Cruciferous weed isolates of Xanthomonas campestris yield insight into pathovar genomic relationships and genetic determinants of host- and tissue-specificity

Zoe Dubrow, Sara Carpenter, Morgan Carter, Ayress Grinage, Carine Gris, E. Lauber, Jules Butchacas, Jonathan Jacobs, Christine Smart, Matthew Tancos, et al.

MPMI Vol. 35, No. 9, 2022, pp. 791–802, https://doi.org/10.1094/MPMI-01-22-0024-R

Cruciferous Weed Isolates of *Xanthomonas campestris* Yield Insight into Pathovar Genomic Relationships and Genetic Determinants of Host and Tissue Specificity

Zoë E. Dubrow,[1] Sara C. D. Carpenter,[1] Morgan E. Carter,[1,2] Ayress Grinage,[3] Carine Gris,[4] Emmanuelle Lauber,[4] Jules Butchachas,[5] Jonathan M. Jacobs,[5] Christine D. Smart,[1] Matthew A. Tancos,[6] Laurent D. Noël,[4] and Adam J. Bogdanove[1,†]

[1] Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, U.S.A.
[2] School of Plant Sciences, University of Arizona, Tucson, AZ, U.S.A.
[3] Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, U.S.A.
[4] LIPME, Université de Toulouse, INRAE, CNRS, Université Paul Sabatier, Castanet-Tolosan, France
[5] Department of Plant Pathology, The Ohio State University, Columbus, OH, U.S.A.
[6] Foreign Disease-Weed Science Research Unit, United States Department of Agriculture-Agricultural Research Service, Frederick, MD, U.S.A.

Accepted for publication 3 May 2022.

Pathovars of *Xanthomonas campestris* cause distinct diseases on different brassicaceous hosts. The genomic relationships among pathovars as well as the genetic determinants of host range and tissue specificity remain poorly understood despite decades of research. Here, leveraging advances in multiplexed long-read technology, we fully sequenced the genomes of a collection of *X. campestris* strains isolated from cruciferous crops and weeds in New York and California as well as strains from global collections, to investigate pathovar relationships and candidate genes for host- and tissue-specificity. Pathogenicity assays and genomic comparisons across this collection and publicly available *X. campestris* genomes revealed a correlation between pathovar and genomic relatedness and provide support for *X. campestris* pv. *barbareae*, the validity of which had been questioned. Linking strain host range with type III effector repertoires identified AvrAC (also 'XopAC') as a candidate host-range determinant, preventing infection of *Matthiola incana*, and this was confirmed experimentally. Furthermore, the presence of a copy of the cellobiosidase gene *cbsA* with coding sequence for a signal peptide was found to correlate with the ability to infect vascular tissues, in agreement with a previous study of diverse *Xanthomonas* species; however, heterologous expression in strains lacking the gene gave mixed results, indicating that factors in addition to *cbsA* influence tissue specificity of *X. campestris* pathovars.

Keywords: AvrAC, black rot, *Brassica*, cabbage, cellobiosidase, pathovar, type III effectors, *Xanthomonas campestris*

Xanthomonas campestris is a destructive species of plant-pathogenic bacteria causing disease in brassicaceous plants (e.g., cabbage, cruciferous weeds, and ornamentals). The species comprises multiple pathogenic variants (pathovars) that cause distinct diseases on different plant hosts. Within the genus *Xanthomonas*, multiple changes to species and pathovar nomenclature have been made and other changes have been proposed, reflecting sometimes conflicting views on what pathovars belong in *X. campestris*. Vauterin and colleagues (1995) restricted *Xanthomonas campestris* to only pathovars infecting *Brassica* hosts (pathovars *campestris*, *raphani*, *incanae*, *barbareae*, *armoraciae*, and *abberans*). However, recent phylogenetic data and pathogenicity assays have suggested that this pathovar structure should be revised, with strains belonging to pathovars *abberans*, *armoraciae*, and *barbareae* reassigned as *campestris*, *raphani*, or *incanae*, and any strains not known to be pathogenic designated as "nonpathogenic" (Fargier and Manceau 2007; Fargier et al. 2011). *Xanthomonas campestris* pv. *campestris* is the most agriculturally important of these pathovars. It causes black rot, a globally destructive vascular disease of *Brassica oleracea*, particularly important in cabbage and cauliflower. Black rot is identified by chlorotic and necrotic V-shaped lesions extending from hydathodes and wounds and by blackening of veins. Pathovar *raphani* is a nonvascular pathogen and causes leaf spot on brassicaceous and solanaceous crops. *X. campestris* pv. *incanae* does not cause disease on *Brassica* crop plants but is a vascular pathogen of ornamental crucifers including *Matthiola* and *Erysimum* spp. Some characterized *X. campestris* pv. *incanae* strains (e.g., type strain CFBP 2527) are pathogens of *Matthiola* spp., while others (e.g., CFBP 1606) infect only

[†]Corresponding author: A. J. Bogdanove; ajb7@cornell.edu

Current affiliation for Zoë E. Dubrow: Pairwise Plants, Durham, NC, U.S.A.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture. USDA is an equal opportunity provider and employer.

Funding: This work was supported by a USDA NIFA Pre-doctoral Research Fellowship (2019-67011-29501 to Z. E. Dubrow), the New York Cabbage Research and Development Program (contract number 11404 to C. D. Smart and A. J. Bogdanove), the French Laboratory of Excellence project 'TULIP' (ANR-10-LABX-41 and ANR-11-IDEX-0002-02 to L. D. Noël), and USDA-ARS (Project 8044-22000-047-00D, M. A. Tancos).

***e*-Xtra:** Supplementary material is available online.

The author(s) declare no conflict of interest.

Erysimum spp. (Fargier and Manceau 2007; Fargier et al. 2011). The basis for this difference in host specificity is unknown. Many of the *X. campestris* nonpathogenic strains lack a type III secretion system (T3SS), which is essential for pathogenicity (Meline et al. 2019; Lee et al. 2020), while others appear to have a full virulence gene repertoire (Fargier and Manceau 2007; Roux et al. 2015). We recently distinguished the latter as *X. campestris* "unknown pathogenicity" due to their high, whole-genome average nucleotide identity (ANI) to pathogenic strains but lack of a known host (Dubrow and Bogdanove 2021).

Despite causing different diseases (or no disease) and exhibiting different host ranges, *X. campestris* pvs. *campestris*, *raphani*, *incanae*, and *X. campestris* nonpathogenic or unknown pathogenicity strains have greater than 96% ANI (Dubrow and Bogdanove 2021). Apart from the lack of the T3SS in *X. campestris* nonpathogenic strains, the molecular mechanisms underlying these differences in pathogenicity are unknown. It has been hypothesized that divergence in host range is due to differences in virulence gene content, such as presence or absence of type III–secreted effectors (T3Es) (Roux et al. 2015). A comparison of the T3E repertoires of eight *X. campestris* pv. *campestris*, two *X. campestris* pv. *raphani*, two *X. campestris* pv. *incanae*, and an *X. campestris* unknown pathogenicity strain revealed multiple, though often unconserved, differences in T3E content across these groups (Bolot et al. 2013b). Also, whole-genome sequences of 10 *Xanthomonas* strains representing vascular and nonvascular pathogens of dicot or monocot hosts, including one *X. campestris* pv. *raphani* and three *X. campestris* pv. *campestris* strains, were compared to gain insight into tissue specificity as well as host specificity (Bogdanove et al. 2011), but no specific genes were identified as determinants. Recently however, a comparison of 54 *Xanthomonas* and five *Xylella* genomes (Gluck-Thaler et al. 2020) uncovered significant association of the *cbsA* gene, encoding a predicted extracellular cellobiosidase, with vascular pathogenesis, and experimental results support a contributing role. Transfer of the gene from a vascular pathovar to a nonvascular pathovar of *X. translucens* enabled vascular pathogenesis by the latter, a knockout of the gene in the vascular *X. translucens* pathovar expanded symptom development to the adjacent nonvascular tissue, and knockouts of the gene in the vascular pathogens *X. oryzae* pv. *oryzae*, *Ralstonia solanacearum*, and *Xylella fastidiosa* impaired pathogenicity (Gluck-Thaler et al. 2020; Jha et al. 2007; Liu et al. 2005).

X. campestris pv. *campestris* does not appear to have a strongly geographically based population structure, and strains isolated in a particular region, for example, in the northeastern United States state of New York, have been observed not to persist from year to year, together indicating the prevalence of dissemination on seed (Denance et al. 2018; Fargier et al. 2011; Lange et al. 2016). Cultural practices such as seed saving, insufficient sanitation of transplant production facilities, or lack of field rotation may result in endemic *X. campestris* populations, but this is not common (Bella et al. 2019).

In addition to infecting crops, *X. campestris* is often associated with cruciferous weeds. However, studies on collections of strains isolated from cruciferous weeds in Germany (Krauthausen et al. 2018), New York (Lange et al. 2022), and California (Ignatov et al. 2007) support the conclusion that weeds are not a significant source of *X. campestris* pv. *campestris* inoculum to crops. While certain weed isolates were observed to be pathogenic in cabbage (particularly those from the California collection), in each study, the weed isolates grouped separately from *X. campestris* pv. *campestris* crop strains by multilocus sequence analysis and most were not pathogenic to crop plants.

Though apparently not a significant cause of crop disease outbreaks, weed isolates nonetheless are a potential resource for probing pathovar relationships and genetic determinants of host range and tissue specificity, a better understanding of which could lead to the discovery of new mechanisms of resistance to be bred into economically important cruciferous species. To this end, we first selected 24 strains from the California weed isolate collection (Ignatov et al. 2007), nine strains from New York weed isolate collections (Burkholder 1941; Lange et al. 2022), 20 crop isolates of *X. campestris* pv. *campestris* from the eastern United States (18 from New York and two from Michigan) (Lange et al. 2016), and two *X. campestris* pv. *campestris* isolates, two *X. campestris* pv. *incanae* isolates, and one *X. campestris* pv. *raphani* isolate originating from various locations from public collections (Supplementary Table S1). We then completed whole-genome sequencing using single molecule long-read technology (SMRT) and determined relationships among these and 49 publicly available *X. campestris* genomes, including one of a weed isolate from the eastern United States state of Maryland we recently reported (Tancos et al. 2022). Next, we assayed pathogenicity and tissue-specificity on representative subsets of the California weed isolates and Eastern United States crop isolates that we sequenced, all but one of the weed isolates from the eastern United States that we sequenced (strain CFBP 5824) (Burkholder 1941), plus the weed isolate from Maryland, and several reference strains, on a panel of plant species, to determine pathovar and to identify associations of pathovar and genetic relatedness. Finally, we analyzed T3E gene content to determine any relationships to host or tissue specificity and the presence or absence of the *cbsA* gene to determine its relationship to tissue specificity in *X. campestris*.

RESULTS AND DISCUSSION

Fifty-nine new complete *X. campestris* genome assemblies.

SMRT sequencing was completed for 59 *X. campestris* strains in three separate runs. Coverage ranged from 176 to 1,030×. Each genome was circular and complete after assembly, consisting of an approximately 5-Mb chromosome and up to three plasmids of varying sizes, but plasmids were not well-conserved among strains. Average GC content of each was approximately 60 to 65%. The assemblies and raw sequences for all strains have been deposited in GenBank under BioProject PRJNA689092 (Supplementary Table S1 lists accession numbers for individual strains).

Isolate pathovar designations and support for *X. campestris* pv. *barbareae*.

To determine pathovar designations for the newly sequenced isolates, we performed pathogenicity assays for representative strains on an array of hosts, including the sequenced kale-like *Brassica oleracea* var. *alboglabra* TO1000, to identify *X. campestris* pv. *campestris* strains, the ornamentals *Erysimum cheiri* and *Matthiola incana*, to identify *X. campestris* pv. *incanae* strains, *Solanum lycopersicum*, to identify *X. campestris* pv. *raphani* strains (tomato is a host of *X. campestris* pv. *raphani* but no other *X. campestris* pathovar), as well as two cruciferous weed species found in New York state, *Barbarea vulgaris* and *Sinapis arvensis* (Fig. 1; Supplementary Table S2). We assayed 48 strains total, including 13 of the California weed isolates selected to represent the genetic diversity of that group (Fig. 2), 12 of the New York crop isolates plus two more that were not sequenced, all nine of the weed isolates from New York plus the Maryland weed isolate, as well as three *X. campestris* pv. *campestris*, four *X. campestris* pv. *raphani*, and four *X. campestris* pv. *incanae* strains that were previously sequenced. To inoculate leaves, we used a clip-and-dip technique (discussed below) that allows for invasion by vascular or nonvascular colonizers.

As expected, previously characterized agronomic *X. campestris* pv. *campestris* strains caused vascular symptoms (expanding chlorotic and necrotic lesions) on *Brassica oleracea* and *Sinapis arvensis*. *X. campestris* pv. *raphani* strains caused spots on leaves of *Brassica oleracea*, *Solanum lycopersicum*, and *Sinapis arvensis*, though certain strains caused symptoms on the ornamental or other weed hosts (Supplementary Table S2). *X. campestris* pv. *incanae* strains caused vascular symptoms on either *M. incana*, *E. cheiri*, or both, but no symptoms on *Brassica oleracea* or *Barbarea vulgaris*. *X. campestris* pv. *incanae* CFPB1606 was observed only to cause symptoms on *E. cheiri* and not *M. incana* (Supplementary Table S2), confirming previously reported results (Fargier et al. 2011).

The weed isolates from California, like the agronomic *X. campestris* pv. *campestris* strains, elicited vascular symptoms on *Brassica oleracea*, as observed by Ignatov and colleagues (2007), and *Sinapis arvensis* and were therefore designated as *X. campestris* pv. *campestris*. In contrast to previously characterized agronomic *X. campestris* pv. *campestris* strains, of which a small percentage caused disease on *M. incana* (three out of

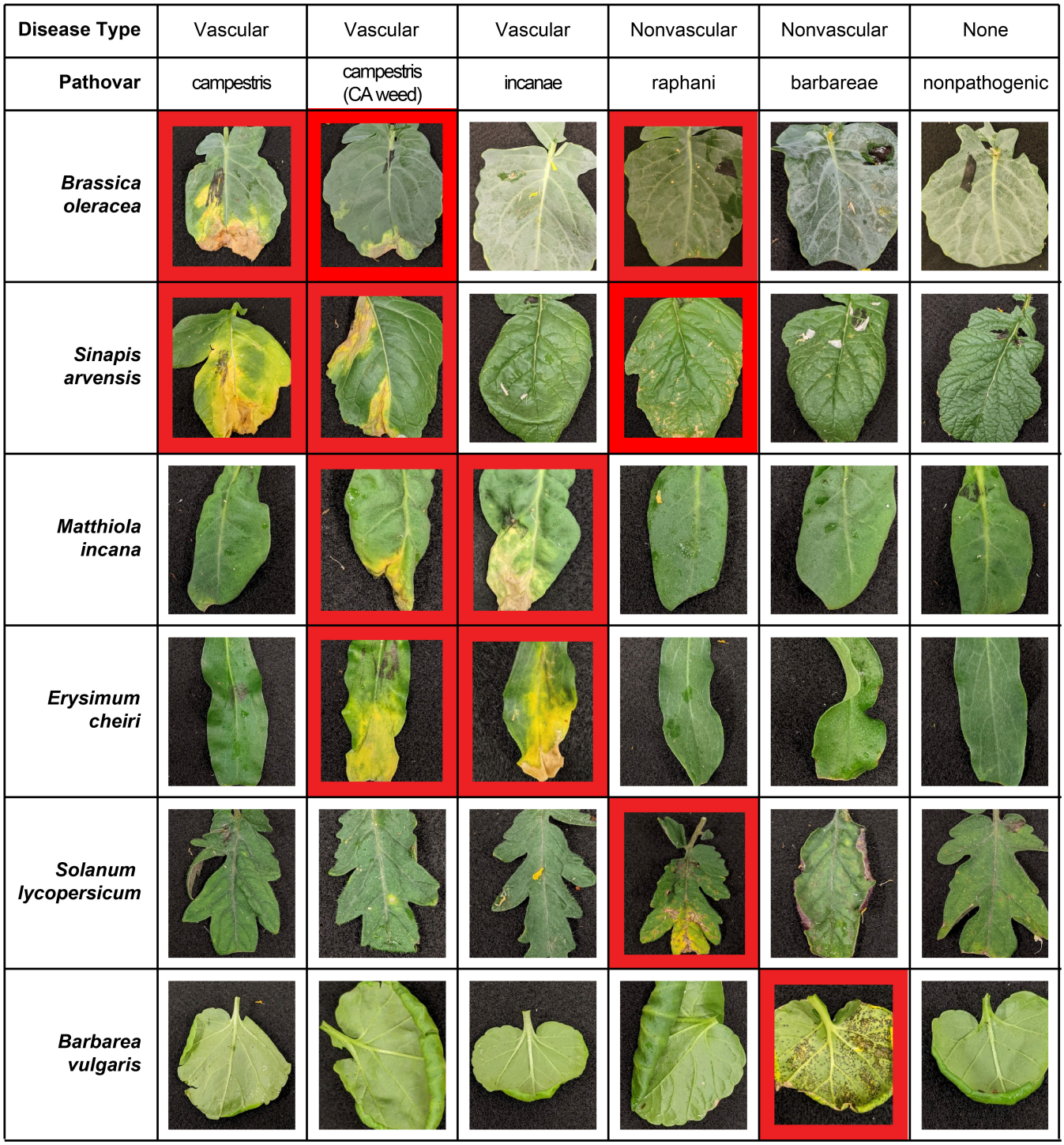

Fig. 1. Symptoms and host specificity of infection by representative strains of *Xanthomonas campestris* pathovars and subgroups examined in this study. Leaves of *Brassica oleracea*, selected brassicaceous weed and ornamental species, and tomato 14 days after clip-and-dip inoculation with a representative strain of each pathovar (or subgroup). Red borders highlight interactions that resulted in disease symptoms.

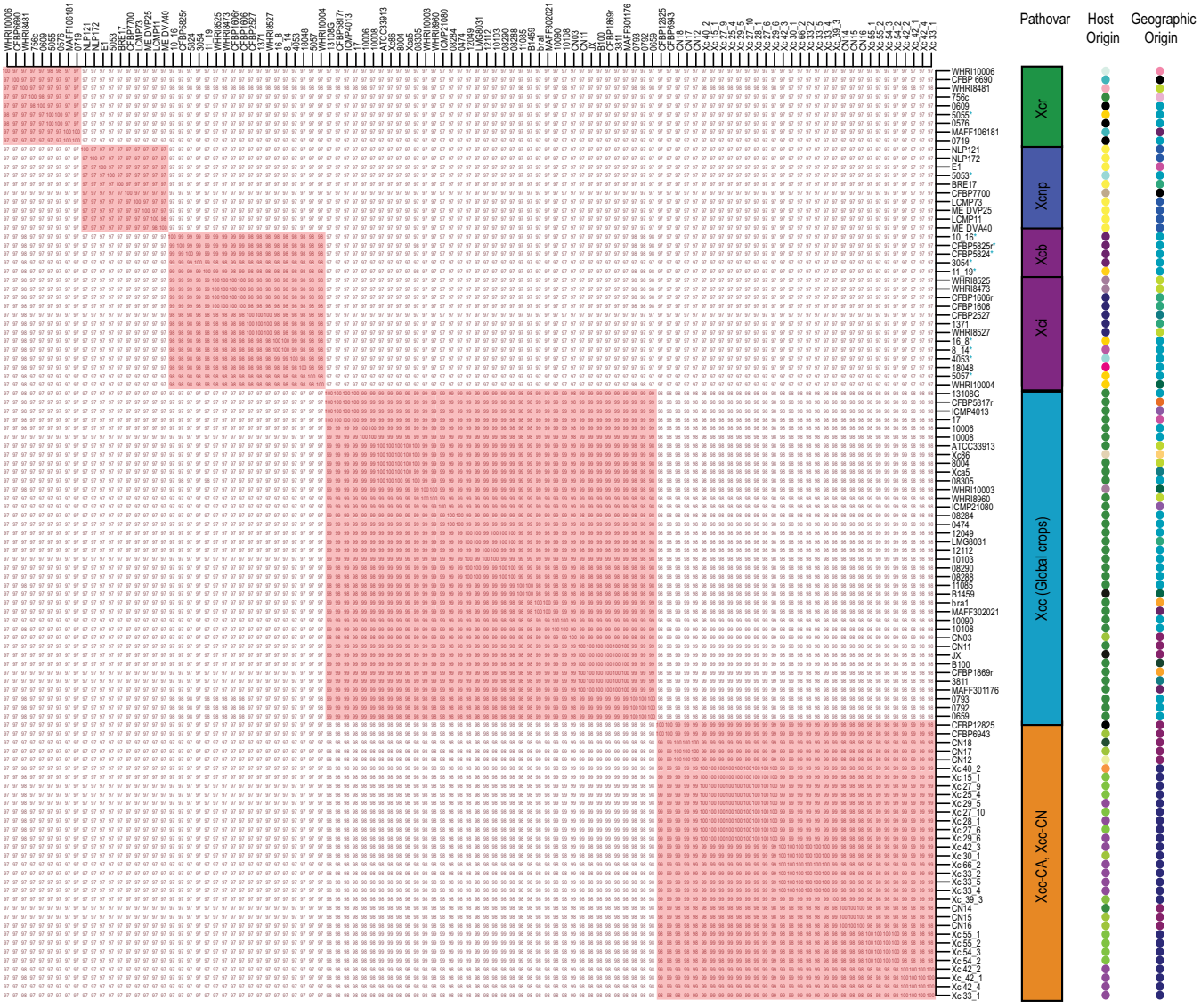

Fig. 2. Grouping of *Xanthomonas campestris* strains based on average nucleotide identity (ANI) is congruent with grouping by pathovar. ANI matrix of 109 *X. campestris* genomes constructed using the Enveomics Collection Toolbox genome-based distance matrix calculator. Red boxes in the matrix indicate compact clusters of closely related genomes calculated using Ward's method as part of the Enveomics Toolbox. Colored circles depict geographic origin and host species from which the strain was isolated. *Xcc* = *X. campestris* pv. *campestris*; *Xci* = *X. campestris* pv. *incanae*; *Xcb* = *X. campestris* pv. *barbareae*; *Xcnp* = *X. campestris* nonpathogenic strain; *Xcr* = *X. campestris* pv. *raphani*; *Xcc*-CA denotes strains isolated in California (Ignatov et al. 2007) and *Xcc*-CN denotes strains isolated in China (He et al. 2007). Asterisks indicate New York weed isolates.

17 strains), *E. cheiri* (seven out of 17 strains), or both ornamentals (three out of 17 strains), all but one California weed isolate caused disease on both ornamental crucifers. That isolate, 40-2, was pathogenic on *M. incana* but not *E. cheiri* (Supplementary Table S2).

Among the ten weed isolates from the eastern United States that were tested, four from New York and the single Maryland isolate caused vascular symptoms and four New York isolates caused spots. One New York weed isolate, 5053, caused no symptoms. Three of the vascular strains caused disease on *M. incana*, all five were pathogenic on *E. cheiri*, and none caused symptoms on *Brassica oleracea*. We therefore designated these strains as *X. campestris* pv. *incanae*. One of the spot-causing isolates, 5055, did so on *Brassica oleracea* and *Sinapis arvensis* but not on any other plant species and was therefore designated as *X. campestris* pv. *raphani*. These results were consistent with observed phenotypes in cabbage for these strains in a previous study, except for 16_8, which we did not observe to cause V-shaped lesions on cabbage, while Lange and colleagues (2022) did, perhaps due to the use of a different *Brassica oleracea* cultivar or inoculation method. The remaining three strains, 10_16, 11_19, and 3054, caused symptoms only on *Barbarea vulgaris* (Fig. 1). These symptoms, severe black and water-soaked spots, were strikingly similar to those of "black rot of *Barbarea vulgaris*," also caused by *X. campestris* isolates from New York weeds (Burkholder 1941). Two of those original isolates, now CFBP 5825 and CFBP 5826 (originally accessioned as ICPB XB2 and XB1, respectively), were designated as nonpathogenic *Xanthomonas campestris* (Fargier et al. 2011; Dubrow and Bogdanove 2021). Sequencing of a rifampicin-resistant derivative of CFBP 5825 ("CFBP 5825R") showed it has a T3SS and effector genes (Roux et al. 2015), so we previously considered it an *X. campestris* unknown pathogenicity strain (Dubrow and Bogdanove 2021). Fargier and colleagues 2011 did not find CFBP 5825 to be pathogenic on *Barbarea vulgaris* and they proposed that *X. campestris* pv. *barbareae*, as described by Vauterin et al. (1995), be abandoned, as it had not been confirmed in any other published study since its discovery (Fargier and Manceau 2007; Fargier et al. 2011). We did not have strains CFBP 5825 or CFBP 5826 to test on the New York cultivar of *Barbarea vulgaris.* However, because the sequence of CFBP 5825R is available and we had included in our sequencing another of the isolates from New York weeds that was reported to possibly cause weak spotting on *Barbarea vulgaris* (CFBP 5824 [Burkholder 1941]), we were able to determine, as detailed further below, that these strains are genetically closely related to the more recent New York isolates causing disease on *Barbarea vulgaris* (Fig. 2; Supplementary Fig. S1). Our results with this new group of strains isolated in New York and causing spot symptoms on *Barbarea vulgaris* but no other tested species support the *X. campestris* pv. *barbareae* pathovar designation.

Associations of pathovar and genetic relatedness.

To determine whether strains of the same pathovar are genetically more closely related to each other than to other strains, the 59 new assemblies and 49 published *X. campestris* genome sequences were first used to build an ANI matrix (Fig. 2). ANI was greater than 96% in all pairwise comparisons, surpassing the 95% ANI threshold for species-level identification. The strains nonetheless group into five clusters, containing i) all California weed isolates (*X. campestris* pv. *campestris*) and some *X. campestris* pv. *campestris* strains from China, ii) all New York *X. campestris* pv. *campestris* crop isolates and other *X. campestris* pv. *campestris* strains from China and elsewhere around the globe, iii) all *X. campestris* pv. *incanae* and *X. campestris* unknown pathogenicity (or rather *X. campestris* pv. *barbareae*) strains, including those isolated from weeds in New York and Maryland as well as CFBP 5824 and CFBP 5825R representing the isolates described in 1940 (Burkholder 1941), iv) all *X. campestris* pv. *raphani* strains, including those isolated from weeds and crops in New York, and v) all *X. campestris* nonpathogenic strains, including the New York weed isolate 5053. This result reveals a strong association of pathovar with genetic similarity and supports the use of genetic relatedness to predict host range and tissue specificity of new *X. campestris* isolates.

Comparing strain relatedness with the original plant host and geographic origin provides potentially useful insight. For example, one of the *X. campestris* unknown pathogenicity/*X. campestris* pv. *barbareae* strains was found on *Capsella bursa-pastoris*. Thus, these strains, all of which were isolated in New York and all of which infect *Barbarea vulgaris*, may have a broader host range than we observed, as pathogens or as endo- or epiphytes. As another example, the relationship of the California weed isolates to strains from China may be informative. In the previous study of the California isolates, it was noted that all were pathogens on both the weeds they were isolated from and *Brassica oleracea* but that they grouped separately from crop isolates tested (Ignatov et al. 2007). The expanded analysis here, including all available *X. campestris* genomes, reveals that the California weed isolates are related specifically to strains from China isolated from *Brassica rapa*, *Brassica juncea*, and *Brassica napus* rather than from *Brassica oleracea* as most other represented *X. campestris* pv. *campestris* strains were. This observation suggests that, in the strains isolated in China and in the California weed isolates, there are i) adaptations that make them better able to infect the other crop brassicas and possibly weed species or ii) avirulence or other limiting factors that make them less fit on *Brassica oleracea*. The latter is consistent with the conclusion of the authors of the original study that weeds are not a major source of inoculum for disease on *Brassica oleracea* crops in California, though it remains possible that such weed strains might spread to such crops under certain environmental conditions (Ignatov et al. 2007).

The California weed isolate collection contrasts with the New York weed isolate collection and the one from Germany in that the isolates from California were pathogenic on *Brassica oleracea* while none of the New York weed isolates tested or the Maryland weed isolate were, and only a few of the isolates in the German collection were observed to cause disease on a crop species (Ignatov et al. 2007; Krauthausen et al. 2018; our results). These differences may be the result of sampling bias, however. Drawing firm conclusions about the origin of the weed-associated *X. campestris* pv. *campestris* strains and the overall *X. campestris* population on weeds in California would benefit from further sampling.

Next, to examine phylogeny, the whole-genome sequences were used to construct a maximum likelihood (ML) tree (Fig. 3; Supplementary Fig. S1). The tree confirms the groups from ANI analysis and provides additional insight into relationships within those groups. For example, the California weed isolates and the related *X. campestris* pv. *campestris* strains from China did not separate out from the rest of the *X. campestris* pv. *campestris* crop isolates monophyletically. However, a large subgroup of those strains does form a single clade. Also, the one Maryland and four New York weed isolates designated as *X. campestris* pv. *incanae* and the three New York weed isolates designated as *X. campestris* pv. *barbareae* form a clade with all other *X. campestris* pv. *incanae* and *X. campestris* unknown pathogenicity or *X. campestris* pv. *barbareae* strains, with the *X. campestris* unknown pathogenicity and *X. campestris* pv. *barbareae* strains forming a subclade within it. As noted earlier, CFBP 5824 and CFBP 5825R, representing the early New York weed isolates (Burkholder 1941), also reside in this *X. campestris* pv. *barbareae* clade. Burkholder (1941) classified CFBP 5824 as

pathovar *armoraciae*, causing symptoms on horseradish and "no or very slight" symptoms on *Barbarea vulgaris.* Also, although CFBP 5825 was described as causing a spot disease on *Barbarea vulgaris* (Burkholder 1941), Fargier and Manceau (2007) did not observe symptoms following inoculation of CFBP 5825 to *Barbarea vulgaris*. We surmise that *Barbarea vulgaris* accessions differ in their susceptibility to different strains in this group. Specifically, the accession used by Fargier and Manceau (2007) from a Swiss collection may not be susceptible to CFBP 5825, while the accession we used, which originated in New York and is likely more genetically similar to the plants initially used by Burkholder (1941), clearly is. It would be of interest to assay CFBP 5824, CFBP 5825R, and the other strains in this group on a collection of *Barbarea vulgaris* accessions to determine whether there is indeed variation across different strain and accession combinations. With respect to CFBP 5825, alternative explanations for the difference between results of Burkholder (1941) and those of Fargier and Manceau (2007) include attenuation of the strain during subculturing and differences in assay conditions. Though re-examination of the pathogenicity of CFBP 5824 and CFBP 5825 on *Barbarea vulgaris* is warranted, overall, the information available and our results strongly suggest a shared genetic basis for pathovar *barbareae*, as a descendant of *X. campestris* pv. *incanae*. Finally, the single New York isolate designated as *X. campestris* pv. *raphani* grouped in a clade with all other *X. campestris* pv. *raphani* strains as well as all *X. campestris* nonpathogenic strains, with the latter forming a single subclade. Included in the latter is the sole New York weed isolate found to be nonpathogenic, 5053. A BLAST search of the 5053 genome confirmed the absence of a T3SS and T3Es. These relationships suggest that loss of the T3SS and associated T3E genes from an *X. campestris* pv. *raphani* progenitor may have given rise to the *X. campestris* nonpathogenic strains or, though seemingly less likely, that lack of a T3SS and T3Es could be the ancestral state and *X. campestris* nonpathogenic strains never gained these genes.

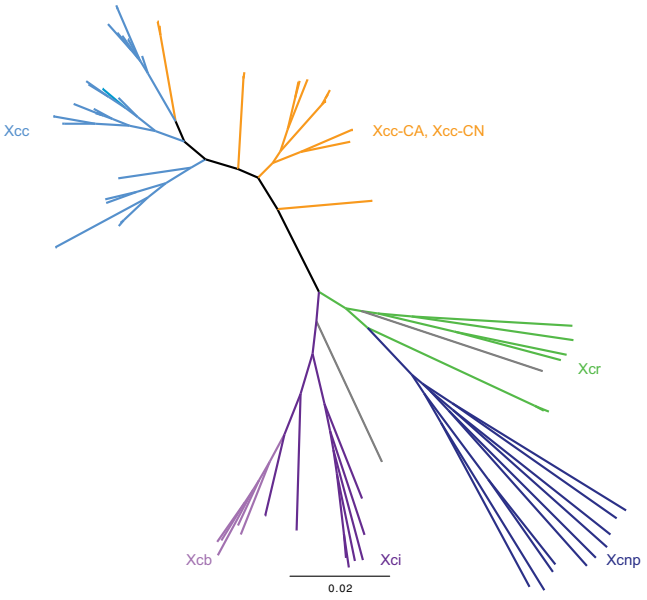

Fig. 3. Maximum likelihood tree of *Xanthomonas campestris* genomes sequenced in this study and publicly available genomes reflects pathovar relationships. Representing 108 total strains, the tree was constructed using the reference sequence alignment-based phylogeny builder (REALPHY) (Bertels et al. 2014). Orthologous single nucleotide polymorphisms were called within REALPHY using the *X. campestris* pv. *campestris* (*Xcc*) B100, ATCC33913, 8004, and Xca5 genomes as references. The resulting alignments were then merged as a part of the REALPHY pipeline. This merged alignment was used to reconstruct the phylogeny using RAxML HPC Blackbox on the CIPRES portal. Branch colors are based on pathovar classification. Gray branches indicate strains for which pathotype was not tested in this or other studies but was inferred to be nonpathogenic due to lack of a type III secretion system.

T3E repertoires.

Due to distinct differences in pathovar tissue specificities and host ranges despite their high genetic percent identity, we hypothesized that there may be pathovar-specific differences in T3E repertoires. Using TBLASTN and a cut-off of 60% identity and 60% coverage, we queried the whole-genome sequences with a list of known *Xanthomonas* T3Es (available online from The *Xanthomonas* Resource) (Supplementary Table S3). The genomes were found to harbor between zero and 19 T3E genes, with *X. campestris* nonpathogenic strains containing none, *X. campestris* pv. *raphani* strains averaging about six, and *X. campestris* pv. *incanae*, *X. campestris* pv. *barbareae*, and *X. campestris* pv. *campestris* strains averaging 14 to 16. While overall patterns of T3E gene content reflect phylogeny at the pathovar level, there are no conserved pathovar-specific T3E genes or individual T3E genes that distinguish vascular and nonvascular pathovars. However, for each pathovar, there are T3E genes that are present in all strains, and T3E genes that are absent from all strains, as has been observed by others (Roux et al. 2015). Most striking in this regard are the *X. campestris* pv. *raphani* strains, all of which have the same six T3E genes and few or no others (Supplementary Table S3).

Sequences of strains CFBP 6690 and WHRI8481, which were accessioned as *X. campestris* pv. *raphani* and *X. campestris*, respectively, and which genetically group with *X. campestris* pv. *raphani* via ANI (though WHRI8481 grouped with *X. campestris* pv. *incanae* strains in the ML tree [Fig. 3]), lack T3E genes entirely. Further investigation revealed that both also lack a T3SS and are therefore likely to be *X. campestris* nonpathogenic strains, though we did not test this prediction with pathogenicity assays. Since CFBP 6690 and WHRI8481 group distinctly from the *X. campestris* nonpathogenic strain clade, evolution toward an endo- or epiphytic lifestyle via loss of the T3SS and associated T3E genes may have occurred in multiple lineages.

Transcription activator-like effector (TALE) variation.

Some *Xanthomonas campestris* strains encode TALEs, a class of DNA-binding T3E that directly activate specific host genes, some of which are 'susceptibility' genes that contribute to disease. TALE structure is well-conserved among homologs, and target specificity depends on pairs of hypervariable amino acids in a central repeat region, termed repeat variable di-residues, that specify individual DNA bases (Boch et al. 2009; Moscou and Bogdanove 2009). The central repeats of TALEs are usually each 33 to 35 amino acids long. Thus, *tal* genes, in which the repeats are 99 to 105 bp, are often missed or misassembled in short read–derived genomes. The published genomes included in our analysis that were assembled from short-read data, therefore, may not accurately reflect *tal* gene content. Furthermore, as TALEs are of particular interest to our research groups, in this and previous studies (e.g., Denance et al. 2018), toward capturing the diversity of *tal* genes present in *X. campestris* populations, strains were selected for sequencing based, in part, on prescreening for *tal* genes by PCR or Western blot analysis; therefore, the proportion of strains in the sequenced collection that harbor *tal* genes, particularly *X. campestris* pv. *campestris* strains, may not represent that in nature.

With these two caveats, analysis of TALE gene content (Supplementary Table S3) revealed the following. The *X. campestris* pv. *campestris* and *X. campestris* pv. *incanae* strains contain

up to four TALE genes, though some strains contain none. Indeed, no TALE genes were found in any California *X. campestris* pv. *campestris* weed isolate from the sequenced collection. TALE genes are present in all the *X. campestris* pv. *barbareae* genome assemblies that were generated using long-read technology. Yet, none of the *X. campestris* pv. *raphani* genome assemblies contains TALE sequences, including the long-read assemblies. This was unexpected since *X. campestris* pv. *raphani* and *X. campestris* pv. *campestris* are often found in the same field or even on the same plant, and the TALE genes in *X. campestris* pv. *campestris* are found either on plasmids, in association with mobile elements, or both, which suggests that they could be horizontally transferred.

While TALE function has been investigated more thoroughly in other *Xanthomonas* species, TALEs are a recent discovery in *X. campestris* pv. *campestris* (Denance et al. 2018). Three TALEs (Hax2, Hax3, and Hax4) from *X. campestris* pv. *campestris* Xca5 (previously classified as pathovar *armoraciae*) were found to collectively contribute to virulence (Bolot et al. 2013a; Kay et al. 2005), but susceptibility gene targets of these or other *X. campestris* pv. *campestris* TALEs have not yet been identified. More research is needed to dissect the significance of the unusual TALE distribution in *X. campestris* and any virulence functions of TALEs in strains of different pathovars.

AvrAC restricts host range on *M. incana*.

AvrAC (also called XopAC) is a T3E found so far only in *X. campestris* (Xu et al. 2008). *X. campestris* pv. *campestris* strains expressing AvrAC are unable to cause disease on *Arabidopsis thaliana* accession Col-0 due to recognition of the effector by the ZAR1 resistance protein complex and subsequent activation of plant defense, characterized by the plant hypersensitive reaction (HR), a rapid, localized, cell death (Adachi et al. 2019; Feng and Tang 2019; Guy et al. 2013b). AvrAC is a uridylyl transferase, and its avirulence function depends on this activity (Feng et al. 2012; Wang et al. 2015). The *avrAC* gene is present in the majority of the *X. campestris* pv. *campestris* strains isolated from *Brassica oleracea* worldwide but missing from all but one of the *X. campestris* pv. *campestris* isolates from California weeds (Supplementary Table S3), 40_2. As noted above, isolate 40_2 is the only California weed isolate that did not cause disease on *M. incana* (Supplementary Table S2). The *avrAC* gene is present in all the *X. campestris* pv. *barbareae* and *X. campestris* pv. *raphani* strains, but only in the subset of *X. campestris* pv. *incanae* strains that did not cause disease on *M. incana* (e.g., CFBP 1606). The *avrAC* gene is in fact absent from every *X. campestris* strain that did cause symptoms on *M. incana* (Supplementary Table S2). These observations suggest that AvrAC acts as an avirulence protein when delivered by the pathogen into *M. incana* cells.

To explore this hypothesis further, we inoculated *M. incana* leaves, using syringe infiltration, with a selection of *X. campestris* pv. *campestris* strains that contain or lack the *avrAC* gene, including the one California weed isolate that contains it, a pair of *X. campestris* pv. *incanae* strains with and without the gene, and an *X. campestris* pv. *barbareae* strain (which has it) (Fig. 4A). By 28 h after inoculation, HR was apparent in all leaves inoculated with any *avrAC*-containing strain but not in those inoculated with a strain lacking *avrAC*. As a more conclusive test, we assayed, alongside the *avrAC*-containing

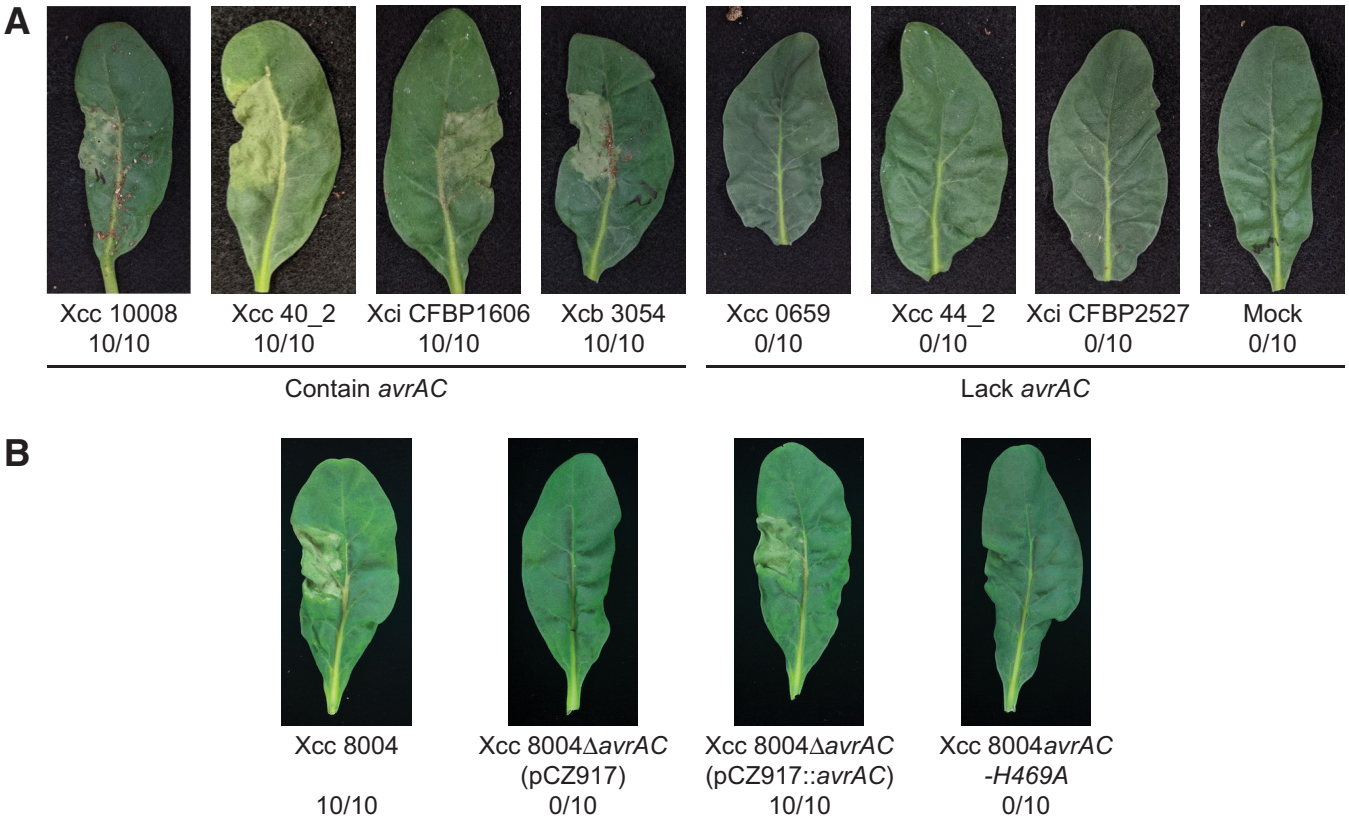

Fig. 4. The *avrAC* gene causes avirulence on *Matthiola incana*. **A,** Results in *M. incana* leaves 28 h after syringe infiltration with strains of *Xanthomonas campestris* pv. *campestris* (*Xcc*), *X. campestris* pv. *incanae* (*Xci*), and *X. campestris* pv. *barbareae* (*Xcb*) naturally containing or lacking *avrAC* or mock inoculum and **B,** 48 h after syringe infiltration with *X. campestris* pv. *campestris* 8004, 8004Δ*avrAC*(pCZ917), 8004Δ*avrAC*(pCZ917:*avrAC*), or 8004*avrAC-H469A*. Ten replicate inoculations were carried out, one per leaf, for each. Representative leaves are shown, with the tally of leaves showing HR indicated below. The experiments were repeated once (A) or twice (B) with equivalent results.

X. campestris pv. *campestris* strain 8004, the following strains generated in an earlier study (Guy et al. 2013a): an *avrAC* knockout derivative of 8004 carrying the empty plasmid vector pCZ917, the derivative carrying *avrAC* on the plasmid, and an 8004 derivative with the endogenous *avrAC* gene replaced by *avrAC-H469A*, which encodes an enzymatically inactive protein. This experiment was scored at 48 h. The results (Fig. 4B) confirmed that AvrAC causes avirulence on *M. incanae* and demonstrated that this function depends on its enzymatic activity, together suggesting the presence of ZAR1 or a close functional analog in *M. incanae*.

While we only tested pathogenicity on one cultivar of *M. incana*, other researchers reported that *X. campestris* pv. *incanae* CFBP 1606 expressing AvrAC is avirulent on a different *M. incana* cultivar (Fargier and Manceau 2007). Thus, AvrAC-specific resistance may be characteristic of *M. incana* and may be present in other, closely related species. No *M. incana* genome sequence is currently available, so it is unknown whether *ZAR1* or some other gene is responsible for the resistance. *Arabidopsis thaliana* and *M. incana* are more genetically similar to *Brassica* crop species that lack AvrAC-specific resistance than to each other (Fig. 5), so the resistance may have convergently evolved. Alternatively, *ZAR1*, which is an ancestral nucleotide-binding domain and leucine-rich repeat containing (*NLR*) gene present in many plant species, or another gene required for its function may have been lost in the lineage that gave rise to the crop species.

The T3E analysis we carried out suggests that the majority of *X. campestris* pv. *campestris* strains isolated from *Brassica oleracea* around the world contain *avrAC* (Supplementary Table S3). Thus, introduction of the gene or genes needed for ZAR1-mediated resistance into cabbage by breeding or genetic transformation could provide broadly effective disease control. If pathogen loss of the effector to evade detection were to reduce virulence, the resistance would also likely be relatively durable. In *Arabidopsis*, AvrAC has been shown to inhibit pattern-triggered plant immunity (PTI) by uridylating Bik1 (Feng et al. 2012; Meng and Zhang 2013) but its contribution to virulence, under controlled conditions, depends on the strain (Guy et al. 2013b). Whether AvrAC is broadly important to *X. campestris* pv. *campestris* virulence on cabbage under field conditions is yet to be tested. Even so, because *X. campestris* pv. *campestris* has not been found to persist in fields in crop production areas such as New York (Lange et al. 2016), the likelihood of resistance-breaking strains evolving from local populations is low. However, if F1 seed is repeatedly generated in an area where the pathogen does persist year-round, resistance may break down through pathogen loss of *avrAC*.

Fig. 5. Relationships among the host species of origin for the *Xanthomonas campestris* strains examined in this study. The hierarchical tree was generated using the NCBI taxonomy browser.

Importantly, though, through interactions with distinct receptor-like cytoplasmic kinases (RLCKs), ZAR1 mediates recognition of several *Pseudomonas* effectors, including HopBA1, HopF1/HopF2, HopO1, HopX1, and HopZ1 (Laflamme et al. 2020). Further, ZAR1 of *Nicotiana benthamiana* mediates recognition of the *X. populans* effector XopJ4, a member of the HopZ/YopJ superfamily (Schultink et al. 2019). Homologs of HopX1 and three members of the HopZ/YopJ superfamily, XopE and XopJ1, Xop-J3, and Xop-J5, respectively, are encoded among the *X. campestris* genomes examined here (Supplementary Table S3). The *xopE* gene is found in most of the *X. campestris* pv. *campestris* genomes, many of the *X. campestris* pv. *incanae* genomes, and one of the *X. campestris* pv. *barbareae* genomes. The *xopJ1* and *xop-J3* genes are present in one or two of the *X. campestris* pv. *barbareae* genomes, and *xopJ5* is found in many *X. campestris* pv. *campestris*, several *X. campestris* pv. *incanae*, and one of the *X. campestris* pv. *barbareae* genomes. Thus, ZAR1 may be effective even against strains that lack AvrAC. Notably however, *X. campestris* pv. *campestris* 8004 contains both *xopE* and *xopJ5*, yet the 8004 *avrAC* knockout mutant grew in population and caused symptoms in *M. incana* leaves following infiltration and pinprick inoculation, respectively (Supplementary Fig. S2). Though we did not repeat the experiment, this result suggests that those Xops are not detected in *M. incanae*, and it underscores the likelihood that the breadth of ZAR1 activity will depend on host genetic background, specifically on the presence of cognate RLCK genes.

Involvement of *cbsA* in *X. campestris* tissue specificity.

To determine whether the presence of *cbsA* correlates with differences in tissue specificity in *X. campestris*, we searched all available *X. campestris* genomes for *cbsA*. Two copies of *cbsA* were identified. One copy, with a 1,701-bp coding sequence ("$cbsA_{1701}$"), was found in all *X. campestris* pv. *campestris* and *X. campestris* pv. *incanae* strains (i.e., all vascular disease-causing strains) and was absent from all *X. campestris* pv. *barbareae*, *X. campestris* nonpathogenic, and *X. campestris* pv. *raphani* strains (i.e., all nonvascular or nonpathogenic strains) (Supplementary Table S4). This copy shows >99% nucleotide identity across the strains that have it. A second copy of *cbsA*, lacking 210 bp coding for the N-terminal signal peptide found in $cbsA_{1701}$, was found in all *X. campestris* strains, with slightly less sequence conservation, at >95% nucleotide identity.

To test whether $cbsA_{1701}$ is, indeed, a determinant of vascular pathogenicity in *X. campestris*, we expressed a clone of the gene from *X. campestris* pv. *campestris* 8004 under its native promoter in *X. campestris* pv. *barbareae* strains 3054 and 11_19 and in *X. campestris* pv. *raphani* 756c and compared these transformants to transformants carrying the empty vector. We hypothesized that the gene would convert the *X. campestris* pv. *incanae* and *X. campestris* pv. *raphani* strains to vascular pathogens on their hosts. Because *X. campestris* pv. *barbareae* is closely related to *X. campestris* pv. *incanae*, which is a vascular pathogen of *E. cheiri*, we further hypothesized that the addition of $cbsA_{1701}$ might expand the host range of *X. campestris* pv. *barbareae* to include *E. cheiri*. Syringe infiltration of *E. cheiri* leaves with wild-type *X. campestris* pv. *barbareae* 11_19 resulted in a HR, while *X. campestris* pv. *barbareae* 3054 did not cause HR,

suggesting the presence of an avirulence factor in 11_19 but not 3054, recognized by *E. cheiri* (Supplementary Fig. S3A and B). Based on this result, we inoculated the *X. campestris* pv. *barbareae* 11_19 and 3054 transformants to *Barbarea vulgaris* and the 3054 transformants to *E. cheiri.* We inoculated the *X. campestris* pv. *raphani* transformants to *Brassica oleracea* TO1000. For all inoculations, we used the clip-and-dip method.

On *Barbarea vulgaris*, $cbsA_{1701}$ made no difference to symptoms caused by *X. campestris* pv. *barbareae* 3054 or *X. campestris* pv. *barbareae* 11_19, each causing only spots (Supplementary Fig. S3C to F). On *E. cheiri*, however, *X. campestris* pv. *barbareae* 3054 carrying the $cbsA_{1701}$ plasmid (pBBR1::$cbsA_{1701}$) caused chlorotic and necrotic lesions extending from the clipped end of the leaf (Fig. 6), similar in appearance to symptoms caused by *X. campestris* pv. *incanae* (Fig. 1), though slower to develop. In *X. campestris* pv. *raphani* 756c inoculated to *Brassica oleracea* TO1000, as in the *X. campestris* pv. *barbareae* strains inoculated to *Barbarea vulgaris*, $cbsA_{1701}$ had no effect on symptom development, with both the $cbsA_{1701}$ transformant and the empty vector control causing spots only (Fig. 6).

While genomic data show a perfect correlation between the presence of the $cbsA_{1701}$ gene and vascular pathogenicity and heterologous expression of $cbsA_{1701}$ conferred apparent weak vascular pathogenicity to one *X. campestris* pv. *barbareae* strain on one of two *Brassica* species tested, additional factors must be involved in vascular versus nonvascular disease development. These may be species-specific avirulence factors, as suggested by the avirulence of *X. campestris* pv. *barbareae* 11_19 on *E. cheiri* (Supplementary Fig. S3A and B). Alternatively and not mutually exclusive, they may be positive-acting factors present in some strains and not others, as suggested for other *Xanthomonas* species (Gluck-Thaler et al. 2020). This seems particularly plausible for *X. campestris* pv. *raphani*, given that *X. campestris* pv. *raphani* strains have a highly reduced effector repertoire in comparison to *X. campestris* pv. *campestris* (Supplementary Table S3). Host factors also could play a role.

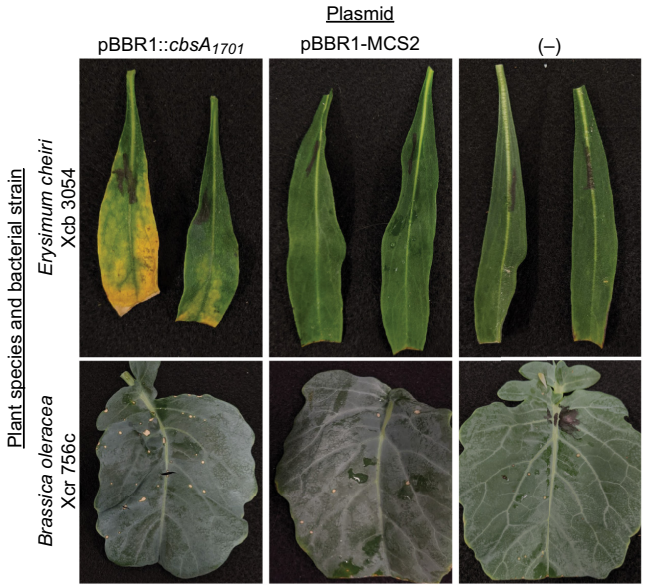

Fig. 6. Heterologous expression of $cbsA_{1701}$ confers apparent vascular pathogenicity to *Xanthomonas campestris* pv. *barbareae* 3054 on *Erysimum cheiri* but not to *X. campestris* pv. *raphani* 756c on *Brassica oleracea*. *E. cheiri* and *Brassica oleracea* TO1000 leaves are shown 21 days after clip-and-dip inoculation with the *X. campestris* pv. *barbareae* and *X. campestris* pv. *raphani* strains carrying one or both pBBR1::$cbsA_{1701}$ and the empty vector or with the wild-type strains.

Summary and conclusions.

Genome sequencing of 59 additional *X. campestris* genomes expanded the existing collection to include a more diverse array of strains isolated from weed and crop species. Pathogenicity assays on selected *Brassica* species and tomato were used to determine pathovar. Strain groupings by pathovar were reflected in whole genome sequence comparisons. Pathogenicity profiles and genetic relationships of strains isolated from weeds in New York that formed a subclade in the larger group of *X. campestris* pv. *incanae* strains provided support for pathovar *barbareae*, rejection of which had been proposed. Mining of the genomic data in the context of the pathovar information provided new insight also into virulence factors and determinants of host and tissue specificity.

Three conclusions emerged regarding tissue-specificity and host range of *X. campestris*. First, nonpathogenic isolates do not have a T3SS, while pathogenic isolates do (Arlat et al. 1991). We determined that a group of *X. campestris* strains isolated in New York cause spot symptoms on *Barbarea vulgaris.* These strains are closely related to ones isolated from New York weeds in the early 1940s, including CFBP 5825, which was originally reported to be the type strain for *X. campestris* pv. *barbareae*. While CFBP 5825 had most recently been considered nonpathogenic and pv. *barbareae* removed from the pathovar naming scheme, our phenotyping provides evidence that *barbareae* is a true pathovar represented by CFBP 5825 and additional recently isolated New York weed strains, which all contain a T3SS and T3E repertoires. It remains to be seen if strains CFBP 6690 and WHRI8481, which genetically group with *X. campestris* pv. *raphani* and *X. campestris* pv. *incanae*, respectively, but lack a T3SS, are pathogenic, though we would expect not. Based on our data, within *X. campestris*, in contrast to some other *Xanthomonas* species (Jacobs et al. 2015; Pieretti et al. 2009), presence of a T3SS and T3Es is tied to pathogenicity and absence to a nonpathogenic endo- or epiphytic lifestyle.

Second, while no clear associations of T3E content and tissue specificity were apparent, the T3E AvrAC appears to act as a host-range determinant, preventing infection of the ornamental species *M. incana*. *Arabidopsis thaliana* Col-0 recognizes AvrAC by virtue of the NLR protein ZAR1 and associated proteins, but *Arabidopsis* is less closely related to *M. incana* than it is to *Brassica oleracea* (Fig. 5). A recent phylogenetic analysis of NLR proteins from diverse plant species suggests that ZAR1 may be the most conserved coiled-coil NLR in angiosperms (Adachi et al. 2020). ZAR1 may be present in additional *Brassica* species and could be leveraged for breeding resistance to black rot. By expanding the pool of sequenced *X. campestris* pv. *campestris* genomes, our study suggests that the majority of *X. campestris* pv. *campestris* crop isolates have AvrAC, as well as other potential elicitors of ZAR1-mediated resistance (i.e., XopE and XopJ5) and that such resistance could be broadly effective.

Last, the observed presence of $cbsA_{1701}$ in vascular but not nonvascular *Xanthomonas* pathogens (Gluck-Thaler et al. 2020) extends to *X. campestris* pathovars, including all of the newly sequenced strains in this study. All strains of *X. campestris* that cause vascular disease, i.e., strains of *X. campestris* pv. *campestris* and *X. campestris* pv. *incanae*, have $cbsA_{1701}$, while all nonvascular or nonpathogenic strains, *X. campestris* pv. *raphani*, *X. campestris* pv. *barbareae*, and *X. campestris* nonpathogenic strains, lack the gene. Experimentally, transfer of $cbsA_{1701}$ to one *X. campestris* pv. *barbareae* strain led to symptoms similar in appearance to those caused by the vascular pathogen *X. campestris* pv. *incanae* in *E. cheiri*. However, a similar result was not observed for a second *X. campestris* pv. *barbareae* strain, and introduction of $cbsA_{1701}$ did not preclude *X. campestris* pv. *barbareae* strains from causing nonvascu-

lar symptoms in *Barbarea vulgaris*. Additionally, transfer of *cbsA*$_{1701}$ to *X. campestris* pv. *raphani* did not alter tissue specificity. These data bolster conclusions from previous studies that *cbsA*$_{1701}$ may contribute but that there are additional factors involved in vascular versus nonvascular disease development (Cerutti et al. 2017; Gluck-Thaler et al. 2020; Jha et al. 2007).

In addition to the insight gained from our study, we expect the new genome sequences and phenotypic data we have presented to facilitate future studies to further improve understanding of *X. campestris* pathogenicity. Investigations of host range determinants and novel factors involved in tissue specificity will be especially valuable in developing strategies for improved resistance in the field.

MATERIALS AND METHODS

Bacterial strains and plant varieties used.

Wild-type *X. campestris* strains used or referenced are listed in Supplementary Table S1. Plants used include *Barbarea vulgaris* and *Sinapis arvensis* (Cornell University Weed Garden), *Brassica oleracea* var. oleracea TO1000 (Vilmorin, Salinas, CA, U.S.A.), *Solanum lycopersicum* cv. Moneymaker (Eden Brothers, Arden, NC, U.S.A.), *Matthiola incana* cv. Iron Cherry (Johnny's Selected Seed, Fairfield, ME, U.S.A.), and *Erysimum cheiri* cv. English Wallflower (Eden Brothers, Arden, NC, U.S.A.).

Genome sequencing.

Bacteria were cultured at 28°C with shaking (225 rpm) for 24 h in nutrient broth (BD Biosciences, San Jose, CA, U.S.A.), and genomic DNA was isolated using the MasterPure gram-positive DNA purification kit (Lucigen, Middleton, WI, U.S.A.), according to manufacturer instructions. Ten-kilobase genomic libraries were prepared and size-selected as previously described (Booher et al. 2015). The libraries were multiplexed on a SMRT cell (10, 24, or 48 libraries per cell) on a SEQUEL I or SEQUEL II machine (Pacific Biosciences, Menlo Park, CA, U.S.A.) at the Icahn School of Medicine at Mt. Sinai (New York). The genomes were assembled using the HGAP assembler version 4.0 or the PacBio Microbial Assembler and were annotated via the National Center for Biotechnology Information (NCBI) with the Prokaryotic Genome Annotation Pipeline (Chin et al. 2013; Li et al. 2021).

Pathogenicity assays.

Strains were grown on glucose yeast extract (GYE) (5 g of glucose and 10 g of yeast extract per liter) agar plates for 24 h before being collected with a spatula and resuspended in 10 mM $MgCl_2$. The cell suspension was adjusted to an optical density at 600 nm (OD_{600}) of 0.1 (about 10^8 CFU/mL) in 50 ml, and Silwet (Phytotech Labs, Lenexa, KS) was added to 0.0125%. Plants were grown in 4-inch pots in a chamber at 90% relative humidity for 4 weeks with 12-h days at 25°C and 12-h nights at 21°C. Plants were inoculated by clipping two leaves and dipping the entire leaves in the prepared inoculum for 10 s. Plants were then returned to the growth chamber for 14 days, after which leaves were observed for the presence of vascular disease or bacterial spot symptoms. This clip-and-dip inoculation technique was used to allow invasion by vascular pathogens through hydathodes and cut veins and by nonvascular pathogens through stomates. A minimum of three technical replicates were completed for each inoculation and all experiments were repeated at least twice. A negative control of 10 mM $MgCl_2$ was included for inoculation in each round of experiments. For the assays in *M. incana* presented in Supplementary Figure S2 only, to assay disease development, fully expanded leaves of 6-week-old plants grown in a greenhouse were inoculated by piercing the middle of the central vein with a bacterial inoculum at 10^8 CFU/ml ($OD_{600} = 0.1$) in 1 mM $MgCl_2$. After inoculation, plants were moved to a growth chamber at 22°C with 70% humidity and an 8-h light period. Disease symptoms were observed 10 days after inoculation. For determination of bacterial populations in *M. incana*, leaves were infiltrated using a 1-ml needleless syringe with a bacterial inoculum at 10^6 CFU/ml ($OD_{600} = 0.001$) in 1 mM $MgCl_2$ and were then moved to a growth chamber as above. After 3 days, leaf discs were sampled using a 0.65 cm diameter cork borer (surface area 0.33 cm^2) and were individually ground using a TissueLyser MM 400 grinder (Retsch, Hann, Germany), two times for 30 s at a frequency of 30 per second with two glass beads (diameter 4.5 mm) in 200 µl of sterile water. The homogenates were serially diluted in sterile water and 5-µl drops were spotted three times on MOKA (Blanvillain et al. 2007) agar supplemented with rifampicin (50 µg/ml) and pimaricin (30 µg/ml). Plates were incubated at 28°C for 48 h, and colonies were enumerated in spots containing 1 to 30 colonies. Bacterial densities in leaves were calculated as log CFU per square centimeter.

Avirulence assays.

Bacteria were either grown on GYE agar plates and harvested as described above (Fig. 4A) or were grown in liquid GYE to late log phase (Fig. 4B) and were harvested by centrifugation, then were resuspended to $OD_{600} = 0.2$ in 10 mM $MgCl_2$. Leaves at approximately the middle of the stem of 4- to 6-week-old *Matthiola incana* plants, four to six leaves per plant, were pricked with a pin on the abaxial side and were infiltrated through the pinhole with about 0.1 ml of bacterial suspension or 10 mM $MgCl_2$, using a needleless syringe. For each inoculum, 10 leaves from 10 different plants were used. Inoculated plants were incubated in a growth chamber at 25°C and ambient humidity under constant light and were monitored for up to 48 h for development of the HR.

Phylogeny and ANI analysis.

ML trees were created from whole-genome sequences using REALPHY (Bertels et al. 2014). Orthologous single nucleotide polymorphisms were called within REALPHY, using *X. campestris* pv. *campestris* reference genomes B100, ATCC33913, 8004, and Xca5, and the resulting alignments were merged as a part of the REALPHY pipeline. This merged alignment was used to reconstruct the phylogeny using RAxML HPC Blackbox on the CIPRES Portal (Miller et al. 2010). For the rooted tree (Supplementary Fig. S1) *X. oryzae* PXO99A (GenBank CP000967.2) was used as the outgroup. Support was assessed using 1,000 bootstrap pseudoreplicates. The trees were visualized in FigTree and coloring was added in FigTree or Adobe Illustrator (Adobe Inc., San Jose, CA, U.S.A.). ANI matrices were made with the Enveomics Toolkit ANI Matrix calculator (Rodriguez-R and Konstantinidis 2016).

Effector comparisons.

Presence or absence of T3E genes and *cbsA* was determined by a TBLASTN search using the genomes as a BLAST database in Geneious Prime (Biomatters, San Diego, CA, U.S.A.) queried with amino acid sequences of known *Xanthomonas* T3E proteins (The *Xanthomonas* Resource) and those encoded by the two *cbsA* genes in *X. campestris* pv. *campestris* (Genbank Protein accession numbers QCX69893.1 and QCX65459.1). Hits with greater than 60% query coverage and 60% identity were listed as present. TALE sequences were annotated manually by searching for the TALE repeat motifs in chromosome and plasmid sequences. Some effectors within the same family (e.g., XopE1

and XopE2) were grouped together under the family name (i.e., XopE) due to high sequence similarity.

Cloning of *cbsA* and *X. campestris* transformation.

The 1,701-bp open reading frame of $cbsA_{1701}$ and 1,000 bp upstream of the gene were amplified from *X. campestris* pv. *campestris* 8004 and were cloned into a mini-Tn*7* vector, were subcloned into pCZ1013, and were then transferred into pBBR1-MCS2 (Kovach et al. 1995), using *Eco*RI and *Hin*dIII, yielding pBBR1::$cbsA_{1701}$. *X. campestris* competent cells were transformed by electroporation as described (White and Gonzalez 1991). Transformants were selected on GYE amended with 15 μg of kanamycin per milliliter.

Data availability.

Genome data generated in the current study are available through GenBank (project number PRJNA689092). Individual strain accession numbers are listed in Supplementary Table S1.

ACKNOWLEDGMENTS

The authors thank L.-M. Rodríguez, H. Lange, and M. Arlat for advice on design and execution of the study. We also thank K. Howard, director of the Cornell Weed Garden, for providing rare seed for cruciferous weeds found in New York, and M. Paauw for providing various *X. campestris* pv. *campestris* genomic DNA samples for sequencing. Finally, we thank A. Read, S. Baruah, and T. Bauer for critical feedback on the manuscript and assistance with inoculations.

AUTHOR-RECOMMENDED INTERNET RESOURCES

CIPRES portal: https://www.phylo.org/index.php
Enveomics Collection Toolbox genome based distance matrix calculator: http://enve-omics.ce.gatech.edu/g-matrix
NCBI taxonomy browser: https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi
The *Xanthomonas* Resource: http://www.biopred.net/xanthomonas/t3e.html

LITERATURE CITED

Adachi, H., Kamoun, S., and Maqbool, A. 2019. A resistosome-activated 'death switch'. Nat. Plants 5:457-458.

Adachi, H., Sakai, T., Kourelis, J., Maqbool, A., and Kamoun, S. 2020. Jurassic NLR: Conserved and dynamic evolutionary features of the atypically ancient immune receptor ZAR1. bioRxiv:2020.2010.2012.333484.

Arlat, M., Gough, C. L., Barber, C. E., Boucher, C., and Daniels, M. J. 1991. *Xanthomonas campestris* contains a cluster of *hrp* genes related to the larger *hrp* cluster of *Pseudomonas solanacearum*. Mol. Plant-Microbe Interact. 4:593-601.

Bella, P., Moretti, C., Licciardello, G., Strano, C. P., Pulvirenti, A., Alaimo, S., Zaccardelli, M., Branca, F., Buonaurio, R., Vicente, J. G., and Catara, V. 2019. Multilocus sequence typing analysis of Italian *Xanthomonas campestris* pv. *campestris* strains suggests the evolution of local endemic populations of the pathogen and does not correlate with race distribution. Plant Pathol. 68:278-287.

Bertels, F., Silander, O. K., Pachkov, M., Rainey, P. B., and van Nimwegen, E. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. Mol. Biol. Evol. 31:1077-1088.

Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denance, N., Vasse, J., Lauber, E., and Arlat, M. 2007. Plant carbohydrate scavenging through *tonB*-dependent receptors: A feature shared by phytopathogenic and aquatic bacteria. PLoS One 2:e224.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509-1512.

Bogdanove, A. J., Koebnik, R., Lu, H., Furutani, A., Angiuoli, S. V., Patil, P. B., Van Sluys, M. A., Ryan, R. P., Meyer, D. F., Han, S. W., Aparna, G., Rajaram, M., Delcher, A. L., Phillippy, A. M., Puiu, D., Schatz, M. C., Shumway, M., Sommer, D. D., Trapnell, C., Benahmed, F., Dimitrov, G., Madupu, R., Radune, D., Sullivan, S., Jha, G., Ishihara, H., Lee, S. W., Pandey, A., Sharma, V., Sriariyanun, M., Szurek, B., Vera-Cruz, C. M., Dorman, K. S., Ronald, P. C., Verdier, V., Dow, J. M., Sonti, R. V., Tsuge, S., Brendel, V. P., Rabinowicz, P. D., Leach, J. E., White, F. F., and Salzberg, S. L. 2011. Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas spp.* J. Bacteriol. 193:5450-5464.

Bolot, S., Guy, E., Carrere, S., Barbe, V., Arlat, M., and Noel, L. D. 2013a. Genome sequence of *Xanthomonas campestris* pv. *campestris* strain Xca5. Genome Announc. 1:e00032-00012.

Bolot, S., Roux, B., Carrere, S., Jiang, B. L., Tang, J. L., Arlat, M., and Noel, L. D. 2013b. Genome sequences of three atypical *Xanthomonas campestris* pv. *campestris* strains, CN14, CN15, and CN16. Genome Announc. 1:e00465-00413.

Booher, N. J., Carpenter, S. C., Sebra, R. P., Wang, L., Salzberg, S. L., Leach, J. E., and Bogdanove, A. J. 2015. Single molecule real-time sequencing of *Xanthomonas oryzae* genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. Microb. Genom. 1:e000032.

Burkholder, W. H. 1941. The black rot of *Barbarea vulgaris*. Phytopathology 31:347-348.

Cerutti, A., Jauneau, A., Auriac, M. C., Lauber, E., Martinez, Y., Chiarenza, S., Leonhardt, N., Berthome, R., and Noel, L. D. 2017. Immunity at cauliflower hydathodes controls systemic infection by *Xanthomonas campestris* pv *campestris*. Plant Physiol. 174:700-716.

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., and Korlach, J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10:563-569.

Denance, N., Szurek, B., Doyle, E. L., Lauber, E., Fontaine-Bodin, L., Carrere, S., Guy, E., Hajri, A., Cerutti, A., Boureau, T., Poussier, S., Arlat, M., Bogdanove, A. J., and Noel, L. D. 2018. Two ancestral genes shaped the *Xanthomonas campestris* TAL effector gene repertoire. New Phytol. 219:391-407.

Dubrow, Z. E., and Bogdanove, A. J. 2021. Genomic insights advance the fight against black rot of crucifers. J. Gen. Plant Pathol. 87:127-136.

Fargier, E., Fischer-Le Saux, M., and Manceau, C. 2011. A multilocus sequence analysis of *Xanthomonas campestris* reveals a complex structure within crucifer-attacking pathovars of this species. Syst. Appl. Microbiol. 34:156-165.

Fargier, E., and Manceau, C. 2007. Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *X. campestris* pv. *campestris*. Plant Pathol. 56:805-818.

Feng, B., and Tang, D. 2019. Mechanism of plant immune activation and signaling: Insight from the first solved plant resistosome structure. J. Integr. Plant Biol. 61:902-907.

Feng, F., Yang, F., Rong, W., Wu, X., Zhang, J., Chen, S., He, C., and Zhou, J.-M. 2012. A *Xanthomonas* uridine 5′-monophosphate transferase inhibits plant immune kinases. Nature 485:114-118.

Gluck-Thaler, E., Cerutti, A., Perez-Quintero, A. L., Butchacas, J., Roman-Reyna, V., Madhavan, V. N., Shantharaj, D., Merfa, M. V., Pesce, C., Jauneau, A., Vancheva, T., Lang, J. M., Allen, C., Verdier, V., Gagnevin, L., Szurek, B., Beckham, G. T., De La Fuente, L., Patel, H. K., Sonti, R. V., Bragard, C., Leach, J. E., Noel, L. D., Slot, J. C., Koebnik, R., and Jacobs, J. M. 2020. Repeated gain and loss of a single gene modulates the evolution of vascular plant pathogen lifestyles. Sci. Adv. 6:eabc4516.

Guy, E., Genissel, A., Hajri, A., Chabannes, M., David, P., Carrere, S., Lautier, M., Roux, B., Boureau, T., Arlat, M., Poussier, S., and Noël, L. D. 2013b. Natural genetic variation of *Xanthomonas campestris* pv. *campestris* pathogenicity on *Arabidopsis* revealed by association and reverse genetics. mBio 4:e00538-00512.

Guy, E., Lautier, M., Chabannes, M., Roux, B., Lauber, E., Arlat, M., and Noel, L. D. 2013a. *xopAC*-triggered immunity against *Xanthomonas* depends on *Arabidopsis* receptor-like cytoplasmic kinase genes *PBL2* and *RIPK*. PLoS One 8:e73469.

He, Y.-Q., Zhang, L., Jiang, B.-L., Zhang, Z.-C., Xu, R.-Q., Tang, D.-J., Qin, J., Jiang, W., Zhang, X., Liao, J., Cao, J.-R., Zhang, S.-S., Wei, M.-L., Liang, X.-X., Lu, G.-T., Feng, J.-X., Chen, B., Cheng, J., and Tang, J.-L. 2007. Comparative and functional genomics reveals genetic diversity and determinants of host specificity among reference strains and a large collection of Chinese isolates of the phytopathogen *Xanthomonas campestris* pv. *campestris*. Genome Biol. 8:R218.

Ignatov, A., Sechler, A., Schuenzel, E. L., Agarkova, I., Oliver, B., Vidaver, A. K., and Schaad, N. W. 2007. Genetic diversity in populations of *Xanthomonas campestris* pv. *campestris* in cruciferous weeds in central coastal California. Phytopathology 97:803-812.

Jacobs, J. M., Pesce, C., Lefeuvre, P., and Koebnik, R. 2015. Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. Front. Plant Sci. 6:431.

Jha, G., Rajeshwari, R., and Sonti, R. V. 2007. Functional interplay between two *Xanthomonas oryzae* pv. oryzae secretion systems in modulating virulence on rice. Mol. Plant-Microbe Interact. 20:31-40.

Kay, S., Boch, J., and Bonas, U. 2005. Characterization of AvrBs3-like effectors from a *Brassicaceae* pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. Mol. Plant-Microbe Interact. 18:838-848.

Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., 2nd, and Peterson, K. M. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene 166:175-176.

Krauthausen, H.-J., Hörner, G., Zimmermann, S., Voegele, R., and Brändle, F. 2018. Competence of *Xanthomonas campestris* from cruciferous weeds and wallflower (*Erysimum cheiri)* to induce black rot in cabbage. Eur. J. Plant Pathol. 151:275-289.

Laflamme, B., Dillon, M. M., Martel, A., Almeida, R. N. D., Desveaux, D., and Guttman, D. S. 2020. The pan-genome effector-triggered immunity landscape of a host-pathogen interaction. Science 367:763-768.

Lange, H. W., Tancos, M. A., Carlson, M. O., and Smart, C. D. 2016. Diversity of *Xanthomonas campestris* isolates from symptomatic crucifers in New York State. Phytopathology 106:113-122.

Lange, H. W., Tancos, M. A., and Smart, C. D. 2022. Cruciferous weeds do not act as major reservoirs of inoculum for black rot outbreaks in New York state. Plant Dis. 106:174-181.

Lee, Y. A., Yang, P. Y., and Huang, S. C. 2020. Characterization, phylogeny, and genome analyses of nonpathogenic *Xanthomonas campestris* strains isolated from *Brassica* seeds. Phytopathology 110:981-988.

Li, W., O'Neill, K. R., Haft, D. H., DiCuccio, M., Chetvernin, V., Badretdin, A., Coulouris, G., Chitsaz, F., Derbyshire, M. K., Durkin, A. S., Gonzales, N. R., Gwadz, M., Lanczycki, C. J., Song, J. S., Thanki, N., Wang, J., Yamashita, R. A., Yang, M., Zheng, C., Marchler-Bauer, A., and Thibaud-Nissen, F. 2021. RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res. 49:D1020-D1028.

Liu, H., Zhang, S., Schell, M. A., and Denny, T. P. 2005. Pyramiding unmarked deletions in *Ralstonia solanacearum* shows that secreted proteins in addition to plant cell-wall-degrading enzymes contribute to virulence. Mol. Plant-Microbe Interact. 18:1296-1305.

Meline, V., Delage, W., Brin, C., Li-Marchetti, C., Sochard, D., Arlat, M., Rousseau, C., Darrasse, A., Briand, M., Lebreton, G., Portier, P., Fischer-Le Saux, M., Durand, K., Jacques, M. A., Belin, E., and Boureau, T. 2019. Role of the acquisition of a type 3 secretion system in the emergence of novel pathogenic strains of *Xanthomonas*. Mol. Plant Pathol. 20:33-50.

Meng, X., and Zhang, S. 2013. MAPK cascades in plant disease resistance signaling. Annu. Rev. Phytopathol. 51:245-266.

Miller, M., Pfeiffer, W., and Schwartz, T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. Pages 11572-11578 in: Gateway Computing Environments Workshop (GCE), 2010. Institute of Electrical and Electronics Engineers, Piscataway, NJ.

Moscou, M. J., and Bogdanove, A. J. 2009. A simple cipher governs DNA recognition by TAL effectors. Science 326:1501-1501.

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., Couloux, A., Darrasse, A., Gouzy, J., Jacques, M. A., Lauber, E., Manceau, C., Mangenot, S., Poussier, S., Segurens, B., Szurek, B., Verdier, V., Arlat, M., and Rott, P. 2009. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. BMC Genomics 10:616.

Rodriguez-R, L. M., and Konstantinidis, K. T. 2016. The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr. 4:e1900v1901.

Roux, B., Bolot, S., Guy, E., Denance, N., Lautier, M., Jardinaud, M. F., Fischer-Le Saux, M., Portier, P., Jacques, M. A., Gagnevin, L., Pruvost, O., Lauber, E., Arlat, M., Carrere, S., Koebnik, R., and Noel, L. D. 2015. Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectome. BMC Genomics 16:975.

Schultink, A., Qi, T., Bally, J., and Staskawicz, B. 2019. Using forward genetics in *Nicotiana benthamiana* to uncover the immune signaling pathway mediating recognition of the *Xanthomonas perforans* effector XopJ4. New Phytol. 221:1001-1009.

Tancos, M. A., Dubrow, Z. E., Carpenter, S. C. D., and Bogdanove, A. J. 2022. Genome sequence of *Xanthomonas campestris* strain FDWSRU 18048, an emerging pathogen of nonnative, invasive garlic mustard (*Alliaria petiolata*). Microbiology Resource Announcements: in press.

Vauterin, L., Hoste, B., Kersters, K., and Swings, J. 1995. Reclassification of *Xanthomonas*. Int. J. Syst. Evol. Microbiol. 45:472-489.

Wang, G., Roux, B., Feng, F., Guy, E., Li, L., Li, N., Zhang, X., Lautier, M., Jardinaud, M. F., Chabannes, M., Arlat, M., Chen, S., He, C., Noel, L. D., and Zhou, J. M. 2015. The decoy substrate of a pathogen effector and a pseudokinase specify pathogen-induced modified-self recognition and immunity in plants. Cell Host Microbe 18:285-295.

White, T. J., and Gonzalez, C. F. 1991. Application of electroporation for efficient transformation of *Xanthomonas campestris* pv. oryzae. Phytopathology 81:521-524.

Xu, R.-Q., Blanvillain, S., Feng, J.-X., Jiang, B.-L., Li, X.-Z., Wei, H.-Y., Kroj, T., Lauber, E., Roby, D., Chen, B., He, Y.-Q., Lu, G.-T., Tang, D.-J., Vasse, J., Arlat, M., and Tang, J.-L. 2008. AvrAC Xcc8004, a type III effector with a leucine-rich repeat domain from *Xanthomonas campestris* pathovar *campestris* confers avirulence in vascular tissues of *Arabidopsis thaliana* ecotype Col-0. J. Bacteriol. 190:343-355.