RESOURCE ANNOUNCEMENT

Whole-Genome Sequence Resource of Indian Race 4 of *Xanthomonas campestris* pv. *campestris*, the Causal Agent of Black Rot Disease of *Brassica oleracea* var. *capitata*

Dinesh Singh,[1,†] Amit K. Kesharwani,[1,2] Kalpana Singh,[3] Sarika Jaiswal,[3] Mir Asif Iquebal,[3] Neelam Geat,[1] and Anupama S. Avasthi[2]

[1] Plant Bacteriology Lab, Division of Plant Pathology, ICAR, Indian Agricultural Research Institute, New Delhi 110012, India
[2] Amity Institute of Biotechnology, Amity University, Noida 201303, Uttar Pradesh, India
[3] Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012, India

Funding
This study was funded by the Science and Engineering Research Board-Department of Science and Technology (SERB-DST), grant numbers EMR/2016/005238, Government of India.

Keywords
crucifers, genome, genomics, pathogen diversity, prokaryotes, race, *Xanthomonas campestris* pv. *campestris*

Xanthomonas campestris pv. *campestris* is a causative agent of black rot disease of cruciferous crops. A whole-genome sequence of any race of *X. campestris* pv. *campestris* has not been reported from India. The isolate Xcc-C7, race 4, of *X. campestris* pv. *campestris* was isolated from cabbage (*Brassica oleracea* var. *capitata*) from Bengaluru, in the southern parts of India. Whole-genome sequence data were generated by the next-generation sequencing-based single-molecule real-time sequencing (SMRT) techniques. This study will improve our knowledge of genomic diversity in *X. campestris* pv. *campestris* and pave the way for research on host–pathogen interactions (crucifer crops–*X. campestris* pv. *campestris*) to develop resistance in cultivated Brassicaceae crops.

X. campestris pv. *campestris* (Pammel) Dowson is a Gram-negative, rod-shaped pathogenic bacterium that causes black rot disease in many cruciferous crops, worldwide. It is a vascular disease developing 'V'-shaped yellow lesions accompanied by blackened veins at margin of the leaf (Fig. 1A). It is genetically diversified by the specific host range of crucifer crops, including *B. oleracea* vegetables (cole crops such as cabbage, cauliflower, brussels sprout, broccoli, knoll khol, and kale), oilseed crops, ornamentals, and weeds (Singh et al. 2016; Vicente et al. 2001).

The annual worldwide spread of black rot disease limits the yield of cole crops under favorable conditions (Singh and Dhar 2011; Singh et al. 2014). The diversity within *X. campestris* pv. *campestris* is reported by several researchers (Jensen et al. 2010; Singh and Dhar 2011; Singh et al. 2016). In fact, the pathovar is subdivided into nine races based on host–pathogen interactions (Vicente et al. 2001). According to Singh et al. (2016), only races 1, 4, and 6 of *X. campestris* pv. *campestris* of crucifer crops have been identified in India, with races 1 and 4 occurring predominately worldwide. Races 2, 3, and 5 seem to be rare (Griffiths et al. 2009; Singh et al. 2016; Vicente et al. 2002), although, race 6 was identified from *B. rapa* (Vicente et al. 2006) and *B. oleracea* var. *capitata* (Singh et al. 2016). The whole-genome sequences of three *X. campestris* pv. *campestris* strains have been reported—namely, B100, United Kingdom (race 1, AM920689), *B. oleracea* (Vorhölter et al. 2008); ATCC33913, United Kingdom (race 3, AE008922), *B. oleracea* var. *gemmifera* (da Silva et al. 2002); and 8004, China (race 9, CP000050), *B. oleracea* var. *botrytis* (Qian et al. 2005). Furthermore, Bolot et al. (2015) reported the whole-genome sequence of *X. campestris* pv. *campestris* race 4 (CM002673)

[†]Corresponding author: D. Singh; dinesh_iari@rediffmail.com

D. Singh and A. K. Kesharwani contributed equally to this work.

The author(s) declare no conflict of interest.

Accepted for publication 2 February 2022.

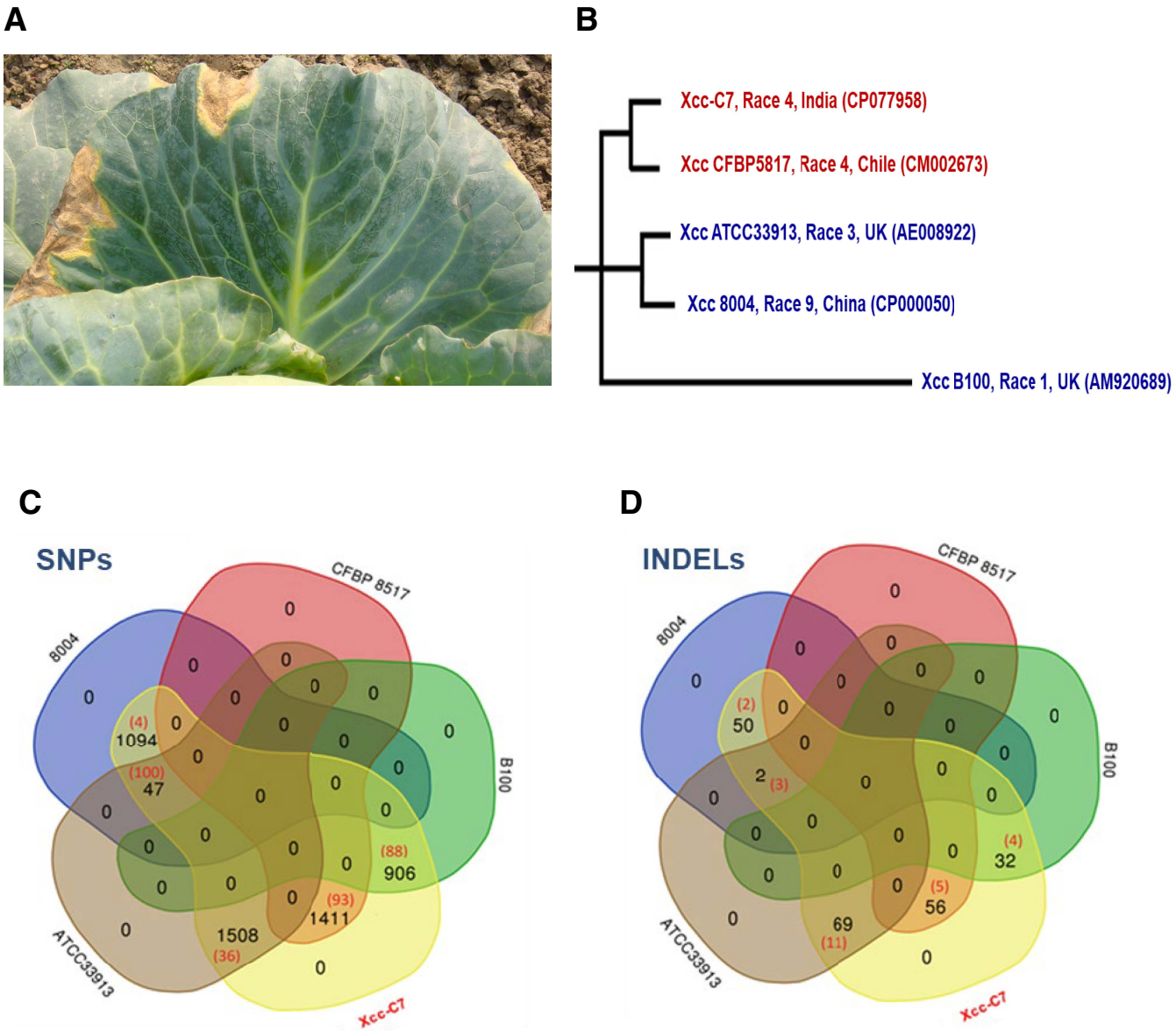

Fig. 1. A, Symptoms of typical black rot V-shaped lesion on a cabbage leaf; **B,** phylogenetic relationship between five strains of *Xanthomonas campestris* pv. *campestris*; **C,** single-nucleotide polymorphisms (SNPs) of five strains of *X. campestris* pv. *campestris* (genes having abundance of SNPs are shown in parenthesis); and **D,** insertions and deletions (INDELs) among the five different strains of *X. campestris* pv. *campestris* (genes having abundance of INDELs are shown in parenthesis).

isolated from cauliflower (*B. oleracea* var. *botrytis*), Chile. In this study, the isolate Xcc-C7 of *X. campestris* pv. *campestris* was isolated from black-rot-infected leaves of cabbage (*B. oleracea* var. *capitata*). Total genomic DNA was extracted (Qiagen, Germany) and completely sequenced using the PacBio sequel system having a sequencing depth is 117.46× through the PacBio SMRT sequencing technique (Pacific Biosciences Inc., Menlo Park, CA, U.S.A.). In total, 11,433 polymerase reads (N_{50}) were obtained encompassing 601,528,411 bp of the whole-genome sequence of Indian isolate Xcc-C7, race 4. The de novo assembly was performed using Flye assembler (Kolmogorov et al. 2019) and assembly statistics were calculated using QUAST (Gurevich et al. 2013). Assembly was further passed through BUSCO (Simão et al. 2015). The quality assessment of generated de novo assembly was performed using CheckM (Parks et al. 2015) in order to check completeness and contamination in

Table 1. Comparative statistics of genomic assembly of *Xanthomonas campestris* pv. *campestris* Indian strain Xcc-C7, race 4 with assemblies of other strains reported from other countries

	***X. campestris* pv. *campestris* strains**				
Features	**Xcc-C7**	**B100**	**ATCC33913**	**8004**	**CFBP5817**
Country	India	United Kingdom	United Kingdom	China	Chile
Race	Race 4	Race 1	Race 3	Race 9	Race 4
Accession number	CP077958	AM920689	AE008922	CP000050	CM002673
Length (bp)	5,121,051	5,079,002	5,076,188	5,148,708	4,918,955
G + C content (%)	64.98	65.00	63.7	64.94	65.03
Total genes	4,493	4,448	4,430	4,462	4,347
Total protein-coding genes	4,347	4,299	4,277	4,309	4,200
Genes encoding proteins with known function	4,217	4,180	4,163	4,190	4,063
Genes encoding hypothetical proteins (i.e., pseudogenes)	130	119	114	119	137
Total RNA-encoding genes	146	149	153	153	147
Ribosomal RNA-encoding genes	6	6	6	6	3
Transfer RNA-encoding genes	54	54	54	54	52
Noncoding RNA-encoding genes	86	89	93	93	92
Plasmids	0	0	0	0	0
Single-nucleotide polymorphisms (SNPs)	4,966	906	1,555	1,141	1,411
Genes having an abundance of SNPs	321	88	136	104	93
Insertions and deletions (INDELs)	209	32	71	52	56
Genes having abundance of INDELs	25	4	14	5	5
Reference	Present study	Vorhölter et al. 2008	da Silva et al. 2002	Qian et al. 2005	Bolot et al. 2015

assembly by comparing with marker lineage of *Gammaproteobacteria* along with 67 genomes, 481 markers, and 276 marker sets. Glimmer, version 3 (Delcher et al. 1999) was used to predict the total protein-coding genes (CDS). The complete genome assembly was composed of 5,121,051 bp and has a total GC content of 64.98%. Of 4,493 total genes, 4,347 CDS and 4,217 genes with assigned function were predicted in Xcc-C7. The profiles of 146 RNA-encoding genes were identified, including 6 ribosomal RNA-encoding genes, 54 transfer RNA-encoding genes, and 86 noncoding RNAs. This whole-genome sequence data has been deposited in the NCBI GenBank database (CP077958).

The comparative analysis of genomic features of Indian race 4 with genomes of four other strains (namely, *X. campestris* pv. *campestris* B100, 8004, CFBP5817, and ATCC33913) elucidates that the Indian race 4 is closely related to the Chilean strain CFBP5817 (Fig. 1C). The genome size, pseudogenes, and RNA-encoding genes of Indian race 4 (approximately 5.1 Mb,130, and 146, respectively) were found comparable with Chilean strain CFBP5817, race 4 (approximately 4.9 Mb, 137, and 147, respectively) (Table 1). Furthermore, 19 genes of Xcc-C7—namely, *pilC* (XC_1057), *iroN* (XC_0123), *nagA* (XC_0754), *rmlA* (XC_3612), *hrcJ* (XC_3009), *hrcV* (XC_3013), *hrpG* (XC_3077), *clp* (XC_0486), *rpfG* (XC_2335), *leuC* (XC_0833), *leuA* (XC_0837), *argC* (XC_1875), *hisF* (XC_2375), *aroA* (XC_2643), *purC* (XC_0467), *fadB* (XC_3420), *uptB* (XC_3641), *icd* (XC_3854), and *ppsA* (XC_1952)—were found in common with genes of strain 8004, which were validated genes of pathogenicity. These genes could be the putative candidate genes that may play a role in adaptation of *X. campestris* pv. *campestris* to cabbage as a host, along with other genes. The phylogenetic tree of whole-genome assemblies of the five strains showed that Xcc-C7 is closely related to CFBP5817, as compared with B100, ATCC33913, and 8004, suggesting that both strains CFBP5817 and Xcc-C7 belong to *X. campestris* pv. *campestris* race 4 (Fig. 1B). In total, 4,966 single-nucleotide polymorphisms (SNPs) and 209 insertions and deletions (INDELs) were extracted from Xcc-C7, race 4 with respect to other strains used in present study. The Venn diagram shows overlapping SNPs and INDELs among these strains (Fig. 1C and D). The quality assessment of generated de novo assembly using CheckM showed the completeness assembly as 99.64%, confirming it to be a good-quality assembly and 0.06% contamination. There were missing markers, single-copy markers, and different grades of marker heterogeneity. Only one marker was missing, while the remaining 480 markers were the single-copy markers in our Xcc-C7 genome assembly.

Data availability. The assembled complete genome sequence of the Xcc-C7 isolate of *X. campestris* pv. *campestris* has been deposited in the GenBank database under accession number CP077958. The accession number of the BioProject is PRJNA736230 and the BioSample is SAMN19613641.

Acknowledgments

The authors are thankful to the Head, Division of Plant Pathology, ICAR-IARI, New Delhi, India. We thank the Amity Institute of Biotechnology, Amity University, Noida, Uttar Pradesh, India, in which Amit K. Kesharwani is registered for his Ph.D. program.

Literature Cited

Bolot, S., Cerutti, A., Carrère, S., Arlat, M., Fischer-Le Saux, M., Portier, P., Poussier, S., Jacques, M.-A., and Noël, L. D. 2015. Genome sequences of the race 1 and race 4 *Xanthomonas campestris* pv. *campestris* strains CFBP 1869 and CFBP 5817. Genome Announce. 3:e01023-15.

da Silva, A. C. R., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., Monteiro-Vitorello, C. B., Van Sluys, M. A., Almeida, N. F., Alves, L. M. C., do Amaral, A. M., Bertolini, M. C., Camargo, L. E. A., Camarotte, G., Cannavan, F., Cardozo, J., Chambergo, F., Ciapina, L. P., Cicarelli, R. M. B., Coutinho, L. L., Cursino-Santos, J. R., El-Dorry, H., Faria, J. B., Ferreira, A. J. S., Ferreira, R. C. C., Ferro, M. I. T., Formighieri, E. F., Franco, M. C., Greggio, C. C., Gruber, A., Katsuyama, A. M., Kishi, L. T., Leite, R. P., Lemos, E. G. M., Lemos, M. V. F., Locali, E. C., Machado, M. A., Madiera, A. M. B. N., Martinez-Rossi, N. M., Martins, E. C., Meidanis, J., Menck, C. F. M., Miyaki, C. Y., Moon, D. H., Moreira, L. M., Novo, M. T. M., Okura, V. K., Oliveira, M. C., Oliveira, V. R., Pereira, H. A., Rossi, A., Sena, J. A. D., Silva, C., de Souza, R. F., Spinola, L. A. F., Takita, M. A., Tamura, R. E., Teixeira, E. C., Tezza, R. I. D., Trindade dos Santos, M., Truffi, D., Tsai, S. M., White, F. F., Setubal, J. C., and Kitajima, J. P. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature 417:459-463.

Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27:4636-4641.

Griffiths, P. D., Marek, L. F., and Robertson, L. D. 2009. Identification of crucifer accessions from the NC-7 and NE-9 plant introduction collections that are resistant to black rot (*Xanthomonas campestris* pv. *campestris*) races 1 and 4. HortScience 44:284-288.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075 https://www.ncbi.nlm.nih.gov/assembly/GCF_009177345.1/

Jensen, B. D., Vicente, J. G., Manandhar, H. K., and Roberts, S. J. 2010. Occurrence and diversity of *Xanthomonas campestris* pv. *campestris* in vegetable brassica fields in Nepal. Plant Dis. 94:298-305.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. 2019. Assembly of long error-prone reads using repeat graphs. Nat. Biotechnol. 37:540-546.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. 2015. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25:1043-1055.

Qian, W., Jia, Y., Ren, S.-X., He, Y.-Q., Feng, J.-X., Lu, L.-F., Sun, Q., Ying, G., Tang, D.-J., Tang, H., Wu, W., Hao, P., Wang, L., Jiang, B.-L., Zeng, S., Gu, W.-Y., Lu, G., Rong, L., Tian, Y., Yao, Z., Fu, G., Chen, B., Fang, R., Qiang, B., Chen, Z., Zhao, G.-P., Tang, J.-L., and He, C. 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. Genome Res. 15:757-767.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212.

Singh, D., and Dhar, S. 2011. Bio-PCR based diagnosis of *Xanthomonas campestris* pv. *campestris* in black rot of infected leaves of crucifers. Indian Phytopathol. 64:7-11.

Singh, D., Raghavendra, B. T., Rathaur, P. S., Singh, H., Raghuwanshi, R., and Singh, R. P. 2014. Detection of black rot disease causing pathogen *Xanthomonas campestris* pv. *campestris* by bio-PCR from seeds and plant parts of cole crops. Seed Sci. Technol. 42:36-46.

Singh, D., Rathaur, P. S., and Vicente, J. G. 2016. Characterization, genetic diversity and distribution of *Xanthomonas campestris* pv. *campestris* races causing black rot disease in cruciferous crops of India. Plant Pathol. 65:1411-1418.

Vicente, J. G., Conway, J., Roberts, S. J., and Taylor, J. D. 2001. Identification and origin of *Xanthomonas campestris* pv. *campestris* races and related pathovars. Phytopathol. 91:492-499.

Vicente, J. G., Everett, B., and Roberts, S. J. 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.

Vicente, J. G., Taylor, J. D., Sharpe, A. G., Parkin, I. A. P., Lydiate, D. J., and King, G. J. 2002. Inheritance of race-specific resistance to *Xanthomonas campestris* pv. *campestris* in *Brassica* genomes. Phytopathology 92:1134-1141.

Vorhölter, F.-J., Schneiker, S., Goesmann, A., Krause, L., Bekel, T., Kaiser, O., Linke, B., Patschkowski, T., Rückert, C., Schmid, J., Sidhu, V. K., Sieber, V., Tauch, A., Watt, S. A., Weisshaar, B., Becker, A., Niehaus, K., and Pühler, 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.