

SUPPLEMENTARY APPENDIX

The need for speed: ultra-rapid high-resolution outbreak analysis in a front-line hospital microbiology laboratory

Author names

Max Bloomfield^{1,2}, Sarah Bakker³, Megan Burton¹, Kristin Dyet³, Alexandra Eustace³, Samantha Hutton¹, Jamaal Jeram¹, Donia Macartney-Coxson³, David J. Winter³, Rhys T. White³

Affiliations(s)

¹Awanui Labs Wellington, Department of Microbiology and Molecular Pathology, Wellington 6021, New Zealand

²Te Whatu Ora/Health New Zealand, Infection Prevention and Control, Capital, Coast & Hutt Valley, Wellington 6021, New Zealand

³Institute of Environmental Science and Research, Health Group, Porirua 5022, New Zealand

Corresponding author and email address

*Corresponding author: Rhys T. White, Institute of Environmental Science and Research, Wellington, New Zealand; Telephone: +64-4-914-0700; E-mail: rhys.white@esr.cri.nz

This file includes the following:

Supplementary Methods.

Supplementary Figure S1. Estimated sequencing depth against computational time for *Staphylococcus aureus* genomic analysis

Supplementary Figure S2. Comparison of phylogenetic analysis for the *Staphylococcus aureus* sequence type (ST)97 outbreak.

Supplementary Figure S3. Estimated sequencing depth against computational time for *Klebsiella variicola* genomic analysis.

Supplementary Figure S4. Comparison of phylogenetic analysis for the *Klebsiella variicola* sequence type (ST)6385 outbreak.

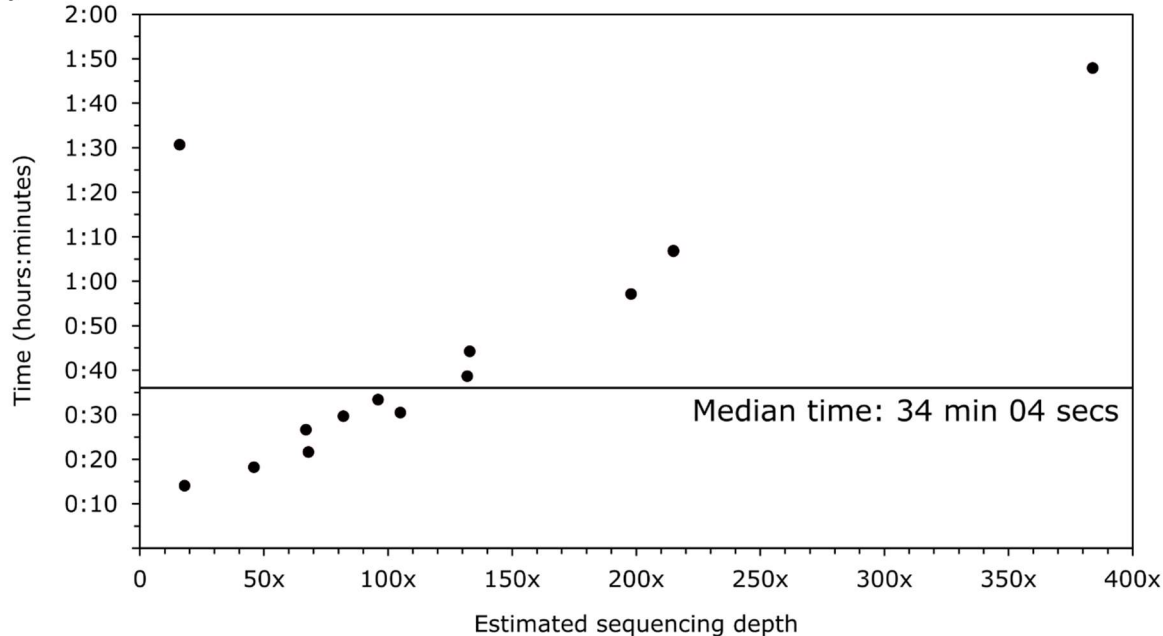
Supplementary Methods

Analyses using established bioinformatic workflow

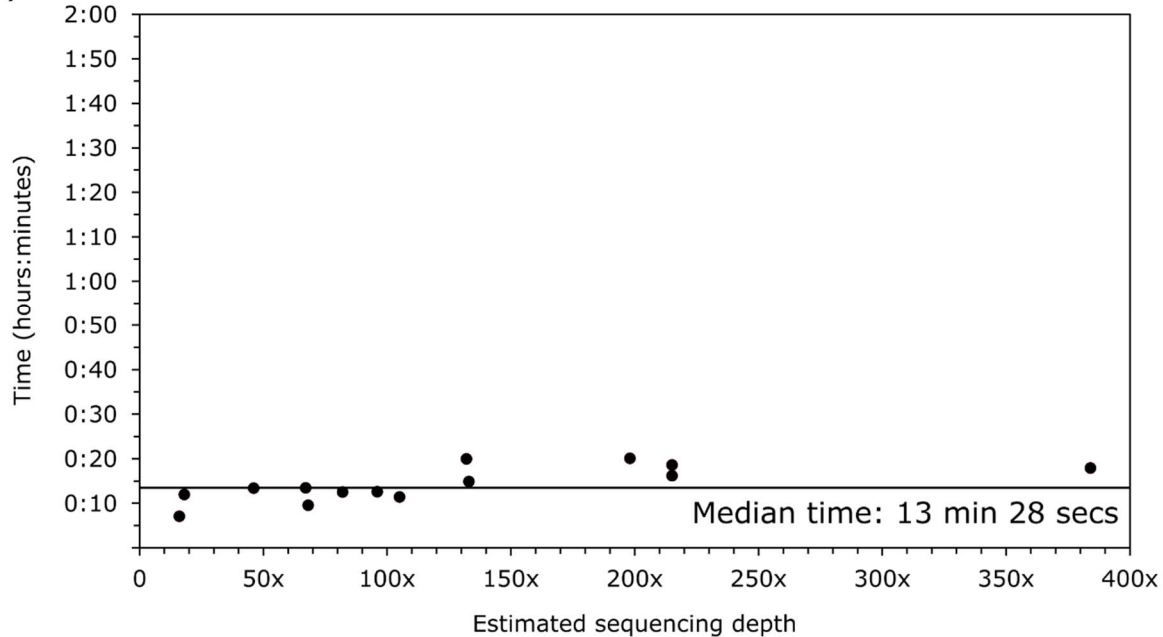
Basecalled nanopore reads underwent quality assessment (NanoStat v1.6.0 (1)) and filtering (NanoFilt v1.1.1 (1)), with 52 nucleotides trimmed from read ends and a Q7 quality threshold. Taxonomic profiling used Kraken v2.1.3 (2) with the NCBI RefSeq Standard database (<https://benlangmead.github.io/aws-indexes/k2>, accessed on 02 January 2025 (3)). The database contained references for archaea, bacteria, human, viruses, plasmids, and the ‘UniVec core’ subset of the UniVec database (a database of vector, adaptor, linker and primer sequences). *De novo* assembly was performed using Flye v2.9.4 (4, 5) (genome size: 2.8 Mb for *Staphylococcus aureus* or 5.5 Mb for *Klebsiella variicola*; three polishing iterations). Read alignment was done with minimap2 v2.28 (6, 7) (configured for long-read data with the ‘map-ont’ preset), followed by polishing to correct single-nucleotide variants (SNVs) and insertions and deletions (INDELs) using racon v1.4.3 (8) and Medaka v1.11.3 (9). After circularisation with Circlator v1.5.5 (10) and final Medaka polishing, assembly quality was assessed using QUAST v5.0.2 (11). Detailed parameters are consistent with our previous work (12). Strain typing included multi-locus sequence typing (MLST) analysis using MLST v2.23 (13) with default settings to query the assemblies against either the *Klebsiella* PasteurMLST sequence definition database (14), or the *S. aureus* typing database (15) hosted on BIGSdb v1.47.0 (16). Virulence genes, acquired antibiotic resistance genes, and mutations conferring resistance to antibiotics were identified using AMRfinderplus v3.12.8 with database version 2024-01-31.1 (17). MOB-suite v3.1.7 was used to identify plasmid replicons (18). Genome assemblies were aligned to create a core-genome alignment using Parsnp v1.7.4 (19). Resulting SNV alignments were used to reconstruct phylogenies. RaxML v8.2.12 (20) built phylogenetic trees using the maximum-likelihood method with GTR-GAMMA correction (optimising 20 distinct, randomised maximum-parsimony trees before adding 1,000 bootstrap replicates).

Supplementary Figures

(A) Manual workflow

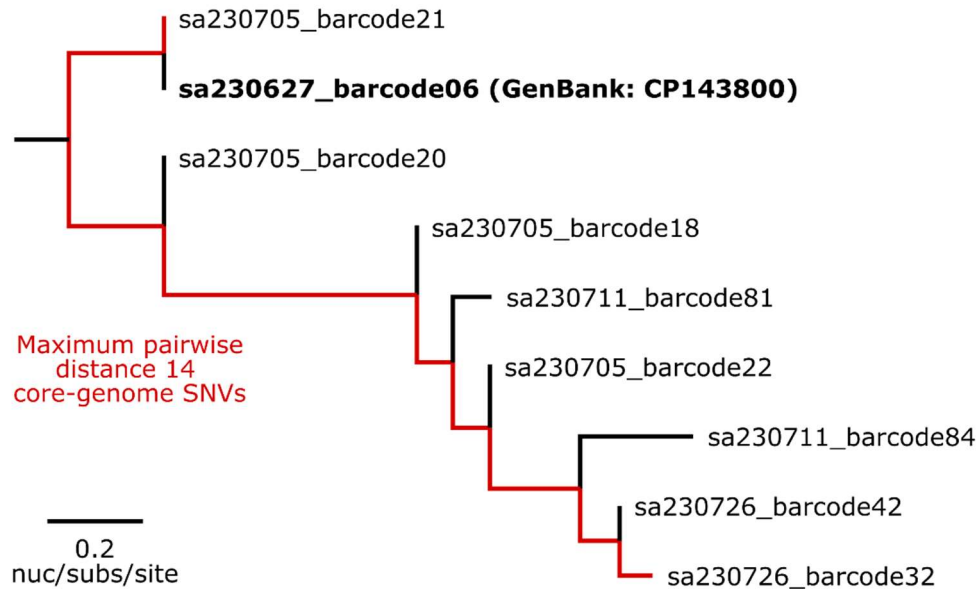


(B) Solu Genomics Platform

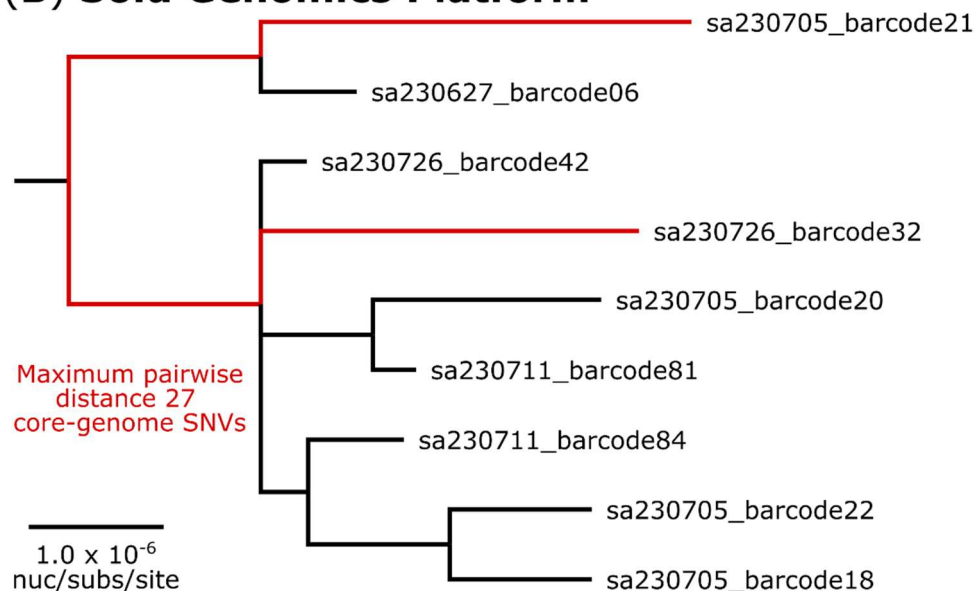


Supplementary Figure S1. Estimated sequencing depth against computational time for *Staphylococcus aureus* genomic analysis. Processing times for the (A) manual bioinformatic analyses, and (B) Solu Genomics cloud-based platform. Horizontal lines indicate median times for each process.

(A) Manual workflow

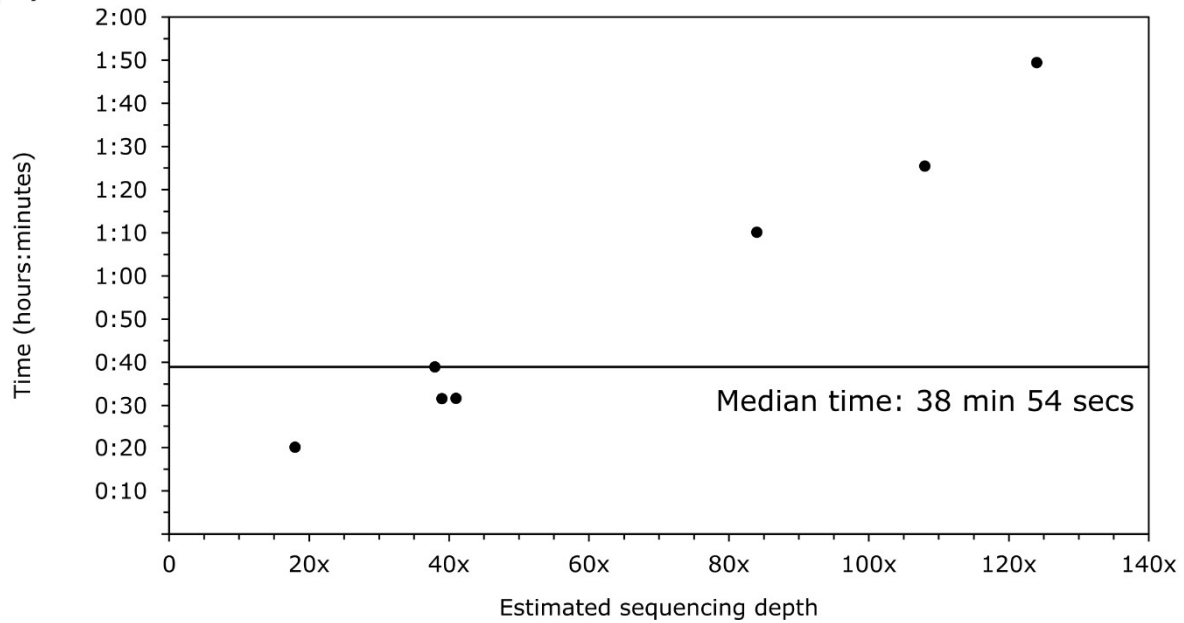


(B) Solu Genomics Platform

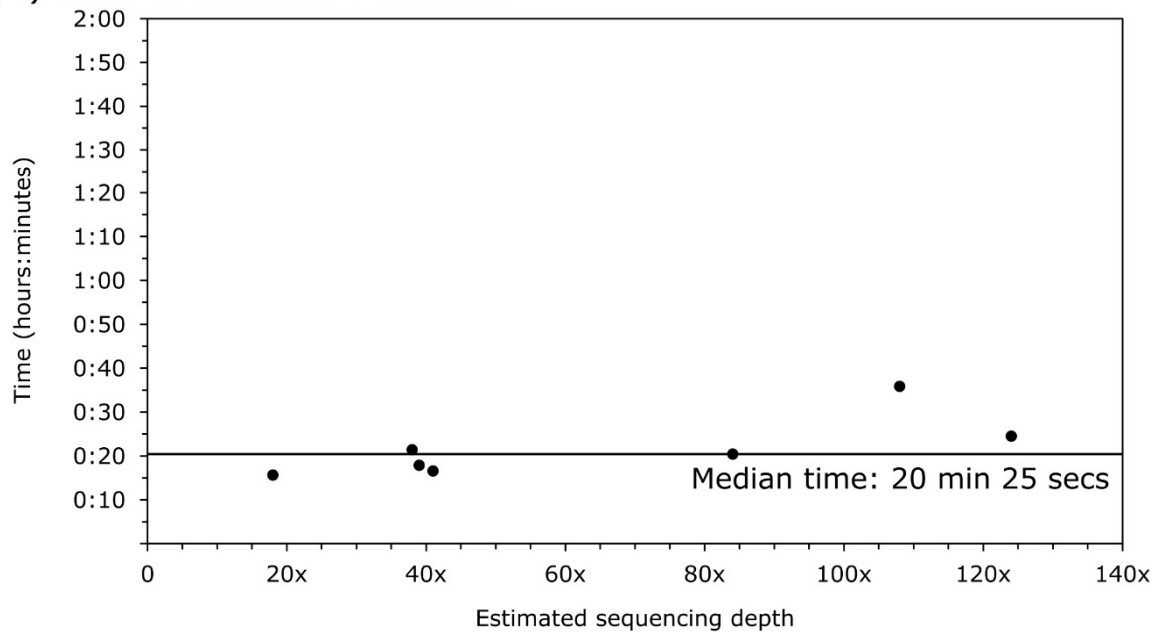


Supplementary Figure S2. Comparison of phylogenetic analysis for the *Staphylococcus aureus* sequence type (ST)97 outbreak. (A) Maximum likelihood generated from the manual bioinformatic analyses. The phylogeny was inferred from 16 core-genome single-nucleotide variants (SNVs) from nine genomes. The multiple sequence alignment was 2,753,159 bp in length. SNVs were called against the 2,753,159 bp chromosome of strain sa230627barcode06 (GenBank: CP143800). (B) Maximum likelihood generated from the Solu Genomics cloud-based platform. The phylogeny was inferred from 48,855 core-genome SNVs from 21 assembled genomes. The multiple sequence alignment was 2,821,361 bp in length. SNVs were called against the 2,821,361 bp chromosome of strain NCTC 8325 (GenBank: CP000253). While the species-level tree is not shown, this zoomed-in view highlights only the ST97 outbreak genomes.

(A) Manual workflow

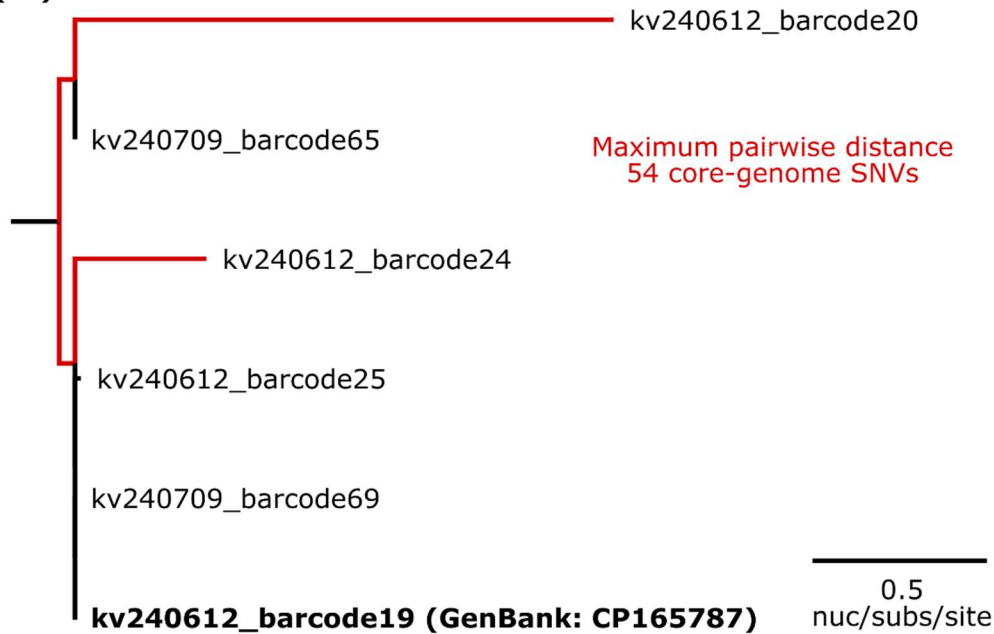


(B) Solu Genomics Platform

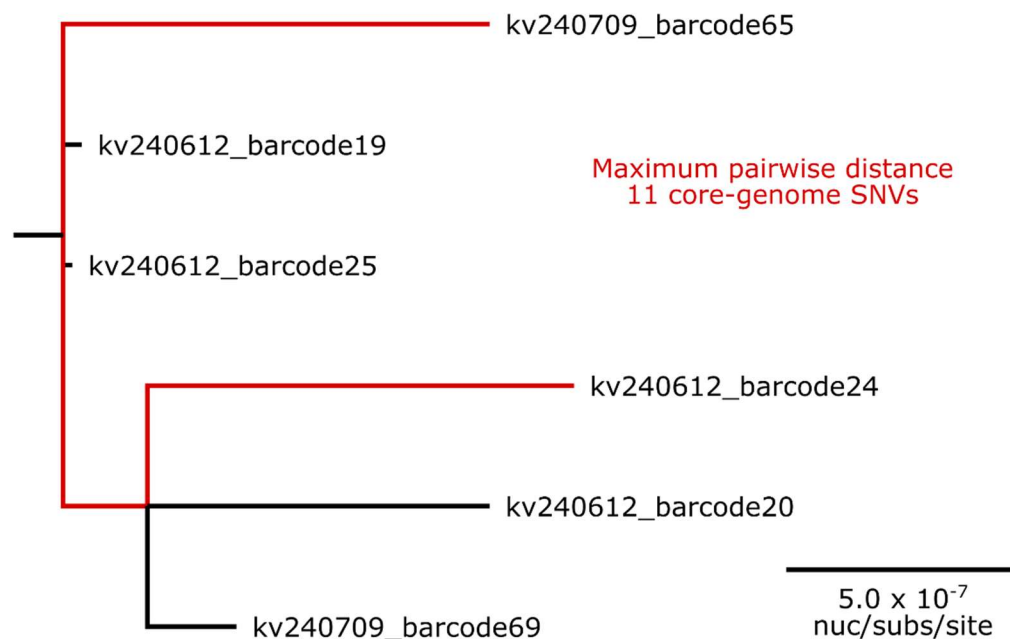


Supplementary Figure S3. Estimated sequencing depth against computational time for *Klebsiella variicola* genomic analysis. Processing times for the (A) manual bioinformatic analyses, and (B) Solu Genomics cloud-based platform. Horizontal lines indicate median times for each process.

(A) Manual workflow



(B) Solu Genomics Platform



Supplementary Figure S4. Comparison of phylogenetic analysis for the *Klebsiella variicola* sequence type (ST)6385 outbreak. (A) Maximum likelihood generated from the manual bioinformatic analyses. The phylogeny was inferred from 57 core-genome single-nucleotide variants (SNVs) from six genomes. The multiple sequence alignment was 5,500,654 bp in length. SNVs were called against the 5,500,654 bp chromosome of strain kv240612_barcode19 (GenBank: CP165787). (B) Maximum likelihood generated from the Solu Genomics cloud-based platform (reference free). The phylogeny was inferred from 87,600 core-genome SNVs from 50 assembled genomes. The multiple sequence alignment was 3,327,801 bp in length. While the species-level tree is not shown, this zoomed-in view highlights only the ST6385 outbreak genomes.

References

1. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666-2669 doi: [10.1093/bioinformatics/bty149](https://doi.org/10.1093/bioinformatics/bty149)
2. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15:R46 doi: [10.1186/gb-2014-15-3-r46](https://doi.org/10.1186/gb-2014-15-3-r46)
3. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, Feolo M, Fingerman IM, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Phan L, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, Wilbur WJ, Yaschenko E, Ye J. 2009. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 37:D5-D15 doi: [10.1093/nar/gkp382](https://doi.org/10.1093/nar/gkp382)
4. Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. 2016. Assembly of long error-prone reads using de Bruijn graphs. *Proc Natl Acad Sci USA* 113:E8396-E8405 doi: [10.1073/pnas.1604560113](https://doi.org/10.1073/pnas.1604560113)
5. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540-546 doi: [10.1038/s41587-019-0072-8](https://doi.org/10.1038/s41587-019-0072-8)
6. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094-3100 doi: [10.1093/bioinformatics/bty191](https://doi.org/10.1093/bioinformatics/bty191)
7. Li H. 2021. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 37:4572-4574 doi: [10.1093/bioinformatics/btab705](https://doi.org/10.1093/bioinformatics/btab705)
8. Vaser R, Sovic I, Nagarajan N, Sikic M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 27:737-746 doi: [10.1101/gr.214270.116](https://doi.org/10.1101/gr.214270.116)
9. Oxford Nanopore Technologies. medaka. Github. <https://github.com/nanoporetech/medaka>.
10. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294 doi: [10.1186/s13059-015-0849-0](https://doi.org/10.1186/s13059-015-0849-0)
11. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072-1075 doi: [10.1093/bioinformatics/btt086](https://doi.org/10.1093/bioinformatics/btt086)
12. White RT, Bakker S, Burton M, Castro ML, Couldrey C, Dyet K, Eustace A, Harland C, Hutton S, Macartney-Coxson D, Tarring C, Velasco C, Voss EM, Williamson J, Bloomfield M. 2024. Rapid identification and subsequent contextualization of an outbreak of methicillin-resistant *Staphylococcus aureus* in a neonatal intensive care unit using nanopore sequencing. *Microb Genom* 10:001273 doi: [10.1099/mgen.0.001273](https://doi.org/10.1099/mgen.0.001273)
13. Seemann T. mlst. Github. <https://github.com/tseemann/mlst>.
14. Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S. 2005. Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *J Clin Microbiol* 43:4178-4182 doi: [10.1128/jcm.43.8.4178-4182.2005](https://doi.org/10.1128/jcm.43.8.4178-4182.2005)

15. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG. 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol* 38:1008-1015 doi: [10.1128/jcm.38.3.1008-1015.2000](https://doi.org/10.1128/jcm.38.3.1008-1015.2000)
16. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3:124 doi: [10.12688/wellcomeopenres.14826.1](https://doi.org/10.12688/wellcomeopenres.14826.1)
17. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, Hoffmann M, Pettengill JB, Prasad AB, Tillman GE, Tyson GH, Klimke W. 2021. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* 11:12728 doi: [10.1038/s41598-021-91456-0](https://doi.org/10.1038/s41598-021-91456-0)
18. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom* 4:e000206 doi: [10.1099/mgen.0.000206](https://doi.org/10.1099/mgen.0.000206)
19. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 15:524 doi: [10.1186/s13059-014-0524-x](https://doi.org/10.1186/s13059-014-0524-x)
20. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313 doi: [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033)