Nearly Complete Genome Assembly of the Pinewood Nematode *Bursaphelenchus xylophilus* Strain Ka4C1

Mehmet Dayi,a,b Simo Sun,a Yasunobu Maeda,a Ryusei Tanaka,a Akemi Yoshida,a Isheng Jason Tsai,d Taisei Kikuchi,a

aDepartment of Infectious Diseases, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan
bForestry Vocational School, Duzce University, Duzce, Turkey
cLaboratory of Genomics, Frontier Science Research Center, University of Miyazaki, Miyazaki, Japan
dBiodiversity Research Center, Academia Sinica, Taipei, Taiwan

**ABSTRACT** *Bursaphelenchus xylophilus* has been destroying pine forests in East Asia and western Europe. Here, we report its nearly complete genomic sequence containing five ~12-Mb scaffolds and one ~15-Mb scaffold representing six chromosomes. Large repeat regions that were previously unidentified are now reasonably integrated, particularly in the ~15-Mb scaffold.

*Bursaphelenchus xylophilus* causes pine wilt disease and has caused extensive damage to pine forests in East Asia and western Europe (1). A 75-Mbp *B. xylophilus* draft genome (v1.2) was previously sequenced using 454 and Illumina GAII technologies with 5,527 scaffolds and an \( N_{50} \) value of 950 kb (2). Here, we provide a nearly complete genome sequence produced through Nanopore long-read sequencing and Hi-C scaffolding.

The *B. xylophilus* strain Ka4C1, maintained at the University of Miyazaki, was cultured for 7 days on *Botrytis cinerea* that was grown on autoclaved barley grains. Mixed-stage worms were collected from the culture using the modified Baermann funnel technique (3). Briefly, worm culture was suspended in distilled water (dH2O) complemented with streptomycin, amphotericin B, and penicillin (antibiotic/antimycotic [anti/anti]; Gibco), and live worms were passed through a Kimwipe-lined (Crecia) sieve followed by discontinuous sucrose gradient centrifugation to remove culture debris (4). The worms were transferred to a worm lysis solution (Qiagen buffer G2 with \( 800 \text{ mg/ml} \) proteinase K, \( 50 \text{ mM} \) dithiothreitol [Wako], and \( 0.5 \text{ mg/ml} \) RNase A [Invitrogen]) and incubated at 55°C for 4 h after two freeze-thaw treatments. High-molecular-weight genomic DNA was spooled from ethanol precipitation following phenol-chloroform extraction and dissolved in 10 mM Tris (pH 8.0). A Nanopore library was prepared using \( 1 \mu g \) genomic DNA using a ligation sequencing kit SQK-L5K109 (Oxford Nanopore Technologies) according to the manufacturer’s protocol. A single 24-h sequencing run was performed with MinION R9.4.1 flow cells to obtain 2.7 Gbp of sequence data (182,569 reads; \( N_{50} \), 27 kbp). The Nanopore reads were base called to generate FASTQ files using the Guppy v4.0.15 basecaller (Oxford Nanopore Technologies) with the supplied dna_r9.4.1_450bps_hac configuration and were quality checked using NanoPlot v1.31.0 (5). An Illumina paired-end sequencing library was prepared from \( 1 \mu g \) of DNA using a TruSeq DNA sample preparation kit according to the manufacturer’s instructions. A total of 6.0 Gbp of paired-end reads (100 bp \( \times 2 \) ) were generated by library sequencing on an Illumina HiSeq 2000 instrument according to the manufacturer’s protocol. The raw Illumina sequence data were used for generating the genome assembly after removing the adapters and low-quality and duplicate reads using the Real-Time Analysis (RTA) v1.12.4.2 analysis pipeline (Illumina). The Hi-C library was prepared from \( \sim 10,000 \) fresh worms using an Arima-HiC kit (Arima Genomics) and a
Collibri ES DNA library prep kit (Thermo Fisher Scientific) according to the manufacturers’ protocols and was sequenced using a MiSeq instrument with the MiSeq reagent kit v3 (101 cycles × 2), and the 3.8 million short reads were quality checked using the Hi-C quality control pipeline (https://phasegenomics.github.io/2019/09/19/hic-alignment-and-qc.html).

The Nanopore long reads were assembled using Flye v2.7.1 (6) with the parameters –genome-size 75M and –iteration 4. After base correction by two rounds of Pilon v1.23 (7) with the Illumina paired-end reads, the assembly was further scaffolded using the 3D-DNA pipeline v180114 (8) without a misjoin correction process, and the chromosome-length scaffolds were extracted via manual curation using Juicebox v1.11.08 (9).

The final assembly was 78.3 Mbp long, which is 3.7 and 2.4 Mbp bigger than the previous v1.2 (2) and v2.0 (10) assemblies, respectively, and showed slightly better CEGMA v2.5 completeness values (Table 1). However, a G+C content of 40.4% was consistent for all assemblies. The new assembly contains six large scaffolds (five ~12-Mbp and one ~15-Mbp scaffolds) representing six chromosomes and five unassigned, 20-kb contigs, giving an N50 value of 12.8 Mbp. The largest scaffold contains large repeat sequence regions, which were unidentified in previous assemblies. The assembly stats of the old and new versions are summarized in Table 1.

**Data availability.** The *B. xylophilus* v5 assembly was deposited in DDBJ/EMBL/ENA/GenBank under project PRJEB40022. The raw Illumina, Nanopore, and Hi-C read data are available in the DDBJ Sequence Read Archive under the accession numbers DRR067231, DRR243685, and DRR243686, respectively.

**ACKNOWLEDGMENTS**

We thank Asuka Konosu and Melis Konno for technical assistance.

This work was funded in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI grant numbers 26460510 and 19H03212 and JST CREST grant number JPMJCR1857.

**REFERENCES**


