

Genome Sequence of *Xanthomonas campestris* pv. *campestris* Strain Xca5

Stéphanie Bolot,^{a,b} Endrick Guy,^{a,b} Sébastien Carrere,^{a,b} Valérie Barbe,^c Matthieu Arlat,^{a,b,d} Laurent D. Noël^{a,b}

National Institute for Agricultural Research (INRA), Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR 441, Castanet-Tolosan, France^a; Centre National de la Recherche Scientifique (CNRS), LIPM, UMR 2594, Castanet-Tolosan, France^b; Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), IG, Genoscope, Évry, France^c; Université de Toulouse, Université Paul Sabatier, Toulouse, France^d

S.B. and E.G. contributed equally to this article.

An annotated high-quality draft genome sequence for *Xanthomonas campestris* pv. *campestris* race 1 strain Xca5 (originally described as *X. campestris* pv. *armoraciae*), the causal agent of black rot on *Brassicaceae* plants, has been determined. This genome sequence is a valuable resource for comparative genomics within the *campestris* pathovar.

Received 26 November 2012 Accepted 5 December 2012 Published 7 February 2013

Citation Bolot S, Guy E, Carrere S, Barbe V, Arlat M, Noël LD. 2013. Genome sequence of *Xanthomonas campestris* pv. *campestris* strain Xca5. *Genome Announc.* 1(1):e00032-12. doi:10.1128/genomeA.00032-12.

Copyright © 2013 Bolot et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Laurent D. Noël, laurent.noel@toulouse.inra.fr.

Xanthomonas campestris pv. *campestris* is the causal agent of black rot on a wide range of *Brassicaceae* plants, including vegetable crops, such as cabbages, ornamental crucifers, and weeds, as well as the model plant *Arabidopsis thaliana*. The bacteria are seed transmitted and enter the plant vascular system by the hydathodes, causing V-shaped lesions, vein blackening, and leaf tissue necrosis. *X. campestris* pv. *campestris* strains have been sorted into 9 physiological races based on their interactions with diverse *Brassicaceae* plants. *X. campestris* pv. *campestris* genome sequences for races 3 and 9 are available (1, 2). Yet, no strains from *X. campestris* pv. *campestris* race 1 have had their genome sequences determined, despite belonging to probably the most represented *X. campestris* pv. *campestris* race worldwide (3).

Xca5 is an American race 1 *X. campestris* pv. *campestris* strain originally classified as pathovar *armoraciae* by M. Daniels (Sainsbury Laboratory, Norwich, United Kingdom). Yet, pathogenicity tests (3), as well as multilocus sequence analyses (data not shown), clearly identified this strain as a bona fide *X. campestris* pv. *campestris* strain. Xca5 genome shotgun sequencing was performed on a GAIIx Illumina platform. A total of 28,569,308 76-bp paired-end reads corresponding to 4,342,534,816 bp and 868-fold coverage were obtained. Genome assembly was performed using a combination of Short Oligonucleotide Analysis Package (SOAP) *de novo* (4) and Velvet (5) assemblers and yielded 130 contigs that were >500 bp and had an N_{50} of 121,265 bp. The average contig size was 38,385 bp, and the largest was 447,475 bp long, for a total genome size of 4,990,056 bp. One hundred twenty of those contigs were further organized into 4 pseudomolecules. The largest one corresponds to the chromosome (4,905,337 bp; 65.2% G+C content) based on *X. campestris* pv. *campestris* 8004 chromosomal organization. The remaining pseudomolecules match known *Xanthomonas* plasmid sequences. Due to the highly repetitive nature of the structures of known transcriptional activator-like (TAL) protein-encoding loci *hax2* (plasmid borne), *hax3*, and *hax4*, their sequences (accession no. AY993937, AY993938, and

AY993939, respectively) could not be assembled automatically (6) and were manually added to the final submission. In total, the genome is composed of 17 pseudomolecules/contigs totaling 5,001,025 bp. Annotation transfer was performed using the Rapid Annotation Transfer Tool (RATT) (7) with *X. campestris* pv. *campestris* B100, 8004, and ATCC 33913 as references. *De novo* annotation was performed on remaining areas using FramedD (8) and was inspected manually. We identified 4,592 coding sequences (CDSs), 51 tRNA genes, and 2 rRNA genes.

Phylogenetic analyses based on the core genome sequence shared with the 3 available *X. campestris* pv. *campestris* reference genome sequences and the genome sequence of *X. campestris* pv. *raphani* strain 756C were performed using Unus (9). These analyses showed that, indeed, Xca5 is most closely related to *X. campestris* pv. *campestris* strains 8004 and ATCC 33913. Using OrthoMCL (percent match cutoff, 80; blast parameter, F = false) (10), Xca5 was found to share 3,711 CDSs with the 3 *X. campestris* pv. *campestris* reference strains. Despite the availability of the 3 *X. campestris* pv. *campestris* reference genome sequences, this genome sequence represents the first *X. campestris* pv. *campestris* strain with TAL protein genes and plasmids, suggesting that much more genomic diversity might be expected at the intraspecific level than is anticipated from the available *X. campestris* pv. *campestris* genome sequences (1, 2, 11).

Nucleotide sequence accession numbers. The Whole Genome Shotgun project has been deposited at EMBL under the accession no. CAOR01000001 to CAOR01000130.

ACKNOWLEDGMENTS

This study was funded by the Laboratoire d'Excellence (LABEX) TULIP (ANR-10-LABX-41), Genoscope grant, AP10/11-N° 28_Rott, a Ph.D. grant from the French Ministry of National Education and Research and French Guiana to E.G., and an Agence Nationale de la Recherche–Jeunes Chercheurs grant (Xopaque ANR-10-JCJC-1703-01) to L.D.N.

REFERENCES

1. da Silva AC, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys MA, Almeida NF, Alves LM, do Amaral AM, Bertolini MC, Camargo LE, Camarotte G, Cannavan F, Cardozo J, Chambergo F, Ciapina LP, Cicarelli RM, Coutinho LL, Cursino-Santos JR, El-Dorry H, Faria JB, Ferreira AJ, Ferreira RC, Ferro MI, Formighieri EF, Franco MC, Greggio CC, Gruber A, Katsuyama AM, Kishi LT, Leite RP, Lemos EG, Lemos MV, Locali EC, Machado MA, Madeira AM, Martinez-Rossi NM, Martins EC, Meidanis J, Menck CF, Miyaki CY, Moon DH, Moreira LM, Novo MT, Okura VK, Oliveira MC, Oliveira VR, Pereira HA, Rossi A, Sena JA, Silva C, de Souza RF, Spinola LA, Takita MA, Tamura RE, Teixeira EC, Tezza RI, Trindade dos Santos M, Truffi D, Tsai SM, White FF, Setubal JC, Kitajima JP. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417:459–463.
2. Qian W, Jia Y, Ren SX, He YQ, Feng JX, Lu LF, Sun Q, Ying G, Tang DJ, Tang H, Wu W, Hao P, Wang L, Jiang BL, Zeng S, Gu WY, Lu G, Rong L, Tian Y, Yao Z, Fu G, Chen B, Fang R, Qiang B, Chen Z, Zhao GP, Tang JL, He C. 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Res.* 15:757–767.
3. Vicente JG, Conway J, Roberts SJ, Taylor JD. 2001. Identification and origin of *Xanthomonas campestris* pv. *campestris* races and related pathogens. *Phytopathology* 91:492–499.
4. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272.
5. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
6. Kay S, Boch J, Bonas U. 2005. Characterization of AvrBs3-like effectors from a *Brassicaceae* pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. *Mol. Plant Microbe Interact.* 18:838–848.
7. Otto TD, Dillon GP, Degraeve WS, Berriman M. 2011. RATT: rapid annotation transfer tool. *Nucleic Acids Res.* 39:e57.
8. Schiex T, Gouzy J, Moisan A, de Oliveira Y. 2003. FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. *Nucleic Acids Res.* 31:3738–3741.
9. Rodriguez-R LM, Grajales A, Arrieta-Ortiz ML, Salazar C, Restrepo S, Bernal A. 2012. Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiol.* 12:43.
10. Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
11. Vorholter FJ, Schneiker S, Goesmann A, Krause L, Bekel T, Kaiser O, Linke B, Patschkowski T, Ruckert C, Schmid J, Sidhu VK, Sieber V, Tauch A, Watt SA, Weisshaar B, Becker A, Niehaus K, Puhler A. 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.