separate 1-gal plastic bags, and stored at 4°C for a maximum of 3 days prior to performing bacterial isolations. The reference BuffGH strain (formerly PSU-1) was isolated in 2007 from an undomesticated wild gourd *C. pepo* subsp. *texana* plant growing at the Rock Springs Experimental Station in Rock Springs, PA (30). These *C. pepo* subsp. *texana* seeds were originally collected from wild populations in New Mexico and Texas and were greenhouse cultivated and then field transplanted for academic research at Pennsylvania State University in University Park, PA (reviewed in reference 87). Isolates collected in 2007 to 2009 were acquired from the authors of reference 51, were collected according to the protocol described there, and are stored at Iowa State University in Ames, IA. *E. tracheiphila* isolates from 2015 were collected by first washing external dirt and debris from symptomatic vines with tap water and then surface sterilizing the cleaned vines with 70% ethanol. Sterilized vines were cut into 3- to 4-in. sections between nodes with sterile razor blades, and 1/2 in. of the vine sections was soaked in 3 ml of autoclaved Milli-Q water in 15-ml Falcon tubes until pure *E. tracheiphila* could be seen on the cut surface (Fig. 1B). Sterile loops were then used to transfer *E. tracheiphila* ooze (Fig. 1B) to King's B (KB) agar plates (1 liter: 20 g protease peptone no. 3, 10 ml glycerol, 1.5 g $MgSO_4 \cdot 7H_2O$, 1.5 g KH_2PO_4, 15 g Bacto agar). Single isolates were restreaked, and then single colonies from the restreaked plates were grown in shaken liquid KB broth at 25°C for 48 h and cryogenically preserved with 15% glycerol.

DNA extraction, library preparation and whole-genome sequencing. Single colonies from cryogenically preserved glycerol stock were grown on KB agar plates, and single colonies were grown in

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

liquid KB for 36 to 48 h or until the OD_{600} reached 1. DNA from liquid cultures was extracted with Promega DNA Wizard (Promega, Madison, WI) according to the manufacturer's instructions.

Libraries of the genomic DNA for isolates listed in Table S1 were generated using a Nextera DNA sample preparation kit (Illumina, San Diego, CA). The libraries were amplified for 8 cycles using the Kapa HiFi library amplification kit (Kapa Biosystems, Wilmington, MA), and the size selection was performed using AMPure XP beads (Agencourt Bioscience Corp., Beverly, MA). Library concentrations were measured using a Qubit DNA quantification kit (Life Technologies, Carlsbad, CA), and the fragment size range detection (100 to 400 bp) was performed using the TapeStation 2200 (Agilent Technologies, Santa Clara, CA). Libraries were pooled using Nextera index kits, and 150-bp paired-end reads were generated with an Illumina HiSeq 2500 sequencing system. Assembly metrics of all strains sequenced for this study were determined with QUAST, with standard settings that retain only contigs larger than 500 bp (108).

Transformation of *Erwinia tracheiphila* with an mCherry-expressing plasmid. *E. tracheiphila* strain BuffGH was used for visualization of *E. tracheiphila* in the xylem of infected squash seedlings. Plasmid pMP7605 carrying a constitutively expressed mCherry gene was electroporated into competent *E. tracheiphila* cells. For this, we followed protocols described previously (109). Briefly, competent *E. tracheiphila* cells were prepared by growing *E. tracheiphila* in 200 ml KB to an OD_{600} of 0.02. Subsequently, cells were washed using decreasing volumes, once with chilled sterile Milli-Q water and twice with 10% glycerol, and finally resuspended in 2 ml of 10% glycerol. For electroporation, a 40-μl aliquot of competent cells was mixed with 4 μl of plasmid DNA, placed in an 0.2-cm cuvette, and electroporated at 2.5 kV for 5.2 to 5.8 ms. Electroporated cells were immediately transferred to 3 ml KB liquid and incubated at room temperature without shaking for 1 h. A cell pellet was obtained, resuspended in 100 μl of medium, and then plated in KB agar with ampicillin (100 μg/ml). Colonies of fluorescent *E. tracheiphila* were obtained after 5 days at room temperature.

Genome assembly and annotation. Adapter trimming and quality filtering of raw Illumina reads were performed using the FastX toolkit 0.0.13.2 (136), SeqTK 1.0 (https://github.com/lh3/seqtk/), and FastQC 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Both mapping and *de novo* assemblies were then generated for each sequenced isolate. For the *de novo* assemblies, SPAdes 3.1.1 was used with default parameters to assemble the quality-filtered, adapter-trimmed, paired-end reads using k-mer sizes of 21, 33, and 55 and the –careful parameter (110). For *ab initio* annotations of the assembled *de novo* whole-genome sequences, Prokka version 1.11 was used with default parameters (111). For the mapping-based assemblies, Mira 4.1 (112) was used to map quality-filtered, adapter-trimmed, paired-end reads from each isolate to the BuffGH PacBio reference strain (30). The functional annotations of all coding sequences (including pseudogenes) were transferred to each genome from the manually curated annotation of the reference BuffGH genome (20), using the RATT function in PAGIT 1.0 (113). We assumed that all pseudogenes are the same in all isolates, which will be confirmable only with long-read PacBio sequencing of these isolates followed by manual annotations.

Phylogenetic relationships of *Erwinia tracheiphila* isolates. Orthologous gene families present in all *E. tracheiphila* isolates were identified from the *de novo* assemblies with OrthoMCL (114) through an all-versus-all BLASTP 2.2.28+ search with an E value cutoff of 10^{-5}. The orthologous genes were aligned using MAFFT 6.853 (115). The gene alignments were trimmed with trimAl version 1.2 using the "automated1" option (116). The individual gene alignments were concatenated into the core genome alignment using the publicly available script at https://doi.org/10.5281/zenodo.1318245 (last accessed 8 September 2014). The 237,634-amino-acid concatenated core genome alignment used to reconstruct the network analysis in Fig. 2A is included in the supplementary file at https://figshare.com/projects/Recent_emergence_of_a_virulent_phytopathogen/35108. The evolutionary relationships among the isolates were reconstructed and visualized in SplitsTree v 4.13.1 (47) using the core genome alignment as input.

Determination of within-cluster diversity. The genes from the manual annotations transferred to the mapped assemblies were used in an all-versus-all BLASTP 2.2.28+ (117) search with an E value cutoff of 10^{-5}. OrthoMCL (114) was run separately for all the isolates within each lineage to identify the core orthologous gene families within each of *Et-melo, Et-C1,* and *Et-C2*. For population genetics analyses, the core genes shared by all isolates within each of the three lineages were designated either Intact, meaning that they are putatively functional based on the manually curated annotations in reference 20, or Pseudogenized/Repetitive, meaning that they either are predicted to be pseudogenes or were predicted to be mobile DNA (genes from bacteriophage, insertion sequences, plasmids, or transposases). The Pseudogenized/Repetitive genes from bacteriophage, insertion sequences, plasmids, or transposases were determined by domain assignments with PfamScan 1.5 (118), ISfinder (January 2015 update) (119), and PHAST (120) as described in reference 20. For *Et-C1* and *Et-melo* clusters sampled at multiple time points, two groups were created: isolates collected from 2008 to 2010 and those collected in 2015. Genetic diversity was quantified for each cluster using Watterson's estimator θ_W per site (121), where θ_W estimates $2N_e\mu$, where N_e is the effective population size and μ is the mutation rate.

For recombination estimates, quality-filtered reads were mapped to the reference BuffGH sequence (30) with the Burrows-Wheeler alignment (BWA) tool 0.7.4 (122), a pileup was created with SAMtools 0.1.18 (123), and variants were called with VCFtools 0.1.9 if the Phred quality score of the variant site was greater than or equal to 60 (124). Single nucleotide polymorphisms (SNPs) were not called if (i) within 9 bp (three codons) of each other and (ii) with less than 10× coverage or (iii) with more than 150× coverage, since short Illumina reads cannot be accurately placed over repetitive regions. Recombination rates within each pathovar were estimated by using the VCF_to_FASTA.sh (see Text S1 in the supplemental material) script to create whole-genome alignments compatible with Gubbins 2.1.0 (125), which was run for a standard 10 iterations.

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

Pangenome identification. The Micropan package (126) in R 3.2 (127) was used to identify the core and pangenome of *de novo E. tracheiphila* isolate assemblies. *De novo* assemblies (see "Genome assembly and annotation" above) were used to ensure that the entire repertoire of genes present per isolate was included, and the pangenome estimates would not be biased with mapping assemblies based on what was present in the reference genome. The groups.txt output file from the OrthoMCL clustering of protein sequences of the *de novo* assemblies (see "Phylogenetic relationships of *Erwinia tracheiphila* isolates" above) and custom R scripts (127) were used to identify genes that were "rare" (present in fewer than 5% of isolates) or "core" (present in more than 95% of the sequenced isolates).

Functional comparison of core and rare genes. The *ab initio*-predicted genes from each *E. tracheiphila* sequenced isolate were searched against the Clusters of Orthologous Groups (COG) database (2014 update) (55) using BLASTP 2.2.28+ (117). Only the top-scoring match (per gene) with an E value of $<10^{-5}$ was kept. Each gene was assigned a COG category of the first functional category of the top-scoring match. Genes without significant matches to any sequence in the COG database were not assigned a functional category. A one-way Fisher exact test with corrections for multiple comparisons was used to identify the COG categories enriched in each cluster and graphed with ggplot2 in R (127).

Identification of T3SS virulence genes and reconstruction of effector gene phylogenetic trees. The *ab initio* coding sequences predicted by Prokka from each *E. tracheiphila* isolate were compared against a manually curated version of the *Pseudomonas* Hop protein effector database (http://www.pseudomonas-syringae.org/T3SS-Hops.xls; accessed 28 August 2015, with additional non-*Pseudomonas* *hrp*T3SS effectors manually added) using BLASTP with an E value cutoff of 10^{-5} (Text S2). The presence and absence of effector genes were visualized with gplots (128) in R 3.2 (127).

To reconstruct the phylogeny of the cluster-specific effector genes identified in *E. tracheiphila*, the amino acid sequence of each gene was used as a BLASTP query against the *nr* database (117). An E value cutoff of 10^{-5} was used to acquire a phylogenetically representative sample of homologs. The sequences were aligned with MAFFT v. 6.853 (115) and trimmed with trimAl 1.2 (116). The maximum-likelihood phylogeny of each aligned gene was reconstructed using RAxML 8.2.4 (129) as implemented on the CIPRES server (130), under the GTR+CAT model and with 100 bootstrap replicates. Bootstrapped pseudoreplicates were summarized with sumTrees.py 4.0.0 (131), and the bootstrap consensus tree was visualized with FigTree 1.4.2 (132).

Cross-inoculation experiments. Seven *E. tracheiphila* isolates from the three different phylogenetic clusters were randomly chosen and used for testing virulence (i.e., the degree of harm) of isolates from each cluster against susceptible host plant species. From *Et-melo*, the experimental isolates were HPsat9.4, RHmelo4.4, and RHsat4.1; from *Et-C1*, the isolates were BHKY, HPcuc8.2, and PAsat11typeII; and from *Et-C2*, the isolate was UVM3.1. Single colonies of each isolate were grown in liquid KB for 24 h until mid-exponential phase, and then all strains were diluted to an OD_{600} of 0.3. For the inoculations, 25 μl of culture from each isolate was then applied to a small break in a single leaf petiole of 2-week-old seedlings at the two-leaf stage. Plants were observed several times weekly for the initial appearance of wilt symptoms in the inoculated leaf, spread of symptoms to a second leaf, and plant death within a 25-day experimental period, according to references 22, 25, 28, 43, 44, 104, and 133. Plant death was scored when all leaves were determined to be too desiccated to support beetle vector herbivory, which is necessary for acquisition of *E. tracheiphila* by beetle vectors and subsequent transmission to healthy hosts. At this stage of infection, the leaves are also too desiccated for photosynthesis. For the statistical analysis, *Et-C1* and *Et-C2* inoculation data were not statistically different, and because they were not different and they share the same host plant range, these were combined. A one-way analysis of variance (ANOVA) with either "days until first wilt symptoms" or "days until death" as the response and "host plant species" was conducted for both *Et-melo* data and *Et-C1* plus *Et-C2* data with model statement aov(lm(Growth~Cluster)) as implemented in R 3.2 (127).

Effects of temperature on *in vitro* growth rate. Twelve *E. tracheiphila* isolates from the three different phylogenetic clusters were randomly chosen for testing the effect of temperature on *in vitro* bacterial growth. The isolates from *Et-C1* are HPcuc8.2, PAsat3.1, PAsat2.3, and BHKY; those from *Et-C2* are ConPepo4M2, ConPepo4M1, and UVM3.1; and those from *Et-melo* are RHmelo2.1, RHmelo4.4, RHsat4.1, PAsat11typeII, and HPsat9.4. The starting cultures were prepared by inoculating a single colony of each *E. tracheiphila* isolate into 4 ml of KB medium, which was grown at room temperature with shaking for 48 h until stationary phase. One milliliter of each culture was then pelleted, washed with 1 ml of fresh KB medium, and resuspended in the same volume. The washed cell suspensions were then diluted in fresh KB medium to an OD_{600} of 0.04, and then four 300-μl replicates of each diluted cell suspension were placed in a single well of an optically clear 96-well plate. The 96-well plate was placed in a plate reader (Spectra Max V2), and absorbance (OD_{600}) was measured every 5 h over a total 40-h experimental period at 28°C, 30°C, 33°C, and 37°C, with noninoculated KB medium used as a negative control. A two-way ANOVA to test the effects of temperature, cluster, and their interaction term on the OD_{600} concentration at different temperatures used the model statement OD_{600} = temperature + cluster + temperature × cluster as implemented in R (127).

Effects of temperate and subtropical temperatures on *in planta* virulence. Single colonies of two *E. tracheiphila* isolates—one derived from a field-infected cucumber (HPcuc8.2, cluster *Et-C1*) and one from a field-infected squash (BHKY, cluster *Et-C1*)—were each grown in liquid KB to an overnight exponential-phase concentration of an OD_{600} of 0.3. The squash origin isolate was inoculated into 2-week-old squash seedlings at the two-leaf stage, and the cucumber origin isolate was inoculated into 2-week-old cucumber plants at the two-leaf stage. All plants were inoculated with 25 μl of bacterial inoculum placed on a single petiole wound. The seed varieties used were *Cucurbita pepo* 'Dixie' squash and *Cucumis sativus* 'Marketmore' cucumber from Johnny's Selected Seeds (Winslow, ME).

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

Average July temperatures for Texas and Massachusetts were determined by a Google search to be 33°C for day and 23°C for night for Texas and 27°C for day and 18°C for night for Massachusetts. All plants were kept in programmable Conviron growth chambers with a 16-h-light/8-h-dark cycle and 60% relative humidity. Plants were observed several times weekly for the initial appearance of wilt symptoms in the inoculated leaf, spread of symptoms to a second leaf, and plant death within a 25-day experimental period, according to references 22, 25, 28, 43, 44, 104, and 133. The sample sizes used in this experiment are $n = 44$ for Texas 'Dixie' squash, $n = 44$ for Massachusetts 'Dixie' squash, $n = 34$ for Massachusetts 'Marketmore' cucumbers, and $n = 36$ for Texas 'Marketmore' cucumbers. A one-way ANOVA to test the effects of host species (either cucumber or squash) at both Texas and Massachusetts temperatures used the model statements death = state + host species and wilt = state + host species as implemented in R (127).

Data availability. Raw reads from the sequenced isolates (Table S1) are available at the NCBI BioProject PRJNA272881, SRA no. SRP056142. The sequence filtering and analysis pipeline, Micropan parameters for pangenome analysis, modified Hop *hrp*T3SS database, and 'VCF_to_FASTA.sh' script used to create FASTA alignments of variant calls for recombination analysis in Gubbins are available via Figshare Project 35108 (https://figshare.com/projects/Recent_emergence_of_a_virulent_phytopathogen/35108). The concatenated core genome alignment file (237,634 amino acids) used to reconstruct the network analysis in Fig. 2A and Fig. S1 can be found at https://figshare.com/projects/Recent_emergence_of_a_virulent_phytopathogen/35108.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.01307-18.

TEXT S1, TXT file, 0.1 MB.
TEXT S2, TXT file, 0.2 MB.
FIG S1, PDF file, 0.2 MB.
FIG S2, PDF file, 0.1 MB.
TABLE S1, XLSX file, 0.1 MB.
TABLE S2, XLS file, 0.05 MB.

ACKNOWLEDGMENTS

This study was made possible by NSF postdoctoral fellowship DBI-1202736 to L.R.S. and NIH grant GM58213 to R.K.; O.Z. was in part supported by a Simons Investigator Award from the Simons Foundation.

We thank Bob Freedman and Aaron Kitzmiller at Harvard FAS for computational advice and support; Harvard Odyssey Computational Resources; Taj Azarian for helpful advice and discussion; Scott Chimileski for the Fig. 1B image; Rob Dunn for constructive comments on the manuscript; the Weld Research Building at the Arnold Arboretum for staff support, confocal microscope access, and growth chamber facilities; Miguel Coehlo for Nextera library preparation help; and Eric Alm for providing the original suggestion to sequence dozens of isolates. Infected plant samples were provided by Christian Herter Community Garden, Verrill Farms, Dana Roberts, Kristy's Barn, the University of Vermont Horticultural Farm, and Green Valley Farm. We thank them and all other farms and individuals who contributed isolates. We thank Nora Mishanec and Laura Jenny for laboratory and greenhouse assistance.

Mention of commercial products and organizations in this work is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned.

L.R.S. conceived of the study; R.K. supervised the development of the project; L.R.S., J.N.P., B.J.A., E.D.S., and K.H. performed computational analysis; L.R.S. and J.R. did laboratory and greenhouse experiments; all authors contributed to the interpretation of data; L.R.S. drafted the original article; and J.N.P., B.J.A., E.D.S., O.Z., N.E.P., J.R., V.K.C., K.H., and R.K. contributed revisions.

REFERENCES

1. Mira A, Pushker R, Rodríguez-Valera F. 2006. The Neolithic revolution of bacterial genomes. Trends Microbiol 14:200–206. https://doi.org/10.1016/j.tim.2006.03.001.
2. Wolfe ND, Dunavan CP, Diamond J. 2007. Origins of major human infectious diseases. Nature 447:279–283. https://doi.org/10.1038/nature05775.
3. Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E, Beerli P, Su X-Z. 2003. Early origin and recent

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

expansion of *Plasmodium falciparum*. Science 300:318–321. https://doi.org/10.1126/science.1081449.
4. Roberts L. 1989. Disease and death in the New World. Science 246:1245. https://doi.org/10.1126/science.2479984.
5. Achtman M. 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol 62:53–70. https://doi.org/10.1146/annurev.micro.62.081307.162832.
6. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008. Global trends in emerging infectious diseases. Nature 451:990–993. https://doi.org/10.1038/nature06536.
7. Taylor LH, Latham SM, Woolhouse MEJ. 2001. Risk factors for human disease emergence. Philos Trans R Soc Lond B Biol Sci 356:983–989. https://doi.org/10.1098/rstb.2001.0888.
8. Joly D, Johnson CK, Goldstein T, Anthony S, Karesh W, Daszak P, Wolfe N, Murray S, Mazet J. 2016. The first phase of PREDICT: surveillance for emerging infectious zoonotic diseases of wildlife origin (2009–2014). Int J Infect Dis 53:31–32. https://doi.org/10.1016/j.ijid.2016.11.086.
9. Vitousek PM, Mooney HA, Lubchenco J, Melillo JM. 1997. Human domination of Earth's ecosystems. Science 277:494–499. https://doi.org/10.1126/science.277.5325.494.
10. Zhu Y, Chen H, Fan J, Wang Y, Li Y, Chen J, Fan J, Yang S, Hu L, Leung H, Mew TW, Teng PS, Wang Z, Mundt CC. 2000. Genetic diversity and disease control in rice. Nature 406:718–722. https://doi.org/10.1038/35021046.
11. Roossinck MJ, García-Arenal F. 2015. Ecosystem simplification, biodiversity loss and plant virus emergence. Curr Opin Virol 10:56–62. https://doi.org/10.1016/j.coviro.2015.01.005.
12. Mann CC. 2011. 1493: uncovering the new world Columbus created. Knopf Doubleday Publishing Group, New York, NY.
13. Khoury CK, Achicanoy HA, Bjorkman AD, Navarro-Racines C, Guarino L, Flores-Palacios X, Engels JMM, Wiersema JH, Dempewolf H, Sotelo S, Ramírez-Villegas J, Castañeda-Álvarez NP, Fowler C, Jarvis A, Rieseberg LH, Struik PC. 2016. Origins of food crops connect countries worldwide. Proc R Soc B Biol Sci 283:20160792. https://doi.org/10.1098/rspb.2016.0792.
14. Chen YH. 2016. Crop domestication, global human-mediated migration, and the unresolved role of geography in pest control. Elementa 4:000106. https://doi.org/10.12952/journal.elementa.000106.
15. Power AG, Mitchell CE. 2004. Pathogen spillover in disease epidemics. Am Nat 164:S79–S89. https://doi.org/10.1086/424610.
16. McCann HC, Rikkerink EHA, Bertels F, Fiers M, Lu A, Rees-George J, Andersen MT, Gleave AP, Haubold B, Wohlers MW, Guttman DS, Wang PW, Straub C, Vanneste JL, Vanneste J, Rainey PB, Templeton MD. 2013. Genomic analysis of the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae* provides insight into the origins of an emergent plant disease. PLoS Pathog 9:e1003503. https://doi.org/10.1371/annotation/af157ddc-200a-4105-b243-3f01251cc677.
17. Savory EA, Fuller SL, Weisberg AJ, Thomas WJ, Gordon MI, Stevens DM, Creason AL, Belcher MS, Serdani M, Wiseman MS, Grünwald NJ, Putnam ML, Chang JH. 2017. Evolutionary transitions between beneficial and phytopathogenic rhodococcus challenge disease management. Elife 6:e30925. https://doi.org/10.7554/eLife.30925.
18. Alexander HM, Bruns E, Schebor H, Malmstrom CM. 2017. Crop-associated virus infection in a native perennial grass: reduction in plant fitness and dynamic patterns of virus detection. J Ecol 105:1021–1031. https://doi.org/10.1111/1365-2745.12723.
19. Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr Opin Genet Dev 14:627–633. https://doi.org/10.1016/j.gde.2004.09.003.
20. Shapiro LR, Scully ED, Straub TJ, Park J, Stephenson AG, Beattie GA, Gleason ML, Kolter R, Coelho MC, Moraes CMD, Mescher MC, Zhaxybayeva O. 2016. Horizontal gene acquisitions, mobile element proliferation, and genome decay in the host-restricted plant pathogen *Erwinia tracheiphila*. Genome Biol Evol 8:649–664. https://doi.org/10.1093/gbe/evw016.
21. Yao C, Zehnder G, Bauske E, Kloepper J. 1996. Relationship between cucumber beetle (Coleoptera: Chrysomelidae) density and incidence of bacterial wilt of cucurbits. J Econ Entomol 89:510–514. https://doi.org/10.1093/jee/89.2.510.
22. Brust GE. 1997. Interaction of *Erwinia tracheiphila* and muskmelon plants. Environ Entomol 26:849. https://doi.org/10.1093/ee/26.4.849.
23. Brust GE. 1997. Seasonal variation in percentage of striped cucumber beetles (Coleoptera: Chrysomelidae) that vector *Erwinia tracheiphila*. Environ Entomol 26:580–584. https://doi.org/10.1093/ee/26.3.580.
24. Brust GE, Foster RE. 1999. New economic threshold for striped cucumber beetle (Coleoptera: Chrysomelidae) in cantaloupe in the Midwest. J Econ Entomol 92:936–940. https://doi.org/10.1093/jee/92.4.936.
25. Smith EF. 1920. An introduction to bacterial diseases of plants. WB Saunders, Philadelphia, PA.
26. Rand FV, Enlows EMA. 1920. Bacterial wilt of cucurbits. US Department of Agriculture, Washington, DC.
27. Rand FV, Enlows EMA. 1920. Further studies of bacterial wilt of cucurbits. US Department of Agriculture bulletin. US Department of Agriculture, Washington, DC.
28. Vrisman CM, Deblais L, Rajashekara G, Miller SA. 2016. Differential colonization dynamics of cucurbit hosts by *Erwinia tracheiphila*. Phytopathology 106:684–692. https://doi.org/10.1094/PHYTO-11-15-0289-R.
29. Rojas ES, Batzer JC, Beattie GA, Fleischer SJ, Shapiro LR, Williams MA, Bessin R, Bruton BD, Boucher TJ, Jesse LCH, Gleason ML. 2015. Bacterial wilt of cucurbits: resurrecting a classic pathosystem. Plant Dis 99:564–574. https://doi.org/10.1094/PDIS-10-14-1068-FE.
30. Shapiro LR, Scully ED, Roberts D, Straub TJ, Geib SM, Park J, Stephenson AG, Rojas ES, Liu Q, Beattie G, Gleason M, Moraes CMD, Mescher MC, Fleischer SJ, Kolter R, Pierce N, Zhaxybayeva O. 2015. Draft genome sequence of *Erwinia tracheiphila*, an economically important bacterial pathogen of cucurbits. Genome Announc 3:e00482-15. https://doi.org/10.1128/genomeA.00482-15.
31. Nee M. 1990. The domestication of *Cucurbita* (Cucurbitaceae). Econ Bot 44:56–68. https://doi.org/10.1007/BF02860475.
32. Sun H, Wu S, Zhang G, Jiao C, Guo S, Ren Y, Zhang J, Zhang H, Gong G, Jia Z, Zhang F, Tian J, Lucas WJ, Doyle JJ, Li H, Fei Z, Xu Y. 2017. Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. Mol Plant 10:1293–1306. https://doi.org/10.1016/j.molp.2017.09.003.
33. Sebastian P, Schaefer H, Telford IRH, Renner SS. 2010. Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. Proc Natl Acad Sci U S A 107:14269–14273. https://doi.org/10.1073/pnas.1005338107.
34. Bisognin DA. 2002. Origin and evolution of cultivated cucurbits. Cienc Rural 32:715–723. https://doi.org/10.1590/S0103-84782002000400028.
35. Decker DS. 1988. Origin(s), evolution, and systematics of *Cucurbita pepo* (Cucurbitaceae). Econ Bot 42:4–15. https://doi.org/10.1007/BF02859022.
36. Whitaker TW, Bemis W. 1975. Origin and evolution of the cultivated *Cucurbita*. Bull Torrey Bot Club 102:362–368. https://doi.org/10.2307/2484762.
37. Hurd PD, Linsley EG, Whitaker T. 1971. Squash and gourd bees (*Peponapis, Xenoglossa*) and the origin of the cultivated *Cucurbita*. Evolution 25:218. https://doi.org/10.1111/j.1558-5646.1971.tb01874.x.
38. Newsom LA, Trieu DA. 2011. Fusion gardens: native North America and the Columbian Exchange, p 557–576. *In* Smith B (ed), Subsistence economies of indigenous North American societies: a handbook. Rowman and Littlefield Publishing Group, Lanham, MD.
39. Rand FV, Enlows E. 1916. Transmission and control of bacterial wilt of cucurbits. J Agric Res 6:417–434.
40. Gould GE. 1944. The biology and control of the striped cucumber beetle. Bull Purdue Univ Agric Exp Stn 490:1–28.
41. Rand FV, Cash LC. 1920. Some insect relations of Bacillus tracheiphilus Erw. Sm. Phytopathology 10:133–140.
42. Sasu M, Seidl-Adams I, Wall K, Winsor J, Stephenson A. 2010. Floral transmission of *Erwinia tracheiphila* by cucumber beetles in a wild *Cucurbita pepo*. Environ Entomol 39:140–148. https://doi.org/10.1603/EN09190.
43. Shapiro L, De Moraes CM, Stephenson AG, Mescher MC. 2012. Pathogen effects on vegetative and floral odours mediate vector attraction and host exposure in a complex pathosystem. Ecol Lett 15:1430–1438. https://doi.org/10.1111/ele.12001.
44. Shapiro LR, Salvaudon L, Mauck KE, Pulido H, De Moraes CM, Stephenson AG, Mescher MC. 2013. Disease interactions in a shared host plant: effects of pre-existing viral infection on cucurbit plant defense responses and resistance to bacterial wilt disease. PLoS One 8:e77393. https://doi.org/10.1371/journal.pone.0077393.
45. Shapiro LR, Seidl-Adams I, De Moraes C, Stephenson A, Mescher M. 2014. Dynamics of short- and long-term association between a bacterial plant pathogen and its arthropod vector. Sci Rep 4:4155. https://doi.org/10.1038/srep04155.
46. Huson DH, Bryant D. 2006. Application of phylogenetic networks in

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

evolutionary studies. Mol Biol Evol 23:254–267. https://doi.org/10.1093/molbev/msj030.

47. Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73. https://doi.org/10.1093/bioinformatics/14.1.68.
48. Vauterin L, Swings J. 1997. Are classification and phytopathological diversity compatible in *Xanthomonas*? J Ind Microbiol Biotechnol 19:77–82. https://doi.org/10.1038/sj.jim.2900433.
49. Sawada H, Suzuki F, Matsuda I, Saitou N. 1999. Phylogenetic analysis of *Pseudomonas syringae* pathovars suggests the horizontal gene transfer of *argK* and the evolutionary stability of *hrp* gene cluster. J Mol Evol 49:627–644. https://doi.org/10.1007/PL00006584.
50. Sarkar SF, Guttman DS. 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. Appl Environ Microbiol 70:1999–2012. https://doi.org/10.1128/AEM.70.4.1999-2012.2004.
51. Rojas ES, Dixon PM, Batzer JC, Gleason ML. 2013. Genetic and virulence variability among *Erwinia tracheiphila* strains recovered from different cucurbit hosts. Phytopathology 103:900–905. https://doi.org/10.1094/PHYTO-11-12-0301-R.
52. Sanjur OI, Piperno DR, Andres TC, Wessel-Beaver L. 2002. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: implications for crop plant evolution and areas of origin. Proc Natl Acad Sci U S A 99:535–540. https://doi.org/10.1073/pnas.012577299.
53. Biller SJ, Berube PM, Lindell D, Chisholm SW. 2015. *Prochlorococcus*: the structure and function of collective diversity. Nat Rev Microbiol 13:13. https://doi.org/10.1038/nrmicro3378.
54. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW. 2015. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics 15:141–161. https://doi.org/10.1007/s10142-015-0433-4.
55. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261. https://doi.org/10.1093/nar/gku1223.
56. Vlot AC, Dempsey DMA, Klessig DF. 2009. Salicylic acid, a multifaceted hormone to combat disease. Annu Rev Phytopathol 47:177–206. https://doi.org/10.1146/annurev.phyto.050908.135202.
57. Nissinen RM, Ytterberg JA, Bogdanove AJ, van Wijk KJ, Beer SV. 2007. Analyses of the secretomes of *Erwinia amylovora* and selected hrp mutants reveal novel type III secreted proteins and an effect of HrpJ on extracellular harpin levels. Mol Plant Pathol 8:55–67. https://doi.org/10.1111/j.1364-3703.2006.00370.x.
58. Bocsanczy AM, Schneider DJ, DeClerck GA, Cartinhour S, Beer SV. 2012. HopX1 in *Erwinia amylovora* functions as an avirulence protein in apple and is regulated by HrpL. J Bacteriol 194:553–560. https://doi.org/10.1128/JB.05065-11.
59. Petty NK, Feltwell T, Pickard D, Clare S, Toribio AL, Fookes M, Roberts K, Monson R, Nair S, Kingsley RA, Bulgin R, Wiles S, Goulding D, Keane T, Corton C, Lennard N, Harris D, Willey D, Rance R, Yu L, Choudhary JS, Churcher C, Quail MA, Parkhill J, Frankel G, Dougan G, Salmond GPC, Thomson NR. 2011. *Citrobacter rodentium* is an unstable pathogen showing evidence of significant genomic flux. PLoS Pathog 7:e1002018. https://doi.org/10.1371/journal.ppat.1002018.
60. Goel AK, Lundberg D, Torres MA, Matthews R, Akimoto-Tomiyama C, Farmer L, Dangl JL, Grant SR. 2008. The *Pseudomonas syringae* type III effector HopAM1 enhances virulence on water-stressed plants. Mol Plant Microbe Interact 21:361–370. https://doi.org/10.1094/MPMI-21-3-0361.
61. Washington EJ, Mukhtar MS, Finkel OM, Wan L, Banfield MJ, Kieber JJ, Dangl JL. 2016. Pseudomonas syringae type III effector HopAF1 suppresses plant immunity by targeting methionine recycling to block ethylene induction. Proc Natl Acad Sci U S A 113:E3577–E3586. https://doi.org/10.1073/pnas.1606322113.
62. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, Connor TR, Harris SR, Fairley D, Bamford KB, D'Arc S, Brazier J, Brown D, Coia JE, Douce G, Gerding D, Kim HJ, Koh TH, Kato H, Senoh M, Louie T, Michell S, Butt E, Peacock SJ, Brown NM, Riley T, Songer G, Wilcox M, Pirmohamed M, Kuijper E, Hawkey P, Wren BW, Dougan G, Parkhill J, Lawley TD. 2013. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. Nat Genet 45:109–113. https://doi.org/10.1038/ng.2478.
63. Stukenbrock EH, Bataillon T. 2012. A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. PLoS Pathog 8:e1002893. https://doi.org/10.1371/journal.ppat.1002893.
64. McDonald BA, Stukenbrock EH. 2016. Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. Philos Trans R Soc B Biol Sci 371:20160026. https://doi.org/10.1098/rstb.2016.0026.
65. Foley JA, Defries R, Asner GP, Barford C, Bonan G, Carpenter SR, Chapin FS, Coe MT, Daily GC, Gibbs HK, Helkowski JH, Holloway T, Howard EA, Kucharik CJ, Monfreda C, Patz JA, Prentice IC, Ramankutty N, Snyder PK. 2005. Global consequences of land use. Science 309:570–574. https://doi.org/10.1126/science.1111772.
66. Petersen JB, Sidell NA. 1996. Mid-Holocene evidence of *Cucurbita* sp. from Central Maine. Am Antiq 61:685–698. https://doi.org/10.2307/282011.
67. Selby AD. 1899. Further studies of cucumber, melon and tomato diseases, with experiments. Ohio Agricultural Experiment Station, Wooster, OH.
68. Smith EF. 1896. The path of the water current in cucumber plants (continued). Am Nat 30:554–562. https://doi.org/10.1086/276429.
69. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MTG, Churcher CM, Bentley SD, Mungall KL, Cerdeño-Tárraga AM, Temple L, James K, Harris B, Quail MA, Achtman M, Atkin R, Baker S, Basham D, Bason N, Cherevach I, Chillingworth T, Collins M, Cronin A, Davis P, Doggett J, Feltwell T, Goble A, Hamlin N, Hauser H, Holroyd S, Jagels K, Leather S, Moule S, Norberczak H, O'Neil S, Ormond D, Price C, Rabbinowitsch E, Rutter S, Sanders M, Saunders D, Seeger K, Sharp S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Unwin L, Whitehead S, Barrell BG, Maskell DJ. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis, Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat Genet 35:32–40. https://doi.org/10.1038/ng1227.
70. Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A 96:14043–14048. https://doi.org/10.1073/pnas.96.24.14043.
71. Cleaveland S, Haydon D, Taylor L. 2007. Overviews of pathogen emergence: which pathogens emerge, when and why?, p 85–111. *In* Childs JE, Mackenzie JS, Richt JA (ed), Wildlife and emerging zoonotic diseases: the biology, circumstances and consequences of cross-species transmission. Springer, Berlin, Germany.
72. Ellis J, Dodds P, Pryor T. 2000. Structure, function and evolution of plant disease resistance genes. Curr Opin Plant Biol 3:278–284. https://doi.org/10.1016/S1369-5266(00)00080-7.
73. Thrall PH, Burdon JJ. 2003. Evolution of virulence in a plant host-pathogen metapopulation. Science 299:1735–1737. https://doi.org/10.1126/science.1080070.
74. Wanderley-Nogueira A, Bezerra-Neto J, Kido E, Araujo F, Amorim L, Crovella S, Benko-Iseppon A. 2017. Plant elite squad: first defense line and resistance genes—identification, diversity and functional roles. Curr Protein Peptide Sci 18:294–310. https://doi.org/10.2174/1389203717666160724193045.
75. Wei G, Tian P, Zhang F, Qin H, Miao H, Chen Q, Hu Z, Cao L, Wang M, Gu X. 2016. Integrative analyses of non-targeted volatile profiling and transcriptome data provide molecular insight into VOC diversity in cucumber plants (*Cucumis sativus* L). Plant Physiol 172:603–618. https://doi.org/10.1104/pp.16.01051.
76. Lv J, Qi J, Shi Q, Shen D, Zhang S, Shao G, Li H, Sun Z, Weng Y, Shang Y, Gu X, Li X, Zhu X, Zhang J, van Treuren R, van Dooijeweert W, Zhang Z, Huang S. 2012. Genetic diversity and population structure of cucumber (*Cucumis sativus* L.). PLoS One 7:e46919. https://doi.org/10.1371/journal.pone.0046919.
77. Tabashnik BE, Brévault T, Carrière Y. 2013. Insect resistance to Bt crops: lessons from the first billion acres. Nat Biotechnol 31:510–521. https://doi.org/10.1038/nbt.2597.
78. Gassmann AJ, Petzold-Maxwell JL, Keweshan RS, Dunbar MW. 2011. Field-evolved resistance to Bt maize by western corn rootworm. PLoS One 6:e22629. https://doi.org/10.1371/journal.pone.0022629.
79. Vinatzer BA, Monteil CL, Clarke CR. 2014. Harnessing population genomics to understand how bacterial pathogens emerge, adapt to crop hosts, and disseminate. Annu Rev Phytopathol 52:19–43. https://doi.org/10.1146/annurev-phyto-102313-045907.
80. Pinel-Galzi A, Traoré O, Séré Y, Hébrard E, Fargette D. 2015. The biogeog-

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

raphy of viral emergence: rice yellow mottle virus as a case study. Curr Opin Virol 10:7–13. https://doi.org/10.1016/j.coviro.2014.12.002.
81. Krysan JL, Miller TA. 1986. Methods for the study of pest Diabrotica. Springer Verlag, New York, NY.
82. Cai R, Lewis J, Yan S, Liu H, Clarke CR, Campanile F, Almeida NF, Studholme DJ, Lindeberg M, Schneider D, Zaccardelli M, Setubal JC, Morales-Lizcano NP, Bernal A, Coaker G, Baker C, Bender CL, Leman S, Vinatzer BA. 2011. The plant pathogen *Pseudomonas syringae* pv. *tomato* is genetically monomorphic and under strong selection to evade tomato immunity. PLoS Pathog 7:e1002130. https://doi.org/10.1371/journal.ppat.1002130.
83. Lefeuvre P, Martin DP, Harkins G, Lemey P, Gray AJA, Meredith S, Lakay F, Monjane A, Lett J-M, Varsani A, Heydarnejad J. 2010. The spread of tomato yellow leaf curl virus from the Middle East to the world. PLoS Pathog 6:e1001164. https://doi.org/10.1371/journal.ppat.1001164.
84. Goodwin SB, Cohen BA, Fry WE. 1994. Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. Proc Natl Acad Sci U S A 91:11591–11595. https://doi.org/10.1073/pnas.91.24.11591.
85. Gladieux P, Ravel S, Rieux A, Cros-Arteil S, Adreit H, Milazzo J, Thierry M, Fournier E, Terauchi R, Tharreau D. 2018. Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. mBio 9:e01806-17. https://doi.org/10.1128/mBio.01806-17.
86. Bates DM, Robinson RW, Jeffrey C. 1990. Biology and utilization of the Cucurbitaceae. Cornell University Press, New York, NY.
87. Shapiro LR, Mauck KE. 2018. Chemically-mediated interactions among cucurbits, insects and microbes, p 55–90. *In* Tabata J (ed), Chemical ecology of insects. CRC Press, Boca Raton, FL.
88. Ferguson JE, Metcalf RL. 1985. Cucurbitacins. J Chem Ecol 11:311–318. https://doi.org/10.1007/BF01411417.
89. Metcalf R, Lampman R. 1989. The chemical ecology of diabroticites and cucurbitaceae. Experientia 45:240–247. https://doi.org/10.1007/BF01951810.
90. Tallamy DW, Stull J, Ehresman NP, Gorski PM, Mason CE. 1997. Cucurbitacins as feeding and oviposition deterrents to insects. Environ Entomol 26:678–683. https://doi.org/10.1093/ee/26.3.678.
91. Metcalf RL, Metcalf RA, Rhodes AM. 1980. Cucurbitacins as kairomones for diabroticite beetles. Proc Natl Acad Sci U S A 77:3769–3772. https://doi.org/10.1073/pnas.77.7.3769.
92. Kistler L, Newsom LA, Ryan TM, Clarke AC, Smith BD, Perry GH. 2015. Gourds and squashes (*Cucurbita* spp.) adapted to megafaunal extinction and ecological anachronism through domestication. Proc Natl Acad Sci U S A 112:15107–15112. https://doi.org/10.1073/pnas.1516109112.
93. Chambliss OL, Jones CM. 1966. Cucurbitacins: specific insect attractants in Cucurbitaceae. Science 153:1392–1393. https://doi.org/10.1126/science.153.3742.1392.
94. Andersen JF, Plattner RD, Weisleder D. 1988. Metabolic transformations of cucurbitacins by *Diabrotica virgifera virgifera* Leconte and *D. undecimpunctata howardi* Barber. Insect Biochem 18:71–77. https://doi.org/10.1016/0020-1790(88)90038-8.
95. Barber H. 1946. Diabrotica and two new genera (Coleoptera, Chrysomelidae). Proc Entomol Soc Wash 49:151–161.
96. Munroe DD, Smith RF. 1980. A revision of the systematics of *Acalymma* sensu stricto Barber (Coleoptera: Chrysomelidae) from North America including Mexico. Mem Entomol Soc Can 112:1–92. https://doi.org/10.4039/entm112112fv.
97. Tallamy DW, Gorski PM. 1997. Long-and short-term effect of cucurbitacin consumption on *Acalymma vittatum* (Coleoptera: Chrysomelidae) fitness. Environ Entomol 26:672–677. https://doi.org/10.1093/ee/26.3.672.
98. de Mackiewicz D, Gildow FE, Blua M, Fleischer SJ, Lukezic FL. 1998. Herbaceous weeds are not ecologically important reservoirs of *Erwinia tracheiphila*. Plant Dis 82:521–529. https://doi.org/10.1094/PDIS.1998.82.5.521.
99. López-Uribe MM, Cane JH, Minckley RL, Danforth BN. 2016. Crop domestication facilitated rapid geographical expansion of a specialist pollinator, the squash bee *Peponapis pruinosa*. Proc Biol Sci 283: 20160443. https://doi.org/10.1098/rspb.2016.0443.
100. Joseph JG, Douglas WT, Edward GR, Anthony IC. 2008. Molecular phylogeny of rootworms and related galerucine beetles (Coleoptera: Chrysomelidae). Zool Scripta 37:195–222. https://doi.org/10.1111/j.1463-6409.2007.00320.x.
101. Eben A, Espinosa de Los Monteros A. 2013. Tempo and mode of evolutionary radiation in Diabroticina beetles (genera *Acalymma*, *Cerotoma*, and *Diabrotica*). Zookeys (332):207–231. https://doi.org/10.3897/zookeys.332.5220.
102. Rand FV. 1915. Dissemination of bacterial wilt of cucurbits. J Agric Res 5:257–260.
103. Leach JG. 1964. Observations on cucumber beetles as vectors of cucurbit wilt. Phytopathology 54:606–607.
104. Brust GE. 1997. Differential susceptibility of pumpkins to bacterial wilt related to plant growth stage and cultivar. Crop Prot 16:411–414. https://doi.org/10.1016/S0261-2194(97)00020-3.
105. Fleischer SJ, de Mackiewicz D, Gildow FE, Lukezic FL. 1999. Serological estimates of the seasonal dynamics of *Erwinia tracheiphila* in *Acalymma vittata* (Coleoptera: Chrysomelidae). Environ Entomol 28:470–476. https://doi.org/10.1093/ee/28.3.470.
106. Garcia-Salazar C, Gildow FE, Fleischer SJ, Cox-Foster D, Lukezic FL. 2000. ELISA versus immunolocalization to determine the association of *Erwinia tracheiphila* in *Acalymma vittatum* (Cleoptera: Chrysomelidae). Environ Entomol 29:542–550. https://doi.org/10.1603/0046-225X-29.3.542.
107. Sanogo S, Etarock BF, Clary M. 2011. First report of bacterial wilt caused by *Erwinia tracheiphila* on pumpkin and watermelon in New Mexico. Plant Dis 95:1583–1583. https://doi.org/10.1094/PDIS-06-11-0507.
108. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
109. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645. https://doi.org/10.1073/pnas.120163297.
110. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
111. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
112. Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer Science and Biology. Proceedings of the German Conference on Bioinformatics, GCB '99, Hannover, Germany.
113. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc 7:1260–1284. https://doi.org/10.1038/nprot.2012.068.
114. Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189. https://doi.org/10.1101/gr.1224503.
115. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436.
116. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
117. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
118. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. The pfam protein families database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223.
119. Siguier P, Pérochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32–D36. https://doi.org/10.1093/nar/gkj014.
120. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:W347–W352. https://doi.org/10.1093/nar/gkr485.
121. Watterson G. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276. https://doi.org/10.1016/0040-5809(75)90020-9.
122. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
123. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest

and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

124. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.
125. Hou S, Mu R, Ma G, Xu X, Zhang C, Yang Y, Wu D. 2011. *Pseudomonas syringae* pv. *phaseolicola* effector HopF1 inhibits pathogen-associated molecular pattern-triggered immunity in a RIN4-independent manner in common bean (*Phaseolus vulgaris*). FEMS Microbiol Lett 323:35–43. https://doi.org/10.1111/j.1574-6968.2011.02356.x.
126. Snipen L, Liland KH. 2015. Micropan: an R-package for microbial pan-genomics. BMC Bioinformatics 16:79. https://doi.org/10.1186/s12859-015-0517-0.
127. R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
128. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M. 2015. gplots: various R programming tools for plotting data. R package version 2.17.0. https://cran.r-project.org/web/packages/gplots/index.html.
129. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690. https://doi.org/10.1093/bioinformatics/btl446.
130. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, p 1–8. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA.
131. Sukumaran J, Holder MT. 2010. DendroPy: a python library for phylogenetic computing. Bioinformatics 26:1569–1571. https://doi.org/10.1093/bioinformatics/btq228.
132. Rambaut A. 2006. FigTree, a graphical viewer of phylogenetic trees and as a program for producing publication-ready figures. http://tree.bio.ed.ac.uk/software/figtree/.
133. Sasu MA, Ferrari MJ, Du D, Winsor JA, Stephenson AG. 2009. Indirect costs of a nontarget pathogen mitigate the direct benefits of a virus-resistant transgene in wild *Cucurbita*. Proc Natl Acad Sci U S A 106:19067–19071. https://doi.org/10.1073/pnas.0905106106.
134. Hartl DL, Clark AG. 1997. Principles of population genetics. Sinauer Associates, Sunderland, MA.
135. Lemieux JE, Tran AD, Freimark L, Schaffner SF, Goethert H, Andersen KG, Bazner S, Li A, McGrath G, Sloan L, Vannier E, Milner D, Pritt B, Rosenberg E, Telford S, Bailey JA, Sabeti PC. 2016. A global map of genetic diversity in *Babesia microti* reveals strong population structure and identifies variants associated with clinical relapse. Nat Microbiol 1:16079. https://doi.org/10.1038/nmicrobiol.2016.79.
136. Pearson WR, Wood T, Zhang Z, Miller W. 1997. Comparison of DNA sequences with protein sequences. Genomics 46:24–36. https://doi.org/10.1006/geno.1997.4995.

Downloaded from http://mbio.asm.org/ on February 25, 2019 by guest