Phytopathology • 2020 • 110:1161-1173 • https://doi.org/10.1094/PHYTO-03-19-0077-R

Pathogen Spotlight

Spotlight: Bacteriology

e-Xtra*

Genomic Acquisitions in Emerging Populations of *Xanthomonas vasicola* pv. *vasculorum* Infecting Corn in the United States and Argentina

Alvaro L. Perez-Quintero,[1,†] Mary Ortiz-Castro,[1] Jillian M. Lang,[1] Adrien Rieux,[2] Guangxi Wu,[1] Sanzhen Liu,[3] Toni A. Chapman,[4] Christine Chang,[5] Janet Ziegle,[5] Zhao Peng,[6] Frank F. White,[6] Maria Cristina Plazas,[7] Jan E. Leach,[1] and Kirk Broders[1,8,†]

[1] Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO, U.S.A.
[2] CIRAD, UMR PVBMT, St. Pierre, Réunion, France
[3] Department of Plant Pathology, Kansas State University, Manhattan, KS, U.S.A.
[4] Biosecurity and Food Safety, NSW Department of Primary Industries, Elizabeth Macarthur Agricultural Institute, Menangle, NSW, Australia
[5] Pacific Biosciences, Menlo Park, CA, U.S.A.
[6] Department of Plant Pathology, University of Florida, Gainesville, FL, U.S.A.
[7] Laboratorio de Fitopatología y Microbiología, Universidad Católica de Córdoba, Ob. Trejo 323, Córdoba, Argentina
[8] Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Ancon, Republic of Panamá
Accepted for publication 5 February 2020.

ABSTRACT

Xanthomonas vasicola pv. *vasculorum* is an emerging bacterial plant pathogen that causes bacterial leaf streak on corn. First described in South Africa in 1949, reports of this pathogen have greatly increased in the past years in South America and in the United States. The rapid spread of this disease in North and South America may be due to more favorable environmental conditions, susceptible hosts and/or genomic changes that favored the spread. To understand whether genetic mechanisms exist behind the recent spread of *X. vasicola* pv. *vasculorum*, we used comparative genomics to identify gene acquisitions in *X. vasicola* pv. *vasculorum* genomes from the United States and Argentina. We sequenced 41 genomes of *X. vasicola* pv. *vasculorum* and the related sorghum-infecting *X. vasicola* pv. *holcicola* and performed comparative analyses against all available *X. vasicola* genomes. Time-measured phylogenetic analyses showed that *X. vasicola* pv. *vasculorum* strains from the United States and Argentina are closely related and arose from two introductions to North and South America. Gene content comparisons identified clusters of genes enriched in corn *X. vasicola* pv. *vasculorum* that showed evidence of horizontal transfer including one cluster corresponding to a prophage found in all *X. vasicola* pv. *vasculorum* strains from the United States and Argentina as well as in *X. vasicola* pv. *holcicola* strains. In this work, we explore the genomes of an emerging phytopathogen population as a first step toward identifying genetic changes associated with the emergence. The acquisitions identified may contain virulence determinants or other factors associated with the spread of *X. vasicola* pv. *vasculorum* in North and South America and will be the subject of future work.

Keywords: bacteriology, corn, genetics and resistance, horizontal gene transfer, population biology, *Xanthomonas vasicola* pv. *vasculorum*, *Xanthomonas vasicola* pv. *holcicola*

Given the large number of acres and economic importance of corn production in the United States, there are important implications to the emergence and spread of new diseases. In the United States, bacterial leaf streak (BLS) of corn was first observed in Nebraska in 2014 and became widespread by 2016 (Korus et al. 2017). The disease now occurs in most of the corn growing regions of the United States including the top three corn producing states: Illinois, Iowa, and Nebraska (Korus et al. 2017; USDA-NASS 2017) as well as Colorado and Kansas, with field disease incidence levels reported to be above 90% and disease severity reaching greater than 50% of leaf area infected (Broders 2017). Understanding how this disease originated and what favors its spread is a crucial step to preventing future losses.

Caused by *X. vasicola* pv. *vasculorum*, BLS was first described in 1949 on corn in South Africa (Dyer 1949), but prior to 2016 it had not been documented in any other country. It is unknown how this organism was introduced to the United States or if it was already present but latent. The only other reports of BLS of corn outside of South Africa and the United States were in Argentina and Brazil in 2017 and 2018 (Leite et al. 2018; Plazas et al. 2017). While the official report of the disease in Argentina is relatively recent, the symptoms of BLS were first observed in 2010 in the Cordoba province and have since spread to nine other corn-producing provinces, including provinces that border the corn growing regions in Brazil and Paraguay (Leite et al. 2018; Plazas et al. 2017). It is still unclear why *X. vasicola* pv. *vasculorum* continues to spread in North and South America or how severe future epidemics may become.

A significant amount of confusion existed around the taxonomic classification of this bacterium. The nomenclature had gone through several changes, from *X. campestris* pv. *zeae* to *X. vasicola* pv. *zeae* to its current designation as *X. vasicola* pv. *vasculorum* (Bradbury 1986; Coutinho and Wallis 1991; Lang et al. 2017; Qhobela et al. 1990; Sanko et al. 2018). The *X. vasicola* species can be divided into five groups including three described pathovars: (i) *X. vasicola* pv. *vasculorum* infecting corn and sugarcane, (ii) *X. vasicola* pv. *holcicola*, commonly infecting sorghum, (iii) *X. vasicola* pv. *musacearum* infecting enset and banana, (iv) a group of strains isolated from *Tripsacum laxum*, and (v) strains isolated from areca nut (previously *X. campestris* pv. *arecae*) (Lang et al. 2017; Studholme et al. 2020).

†Corresponding authors: A. L. Perez-Quintero; alperezqui@gmail.com, and K. Broders; brodersk@si.edu

Funding: This work was funded by grants from the Colorado Corn Administrative Committee, the Animal and Plant Health Inspection Service (project #6.0533.01), and the Foundation for Food and Agriculture Research (project #544722).

*The *e*-Xtra logo stands for "electronic extra" and indicates that eight supplementary figures and four supplementary tables are published online.

The author(s) declare no conflict of interest.

The term pathovar refers to a strain or set of strains with the same or similar characteristics, differentiated at the infrasubspecific level from other strains of the same species or subspecies on the basis of distinctive pathogenicity to one or more plant hosts (Young et al. 1991). While the described pathovars of *X. vasicola* seem to have defined host preferences, their host ranges may be broader than initially claimed. *X. vasicola* pv. *holcicola* and *X. vasicola* pv. *vasculorum*, in particular, may have overlapping host ranges. Under laboratory conditions, isolates of *X. vasicola* pv. *vasculorum* from corn and sugarcane caused disease on corn, sugarcane, and sorghum but were most virulent on corn and sugarcane (Lang et al. 2017). Similarly, when infiltrated into leaves, *X. vasicola* pv. *holcicola* infected corn, sorghum, and sugarcane but caused more disease on sorghum (Lang et al. 2017). *X. vasicola* pv. *vasculorum* has not been isolated from sorghum, while *X. vasicola* pv. *holcicola* has occasionally been isolated from corn in the field (Moffett 1983; Péros et al. 1994). Upon inoculation in the greenhouse, *X. vasicola* pv. *vasculorum* isolates from the United States can infect 16 hosts, mostly monocots such as rice, oats, and big blue stem and the dicot yellow nutsedge (Hartman et al. 2020). Field studies confirmed these results for big blue stem and bristly foxtail as hosts in a natural inoculum system (Hartman et al. 2020). At least two host jumps have been hypothesized for *X. vasicola*, i.e., from grasses to banana (Tushemereirwe et al. 2004) and from sugarcane to *Eucalyptus* spp, a dicot (Coutinho et al. 2015), suggesting a remarkable adaptive ability for the species.

The spread of *X. vasicola* pv. *vasculorum* in the United States and Argentina was rapid. How and why the populations expanded so quickly in two countries located on the opposite side of the equator at approximately the same time, while the disease remained rarely documented in South Africa during the same time period, is intriguing. Reasons for these disparate observations could include the occurrence of more favorable environmental conditions and susceptible corn germplasm in North and South America versus South Africa, differences in surveillance and reporting of the pathogen, and/or, as we hypothesize here, the acquisition of genetic features that favored infection or spread or virulence. In this study, we employed a comparative genomics approach to identify genetic changes associated to these emerging populations.

MATERIALS AND METHODS

Strain collection and molecular detection. Isolation of *X. vasicola* pv. *vasculorum* from corn leaves was performed as in Lang et al. (2017) with minor modifications. Instead of placing the tissue in distilled water, fresh tissues were dissolved in 1 ml of 10 mM $MgCl_2$, macerated with sterile pellet pestle, and incubated for at least 1.5 h at room temperature. For bacterial isolation, one loop-fill (10 µl) of solution was spread onto nutrient agar. Plates were incubated at 28°C for 2 days. Single characteristic bright yellow colonies were selected and restreaked for further isolation until pure colonies were obtained. Samples from the United States were collected across several fields in Colorado, Iowa, Kansas, and Nebraska. Samples from Argentina were collected from fields located in San Luis, Córdoba, and Santa Fe states (Supplementary Table S1).

South African *X. vasicola* pv. *vasculorum* strains were obtained from the L. E. Claflin collection (Qhobela et al. 1990) maintained at Colorado State University. Australian *X. vasicola* pv. *holcicola* strains were obtained from the NSW Department of Primary Industries Plant Pathology and Mycology Herbarium Culture Collection (https://www.dpi.nsw.gov.au/about-us/services/collections/herbarium).

Molecular detection of *X. vasicola* pv. *vasculorum* was performed following one of these procedures: For the cases using colony PCR, one single colony was suspended in 10 µl of sterile water and boiled at 95°C for 5 min. First, colony PCR of suspected *X. vasicola* pv. *vasculorum* samples were performed using Xvv3 or Xvv5 primers as described previously (Lang et al. 2017). To further confirm isolates, a second method using 16S rRNA gene and a housekeeping gene, *atpD* (ATP synthase β chain), was used to identify bacteria to species level. PCR reactions for 16S rRNA (50 µl) contained 2 µl of boiled DNA template, 0.2 µM of each primer (Supplementary Table S4), 1× GoTaq reaction buffer, 2 mM $MgCl_2$, 0.2 mM dNTP, and 0.25 unit/µl of GoTaq DNA polymerase enzyme (Promega, Madison, WI). The cycling conditions were as follows: initial denaturation at 94°C for 3 min, following 35 cycles of 94°C for 45 s, 50°C for 1 min, and 72°C for 1:30 min, and the final extension period at 72°C for 10 min. PCR fragments were separated in a 1.5% agarose gel for 45 min at 90 V, and fragments were extracted and purified using the DNA clean & concentrator kit (ZYMO Research). Sequencing was performed with 5 ng/µl of each PCR product at Quintara Biosciences (Fort Collins, CO) and analyzed using Geneious software (version 10.0.7). Sequence identities to the genus level were determined using BLASTn from the NCBI database.

Genome sequencing, assembly, and data collection. Genomic DNA for the *Xanthomonas* positive samples was extracted using Easy-DNA kit (Invitrogen), and PCR amplification of the *atpD* gene was carried out for further confirmation to the species level. PCR reactions for *atpD* gene (40 µl) contained 25 ng/µl of DNA template, 0.4 µM of each primer (Supplementary Table S4), 1× GoTaq reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, and 0.1 unit/µl of GoTaq DNA polymerase enzyme. Cycling conditions were performed as described by Fargier et al. (2011).

For 24 *X. vasicola* pv. *vasculorum* strains and one *X. vasicola* pv. *holcicola* strain from the United States, Illumina sequencing was performed by BGI (https://www.bgi.com/) using HiSeq 4000 with paired-end 100 bp reads. All Illumina reads were first trimmed with Trimmomatic (PE ILLUMINACLIP:CO-2.adapters.fa LEADING:2 TRAILING:2 SLIDINGWINDOW:4:2 MINLEN:30) (Bolger et al. 2014) and then assembled into scaffolds using SPAdes (Bankevich et al. 2012) with default settings. For 11 Australian *X. vasicola* pv. *holcicola* strains, sequencing was performed using Illumina Miseq and assembled using the A5 pipeline (Coil et al. 2015).

Five strains (CO-5, XV1601, NE744, Mex-1, and ZCP611) were sequenced using long read, single molecule real-time sequencing (SMRT Sequel, PacBio, Menlo Park, CA). SMRT read sequences were assembled using HGAP v4 (Chin et al. 2013). Genomes were circularized using Circlator (Hunt et al. 2015). For XV1601, Illumina reads were also available and were used to polish the PacBio assembly using Canu v1.3-r7616 (Koren et al. 2017). All generated genomes have been deposited in NCBI under BioProject PRJNA413069 (Supplementary Table S1).

We obtained all available assemblies of *X. vasicola* (NCBI txid56459) and *X. campestris* pv. *musacearum* (NCBI txid454958, here referred to as *X. vasicola* pv. *musacearum*) as of November 2018 (Supplementary Table S1). Assemblies with an N_{50} of minimum 10 kbp were kept (thus excluding *X. vasicola* pv. *vasculorum* strains NCPPB 895 and NCPPB 890). Ten recently published *X. vasicola* pv. *vasculorum* strains form South Africa (named *Xanthomonas vasicola* pv. *zeae*) (Sanko et al. 2018) were reassembled from available Illumina raw reads (Biosample accessions SAMN10286417 to SAMN10286426) using Unicycler v0.4.8-beta (–mode bold), which functions as a SPAdes-optimizer (Wick et al. 2017) (Supplementary Fig. S1). The reassembled genomes have been submitted as Third PArty (TPA) submissions in the International Nucleotide Sequence Database Collaboration (http://www.insdc.org/) (accession numbers pending) and are available upon request.

Phylogenetic analyses. Phylogenetic trees were obtained with various methods using whole genome data. CSI Phylogeny 1.4 (Kaas et al. 2014) (default parameters) was used to obtain trees based on core genome single nucleotide polymorphisms (SNPs), with *X. oryzae* pv. *oryzae* PXO99A (Fig. 1A) or *X. vasicola* pv.

holcicola NCPPB 1060 (Supplementary Fig. S2A) used as reference for SNP calling. KSNP3 (Gardner et al. 2015) was used to obtain maximum-likelihood (ML) trees based on pan-genome SNPs from identified k-mers (K-mer size = 21).

Multilocus sequence alignments (MLSA) neighbor-joining trees were obtained by identifying 31 housekeeping genes using AMPHORA v2 (Kerepesi et al. 2014), creating multiple alignments form their concatenated sequences using MUSCLE v3.8.31 (Edgar 2004), and generating the trees using functions of the R package phangorn (pml, optim.pml [model = "Blosum62"] and bootstrap.pml [bs=100]) (Schliep 2011). Average nucleotide identity (ANI) values were obtained using the ANI-matrix script from the enveomics collection (v1.3) (Rodriguez-R and Konstantinidis 2016).

The obtained CSI phylogeny ML tree of strains with known date of isolation ($n = 92$) spanning 75 years of evolution (1942 to 2017) was used to investigate the presence of temporal signal. For this purpose, we computed the linear regression between sample age and root-to-tip distances at every internal node of the tree using a custom R script sourcing the distRoot function of the adephylo R package (Jombart et al. 2010). One internal node (the *X. vasicola* pv. *vasculorum* United States + Argentina node) at which both the linear regression was significant and the slope positive was assumed to contain detectable amount of evolutionary change and was

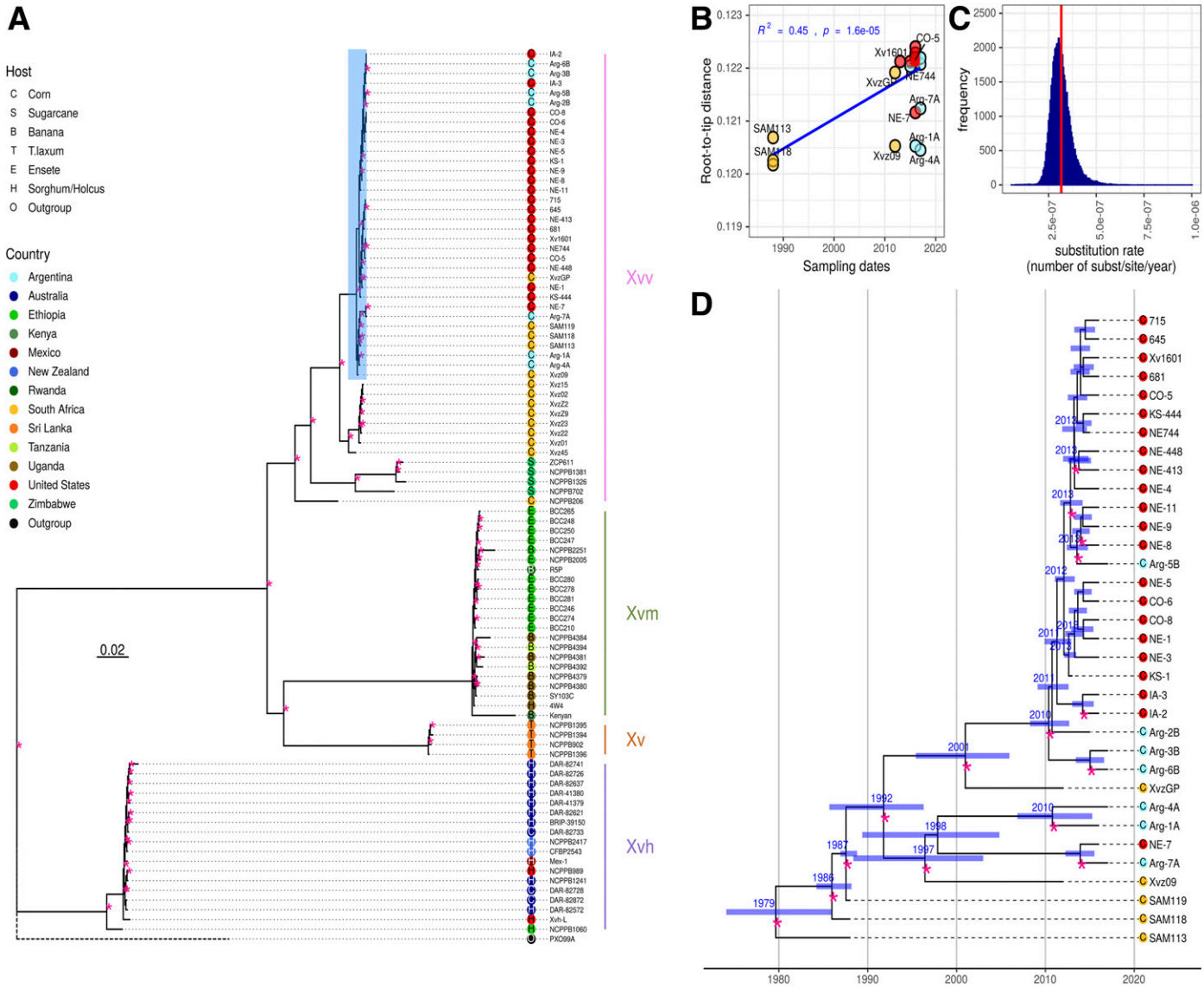

Fig. 1. Phylogeny of *Xanthomonas vasicola* (Xv). **A,** Maximum-likelihood (ML) tree based on core genome single nucleotide polymorphisms (SNPs; 51,969 total) from 91 draft and complete *X. vasicola* genomes. *X. oryzae* strain PXO99A serves as a reference for SNP calling and as the outgroup. ML tree was built using CSI phylogeny (Kaas et al. 2014). The three main pathovars (Xvh = *X. vasicola* pv. *holcicola*, Xvm = *X. vasicola* pv. *musacearum*, and Xvv = *X. vasicola* pv. *vasculorum*), and a set of unassigned strains (Xv), are indicated by solid lines. Colors/shading in tree tips indicate country of isolation and tip letters indicate the plant host of the isolate. Bar shows the tree scale. Dotted line to the out-group (PXO99A) indicates this distance was scaled down 20-fold to improve visualization. Colors in the tip letters (black or white) are for readability and do not indicate any feature. Seventy-three percent of the nodes had support over 0.8 as calculated by CSI phylogeny and are indicated with asterisks. Clade highlighted in a blue/gray rectangle indicates the monophyletic group at which temporal signal was detected and tip-dating inferences performed. This clade contains all contemporary corn *X. vasicola* pv. *vasculorum* (Xvv) strains from the United States and Argentina used in this study. **B,** Root-to-tip regression displaying temporal signal (accumulation of SNPs with time) within the clade highlighted in blue or gray in A. Each point represents a strain in the clade; solid diagonal line indicates a linear regression. **C,** Posterior distribution of substitution rate estimations (in number of substitutions/site/year) in the highlighted clade in A; horizontal line indicates the mean. **D,** Tip-dated Bayesian phylogeny of the strains in the highlighted clade in A obtained from BEAST (Drummond and Rambaut 2007) analysis of nonrecombinant SNPs (243) in this clade. Horizontal axis shows dates in years. Estimated divergence dates for internal nodes prior to 2014 are shown in horizontal blue/gray bars and represent 95% highest posterior density (HPD) for node ages. Asterisks indicate nodes with posterior probabilities over 0.8.

considered as a suitable evolutionary scale for tip-dating to be performed.

A core SNP alignment was then made for the *X. vasicola* pv. *vasculorum* clade containing U.S. and Argentina strains using CSI phylogeny (Kaas et al. 2014) with the *X. vasicola* pv. *vasculorum* CO-5 genome as a reference for SNP calling (1,204 SNPs total). SNPs in recombinant sites (961) were detected using ClonalFrameML (Didelot and Wilson 2015) and excluded from the dating inferences. Tip-dating inferences were performed using BEAST v1.8.4 (Drummond and Rambaut 2007). Nucleotide substitution rate was simulated with a general time-reversible (GTR) substitution model of evolution. Rate variation among sites was modeled with a discrete gamma distribution with four rate categories. We assumed an uncorrelated lognormal relaxed clock to account for rate variation among lineages. To minimize prior assumptions about demographic history, we adopted an extended Bayesian skyline plot approach in order to integrate data over different coalescent histories. The tree was calibrated using tip-dates only. To do so, we applied flat priors (i.e., uniform distributions) for the substitution rate (1×10^{-12} to 1×10^{-2} substitutions/site/year) as well as for the age of any internal node in the tree. We ran five independent chains in which samples were drawn every 5,000 Markov chain Monte Carlo steps from a total of 50,000,000 steps, after a discarded burn-in of 5,000,000 steps. Convergence to the stationary distribution and sufficient sampling and mixing were checked by inspection of posterior samples (effective sample size >200). Parameter estimation was based on the samples combined from the different chains. The best-supported tree was estimated from the combined samples using the maximum clade credibility method implemented in TreeAnnotator (Drummond and Rambaut 2007).

An SNP tree was also created using only the Cluster E/Prophage region of *X. vasicola* pv. *vasculorum* CO-5 as reference, using CSI phylogeny (Kaas et al. 2014) (all filters disabled, 884 SNPs total). This tree and SNP alignment was used to define types of the prophage region. Four types were defined by applying the cutree function in R (h = 1.5).

Genome annotation and ortholog identification. All assemblies were automatically annotated using Prokka v1.14-dev (–rfam) (Seemann 2014). Ortholog groups from Prokka-annotated proteins were identified using OrthoFinder v. 2.2.6 (–M dendroblast, –S blast) (Emms and Kelly 2015) and Pan-X (default parameters) (Ding et al. 2018). Similar ortholog gene groups with similar distribution were found with both strategies (Pan-X = 6,155 groups and 2,163 unassigned genes, OrthoFinder = 6,084 groups and 1,896 unassigned genes). OrthoFinder results were kept for further analyses. These results were explored using an UpSet plot, finding 3,616 genes were present in all the *X. vasicola* groups and 44 genes are found exclusively in corn *X. vasicola* pv. *vasculorum* (Supplementary Fig. S4). Nineteen genes were shared exclusively between U.S. and Argentinian strains, while two and eight genes were found uniquely in U.S. and Argentinian strains, respectively (Supplementary Fig. S4).

Identification of enriched regions. Sixty-four orthologs were shared between *X. vasicola* pv. *holcicola* and U.S. and Argentina *X. vasicola* pv. *vasculorum* (Supplementary Fig. S4) but not South African *X. vasicola* pv. *vasculorum*, indicating that genes may be associated with the U.S. and Argentina *X. vasicola* pv. *vasculorum* population without being exclusive to it. To uncover such enriched genes, a hypergeometric test was designed and applied to each ortholog gene group identified with OrthoFinder to look for over- or underrepresented genes in corn U.S. and Argentina *X. vasicola* pv. *vasculorum* strains. The test was applied using the function phyper (q, m, n, k, log.p = FALSE) in R, where for each ortholog: q = United States + Argentina *X. vasicola* pv. *vasculorum* strains that contain the gene, m = total number of United States + Argentina *X. vasicola* pv. *vasculorum* strains, n = number of strains in the comparison group, and k = total strains that contain the gene in both groups.

The test was applied for each gene in both directions for overrepresentation and underrepresentation (phyper parameters q = q – 1 + lower.tail = FALSE, and lower.tail = TRUE, respectively), and the lowest *P* value of the two tests was chosen (if the lowest *P* value was for the underrepresentation test, it was multiplied by –1 to differentiate them). The test was applied 100 times for each gene, each time changing the comparison group by randomly selecting a group of strains from the comparison groups of a random size between 10 and 62 (total *X. vasicola* strains that are not *X. vasicola* pv. *vasculorum* from the United States or Argentina). The average *P* value of the 100 tests was taken for each gene, and a correction for multiple testing (p.adjust function in R, method BH [Benjamini and Hochberg 1995]) was applied to the *P* values obtained for all genes. Genes with an absolute adjusted $P < 0.01$ were considered as enriched in the U.S. and Argentina *X. vasicola* pv. *vasculorum*.

The pan-genome-wide association studies (pan-GWAS) strategy Scoary (Brynildsrud et al. 2016) was also used (default parameters, with no input tree) to determine enrichment of genes in the U.S. and Argentina *X. vasicola* pv. *vasculorum* strains; for this, the program was run using a matrix of counts for ortholog groups in each genome obtained with OrthoFinder (Emms and Kelly 2015) and using the classification of *X. vasicola* pv. *vasculorum* United States + Argentina versus other as the trait of interest. Genes determined as enriched in *X. vasicola* pv. *vasculorum* United States + Argentina with an adjusted $P < 0.01$ (Bonferroni *P*) were kept for analysis.

The position of the selected genes in the genome of the strain CO-5 was then used to establish clusters. Groups of more than 10 overrepresented genes found less than 5 kb from each other were considered a cluster and assigned a letter (A to E) according to their distance to the replication origin, using the *dnaA* gene as reference in its 5′ to 3′ orientation in the plus strand.

Annotation of genomic regions. Protein sequences of the *X. vasicola* pv. *vasculorum* CO-5 strain were further annotated using Blast2GO v5.2.5 (Conesa et al. 2005) by combining hits against the NCBI nr-protein database (BLAST-p fast, e-value 0.01, number of hits 10), InterPro, Gene Ontology (GOs)terms, and KEGG enzyme codes (default parameters). Enrichment of GO terms was assessed for the different groups using a hyper-geometric test as implemented in the GoFuncR package (Grote 2018).

Genomic islands were predicted using the IslandViewer 4 suite, which compiles parametric and phylogenetic methods for genomic island prediction (Bertelli et al. 2017). Insertion sequences were identified using ISEScan (v1.6) (default parameters) (Xie and Tang 2017). Possible prophages were identified using PHASTER (Arndt et al. 2016). Novel type III (T3) secreted proteins were predicted using EffectiveDB (default parameters + plant model for Predotar). This suite allows identification of putative T3Es based on prediction of secretion signals, T3 chaperone binding domains, eukaryotic-like domains, and eukaryotic subcellular localization (Eichinger et al. 2016), and results were filtered to keep proteins with an EffectiveT3 (signal peptide) of minimum 0.9999, plus any additional predictions with other methods. In order to be able to accurately correlate the genomic position of clusters and annotations, and because IslandViewer 4 (Xie and Tang 2017) and PHASTER (Arndt et al. 2016) are better suited for complete genomes, these analyses were made using Prokka-annotated files only for strains with complete genomes.

Known T3 effectors were identified by BLASTp (v. 2.6.0+, results were filtered keeping hits with e-value of <0.0001, >30% identity in >40% the query length) (Altschul et al. 1997) of consensus effectors sequences obtained from http://xanthomonas.org/ against the protein sequences of *X. vasicola* strains obtained using Prokka.

Kaiju (-r species -l super kingdom, family, genus, species) (Menzel et al. 2016) was used to annotate possible taxonomic origin of cluster genes using the CO-5 sequences against the progenomes database (Mende et al. 2017).

Visualization and other analyses. Most figures were generated using R (R Core Team 2013). Phylogenetic trees and multiple sequence alignments were visualized using the ggtree package (Yu et al. 2017). Linear genomic region visualizations were generated using ggbio (Yin et al. 2012). Genomic maps and synteny visualizations were generated with genoPlotR (Guy et al. 2010). Heatmaps were generated using pheatmap (Kolde and Kolde 2015). Upset plot was generated using UpsetR (Conway et al. 2017). Taxonomy treemap visualizations were made with treemap (Tennekes and Ellis 2017). Comparisons of genomic regions and definitions of clusters were made using GenomicRanges (Lawrence et al. 2013). Genomic alignments were visualized using Mauve v. Jan-19-2018 (Darling et al. 2004).

RESULTS

***X. vasicola* pv. *vasculorum* strains from the United States and Argentina are closely related and arose from two introduction events.** Draft genome assemblies were generated using short-read sequencing (Illumina Miseq) for *X. vasicola* pv. *vasculorum* strains isolated from corn between 2016 and 2017 in the states of Colorado, Iowa, Kansas, and Nebraska in the United States (15 strains) and in the states of Cordoba, Santa Fe, and San Luis in Argentina (7 strains). Draft genomes were also generated for available *X. vasicola* pv. *vasculorum* isolates from South Africa (2 strains from corn) and for *X. vasicola* pv. *holcicola* isolates from the United States (1 strain, from sorghum) and Australia (8 strains from sorghum, 3 from corn). The draft genomes were composed of 158 to 827 contigs (average 243 contigs), had average N_{50} values of 120 kb, and had estimated completeness over 96% (Supplementary Table S1).

Fully assembled genomes were generated using long read sequencing (Pacbio Sequel) for three U.S. *X. vasicola* pv. *vasculorum* isolates recovered from corn, one sugarcane *X. vasicola* pv. *vasculorum* isolate from Zimbabwe, and one sorghum *X. vasicola* pv. *holcicola* isolate from Mexico. With the exception of the sugarcane strain (ZCP611), these genomes were each assembled to a single chromosome ~4.9 Mbp in size. Strain ZCP611 contained a ~4.0 Mb chromosome and a 76.6 kb contig corresponding to a possible plasmid, as determined by mlplasmids (Arredondo-Alonso et al. 2018). This dataset represents a total of 41 new *X. vasicola* genomes (Supplementary Table S1).

All other available *X. vasicola* genomes were obtained from the NCBI for comparative analyses. For 10 recently published *X. vasicola* pv. *vasculorum* strains from South Africa (Sanko et al. 2018), the available assemblies ranged in size from 3.8 to 4.5 Mbp, significantly less than the average size for other *X. vasicola* genomes (~4.9 Mbp), and alignments to reference genomes revealed large fragments missing from the assemblies (Supplementary Fig. S1). These genomes were thus reassembled from available Illumina raw reads (Biosample accessions SAMN10286417 to SAMN10286426). The new assemblies had expected sizes and were not missing large regions and were thus kept for analysis in this work (Supplementary Fig. S1).

An ML phylogeny was built using core genome SNPs (Kaas et al. 2014), identified in the newly sequenced genomes as well as all available *X. vasicola* genomes, adding *X. oryzae* pv. *oryzae* strain PXO99A as an outgroup (92 genomes in total, 51,969 SNPs) (Supplementary Table S1). The tree reveals division of the three described pathovars in the species: *holcicola*, *musacearum*, and *vasculorum* in three distinct clades. An additional clade consists of the genomes of strains isolated from *Tripsacum laxum*, not associated yet to any described pathovar since pathogenicity characterization has not been performed (Studholme et al. 2020); these strains are referred to here as *X. vasicola* (Fig. 1A). Most *X. vasicola* pv. *vasculorum* strains from corn formed a closely related group separate from strains isolated from sugarcane, with the exception of strain NCPPB 206, a weakly virulent isolate from South Africa collected in 1948 (Dyer 1949; Lang et al. 2017). The tree showed that all *X. vasicola* pv. *vasculorum* strains from the United States and Argentina form a single clade along with five *X. vasicola* pv. *vasculorum* strains from South Africa (Fig. 1A). Similar groupings were seen in phylogenetic trees based on only *X. vasicola* SNPs (excluding *X. oryzae* PXO99A), pan-genome SNPs, MLSA (Supplementary Fig. S2), and average nucleotide identity (Supplementary Fig. S3), one difference being that in the MLSA tree, *X. vasicola* pv. *musacearum* strains grouped together with sugarcane *X. vasicola* pv. *vasculorum* strains and the corn *X. vasicola* pv. *vasculorum* strain NCPPB 206, suggesting possible recombination between these *X. vasicola* pv. *vasculorum* strains and *X. vasicola* pv. *musacearum* (Supplementary Fig. S2).

The core SNP ML phylogeny was used to investigate the temporal evolution of the *X. vasicola* species. Sufficient temporal signal (i.e., detectable amount of evolutionary change) was found only at the scale of the clade containing all contemporary U.S. and Argentinian *X. vasicola* pv. *vasculorum* strains (Fig. 1B). Multiple factors can contribute to the lack of significant temporal signal in the other clades, including overall genetic structure, a limited set of samples, and the short-span between the collection dates and the most recent common ancestor in each clade (Murray et al. 2016; Vanhove et al. 2019). Core nonrecombining SNPs were then identified in the clade containing the *X. vasicola* pv. *vasculorum* strains from the United States and Argentina (34 genomes, 243 SNPs). Performing tip-dating inference using BEAST (Drummond and Rambaut 2007) at this evolutionary scale allowed inference of a substitution rate of 3.17×10^{-7} per site per year (95% confidence interval [CI], 2.22×10^{-7} to 4.22×10^{-7}), as well as the age of every internal node in the tree (Fig. 1C). Given a likely South African origin for these strains, the resulting dated tree reveals two possible introduction events of corn *X. vasicola* pv. *vasculorum* strains to North and South America, one around 1997 to 1998, from which strains Arg-1A, Arg-4-A, Arg-7-A, and NE-7 originate, and one event around 2001 to 2010 involving all other Argentinian and U.S. *X. vasicola* pv. *vasculorum* strains (Fig. 1D). For the clade corresponding to the second event, Argentinian strains Arg-2B, Arg-6B, and Arg3-B are found at the root of the clade suggesting the introduction took place first in South America, at least for this event.

Overall, U.S. and Argentinian *X. vasicola* pv. *vasculorum* strains were not separated into two different clades in the dated tree. A possible explanation is that after the introduction of *X. vasicola* pv. *vasculorum* to North and South America, the population went through a bottleneck resulting in reduced genetic diversity, additionally the *X. vasicola* pv. *vasculorum* populations in these two countries may have not been geographically isolated long enough to diverge, or continuous genetic exchanges between them may have occurred. More detailed population genetics analyses, using extensive and randomized sampling, are needed to understand the population dynamics of *X. vasicola* pv. *vasculorum.*

***X. vasicola* pv. *vasculorum* genomes from the United States and Argentina contain clusters of genes often absent in other *X. vasicola* genomes.** We annotated all coding regions and found groups of orthologous genes across all analyzed genomes. Overall, 3,616 genes were present in all *X. vasicola* and 19 genes were found exclusively in *X. vasicola* pv. *vasculorum* from both the United States and Argentina (Supplementary Fig. S4).

To further uncover genes associated with the U.S. and Argentina *X. vasicola* pv. *vasculorum* population, we calculated statistical enrichment of genes in the U.S. and Argentina strains compared with all other *X. vasicola* genomes. For this, we performed a hypergeometric test for each ortholog group. In this test, we compared the presence and copy number of each gene in the *X. vasicola* pv. *vasculorum* genomes from the United States and Argentina against sets of other randomly selected *X. vasicola* genomes. In total, the test identified 335 total enriched genes (Supplementary Table S2). We also used the pan-GWAS strategy Scoary (Brynildsrud et al.

2016), which uncovered 399 enriched genes when comparing U.S. and Argentina strains to other *X. vasicola* genomes (Supplementary Table S2). The 330 genes common to both strategies were kept for further analyses, and out of these, 281 were overrepresented in *X. vasicola* pv. *vasculorum* from the United States and Argentina, while 49 were underrepresented (Supplementary Table S2). The overrepresented genes were enriched in annotations related to nucleic acid binding activity and involvement in DNA metabolism, recombination, and transposition (Supplementary Fig. S5), while no annotations were enriched in the underrepresented genes.

In the reference genome *X. vasicola* pv. *vasculorum* CO-5, multiple enriched genes clustered together (Supplementary Fig. S6). Five clusters (named A to E) contained 138 of the overrepresented genes, with two large clusters containing 41 (Cluster C) and 57 (Cluster E) genes, respectively (Fig. 2).

Genes from Clusters A to D were found in all the U.S. and Argentina *X. vasicola* pv. *vasculorum* strains and other corn *X. vasicola* pv. *vasculorum*, but less frequent in the other pathovars (Fig. 2). We hypothesize that these clusters could contain genes associated with differentiation of the *X. vasicola* pv. *vasculorum* pathovar. Meanwhile, Cluster E is found in the *X. vasicola* pv. *vasculorum* from the United States and Argentina and mostly absent in *X. vasicola* pv. *vasculorum* strains from South Africa. We propose Cluster E may be closely linked to adaptation of this population. Most of the genes in Cluster E are also found in all *X. vasicola* pv. *holcicola* strains, and one contemporary *X. vasicola* pv. *holcicola* strain from the United States (*Xvh-L*) contains all five clusters, indicating possible genomic exchanges between *X. vasicola* pv. *holcicola* and *X. vasicola* pv. *vasculorum* (Fig. 2).

Clusters of genes in *X. vasicola* pv. *vasculorum* are horizontally transferred genomic islands. The gene clusters were defined based on their genomic position in the *X. vasicola* pv. *vasculorum* strain CO-5. We then compared their organization and genomic features in a set of other seven complete *X. vasicola* genomes. When present, these genes were also clustered in all the examined strains and were found in similar genomic locations (Figs. 3 and 4).

All of the identified clusters were predicted to be in genomic islands, using the IslandViewer4suite (Bertelli et al. 2017), this suggested that the clusters were acquired by horizontal transfer (Fig. 3). Clusters C and D contained multiple insertion sequences (Fig. 3), possibly related to their mobility, and Cluster C was particularly enriched in insertion sequences and transposition-associated genes (Fig. 3 and Supplementary Fig. S5).

To assess a possible association between these clusters and pathogenicity, we did a BLAST search (Altschul et al. 1997) of the gene sequences against a set of all described *Xanthomonas* T3 effectors (http://xanthomonas.org/t3e.html). None of the genes in the clusters, nor any of the over- or underrepresented genes, matched against known *Xanthomonas* T3 effectors. Furthermore, no specific association was found between effector presence/absence and U.S. *X. vasicola* pv. *vasculorum* or corn *X. vasicola* pv. *vasculorum*, in general, with the possible exception of XopG1, an M27 zinc protease (White et al. 2009), which is absent in most *X. vasicola* pv. *vasculorum* strains and present in other *X. vasicola* pathovars (Supplementary Fig. S7).

We then predicted ~450 putative novel T3-secreted proteins in each genome having, at minimum, a predicted N-terminal signal peptide as determined by EffectiveT3 from the suite EffectiveDB (Eichinger et al. 2016) (Fig. 3). These predictions represent approximately 10% of the annotated genes in each genomes and likely include many false positives. Some predicted T3-secreted proteins were found encoded by genes in the clusters, including nine genes in Cluster E, annotated as producing various hypothetical proteins, an HTH (helix-turn-helix) transcriptional regulator, and two methyltransferases (Fig. 4, Supplementary Table S3). These hypothetical T3-secreted proteins may be the result of over-prediction but may also represent potential candidates for virulence determinants for these strains and should be tested experimentally.

Since these clusters are likely to have been horizontally transferred, we attempted to find the taxonomic origin of the transfer by using the meta-genomics annotation strategy Kaiju (Menzel et al. 2016) to find the closest match for each gene from the CO-5 strains in the progenomes database (Mende et al. 2017), a database of representative microbial genomes that does not include *X. vasicola* pv. *vasculorum* or *X. vasicola* pv. *holcicola*. As a whole, over 82.7% of the CO-5 genes were effectively assigned to *Xanthomonas* sp., with most genes assigned to *X. vasicola* pv. *musacearum* (*X. campestris* in the progenomes database) (Supplementary Fig. S8). In contrast, the gene clusters contained sequences from different taxonomic groups. In Cluster A, 41.7% of the genes were assigned to *Pantoea ananatis* (Supplementary Fig. S8), which is frequently isolated along with *X. vasicola* pv. *vasculorum* from lesions on corn (Lang et al. 2017; Coutinho et al. 2015). All other clusters contained a large percentage of genes assigned to *Xanthomonas* (38.9 to 63.6%), but other taxa were also found. In Cluster C, for instance, 40% of the genes were assigned to taxa other than *Xanthomonas* including *Sphingobium*, *Achromobacter*, *Methylomonas*, and *Cupriavidus* (Supplementary Fig. S8).

Curiously, Clusters B and E contained genes that matched phages in the families *Inoviridiae* (16.7%) and *Caudoviridae* (7%), respectively (Supplementary Fig. S8). Both clusters were not associated to insertion sequences (Fig. 3) so their acquisition could have been mediated by phage transmission. And Cluster E was also enriched in GO terms associated to viral life cycles (Supplementary Fig. S5).

***X. vasicola* pv. *vasculorum* from the United States and Argentina contain a prophage region shared by *X. vasicola* pv. *holcicola* strains.** The finding of phage-related sequences prompted us to scan the genomes for additional phage sequences using PHASTER (Arndt et al. 2016). Cluster B was identified as an incomplete or questionable prophage. And Cluster E corresponded to an intact prophage of ~45 kb that encoded multiple phage-related proteins (Figs. 4 and 5A). Many of the proteins encoded in the Cluster E prophage had similarities to proteins from the *Xanthomonas*-infecting phages Cp1, Cp2, Xp15, as well as phages infecting other bacteria (Fig. 5A). In all complete genomes analyzed, when present, the prophage is inserted in the same genomic location: on the 5′ end of the highly conserved tRNA-dihydrouridine synthase A (*dusA*) gene (Farrugia et al. 2015) (Fig. 4).

The Cluster E prophage genes were found in all U.S. and Argentinian corn *X. vasicola* pv. *vasculorum*, one contemporary South African strain (Xvz45), and all *X. vasicola* pv. *holcicola* isolates, while being absent in *X. vasicola* pv. *musacearum*, *X. vasicola*, and most South African *X. vasicola* pv. *vasculorum* (Fig. 2). The Cluster E prophage is also absent from a recently reclassified genome of a strain isolated from Areca nut that is closely related to *X. vasicola* pv. *vasculorum*, which was not included in our analyses (Bradbury 1986; Studholme et al. 2020).

To understand the evolutionary history of the Cluster E prophage, we identified SNPs in this region across all the genomes that contained it (878 SNPs) and created a phylogeny, which enabled identification of four major types (Fig. 5B). Type 4, found only in the *X. vasicola* pv. *holcicola* strain NCPPB 1060 (isolated in 1961 in Ethiopia), was highly divergent from the other types. Types 1 to 3 were found in strains collected in South Africa, Oceania, and North and South America from 1942 to 2017, and contained only 216 SNPs between them, suggesting conservation of this region (Fig. 5B). All four types were found in *X. vasicola* pv. *holcicola* strains, with no clear relation between prophage types and the location or year of isolation of the strain. Corn *X. vasicola* pv. *vasculorum* strains contained either type 2, found in strains Xvz45, Arg-7A, and NE-7, or type 1, found in all other *X. vasicola* pv. *vasculorum* strains in the United States and Argentina (Fig. 5B and C).

Multiple scenarios could explain the distribution of the prophage and its types in *X. vasicola* pv. *vasculorum*: (i) the prophage existed in the common ancestor of all ancestral *X. vasicola* strains and then

was lost in multiple clades while remaining in *X. vasicola* pv. *holcicola* and certain *X. vasicola* pv. *vasculorum*. This scenario would require multiple losses across the phylogeny (at least nine). These multiple losses can occur if for example transitions between lysogenic and lytic cycles were common for this prophage and/or if this region conferred a disadvantage that was selected against in some clades (Ramisetty and Sudhakari 2019) (Fig. 5C). (ii) The prophage was acquired in a common ancestor for all *X. vasicola* pv. *holcicola* strains, and more recently, types 1 and 2 were horizontally transferred to ancestors of U.S. and Argentina *X. vasicola* pv.

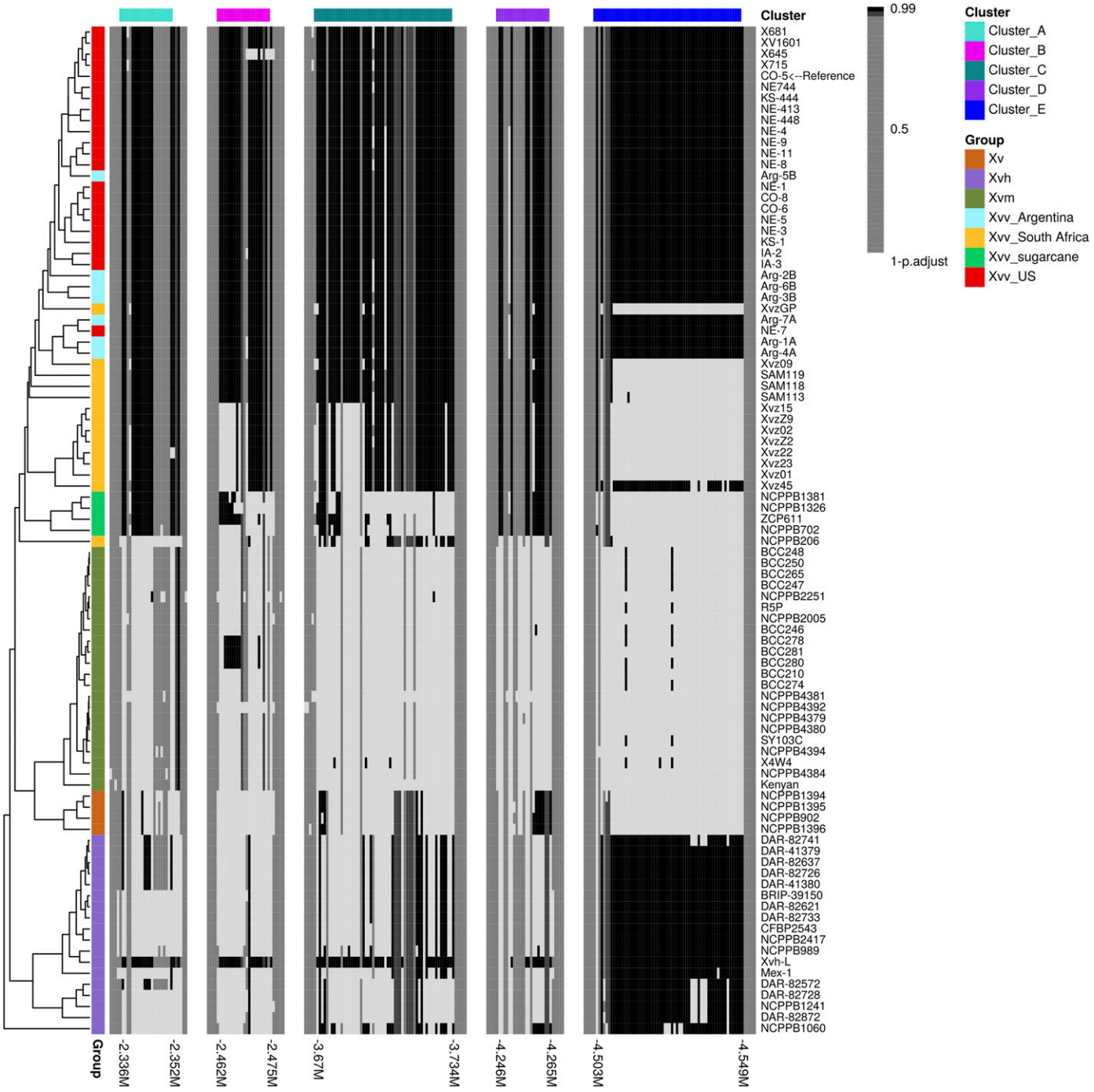

Fig. 2. Clusters of overrepresented genes in *Xanthomonas vasicola* pv. *vasculorum* strains from the United States and Argentina. Enriched genes were identified in *X. vasicola* pv. *vasculorum* strains from the United States and Argentina when compared with other *X. vasicola* strains using both a hypergeometric test and the pan-genome-wide association studies (pan-GWAS) strategy Scoary. A total of 138 of these genes grouped into five clusters according to their genomic location in the genome of the U.S. strain *X. vasicola* pv. *vasculorum* CO-5. Clusters were defined as a group of at least 10 enriched genes, with *P* values for both tests of <0.01, found at a maximum distance of 5 kb from each other. The heatmaps show the distribution of orthologs of genes of the five CO-5 clusters across the *X. vasicola* phylogeny, each cluster indicated by shaded bars on top of the heatmap. Light gray indicates an ortholog is absent from the respective genomes. While the dark gray to black scale indicates genes that are present in the genomes and their adjusted *P* value indicating their association to the *X. vasicola* pv. *vasculorum* U.S. and Argentina population, as determined by the hypergeometric test. Columns in the heatmap are ordered according to the genomic position of each gene in the *X. vasicola* pv. *vasculorum* CO-5 genome, with positions indicated in million base pairs (M) below the heatmaps. Genes may not be in the same order in other genomes. Five genes flanking each cluster on both sides in the *X. vasicola* pv. *vasculorum* CO-5 genomes are also shown for comparison. Shaded bars at the left of the heatmap indicate the group (host or country of isolation for *X. vasicola* pv. *vasculorum*, pathovar for others) of the strain. The cladogram to the left corresponds to core single nucleotide polymorphism phylogenies in Figure 1D and Supplementary Fig. S1.

vasculorum and the South African strain *Xvz45*. This scenario would require at least four horizontal gene transfer events (Fig. 5C). Horizontal gene transfer of the prophage from the *X. vasicola* pv. *vasculorum* clade to all *X. vasicola* pv. *holcicola* is also possible but less likely in this scenario given the temporal distribution of the strains (Fig. 5C). (iii) The prophage region was independently acquired in *X. vasicola* pv. *vasculorum* and *X. vasicola* pv. *holcicola* from viruses infecting these pathovars, with possibly four acquisition events occurring in *X. vasicola* pv. *vasculorum* to account for the presence of prophage types 1 and 2. There is however not enough information as to the viral origin of this region or its subsequent distribution (Fig. 5C).

We believe the complex distribution of the prophage types, mainly the fact that similar types are found in phylogenetically distant strains, favors scenario 2 (horizontal gene transfer) as the simpler explanation. Horizontal gene transfer events can occur between these pathovars as shown by the presence of multiple *X. vasicola* pv. *vasculorum* regions in the *X. vasicola* pv. *holcicola* strain *Xvh-L* (Fig. 2) and considering that *X. vasicola* pv. *holcicola* has been occasionally isolated from corn (Moffett 1983; Péros et al. 1994). We cannot however dismiss that the other scenarios presented, or a combination of them, may have occurred. The neighboring region of the prophage is highly conserved and prevents us from further discriminating between the possible scenarios.

Overall, our results identified genomic regions associated with *X. vasicola* pv. *vasculorum* infecting corn, and a prophage region enriched in the emerging *X. vasicola* pv. *vasculorum* population in North and South America that might have been transferred from *X. vasicola* pv. *holcicola*. Whether the acquisition of these regions played any role in the recent spread of *X. vasicola* pv. *vasculorum* in the United States and Argentina is intriguing and will be the explored in future work.

DISCUSSION

We showed that the emerging populations of *X. vasicola* pv. *vasculorum* infecting corn in Argentina and the United States are genetically related, and we highlighted genomic regions enriched in these populations. In phylogenetic analyses, there was no distinct separation between Argentinian and U.S. strains, suggesting *X. vasicola* pv. *vasculorum* strains from these countries may be the result of a genetic bottleneck, have not significantly diverged since their introduction, and/or genetic exchanges have continuously existed between them. Tip-dating analyses revealed two

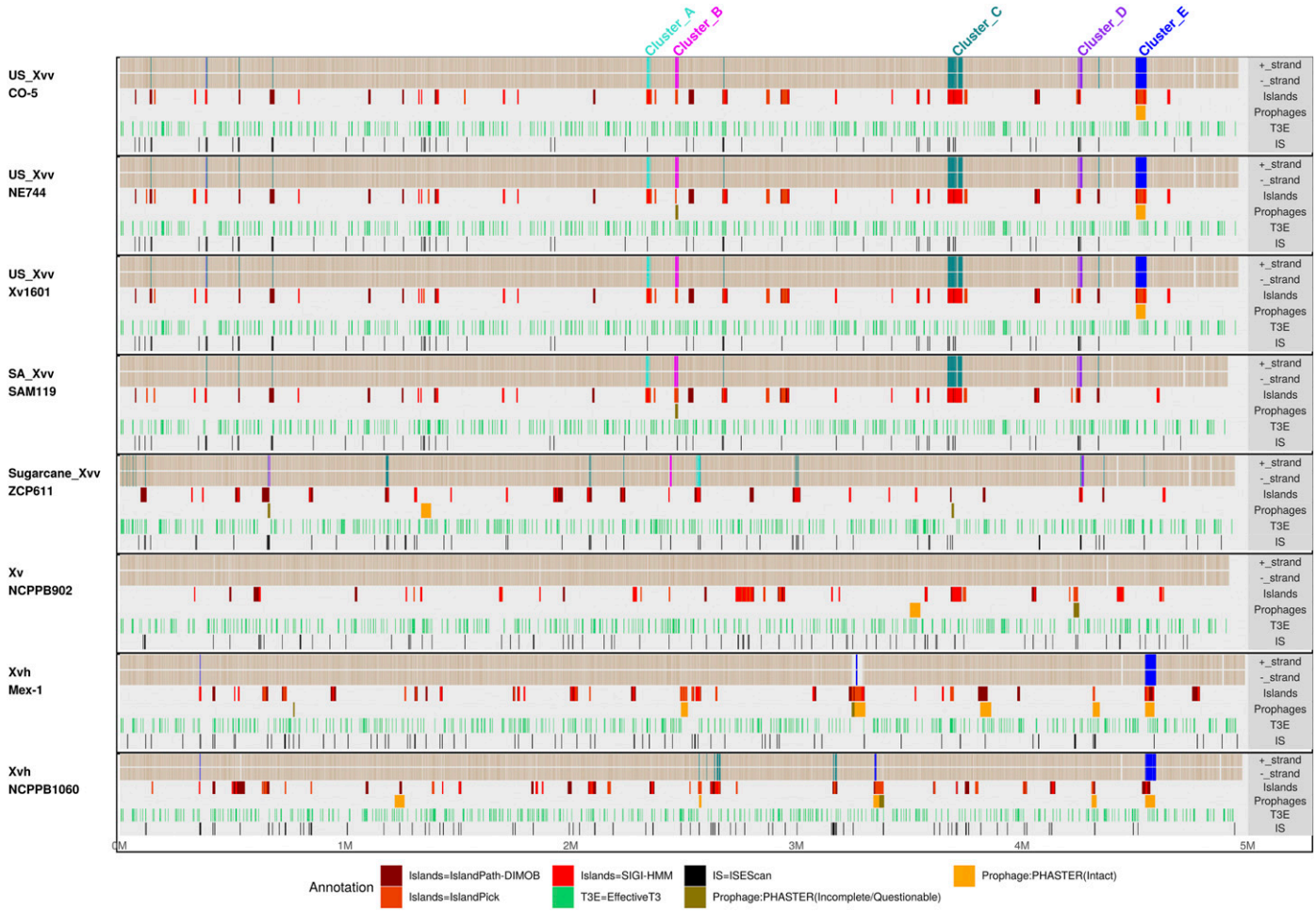

Fig. 3. Genomic features associated with *Xanthomonas vasicola* pv. *vasculorum* clusters in complete *X. vasicola* genomes. Linear representation of eight complete *X. vasicola* genomes with annotated regions of interest. Genomic scale in million base pairs (M) is shown at the bottom of the figure. For each genome, horizontal bars represent the following from top to bottom. (i) Presence of annotated genes in each genome strand (+ and −) are shown in beige or light gray. Clusters of genes identified as overrepresented in *X. vasicola* pv. *vasculorum* from the United States and Argentina, and their orthologs, are shown in colors or shades of gray. (ii) Predicted Genomic Islands using three methods integrated in Island Viewer are shown in red/gray: SIGI-HMM is based on sequence composition, IslandPath-DIMOB is based on dinucleotide bias and presence of mobility genes, and IslandPick is based on phylogenetic comparisons (Bertelli et al. 2017), and multiple predictions for same region are shown overlapped. (iii) Prophage regions as identified by PHASTER (Arndt et al. 2016) and classified as intact or incomplete/questionable. (iv) Genes predicted to encode type 3 secreted proteins, having a significant prediction score with EffectiveT3 from the EffectiveDB suite (Eichinger et al. 2016) are shown in green/light gray. (v) Insertion sequence (IS) elements as identified using ISEScan (Xie and Tang 2017) are shown in black.

possible introduction events of corn *X. vasicola* pv. *vasculorum* strains to North and South America, one around 1997 to 1998 and one around 2001 to 2010, both events years before the detection of the epidemic (Korus et al. 2017; Plazas et al. 2017). The clade corresponding to the 2001 to 2010 introduction contains most of the sampled strains and has Argentinian strains at its root, suggesting this introduction occurred first in South America. It is harder to determine whether this is also the case for the 1997 to 1998 introduction since only one of the sampled U.S. strains falls within this clade.

The rapid spread of the disease in the United States and the possible ongoing genetic exchange between *X. vasicola* pv. *vasculorum* populations in these distant countries may have been facilitated by human activity. Corn breeders in the United States accelerate product development by maintaining year-long operations and have increasingly adopted the practice of using winter nurseries for breeding and seed production (Brewbaker 2003; Butruille et al. 2015); many of these winter nurseries are located in the southern hemisphere, including South America (Butruille et al. 2015; Zaworski 2016). Additionally, corn production and export in South America has experienced considerable growth in the last decades (Meade et al. 2016). Although it is still unknown whether *X. vasicola* pv. *vasculorum* is transmitted by seeds, current practices likely allow enough exchange of contaminated material such that an adapted population can spread quickly between continents.

A common pathway that could lead to phytopathogenic bacteria epidemics is the introduction of exotic genotypes into environments that are favorable for the spread of the bacteria in the plant community, due to the availability of, for example, plant species susceptible to the infection (Almeida and Nunney 2015). It is possible that this is the case for the recent *X. vasicola* pv. *vasculorum* spread and that differences in the cultivated varieties or agricultural practices in the United States and Argentina compared with South Africa are enough to explain the epidemic. However, outbreaks of plant pathogens have also often been associated with the acquisition, particularly through recombination or horizontal gene transfer, of pathogenicity determinants that can confer an advantage to an emerging population (Monteil et al. 2016; Vinatzer et al. 2014). In *Xanthomonas* species, aggressiveness or virulence of emerging populations has been shown to be related to such genetic changes (Saux et al. 2015; Vera Cruz et al. 2000).

It is unknown whether the fast-spreading *X. vasicola* pv. *vasculorum* from the United States and Argentina have any phenotypical differences to South African strains in terms of plant colonization. Initial experiments showed no difference in symptoms in corn (hybrid DKC 61-88) between a contemporary U.S. strain (NE-744), a South African strain (SAM119), and four *X. vasicola* pv. *holcicola* strains (Lang et al. 2017). These tests were limited to a few strains in only one genetic background of maize. Further experiments are needed comparing multiple strains and using an expanded panel of corn varieties, including varieties grown

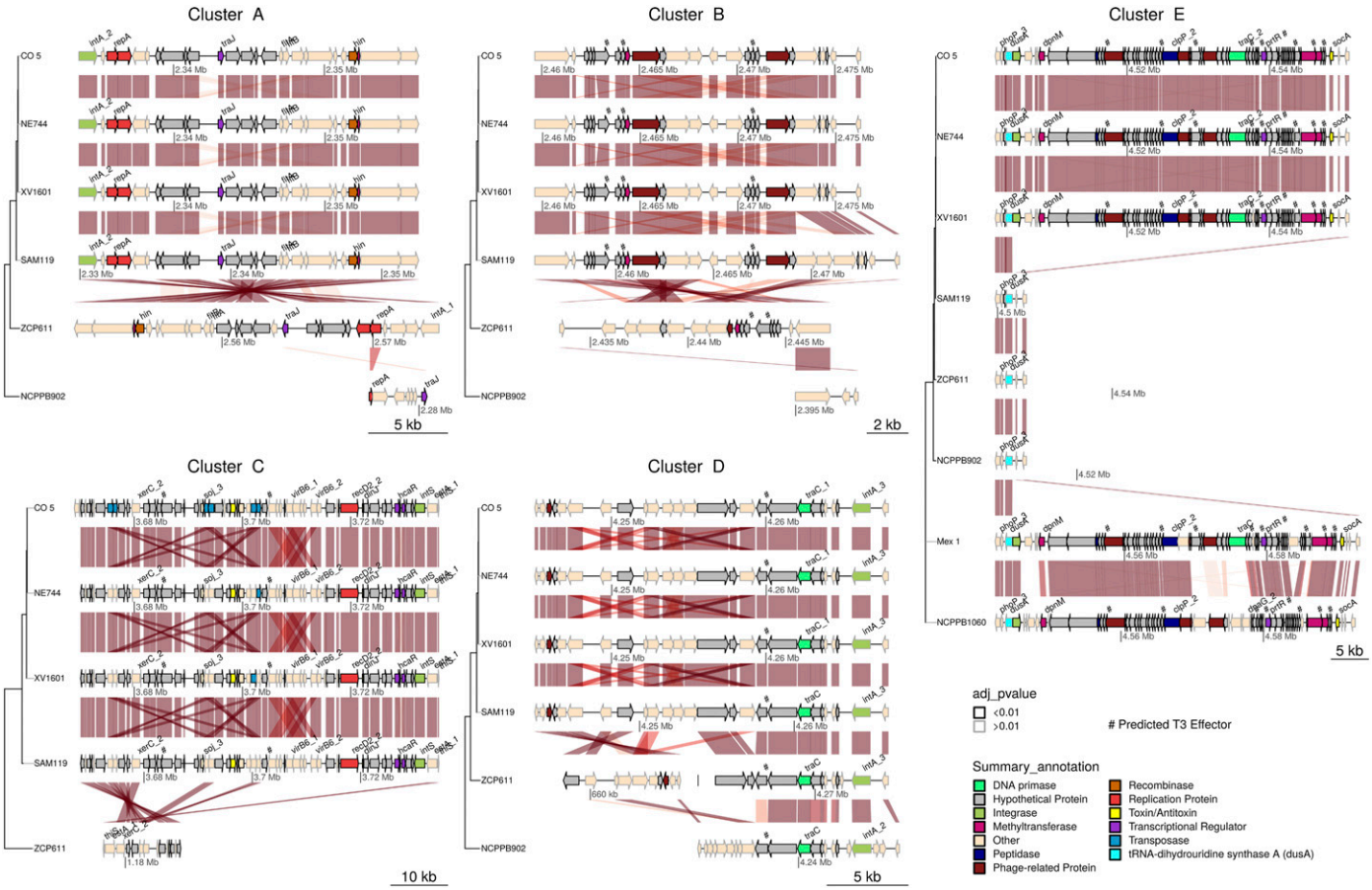

Fig. 4. Gene maps of *Xanthomonas vasicola* pv. *vasculorum* clusters in complete *X. vasicola* genomes. For each cluster, the defined overrepresented regions in the reference genome *X. vasicola* pv. *vasculorum* CO-5 are shown in addition to two flanking genes, along with the corresponding orthologous regions in other complete genomes. Genes are drawn as arrows according to their orientation (right = plus strand, left = minus strand). Distance of the genomic region to the replication origin is indicated in million base pairs (Mb). Arrows are colored according to broad categories based on Blast2GO annotation (Supplementary Table S3). The color of arrow borders indicates whether each gene was determined as overrepresented (adjusted *P* value of <0.01) in *X. vasicola* pv. *vasculorum* from the United States and Argentina (black = yes, gray = no). Red crisscrossed areas connecting genes between genomes indicate similarity from low (light) to high (dark) as determined by pairwise protein BLAST. Gene names assigned by Prokka (Seemann 2014) are shown for each region. Trees to the left of each cluster correspond to the single nucleotide polymorphisms-based phylogeny in Figure 1, keeping only the nodes shown. Predicted T3 Effector genes, containing at least a predicted T3 secretion signal as determined by EffectiveT3 (Eichinger et al. 2016), are indicated with # symbol.

in the three continents where *X. vasicola* pv. *vasculorum* is present. Such experiments are hindered, however, because no contemporary collection of South African strains of *X. vasicola* pv. *vasculorum* is available. If found, phenotypical differences between these *X. vasicola* pv. *vasculorum* populations could suggest a process of niche adaptation upon their introduction to North and South America, such adaptation can be mediated by adaptive horizontal gene transfers as the ones hypothesized here (Wiedenbeck and Cohan 2011). Future work should also consider how other, if any, population and evolutionary processes (such as mutation and recombination rates) were altered in *X. vasicola* pv. *vasculorum* during niche adaptation (Sheppard et al. 2018).

We identified five clusters of genes that were overrepresented in *X. vasicola* pv. *vasculorum* strains from the United States and Argentina. These gene clusters overlapped with predicted genomic island regions, consistent with acquisition through horizontal gene transfer. Several clusters may represent important genomic acquisitions, if not for the current emerging population, for corn *X. vasicola* pv. *vasculorum* strains overall. Cluster C, for instance, consists of ~44 genes found in all contemporary corn *X. vasicola* pv. *vasculorum* strains and some *X. vasicola* pv. *holcicola* strains. This cluster is enriched in sequences associated with mobility, including insertion sequences, genes coding transposases, and various DNA binding proteins, and it contains five genes coding predicted T3 secreted proteins (Fig. 4). Taxonomic analyses revealed that a large percentage of genes in this cluster match other taxa including *Pseudomonas*, S*phingobium*, and various *Burkholderiales.*

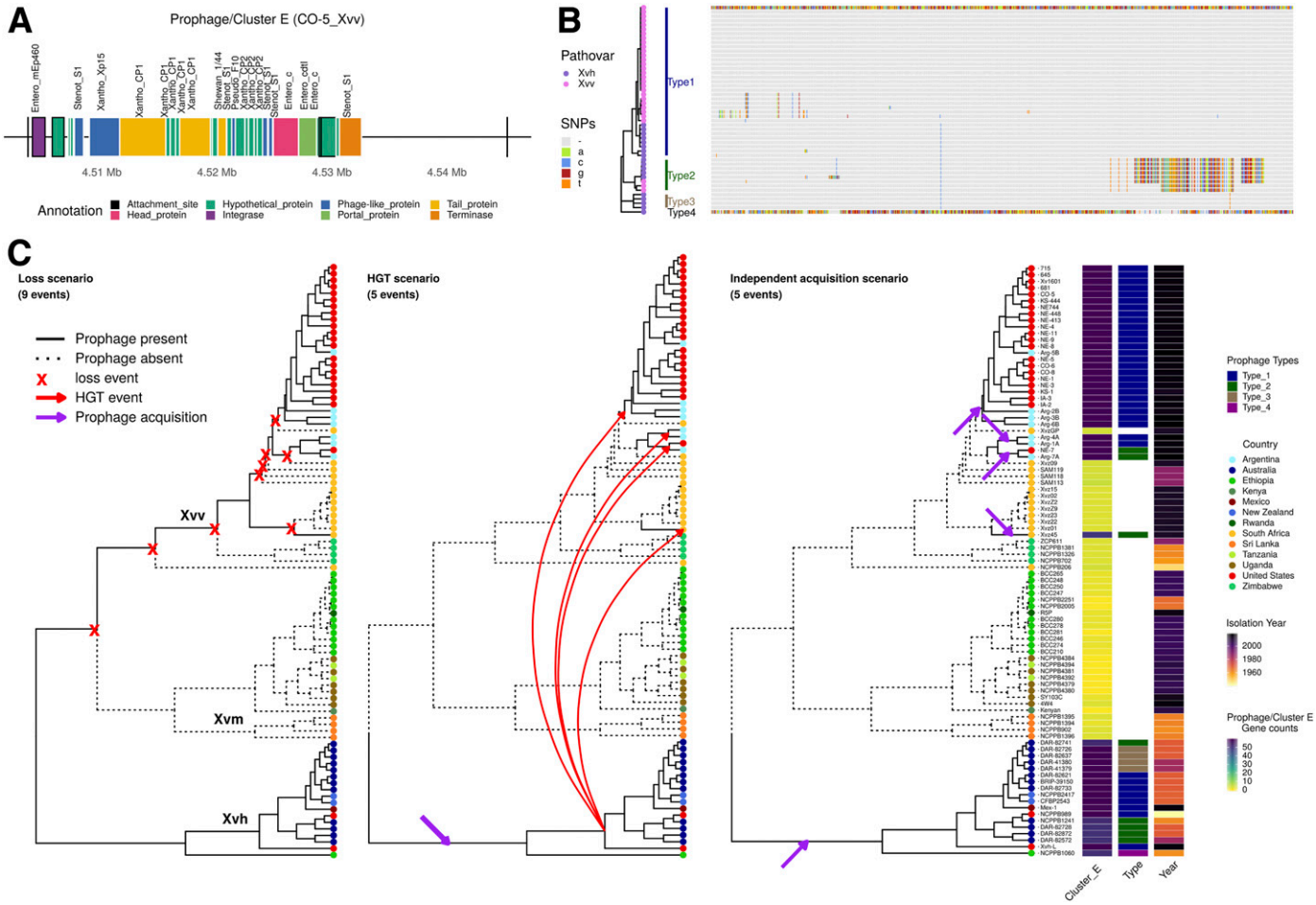

Fig. 5. Possible evolutionary histories of the Cluster E prophage. **A,** Diagram showing the genes in the predicted prophage corresponding to Cluster E in the genome of *Xanthomonas vasicola* pv. *vasculorum* CO-5. Genes in the plus strand are shown with a black border and genes in the minus strand with a white border. For each gene, the top hit against known genes in the Virus and Prophage database of PHASTER (Arndt et al. 2016) is shown. Genes with no annotation had no significant hits. Colors for each gene indicate functional annotation as produced by PHASTER (Arndt et al. 2016). **B,** Single nucleotide polymorphisms (SNPs) were identified in the prophage/Cluster E region using the *X. vasicola* pv. *vasculorum* CO-5 genome as reference, excluding indels. The tree to the left is a cladogram based on a maximum-likelihood tree produced by CSI phylogeny using these SNPs (Kaas et al. 2014). Tip colors indicate the pathovar of each genome used. The alignment to the right shows these SNPs (878 total) using the *X. vasicola* pv. *vasculorum* CO-5 sequence as reference at the top; light gray colors indicate for each genome that the corresponding nucleotide is identical to that in *X. vasicola* pv. *vasculorum* CO-5, while variants are shown in colors. Five prophage types were defined based on the tree structure and are indicated to the right of the tree. **C,** Three possible scenarios explaining the prophage distribution in *X. vasicola* pv. *vasculorum* strains if events of only one type were predominant: prophage loss, horizontal gene transfer (HGT), or independent acquisition. Mutation is not shown and for simplicity, events leading to the distribution of the prophage types in the *X. vasicola* pv. *holcicola* clade are not considered. The trees are cladograms based on the core SNP phylogenies in Figure 1. Tip points represent the geographic origin of each strain. Heatmaps to the right show for each genome, respectively, the year of isolation, the number of genes in the genome that are orthologs of Cluster E genes, and the prophage type found in the genome as defined in C. From left to right, the trees represent the following: (i) A loss scenario where the prophage is found in the common ancestor to all *X. vasicola* strains and then is lost in multiple clades. In this scenario multiple prophage types are retained in *X. vasicola* pv. *holcicola* while only two are kept in *X. vasicola* pv. *vasculorum.* (ii) A scenario where the prophage is acquired in an ancestor of the *X. vasicola* pv. *holcicola* strains, it undergoes diversification, and then types 1 and 2 are transferred horizontally to *X. vasicola* pv. *vasculorum*. The arrow origin for the transfers is placed at the smallest clade that includes all *X. vasicola* pv. *holcicola* strains with types 1 or 2. (iii) A scenario where the prophage independently integrated upon viral infection in the genomes of ancestors of the different clades that contain it; different integration events are shown for the types 1 and 2 in *X. vasicola* pv. *vasculorum*.

Cluster A contained eight genes found in *P. ananatis*, including genes involved in replication (replication proteins A and C) and conjugation (P-type conjugative transfer proteins). *P. ananatis* is the causal agent of brown stalk rot of corn (Goszczynska et al. 2007), but it is also a versatile organism able to infect monocotyledonous and dicotyledonous hosts, and a common epiphyte and endophyte (Coutinho and Venter 2009). In corn, *P. ananatis* has also been reported to cause necrotic spots and streaks, and it has been shown to be the dominant bacteria with ice nucleation activity in corn leaves while causing frost injuries in several other plant species (Coutinho and Venter 2009). *P. ananatis* was documented in association with *X. vasicola* pv. *vasculorum* on *Eucalyptus* in South Africa (Coutinho et al. 2015) and with *X. vasicola* pv. *vasculorum* BLS symptomatic corn in the United States (Lang et al. 2017). However, *Pantoea* strains that were isolated along with *X. vasicola* pv. *vasculorum*, when inoculated alone, were unable to cause disease in corn (Lang et al. 2017), and brown stalk rot symptoms have not been reported on plants infected with *X. vasicola* pv. *vasculorum*. The relationship between *X. vasicola* pv. *vasculorum* and *P. ananatis* is likely complex, and it is intriguing what role, if any, the genes *X. vasicola* pv. *vasculorum* acquired from its partner play in their interaction, or in the interaction of these bacteria with their host.

We focused on Cluster E because, among *X. vasicola* pv. *vasculorum*, this cluster is enriched in the Argentina and U.S. *X. vasicola* pv. *vasculorum* populations, and its acquisition may be related to their emergence. However, it is also possible that the prophage exists in the South African *X. vasicola* pv. *vasculorum* population in a higher proportion than reflected in the publicly available genomes, and that the association of this region with the North and South American population is incidental. As in, the U.S. and Argentina populations are the result of a bottleneck effect where they are descendants from yet unsampled South African strains containing the prophage (both type 1 and 2). Further sampling and sequencing of contemporary strains is needed to address this possibility.

Prophages are temperate (noninfective or nonlytic) viruses that are integrated into bacterial genomes and can play an important role in bacterial genome evolution (Varani et al. 2013). The prophage region in Cluster E was shared by *X. vasicola* pv. *holcicola* and U.S. and Argentinian *X. vasicola* pv. *vasculorum*, and contained genes resembling elements of *Xanthomonas*-infecting bacteriophages CP1, CP2, and Xp15, although there is no complete homology to any identified bacteriophages. From its insertion site in *X. vasicola* pv. *holcicola* and *X. vasicola* pv. *vasculorum* genomes, the Cluster E prophage seems to be part of a wide class of genomic islands characterized by their insertion in the 5′ end of the *dusA* gene, mediated by specific tyrosine recombinases known as *dusA*-associated integrases found in alpha, beta, and gammaproteobacteria (Farrugia et al. 2015), a homolog of this type of integrase is found and highly conserved in the Cluster E prophage (Fig. 4). These genomic islands seem to be highly mobile and capable of excision as a circularized intermediate (Farrugia et al. 2015). This could explain both frequent losses and acquisitions through the *X. vasicola* phylogeny.

Prophages are important vehicles for horizontal gene transfer, they can promote recombination and rearrangements in the bacterial genome, and they often carry additional nonessential cargo genes (morons) that may confer new phenotypic properties to the bacteria (Brüssow et al. 2004; Varani et al. 2013). Prophages can carry virulence factors or factors that enhance bacterial fitness (Brüssow et al. 2004; Figueroa-Bossi et al. 2001), although reduction of virulence has also been reported (Ahmad et al. 2014). Prophages harboring elements conferring virulence activity were found in different plant pathogenic bacteria including *Xylella* sp. (Varani et al. 2008) and '*Candidatus* Liberibacter asiaticus' (Jain et al. 2015). In *X. arboricola*, strains pathogenic on walnut carry a higher number (and also a different repertoire) of prophages than nonpathogenic strains (Cesbron et al. 2015).

The Cluster E prophage contains various genes coding nonphage related hypothetical proteins with unknown function, as well as proteins predicted to be T3-secreted, peptidases, and transcription factors. Many of these genes were highly conserved between geographically and temporarily distant *X. vasicola* pv. *holcicola* and *X. vasicola* pv. *vasculorum* strains, suggesting this region is not subject to prophage decay and may confer some advantage to the bacteria (Brüssow et al. 2004). Further characterization of these genes, as well as of other overrepresented *X. vasicola* pv. *vasculorum* genes that were not assigned to clusters, is needed to establish a possible role in virulence.

We used comparative genomics to generate hypotheses about the origin of an epidemic and the genetic mechanisms associated with pathogen population spread. Based on our findings, we identify candidate genetic factors that will be the subject of future work to understand the lifestyle and evolution of *X. vasicola* pv. *vasculorum* and related bacteria.

ACKNOWLEDGMENTS

We thank collaborators Tamra Jackson-Ziems, Terra Hartman, Silvina Areas, and Garry Munkvold for providing isolates of *X. vasicola* pv. *vasculorum* from Nebraska and Iowa, and Emile Gluck-Thaler for guidance and discussions on evolutionary analyses.

LITERATURE CITED

Ahmad, A. A., Askora, A., Kawasaki, T., Fujie, M., and Yamada, T. 2014. The filamentous phage XacF1 causes loss of virulence in *Xanthomonas axonopodis* pv. *citri*, the causative agent of citrus canker disease. Front. Microbiol. 5:321.

Almeida, R. P. P., and Nunney, L. 2015. How do plant diseases caused by *Xylella fastidiosa* emerge? Plant Dis. 99:1457-1467.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D. S. 2016. PHASTER: A better, faster version of the PHAST phage search tool. Nucleic Acids Res. 45:W16-W21.

Arredondo-Alonso, S., Rogers, M. R. C., Braat, J. C., Verschuuren, T. D., Top, J., Corander, J., Willems, R. J. L., and Schürch, A. C. 2018. mlplasmids: A user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. Microbiol. Genom. 4:e000224.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prijibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455-477.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 57:289-300.

Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winsor, G. L., and Brinkman, F. S. L. 2017. IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 45:W30-W35.

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Bradbury, J. F. 1986. Guide to Plant Pathogenic Bacteria. CAB International, Wallingford, UK.

Brewbaker, J. L. 2003. Corn Production in the Tropics: The Hawaii Experience. University of Hawaii, Honolulu, HI.

Broders, K. 2017. Status of bacterial leaf streak of corn in the United States. Proceedings of the Integrated Crop Management Conference, Iowa State University, Digital Press.

Brüssow, H., Canchaya, C., and Hardt, W.-D. 2004. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. 68:560-602.

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 17:238.

Butruille, D. V., Birru, F. H., Boerboom, M. L., Cargill, E. J., Davis, D. A., Dhungana, P., et al. 2015. Maize breeding in the United States: Views from within Monsanto. Pages 199-282 in: Plant Breeding Reviews, Vol. 39. John Wiley & Sons, Ltd., New York.

Cesbron, S., Briand, M., Essakhi, S., Gironde, S., Boureau, T., Manceau, C., Fischer-Le Saux, M., and Jacques, M.-A. 2015. Comparative genomics of

pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. Front. Plant Sci. 6: 1126.

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.

Coil, D., Jospin, G., and Darling, A. E. 2015. A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31:587-589.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Conway, J. R., Lex, A., and Gehlenborg, N. 2017. UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics 33: 2938-2940.

Coutinho, T. A., van der Westhuizen, L., Roux, J., McFarlane, S. A., and Venter, S. N. 2015. Significant host jump of *Xanthomonas vasicola* from sugarcane to a *Eucalyptus grandis* clone in South Africa. Plant Pathol. 64: 576-581.

Coutinho, T. A., and Venter, S. N. 2009. *Pantoea ananatis*: An unconventional plant pathogen. Mol. Plant Pathol. 10:325-335.

Coutinho, T. A., and Wallis, F. M. 1991. Bacterial streak disease of maize (*Zea mays* L.) in South Africa. J. Phytopathol. 133:112.

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394-1403.

Didelot, X., and Wilson, D. J. 2015. Clonalframeml: Efficient inference of recombination in whole bacterial genomes. PLOS Comput. Biol. 11:e1004041.

Ding, W., Baumdicker, F., and Neher, R. A. 2018. panX: Pan-genome analysis and exploration. Nucleic Acids Res. 46:e5.

Drummond, A. J., and Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Dyer, R. A. 1949. Botanical surveys and control of plant diseases. Farming S. Afr. 24:119-121.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1797.

Eichinger, V., Nussbaumer, T., Platzer, A., Jehl, M.-A., Arnold, R., and Rattei, T. 2016. EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. Nucleic Acids Res. 44:D669-D674.

Emms, D. M., and Kelly, S. 2015. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.

Fargier, E., Saux, M. F.-L., 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.

Farrugia, D. N., Elbourne, L. D. H., Mabbutt, B. C., and Paulsen, I. T. 2015. A novel family of integrases associated with prophages and genomic islands integrated within the tRNA-dihydrouridine synthase A (*dusA*) gene. Nucleic Acids Res. 43:4547-4557.

Figueroa-Bossi, N., Uzzau, S., Maloriol, D., and Bossi, L. 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. Mol. Microbiol. 39:260-272.

Gardner, S. N., Slezak, T., and Hall, B. G. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics 31:2877-2878.

Goszczynska, T., Botha, W. J., Venter, S. N., and Coutinho, T. A. 2007. Isolation and identification of the causal agent of brown stalk rot, a new disease of maize in South Africa. Plant Dis. 91:711-718.

Grote, S. 2018. GOfuncR: Gene ontology enrichment using FUNC. R package version 1. https://bioconductor.org/packages/release/bioc/vignettes/GOfuncR/inst/doc/GOfuncR.html

Guy, L., Roat Kultima, J., and Andersson, S. G. E. 2010. genoPlotR: Comparative gene and genome visualization in R. Bioinformatics 26:2334-2335.

Hartman, T., Tharnish, B., Harbour, J., Yuen, G. Y., and Jackson-Ziems, T. A. 2020. Alternative hosts in the families Poaceae and Cyperaceae for *Xanthomonas vasicola* pv. *vasculorum*, causal agent of bacterial leaf streak of corn. Phytopathology. 110:1147-1152.

Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., and Harris, S. R. 2015. Circlator: Automated circularization of genome assemblies using long sequencing reads. Genome Biol. 16:294.

Jain, M., Fleites, L. A., and Gabriel, D. W. 2015. Prophage-encoded peroxidase in 'C*andidatus* Liberibacter asiaticus' is a secreted effector that suppresses plant defenses. Mol. Plant-Microbe Interact. 28:1330-1337.

Jombart, T., Balloux, F., and Dray, S. 2010. adephylo: New tools for investigating the phylogenetic signal in biological traits. Bioinformatics 26: 1907-1909.

Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M., and Lund, O. 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One 9:e104984.

Kerepesi, C., Bánky, D., and Grolmusz, V. 2014. AmphoraNet: The webserver implementation of the AMPHORA2 metagenomic workflow suite. Gene 533:538-540.

Kolde, R., and Kolde, M. R. 2015. Package 'pheatmap'. R Package version 1. https://CRAN.R-project.org/package=pheatmap

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722-736.

Korus, K., Lang, J. M., Adesemoye, A. O., Block, C. C., Pal, N., Leach, J. E., and Jackson-Ziems, T. A. 2017. First report of *Xanthomonas vasicola* causing bacterial leaf streak on corn in the United States. Plant Dis. 101: 1030.

Lang, J. M., DuCharme, E., Ibarra Caballero, J., Luna, E., Hartman, T., Ortiz-Castro, M., Korus, K., Rascoe, J., Jackson-Ziems, T. A., Broders, K., and Leach, J. E. 2017. Detection and characterization of *Xanthomonas vasicola* pv. *vasculorum* (Cobb 1894) comb. nov. causing bacterial leaf streak of corn in the United States. Phytopathology 107:1312-1321.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., and Carey, V. J. 2013. Software for computing and annotating genomic ranges. PLOS Comput. Biol. 9:e1003118.

Leite, R. P., Custódio, A. P., Madalosso, T., Robaina, R. R., Duin, I. M., and Sugahara, V. H. 2018. First report of the occurrence of bacterial leaf streak of corn caused by *Xanthomonas vasicola* pv. *vasculorum* in Brazil. Plant Dis. 103:145.

Meade, B., Puricelli, E., McBride, W. D., Valdes, C., Hoffman, L., Foreman, L., and Dohlman, E. 2016. Corn and soybean production costs and export competitiveness in Argentina, Brazil, and the United States. United States Department of Agriculture, Economic Research Service. https://ideas.repec.org/p/ags/uersib/262143.html

Mende, D. R., Letunic, I., Huerta-Cepas, J., Li, S. S., Forslund, K., Sunagawa, S., and Bork, P. 2017. proGenomes: A resource for consistent functional and taxonomic annotations of prokaryotic genomes. Nucleic Acids Res. 45: D529-D534.

Menzel, P., Ng, K. L., and Krogh, A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 7:11257.

Moffett, M. L. 1983. Bacterial plant pathogens recorded in Australia. Pages 317-336 in: Plant Bacterial Diseases: A Diagnostic Guide. Academic Press, Sydney.

Monteil, C. L., Yahara, K., Studholme, D. J., Mageiros, L., Méric, G., Swingle, B., Morris, C. E., Vinatzer, B. A., and Sheppard, S. K. 2016. Population-genomic insights into emergence, crop adaptation and dissemination of *Pseudomonas syringae* pathogens. Microbiol. Genom. 2:e000089.

Murray, G. G. R., Wang, F., Harrison, E. M., Paterson, G. K., Mather, A. E., Harris, S. R., Holmes, M. A., Rambaut, A., and Welch, J. J. 2016. The effect of genetic structure on molecular dating and tests for temporal signal. Methods Ecol. Evol. 7:80-89.

Péros, J. P., Girard, J. C., Lombard, H., Janse, J. D., and Berthier, Y. 1994. Variability of *Xanthomonas campestris* pv. *vasculorum* from sugarcane and other Gramineae in Reunion Island. Characterization of a different Xanthomonad. J. Phytopathol. 142:177-188.

Plazas, M. C., De Rossi, R. L., Brücher, E., Guerra, F. A., Vilaró, M., Guerra, G. D., Wu, G., Ortiz-Castro, M. C., and Broders, K. 2017. First report of *Xanthomonas vasicola* pv. *vasculorum* causing bacteria leaf streak of maize (*Zea mays*) in Argentina. Plant Dis. 102:1026.

Qhobela, M., Claflin, L. E., and Nowell, D. C. 1990. Evidence that *Xanthomonas campestris* pv. *zeae* can be distinguished from other pathovars capable of infecting maize by restriction fragment length polymorphism of genomic DNA. Can. J. Plant Pathol. 12:183-186.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramisetty, B., and Sudhakari, P. A. 2019. Bacterial 'grounded' prophages: Hotspots for genetic renovation and innovation. Front. Genet. 10:65.

Rodriguez-R, L. M., and Konstantinidis, K. T. 2016. The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900v1. https://peerj.com/preprints/1900/

Sanko, T. J., Kraemer, A. S., Niemann, N., Gupta, A. K., Flett, B. C., Mienie, C., and Bezuidenhout, C. C. 2018. Draft genome assemblages of 10 *Xanthomonas vasicola* pv. *zeae* strains, pathogens causing leaf streak disease of maize in South Africa. Genome Announc. 6:e00532-e18.

Saux, M. F.-L., Bonneau, S., Essakhi, S., Manceau, C., and Jacques, M.-A. 2015. Aggressive emerging pathovars of *Xanthomonas arboricola* represent widespread epidemic clones distinct from poorly pathogenic strains, as revealed by multilocus sequence typing. Appl. Environ. Microbiol. 81: 4651-4668.

Schliep, K. P. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics 27: 592-593.

Seemann, T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069.

Sheppard, S. K., Guttman, D. S., and Fitzgerald, J. R. 2018. Population genomics of bacterial host adaptation. Nat. Rev. Genet. 19:549-565.

Studholme, D. J., Wicker, E., Abrare, S. M., Aspin, A., Bogdanove, A., Broders, K., Dubrow, Z., Grant, M., Jones, J. B., Karamura, G., Lang, J., Leach, J., Mahuku, G., Nakato, G. V., Coutinho, T., Smith, J., and Bull, C. T. 2020. Transfer of *Xanthomonas campestris* pv. *arecae* and *X. campestris* pv. *musacearum* to *X. vasicola* (Vauterin) as *X. vasicola* pv. *arecae* comb. nov. and *X. vasicola* pv. *musacearum* comb. nov. and description of *X. vasicola* pv. *vasculorum* pv. nov. Phytopathology. 110:1153–1160.

Tennekes, M., and Ellis, P. 2017. treemap: Treemap Visualization. https://cran.r-project.org/web/packages/treemap/index.html

Tushemereirwe, W., Kangire, A., Ssekiwoko, F., Offord, L. C., Crozier, J., Boa, E., Rutherford, M., and Smith, J. J. 2004. First report of *Xanthomonas campestris* pv. *musacearum* on banana in Uganda. Plant Pathol. 53:802.

USDA-NASS. 2017. Crop production summary 2016, United States Department of Agriculture, National Agricultural Statistics Service. Washington, D.C.: United States Department of Agriculture, Statistical Reporting Service, Crop Reporting Board. https://usda.library.cornell.edu/concern/publications/k3569432s?locale=en

Vanhove, M., Retchless, A. C., Sicard, A., Rieux, A., Coletta-Filho, H. D., Fuente, L. D. L., Stenger, D. C., and Almeida, R. P. P. 2019. Genomic diversity and recombination among *Xylella fastidiosa* subspecies. Appl. Environ. Microbiol. 85:e02972-e18.

Varani, A. M., Monteiro-Vitorello, C. B., Nakaya, H. I., and Van Sluys, M.-A. 2013. The role of prophage in plant-pathogenic bacteria. Annu. Rev. Phytopathol. 51:429-451.

Varani, A. M., Souza, R. C., Nakaya, H. I., de Lima, W. C., de Almeida, L. G. P., Kitajima, E. W., Chen, J., Civerolo, E., Vasconcelos, A. T. R., and Van Sluys, M.-A. 2008. Origins of the *Xylella fastidiosa* prophage-like regions and their impact in genome differentiation. PLoS One 3:e4059.

Vera Cruz, C. M., Bai, J., Oña, I., Leung, H., Nelson, R. J., Mew, T.-W., and Leach, J. E. 2000. Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. Proc. Natl. Acad. Sci. USA 97: 13500-13505.

Vinatzer, B. A., Monteil, C. L., and Clarke, C. R. 2014. Harnessing population genomics to understand how bacterial pathogens emerge, adapt to crop hosts, and disseminate. Annu. Rev. Phytopathol. 52:19-43.

White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. 2009. The type III effectors of *Xanthomonas*. Mol. Plant Pathol. 10:749-766.

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput. Biol. 13:e1005595.

Wiedenbeck, J., and Cohan, F. M. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. 35:957-976.

Xie, Z., and Tang, H. 2017. ISEScan: Automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics 33:3340-3347.

Yin, T., Cook, D., and Lawrence, M. 2012. ggbio: An R package for extending the grammar of graphics for genomic data. Genome Biol. 13:R77.

Young, J. M., Bradbury, J. F., Davis, R. E., Dickey, R. S., Ercolani, G. L., Hayward, A. C., and Vidader, A. K. 1991. Nomenclatural revisions of plant pathogenic bacteria and list of names 1980-1988. Rev. Plant Pathol. 70: 211-221.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. 2017. ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8:28-36.

Zaworski, F. 2016. Winter Breeding Programs in South America. SeedWorld. https://seedworld.com/winter-breeding-programs-south-america/