



Genome Sequences of 17 Strains from Eight Races of *Xanthomonas campestris* pv. *campestris*

 Caroline Bellenot,^{a,b,c}  Sébastien Carrère,^{a,b,c} Carine Gris,^{a,b,c}  Laurent D. Noël,^{a,b,c}  Matthieu Arlat^{a,b,c}

^aINRAE, UMR 441, Laboratoire des Interactions Plantes-Microbes-Environnement, Castanet-Tolosan, France

^bCNRS, UMR 2594, Laboratoire des Interactions Plantes-Microbes-Environnement, Castanet-Tolosan, France

^cUniversité de Toulouse, Université Paul Sabatier, Toulouse, France

ABSTRACT *Xanthomonas campestris* pv. *campestris* is a group of phytopathogenic bacteria causing black rot disease on Brassicaceae crops. Here, we report on draft genome sequences of 17 strains representing eight of nine known races of this pathogen, including the pathotype strain CFBP 6865.

Genomic analysis is a powerful approach to study the diversity and evolution of virulence determinants of pathogenic bacteria (1, 2). *Xanthomonas campestris* pv. *campestris* is the causal agent of black rot disease, affecting many crop plants from the Brassicaceae family (3). *X. campestris* pv. *campestris* strains have been organized into nine races based on symptoms caused on a set of host and nonhost plants (3). Amplified fragment length polymorphism (AFLP) allowed the classification of *X. campestris* pv. *campestris* strains into seven genomic clades (clades A to G) (4), which is not congruent with the race classification. Here, we present the genome sequences of 17 *X. campestris* pv. *campestris* strains (Table 1), including the pathotype strain CFBP 6865. These strains were isolated from different crop varieties or subspecies of *Brassica oleracea* and *Brassica rapa* between 1958 and 2002, in Australia, Belgium, China, France, or Germany. They belong to four of the seven genomic clades and are representative of eight of the nine known races. All of these strains were obtained from the International Centre for Microbial Resources-French Collection for Plant-Associated Bacteria (CIRM-CFBP), National Research Institute for Agriculture, Food, and the Environment (INRAE).

The *X. campestris* pv. *campestris* strains were grown overnight in MOKA-rich medium (4 g/L yeast extract, 8 g/L Casamino Acids, 2 g/L K₂HPO₄, and 0.3 g/L MgSO₄·7H₂O) at 28°C. From these cultures, genomic DNA was extracted with a Wizard genomic DNA purification kit (Promega). Genomic DNA libraries were prepared using a NEXTflex PCR-free DNA-sequencing kit (PerkinElmer) for most strains and a NEBNext kit (New England BioLabs) for strains CFBP 119, CFBP 4953, CFBP 4956, CFBP 6863, CN19, and CN20. All libraries were constructed following the manufacturers' instructions. *X. campestris* pv. *campestris* total genomic DNA (including chromosome and plasmids) was sequenced on an Illumina HiSeq 2000 platform (2 × 101-bp paired-end reads) or a Genome Analyzer_{IIx} system (2 × 76-bp paired-end reads) (for strains CFBP 119, CFBP 4953, CFBP 4956, CFBP 6863, CN19, and CN20). Several assemblies were produced using the SOAPdenovo assembler (5) with incremental *k*-mer sizes (starting with a *k*-mer length corresponding to one-third of the read length, to a maximum of 99 nucleotides [nt], with a step of 4 nt). The resulting contigs were scaffolded using the Velvet assembler (6) (parameters: `-cov_cutoff 5 -min_contig_length 100 -max_divergence 0.05 -exportFiltered yes -exp_cov auto`) using the same *k*-mer range. A final round with SOAPGapCloser (parameter: `-p 31`) was performed to fill in the gaps, and the remaining redundancy (misassembled contigs) was removed using MegaBLAST-based identification. Genome sequences were annotated using the default parameters of the EuGene-P annotation pipeline to identify RNAs and protein-coding genes (7). The completeness of the genomes was assessed with BUSCO v5.1.2 (default parameters), using the *xanthomonadales_odb10* data set (8, 9) (Table 1).

Editor David A. Baltus, University of Arizona

Copyright © 2022 Bellenot et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Matthieu Arlat, matthieu.arlat@inrae.fr.

The authors declare no conflict of interest.

Received 18 March 2022

Accepted 16 May 2022

Published 13 June 2022

TABLE 1 Description of the *Xanthomonas campestris* pv. *campestris* strains reported and statistics for the sequencing and assembly of their genomes

Strain	CFBP designation ^a	Clade ^b	Ra ce ^c	Date of isolation	Host	Country of isolation	No. of ds	Genome size (bp)	Genome coverage (x)	No. of contigs	M ₅₀ (bp)	GC content (%)	BUSCO completeness (%)	No. of coding sequences	GenBank accession no.	SRA accession no.	DOI no.
147	CFBP 7156	A	4	1967	<i>Brassica oleracea</i>	France	10,035,656	5,060,825	320	141	80,681	65.0	99.3	4,305	AVEL000000000.1/ SRP322641	10.25794/reference/um8maue7	
CFBP 1119	CFBP 1119	A	1	1967	<i>Brassica oleracea</i> var. botrytis	France	35,046,928	5,008,837	1,063	74	183,459	64.6	99.9	4,220	CBVE000000000.1/ SRP322638	10.25794/reference/njwrm70	
CFBP 1124	CFBP 1124	C	8	1967	<i>Brassica oleracea</i> var. botrytis	France	7,211,603	4,997,962	240	103	149,591	65.6	99.7	4,436	AVEV000000000.1/ SRP322639	10.25794/reference/yhf1qnob	
CFBP 1712	CFBP 1712	E	5	1975	<i>Brassica oleracea</i> cv. Capitata	France	17,215,274	5,015,961	553	129	79,498	64.9	98.6	4,461	AVDC000000000.1/ SRP322649	10.25794/reference/dhvsif18	
CFBP 1713	CFBP 1713	E	5	1975	<i>Brassica oleracea</i> var. botrytis	France	18,408,662	5,041,610	739	89	151,632	64.4	99.4	4,491	AVDM000000000.1/ SRP322642	10.25794/reference/rmbysyly	
CFBP 4953	CFBP 4953		7	1999	<i>Brassica oleracea</i> var. botrytis cv. Cortes	Belgium	32,918,799	5,052,992	991	83	163,533	64.7	99.9	4,562	CBVF000000000.1/ SRP322648	10.25794/reference/hnmw_mnq	
CFBP 4954	CFBP 4954		6	1999	<i>Brassica oleracea</i> var. botrytis cv. Aviso	Belgium	6,433,757	5,073,734	204	149	66,639	65.1	99.4	4,496	AVDE000000000.1/ SRP322643	10.25794/reference/fkxlurkw	
CFBP 4955	CFBP 4955	E	9	1999	<i>Brassica oleracea</i> var. botrytis cv. Aviso	Belgium	4,239,839	5,136,167	138	142	84,609	65.1	99.7	4,600	AVDF000000000.1/ SRP322645	10.25794/reference/c02meut	
CFBP 4956	CFBP 4956		4	1999	<i>Brassica oleracea</i> var. botrytis cv. Spacestar	Belgium	37,480,089	4,965,880	1,148	125	109,963	64.5	99.8	4,396	CBVG000000000.1/ SRP322646	10.25794/reference/nx9gmcmt	
CFBP 5130	CFBP 5130	A	7		<i>Brassica</i> sp.		9,232,935	5,056,538	295	129	93,783	64.9	99.7	4,561	AVDG000000000.1/ SRP322647	10.25794/reference/_016f91q	
CFBP 5683	CFBP 5683	E	3	1979	<i>Brassica oleracea</i>	France	2,579,478	5,030,566	82	184	51,279	65.4	98.8	4,492	AVDH000000000.1/ SRP322644	10.25794/reference/0sr1ne3k	
CFBP 6863	CFBP 6863		9	1958	<i>Brassica oleracea</i> var. botrytis	Germany	32,059,044	5,001,025	975	94	173,153	64.5	99.7	4,458	CBVH000000000.1/ SRP322661	10.25794/reference/shkh_1 dj	
CFBP 6865R ^d	CFBP 6865		5	1975	<i>Brassica oleracea</i> var. capitata	Australia	16,397,066	5,166,940	916	24	415,646	63.7	99.6	4,657	AVDT000000000.1/ SRX11066177	10.25794/reference/uce57zl_	
CN01	CFBP 8237	A	1	2002	<i>Brassica rapa</i> subsp. pekinensis	China	20,801,984	5,030,897	836	75	179,901	64.3	99.9	4,543	AVDN000000000.1/ SRP322694	10.25794/reference/zzeitq6q	
CN10	CFBP 8244	B	7	2002	<i>Brassica rapa</i> subsp. pekinensis	China	19,211,549	4,958,955	784	101	124,175	64.5	99.5	4,411	AVDR000000000.1/ SRP322708	10.25794/reference/h5nghmsn	
CN19		B	4			China	32,316,246	4,958,083	992	113	115,047	64.4	99.5	4,422	CBVC000000000.1/ SRP322706	10.25794/reference/ez2luoj	
CN20	CFBP 8251		1	2002	<i>Brassica oleracea</i> var. alboglabra	China	31,554,774	5,084,276	944	115	116,198	64.5	99.6	4,551	CBVD000000000.1/ SRP322707	10.25794/reference/e3ex38hh	

^a Strains are available in the CIRM-CFBP (www.cirm-cfbp.fr/page/Home).

^b Clades based on AFLP of *X. campestris* pv. *campestris* strains reported by Guy et al. (4).

^c Races as described previously by Vicente and Holub (3).

^d Pathotype strain.

Data availability. The genome sequences were deposited in GenBank, and the raw reads were deposited in the Sequence Read Archive (SRA); their accession numbers are listed in Table 1. The genome annotations are available under the DOI numbers provided in Table 1.

ACKNOWLEDGMENTS

We acknowledge the CIRM-CFBP for providing the strains. We are grateful to Perrine David (HM Clause, France) for the race typing of *X. campestris* pv. *campestris* strains performed in the framework of the Brassichax project (Limagrain).

C.B. was funded by a PhD grant from the ANR NEPHRON (grant ANR-18-CE20-0020-01) and CNRS. S.C., C.G., L.D.N., and M.A. were supported by the ANR XANTHOMIX (grant ANR-2010-GENM-013-02) and NEPHRON (grant ANR-18-CE20-0020-01). This study is set within the frameworks of the Laboratoires d'Excellences (LABEX) TULIP (ANR-10-LABX-41) and of the École Universitaire de Recherche (EUR) TULIP-GS (ANR-18-EURE-0019).

REFERENCES

- Allen JP, Snitkin E, Pincus NB, Hauser AR. 2021. Forest and trees: exploring bacterial virulence with genome-wide association studies and machine learning. *Trends Microbiol* 29:621–633. <https://doi.org/10.1016/j.tim.2020.12.002>.
- Straub C, Colombi E, McCann HC. 2021. Population genomics of bacterial plant pathogens. *Phytopathology* 111:23–31. <https://doi.org/10.1094/PHYTO-09-20-0412-RVW>.
- Vicente JG, Holub EB. 2013. *Xanthomonas campestris* pv. *campestris* (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Mol Plant Pathol* 14:2–18. <https://doi.org/10.1111/j.1364-3703.2012.00833.x>.
- Guy E, Genissel A, Hajri A, Chabannes M, David P, Carrere S, Lautier M, Roux B, Boureau T, Arlat M, Poussier S, Noël LD. 2013. Natural genetic variation of *Xanthomonas campestris* pv. *campestris* pathogenicity on *Arabidopsis* revealed by association and reverse genetics. *mBio* 4:e00538-12. <https://doi.org/10.1128/mBio.00538-12>.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. <https://doi.org/10.1186/2047-217X-1-18>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
- Sallet E, Roux B, Sauviac L, Jardinaud M-F, Carrère S, Faraut T, de Carvalho-Niebel F, Gouzy J, Gamas P, Capela D, Bruand C, Schiex T. 2013. Next-generation annotation of prokaryotic genomes with EuGene-P: application to *Sinorhizobium meliloti* 2011. *DNA Res* 20:339–354. <https://doi.org/10.1093/dnares/dst014>.
- Seppy M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness, p 227–245. *In* Kollmar M (ed), *Gene prediction: methods and protocols*. Springer, New York, NY.
- Waterhouse RM, Seppy M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35:543–548. <https://doi.org/10.1093/molbev/msx319>.