Journal of Bacteriology

Two New Complete Genome Sequences Offer Insight into Host and Tissue Specificity of Plant Pathogenic Xanthomonas spp.

Adam J. Bogdanove, Ralf Koebnik, Hong Lu, Ayako Furutani, Samuel V. Angiuoli, Prabhu B. Patil, Marie-Anne Van Sluys, Robert P. Ryan, Damien F. Meyer, Sang-Wook Han, Gudlur Aparna, Misha Rajaram, Arthur L. Delcher, Adam M. Phillippy, Daniela Puiu, Michael C. Schatz, Martin Shumway, Daniel D. Sommer, Cole Trapnell, Faiza Benahmed, George Dimitrov, Ramana Madupu, Diana Radune, Steven Sullivan, Gopaljee Jha, Hiromichi Ishihara, Sang-Won Lee, Alok Pandey, Vikas Sharma, Malinee Sriariyanun, Boris Szurek, Casiana M. Vera-Cruz, Karin S. Dorman, Pamela C. Ronald, Valérie Verdier, J. Maxwell Dow, Ramesh V. Sonti, Seiji Tsuge, Volker P. Brendel, Pablo D. Rabinowicz, Jan E. Leach, Frank F. White and Steven L. Salzberg

J. Bacteriol. **2011, 193(19):5450. DOI: 10.1128/JB.05262-11. Published Ahead of Print 22 July 2011.**

JOURNAL OF BACTERIOLOGY, Oct. 2011, p. 5450–5464 Vol. 193, No. 19
0021-9193/11/$12.00 doi:10.1128/JB.05262-11

Two New Complete Genome Sequences Offer Insight into Host and Tissue Specificity of Plant Pathogenic *Xanthomonas* spp.▽†

Adam J. Bogdanove,[1]* Ralf Koebnik,[2] Hong Lu,[3] Ayako Furutani,[4,5] Samuel V. Angiuoli,[6] Prabhu B. Patil,[7] Marie-Anne Van Sluys,[8,9] Robert P. Ryan,[10] Damien F. Meyer,[1,11] Sang-Wook Han,[12] Gudlur Aparna,[13,14] Misha Rajaram,[15,16] Arthur L. Delcher,[6,29] Adam M. Phillippy,[6,17] Daniela Puiu,[6] Michael C. Schatz,[6,18] Martin Shumway,[19] Daniel D. Sommer,[6] Cole Trapnell,[1,20] Faiza Benahmed,[21,22] George Dimitrov,[21] Ramana Madupu,[21] Diana Radune,[21] Steven Sullivan,[21,23] Gopaljee Jha,[13,24] Hiromichi Ishihara,[8,25] Sang-Won Lee,[12,26] Alok Pandey,[13] Vikas Sharma,[7] Malinee Sriariyanun,[12] Boris Szurek,[2] Casiana M. Vera-Cruz,[27] Karin S. Dorman,[15] Pamela C. Ronald,[12] Valérie Verdier,[2] J. Maxwell Dow,[10] Ramesh V. Sonti,[13] Seiji Tsuge,[28] Volker P. Brendel,[3,15] Pablo D. Rabinowicz,[21,29,30] Jan E. Leach,[8] Frank F. White,[31] and Steven L. Salzberg[6,32]

Department of Plant Pathology, Iowa State University, Ames, Iowa 50011[1]; Institut de Recherche pour le Développement, UMR RPB IRD-CIRAD-UM2, 911 Avenue Agropolis BP 64501, 34394 Montpellier Cedex 5, France[2]; Department of Genetics Cell and Developmental Biology, Iowa State University, Ames, Iowa 50011[3]; Department of Genetic Resources, National Institute of Agrobiological Sciences, Kannondai, Tsukuba 305-8602, Japan[4]; Gene Research Center, Ibaraki University, Ami, Ibaraki 300-0393, Japan[5]; Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland 20742[6]; Institute of Microbial Technology, Council of Scientific and Industrial Research, Sector 39A, Chandigarh 160036, India[7]; Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, Colorado 80523[8]; Departamento de Botânica, IB-USP, Sao Paulo, SP, Brazil[9]; BIOMERIT Research Centre, BioSciences Institute, University College Cork, Cork, Ireland[10]; CIRAD/INRA, Control of Emerging and Exotic Animal Diseases, Domaine de Duclos, Prise d'Eau 97170 Petit-Bourg, Guadeloupe, France[11]; Department of Plant Pathology, University of California, Davis, California 95616[12]; Centre for Cellular and Molecular Biology, Council of Scientific and Industrial Research, Hyderabad, India[13]; La Jolla Institute for Allergy and Immunology, La Jolla, California 92037[14]; Department of Statistics, Iowa State University, Ames, Iowa 50011[15]; Cardiovascular Research Institute, University of California, San Francisco, San Francisco, California 94143[16]; National Biodefense Analysis and Countermeasures Center, Frederick, Maryland 21702[17]; Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724[18]; National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland 20892[19]; Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138[20]; The Institute for Genomic Research, Rockville, Maryland 20850[21]; Division of Food and Animal Microbiology, Office of Research, Center for Veterinary Medicine, U.S. Food and Drug Administration, Laurel, Maryland 20708[22]; Center for Genomics and Systems Biology, New York University, 1009 Silver Building, 100 Washington Square East, New York, New York 10003[23]; The Institute of Himalayan Bioresouce Technology, Council of Scientific and Industrial Research, Palampur, India[24]; Department of Material Chemistry, Graduate School of Natural Science and Technology, Okayama University, Tsushima-naka 3-1-1, Okayama 700-8530, Japan[25]; Department of Plant Molecular Systems Biotechnology and Crop Biotech Institute, Kyung Hee University, Yongin 446-701, South Korea[26]; Plant Breeding, Genetics, and Biotechnology Division, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines[27]; Laboratory of Plant Pathology, Kyoto Prefectural University, Sakyo, Kyoto 606-8522 Japan[28]; Institute for Genome Sciences (IGS)[29] and Department of Biochemistry and Molecular Biology,[30] University of Maryland School of Medicine, Baltimore, Maryland 21201; Department of Plant Pathology, Kansas State University, Manhattan, Kansas 66506[31]; and McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205[32]

Received 16 May 2011/Accepted 11 July 2011

***Xanthomonas* is a large genus of bacteria that collectively cause disease on more than 300 plant species. The broad host range of the genus contrasts with stringent host and tissue specificity for individual species and pathovars. Whole-genome sequences of *Xanthomonas campestris* pv. raphani strain 756C and *X. oryzae* pv. oryzicola strain BLS256, pathogens that infect the mesophyll tissue of the leading models for plant biology, *Arabidopsis thaliana* and rice, respectively, were determined and provided insight into the genetic determinants of host and tissue specificity. Comparisons were made with genomes of closely related strains that infect the vascular tissue of the same hosts and across a larger collection of complete *Xanthomonas* genomes. The results suggest a model in which complex sets of adaptations at the level of gene content account for host specificity and subtler adaptations at the level of amino acid or noncoding regulatory nucleotide sequence determine tissue specificity.**

The genus *Xanthomonas* is a member of the class *Gammaproteobacteria* and consists of 20 plant-associated species, many of which cause important diseases of crops and ornamentals. Individual species comprise multiple pathogenic variants (pathovars [pv.]). Collectively, members of the genus cause disease on at least 124 monocot species and 268 dicot species, including fruit and nut trees, solanaceous and brassicaceous plants, and cereals (32). They cause a variety of symptoms, including necrosis, cankers, spots, and blight, and they affect a variety of plant parts, including leaves, stems, and fruits (47). The broad host range of the genus contrasts strikingly with the

* Corresponding author. Mailing address: Department of Plant Pathology, Iowa State University, 351 Bessey Hall, Iowa State University, Ames, IA 50011. Phone: (515) 294-3421. Fax: (515) 294-9420. E-mail: ajbog@iastate.edu.

† Supplemental material for this article may be found at http://jb.asm.org/.

▽ Published ahead of print on 22 July 2011.

narrow host ranges of individual species and pathovars (96), which also exhibit a marked tissue specificity by colonizing either the xylem or the intercellular spaces of nonvascular, mesophyll tissues.

Morphological and physiological uniformity within *Xanthomonas* has hampered the establishment of a stable taxonomy reflective both of the phenotypic diversity and of the evolutionary relationships within the genus (32, 86). The current taxonomy, proposed by Vauterin et al. (96) and later substantiated and refined by Rademaker et al. (67), however, is robust. On the basis of molecular genotyping, 20 species (genomic groups) are recognized, with each comprising one to several pathovars. Within this framework, examples are found both of apparent convergent evolution with regard to pathogenic traits, e.g., isolates in different genomic groups that infect the same host(s), and of divergent evolution, e.g., isolates in the same genomic group that infect different hosts or the same host differently (67). By establishing molecular genetic relationships among the more than 140 known members of the genus with distinctive pathogenic characteristics, the current taxonomy sets the stage for informed comparisons directed toward identifying unique determinants of host and tissue specificity as well as pathogenicity factors that are universally important in plant disease.

Complete genome sequences of nine *Xanthomonas* strains, representing pathovars within four species, have been published. The strains are strain 306 of *X. axonopodis* pv. citri, which causes citrus canker (18); strain 85-10 of *X. axonopodis* pv. vesicatoria, the bacterial spot pathogen of pepper and tomato, formerly a pathovar of *X. campestris* (92); strains 8004, ATCC 33913, and B100 of *X. campestris* pv. campestris, the causal agent of black rot in crucifers, including the model plant *Arabidopsis thaliana* (here *Arabidopsis*) (18, 66, 98); strains KACC10331, MAFF311018, and P*XO*99^{A} of *X. oryzae* pv. oryzae, which is responsible for bacterial blight of rice (44, 59, 74); and strain GPE PC73 of *X. albilineans*, a more distantly related, xylem-limited pathogen of sugarcane with a markedly reduced genome size relative to other *Xanthomonas* spp. (62).

We report here the complete genome sequence of the nonvascular counterpart of *X. campestris* pv. campestris, namely, *X. campestris* pv. raphani (strain 756C, formerly classified as *X. campestris* pv. armoraciae), which causes bacterial spot of crucifers, including *Arabidopsis*, and the complete genome sequence of the nonvascular counterpart of *X. oryzae* pv. oryzae, namely, *X. oryzae* pv. oryzicola (strain BLS256), which causes bacterial leaf streak of rice. In addition to facilitating functional genomics and DNA-based diagnostics important for understanding and controlling the respective diseases, these genome sequences enable key comparisons within the vascular and nonvascular pairs to shed light on tissue specificity and across the pairs to provide insight into host specificity. They also enable broader comparisons across all the available complete genome sequences to identify candidate specificity determinants as well as candidate genes fundamental to pathogenesis, irrespective of the host and tissue infected. Results of such comparisons are also presented.

MATERIALS AND METHODS

Sequencing. Bacterial genomic DNA was randomly sheared by nebulization, end repaired with consecutive BAL31 nuclease and T4 DNA polymerase treatments, and size selected using gel electrophoresis on 1% low-melting-point agarose. After ligation to BstXI adapters, DNA was purified by three rounds of gel electrophoresis to remove excess adapters and the fragments were ligated into the vector pHOS2 (a modified pBR322 vector) linearized with BstXI. The pHOS2 plasmid contains two BstXI cloning sites immediately flanked by sequencing primer binding sites. These features reduce the frequency of nonrecombinant clones and reduce the amount of vector sequences at the end of the reads. The ligation reaction products were electroporated into *Escherichia coli*. We constructed several shotgun libraries with average insert sizes of 4.5, 7, 11, and 13.5 kb for *X. oryzae* pv. oryzicola BLS256 and 4, 5, and 9 kb for *X. campestris* pv. raphani 756C. Clones were plated onto large-format (16- by 16-cm) diffusion plates prepared by layering 150 ml of fresh antibiotic-free agar onto a previously set 50-ml layer of agar containing antibiotic. Colonies were picked for template preparation, inoculated into 384-well blocks containing liquid medium, and incubated overnight with shaking. High-purity plasmid DNA was prepared using a DNA purification robotic workstation custom-built by Thermo CRS (Thermo Scientific) and based on the alkaline lysis miniprep (75) and isopropanol precipitation. DNA precipitate was washed with 70% ethanol, dried, and resuspended in 10 mM Tris-HCl buffer containing a trace of blue dextran. The yield of plasmid DNA was approximately 600 to 800 ng per clone, providing sufficient DNA for at least four sequencing reactions per template. Sequencing was done using a dideoxy sequencing method (76). Two 384-well cycle sequencing reaction plates were prepared from each plate of plasmid template DNA for opposite-end, paired-sequence reads. Sequencing reactions were completed using the BigDye Terminator chemistry and standard M13 forward and reverse primers. Reaction mixtures, thermal cycling profiles, and electrophoresis conditions were optimized to reduce the volume of the BigDye Terminator mix and to extend read lengths on the AB3730xl sequencers (Applied Biosystems). Sequencing reactions were set up by the Biomek FX pipetting workstations. Robots were used to aliquot and combine templates with reaction mixtures consisting of deoxy- and fluorescently labeled dideoxynucleotides, DNA polymerase, sequencing primers, and reaction buffer in a 5-μl volume. After 30 to 40 consecutive cycles of amplification, reaction products were precipitated by isopropanol, dried at room temperature, resuspended in water, and transferred to an AB3730xl sequencer. A total of 58,510 and 75,159 successful *X. oryzae* pv. oryzicola BLS256 and *X. campestris* pv. raphani 756C reads, respectively, were produced. The average read lengths were 859 bp for *X. oryzae* pv. oryzicola BLS256 and 796 bp for *X. campestris* pv. raphani 756C. After initial assembly, gaps were closed in each genome by primer walking on plasmid templates, sequencing genomic PCR products that spanned the gaps, and transposon insertion and sequencing of selected shotgun clones.

Assembly and annotation. Multiple rounds of assembly were performed, beginning with the shotgun reads and, later, including additional finishing reads. In the final assemblies, 74,930 reads (for *X. campestris* pv. raphani 756C) and 57,520 reads (for *X. oryzae* pv. oryzicola BLS256) were trimmed to remove vector and low-quality sequences and then assembled using Celera Assembler (57). The assemblies were evaluated using the Hawkeye assembly diagnosis software (77) to identify and correct any collapsed repeats. Protein-coding genes were identified using the Glimmer, version 3.0, program (20, 21), which includes an algorithm to identify ribosome binding sites for each gene. Transcription terminators were predicted using the TransTermHP tool (42) with parameter settings expected to yield over 90% accuracy. Transfer RNAs were identified with the tRNAScanSE server (51). Regions with neither Glimmer predictions nor RNA genes were searched in all six frames using BLASTx (2) to identify any missed proteins, and all annotations were manually curated as described previously (34). The origin and terminus of replication were determined using GC skew analysis (50), which for *X. campestris* pv. raphani 756C indicates an origin near position 1 and a terminus near 2,460 kb. The chromosome replication initiator gene *dnaA*, which is commonly found near the origin, is at position 201 in the *X. campestris* pv. raphani 756C genome and at position 43 in the *X. oryzae* pv. oryzicola BLS256 genome. Oligomer skew analysis, which identifies 8-mers preferentially located on the leading strand (73), indicates an origin for *X. campestris* pv. raphani 756C at position 99 and a terminus at 2,471 kb, based on multiple 8-mers, including CCCTGCCC and CGCCCTGC. These 8-mers occur 409/468 (87.4%) and 856/1,205 times on the leading strand; for CCCTGCCC the likelihood that this occurred by chance is 2.7×10^{-58}. Similarly, in *X. oryzae* pv. oryzicola BLS256, the 8-mer CCCTGCCC occurs 325/368 (88.3%) times on the leading strand and indicates an origin at position 57 kb and a terminus at 2,473 kb. (The oligomer skew software is available at http://cbcb.umd.edu/~salzberg/docs/skew/.)

Genome sequences compared. The genome sequences compared in this study are listed in Table 1. In the tables and figures throughout, where sequences for multiple strains of the same pathovar were identical, the sequence for only one strain is noted. The genome of *X. albilineans* GPE PC73, which is reduced and divergent relative to the other genomes, was not included.

TABLE 1. Complete *Xanthomonas* genome sequences compared in this study

Organism	Disease	Size (Mb)	Components	GenBank Accession no.	Reference	Class.[a]
X. axonopodis pv. citri 306	Citrus canker	5.27	Circular chromosome (5,175,554 bp); plasmids pXAC64 (64,920 bp) and pXAC33 (33700 bp)	NC_003919, NC_003922, NC_003921	18	ND
X. axonopodis pv. vesicatoria 85–10	Bacterial spot of pepper and tomato	5.42	Circular chromosome (5,178,466 bp); plasmids pXCV183 (182,572 bp), pXCV38 (38,116 bp), pXCV19 (19,146 bp), and pXCV2 (1,852 bp)	NC_007508, NC_007507, NC_007506, NC_007505, NC_007504	92	ND
X. campestris pv. campestris 8004	Black rot of crucifers	5.15	Circular chromosome	NC_007086	66	VD
X. campestris pv. campestris ATCC 33913	Black rot of crucifers	5.08	Circular chromosome	NC_003902	18	VD
X. campestris pv. campestris B100	Black rot of crucifers	5.08	Circular chromosome	NC_003902	98	VD
X. campestris pv. raphani 756C	Leaf spot of crucifers and solanaceous species	4.94	Circular chromosome	CP002789	This work	ND
X. oryzae pv. oryzicola BLS256	Bacterial leaf streak of rice	4.83	Circular chromosome	AAQN01000001	This work	NM
X. oryzae pv. oryzae KACC10331	Bacterial blight of rice	4.94	Circular chromosome	NC_006834	44	VM
X. oryzae pv. oryzae MAFF 311018	Bacterial blight of rice	4.94	Circular chromosome	NC_007705	59	VM
X. oryzae pv. oryzae PXO99^A	Bacterial blight of rice	5.24	Circular chromosome	NC_010717	74	VM

[a] Class., classifications: D, pathogen of dicot(s); M, pathogen of monocot(s); N, infects the nonvascular tissue; V, infects the vascular tissue.

Phylogenomic analysis. The complete nucleotide sequences of a set of 31 phylogenetic marker genes (primarily genes involved in transcription, translation, and replication [103]) were extracted from each of the genomes and aligned to generate a maximum-likelihood tree with 1,000 bootstrap replicates using the MEGA5 program and applying the GTR + G (gamma distribution) model of nucleotide substitution (89; http://www.megasoftware.net). Sequences from *Stenotrophomonas maltophilia* strain K279a (15) were used as an outgroup.

Genome alignments. Whole-genome alignments were carried out using the MAUVE program with default parameters (16).

TAL effector phylogenetic analysis. Nucleotide sequences of transcription activator-like (TAL) effector genes from *X. oryzae* pv. oryzicola BLS256, *X. oryzae* pv. oryzae MAFF 311018, and *X. oryzae* pv. oryzae PXO99^A and of *pthA* of *X. axonopodis* pv. citri (gi 899439), with the repeat regions removed, were aligned using the software MAFFT (40) with the L-INS-I utility. Pseudogenes were excluded. The repeat region was identified as starting after the sequence CCC CCTGAAC and ending before (A/G)GCATTGTTGC. The Phyml algorithm (30) was used to construct the maximum-likelihood phylogenetic tree on the basis of the alignment with 100 bootstrap replicates. The jModelTest tool (63) identified the TPM1uf+I+G model to be the best according to the Bayesian information criterion, and this model was used by Phyml. TPM1uf is the Kimura 3-parameter model (K81) (41) with unequal base frequencies (uf), invariant sites (I), and gamma-distributed rates (G). Maximum-likelihood phylogenetic trees were also constructed from amino acid alignments using the JTT+G+F model, selected by the ProtTest program (1).

Genome-wide identification of strain category-specific and essential genes. The annotated protein sequences of each genome in Table 1, 44,658 protein sequences total, were pairwise compared by using BLASTp (11) with an E-value threshold of 1e−10, complexity filter inactivated, and output formatted for the OrthoMCL program (option 8). The BLASTp output was processed with the OrthoMCL program using default parameters (48), resulting in 5,285 clusters of homologous proteins, with each cluster consisting of at least two sequences, which were not necessarily from different strains. Unique sequences were then included as singlets with the set of clusters. Finally, all sequences annotated as transposon related were collected in a separate BLAST database, and sequences in the set of clusters were deleted if they matched sequences in the transposon database file at a threshold of 1e−20. Clusters that had more than 50% of their initial members culled were entirely removed. Also removed were short sequences of less than 60 amino acids. The final set of clustered sequences consisted of 7,040 clusters, ranging from 1 cluster with 81 members to 2,112 singlets.

On the basis of the strains represented in them, clusters were associated with 1 of 15 categories derived from the two independent classifications of strains as either vascular (V) or nonvascular (N) in their mode of infection and monocot (M) or dicot (D) with respect to their host(s). To determine genes specific, essential, or specific and essential to different strain categories, Venn diagrams were drawn in R language using the VennDiagram package (13).

In-house computational scripts for automation of the analyses and intermediate output files are available at http://brendelgroup.org/JBac11/.

Nucleotide sequence accession numbers. The complete genome sequences and annotation for *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 have been deposited in GenBank under accession numbers CP002789 and AAQN01000001, respectively. The traces have been deposited in the NCBI Trace Archive (http://www.ncbi.nlm.nih.gov/Traces), and the complete assemblies are in the NCBI Assembly Archive (http://www.ncbi.nlm.nih.gov/Traces/assembly).

RESULTS AND DISCUSSION

Strains sequenced. *X. campestris* pv. raphani 756C is a chloramphenicol-resistant derivate of strain 756 which can be transformed efficiently and is virulent on *Arabidopsis* (38, 39; A. J. Bogdanove, unpublished data). Strain 756 was isolated from cabbage in East Asia prior to 1985 (3, 4). The strain was used in early studies of the role of *hrpX* in bacterial blight and bacterial leaf spot of crucifers (39). Originally classified as *X. campestris* pv. armoraciae, *X. campestris* pv. raphani 756C is among a large collection of *X. campestris* strains that had been isolated from cruciferous plants and classified under six pathovar designations (including *X. campestris* pv. *armoraciae*) that were recently regrouped by Fargier and Manceau into *X. campestris* pv. campestris, which they defined as causing black

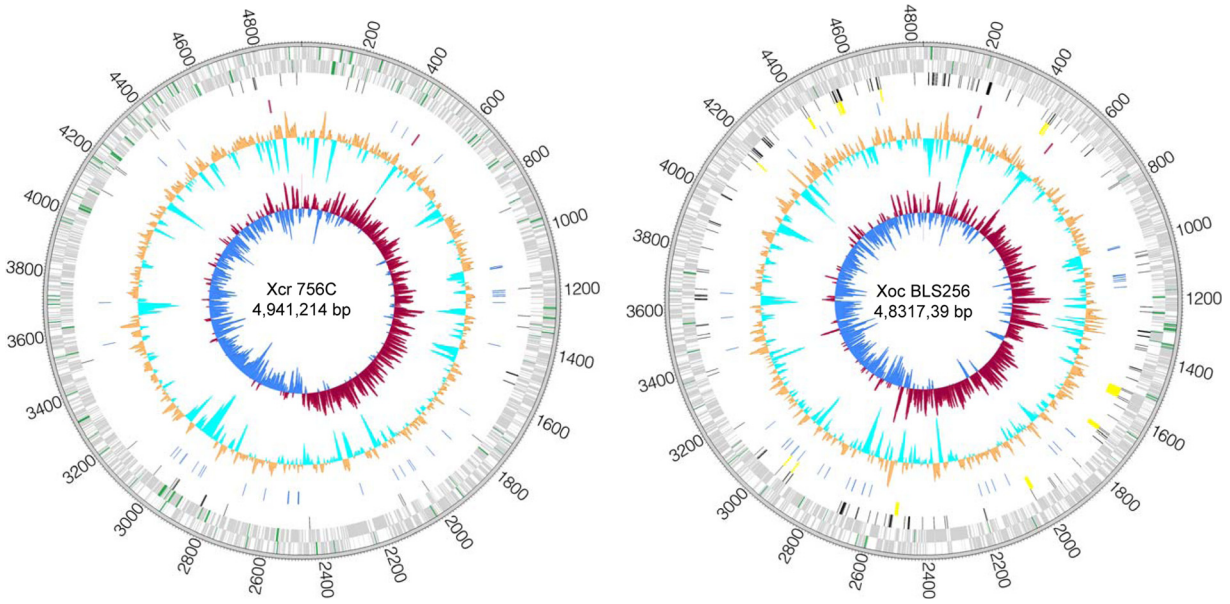

FIG. 1. The *X. campestris* pv. raphani (Xcr) 756C and *X. oryzae* pv. oryzicola (Xoc) BLS256 genomes. Rings illustrate, from outside to inside, genome coordinates (kb), protein-coding genes in forward (outer) and reverse strands, IS elements (black), TAL effector genes (yellow), tRNA (blue) and rRNA (red) genes, percent G+C content, and GC skew. GC skew shows (G − C)/(G + C) in 10-kb windows; positive values indicate the leading strand of replication, and negative values indicate the lagging strand. Protein-coding genes specific and essential (see text) to the strains in Table 1 that are pathogenic to dicots are highlighted (green) in the *X. campestris* pv. raphani circle, and those specific to pathogens of monocots are highlighted in the *X. oryzae* pv. oryzicola circle.

rot on at least one cruciferous species; *X. campestris* pv. incanae, causing bacterial blight on garden stock and wallflower; and *X. campestris* pv. raphani, causing bacterial leaf spot on both cruciferous and solanaceous plants (24). Fargier and colleagues have since shown additional support for this regrouping by multilocus sequence typing (25). Further, the authors identified *X. campestris* pv. raphani 756C as *X. campestris* pv. raphani race 3 (24) using the set of differential cultivars and accessions of *Brassica* spp. established by Vicente et al. (97).

X. oryzae pv. oryzicola BLS256 is a moderately to highly virulent isolate collected in 1984 from a breeding plot at the International Rice Research Institute (IRRI) Experiment Station in the Philippines (C. M. Vera-Cruz, unpublished data). *X. oryzae* pv. oryzicola BLS256 has been used previously in molecular studies. It was shown to contain a homologue of the *hrpXo* gene from *X. oryzae* pv. oryzae PXO99^{A} (69), and it is the original source of the *avrRxo1* avirulence gene that triggers gene-for-gene resistance in maize and transgenic rice plants carrying the *Rxo1* resistance gene (107, 108).

***X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 genomes.** The *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 genomes are single circular chromosomes of 4,941,214 bp and 4,831,739 bp, respectively, with 65.3% and 64.0% GC contents, respectively (Fig. 1). The *X. campestris* pv. raphani 756C genome contains 4,535 annotated protein-coding genes and 57 tRNA genes. The *X. oryzae* pv. oryzicola BLS256 genome contains 4,480 protein-coding genes and 54 tRNA genes. Each genome contains two rRNA operons (Table 2).

Phylogenomic relationships among completely sequenced *Xanthomonas* strains. A genome-based phylogenetic relationship among the completely sequenced *Xanthomonas* strains is shown in Fig. 2. This phylogenomic inference was drawn on the basis of the complete sequences of a set of 31 conserved housekeeping genes that are considered to be rarely subjected to horizontal gene transfer (35). The same set of genes was recently used to establish phylogenomic relationships of 106 bacterial and archaeal genomes (102). The *Xanthomonas* tree is well resolved and robust, with high bootstrap values for all clades. As expected, *X. oryzae* pv. oryzicola BLS256 and all *X. oryzae* pv. oryzae strains together form their own phylogenetic group. In contrast, *X. campestris* pv. raphani 756C, though closely related to the *X. campestris* pv. campestris strains, forms a distinct clade. The tree also reveals that the *X. oryzae* strains are more closely related to the *X. axonopodis* strains than to the *X. campestris* strains.

IS elements and genome plasticity. A striking feature shared by the *Xanthomonas* genomes is an abundance of insertion sequence (IS) elements, which are postulated to be important drivers of *Xanthomonas* genome evolution (55, 92). *X. oryzae* strains have particularly large numbers and the greatest diversity of IS elements yet observed in the genus (74). *X. oryzae* pv.

TABLE 2. Characteristics of the *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 genomes

Characteristic	*X. campestris* pv. raphani 756C	*X. oryzae* pv. BLS256
Length (bp)	4,941,214	4,831,739
GC content (%)	65.28	64.05
No. of protein-coding genes	4,535	4,480
% genes assigned a role category	69.23	72.15
No. of tRNA genes	57	54
No. of rRNA operons	2	2

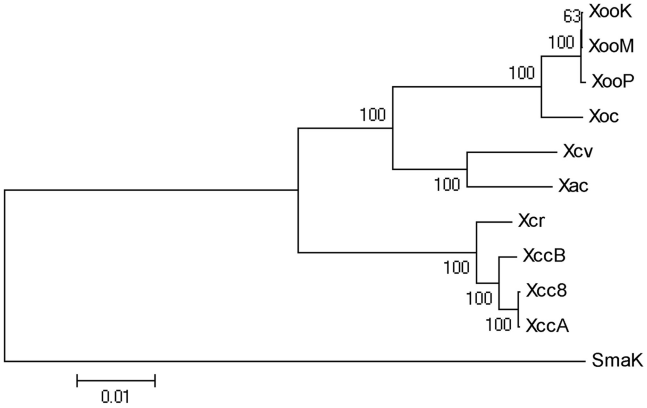

FIG. 2. Phylogenomic relationships among sequenced *Xanthomonas* strains. A maximum likelihood tree based on a set of 31 conserved phylogenomic marker genes (see text) is shown. Bootstrap values are displayed at nodes. *Stenotrophomonas maltophilia* strain K279a (SmaK) was used as an outgroup. XooK, XooM, XooP, *X. oryzae* pv. oryzae strains KACC10331, MAFF311018, and PXO99^{A}, respectively; Xoc, *X. oryzae* pv. oryzicola BLS256; Xcv, *Xanthomonas axonopodis* pv. vesicatoria 85-10; Xac, *X. axonopodis* pv. citri 306; Xac, *X. campestris* pv. raphani 756C; XccB, Xcc8, and XccA, *X. campestris* pv. campestris strains 8004, ATCC33913, and B100, respectively.

oryzicola BLS256 is no exception, with 245 IS elements (including partial ones) representing six distinct families. The *X. campestris* pv. raphani 756C genome carries only 41 IS elements representing three families. When IS element content is compared across sequenced *X. oryzae* and *X. campestris* genomes, the families represented and patterns of amplification are similar within pathovars but distinct across pathovars and across species (see Fig. S1 and Table S1 in the supplemental material). Only three out of the eight IS families detected, IS*3*, IS*4*, and IS*5*, are represented in all strains. Members of the IS*5* family are the most numerous. Four families, IS*256*, IS*30*, IS*630*, and IS*L3*, are represented only in *X. oryzae* strains. IS*1114* (IS*5* family) is the only individual IS element present in all the sequenced *Xanthomonas* genomes, including the *X. axonopodis* pv. citri 306 and *X. axonopodis* pv. vesicatoria 85-10 genomes.

In addition to serving as vectors for lateral gene transfer, IS elements can generate other types of genome modifications, including rearrangements, inversions, and deletions, any of which can lead to acquisition, modification, or loss of gene content. Whole-genome alignments of *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 to their respective vascular counterparts reveal numerous rearrangements and inversions within each group, but these are markedly more prevalent among the *X. oryzae* strains (Fig. 3). The *X. oryzae* strains also show larger and more numerous insertions and deletions. Not surprisingly, most of the breakpoints in the alignments in both groups are associated with IS elements (see Fig. S2 in the supplemental material).

Though not obligate pathogens, conceivably, the *X. oryzae* strains tolerate larger numbers of IS elements and consequent gene disruptions due to a more consistent association with their host plant. Rice is cultivated more extensively than the cruciferous hosts of *X. campestris* and often across multiple seasons per year, providing a more stable environment than *X. campestris* might experience. This could alleviate the requirement for some genes and thereby make the genome more tolerant of IS element proliferation. Alternatively or additionally, intensive cultivation of rice might have imposed selection on *X. oryzae* for the ability to rapidly adapt genetically. Genome plasticity conferred by IS elements may represent a mechanism for this.

Different patterns of IS element distribution in the two otherwise genetically highly similar pathovars support this notion.

T3E genes. Type III effectors (T3Es) are delivered into plant cells via the *hrp* gene-encoded type III secretion (T3S) system, which is required for pathogenesis in both *X. campestris* pv. raphani and *X. oryzae* pv. oryzicola (39, 53). They include proteins that act as virulence factors by promoting disease or avirulence factors by triggering host plant defense, depending on the host genotype. Type III effector genes in the *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 genomes were predicted by BLASTn using a database of identified and candidate T3Es of plant pathogens (http://www.xanthomonas.org).

X. campestris pv. raphani 756C differs markedly with respect to its repertoire of T3E genes from the other *Xanthomonas* strains, which generally contain a core set of 10 (*avrBs2*, *xopF1*, *xopK*, *xopL*, *xopN*, *xopP*, *xopQ*, *xopR*, *xopX*, and *xopZ1*) (http://www.xanthomonas.org). *X. campestris* pv. raphani 756C contains only three of these (*xopF1*, *xopP*, and *xopR*) and four other T3E genes that are less broadly conserved (Table 3). All the T3S genes themselves, including the *hpa* and *hrpW* accessory protein genes, are intact (52). Bordering the *hrp* gene cluster, a novel candidate T3E gene, *xopAT*, was identified on the basis of the presence of a plant-inducible promoter (PIP) box and a properly positioned −10 box-like sequence upstream of the coding sequence (Table 3; see Table S2 in the supplemental material). The PIP box is a conserved promoter element for the coregulation of *hrp* and many T3E genes by the activator HrpX (26, 28, 43, 91, 95). Including *xopAT*, seven of the eight *X. campestris* pv. raphani 756C candidate T3E genes are preceded by the PIP and −10 motifs (see Table S2 in the supplemental material). Like other T3E genes, *xopAT* also has an unusual (low) G+C content suggestive of acquisition by horizontal gene transfer (Table 3). Finally, in common with other T3Es, the encoded protein contains eukaryotic motifs for myristoylation and palmitoylation (93) but otherwise has no similarity to other known proteins. *X. campestris* pv. raphani 756C carries two other genes, *avrXccA1* and *avrXccA2*, that are related to *avrXca* of *X. campestris* pv. raphani strain 1067, which was shown to function as an avirulence gene when expressed heterologously in *X. campestris* pv. campestris 8004 inoculated into *Arabidopsis thaliana* (60). Some features of these genes, however, including an N-terminal signal peptide for type II secretion, suggest that they are not bona fide T3Es. Overall, across the sequenced *Xanthomonas* genomes, the T3E gene content of *X. campestris* pv. raphani 756C is most similar to that of the other *X. campestris* strains. Of particular interest, *X. campestris* pv. raphani 756C contains *xopAC* (synonym *avrAC*), which in *X. campestris* pv. campestris 8004 triggers vascular tissue-specific resistance in *A. thaliana* ecotype Col-0 (104). The gene in *X. campestris* pv. raphani 756C is highly conserved with its homolog in *X. campestris* pv. campestris 8004, including its promoter, which contains the same imper-

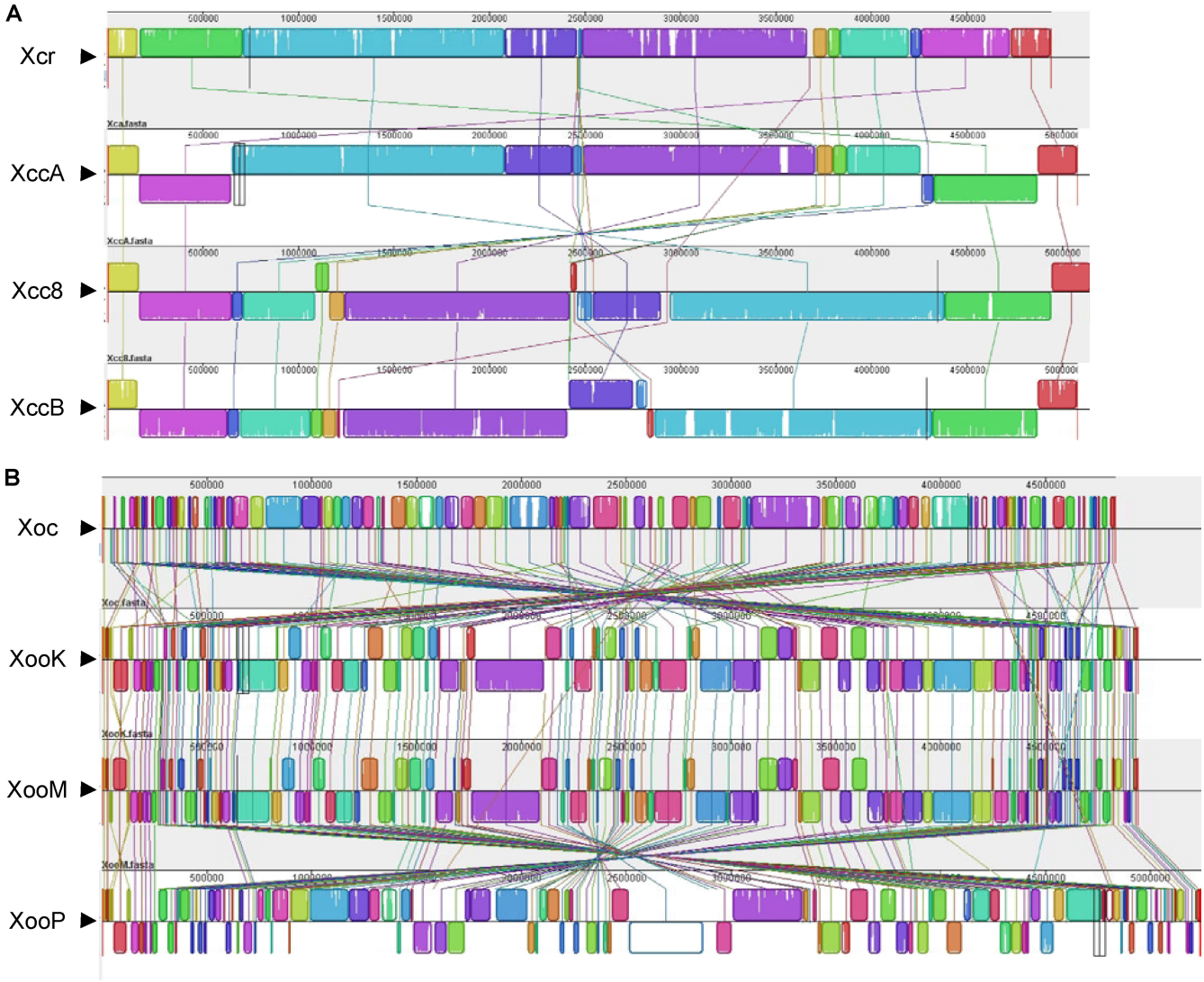

FIG. 3. Alignments of the *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 genomes to those of their vascular counterparts. (A) Whole-genome alignments of *X. campestris* pv. raphani 756C with *X. campestris* pv. campestris 8004, *X. campestris* pv. campestris ATCC 33913, and *X. campestris* pv. campestris B100; (B) whole-genome alignments of *X. oryzae* pv. oryzicola BLS256 with *X. oryzae* pv. oryzae KACC10331, *X. oryzae* pv. oryzae MAFF311018, and *X. oryzae* pv. oryzae PXO99^{A}. Alignments were generated using the MAUVE program (18) with default parameters. Locally colinear blocks, shown as rounded rectangles, represent regions without rearrangement of homologous sequence across genomes. The orientation of the locally colinear blocks, forward or reverse, is indicated by their position above or below the line, respectively. Lines between genomes trace orthologous locally colinear blocks. The minimum locally colinear block weight and the number of locally colinear blocks are shown for each alignment. Inside each locally colinear block, vertical bars represent the average level of conservation of sequence across genomes.

fect PIP box and −10 box-like sequences. An *xopAC* mutant of *X. campestris* pv. campestris 8004, retained full virulence on *A. thaliana* ecotype Kashmir and became virulent on *A. thaliana* ecotype Col-0 following vein inoculation (104). It would be of interest to test whether mutation of the *xopAC* allele in *X. campestris* pv. raphani 756C might enable it to colonize the xylem of its hosts.

Twenty-six candidate T3E genes were identified in the *X. oryzae* pv. oryzicola BLS256 genome, with 19 preceded by the PIP and −10 motifs (Table 4; see Table S3 in the supplemental material). This number does not include the large family of TAL effector genes, discussed in the next section. In addition to each of the 10 core T3E genes, several candidates with homologs in the genomes of only a subset of the *Xanthomonas* strains or other plant pathogenic bacteria were found. One, *xopAF*, is present only in *X. oryzae* pv. oryzicola BLS256 and three *Pseudomonas syringae* genomes examined, though a homolog exists in *X. axonopodis* pv. vesicatoria strain 91-118 (5). With the exception of *xopAF* and four other T3E genes, *xopAJ* (synonym *avrRxo1*), *xopAK*, *xopI*, and *xopO*, all of the *X. oryzae* pv. oryzicola BLS256 T3E genes are present in each of the three *X. oryzae* pv. oryzae genomes, though *xopU* and *xopY* are pseudogenes in *X. oryzae* pv. oryzae KACC10331. *X. oryzae* pv. oryzicola was shown to inhibit the rice defense response to several TAL effector avirulence proteins of *X. oryzae* pv. oryzae; this ability was T3S dependent, indicating that one or

TABLE 3. *X. campestris* pv. raphani 756C genes encoding type III effectors and effector candidates, other type III-secreted proteins, and other candidate avirulence proteins

Gene	*X. campestris* pv. raphani 756C locus tag no.[b]	% GC content	Gene distribution[a]					
			X. axonopodis pv. citri 306	*X. axonopodis* pv. vesicatoria 85-10	*X. campestris* pv. campestris 8004	*X. campestris* pv. campestris ATCC 33913	*X. campestris* pv. campestris B100	*X. oryzae* pv. oryzicola BLS256
avrXccA1	4581	69.5	/	/	+	+	+	/
avrXccA2	2696	62.1	/	/	+	+	+	/
hpa1	1492	58.5	+	+	+	+	+	+
hpaA	1476	67.3	+	+	+	+	+	+
hrpW	1471	63.3	+	+	+	+	+	/
xopAC	2914	56.4	/	/	+	+	+	/
xopAD	1464	62.1	+	Ψ	/	/	/	+
xopAL1	1499	49.4	/	/	+	+	+	/
xopAR	4085.a	46.6	/	/	+	+	+	/
xopAT	1464.a	48.7	/	/	/	/	/	/
xopF1	1470	66.1	/	+	+	+	+	+
xopP	1500	57.3	+	+	+	+	+	+
xopR	4254	63.6	+	+	+	+	+	+

[a] Distribution among the complete *Xanthomonas* genomes and other plant pathogen genomes is shown: +, presence of a homolog; /, absence of a homolog; (+), weak or partial similarity; Ψ, pseudogene due to frameshift or premature stop codon.

more *X. oryzae* pv. oryzicola T3Es interferes with rice resistance gene-mediated effector recognition or signaling (53). The ones missing from the *X. oryzae* pv. oryzae strains are attractive candidates. No virulence function has yet been reported for XopAJ (107), but a homolog of XopAF in *P. syringae* was shown to suppress plant basal immunity (49). Also, HopK1, the *P. syringae* homolog of XopAK, was reported to suppress the defense-associated hypersensitive reaction triggered by T3E HopPsyA in *Nicotiana tabacum* and to induce jasmonic acid-responsive genes (33, 36). Four of the five *X. oryzae* pv. oryzicola BLS256 T3E genes missing from the *X. oryzae* pv. oryzae strains are conserved in *X. axonopodis* pv. vesicatoria 85-10, and as noted, the fifth, *xopAF*, is present in *X. axonopodis* pv. vesicatoria strain 91-118. It is tempting to speculate also therefore that one or more of these play a role in tissue specificity as well. Across the *Xanthomonas* genomes, *xopC2*, which encodes an SKWP-related protein, is present and intact only in the *X. oryzae* strains, suggesting a possible role in *Xanthomonas* adaptation to monocots, though an allele is also present in *Ralstonia solanacearum*, a dicot pathogen.

TAL effector genes. Upon delivery into host cells via the type III secretion system, TAL effectors enter the nucleus, bind to effector-specific sequences in host gene promoters, and drive transcription of host disease susceptibility (*S*) genes that facilitate bacterial colonization, symptom development, or pathogen dissemination (for recent reviews, see references 8, 9, and 100). TAL effector specificity is determined by a central domain of tandem, 33- to 35-amino-acid repeats. A polymorphic pair of residues at positions 12 and 13 in each repeat, the repeat-variable diresidue (RVD), specifies individual nucleotides in the binding site, with the four most common RVDs each preferentially associating with one of the four bases. Binding sites, which correspond in length to the number of repeats in the effector, are referred to as upregulated by TAL effector (UPT) boxes.

Though some *X. campestris* pv. raphani strains and others reported to be *X. campestris* pv. armoraciae carry TAL effector genes, the *X. campestris* pv. raphani 756C genome has none. Among *Xanthomonas* spp. that carry TAL effector genes, *X. oryzae* stands out in terms of the large numbers present in individual strains and in the diversity across strains. *X. oryzae* pv. oryzicola BLS256 is no exception: it contains 26 TAL effector genes and 2 TAL effector pseudogenes (see Table S4 and Fig. S3 in the supplemental material). Several TAL effectors of *X. oryzae* pv. oryzae have been functionally characterized (for a review, see reference 100). They include both virulence factors, which activate host *S* genes, and avirulence factors, so called because they activate host resistance (*R*) genes. Function has yet to be assigned for any *X. oryzae* pv. oryzicola TAL effector, but candidate targets in rice for six *X. oryzae* pv. oryzicola BLS256 effectors were identified on the basis of UPT box nucleotide sequences predicted from the effector RVDs (56).

As in *X. oryzae* pv. oryzae KACC10331, *X. oryzae* pv. oryzae MAFF 311018, and *X. oryzae* pv. oryzae PXO99^{A}, all of the encoded *X. oryzae* pv. oryzicola BLS256 TAL effector repeats are either 33 or 34 amino acids long, except the fifth repeat of Tal9b, which is missing 4 amino acids at the end following the RVD. The numbers of RVDs range from 13 to 27 (see Fig. S3 in the supplemental material). The RVD sequences are distinct from one another and from those of the *X. oryzae* pv. oryzae TAL effectors, suggesting that the two pathovars generally target different host genes. Furthermore, the *X. oryzae* pv. oryzicola BLS256 and *X. oryzae* pv. oryzae TAL effectors appear to have differentiated independently (Fig. 4). The maximum-likelihood tree inferred from an alignment of *X. oryzae* pv. oryzicola BLS256, *X. oryzae* pv. oryzae MAFF 311018, and *X. oryzae* pv. oryzae PXO99^{A} TAL effector nucleotide sequences, with the repeat regions removed and using the *pthA* gene of *X. axonopodis* pv. citri as the outgroup, places all *X. oryzae* pv. oryzicola BLS256 TAL effector sequences in a monophyletic clade with 73% bootstrap support. The *X. oryzae* pv. oryzae TAL effector sequences are also monophyletic, albeit with lower bootstrap support (20%). The low bootstrap support for the *X. oryzae* pv. oryzae clade reflects uncertainty in the location of the outgroup. However, analyses with 17 other

TABLE 3—*Continued*

Gene distribution[a]							
X. oryzae pv. oryzae KACC10331	*X. oryzae* pv. oryzae MAFF 311018	*X. oryzae* pv. oryzae PXO99^A	*P. syringae* pv. tomato DC3000	*Pseudomonas syringae* pv. phaseolicola 1448A	*P. syringae* pv. syringae B728a	*Ralstonia solanacearum* GMI1000	*Erwinia carotovora* subsp. atroseptica SCRI104
/	/	/	/	/	/	/	+
/	/	/	/	/	/	/	/
+	+	+	/	/	/	/	/
+	+	+	/	/	/	/	/
/	/	/	(+)	(+)	(+)	+	(+)
/	/	/	/	/	/	/	/
+	+	+	/	+	/	+	/
/	/	/	/	/	/	/	/
/	/	/	/	/	/	/	/
/	/	/	/	/	/	/	/
+	+	+	/	/	/	/	/
+	+	+	/	/	/	+	/
+	+	+	/	/	/	/	/

members of the *Xanthomonas* TAL effector (79) family as outgroups each consistently split the *X. oryzae* pv. oryzicola BLS256 and *X. oryzae* pv. oryzae sequences into distinct monophyletic groups (data not shown). *X. oryzae* pv. oryzae KACC10331 TAL effectors were not included in the analysis due to a discrepancy in the number of these genes in the published genome and previous estimates by Southern blotting suggesting misassembly (44, 105). Nevertheless, the evidence suggests that a single TAL effector gene existed in the common ancestor of *X. oryzae* pv. oryzicola and *X. oryzae* pv. oryzae that was recruited for specialization and underwent duplication independently in the two lineages. The tree also suggests that duplication occurred both before and after the split of *X. oryzae* pv. oryzae MAFF 311018 and *X. oryzae* pv. oryzae PXO99^A. Though phylogenetic resolution is low within the *X. oryzae* pv. oryzae cluster, there is some indication of gene loss as well. Another possibility is that intragenome recombination, encouraged by the central repeat region, causes inconclusive phylogenetic signals within the *X. oryzae* pv. oryzae lineage. Analysis of amino acid sequence data (data not shown) also shows a monophyletic *X. oryzae* pv. oryzicola clade; however, the *X. oryzae* pv. oryzae clade is split by the outgroup. This tree suggests at least two TAL effector genes in the ancestor, but low bootstrap support and conflicting results from the more informative DNA sequence data (83) argue against this conclusion.

The *X. oryzae* pv. oryzicola BLS256 TAL effector genes are distributed across 12 loci in the genome (see Fig. S3 in the supplemental material), defined by proximity and shared orientation of the genes at each locus. Five of these contain just one gene, three contain two genes, three contain three genes, and the remaining locus contains six genes and the two pseudogenes. In contrast to TAL effector genes in the sequenced *X. oryzae* pv. oryzae genomes, which are typically separated by a conserved 989-bp sequence, the *X. oryzae* pv. oryzicola BLS256 TAL effectors are uniformly preceded immediately by 108 to 134 bp of conserved sequence similar to the start codon proximal end of the 989-bp *X. oryzae* pv. oryzae sequence, but with a 7-bp insertion and 10 substitutions. Though shorter, this sequence suggests that, as in *X. oryzae* pv. oryzae, the loci are not operons but clusters of genes driven individually by a conserved promoter. Loci 1, 2, 9, and 12 are followed by the first 40 bp of this sequence, and genes *tal2a*, *tal3a*, *tal3c*, and *tal8* are followed immediately by the first 28 bp of the sequence, providing evidence of a history of duplication and recombination with endpoints within the sequence. IS elements or fragments flank some of the genes, outside these putative promoter sequences. Curiously, the two pseudogenes that constitute locus 3 of *X. oryzae* pv. oryzae PXO99^A are each immediately preceded not by the 989-bp sequence but by a 109-bp sequence closely related to the *X. oryzae* pv. oryzicola BLS256 sequence carrying a similar 7-bp insertion, suggesting that this degenerate locus shares its origin more directly with the *X. oryzae* pv. oryzicola BLS256 loci, possibly due to a horizontal transfer event.

Genes required for Ax21 activity. The rice pattern recognition receptor XA21 confers resistance to *X. oryzae* pv. oryzae strains that secrete the protein Ax21 (*a*ctivator of *XA21*-mediated immunity) (45, 85). A 17-amino-acid peptide, called axY^{S}22, from the Ax21 N terminus is sufficient to trigger XA21-mediated immunity. Sulfation of tyrosine 22 is critical for this activity (45). Ax21 is encoded by the *ax21* gene. The genes *raxR* and *raxH* and the broadly conserved *phoP* and *phoQ* genes encode two-component regulatory systems that are predicted to control expression of *ax21* (10, 46). Sulfation of Ax21 is thought to be accomplished by the products of *raxP* and *raxQ*, which function in concert to produce 3′-phosphoadenosine 5′-phosphosulfate (PAPS), and by the product of *raxST*, which is similar to mammalian tyrosyl-sulfotransferases. RaxST is hypothesized to catalyze the transfer of the sulfuryl group from PAPS to the tyrosine(s) of Ax21 (31, 80). Finally, *raxA*, *raxB*, and *raxC* encode components of a type I system for Ax21 secretion (19). The *ax21* gene is conserved in all sequenced *Xanthomonas* spp., with over 89% identity, but *rax* gene content varies: *phoP*, *phoQ*, *raxC*, *raxH*, *raxP*, *raxQ*, and *raxR* are present in all strains, but *raxA*, *raxB*, and *raxST* are present only in the *X. oryzae* strains and in *X. axonopodis* pv. vesicatoria 85-10 (see Table S5 in the supplemental material). Furthermore, in *X. oryzae* pv. oryzae KACC10331, which lacks Ax21 activity, there is a frameshift mutation in *raxST* that results in a premature stop codon, and the *raxA* and *raxB* genes

TABLE 4. *X. oryzae* pv. oryzicola BLS256 genes encoding type III effectors and effector candidates and other type III-secreted proteins[a]

Gene	*X. oryzae* pv. oryzicola BLS256 locus tag no.	% GC content	Gene distribution[b]					
			X. axonopodis pv. citri 306	*X. axonopodis* pv. vesicatoria 85-10	*X. campestris* pv. campestris 8004	*X. campestris* pv. campestris ATCC 33913	*X. campestris* pv. campestris B100	*X. campestris* pv. raphani 756C
avrBs2	0146	63.4	+	+	+	+	+	/
hpa1	4432	59.2	+	+	+	+	+	+
hpaA	4450	63.9	+	+	+	+	+	+
xopAA	2511	65.5	/	+	/	/	/	/
xopAB	1662.a	61.3	/	/	/	/	/	/
xopAD	4489	66.9	+	Ψ	/	/	/	+
xopAE	4459	63.7	+	Ψ	/	/	/	/
xopAF	0445	55.6	/	/	/	/	/	/
*xopAH*Ψ	2510.a	63.0	/	/	+	+	+	/
xopAJ	4048	51.1	/	+	/	/	/	/
xopAK	3934	59.7	+	+	/	/	/	/
xopC2	1264	60.7	Ψ	Ψ	/	/	/	/
xopF1	4455	62.4	/	+	+	+	+	+
xopI	0821	64.1	+	+	/	/	/	/
xopK	3274	66.1	+	+	+	+	+	/
xopL	3279	60.1	+	+	+	+	+	/
xopN	0350	63.0	+	+	+	+	+	/
xopO	1098.a	53.1	/	+	/	/	/	/
xopP	1262	60.8	+	+	+	+	+	+
xopP	1263	60.0	+	+	+	+	+	+
xopQ	0099	67.7	+	+	+	+	+	/
xopR	4603	60.4	+	+	+	+	+	+
xopU	2513	66.9	/	/	/	/	/	/
xopV	0649	60.5	+	+	/	/	/	/
xopW	0487.a	53.1	/	/	/	/	/	/
xopX	0618	64.0	+	+	+	+	+	/
xopY	3486.a	66.4	/	/	/	/	/	/
xopZ1	2415	63.3	+	+	+	+	+	/

[a] TAL effector genes are not included in this table.
[b] Distribution among the complete *Xanthomonas* genomes and other plant pathogen genomes is shown: +, presence of a homolog; /, absence of a homolog; (+), weak or partial similarity; Ψ, pseudogene due to frameshift or premature stop codon.

are divergent in their C- and N-terminal coding regions relative to the other *X. oryzae* pv. oryzae strains. In contrast, in *X. oryzae* pv. oryzicola BLS256 all the *rax* genes are present and 95 to 99% identical to their counterparts in *X. oryzae* pv. oryzae PXO99^A. None of these variations, however, correlate with host (species) or tissue specificity.

TCSTS genes. Two-component signal transduction systems (TCSTSs), consisting of a sensor kinase and a response regulatory protein, are used broadly by prokaryotes to modulate gene expression in response to environmental cues (29, 87). TCSTS-related gene content across all the sequenced *Xanthomonas* genomes is generally similar (see Table S6 and Table S7 in the supplemental material). *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 possess 108 and 99 TCSTS-related genes, respectively. However, the *X. campestris* and *X. oryzae* strains are missing several, different sets of TCSTS genes that are present, with a few exceptions, in the *X. axonopodis* strains. Since the *X. axonopodis* strains are predicted to be most closely related to the ancestral strain (Fig. 2), this observation suggests that the lineages leading to the *X. campestris* and *X. oryzae* strains discarded ancestral TCSTS genes no longer necessary for survival as they adapted to their hosts and other features of their environments. Both the *X. campestris* and the *X. oryzae* strains also contain a small number of TCSTS genes specific to their lineages (XCR2315, XCR2343, XCR2344, XCR2730, and orthologs in the *X. campestris* strains and XOC3605, XOC3969, XOC3968, and orthologs in the *X. oryzae* strains). These may have been acquired during differentiation from the other *Xanthomonas* strains and could represent host-specific adaptations.

Specific functions of most *Xanthomonas* TCSTSs are yet to be discerned. Two TCSTSs that have been functionally characterized in *Xanthomonas* are conserved across all the sequenced strains: the PhoP and PhoQ system (corresponding to PXO_02836 and PXO_02837) is required in *X. oryzae* pv. oryzae PXO99^A for sensing low extracellular levels of Ca^{2+} and expressing virulence genes, including the T3S regulator *hrpG* (34), and the ColR and ColH system (corresponding to XC_1049 and XC_1050) in *X. campestris* pv. campestris 8004 also regulates *hrp* genes as well as genes for stress tolerance (106). Shared regulatory substrates of PhoPQ and ColRH suggest the possibility that differential integration and cross talk among shared TCSTSs may also play a role in host or tissue specificity.

Cyclic di-GMP signaling systems. The *rpf* gene cluster positively regulates the synthesis of extracellular enzymes, extracellular polysaccharide, biofilm dispersal, and virulence in *X. campestris* pv. campestris by governing synthesis and perception of the intercellular signal DSF (6, 22, 23, 84, 99, 101). Mutations in this cluster also lead to loss of virulence in *X. axonopodis* pv. citri, *X. oryzae* pv. oryzicola, and *X. oryzae* pv. oryzae (12, 82, 84, 90). RpfG is the response regulator impli-

TABLE 4—*Continued*

Gene distribution[b]							
X. oryzae pv. oryzae KACC10331	*X. oryzae* pv. oryzae MAFF 311018	*X. oryzae* pv. oryzae PXO99^A	*P. syringae* pv. tomato DC3000	*Pseudomonas syringae* pv. phaseolicola 1448A	*P. syringae* pv. syringae B728a	*Ralstonia solanacearum* GMI1000	*Erwinia carotovora* subsp. *atroseptica* SCRI104
+	+	+	/	/	/	/	/
+	+	+	/	/	/	/	/
+	+	+	/	/	/	/	/
+	+	+	/	/	/	/	/
+	+	+	/	/	/	/	/
+	+	+	/	+	/	+	/
+	+	+	/	/	/	(+)	/
/	/	/	+	+	+	/	/
Ψ	Ψ	Ψ	/	+	+	/	/
/	/	/	/	/	/	/	/
/	/	/	(+)	/	/	/	/
+	+	+	/	/	/	+	/
+	+	+	/	/	/	/	/
Ψ	Ψ	Ψ	/	/	/	/	/
+	+	+	/	/	/	/	/
+	+	+	/	/	/	+	/
+	+	+	/	(+)	/	/	/
/	/	/	(+)	+	/	/	/
+	+	+	/	/	/	(+)	/
+	+	+	/	/	/	+	/
+	+	+	+	+	/	+	/
+	+	+	/	/	/	/	/
Ψ	+	+	/	/	/	/	/
+	+	+	/	/	/	+	/
+	+	+	/	/	/	/	/
+	+	+	/	/	/	/	/
Ψ	+	+	/	/	/	/	/
+	+	+	Ψ	+	/	(+)	/

cated in signal transduction in response to DSF (6, 71, 84). RpfG contains an HD-GYP domain and functions as a cyclic di-GMP phosphodiesterase (71). Cyclic di-GMP is a second messenger that was originally identified to be an allosteric activator of cellulose synthesis but is now known to regulate a range of functions in diverse bacteria, including the virulence of several animal and plant pathogens (37, 70). Two other protein domains, named for the conserved amino acid motifs GGDEF and EAL, have been implicated in cyclic di-GMP synthesis and degradation, respectively. HD-GYP, GGDEF, and EAL domains are widely conserved and abundant in bacteria. They are found largely in combination with other signaling domains, suggesting that their activities in cyclic di-GMP turnover can be modulated by environmental cues. The genomes of *X. campestris* pv. campestris 8004 and *X. campestris* pv. campestris ATCC 33913 encode 3 proteins with an HD-GYP domain and 35 proteins with a GGDEF and/or EAL domain. *X. campestris* pv. campestris B100 lacks three genes encoding GGDEF domain proteins that are found in *X. campestris* pv. campestris 8004 and *X. campestris* pv. campestris ATCC 33913. The HD-GYP domain proteins are completely conserved across all the sequenced genomes. However, there are fewer genes encoding GGDEF and/or EAL domain proteins in *X. campestris* pv. raphani 756C than in the *X. campestris* pv. campestris strains and fewer still in *X. oryzae* pv. oryzicola BLS256 and again in the *X. oryzae* pv. oryzae strains. Specifically, of the inventory of genes in *X. campestris* pv. campestris 8004 and *X. campestris* pv. campestris ATCC 33913, one gene is absent in *X. campestris* pv. raphani 756C, five additional genes are absent in *X. oryzae* pv. oryzicola BLS256, and a further five genes (making 11 in total) are absent in the *X. oryzae* pv. oryzae strains. The *X. oryzae* pv. oryzae strains also have an additional gene encoding a GGDEF protein that is not found in the *X. campestris* pv. campestris strains. Functional genomic analysis has revealed that 13 genes encoding HD-GYP, GGDEF, and/or EAL domain proteins significantly contribute to virulence of *X. campestris* pv. campestris 8004 on Chinese radish (72). Intriguingly, 10 of these genes are retained in the *X. oryzae* pv. oryzae genomes (see Table S8 in the supplemental material). There are no clear correlations of GGDEF and EAL domain protein content with tissue or host specificity, but as suggested by differences in IS element content discussed above, it is possible that the reduced number of GGDEF and EAL proteins in the *X. oryzae* clade reflects greater stability and lower complexity in the environmental niche that the *X. oryzae* strains occupy.

TBDRs. TonB-dependent receptors (TBDRs) are outer membrane proteins of proteobacteria involved in iron uptake (64). TBDRs are also carbohydrate transporters (7, 58, 78) and may play a key role in sensing and exploitation of plant-derived carbon sources. In *X. campestris* pv. campestris ATCC 33913, they are involved in utilization of plant cell wall compounds and in virulence (7). Indeed, many of the *X. campestris* pv. campestris ATCC 33913 TBDR genes are linked to genes for carbohydrate uptake and catabolism, as well as signaling (7). TBDRs classically harbor an N-terminal plug domain linked to

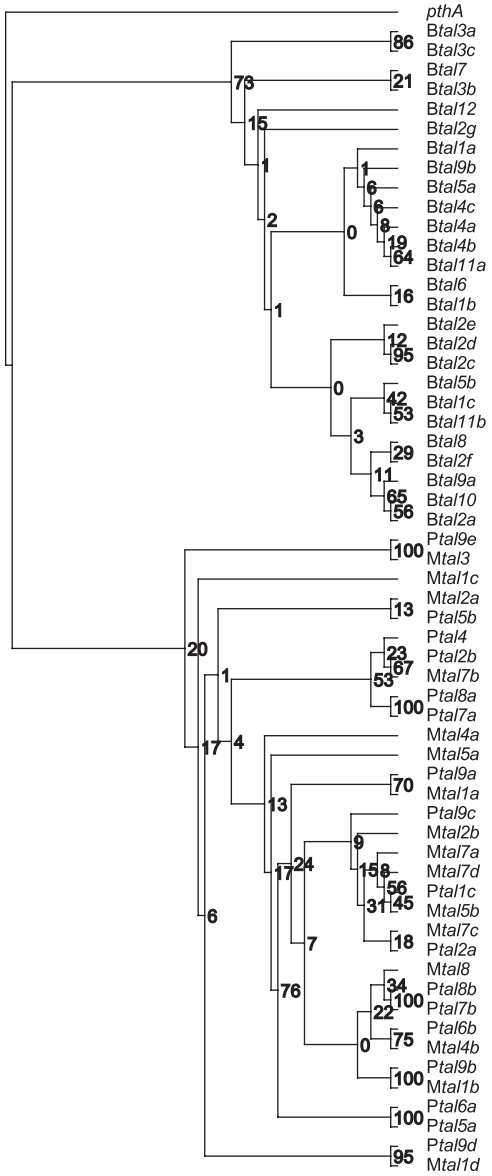

FIG. 4. Phylogenetic analysis of TAL effector genes in three *X. oryzae* strains. Displayed is the phylogeny for the DNA alignment of 60 TAL effector genes from *X. oryzae* pv. oryzicola BLS256, *X. oryzae* pv. oryzae MAFF 311018, and *X. oryzae* pv. oryzae PXO99^{A}. Gene names are preceded by B, M, or P, respectively, to denote their origin and are designated according to Salzberg et al. (74). *X. oryzae* pv. oryzae KACC10331 TAL effector genes were not included (see text). An unrooted tree was estimated by the Phyml program using 100 bootstrap replicates and the TPM1uf +I+G model. The tree is shown rooted at the outgroup, which is the *pthA* gene sequence from *X. axonopodis* pv. citri 306.

a C-terminal β-barrel domain, which is inserted in the bacterial outer membrane (14, 27, 61, 81). At their N termini, TBDRs display a signal peptide for transport across the inner membrane and at the N-terminal end of the plug domain a conserved stretch of amino acids called the TonB box, tLDXVXV in *X. campestris* pv. campestris ATCC 33913 (lowercase indicates a less well conserved residue and X represents any amino acid [7]), that interacts with TonB (61, 81).

As in *X. campestris* pv. campestris ATCC 33913, TBDRs are overrepresented in *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 relative to other Gram-negative bacteria, consistent with the hypothesis that they are an ancient and important class of proteins in the genus. The *X. campestris* pv. raphani 756C genome encodes 70 complete TBDRs and 2 truncated TBDRs, 1 missing the plug domain and the other missing the β-barrel domain (see Table S9 in the supplemental material). *X. oryzae* pv. oryzicola BLS256 contains 36 complete TBDR genes, 2 TBDR genes that lack a detectable TonB box but are otherwise complete, 5 pseudogenes missing the plug domain, and 14 pseudogenes missing the β-barrel domain (see Table S10 in the supplemental material). Three of the pseudogenes missing the plug domain encode a signal peptide and a TonB box, and two missing the β-barrel domain lack a detectable TonB box in the plug domain, indicating that combinations of internal and terminal deletions (or partial duplications) contributed to the large number of pseudogenes in *X. oryzae* pv. oryzicola BLS256. The relative reduction in the number of intact TBDRs in *X. oryzae* pv. oryzicola BLS256 is also reflected in the *X. oryzae* pv. oryzae genomes (see Fig. S4 in the supplemental material). Similar to the GGDEF and EAL domain proteins and consistent with the large numbers of IS elements, smaller numbers of TBDRs in the *X. oryzae* strains again may reflect adaptation to a closer and more stable association with the host.

The TBDR genes in all of the sequenced *Xanthomonas* genomes and two *Xylella fastidiosa* genomes were clustered on the basis of sequence similarity, and a phylogenetic tree was drawn using the neighbor-joining method (see Fig. S4 in the supplemental material). Of the 19 TBDR genes implicated in virulence in *X. campestris* pv. campestris ATCC 33913 (7), 10 are conserved across all the *Xanthomonas* genomes (clade A; see Fig. S4 in the supplemental material), suggesting a general role in virulence. One of these may be particularly important, as it is also conserved in the *X. fastidiosa* genomes (clade A*; see Fig. S4 in the supplemental material). Seven of the *X. campestris* pv. campestris ATCC 33913 genes cluster only with TBDRs from the other *X. campestris* strains and the *X. axonopodis* strains (clade B; see Fig. S4 in the supplemental material). These might be important specifically in pathogenicity for dicots. One of these is conserved in *X. fastidiosa* (clade B*; see Fig. S4 in the supplemental material). One of the *X. campestris* pv. campestris ATCC 33913 TBDRs implicated in virulence clusters with genes in only the *X. campestris* strains (clade C; see Fig. S4 in the supplemental material) and may be important to association with brassicas. The remaining *X. campestris* pv. campestris ATCC 33913 TBDR gene resides in a cluster missing representatives only from *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzae KACC10331 (clade A′; see Fig. S4 in the supplemental material). No *X. oryzae*-specific TBDR gene clusters that would reflect adaptations to monocots or to rice specifically were found, though one clade consists of seven *X. oryzae* pv. oryzicola BLS256 genes and one *X. campestris* pv. raphani 756C gene (clade D; see Fig. S4 in the supplemental material), suggesting amplification in *X. oryzae* pv. oryzicola BLS256 of a TBDR gene common to both strains. No clusters comprise genes only from vascular pathogens or only from nonvascular pathogens.

Adhesins. The *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 genomes encode several outer membrane-associated, fimbrial and nonfimbrial adhesin proteins that are similar to their counterparts in animal pathogenic bacteria and in other plant-pathogenic bacteria (see Table S11 in the supplemental material). One of these nonfimbrial proteins is an ortholog of the previously described XadA (*Xanthomonas* adhesin-like protein A)/XadA1 protein of *X. oryzae* pv. oryzae (68). XadA1 and its paralog, XadB, play an important role in the early stages of rice infection by *X. oryzae* pv. oryzae (17). All sequenced *Xanthomonas* strains encode XadA1, but only the *X. oryzae* strains encode XadB. Another paralog, called XadA2, is present only in *X. axonopodis* pv. citri 306 and *Xanthomonas axonopodis* pv. vesicatoria 85-10 but not in the other sequenced xanthomonads. In addition to XadA1/XadB, the *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 genomes encode another nonfimbrial adhesin-like protein, YapH (*Yersinia* autotransporter-like protein H). Orthologs of YapH are also present in the *X. campestris* pv. campestris and *X. oryzae* pv. oryzae genomes.

A characteristic feature of many adhesins, such as the XadA/XadB and YapH protein families, is the presence of internal amino acid repeats. Differences in the numbers of these repeats and sequence variations in these repeats are observed for the various adhesins of *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256, compared with each other and with the orthologs from the vascular counterparts of these strains (data not shown). The observed differences in the repertoire of adhesins, particularly the sequence variations among them, might contribute to host and tissue specificity. A similar adhesion-based host specificity model was also recently proposed by Mhedbi-Hajri and coworkers, who found a correlation of adhesin repertoires with *Xanthomonas* host specificity as well as evidence of rapid evolution among adhesin genes (54).

Genome-wide identification of candidate tissue and host specificity genes. To identify other genes that may be involved in host specificity or tissue specificity, an *ab initio*, comprehensive comparison across genomes was carried out. All nontransposon-related protein-coding genes were grouped into clusters of homologs, and for each cluster the most restrictive category based on the strains represented in the cluster was identified by using the two 2-way classifications of strains as monocot (M) versus dicot (D) with respect to the host(s) they infect and vascular (V) versus nonvascular (N) with regard to their mode of infection (Table 1). A cluster of homologs was considered to be specific to a category if all cluster members belong to this category, although a homolog may be missing in one or more strains of the category, and none are found elsewhere (see Fig. S5A in the supplemental material). A cluster of homologs was considered to be essential to a category if all strains in the category are represented in the cluster, although there may be additional homologs in other strains (see Fig. S5B in the supplemental material). The intersection of the specific and essential sets defines clusters for which all strains in the category are represented and there are no additional homologs in other strains (Fig. 5). A total of 2,508 clusters are shared in all 10 bacterial strains analyzed; 333 clusters occur only in D strains, 90 occur only in M strains, only 1 cluster appears to be unique to V strains, and none is unique to N strains. Clusters specific and essential to the ND, VD, NM, and VM categories were also found, though they likely suffer bias due to the small numbers and close relatedness of strains within these categories. Because the gene clusters represent putative orthologs, their numbers include cases in which different copy numbers of a gene in different strains result in multiple clusters, some of which are specific to the strains with higher copy numbers of the gene. For example, the 333 D-specific clusters may include a cluster of genes homologous to a category-nonspecific cluster because the gene is present in two (sufficiently divergent) copies in the D group strains and present in one copy in all the rest. Also, since the numbers are based on annotated protein-coding genes, there might be errors due to missed genes in some genomes. To refine the sets of D-, M-, and V-specific and essential gene clusters and to minimize possible errors due to unannotated genes, for each cluster a representative protein sequence was compared to every cluster in the complementary category using BLASTp, and the coding sequence of that protein was compared to the genome sequence of every strain in the complementary category by using tBLASTn, with an E-value threshold of 1e−10 in each case. Only clusters for which the representative protein showed no sequence matches in the complementary category were deemed truly category specific and essential. This refinement resulted in 172 D-specific gene clusters, 54 M clusters, and the 1 V cluster (representative sequences of the 2,508 shared clusters, the refined D-, M-, and V-specific and essential clusters, and the ND-, VD-, NM-, and VM-specific and essential clusters, as well as the complete lists of genes in the D-, M-, and V-specific and essential cluster sets, are given in Data Files S1 through S9 in the supplemental material).

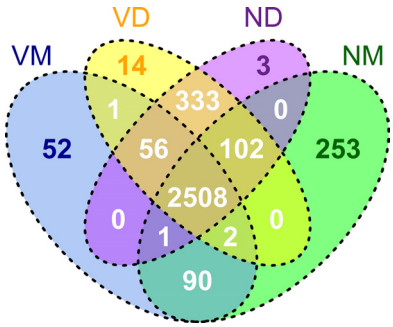

FIG. 5. Distribution of homologous protein clusters specific and essential to different categories of sequenced *Xanthomonas* strains. The cluster sets are not refined; i.e., the numbers are based on annotated genes and include clusters formed by distinct sets of paralogs.

The refined sets of D-, M-, and V-specific and essential clusters were grouped according to functional category. The single V-specific and essential cluster is annotated as a hypothetical protein gene, unfortunately providing no basis for a functional hypothesis. Among the D- and M-specific and essential clusters (see Fig. S6 in the supplemental material), the most highly represented, accounting for a third of the dicot pathogen-specific and essential genes and two-thirds of the monocot pathogen-specific and essential genes, are hypothetical protein genes or genes with no functional category assigned and putative secreted protein genes. Genes for intracellular trafficking and secretion are also highly represented.

Consistent with the distributions of TBDR genes, genes for carbohydrate transport and metabolism are the next largest group for the dicot pathogen clusters, but there is only one member of this functional category unique to the monocot pathogen clusters. Less well represented categories form a diverse assortment. These findings suggest a number of possibilities. Type III effectors or other secreted proteins critical to host or tissue specificity might be among the hypothetical proteins or putative secreted proteins. Differences in carbohydrate sensing, uptake, or catabolism may play key roles. Genes among some of the poorly represented functional categories, including secondary metabolite synthesis, transport, or catabolism, lipid transport and catabolism, and defense mechanisms, might also contribute. Overall, the large number of genes that distinguish dicot from monocot pathogens suggests the possibility that complex sets of adaptations determine host specificity. In contrast, minimal distinguishing differences in gene content between the vascular and nonvascular pathogens indicate that subtler differences, at the level of amino acid or noncoding nucleotide polymorphisms, may be the determinants of tissue specificity.

Concluding remarks. Lu and colleagues (52) found no differences correlating to host or tissue specificity in the gene content of several pathogenesis-associated gene clusters across most of the genomes examined here. The authors concluded that determinative differences likely reside among the secretory and regulatory substrates of those clusters, among polymorphisms within genes or regulatory sequences in the clusters, or in unrelated genes elsewhere in the genome. The complementary, genome-wide comparative analyses presented here, enabled by the completion of the *X. campestris* pv. raphani 756C and *X. oryzae* pv. oryzicola BLS256 genomes, further support this conclusion and indeed point to type III effector genes, cyclic di-GMP signaling protein genes, two-component signaling system genes, TBDRs, and several other genes potentially related to interactions of *Xanthomonas* spp. with plants to be candidate determinants of host and tissue specificity. Comparisons across a broader set of genome sequences may help narrow this list. Recently acquired draft sequences of several *Xanthomonas* genomes will be useful in this regard (40a, 65, 88, 94). At the same time, future improvements of computational methods for large-scale comparisons may reveal polymorphisms within genes and noncoding sequences that will contribute to refined models for host and tissue specificity. Ultimately, in-depth, large-scale functional characterization of candidates will be necessary to pinpoint those that play key roles.

ACKNOWLEDGMENTS

We thank Anne Alvarez for assistance in selecting *X. campestris* pv. raphani 756C as a representative strain, Sophien Kamoun for providing that strain, Fadi Towfic and the BCB Lab of Iowa State University for assistance with TAL effector phylogenetic analysis, and Jonas Gaiarsa for assistance with IS element analysis.

This work was supported by the USDA-NSF Microbial Genome Sequencing Program (award CSREES 20043560015022).

REFERENCES

1. **Abascal, F., R. Zardoya, and D. Posada.** 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21:**2104–2105.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Alvarez, A. M., A. A. Benedict, and C. Y. Mizumoto.** 1985. Identification of xanthomonads and grouping of strains of *Xanthomonas campestris* pv. campestris with monoclonal antibodies. Phytopathology **75:**722–728.
4. **Alvarez, A. M., A. A. Benedict, C. Y. Mizumoto, J. E. Hunter, and D. W. Gabriel.** 1994. Serological, pathological, and genetic diversity among strains of *Xanthomonas campestris* infecting crucifers. Phytopathology **84:**1449–1457.
5. **Astua-Monge, G., et al.** 2000. Resistance of tomato and pepper to T3 strains of *Xanthomonas campestris* pv. vesicatoria is specified by a plant-inducible avirulence gene. Mol. Plant Microbe Interact. **13:**911–921.
6. **Barber, C. E., et al.** 1997. A novel regulatory system required for pathogenicity of *Xanthomonas campestris* is mediated by a small diffusible signal molecule. Mol. Microbiol. **24:**555–566.
7. **Blanvillain, S., et al.** 2007. Plant carbohydrate scavenging through *tonB*-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. PLoS One **2:**e224.
8. **Boch, J., and U. Bonas.** 2010. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. Annu. Rev. Phytopathol. **48:**419–436.
9. **Bogdanove, A. J., S. Schornack, and T. Lahaye.** 2010. TAL effectors: finding plant genes for disease and defense. Curr. Opin. Plant Biol. **13:**394–401.
10. **Burdman, S., Y. Shen, S. W. Lee, Q. Xue, and P. Ronald.** 2004. RaxH/RaxR: a two-component regulatory system in *Xanthomonas oryzae* pv. oryzae required for AvrXa21 activity. Mol. Plant Microbe Interact. **17:**602–612.
11. **Camacho, C., et al.** 2009. BLAST+: architecture and applications. BMC Bioinformatics **10:**421.
12. **Chatterjee, S., and R. V. Sonti.** 2002. *rpfF* mutants of *Xanthomonas oryzae* pv. oryzae are deficient for virulence and growth under low iron conditions. Mol. Plant Microbe Interact. **15:**463–471.
13. **Chen, H., and P. C. Boutros.** 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics **12:**35.
14. **Chimento, D. P., A. K. Mohanty, R. J. Kadner, and M. C. Wiener.** 2003. Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. Nat. Struct. Biol. **10:**394–401.
15. **Crossman, L. C., et al.** 2008. The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. Genome Biol. **9:**R74.
16. **Darling, A. E., B. Mau, and N. T. Perna.** 2010. progressiveMAUVE: multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5:**e11147.
17. **Das, A., N. Rangaraj, and R. V. Sonti.** 2009. Multiple adhesin-like functions of *Xanthomonas oryzae* pv. oryzae are involved in promoting leaf attachment, entry, and virulence on rice. Mol. Plant Microbe Interact. **22:**73–85.
18. **da Silva, A. C., et al.** 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature **417:**459–463.
19. **da Silva, F. G., et al.** 2004. Bacterial genes involved in type I secretion and sulfation are required to elicit the rice Xa21-mediated innate immune response. Mol. Plant Microbe. Interact. **17:**593–601.
20. **Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679.
21. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
22. **Dow, J. M., et al.** 2003. Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. Proc. Natl. Acad. Sci. U. S. A. **100:**10995–11000.
23. **Dow, J. M., J. X. Feng, C. E. Barber, J. L. Tang, and M. J. Daniels.** 2000. Novel genes involved in the regulation of pathogenicity factor production within the *rpf* gene cluster of *Xanthomonas campestris*. Microbiology **146:**885–891.
24. **Fargier, E., and C. Manceau.** 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.
25. **Fargier, E., M. F. Saux, and C. Manceau.** 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.
26. **Fenselau, S., and U. Bonas.** 1995. Sequence and expression analysis of the *hrpB* pathogenicity operon of *Xanthomonas campestris* pv. vesicatoria which encodes eight proteins with similarity to components of the Hrp, Ysc, Spa, and Fli secretion systems. Mol. Plant Microbe Interact. **8:**845–854.
27. **Ferguson, A. D., et al.** 2002. Structural basis of gating by the outer membrane transporter FecA. Science **295:**1715–1719.
28. **Furutani, A., et al.** 2006. Identification of novel HrpXo regulons preceded by two cis-acting elements, a plant-inducible promoter box and a −10 box-like sequence, from the genome database of *Xanthomonas oryzae* pv. oryzae. FEMS Microbiol. Lett. **259:**133–141.
29. **Galperin, M. Y.** 2004. Bacterial signal transduction network in a genomic perspective. Environ. Microbiol. **6:**552–567.
30. **Guindon, S., and O. Gascuel.** 2003. A simple, fast, and accurate algorithm

to estimate large phylogenies by maximum likelihood. Syst. Biol. **52:**696–704.
31. **Han, S. W., S. W. Lee, and P. C. Ronald.** 2011. Secretion, modification, and regulation of Ax21. Curr. Opin. Microbiol. **14:**62–67.
32. **Hayward, A. C.** 1993. The hosts of *Xanthomonas*, p. 1–119. *In* J. G. Swings and E. L. Civerolo (ed.), *Xanthomonas*. Chapman and Hall, London, United Kingdom.
33. **He, P., et al.** 2004. Activation of a COI1-dependent pathway in *Arabidopsis* by *Pseudomonas syringae* type III effectors and coronatine. Plant J. **37:**589–602.
34. **Heidelberg, J. F., et al.** 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406:**477–483.
35. **Jain, R., M. C. Rivera, and J. A. Lake.** 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. U. S. A. **96:**3801–3806.
36. **Jamir, Y., et al.** 2004. Identification of *Pseudomonas syringae* type III effectors that can suppress programmed cell death in plants and yeast. Plant J. **37:**554–565.
37. **Jenal, U., and J. Malone.** 2006. Mechanisms of cyclic-di-GMP signaling in bacteria. Annu. Rev. Genet. **40:**385–407.
38. **Kamoun, S., and C. I. Kado.** 1990. Phenotypic switching affecting chemotaxis, xanthan production, and virulence in *Xanthomonas campestris*. Appl. Environ. Microbiol. **56:**3855–3860.
39. **Kamoun, S., H. Kamdar, E. Tola, and C. I. Kado.** 1992. Incompatible interactions between crucifers and *Xanthomonas campestris* involve a vascular hypersensitive response: role of the *hrpX* locus. Mol. Plant Microbe. Interact. **5:**22–33.
40. **Katoh, K., K. Kuma, H. Toh, and T. Miyata.** 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. **33:**511–518.
40a. **Kimbrel, J. A., S. A. Givan, T. N. Temple, K. B. Johnson, and J. H. Chang.** 2011. Genome sequencing and comparative analysis of the carrot bacterial blight pathogen, *Xanthomonas hortorum* pv. carotae M081, for insights into pathogenicity and applications in molecular diagnostics. Mol. Plant Pathol. **12:**580–594.
41. **Kimura, M.** 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. U. S. A. **78:**454–458.
42. **Kingsford, C. L., K. Ayanbule, and S. L. Salzberg.** 2007. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biol. **8:**R22.
43. **Koebnik, R., A. Kruger, F. Thieme, A. Urban, and U. Bonas.** 2006. Specific binding of the *Xanthomonas campestris* pv. vesicatoria AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. J. Bacteriol. **188:**7652–7660.
44. **Lee, B. M., et al.** 2005. The genome sequence of *Xanthomonas oryzae* pathovar oryzae KACC10331, the bacterial blight pathogen of rice. Nucleic Acids Res. **33:**577–586.
45. **Lee, S. W., et al.** 2009. A type I-secreted, sulfated peptide triggers XA21-mediated innate immunity. Science **326:**850–853.
46. **Lee, S. W., et al.** 2008. The *Xanthomonas oryzae* pv. oryzae PhoPQ two-component system is required for AvrXA21 activity, *hrpG* expression, and virulence. J. Bacteriol. **190:**2183–2197.
47. **Leyns, F., M. De Cleene, J. G. Swings, and J. De Ley.** 1984. The host range of the genus Xanthomonas. Bot. Rev. **50:**308–356.
48. **Li, L., C. J. Stoeckert, Jr., and D. S. Roos.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13:**2178–2189.
49. **Li, X., et al.** 2005. Flagellin induces innate immunity in nonhost interactions that is suppressed by *Pseudomonas syringae* effectors. Proc. Natl. Acad. Sci. U. S. A. **102:**12990–12995.
50. **Lobry, J. R.** 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. Mol. Biol. Evol. **13:**660–665.
51. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of tRNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
52. **Lu, H., et al.** 2008. Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. PLoS One **3:**e3828.
53. **Makino, S., A. Sugio, F. White, and A. J. Bogdanove.** 2006. Inhibition of resistance gene-mediated defense in rice by *Xanthomonas oryzae* pv. oryzicola. Mol. Plant Microbe Interact. **19:**240–249.
54. **Mhedbi-Hajri, N., et al.** 2011. Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. BMC Evol. Biol. **11:**67.
55. **Monteiro-Vitorello, C. B., et al.** 2005. *Xylella* and *Xanthomonas* Mobil'omics. OMICS **9:**146–159.
56. **Moscou, M. J., and A. J. Bogdanove.** 2009. A simple cipher governs DNA recognition by TAL effectors. Science **326:**1501.
57. **Myers, E. W., et al.** 2000. A whole-genome assembly of *Drosophila*. Science **287:**2196–2204.
58. **Neugebauer, H., et al.** 2005. ExbBD-dependent transport of maltodextrins through the novel MalA protein across the outer membrane of *Caulobacter crescentus*. J. Bacteriol. **187:**8300–8311.
59. **Ochiai, H., Y. Inoue, M. Takeya, A. Sasaki, and H. Kaku.** 2005. Genome sequence of *Xanthomonas oryzae* pv. oryzae suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. Jpn. Agric. Res. Q. **39:**275–287.
60. **Parker, J. E., C. E. Barber, M. J. Fan, and M. J. Daniels.** 1993. Interaction of *Xanthomonas campestris* with *Arabidopsis thaliana*: characterization of a gene from *X. c*. pv. raphani that confers avirulence to most *A. thaliana* accessions. Mol. Plant Microbe Interact. **6:**216–224.
61. **Pawelek, P. D., et al.** 2006. Structure of TonB in complex with FhuA, E. coli outer membrane receptor. Science **312:**1399–1402.
62. **Pieretti, I., et al.** 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.
63. **Posada, D.** 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. **25:**1253–1256.
64. **Postle, K., and R. J. Kadner.** 2003. Touch and go: tying TonB to transport. Mol. Microbiol. **49:**869–882.
65. **Potnis, N., et al.** 2011. Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. BMC Genomics **12:**146.
66. **Qian, W., et al.** 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. campestris. Genome Res. **15:**757–767.
67. **Rademaker, J. L. W., et al.** 2005. A comprehensive species to strain taxonomic framework for *Xanthomonas*. Phytopathology **95:**1098–1111.
68. **Ray, S. K., R. Rajeshwari, Y. Sharma, and R. V. Sonti.** 2002. A high-molecular-weight outer membrane protein of *Xanthomonas oryzae* pv. oryzae exhibits similarity to non-fimbrial adhesins of animal pathogenic bacteria and is required for optimum virulence. Mol. Microbiol. **46:**637–647.
69. **Raymundo, A. K., and J. E. Leach.** 1993. Amplification and sequence analysis of the upstream region of the *hrpXo* gene homolog in *Xanthomonas oryzae* pv. oryzicola. Philipp. J. Biotechnol. **4:**29–38.
70. **Romling, U., M. Gomelsky, and M. Y. Galperin.** 2005. C-di-GMP: the dawning of a novel bacterial signalling system. Mol. Microbiol. **57:**629–639.
71. **Ryan, R. P., et al.** 2006. Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. Proc. Natl. Acad. Sci. U. S. A. **103:**6712–6717.
72. **Ryan, R. P., et al.** 2007. Cyclic di-GMP signalling in the virulence and environmental adaptation of *Xanthomonas campestris*. Mol. Microbiol. **63:**429–442.
73. **Salzberg, S. L., A. J. Salzberg, A. R. Kerlavage, and J. F. Tomb.** 1998. Skewed oligomers and origins of replication. Gene **217:**57–67.
74. **Salzberg, S. L., et al.** 2008. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. oryzae PXO99A. BMC Genomics **9:**204.
75. **Sambrook, J., E. F. Fritsch, and T. A. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
76. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. **74:**5463–5467.
77. **Schatz, M. C., A. M. Phillippy, B. Shneiderman, and S. L. Salzberg.** 2007. Hawkeye: an interactive visual analytics tool for genome assemblies. Genome Biol. **8:**R34.
78. **Schauer, K., D. A. Rodionov, and H. de Reuse.** 2008. New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? Trends Biochem. Sci. **33:**330–338.
79. **Schornack, S., K. Peter, U. Bonas, and T. Lahaye.** 2005. Expression levels of *avrBs3*-like genes affect recognition specificity in tomato Bs4- but not in pepper Bs3-mediated perception. Mol. Plant Microbe Interact. **18:**1215–1225.
80. **Shen, Y., P. Sharma, F. Goes da Silva, and P. Ronald.** 2002. The *Xanthomonas oryzae* pv. oryzae *raxP* and *raxQ* genes encode an ATP sulphurylase and adenosine-5′-phosphosulphate kinase that are required for AvrXa21 avirulence activity. Mol. Microbiol. **44:**37–48.
81. **Shultis, D. D., M. D. Purdy, C. N. Banchs, and M. C. Wiener.** 2006. Outer membrane active transport: structure of the BtuB:TonB complex. Science **312:**1396–1399.
82. **Siciliano, F., et al.** 2006. Analysis of the molecular basis of *Xanthomonas axonopodis* pv. *citri* pathogenesis in *Citrus limon.* Electron. J. Biotechnol. **9**. doi:10.2225/vol9-issue3-fulltext-20.
83. **Simmons, M. P., H. Ochoterena, and J. V. Freudenstein.** 2002. Conflict between amino acid and nucleotide characters. Cladistics **18:**200–206.
84. **Slater, H., A. Alvarez-Morales, C. E. Barber, M. J. Daniels, and J. M. Dow.** 2000. A two-component system involving an HD-GYP domain protein links cell-cell signalling to pathogenicity gene expression in *Xanthomonas campestris*. Mol. Microbiol. **38:**986–1003.
85. **Song, W. Y., et al.** 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science **270:**1804–1806.
86. **Starr, M. P.** 1981. The genus *Xanthomonas*, p. 742–763. *In* M. P. Starr, H. Stolp, H. G. Truper, A. Balows, and H. G. Schlegel (ed.), The prokaryotes, vol. 1. Springer Verlag, Berlin, Germany.
87. **Stock, A. M., V. L. Robinson, and P. N. Goudreau.** 2000. Two-component signal transduction. Annu. Rev. Biochem. **69:**183–215.

88. **Studholme, D. J., et al.** 2010. Genome-wide sequencing data reveals virulence factors implicated in banana *Xanthomonas* wilt. FEMS Microbiol. Lett. **310:**182–192.
89. **Tamura, K., et al.** 4 May 2011, posting date. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. doi:10.1093/molbev/msr121.
90. **Tang, J. L., et al.** 1996. Cloning and characterization of the *rpfC* gene of *Xanthomonas oryzae* pv. oryzae: involvement in exopolysaccharide production and virulence to rice. Mol. Plant Microbe Interact. **9:**664–666.
91. **Tang, X., Y. Xiao, and J. M. Zhou.** 2006. Regulation of the type III secretion system in phytopathogenic bacteria. Mol. Plant Microbe Interact. **19:**1159–1166.
92. **Thieme, F., et al.** 2005. Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. vesicatoria revealed by the complete genome sequence. J. Bacteriol. **187:**7254–7266.
93. **Thieme, F., et al.** 2007. New type III effectors from *Xanthomonas campestris* pv. vesicatoria trigger plant reactions dependent on a conserved N-myristoylation motif. Mol. Plant Microbe Interact. **20:**1250–1261.
94. **Triplett, L. R., et al.** 2011. Genomic analysis of *Xanthomonas oryzae* from US rice reveals substantial divergence from known *X. oryzae* pathovars. Appl. Environ. Microbiol. **77:**3930–3937.
95. **Tsuge, S., et al.** 2005. Effects on promoter activity of base substitutions in the cis-acting regulatory element of *HrpXo* regulons in *Xanthomonas oryzae* pv. oryzae. J. Bacteriol. **187:**2308–2314.
96. **Vauterin, L., B. Hoste, K. Kersters, and J. Swings.** 1995. Reclassification of *Xanthomonas*. Int. J. Syst. Bacteriol. **45:**472–489.
97. **Vicente, J. G., B. Everett, and S. J. Roberts.** 2006. Identification of isolates that cause a leaf spot disease of brassicas as *Xanthomonas campestris* pv. raphani and pathogenic and genetic comparison with related pathovars. Phytopathology **96:**735–745.
98. **Vorhölter, F. J., et al.** 2008. The genome of *Xanthomonas campestris* pv. campestris B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. J. Biotechnol. **134:**33–45.
99. **Wang, L. H., et al.** 2004. A bacterial cell-cell communication signal with cross-kingdom structural analogues. Mol. Microbiol. **51:**903–912.
100. **White, F. F., and B. Yang.** 2009. Host and pathogen factors controlling the rice-*Xanthomonas oryzae* interaction. Plant Physiol. **150:**1677–1686.
101. **Wilson, T. J. G., et al.** 1998. The *rpfA* gene of *Xanthomonas campestris* pathovar *campestris*, which is involved in the regulation of pathogenicity factor production, encodes an aconitase. Mol. Microbiol. **28:**961–970.
102. **Wu, D., et al.** 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature **462:**1056–1060.
103. **Wu, M., and J. A. Eisen.** 2008. A simple, fast, and accurate method of phylogenomic inference. Genome Biol. **9:**R151.
104. **Xu, R. Q., et al.** 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.
105. **Yang, B., and F. F. White.** 2004. Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. Mol. Plant Microbe Interact. **17:**1192–1200.
106. **Zhang, S. S., et al.** 2008. A putative colR(XC1049)-colS(XC1050) two-component signal transduction system in *Xanthomonas campestris* positively regulates *hrpC* and *hrpE* operons and is involved in virulence, the hypersensitive response and tolerance to various stresses. Res. Microbiol. **159:**569–578.
107. **Zhao, B., et al.** 2004. The *avrRxo1* gene from the rice pathogen *Xanthamonas oryzae* pv. *oryzicola* confers a nonhost defense reaction on maize with resistance gene *Rxo1*. Mol. Plant Microbe Interact. **17:**771–779.
108. **Zhao, B., et al.** 2005. A maize resistance gene functions against bacterial streak disease in rice. Proc. Natl. Acad. Sci. U. S. A. **102:**15383–15388.