





Complete Genome Sequence of *Streptococcus pneumoniae* Strain BVJ1JL, a Serotype 1 Carriage Isolate from Malawi

 Modupeh Betts,^a Seth Jarvis,^b Aaron Jeffries,^c  Andrea Gori,^a Chrispin Chaguzo,^{d,e,f} Jacqueline Msefula,^g Caroline M. Weight,^a Brenda Kwambana-Adams,^a Neil French,^{f,g} Todd D. Swarthout,^{a,g} Jeremy S. Brown,^h Robert S. Heyderman^{a,g}

^aNIHR Global Health Research Unit on Mucosal Pathogens, Division of Infection and Immunity, University College London, London, United Kingdom

^bUCL Genetics Institute, University College London, London, United Kingdom

^cUniversity of Exeter, Exeter, United Kingdom

^dWellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom

^eDarwin College, University of Cambridge, Cambridge, United Kingdom

^fDepartment of Clinical Infection, Microbiology and Immunology, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom

^gMalawi-Liverpool-Wellcome Trust Clinical Research Programme Blantyre, Blantyre, Malawi

^hDepartment of Respiratory Medicine, Centre for Inflammation and Tissue Repair, University College London, London, United Kingdom

ABSTRACT *Streptococcus pneumoniae* is a leading cause of pneumonia, meningitis, and bacteremia. Serotype 1 is rarely carried but is commonly associated with invasive pneumococcal disease, and in the African “meningitis belt,” it is prone to cause cyclical epidemics. We report the complete genome sequence of *S. pneumoniae* serotype 1 strain BVJ1JL, isolated in Malawi.

Streptococcus pneumoniae, a Gram-positive bacterium, is a leading cause of childhood mortality worldwide (1). At least 100 different capsular serotypes of *S. pneumoniae* have been described (2). Serotype 1 is among the most commonly isolated serotypes from blood or cerebrospinal fluid (CSF) (3, 4).

Strain BVJ1JL was isolated in 2015 from a nasopharyngeal swab (NPS) obtained from a 9-year-old child in Blantyre, Malawi (5). The study protocol was approved by the College of Medicine Research and Ethics Committee, University of Malawi (P.02/15/1677), and the Liverpool School of Tropical Medicine Research Ethics Committee, UK (14.056). The primary NPS, retrieved from storage in skim milk-tryptone-glucose-glycerol (STGG) medium, was plated onto Columbia blood agar (CBA) supplemented with 5% horse blood and incubated overnight at 37°C and 5% CO₂. A single colony was then picked and purified on a fresh CBA plate. *S. pneumoniae* was confirmed by morphology, optochin test, and Gram stain. The capsular type was assessed using a serological latex agglutination test kit (ImmuLex Pneumotest; SSI Diagnostica) and confirmed genomically using PneumoCat (6). DNA was isolated from lawn plate cultures of frozen stocks incubated overnight at 37°C and 5% CO₂ on CBA. The Qiagen Genomic-tip 500/G DNA kit was used to isolate DNA for PacBio sequencing, following the manufacturer’s protocol. Lysis buffer was supplemented with 30 mg/ml lysozyme (Sigma-Aldrich) and 50 units mutanolysin (Sigma-Aldrich). DNA was sheared using a g-TUBE device (Covaris) with a target length of 10 kb, and library preparation was performed according to the protocol “Preparing Multiplexed Microbial Libraries Using SMRTbell Express Template Prep Kit 2.0,” with the Barcoded Overhang Adapter Kit-8A (Pacific Biosciences) at the University of Exeter, UK. The pooled samples were purified and size selected to remove SMRTbell templates of >3 kb using AMPure PB beads (Pacific Biosciences). A 10-h capture using a 1M single-molecule real-time (SMRT) cell was performed on a PacBio Sequel instrument using Sequel v3 chemistry (7). In total, 295,697 reads were generated for BVJ1JL with a read N₅₀ length of

Citation Betts M, Jarvis S, Jeffries A, Gori A, Chaguzo C, Msefula J, Weight CM, Kwambana-Adams B, French N, Swarthout TD, Brown JS, Heyderman RS. 2021. Complete genome sequence of *Streptococcus pneumoniae* strain BVJ1JL, a serotype 1 carriage isolate from Malawi. *Microbiol Resour Announc* 10:e00715-21. <https://doi.org/10.1128/MRA.00715-21>.

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Copyright © 2021 Betts et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Modupeh Betts, modupeh.betts.17@ucl.ac.uk.

Received 14 July 2021

Accepted 4 September 2021

Published 30 September 2021

5,084 bp. The reads were demultiplexed and downsampled using HGAP4 in the SMRTLink v8.0 portal. DNA for Illumina sequencing was isolated using the DNeasy blood and tissue kit (Qiagen) and quantified using Quant-iT PicoGreen double-stranded DNA (dsDNA) kits (Invitrogen) according to the manufacturer's specifications. DNA was fragmented using an EpiSonic sonication system (EpiGenetek), and libraries were constructed using the NEBNext Ultra DNA sample prep master mix kit (New England BioLabs [NEB]) with in-house adapters and barcode tags described at Oxford Genomics Centre, UK (8). The libraries were sequenced on an Illumina HiSeq 4000 instrument as 150-bp paired-end reads. The raw Illumina DNA reads (1,162,988 paired-end reads) were trimmed of low-quality ends and cleaned of adapters using Trimmomatic v0.32 (9). In all, 1,400,000 reads (700,000 paired-end reads) were used for the genome assembly. *De novo* assembly using the PacBio and Illumina reads was performed with the Unicycler v0.4.9b pipeline in bold mode (10), resulting in a single circular contig, as confirmed using Bandage v0.8.1 (11), with a final genome coverage of >500×. The generated assembly was quality assessed using QUAST v5.1.0rc1 (12), and automated annotation was performed using Prokka v1.14.6 (13). The genome assembly was 2,134,668 bp with a G+C content of 39.73%. Prokka v1.14.6 predicted 2,101 coding sequences, 2,238 genes, 12 rRNAs, 59 tRNAs, 59 noncoding RNAs, and 9 riboswitches. Default parameters were used for all software unless otherwise specified.

BVJ1JL belonged to sequence type 5012 (ST5012) and Global Pneumococcal Sequence Cluster 2 (GPSC2) lineage (14). ST5012 is a locus variant of the highly virulent ST217 (15, 16). Genetic analysis predicted susceptibility to penicillin (17) but resistance to chloramphenicol and tetracycline (*cat* and *tetM*, respectively). The macrolide resistance genes *mef* and *ermB* were not detected.

Data availability. The assembled complete genome sequence has been deposited in GenBank under accession number [CP071871](#). The PacBio and Illumina raw reads are available in NCBI Sequence Read Archive (SRA) under accession numbers [SRX10254117](#) and [SRX10254116](#), respectively. The BioProject accession number is [PRJNA695191](#), and the BioSample accession number is [SAMN17602804](#).

ACKNOWLEDGMENTS

This work was funded in part through a UCL Grand Challenges Doctoral Students' Small Grant to M.B. and S.J. M.B., A.G., B.K.-A., T.D.S., and R.S.H. are supported by the NIHR Global Health Research Unit on Mucosal Pathogens using UK aid from the UK Government. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, the Department of Health and Social Care, or the authors' affiliated institutions. The original study that collected the sample in Malawi was funded by a Bill and Melinda Gates Foundation grant (OPP1117653). The PacBio library preparation and sequencing were performed at the University of Exeter, UK, utilizing equipment funded by the UK Medical Research Council (MRC) Clinical Research Infrastructure Initiative (award number MR/M008924/1). We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for Illumina library preparation and sequencing.

REFERENCES

1. Wahl B, O'Brien KL, Greenbaum A, Majumder A, Liu L, Chu Y, Lukšić I, Nair H, McAllister DA, Campbell H, Rudan I, Black R, Knoll MD. 2018. Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type B disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *Lancet Glob Health* 6:e744–e757. [https://doi.org/10.1016/S2214-109X\(18\)30247-X](https://doi.org/10.1016/S2214-109X(18)30247-X).
2. Ganaie F, Saad JS, McGee L, van Tonder AJ, Bentley SD, Lo SW, Gladstone RA, Turner P, Keenan JD, Breiman RF, Nahm MH. 2020. A new pneumococcal capsule type, 10D, is the 100th serotype and has a large *cps* fragment from an oral *Streptococcus*. *mBio* 11:e00937-20. <https://doi.org/10.1128/mBio.00937-20>.
3. Brueggemann AB, Peto TEA, Crook DW, Butler JC, Kristinsson KG, Spratt BG. 2004. Temporal and geographic stability of the serogroup-specific invasive disease potential of *Streptococcus pneumoniae* in children. *J Infect Dis* 190:1203–1211. <https://doi.org/10.1086/423820>.
4. Balsells E, Dagan R, Yildirim I, Gounder PP, Steens A, Muñoz-Almagro C, Mameli C, Kandasamy R, Givon Lavi N, Daprai L, van der Ende A, Trzciński K, Nzenze SA, Meiring S, Foster D, Bulkow LR, Rudolph K, Valero-Rello A, Ducker S, Vestheim DF, von Gottberg A, Pelton SI, Zuccotti G, Pollard AJ, Sanders EAM, Campbell H, Madhi SA, Nair H, Kyaw MH. 2018. The relative invasive disease potential of *Streptococcus pneumoniae* among children after PCV introduction: a systematic review and meta-analysis. *J Infect* 77: 368–378. <https://doi.org/10.1016/j.jinf.2018.06.004>.
5. Swarthout TD, Fronterre C, Lourenço J, Obolski U, Gori A, Bar-Zeev N, Everett D, Kamng'ona AW, Mwalukomo TS, Mataya AA, Mwansambo C, Banda M, Gupta S, Diggle P, French N, Heyderman RS. 2020. High residual

- carriage of vaccine-serotype *Streptococcus pneumoniae* after introduction of pneumococcal conjugate vaccine in Malawi. *Nat Commun* 11: 2222. <https://doi.org/10.1038/s41467-020-15786-9>.
6. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, Fry NK. 2016. Whole genome sequencing of *Streptococcus pneumoniae*: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. *PeerJ* 4:e2477. <https://doi.org/10.7717/peerj.2477>.
 7. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138. <https://doi.org/10.1126/science.1162986>.
 8. Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P. 2013. Improved workflows for high throughput library preparation using the transposome-based Nextera system. *BMC Biotechnol* 13:104. <https://doi.org/10.1186/1472-6750-13-104>.
 9. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 10. 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>.
 11. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31:3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>.
 12. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUILT: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
 13. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
 14. Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, Corander J, Page AJ, Marttinen P, Bentley LJ, Ochoa TJ, Ho PL, Du Plessis M, Cornick JE, Kwambana-Adams B, Benisty R, Nzenze SA, Madhi SA, Hawkins PA, Everett DB, Antonio M, Dagan R, Klugman KP, von Gottberg A, McGee L, Breiman RF, Bentley SD, Global Pneumococcal Sequencing Consortium. 2019. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. *EBioMedicine* 43:338–346. <https://doi.org/10.1016/j.ebiom.2019.04.021>.
 15. Jacques LC, Panagiotou S, Baltazar M, Senghore M, Khandaker S, Xu R, Bricio-Moreno L, Yang M, Dowson CG, Everett DB, Neill DR, Kadioglu A. 2020. Increased pathogenicity of pneumococcal serotype 1 is driven by rapid autolysis and release of pneumolysin. *Nat Commun* 11:1892. <https://doi.org/10.1038/s41467-020-15751-6>.
 16. Bricio-Moreno L, Chaguza C, Yahya R, Shears RK, Cornick JE, Hokamp K, Yang M, Neill DR, French N, Hinton JCD, Everett DB, Kadioglu A. 2020. Lower density and shorter duration of nasopharyngeal carriage by pneumococcal serotype 1 (ST217) may explain its increased invasiveness over other serotypes. *mBio* 11:e00814-20. <https://doi.org/10.1128/mBio.00814-20>.
 17. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Jr, Walker H, Hawkins PA, Tran T, Whitney CG, McGee L, Beall BW. 2016. Penicillin-binding protein transpeptidase signatures for tracking and predicting beta-lactam resistance levels in *Streptococcus pneumoniae*. *mBio* 7:e00756-16. <https://doi.org/10.1128/mBio.00756-16>.