AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Improved Genome Sequence of Australian Methicillin-Resistant *Staphylococcus aureus* Strain JKD6159

Ryan R. Wick,[a] Louise M. Judd,[b] Ian R. Monk,[b] Torsten Seemann,[b] Timothy P. Stinear[b]

[a]Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Australia
[b]Department of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

ABSTRACT *Staphylococcus aureus* strain JKD6159 represents a prominent community-acquired methicillin-resistant *S. aureus* (MRSA) clone in Australia. Here, we report an improved assembly of the original *S. aureus* JKD6159 genome sequence. By using deep sequencing with multiple technologies combined with carefully curated assembly and polishing, we believe the assembly to contain zero errors.

Staphylococcus aureus strain JKD6159 is a methicillin-resistant clone of *S. aureus* (1) belonging to the sequence type 93 (ST93) lineage, which was first reported in Australia but is also found in Europe and New Zealand (2). This strain was isolated in Australia in 2004 from a patient in whom it caused septicemia and multifocal abscesses (3). While its complete genome sequence was originally published in 2010 (1) (NCBI Assembly accession no. GCF_000144955.1), we have since resequenced and reassembled JKD6159 using modern platforms and bioinformatic tools to produce a genome sequence, which we believe to be free of errors.

The isolate was cultured overnight at 37°C (200 rpm) in Bacto Brain Heart Infusion broth (Becton Dickinson), and DNA was extracted using GenFind V3 according to the manufacturer's instructions (Beckman Coulter) using lysozyme and proteinase K without size selection. We generated 1,831,719 reads (5.59 Gbp, N_{50} of 4.2 kbp) using an R10.4 MinION flow cell by using the SQK-NBD112.96 kit. The reads were basecalled and adapter trimmed with Guppy v6.1.7 (dna_r10.4_e8.1_sup model). We performed quality control (QC) by discarding reads <6 kbp and the worst 10% of reads using Filtlong v0.2.0 (4), resulting in 135,671 reads (1.82 Gbp, N_{50} of 15.2 kbp). We also generated 6,844,242 paired-end 150-bp reads (998 Mbp) on an Illumina NextSeq 500 using a Nextera XT preparation. Illumina QC was performed using fastp v0.23.2 (5) with default parameters.

We assembled the long reads using Trycycler v0.5.3, following the "extra-thorough" instructions in Trycycler's documentation (using Canu v2.3 [6], Flye v2.9 [7], miniasm v0.3/Minipolish v0.1.3 [8, 9], NECAT v20200803 [10], NextDenovo v2.5.0/NextPolish v1.4.0 [11, 12], and Raven v1.8.1 [13]). This produced three circular contigs, which were a 2,818,668-bp chromosome, a 43,131-bp phage, and a 20,730-bp plasmid. We then ran Medaka v1.6.0 (14), which made 19 single base pair changes to the chromosome and no changes to the phage or plasmid. Short-read polishing with Polypolish v0.5.0 (15) made 26 single base pair changes to the chromosome and no changes to the phage or plasmid. We then ran POLCA v4.0.9 (16), which made no changes, followed by FMLRC2 v0.1.7 (17), which changed seven regions of the chromosome, but each was manually assessed in Integrative Genomics Viewer (IGV) v2.13.0 (18), determined to be an introduced error, and rejected. For all tools, default parameters were used except where otherwise noted.

The circular phage sequence was identical to an integrated phage in the chromosome. To verify that there were no differences between the circular and integrated phage sequences, we produced a 100× Oxford Nanopore Technologies (ONT) read set with Filtlong v0.2.0, which was 7,630 reads with an N_{50} of 40.6 kbp (long enough for

Editor Irene L. G. Newton, Indiana University, Bloomington

Address correspondence to Ryan R. Wick, rrwick@gmail.com.

The authors declare no conflict of interest.

Received 1 November 2022
Accepted 2 January 2023
Published 18 January 2023

most reads to uniquely align). We then repeated the entire assembly/polishing process, which produced an identical result to our previous assembly. Since the integrated and circular phage sequences were confirmed to be identical, we removed the redundant circular phage. To verify that no small plasmids were excluded, we performed a short-read-first hybrid assembly using Unicycler v0.5.0 (19) but did not find any additional plasmids. Our final assembly had a 2,818,670-bp chromosome and a 20,730-bp plasmid (pSaa6159) with 32.8% GC content. After annotation with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.2, the chromosome contained 2,701 coding sequences, 59 tRNAs, 19 rRNAs, 3 noncoding RNAs (ncRNAs), and 1 transfer-messenger RNA (tmRNA), and pSaa6159 contained 24 coding sequences.

To verify the assembly's accuracy, we produced R9.4.1 MinION reads (SQK-RBK110.96 kit; 255,545 reads, 2.11 Gbp, N_{50} of 22.4 kbp, generated from the same DNA) and repeated the process with Trycycler v0.5.3, Medaka v1.6.0 (181 changes), Polypolish v0.5.0 (49 changes), and POLCA v4.0.9 (one change verified in IGV), and the result was identical to our R10.4-plus-Illumina assembly. Finally, we assembled the *S. aureus* JKD6159 genome using previously sequenced PacBio RS II reads (20) (628,002 reads, 797 Mbp, N_{50} of 2.4 kbp) with Trycycler v0.5.3 and Quiver v2.3.3 (21) (24 changes), and the result was also identical. The fact that three alternative approaches (R10.4-plus-Illumina, R9.4.1-plus-Illumina, and PacBio RS II) had no discrepancies supports our claim that this *S. aureus* JKD6159 assembly contains zero errors.

Data availability. The revised genome sequence for *S. aureus* JKD6159 has been deposited in GenBank with accession number GCF_000144955.2. Sequencing data are available on SRA (Illumina, accession number SRR21386014; ONT R10.4 raw, accession number SRR21386013; ONT R10.4 basecalled, accession number SRR21386012; ONT R9.4.1 raw, accession number SRR21386011; ONT R9.4.1 basecalled, accession number SRR21386010; PacBio RS II raw, accession number ERR1213694; and PacBio RS II basecalled, accession number SRR21386009) and figshare (https://bridges.monash.edu/articles/dataset/S_aureus_JKD6159_sequencing_data/21007033).

ACKNOWLEDGMENT

This work was funded by the National Health and Medical Research Council (GNT1105525).

REFERENCES

1. Chua K, Seemann T, Harrison PF, Davies JK, Coutts SJ, Chen H, Haring V, Moore R, Howden BP, Stinear TP. 2010. Complete genome sequence of *Staphylococcus aureus* strain JKD6159, a unique Australian clone of ST93-IV community methicillin-resistant *Staphylococcus aureus*. J Bacteriol 192:5556–5557. https://doi.org/10.1128/JB.00878-10.
2. van Hal SJ, Steinig EJ, Andersson P, Holden MTG, Harris SR, Nimmo GR, Williamson DA, Heffernan H, Ritchie SR, Kearns AM, Ellington MJ, Dickson E, de Lencastre H, Coombs GW, Bentley SD, Parkhill J, Holt DC, Giffard PM, Tong SYC. 2018. Global scale dissemination of ST93: a divergent *Staphylococcus aureus* epidemic lineage that has recently emerged from remote northern Australia. Front Microbiol 9:1453. https://doi.org/10.3389/fmicb.2018.01453.
3. Chua KYL, Seemann T, Harrison PF, Monagle S, Korman TM, Johnson PDR, Coombs GW, Howden BO, Davies JK, Howden BP, Stinear TP. 2011. The dominant Australian community-acquired methicillin-resistant *Staphylococcus aureus* clone ST93-IV [2B] is highly virulent and genetically distinct. PLoS One 6:e25887. https://doi.org/10.1371/journal.pone.0025887.
4. Wick RR. 2021. Filtlong. github.com/rrwick/Filtlong. Retrieved 14 August 2022.
5. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
6. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
7. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
8. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152.
9. Wick RR, Holt KE. 2019. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. F1000Res 8:2138. https://doi.org/10.12688/f1000research.21782.1.
10. Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, Wang Y-X, Xing J-F, Huang Z-J, Wang D-P, He L-J, Luo F, Wang J-X, Liu Y-Z, Xiao C-L. 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. Nat Commun 12:60. https://doi.org/10.1038/s41467-020-20236-7.
11. Hu J. 2021. NextDenovo. github.com/Nextomics/NextDenovo. Retrieved 14 August 2022.
12. Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics 36:2253–2255. https://doi.org/10.1093/bioinformatics/btz891.
13. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci 1:332–336. https://doi.org/10.1038/s43588-021-00073-4.
14. Wright C, Wykes M. 2022. Medaka. github.com/nanoporetech/medaka. Retrieved 14 August 2022.
15. Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol 18:e1009802. https://doi.org/10.1371/journal.pcbi.1009802.
16. Zimin AV, Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLoS Comput Biol 16:e1007981. https://doi.org/10.1371/journal.pcbi.1007981.
17. Mak QC, Wick RR, Holt JM, Wang JR. 2022. Polishing *de novo* nanopore assemblies of bacteria and eukaryotes with FMLRC2. bioRxiv. https://doi.org/10.1101/2022.07.22.501182.

18. Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017.
19. 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.
20. Monk IR, Tree JJ, Howden BP, Stinear TP, Foster TJ. 2015. Complete bypass of restriction systems for major *Staphylococcus aureus* lineages. mBio 6:e00308-15. https://doi.org/10.1128/mBio.00308-15.
21. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.