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

Nay C. Dia,a,b Fabio Rezzonico,a Theo H. M. Smits,a Joël F. Pothiera

Environmental Genomics and Systems Biology Research Group, Institute for Natural Resource Sciences, Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland
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, hederae, 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, Beaucouzé, 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 µl 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 × 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 Gentra PureGene Yeast/Bact kit protocol (Qiagen, Hilden, Germany). The
### TABLE 1 Genome metrics and accession numbers of the newly sequenced genomes within the Xanthomonas hortorum species-level clade

<table>
<thead>
<tr>
<th>Strain</th>
<th>Bacterial species</th>
<th>Origin (yr)</th>
<th>Host</th>
<th>Genomesize (bp)</th>
<th>G+C content (%)</th>
<th>Total no. of genes</th>
<th>Genomic status (N50 [bp])</th>
<th>Illumina data:</th>
<th>Oxford Nanopore data:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No. of contigs/plasmids</td>
<td>17/15 indexes</td>
<td>Total no. of reads</td>
<td>Avg read length (bp)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Avg read length (bp)</td>
<td>Total no. of reads</td>
<td>Avg coverage (x)</td>
<td>SRA accession no.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFBP 498</td>
<td>X. hortorum pv. vitians</td>
<td>USA (1949)</td>
<td>Lactuca sp.</td>
<td>5,678,543</td>
<td>63.23</td>
<td>4,976 HQ draft (5,365,193)</td>
<td>4/3</td>
<td>N705/S504</td>
<td>1,112,420</td>
</tr>
<tr>
<td>CFBP 7900</td>
<td>X. hortorum pv. carotae</td>
<td>USA (2008)</td>
<td>Daucus carota (seed)</td>
<td>5,149,201</td>
<td>63.77</td>
<td>4,274 HQ draft (2,659,169)</td>
<td>5/ND</td>
<td>N702/S504</td>
<td>1,721,852</td>
</tr>
<tr>
<td>NCPPB 940</td>
<td>X. hortorum pv. taraxaci</td>
<td>USA (1942)</td>
<td>Taraxacum kok-saghyz</td>
<td>5,029,134</td>
<td>63.83</td>
<td>4,316 Complete</td>
<td>2/1</td>
<td>N701/S503</td>
<td>1,197,722</td>
</tr>
</tbody>
</table>

---

*a* BC, barcode; HQ, high-quality; ND, not determined; chr., chromosome.

*b* The culture collections providing strains are abbreviated in the strain names as CFBP (Collection Française de Bactéries Associées aux Plantes, Beaucouzé, France) and NCPPB (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 MiniION 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 MiniION 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 (9). Indels were checked using Idel (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. 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. IZCOZ0_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**


