Data in Brief 61 (2025) 111875

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset for the near-complete hybrid genome assembly of *Citrobacter werkmanii* UPMLSS01 isolated from Borneo river and insights into its native antibiotic resistance repertoire

Sui Sien Leong[a,b,*], Han Ming Gan[c]

[a] *Department of Animal Sciences and Fishery, Faculty of Agricultural and Forestry Sciences, Universiti Putra Malaysia Campus Bintulu Sarawak, Bintulu 97008, Sarawak, Malaysia*
[b] *International Institute of Aquaculture and Aquatic Sciences (I-AQUAS), Universiti Putra Malaysia, 0150 Port Dickson, Negeri Sembilan, Malaysia*
[c] *Patriot Biotech Sdn Bhd, Bandar Sunway, 47500, Malaysia*

ARTICLE INFO

Article history:
Received 2 April 2025
Revised 9 June 2025
Accepted 7 July 2025
Available online 11 July 2025

Dataset link: Microbe sample from Citrobacter werkmanii (Original data)

Keywords:
Citrobacter
Nanopore
Illumina
Beta-lactamase

ABSTRACT

The near-complete genome sequence of a *Citrobacter werkmanii* UPMLSS01, isolated from the Kemena River in Bintulu, Sarawak, Malaysia, is presented. Remarkably, this bacterium exhibited resistance to beta-lactam despite no prior exposure, suggesting an innate resistance mechanism in an environment increasingly impacted by anthropogenic pressures. Genome sequencing was performed using both Illumina (2 × 150 bp paired-end) and Oxford Nanopore technologies, yielding a high-quality de novo assembly with a total length of 4994,936 base pairs, a GC content of 52 %, and an N50 of 3.5 Mb. The assembly comprises three contigs, none of which were identified as plasmid or mobile elements. Genomic analysis revealed that the strain belongs to sequence type 948, while also harboring the *blaCMY* gene presumably responsible for the beta-lactam resistance. Notably, this is the first genome of *Citrobacter werkmanii* reported from Borneo, Malaysia, underscoring a significant milestone that provides a robust baseline for future

* Corresponding author.
E-mail address: leongsuisien@upm.edu.my (S.S. Leong).

https://doi.org/10.1016/j.dib.2025.111875

phylogenomic studies and molecular resistance profiling in both environmental and clinical settings.

Specifications Table

Subject	Biology
Specific subject area	Molecular biology, Microbiology and Genomics
Type of data	Raw Reads (fastq format), Assembled Genome in (fasta format), Phylogenetic Tree and Genome Annotation Figures
Data collection	Genomic DNA was extracted using the Zymobiomic DNA extraction kit. Library preparation was performed using the NEBNext Ultra II library preparation kit for Illumina sequencing and the LSK110 library preparation kit for Oxford Nanopore sequencing. Illumina sequencing was carried out on a partial lane of the NovaSEQ6000 using a paired-end 2 × 150 bp configuration. Oxford Nanopore sequencing was performed on a Flongle flowcell, and basecalling was executed using Guppy (v6.5.7) in super accuracy mode. Long-read only *de novo* assembly was initially conducted with Flye (v2.9), followed by error correction of the assembly using Pilon (v1.24) with Illumina data. The assembled genome was subsequently annotated using Bakta (v1.9.1) and screened against the antimicrobial resistance database with Abricate (v1.0.1).
Data source location	Strain Isolation Source: Kemena River, Bintulu, Sarawak, Malaysia
Data accessibility	Repository name: NCBI Bioproject (https://www.ncbi.nlm.nih.gov/bioproject/1061321) Raw Read Link: https://www.ncbi.nlm.nih.gov/sra/SRX23156860 https://www.ncbi.nlm.nih.gov/sra/SRX23156859 Genome Assembly Link: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_035818355.1/

1. Value of the Data

- First genome of an environmental strain of *Citrobacter werkmanii* from Borneo Malaysia with basal antibiotic resistance profile
- Long-read genome assembly enables more accurate comparative genomics of *Citrobacter werkmanii*
- Baseline for future resistance profiling particularly with regards to beta-lactam resistance in this species in Malaysia

2. Background

Sungai Kemena, a major river system in Bintulu, Sarawak, is increasingly impacted by anthropogenic activities such as construction and wastewater discharge, contributing to microbial contamination. Among the bacteria of interest is *Citrobacter werkmanii*, a species that has been reported infrequently in Malaysia, with the last genome sequencing conducted nearly a decade ago by the Institute for Medical Research (IMR) and was recently employed in a comparative genomics analysis by Aguirre-Sánchez et al. [1]. The absence of recent genomic data highlights a significant gap in understanding its environmental presence, genetic diversity, and potential role in contamination. Unlike clinical strains, environmental isolates of *C. werkmanii* experience different selective pressures, with limited direct exposure to antibiotics. Consequently, they may possess a distinct resistome that reflects basal antimicrobial resistance (AMR) rather than resistance acquired due to clinical antibiotic use [2,3]. Investigating these environmental strains provides a valuable opportunity to explore the natural resistome and assess the potential for horizontal gene transfer of resistance elements within aquatic ecosystems. To address this knowl-

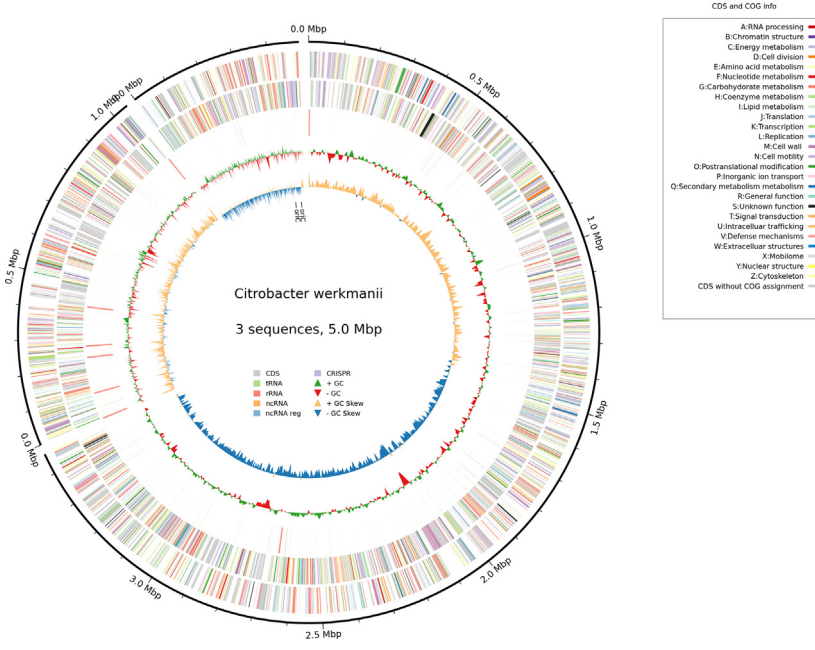

Fig. 1. The near-complete *Citrobacter werkmanii* UPMLSS01 genome spans approximately 5.0 Mbp across three contigs. Clustered rRNA genes (depicted as a red bar in the third ring) are mostly located within the second contig, with some positioned near contig terminal region. From innermost to outermost ring: genome scale in Mbp; GC skew; GC content; non-coding RNA elements; and coding sequences (CDS) on both the forward and reverse strands. CDS are color-coded based on their COG (Clusters of Orthologous Groups) classification.

edge gap, this study aims to sequence the genome of a *C. werkmanii* strain isolated from Kemena River. Genomic analysis will offer insights into its ecological adaptability, genetic composition, and intrinsic AMR mechanisms. Given growing concerns about antibiotic resistance in environmental bacteria, understanding the genomic traits of *C. werkmanii* is crucial for evaluating public health risks and microbial ecology.

3. Data Description

Citrobacter werkmanii strain UPMLSS01 exhibited resistance to Ampicillin (zone of inhibition: 9 mm), Nitrofurantoin (10 mm), and Clavulanic Acid/Amoxicillin (10 mm), based on disk diffusion testing. Whole genome sequencing generated 969.9 Mb and 1.5 Gb data for Nanopore long-read and Illumina short-read, respectively. Specifically, for Nanopore long reads, a total of 678,989 reads with a read length N50 of 1930 bp were generated. Despite a relatively short N50, the long-read-only assembly produced three contigs: JAYMFA010000001.1 (3.46 Mb), JAYMFA010000002.1 (1.04 Mb), and JAYMFA010000003.1 (0.5 Mb), with a total length of 5 Mb and an N50 of 3.5 Mb (Fig. 1). . On the contrary, hybrid assembly utilizing Unicycler v0.5.0 generated 8 contigs with an N50 of 3.14 Mb (Data not shown). As a result, the more contiguous long-read assembly was selected for subsequent analysis. Following error correction, the genome quality as assessed using CheckM v1.2.3 indicates a high completeness of 99.89 % and redundancy of 2.29 %. Genome annotation identified 4769 genes of which 4595 are protein-coding. Assessment of each contig using geNomad v1.11.0 confirmed that none were associated with mobile elements such as plasmids or phages, indicating that all contigs are derived

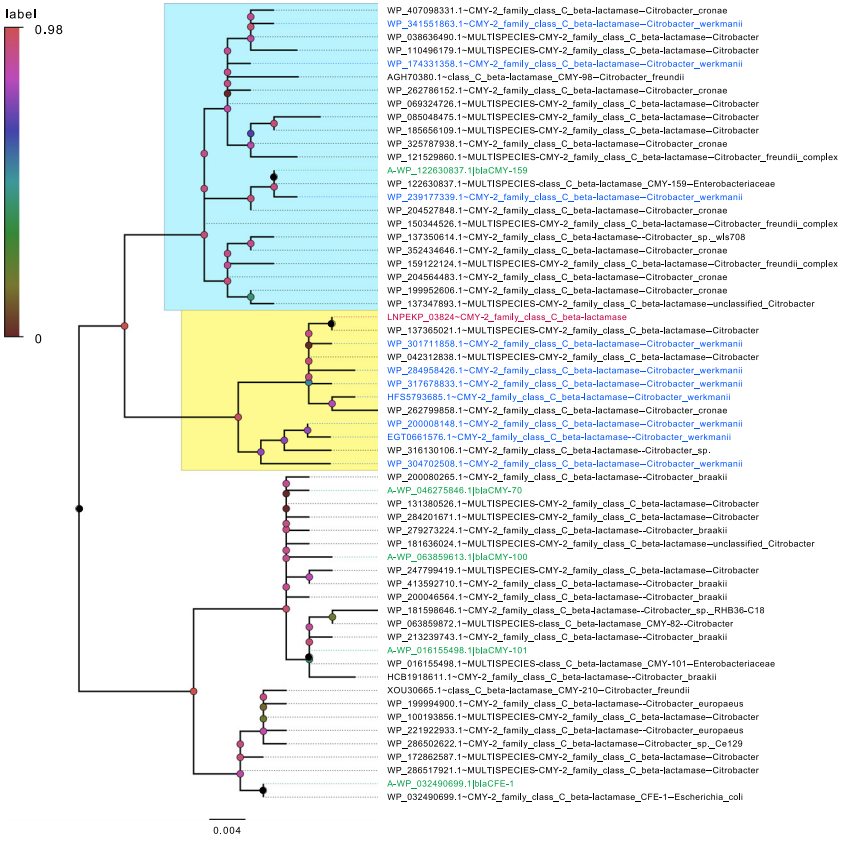

Fig. 2. Maximum Likelihood Phylogenetic Tree of CMY-2 Family Class C Beta-Lactamases
This tree illustrates the evolutionary relationships among CMY-2 family class C beta-lactamases including reference protein variants from the AMRFinder database (green colored taxa) that share high amino acid identity with the class C beta-lactamase from *Citrobacter werkmanii* strain UPMLSS01.. Taxa belonging to *Citrobacter werkmanii* are colored blue, while the strain UPMLSS01 is colored in red. The clade containing *bla-CMY159* type beta-lactamase with the highest protein identity % to UPMLSS01 were highlighted blue while the clade containing UPMLSS01 is shaded yellow. Branch lengths represent the number of substitutions per site, and nodes are colored based on SH-like support values (Shimodaira-Hasegawa support), which indicate the reliability of the tree's branches, with higher values reflecting stronger support for the branching structure.

from the chromosome (data not shown). *In-silico* MLST analysis of strain UPMLSS01 revealed that it belongs to ST-948, with 100 % nucleotide identity across the following loci: *aspC* (95), *clpX* (234), *fadD* (29), *mdh* (262), *dnaG* (23), and *lysP* (266) on contig NZ_JAYMFA010000001.1, located at positions 836,900–837,412; 280,367–280,933; 1781,711–1781,229; 3419,445–3418,897; 3217,939–3217,496 and 2226,554–2227,030 bp, respectively. The *arcA* (15) locus was found on contig NZ_JAYMFA010000002.1, at positions 795,916–796,350 bp. Screening of the predicted protein-coding genes identified a single chromosomally-encoded *bla-CMY* gene located on contig JAYMFA0100000002, spanning positions 503,004 to 504,149 (locus tag: VSO28_RS18930) with the highest nucleotide identity of 94.54 % to *blaCMY-159* (NCBI Accession Number: NG_062243.1; AMRFinderPlus Protein ID: WP_122630837.1). Phylogenetic analysis based on its amino acid sequence reveals that it forms a monophyletic group with several *bla-CMY* proteins predominantly from *Citrobacter werkmanii*. This group then only forms a strongly supported sister clade containing the AMRFinderPlus reference clade containing the bla-CMY-159 protein (Fig. 2, blue clade).

4. Experimental Design, Materials and Methods

4.1. Sample collection and processing

Water samples were collected using a Van Dorn sampler (Lamotte 1077, USA) from designated sites. The four sampling points were S1: (3°11′57.2492″N, 113°2′41.2746″E), S2: (3°11′57.6856″N, 113°2′41.4755″E), S3: (3°11′58.3″N, 113°02′41.2″E) and S4: (3°11′60.0″N, 113°02′40.5"E). Triplicate 1 L samples were stored at 4 °C and transported to the lab. Samples were serially diluted in PBS (Sigma, USA) and plated on nutrient agar (Merck, Germany) with 2 % NaCl, then incubated aerobically at 27 °C for 24 h.

Following incubation, colonies with distinct morphological features: large, circular, convex colonies with smooth edges and pale-gray pigmentation were presumptively identified as *Citrobacter* spp. These colonies were selected based on their consistency with previously reported *Citrobacter werkmanii* characteristics. Selected colonies were subcultured to obtain pure isolates. Conventional microbiological tests, including Gram staining, catalase, and oxidase tests, were performed to support identification. Gram-negative, rod-shaped, catalase-positive, and oxidase-negative isolates were selected for further molecular confirmation.

4.2. Antimicrobial susceptibility testing

Antimicrobial susceptibility testing was performed using the disc diffusion method, and the results were interpreted according to the Clinical and Laboratory Standards Institute (CLSI 2024) guidelines [4]. *Escherichia coli* ATCC 25922 was used as a quality control strain. Ten antibiotic discs (Oxoid, UK) were tested, and the zones of inhibition were measured and categorized as susceptible, intermediate, or resistant based on the CLSI-defined interpretive criteria. *Citrobacter werkmanii* UPMLSS01, resistant to ampicillin, nitrofurantoin, and clavulanic acid/amoxicillin, was selected for whole-genome sequencing.

4.3. Genomic DNA extraction and sequencing

A two-day-old agar culture of *C. werkmanii* UPMLSS01 was scraped, and genomic DNA was extracted using the ZymoBIOMICS DNA kit. For Nanopore sequencing, 1 µg of gDNA was prepared with the LSK110 kit and sequenced on a Flongle flowcell for 24 h. Basecalling was performed with Guppy v5.0.7 (super accuracy mode). For Illumina sequencing, 100 ng of gDNA was fragmented (∼300 bp), prepared with the NEBNext Ultra II kit, and sequenced on a NovaSeq6000 (2 × 150 bp), yielding 1 Gb of data.

4.4. Genome assembly and annotation

Nanopore reads were assembled with Flye v2.9 (“–nano-hq”) [5], polished with Pilon v1.24 [6] using Illumina reads, and analyzed with Seqkit v2.3.0 [7]. A separate hybrid assembly was also performed using Unicycler v0.5.0 [8] utilizing both Illumina short reads and Nanopore long reads (XX). Genome completeness was assessed with CheckM v1.2.3 [9]. In addition, to ensure that the assembled contigs was chromosomally derived and not of plasmid or phage origin, the contigs were analyzed using geNomad v1.11.0 [10] which applies an IGLOO-based neural network to detect non-local sequence motifs directly from nucleotide sequences. Subsequently, genome annotation was performed using Bakta v1.9.1 [11] to generate the annotated input required for circular genome visualization. Antimicrobial resistance genes were identified using Abricate v1.0.1 against ResFinder, CARD, and NCBI databases [12–14]. In silico MLST was performed with the *Citrobacter* spp MLST database (https://pubmlst.org/organisms/citrobacter-spp).

A phylogenetic tree was built using FastTree2 from a MAFFT v7 alignment of Bla-CMY proteins [15,16].

Limitations

Although phenotypic confirmation of antimicrobial resistance (AMR) and the identification of the *blaCMY* gene in the genome have been achieved, further research is necessary to fully validate the functional role of *blaCMY* in conferring resistance.

Ethics Statement

The aurthors have read and follow the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Credit Author Statement

Sui Sien Leong: Conceptualization, Methodology, Software. **Han Ming Gan:** Data curation, Writing, Original draft preparation. **Sui Sien Leong:** Visualization, Investigation. **Han Ming Gan:** Software, Validation. **Sui Sien Leong**, **Han Ming Gan:** Writing- Reviewing and Editing.

Data Availability

Microbe sample from Citrobacter werkmanii (Original data) (NCBI).

Acknowledgements

This work was supported by the Private Grant [grant numbers 6300950].

Declaration of Competing Interest

The authors declare no known competing financial interest.

References

[1] J.R. Aguirre-Sánchez, B. Quiñones, J.A. Ortiz-Muñoz, R. Prieto-Alvarado, I.F. Vega-López, J. Martínez-Urtaza, B.G. Lee, C. Chaidez, Comparative genomic analyses of virulence and antimicrobial resistance in *Citrobacter werkmanii*, an emerging opportunistic pathogen, Microorganisms 11 (2023) 2114, doi:10.3390/microorganisms11082114.
[2] J.F. Delgado-Blas, C. Valenzuela Agüi, E.Marin Rodriguez, C. Serna, N. Montero, C.K.S. Saba, B. Gonzalez-Zorn, Dissemination routes of carbapenem and pan-aminoglycoside resistance mechanisms in hospital and urban wastewater canalizations of Ghana, mSystems 7 (2022) e01019–e01021, doi:10.1128/msystems.01019-21.
[3] S.S. Leong, H. Hamli, E.F. Abdul Rahim, S. Mustafa, Prevalence and diversity of antibiotic resistance heterotrophic bacteria found along the Bintulu rivers, Sarawak, BJRST 13 (2) (2023) 32–41, doi:10.33736/bjrst.5450.2023.
[4] Clinical and Laboratory Standards Institute (CLSI), Performance Standards for antimicrobial susceptibility testingCLSI Supplement M100, 34th ed., Clinical and Laboratory Standards Institute, Wayne, PA, 2024.
[5] M. Kolmogorov, J. Yuan, Y. Lin, P.A. Pevzner, Assembly of long, error-prone reads using repeat graphs, Nat. Biotechnol. 37 (2019) 540–546, doi:10.1038/s41587-019-0072-8.

[6] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C.A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS One 9 (2014) e112963, doi:10.1371/journal.pone.0112963.
[7] W. Shen, S. Le, Y. Li, F. Hu, SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation, PLoS One 11 (2016) e0163962, doi:10.1371/journal.pone.0163962.
[8] R.R. Wick, L.M. Judd, C.L. Gorrie, K.E. Holt, Unicycler: resolving bacterial genome assemblies from short and long sequencing reads, PLoS Comput. Biol. 13 (2017) e1005595, doi:10.1371/journal.pcbi.1005595.
[9] A. Chklovski A, D.H. Parks, B.J. Woodcroft, G.W. Tyson, CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning, Nat Methods 20 (2023) 1203–1212, doi:10.1038/s41592-023-01940-w.
[10] A.P. Camargo, S. Roux, F. Schulz, M. Babinski, Y. Xu, B. Hu, P.S.G. Chain, S. Nayfach, N.C. Kyrpides, Identification of mobile genetic elements with geNomad, Nat. Biotechnol. 42 (2024) 1303–1312, doi:10.1038/s41587-023-01953-y.
[11] O. Schwengers, L. Jelonek, M.A. Dieckmann, S. Beyvers, J. Blom, A. Goesmann, Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification, Microb. Genomics 7 (2021) 000685, doi:10.1099/mgen.0.000685.
[12] V. Bortolaia, R.S. Kaas, E. Ruppe, M.C. Roberts, S. Schwarz, V. Cattoir, A. Philippon, R.L. Allesoe, A.R. Rebelo, A.F. Florensa, L. Fagelhauer, T. Chakraborty, B. Neumann, G. Werner, J.K. Bender, K. Stingl, M. Nguyen, J. Coppens, B.B. Xavier, S. Malhotra-Kumar, H. Westh, M. Pinholt, M.F. Anjum, N.A. Duggett, I. Kempf, S. Nykäsenoja, S. Olkkola, K. Wieczorek, A. Amaro, L. Clemente, J. Mossong, S. Losch, C. Ragimbeau, O. Lund, F.M. Aarestrup, ResFinder 4.0 for predictions of phenotypes from genotypes, J. Antimicrob. Chemother. 75 (2020) 3491–3500, doi:10.1093/jac/dkaa345.
[13] B.P. Alcock, A.R. Raphenya, T.T.Y. Lau, K.K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A.-L.V. Nguyen, A.A. Cheng, S. Liu, S.Y. Min, A. Miroshnichenko, H.-K. Tran, R.E. Werfalli, J.A. Nasir, M. Oloni, D.J. Speicher, A. Florescu, B. Singh, M. Faltyn, A. Hernandez-Koutoucheva, A.N. Sharma, E. Bordeleau, A.C. Pawlowski, H.L. Zubyk, D. Dooley, E. Griffiths, F. Maguire, G.L. Winsor, R.G. Beiko, F.S.L. Brinkman, W.W.L. Hsiao, G.V. Domselaar, A.G. McArthur, CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database, Nucleic Acids Res. 48 (2020) D517–D525, doi:10.1093/nar/gkz935.
[14] M. Feldgarden, V. Brover, N. Gonzalez-Escalona, J.G. Frye, J. Haendiges, D.H. Haft, M. Hoffmann, J.B. Pettengill, A.B. Prasad, G.E. Tillman, G.H. Tyson, W. Klimke, AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence, Sci. Rep. 11 (2021) 12728, doi:10.1038/s41598-021-91456-0.
[15] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2 – Approximately maximum-likelihood trees for large alignments, PLoS One 5 (2010) e9490, doi:10.1371/journal.pone.0009490.
[16] K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform, Nucleic Acids Res. 30 (2002) 3059–3066, doi:10.1093/nar/gkf436.