Data in Brief 62 (2025) 111983

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Comprehensive genomic insights into *Mixta calida* isolated from the faecal sample of a tuberculosis patient in India

Archana Pant[a,*], Jayalakshmi Srinivasan[a], Sureshkumar Venkadesan[b], Sonali Sarkar[c], Senbagavalli Prakash Babu[c], Debasisa Mohanty[b], Bhabatosh Das[d], G. Aneeshkumar Arimbasseri[a]

[a] *Molecular Genetics Lab, National Institute of Immunology, New Delhi, India*
[b] *Bioinformatics Centre, National Institute of Immunology, New Delhi, India*
[c] *Dept. of PSM, Jawaharlal Institute of Postgraduate Medical Education & Research, Puducherry, India*
[d] *Functional Genomics Laboratory, Translational Health Science and Technology Institute, Faridabad, India*

ARTICLE INFO

Article history:
Received 1 May 2025
Revised 6 August 2025
Accepted 6 August 2025
Available online 20 August 2025

Keywords:
Mixta calida
Whole genome sequencing
Antimicrobial resistance
Mobile genetic elements
Gut microbiome

ABSTRACT

Mixta calida, formerly known as *Pantoea calida*, is a motile Gram-negative, facultatively anaerobic bacterium with coccoid rod morphology. Although previously considered non-pathogenic, emerging case studies indicate its potential role in causing serious infections, including bacteraemia, meningitis, sepsis, and implant-associated infections. This study presents the first whole-genome sequence of *M. calida* of Indian origin, isolated from the stool sample of a tuberculosis patient undergoing treatment. Sequencing was performed using the Illumina NextSeq 2000 and Oxford Nanopore PromethION platforms. The genomic data provides valuable insights into the antimicrobial resistance traits and mobile genetic elements of the bacterium, contributing to a deeper understanding of its pathogenic potential.

* Corresponding author.
E-mail address: archanapant@nii.ac.in (A. Pant).

https://doi.org/10.1016/j.dib.2025.111983

Specifications Table

Subject	Biology
Specific subject area	Microbiology, Bacteriology, Genomics
Type of data	Whole-genome sequence, genome assembly, annotated genes, tables, and figures
Data collection	*Mixta calida* was isolated from the stool sample of an Indian tuberculosis patient (age-49/male). The sample was collected during the culmination of ongoing 6 months anti-TB treatment. The genomic DNA was extracted using a GenElute™ Bacterial Genomic DNA Kits from Sigma-Aldrich according to the manufacturer's protocol. Then sequenced using the NextSeq 2000 and Nanopore PromethION Sequencers for both short and long reads. Genome assembly was performed using Unicycler tool. Gene prediction was performed using the Rapid Annotation Subsystem Technology (RAST).
Data source location	Host: *Homo sapiens* Place: Stool Sample Collection- JIPMER, Puducherry, India; Bacterial Isolation: NII, Delhi, India Month and years: March 2023
Data accessibility	Data is deposited in the following repositories and is publicly available. Repository name: NCBI & IBDC Data identification number(s): NCBI- BioProject: PRJNA1249924; SRA: SRR33108476, SRR33108477; Assembly: JBNFWA000000000 & IBDC Study accession- INRP000327 Direct URL to data: NCBI: Genome Assembly: https://www.ncbi.nlm.nih.gov/nuccore/JBNFWA000000000.1/ SRA: https://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=1249924 IBDC: https://ibdc.dbtindia.gov.in/inda/completeStudyDetailsById?studyid=INRP000327
Related research article	NA

1. Value of the Data

- First genome sequence of *Mixta calida* isolated from the faecal sample collected from the gut of Indian human subject.
- Isolated from the stool sample of patient exposed to a long-term antibiotic treatment.
- Enables research on the genomic adaptations of environmental bacteria to host-associated niches.
- Highlights potential virulence and antimicrobial resistance factors with significant public health implications.
- Provides a reference for future comparative and functional genomics.

2. Background

Antimicrobial resistance (AMR) poses a significant global health challenge, further intensified by mobile genetic elements (MGEs) that facilitate horizontal gene transfer (HGT) among closely or distantly related microbial species, including pathogenic bacteria. This process facilitates to the emergence of multidrug-resistant "superbugs" that can evade current antibiotic therapies. TB disease compromises the immune system and disrupts homeostasis and its long-term, broad-spectrum antibiotic treatment disrupts the normal microbial ecology, allowing opportunistic organisms to colonize niches left vacant by beneficial microbes. *M. calida* is a species of Gram-negative, facultatively anaerobic, motile, coccoid rod-shaped bacteria in the Erwiniaceae family. Traditionally regarded as an environmental bacterium, it has recently been recognized as an emerging opportunistic pathogen, with documented infections in immunocompromised individuals [1,2]. This shift in its clinical significance makes it a compelling subject for genomic

investigation. The isolation of *M. calida* from the stool sample of a TB patient suggests it could be the part of gut microflora as rare taxon. Although, we observed a few virulence associated traits in this strain, however, its AMR genes compilation makes it a potential reservoir for HGT of these AMR genes. In this study, we present the whole genome sequence of *M. calida*, offering critical insights into the genomic and metabolic traits, as well as its potential AMR and virulence factors.

3. Data Description

Initial identification of *M. calida* was confirmed through 16S rRNA gene sequencing using a capillary sequencing platform, with sequence validation performed via NCBI-BLAST (v2.15.0) analysis. Whole-genome sequencing was conducted using both Illumina NextSeq 2000 and Oxford Nanopore PromethION platforms to generate high-quality short (total reads: 1,06,27,418) and long (total reads: 6,30,541) reads, respectively. The assembled genome of the bacterium comprises nine contigs with a total size of 4,665,461 bp and a GC content of 56.1 %. The assembly has an N50 value of 3,865,698 bp and L50 of 1 (Fig. 1a). Genome completeness (99.95 %) and contamination (1.80 %) were within the desired range with 627x genome coverage. Annotation through RAST identified 4622 coding sequences and 105 RNAs spanning 338 functional subsystems (Fig. 2, Suppl Table 1). Notably enriched subsystems include protein metabolism (171 ORFs), carbohydrate metabolism (220 ORFs), amino acid metabolism (241 ORFs), and cofactor/vitamin biosynthesis (133 ORFs) (Suppl Table 2). The genome assembly submission to NCBI provided the PGAP annotation with 4352 CDS and 117 RNAs.

AMR gene analysis using Resistance Gene Identifier (RGI v6.0.3), ABRicate-CARD (v1.0.1), and Centre for Genomic Epidemiology (CGE)-ResFinder identified genes conferring resistance to different antibiotic classes like fluoroquinolones, macrolides, tetracyclines, phenicols, and aminoglycosides etc (Fig. 1b, Suppl Table 3). Antimicrobial resistance profiles were confirmed through antimicrobial susceptibility testing (AST) with 20 different antibiotics (Table 1). Based on CLSI/EUCAST guidelines where applicable, the isolate exhibited resistance to four distinct classes of antibiotics, classifying it as a multidrug-resistant (MDR) organism. The key AMR genes identified

Table 1
Antimicrobial susceptibility testing (AST) data and interpretation.

Antibiotic Discs (Conc.)	Zone Observed (mm)	Interpretation
PEN (10)	NZ	Resistant
AMP (10)	8 (MC)	Resistant
CARB (100)	7 (MC)	No information available
IMI (10)	27	Sensitive
CEF (30)	IZ 24 / OZ 29	Sensitive
AZT (30)	IZ 25 / OZ 30	Sensitive
POLY B (300)	14	No information available
SXT (25: 23.7+1.2)	17*	Could be sensitive due to sulfamethoxazole
RIF (5)	13	No information available
LIN (30)	15 (MC)	No information available
LEV (5)	19	Intermediate
CIP (5)	21	Resistant
CHL (30)	NZ	Resistant
TET (5)	13 (MC)	Intermediate
DOX (30)	15	Sensitive
ERY (15)	6	No information available
STR (10)	NZ	Resistant
AMI (30)	19	Sensitive
KAN (30)	21	Sensitive
SPEC (100)	18**	No information available

NZ: no zone; IZ: inner zone; OZ: outer zone; *: zone with fine transparent layer; **: opaque zone; MC: mutant colonies (few colonies growing within the periphery of zone of inhibition).

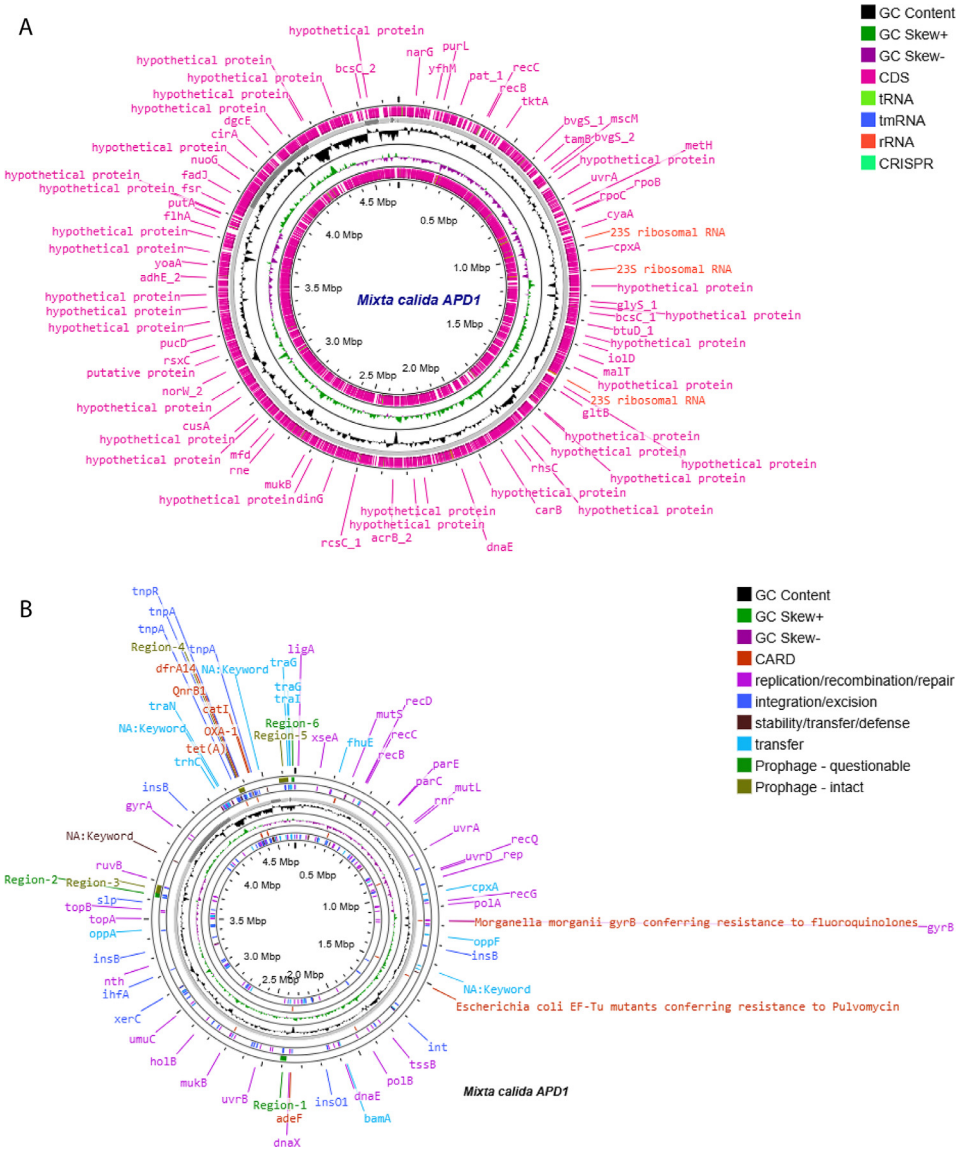

Fig. 1. Genome maps of *M. calida* are built using Proksee-Genome Analysis (https://proksee.ca/). a) Genomic features are represented from the outer to the inner circle, Circle 1: illustrates Prokka annotated Coding sequences (CDS) with tRNA, rRNA, and tmRNA (+); Circle 2: Backbone (Contigs); Circle 3: represents the GC content; Circle 4: displays the GC skew (G-C) / (G + C); Circle 5: Prokka Annotation (-). b) AMR and MGE features are represented from the outer to the inner circle, Circle 1- illustrates phage regions from PHASTEST; Circle 2- represents MGE annotation from mobileOG-db (+); Circle 3- AMR features from CARD RGI (+); Circle 4: Backbone (Contigs); Circle 5: GC Content; Circle 6: GC Skew; Circle 7: CARD RGI Results (-); Circle 8: mobileOG-db Annotation (-).

in strain APD1, i.e. *blaOXA-1, catB3, qnrB1, tetA, aadA,* and *dfrA14* etc are critical AMR drivers, posing significant public health risks by spreading through clinical, environmental, and animal reservoirs. The co-existence of these genes in single isolates (e.g., *Escherichia coli* ST744 with *blaOXA-1, dfrA14, qnrB1, tetA*) creates MDR profiles that resist multiple first-line antibiotics, forcing clinicians to use last-resort drugs with higher toxicity and cost. The variable expression of

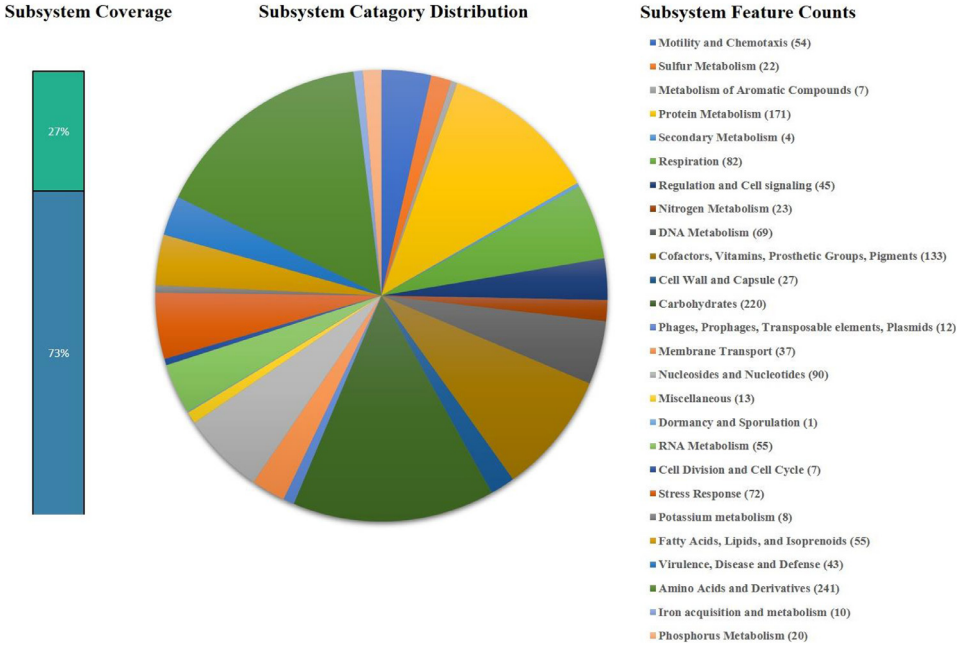

Fig. 2. Analysis of the protein encoding genes (pegs) of the *M. calida* whole-genome sequence assigned to subsystems categories according to the RAST server via SEED viewer. The bar on the left presents the percentage of pegs assigned to subsystems (green) and the pegs which could not be placed into any subsystem (blue). The pie chart in the centre depicts the subsystem category distribution. The coloured categories on the right indicate the subsystem feature counts are generated using subsystem file.

genes (e.g., *qnrB1*'s low-level resistance) complicates susceptibility testing, leading to potential treatment failures. These resistance genes enhance bacterial pathogenicity by enabling survival in hostile environments (e.g., antibiotic-treated patients or environments), increasing their persistence, and facilitating the spread of virulent clones. Often carried on MGEs, these genes are associated with high-risk clones (e.g., *E. coli* ST131, *K. pneumoniae* ST405). AMR genes such as *blaOXA-1, catB3, qnrB1, tetA*, and *dfrA14* were found co-localized with MGEs on Contig 3 (Fig. 1b).

MGEs, including insertion sequences, composite transposons, plasmid regions, and phage-associated regions, in APD1 strain were detected using ABRicate-PlasmidFinder (v1.0.1), ISFinder, PHASTEST, and CGE tools (Suppl Table 4).

Interestingly, only two virulence-associated genes, *terC* and *flhD*, were identified through virulence factor database (ABRicate-VFDB v1.0.1) and CGE-MGE, suggesting limited pathogenic potential (Suppl Table 5).

Whole-genome-based phylogenetic analysis revealed that the bacterium displayed the highest similarity (average nucleotide identity of 99.6 % using FastANI-Proksee) to *Mixta calida* strain DE0300 (Fig. 3a). It clustered more closely with species from the genera *Pantoea, Mixta*, and *Erwinia*—all members of the family Erwiniaceae—than with representatives of the families Yersiniaceae and Enterobacteriaceae (Fig. 3a). We also conducted the whole-genome-based phylogenetic analysis of strain APD1 alongside fourteen globally sourced, human-associated isolates of *M. calida* (retrieved from NCBI), including reference and type strains (DE0300 and DSM22759^T, respectively), to elucidate evolutionary relationships and genetic divergence among the isolates. Strain APD1 exhibited close phylogenetic relatedness to strains DE0300, DSM22759^T, B021323 (isolated from urine) and four strains isolated from infant faeces while showing greater divergence from the remaining isolates (Fig. 3b).

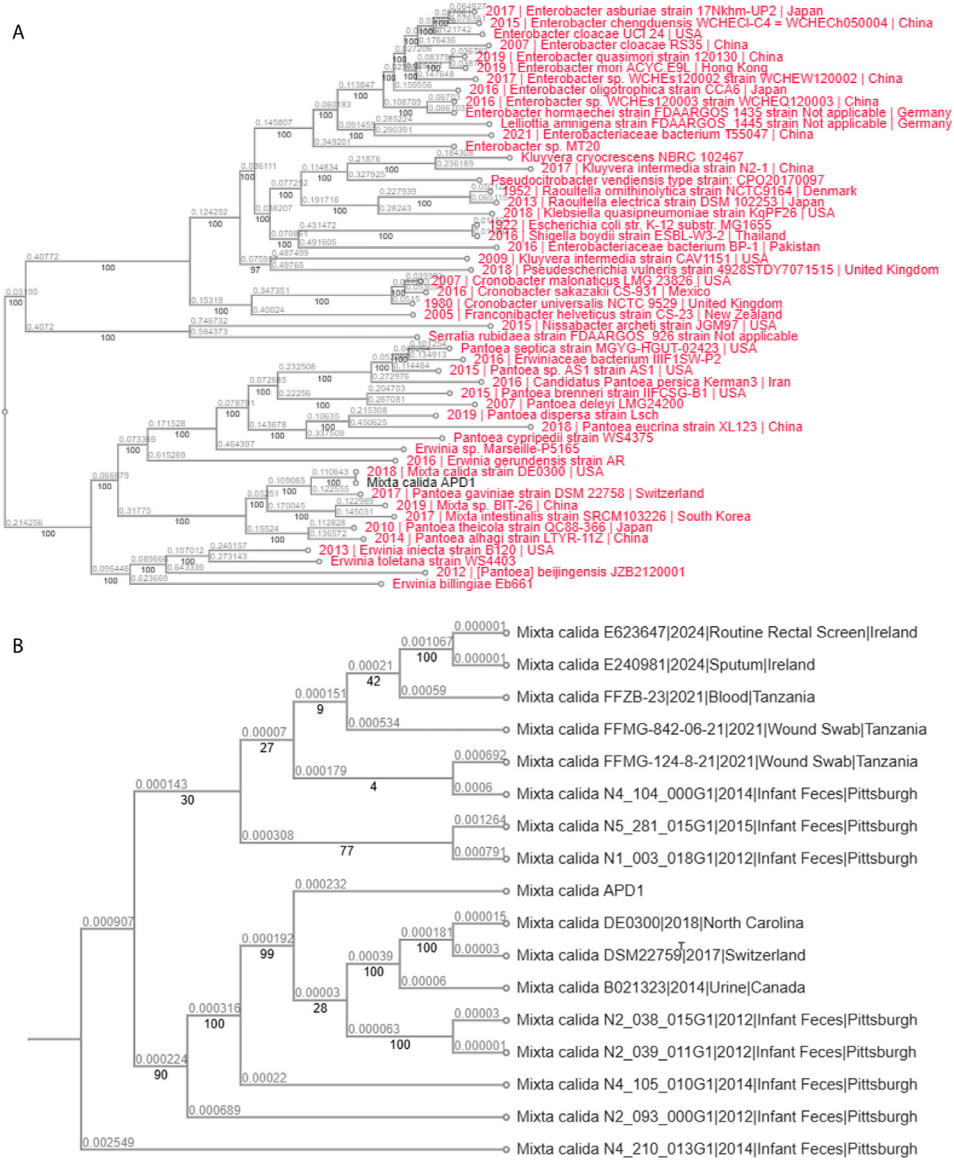

Fig. 3. Whole-genome based phylogenetic tree constructed using the BV-BRC Codon Tree pipelineshowing the phylogenetic relationship of *M. calida* APD1 with a) reference genomes, b) with genomes of global human isolates of *M. calida*. The trees were generated by aligning amino acid and nucleotide sequences from selected PGFams using MUSCLE and BioPython. Tree construction used RAxML with 100 bootstrap replications.

Comparative analysis of AMR genes, virulence factors, and mobile genetic elements (MGEs) revealed that *M. calida* strain APD1, isolated from the human stool sample, harbors a significantly higher number of AMR genes and MGE regions than the reference strain DE0300 and type strain DSM 22759^{T} —both of which were isolated from non-human sources (Fig. 4; Suppl. Tables 6 & 7).

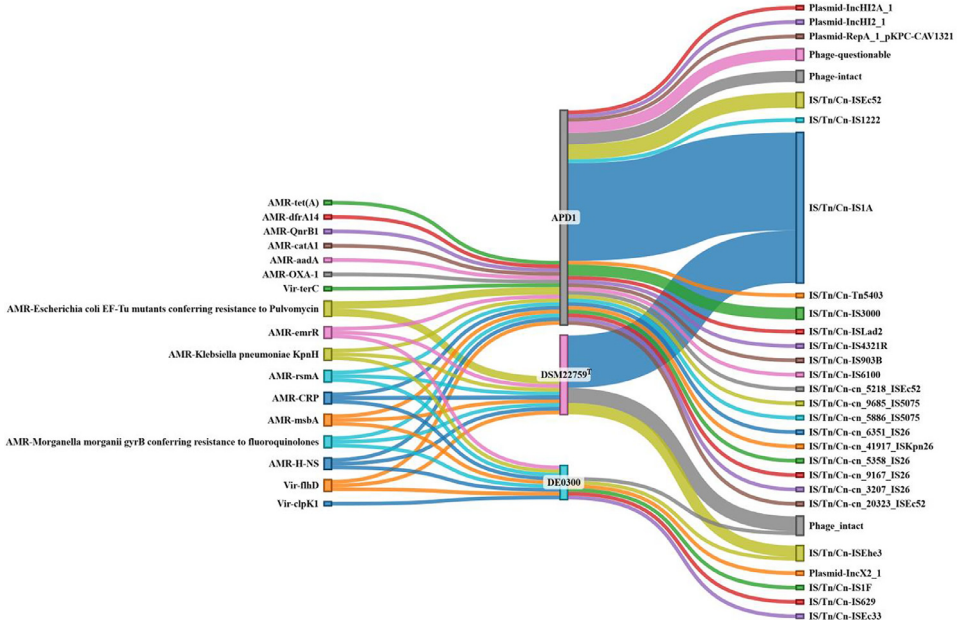

Fig. 4. A Sankey diagram illustrating the distribution of antimicrobial resistance (AMR) genes, virulence factors (left side), and mobile genetic elements (MGEs, right side) in *M. calida* strain APD1 in comparison with two reference genomes, *M. calida* DE0300 and DSM 22759^T, available on the NCBI database. The visualized features include AMR genes (AMR), virulence genes (Vir), plasmid-associated regions (Plasmid), prophage regions (Phage), insertion sequences (IS), transposons (Tn), and composite transposons (Cn). The figure was generated using SankeyMATIC (https://sankeymatic.com/).

4. Experimental Design, Materials and Methods

4.1. Sample collection, bacterial isolation and DNA extraction

The tuberculosis patient (49/ Male) was recruited at JIPMER, Puducherry. Stool sample was collected at sixth month (culmination) of treatment, immediately refrigerated, and transported to the laboratory (JIPMER and then to NII, Delhi), where it was stored at −80 °C until processing. Approximately 500 mg faecal sample was resuspended in 1 ml of Luria Bertani (LB) broth, and further diluted serially in the same broth, and plated on McConkey agar (MCA) plate. Bacterial cells were spread over the surface of the plates using few glass beads of 2.5 mm. The plates were then incubated for 16 to 18 hours at 37 °C. From various morphologically distinct colonies, *M. calida* were pink (lactose-fermenting), tiny and pin-pointed in morphology. A single colony was picked and re-streaked on MCA plate again to confirm morphological distinction and cultured in 5 ml LB broth to prepare glycerol stock. Incubation was performed overnight at 37 °C. 2 ml bacterial culture of *M. calida* in LB broth was used to extract the genomic DNA using GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich) as per manufacturer's protocol. The concentration and quality of extracted DNA were evaluated using a MULTISKAN Sky (Thermo Scientific, USA) and agarose gel electrophoresis. The extracted gDNA was then used for library construction. For 16S rRNA region amplification for capillary sequencing, PCR was performed using Q5® High-Fidelity DNA Polymerase (New England Biolabs) with forward primer (27F): AGAGTTTGATCCTGGCTCAG and reverse primer (1391R): GACGGGCGGTGWGTRCA. Amplification was done in 50 μl reaction volume with 1–10 ng of template DNA and following the reaction conditions: 98 °C-2 min

(1 cycle), 98 °C-10 s, 58.5 °C-15 s, 72 °C-1 min (30 cycle), 72 °C-5 min (1 cycle). Purified the PCR amplicon from agarose gel using QIAquick Gel Extraction Kit (QIAGEN) followed by capillary sequencing using 27F primer.

4.2. Library preparation and sequencing

Whole genome sequencing was performed using a high-throughput Illumina NextSeq 2000 sequencing platform (Illumina, Inc., USA) for short reads and PromethION Sequencer (PromethION24 and Data Acquisition Unit, Oxford Nanopore Technology, UK) for long reads. For Illumina, approximately 100 ng of DNA was used for genomic DNA fragmentation and assembling the library. The QIAseq FX DNA Library Kit was used for pair-end sequencing (Illumina, Inc., USA). Library selection was done using enzyme fragmentation. For Oxford Nanopore sequencing technology (ONT), the DNA samples were cleaned and end-repaired using 1X AMPure XP beads (A63880, Beckmann Coulter, USA) and NEBNext® UltraTM II End Repair/dA-Tailing Module (E7546, New England Biolabs, MA, USA), respectively. Native barcode ligation was performed with Blunt/TA Ligase Master Mix (M0367, New England Biolabs, MA, USA) using barcodes from SQK-NBD114.96 kit (Oxford Nanopore Technology, UK). Barcoded samples were cleaned with 1X AMPure XP beads and quantified using QubitTM dsDNA HS assay kit (Q32851, Thermo Fisher Scientific, USA) on QubitTM 4 Fluorometer (Q33238, Thermo Fisher Scientific, USA).

4.3. Genome assembly and annotation

Genome assembly was performed by integrating short and long reads using the Unicycler tool (v0.5.1), resulting in the generation of nine contigs, after removing the contigs less than 300 bps. The assembled genome quality was analyzed using CheckM tool (v1.2.3) for completeness and contamination at the genome level. Gene prediction was performed using the Rapid Annotation Subsystem Technology (RAST; http://rast.nmpdr.org/), applying the FIG's "Subsystem Approach" via the SEED viewer [3]. Circular genome maps were created using the Proksee - Genome Analysis from the resulting assembled genome (https://proksee.ca/) [4]. NOTE: We have also re-analyzed the sequencing data using the Flye assembler (v2.9.5-b1801), followed by polishing with Illumina short reads using Pilon (v1.24). Upon comparing the results, both assembly pipelines produced highly comparable outcomes in terms of overall genome size (~4.66 Mb) and other assembly quality metrics.

4.4. Phylogenetic analysis

The reference and representative genomes were searched using BV-BRC Similar Genome Finder service. The genomes from human associated global isolates of *M. calida* were identified from NCBI (https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=665913). Whole-genome based phylogenetic tree was constructed via the BV-BRC Codon Tree pipeline, aligning amino acid and nucleotide sequences from selected PGFams using MUSCLE and BioPython (https://www.bv-brc.org/app/PhylogeneticTree). Tree construction used RAxML with 100 bootstrap replications [5].

4.5. Screening of AMR and virulence genes, MGEs and IS elements

These genes/ regions were predicted using various tools (discussed in data description section). The AMR, virulence genes and MGE regions were filtered and selected based on identity ≥80 % and coverage ≥ 65 %. The IS elements were selected based on identity ≥90 %

and alignment length ≥500. Only those IS elements with 100 % of original length were selected after filtering. The phages were selected if the completeness score was ≥ 70. Links for the online available tools used: ABRicate (https://github.com/tseemann/abricate), ISFinder (https://isfinder.biotoul.fr/), CGE (https://www.genomicepidemiology.org/services/), RGI (https://github.com/arpcard/rgi), PHASTEST (https://phastest.ca/submissions/new). The average nucleotide identity between strain DE0300 & APD1 was identified using FastANI-Proksee (https://proksee.ca/tools/fastani; v1.1.0). All tools were executed with default parameters unless stated otherwise.

4.6. Antibiotic sensitivity test

The bacterium was revived from glycerol stock on LB Agar plate. Around 2–5 colonies were picked and dissolved in 0.9 % NaCl. The suspension was then matched with 0.5 Mc Farland solution. 20 antibiotics were taken (Pen, Amp, Carb, Ime, PolyB, Cef, Azt, SXT, Rif, Lin, Ami, Lev, Kan, Spec, Chl, Tet, Dox, Ery, Cip and Str) and placed on center with equal distance between each disc over the bacterial culture spread plate of Mueller Hinton Agar (MHA). The plates were then incubated at 37 °C for 12–14 hours. Observations were made the next day and zones were measured using CLSI (2021) and/or EUCAST (2023) guidelines for Enterobacterales. *E. coli* ATCC 25922 was used as a control.

Data Availability

This Whole Genome Shotgun project has been deposited at:

1) DDBJ/ENA/GenBank under the accession JBNFWA000000000. The version described in this paper is version JBNFWA010000000.
2) IBDC Study Accession: INRP000327.

Limitations

Limited virulence genes identified; further experimental validation is necessary for functional analysis.

Ethics Statement

The study used sample under appropriate institutional approvals.

Credit Author Statement

AP designed the study. AP & JS performed experiments. GA & DM provided research facilities and server space. SS & SB provided the patient's stool sample. AP analyzed the data and wrote the paper. AP and SV performed the computational work and developed figures. AP, GA, BD & DM edited and finalized the article.

Acknowledgements

This study received financial support from Department of Science & Technology (DST), Innovation in Science Pursuit for Inspired Research (INSPIRE)-Faculty Fellowship (Grant Number

DST/INSPIRE/04/2020/001246), Govt. of India (GOI). AP received support from DST-INSPIRE Faculty Fellowship, GOI. AP acknowledges BRIC-National Institute of Immunology, for providing space, and research facilities. We acknowledge the members of NGS-Laboratory, NII for performing Illumina sequencing.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2025.111983.

References

[1] S. Van Hees, S. Keulemans, K. Vanden Driessche, A.-S. Schoonjans, T. Goegebuer, A. Lemmens, A first case of Mixta calida bacteremia and meningitis in a 5-week old child, Infect. Dis. (2024), doi:10.1080/23744235.2024.2391022.

[2] L. Blairon, J. Everaert, B. Baillon, S. Collignon, F. Coenen, R. Cupaiolo, I. Beukinga, M. Tré-Hardy, Mixta calida, not only an environmental bacterium, but a potential opportunistic threat: case report of an osteitis with skin necrosis and mini review of the literature, New Microb. New Infect (101524) (2024), doi:10.1016/j.nmni.2024.101524.

[3] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology, BMC Genom. 9 (2008) 75, doi:10.1186/1471-2164-9-75.

[4] J.R. Grant, E. Enns, E. Marinier, A. Mandal, E.K. Herman, C.Y. Chen, M. Graham, G. Van Domselaar, P. Stothard, Proksee: in-depth characterization and visualization of bacterial genomes, Nucleic Acid. Res 51 (2023) W484–W492, doi:10.1093/nar/gkad326.

[5] R.D. Olson, R. Assaf, T. Brettin, N. Conrad, C. Cucinell, J.J. Davis, D.M. Dempsey, A. Dickerman, E.M. Dietrich, R.W. Kenyon, M. Kuscuoglu, E.J. Lefkowitz, J. Lu, D. Machi, C. Macken, C. Mao, A. Niewiadomska, M. Nguyen, G.J. Olsen, J.C. Overbeek, B. Parrello, V. Parrello, J.S. Porter, G.D. Pusch, M. Shukla, I. Singh, L. Stewart, G. Tan, C. Thomas, M. VanOeffelen, V. Vonstein, Z.S. Wallace, A.S. Warren, A.R. Wattam, F. Xia, H. Yoo, Y. Zhang, C.M. Zmasek, R.H. Scheuermann, R.L. Stevens, Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR, Nucleic Acid. Res 51 (2023) D678–D689, doi:10.1093/nar/gkac1003.