**MATERIALS AND METHODS**

**Illumina sequencing library preparations**

Illumina sequencing library preparations were constructed following the manufacturer’s protocol (NEBNext® Ultra™ DNA Library Prep Kit for Illumina®). For each sample, 1 μg of genomic DNA was randomly fragmented to <500 bp by sonication (Covaris S220). The fragments were treated with End Prep Enzyme Mix for end repairing, 5’ Phosphorylation and dA-tailing in one reaction, followed by a T-A ligation to add adaptors to both ends. The size of adaptor-ligated DNA was selected by using AxyPrep Mag PCR Clean-up (Axygen), and fragments of ~410 bp (with the approximate insert size of 350 bp) were recovered. Each sample was then amplified by PCR for 8 cycles using P5 and P7 primers, with both primers carrying sequences which can anneal with flowcell to perform bridge PCR and P7 primer carrying a six-base index allowing for multiplexing. The PCR products were cleaned up using AxyPrep Mag PCR Clean-up (Axygen), validated by an Agilent 2100 Bioanalyzer (Agilent Technologies, PaloAlto, CA, USA), and quantified by Qubit2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Then libraries with different indexes were multiplexed and loaded on an Illumina HiSeq instrument according to manufacturer’s instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2x150 paired-end (PE) configuration; image analysis and base calling were conducted by the HiSeq Control Software (HCS) + OLB + GAPipeline-1.6 (Illumina) on the HiSeq instrument.

**Bioinformatics analysis of Illumina sequencing data**

The software bcf2fastq (version 2.17.1.14) was applied for the process of the source image data from the output of sequencing, to recognize the image of base calling. After the initial quality analysis, (at the time of sequencing, llumina build software determined whether to reserve or discard a read based on the quality of the first 25 nucleotides of each sequencing fragment), the PF data (Pass Filter Data) were collected from the sequencing samples and stored in a FastQ file, which contains the information of sequencing and the corresponding sequencing quality information. The raw data obtained from two strains as mentioned in the TABLE S3. PF data was preprocessed with filtering the low-quality bases, removing adapters and decontamination. The software cutadapt (version 1.9.1) used to remove the sequences of adaptors, polymerase chain reaction (PCR) primers, remove sequences with N base ratio > 10%, and keep the reads with minimum length of 75 bp (TABLE S4, TABLE S5). The alignment software BWA (version 0.7.12) used for alignment the clean data to the reference genome sequence. The statistical result of alignment was collected, including the number of reads from the reference genome, average depth, and coverage. BWA alignment software is suitable for short sequences, which can map the sequencing short reads to reference genome correctly, and then generate SAM files containing the information of the alignment of a single read to the reference genome sequence (TABLE S6). Mapping results were processed by Picard (V1.119) to remove duplication.