

Sequencing and Bioinformatic Methods

MMWR: Detection of *Paraburkholderia* in Clinical Specimens Associated with Use of Nonsterile Ultrasound Gel for Percutaneous Procedures – United States, Canada, and Israel, May 2023-April 2025

Whole genome sequencing

Study genomes were subject to short-read Illumina sequencing. Sequencing was performed following library preparation according to manufacturer instructions, using different Illumina platforms available at each site. Bioinformatics analysis was performed on raw sequence data as described below. Three genomes were also subject to long-read sequencing using the Oxford Nanopore instrument in Alberta, Canada, as used for generating high-quality closed genomes as reference (see below). Libraries were prepared with a SQK-RBK114-96 (rapid barcoding) and sequenced on an Oxford Nanopore MinION FLO-MIN114, using super-accurate basecalling (v4.3.0) and Dorado v7.6.7. Adapter trimming was ON.

Short read assembly

Raw WGS data (FASTQ files) of the 52 sequenced *Paraburkholderia.fungorum* isolates underwent quality control (QC) using FastQC (v.0.11.9; <https://www.bioinformatics.babraham.ac.uk>), trimming using Trimmomatic (v.0.38) [1], species identification using Kraken2 (v.2.0.8; with the Minikraken v2 database available from <https://benlangmead.github.io/aws-indexes/k2>) [2], down-sampling to 100x depth of coverage (when applicable), and *de novo* assembly using Shovill (v.1.1.0; with spades, v3.14.1; using the parameters ‘--trim’ and ‘--opts “--sc”’) [3]. The resulting genome assemblies underwent QC, species identification and rMLST using checkm2 (v.1.0.1) [4] and gambit (v.1.0.0) [5] and pubMLST[6,7] (last accessed 2025/05/18), respectively. Statistics for short reads (raw reads and assemblies) are provided in the accompanying **Supplementary Tables** file. WGS data were deposited under the bioproject ID: PRJNA1260949 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1260949>).

Hybrid assembly and annotation

Isolate AB-Pf03 was assembled using a short-read-polished long-read assembly approach. Long reads from ONT sequencing were base-called and adapter-trimmed as above. Subsequently, short- and long-reads were used for hybrid assembly using Autocycler (v.0.4.0 with default parameters [8], combining the assemblers Canu (v.2.3) [9]/Flye (v.2.9.6) [10]/Miniasm (v.0.3) [11]/Raven (v.1.8.3) [12] for LR assembly and polypolish (v.0.6.0) [13] for SR polishing). Hybrid assemblies were annotated using Bakta (v.1.8.1; using the light-db)

[14] with default parameters. Results for raw long-read statistics and hybrid assemblies are provided in the accompanying **Supplementary Tables** file.

Phylogenetic analysis

Ad hoc cgMLST with public genomes

Genome assemblies of *P. fungorum* underwent core genome multilocus sequence typing (cgMLST) using chewBBACA (v.3.2.0) [15] [with a Prodigal (v.2.6.3) training file for the NCBI *P. fungorum* reference genome GCF_000961515.1 and including only loci with 100% genome presence]. In the absence of a published typing scheme, an *ad hoc* scheme was created. The first scheme consisted of the 52 study genomes and consisted of 5,219 loci and the second scheme consisted of the same 52 genomes in addition to 20 publicly available genomes downloaded from NCBI, consisting of 2,380 loci. Minimum spanning trees (MST) were generated from the *ad hoc* cgMLST schemes and visualized using GrapeTree [16] (with MSTreeV2).

cgSNPs analysis

Phylogenetic analysis using single-nucleotide polymorphism (SNP) calling was performed with the pipeline SNVPhyl_Nextflow (v.1.0.0 with default parameters; https://github.com/DHQP/SNVPhyl_Nextflow/) using SNVPhyl (v.1.1.0) [17]. The short-read-polished long-read assembly of AB-Pf03 was used as a reference, with plasmid replicons removed. The resulting cgSNPs alignment (snvAlignment.phy; after converting from phylip to fasta using the script from <https://github.com/lterod/phylip2fasta>) and maximum likelihood (ML) phylogeny (phylogeneticTree.newick) were converted to a cgSNPs newick file using pyjar (<https://github.com/simonrharris/pyjar>; last accessed 2025/05/18) [18] and visualized as an MST in GrapeTree (v.1.5.0) [16].

Contact details:

Prof. Jacob Moran-Gilad giladko@post.bgu.ac.il

Paula Snippes Vagnone paula.snippes@state.mn.us

References

1. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114-20. doi:

- 10.1093/bioinformatics/btu170. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.
2. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019 Nov 28;20(1):257. doi: 10.1186/s13059-019-1891-0. PMID: 31779668; PMCID: PMC6883579.
 3. <https://GitHub.com/tseemann/showill>
 4. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods.* 2023 Aug;20(8):1203-1212. doi: 10.1038/s41592-023-01940-w. Epub 2023 Jul 27. Erratum in: *Nat Methods.* 2024 Apr;21(4):735. doi: 10.1038/s41592-024-02248-z. PMID: 37500759.
 5. Lumpe J, Gumbleton L, Gorzalski A, Libuit K, Varghese V, Lloyd T, Tadros F, Arsimendi T, Wagner E, Stephens C, Sevinsky J, Hess D, Pandori M. GAMBIT (Genomic Approximation Method for Bacterial Identification and Tracking): A methodology to rapidly leverage whole genome sequencing of bacterial isolates for clinical identification. *PLoS One.* 2023 Feb 16;18(2):e0277575. doi: 10.1371/journal.pone.0277575. PMID: 36795668; PMCID: PMC9934365.
 6. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, Wimalarathna H, Harrison OB, Sheppard SK, Cody AJ, Maiden MCJ. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology (Reading).* 2012 Apr;158(Pt 4):1005-1015. doi: 10.1099/mic.0.055459-0. Epub 2012 Jan 27. PMID: 22282518; PMCID: PMC3492749.
 7. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 2018 Sep 24;3:124. doi: 10.12688/wellcomeopenres.14826.1. PMID: 30345391; PMCID: PMC6192448.
 8. <https://www.biorxiv.org/content/10.1101/2025.05.12.653612v1>
 9. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* 2017 May;27(5):722-736. doi: 10.1101/gr.215087.116. Epub 2017 Mar 15. PMID: 28298431; PMCID: PMC5411767.
 10. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019 May;37(5):540-546. doi: 10.1038/s41587-019-0072-8. Epub 2019 Apr 1. PMID: 30936562.

11. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*. 2016 Jul 15;32(14):2103-10. doi: 10.1093/bioinformatics/btw152. Epub 2016 Mar 19. PMID: 27153593; PMCID: PMC4937194.
12. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res*. 2019 Dec 23;8:2138. doi: 10.12688/f1000research.21782.4. PMID: 31984131; PMCID: PMC6966772.
13. Bouras G, Judd LM, Edwards RA, Vreugde S, Stinear TP, Wick RR. How low can you go? Short-read polishing of Oxford Nanopore bacterial genome assemblies. *Microb Genom*. 2024 Jun;10(6):001254. doi: 10.1099/mgen.0.001254. PMID: 38833287; PMCID: PMC11261834.
14. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom*. 2021 Nov;7(11):000685. doi: 10.1099/mgen.0.000685. PMID: 34739369; PMCID: PMC8743544.
15. Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, Santos S, Ramirez M, Carriço JA. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microb Genom*. 2018 Mar;4(3):e000166. doi: 10.1099/mgen.0.000166. Epub 2018 Mar 15. PMID: 29543149; PMCID: PMC5885018.
16. Zhou Z, Alikhan NF, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, Carriço JA, Achtman M. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res*. 2018 Sep;28(9):1395-1404. doi: 10.1101/gr.232397.117. Epub 2018 Jul 26. PMID: 30049790; PMCID: PMC6120633.
17. Petkau A, Mabon P, Sieffert C, Knox NC, Cabral J, Iskander M, Iskander M, Weedmark K, Zaheer R, Katz LS, Nadon C, Reimer A, Taboada E, Beiko RG, Hsiao W, Brinkman F, Graham M, Van Domselaar G. SNVPhyl: a single nucleotide variant phylogenomics pipeline for microbial genomic epidemiology. *Microb Genom*. 2017 Jun 8;3(6):e000116. doi: 10.1099/mgen.0.000116. PMID: 29026651; PMCID: PMC5628696.
18. Pupko T, Pe'er I, Shamir R, Graur D. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol Biol Evol*. 2000 Jun;17(6):890-6. doi: 10.1093/oxfordjournals.molbev.a026369. PMID: 10833195.

To contact Content Expert: Jacob Moran-Gilad, MD MPH: giladko@post.bgu.ac.il

Alternate Contact: Paula Snippes Vagnone, MT(ASCP): paula.snippes@state.mn.us