



# Complete or High-Quality Draft Genome Sequences of Six *Xanthomonas hortorum* Strains Sequenced with Short- and Long-Read Technologies

Nay C. Dia,<sup>a,b</sup> Fabio Rezzonico,<sup>a</sup> Theo H. M. Smits,<sup>a</sup>  Joël F. Pothier<sup>a</sup>

<sup>a</sup>Environmental Genomics and Systems Biology Research Group, Institute for Natural Resource Sciences, Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland

<sup>b</sup>Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland

**ABSTRACT** We report the genome sequences of six *Xanthomonas hortorum* species-level clade members, *X. hortorum* pathovars taraxaci, pelargonii, cynarae, and gardneri (complete genome sequences) and *X. hortorum* pathovars carotae and vitians (high-quality draft genome sequences). Both short- and long-read sequencing technologies were used.

The genetic relatedness of the *Xanthomonas hortorum* species-level clade (slc) was established by Parkinson et al. using partial *gyrB* gene sequences (1). The most recently updated taxonomy of the clade (2) includes seven pathovars of *X. hortorum* (cynarae, gardneri, carotae, hederiae, pelargonii, taraxaci, and vitians type B) (1–3).

We report here the whole-genome sequences of six *X. hortorum* strains (Table 1). The strains were isolated between 1942 and 2008 from a wide range of hosts and in various countries (Table 1). The genome sequences published in this work are either complete genome sequences or high-quality draft genome sequences, and all have five contigs or less (Table 1).

The strains were initially obtained as freeze-dried cultures in glass ampoules from two international strain collections abbreviated in the strain names as CFBP (Collection Française de Bactéries Associées aux Plantes, Beaucauzé, France) and NCPPB (National Collection of Plant Pathogenic Bacteria, York, United Kingdom). After revival on nutrient yeast extract glycerol agar plates (4) for 2 days at 28°C, an isolated colony was streaked onto the same medium and grown in the same manner. The strains were then stored in a –80°C ultrafreezer as glycerol stocks in 50% (vol/vol) nutrient yeast extract glycerol broth (4) until further use. The strains were revived as described above, and a single colony was used as starting material for downstream experiments.

Genomic DNA (gDNA) for Illumina MiSeq short-read sequencing was extracted from cells grown overnight at 28°C in nutrient yeast extract glycerol broth (4) using the NucleoSpin tissue kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol with the following specifications: elution buffer was heated at 70°C before use, and the gDNA was eluted with 60  $\mu$ l of this buffer. The quality of the gDNA was checked using a fragment analyzer (Advanced Analytical Technologies, Inc., Ankeny, IA) and quantified using the Quant-iT PicoGreen double-stranded DNA (dsDNA) quantification assay (Thermo Fisher Scientific, Waltham, MA). Library preparation was done using the Nextera XT DNA library prep kit (Illumina, San Diego, CA) following the manufacturer's instructions. Sequencing was performed on a MiSeq Illumina sequencer with 2  $\times$  300-bp paired-end reads using a MiSeq reagent kit version 3 (Illumina) according to the manufacturer's instructions.

Genomic DNA for long-read sequencing was extracted from overnight-grown cells using the Genra PureGene Yeast/Bact kit protocol (Qiagen, Hilden, Germany). The

**Citation** Dia NC, Rezzonico F, Smits THM, Pothier JF. 2020. Complete or high-quality draft genome sequences of six *Xanthomonas hortorum* strains sequenced with short- and long-read technologies. Microbiol Resour Announc 9:e00828-20. <https://doi.org/10.1128/MRA.00828-20>.

**Editor** David A. Baltus, University of Arizona

**Copyright** © 2020 Dia 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 Joël F. Pothier, [joel.pothier@zhaw.ch](mailto:joel.pothier@zhaw.ch).

**Received** 6 August 2020

**Accepted** 15 September 2020

**Published** 8 October 2020

**TABLE 1** Genome metrics and accession numbers of the newly sequenced genomes within the *Xanthomonas hortorum* species-level clade<sup>a</sup>

Strain <sup>b</sup>	Bacterial species	Origin (yr)	Host	Genome size (bp)	G+C content (%)	Total genes (N <sub>50</sub> )	Genome no. of status (N <sub>50</sub> )	No. of contigs/plasmids	Illumina data:				Oxford Nanopore data:				ENA accession no.	
									17/15 indexes	Total no. of reads	Avg read length (bp)	Avg coverage (x)	BC	Total reads no.	Read length (bp)	Avg coverage (x)		SRA accession no. (MIS/Seq/MinION)
CFBP 498	<i>X. hortorum</i> pv. <i>vitians</i>	USA (1949)	<i>Lactuca</i> sp.	5,678,543	63.23	4,976	HQ draft (5,365,193)	4/3	N703/S504	1,112,420	259	38	BC18	27,882	9,898	58	ERR4327117	LR828257 (chr.), LR828258 (p224), LR828259 (p47), LR828260 (p41)
CFBP 2044	<i>X. hortorum</i> pv. <i>cynarae</i>	France (1981)	<i>Cynara scolymus</i> L.	5,119,234	63.67	4,322	Complete	2/1	N705/S517	2,593,908	277	96	BC16	18,849	12,128	55	ERR4326260	LR828251 (chr.), LR828252 (p40)
CFBP 2533 <sup>c</sup>	<i>X. hortorum</i> pv. <i>pelargonii</i>	New Zealand (1974)	<i>Pelargonium peltatum</i>	5,287,542	63.72	4,427	Complete	3/2	N702/S517	3,398,636	261	111	BC15	14,695	8,396	20	ERR4326261, ERR4327119	LR828261 (chr.), LR828262 (p47), LR828263 (p41)
CFBP 7900 <sup>d</sup>	<i>X. hortorum</i> pv. <i>carotae</i>	USA (2008)	<i>Daucus carota</i> (seed)	5,149,201	63.77	4,274	HQ draft (2,659,169)	5/ND	N702/S504	1,721,852	264	65	BC14	7,777	8,728	15	ERR4326262, ERR4327120	CAJDKC000000000
CFBP 8129	<i>X. hortorum</i> pv. <i>gardneri</i>	Costa Rica (1991)	<i>Solanum lycopersicum</i>	5,440,942	63.48	4,655	Complete	4/3	N705/S504	1,653,590	258	27	BC20	45,436	8,073	80	ERR4326263, ERR4327121	LR828253 (chr.), LR828254 (p211), LR828255 (p46), LR828256 (p26)
NCPBP 940 <sup>e</sup>	<i>X. hortorum</i> pv. <i>taraxaci</i>	USA (1942)	<i>Taraxacum kok-saghyz</i>	5,029,134	63.83	4,316	Complete	2/1	N701/S503	1,197,722	269	43	BC17	9,847	7,654	18	ERR4329153, ERR4327122	LR828264 (chr.), LR828265 (p30)

<sup>a</sup>BC: barcode; HQ: high-quality; ND, not determined; chr., chromosome.

<sup>b</sup>The culture collections providing strains are abbreviated in the strain names as CFBP (Collection Française de Bactéries Associées aux Plantes, Beaucauzé, France) and NCPBP (National Collection of Plant Pathogenic Bacteria, York, United Kingdom). Superscript P following a strain name indicates the pathotype strain for the pathovar.

gDNA was quantified and its quality checked as described above. Library preparation and sequencing were performed with the ligation sequencing kit (catalog no. SQK-LSK109; Oxford Nanopore Technologies, Oxford, United Kingdom) and run on an R9.4.1 flow cell with a MinION sequencer. The native barcoding expansion kit (catalog no. XP-NBD114) was used for multiplexing. Base calling was performed using Guppy version 3.0.7 in the “accurate” mode implemented in the MinION release 19.06.8.

A hybrid assembly using the MiSeq and MinION reads was conducted with Unicycler version 0.4.7 (5). To check for misalignments, the MiSeq reads were mapped against the Unicycler assemblies using SeqMan Pro version 12.2 (DNASTar, Madison, WI). Contigs were reordered using the Mauve Contig Mover version 2.3.1 (6) when required (7). The scrub and align options of Unicycler were used to detect chimeras and check for high-error regions, respectively. The manual improvement of the genomes was finalized with Bandage version 0.8.1 (8). The genomes were then annotated using Prokka version 1.14.5 (9). Indels were checked using Ideel (10). All tools were run with default parameters unless otherwise specified.

The sizes of the hybrid assemblies ranged from 5,029,134 to 5,678,543 bp, a size range typically found in *Xanthomonas* genomes (Table 1). The G+C contents of the genomes varied from 63.23% to 63.83%, comparable to other *Xanthomonas* genome G+C contents. The sequenced genomes discussed here will be used for further analysis of evolution within the *X. hortorum* slc.

**Data availability.** The annotated genome sequences of the six *X. hortorum* strains have been deposited in ENA under BioProject no. [PRJEB38812](https://www.ncbi.nlm.nih.gov/bioproject/PRJEB38812). The genome and raw read accession numbers for each isolate are shown in Table 1.

## ACKNOWLEDGMENTS

We thank the HPC team of the School for Life Sciences and Facility Management at ZHAW for the computing resources and support.

This research project was supported by grant no. IZCOZO\_177064 from the Swiss National Foundation (SNF) for Scientific Research.

This article is based upon work from COST Action CA16107 EuroXanth, supported by COST (European Cooperation in Science and Technology).

## REFERENCES

- Parkinson N, Cowie C, Heeney J, Stead D. 2009. Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int J Syst Evol Microbiol* 59:264–274. <https://doi.org/10.1099/ijs.0.65825-0>.
- Morinière L, Burlet A, Rosenthal ER, Nesme X, Portier P, Bull CT, Lavire C, Fischer-Le Saux M, Bertolla F. 2020. Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that *Xanthomonas cynarae* Trébaol et al. 2000 emend. Timilsina et al. 2019 is a later heterotypic synonym of *Xanthomonas hortorum* Vauterin et al. 1995. *Syst Appl Microbiol* 43:126087. <https://doi.org/10.1016/j.syapm.2020.126087>.
- Young JM, Park D-C, Shearman HM, Fargier E. 2008. A multilocus sequence analysis of the genus *Xanthomonas*. *Syst Appl Microbiol* 31:366–377. <https://doi.org/10.1016/j.syapm.2008.06.004>.
- Turner P, Barber C, Daniels M. 1984. Behaviour of the transposons Tn5 and Tn7 in *Xanthomonas campestris* pv. *campestris*. *Mol Gen Genet* 195:101–107. <https://doi.org/10.1007/BF00332731>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* 25:2071–2073. <https://doi.org/10.1093/bioinformatics/btp356>.
- Smits THM. 2019. The importance of genome sequence quality to microbial comparative genomics. *BMC Genomics* 20:662. <https://doi.org/10.1186/s12864-019-6014-5>.
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31:3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Watson M. 2018. Indels are not ideal: quick test for interrupted ORFs in bacterial/microbial genomes. <https://github.com/mw55309/ideel>. Retrieved January 2020.