

Genome description of a potentially novel species of *Rossellomorea* sp. strain H39_3 isolated from the Hindon River, India

Chakresh Kumar¹, Nirupama Saini¹, Anwesha Ghosh² and Punyasloke Bhadury^{1,2,*}

Abstract

In October 2024, a putative novel species, belonging to the genus *Rossellomorea*, designated as strain H39_3, was isolated using Luria–Bertani medium from surface water representing station H39, located on the Hindon River, in close proximity to Gautam Buddha Nagar in Uttar Pradesh, India. The 16S rRNA sequence of this isolate showed 100% identity to *Rossellomorea marisflavi* from the International Nucleotide Sequence Database Collaboration (INSDC) DNA databases (GenBank/DDBJ/ENA). Whole-genome sequencing was undertaken using long-read sequencing with Oxford Nanopore Technologies (ONT) chemistry on the MinION platform, followed by genome annotation against the NCBI Reference Sequence Database (RefSeq) and Genome Taxonomy Database (GTDB) databases. The genome is ~4.46 Mb in size, with a G+C content of 48.64 mol%. The low average nucleotide identity (92.23%) and digital DNA–DNA hybridization (46.1%) values showed affiliation with the reference strain *R. marisflavi*, indicating the isolate as a potential novel species. Functional analysis of the draft genome of this isolate revealed an array of genes, including the presence of *arsC* (arsenate reductase), the assimilatory nitrate reduction pathway and the ability to degrade polysaccharides. The presence of nitrogen metabolizing genes such as *nirB* (nitrite reductase subunit B), along with the ability to break down complex forms of carbon, offers the potential of this strain for application in bioremediation of contaminated river ecosystems.

DATA AVAILABILITY

The whole-genome sequence was uploaded to National Center for Biotechnology Information (NCBI) under the accession no. SAMN47866444. The corresponding raw sequencing reads are available in Sequence Read Archive (SRA) under accession no. SRR34443858.

INTRODUCTION

In the family Bacillaceae (Phylum: Firmicutes), the genus *Rossellomorea* was established in 2020 and includes species that were previously categorized as *Bacillus* [1]. It includes previously described *Bacillus* species that can be reliably identified from all other Bacillaceae species by the presence of exclusively shared conserved signature indels in their protein sequences [2]. With the genus *Rossellomorea*, half of the species have been isolated from diverse marine ecosystems and are heterotrophic, halophilic or halotolerant aerobes [2–4]. The genus *Rossellomorea* comprises Gram-positive, rod-shaped bacteria [5]. It shares a close relationship with the genus *Lactobacillaceae* of lactic acid-producing bacteria. To date, valid names representing six species within

Access Microbiology is an open research platform. Pre-prints, peer review reports, and editorial decisions can be found with the online version of this article. Received 19 April 2025; Accepted 18 December 2025; Published 19 May 2026

Author affiliations: ¹Integrative Taxonomy and Microbial Ecology Research Group, Department of Biological Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur-741246, Nadia, WB, India; ²Centre for Climate and Environmental Studies, Indian Institute of Science Education and Research Kolkata, Mohanpur-741246, Nadia, WB, India.

***Correspondence:** Punyasloke Bhadury, pbhadury@iiserkol.ac.in; pbhadury@gmail.com

Keywords: arsenic; bioremediation; polysaccharide; river; *Rossellomorea*.

Abbreviations: AAI, average amino acid identity; ANI, average nucleotide identity; BGCs, biosynthetic gene clusters; CARD, Comprehensive Antibiotic Resistance Database; CAZymes, carbohydrate active enzymes; CEs, carbohydrate esterases; dDDH, digital DNA–DNA hybridization; FastANI, fast alignment-free computation of whole-genome average nucleotide identity; FESEM, field emission scanning electron microscopy; GBDP, Genome BLAST Distance Phylogeny; gDNA, genomic DNA; GGDC, Genome Genome Distance Calculator; GHs, glycoside hydrolases; INSDC, International Nucleotide Sequence Database Collaboration; LB, Luria–Bertani; PLs, polysaccharide lyases; RGI, resistance gene identifier; SWT, surface water temperature; TYGS, Type (Strain) Genome Server.

Three supplementary figures are available with the online version of this article.

001042.v4 © 2026 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

this genus have been reported. These include the type strain *Rossellomorea aquimaris* and *Rossellomorea marisflavi*, which were first isolated from a tidal flat [5]. Besides, *Rossellomorea arthrocnemi* has been discovered from plant roots [2], *Rossellomorea vietnamensis* from Vietnamese fish sauce [6], *Rossellomorea oryzaecorticis* from rice husks [7] and *Rossellomorea pakistanensis* from a salt mine [8]. Given the versatile distribution of this genus, there are opportunities to further isolate members of this genus from contaminated freshwater ecosystems and investigate their bioremediation potential for restoration of these ecosystems.

METHODS

Sampling and isolation

Surface water sample was collected from the Hindon River, near Gautam Buddha Nagar, Ghaziabad, Uttar Pradesh, India (28° 24' 52.6" N; 077° 28' 53.0" E), during the month of October 2024. The collection was done in a 50 ml sterilized centrifuge tube (Tarsons, India). The Hindon River is reeling from anthropogenic forcings such as continuous flow of untreated municipal sewage and industrial effluents. The *in situ* environmental parameters, such as surface water temperature (SWT), pH, dissolved oxygen, total dissolved solids, electrical conductivity, water depth and water velocity, were also measured at the time of sampling. SWT was measured using a digital thermometer (Digi-sense RTD Meter 20250–95, Cole-Parmer, United States of America), and pH was recorded using an EcoTestr pH Meter (Eutech Instrument Pte Ltd., Singapore). The water sample was immediately transferred to the laboratory under controlled temperature conditions. Subsequently, 20 µl of the collected surface water was spread on Luria–Bertani (LB) agar plates. The pH of the culture medium was adjusted to 7.7 to mimic the recorded *in situ* environmental parameters. These plates were incubated at 37 °C for overnight growth, mimicking the SWT measured at the time of collection. An uninoculated blank control plate was also set up in parallel to monitor for any potential contamination introduced during the experimental procedure. Further, a pure bacterial isolate strain named H39_3 was obtained by streak-plate isolation. The colony from the mother plate was re-streaked twice, first to confirm purity and then to obtain the final pure culture for downstream analysis.

Morphological characteristics

The isolated strain H39_3 was examined under a bright-field microscope (Bx53, Olympus, Japan; 1,000× magnification) to determine the cell shape and Gram staining characteristics. Furthermore, field emission scanning electron microscopy (FESEM) of strain H39_3 was undertaken to confirm the morphological characteristics. Briefly, a fresh culture of strain H39_3 was set up in LB broth, and subsequently, the cells were harvested from an overnight culture by centrifugation at 5,000 r.p.m. The cell pellet was washed using PBS. The washed cell pellet was soaked in a 2.5% glutaraldehyde solution (prepared in PBS) for 8 h at 4 °C. This was followed by subsequent post-fixation with 1% osmium tetroxide. Stepwise dehydration was carried out using ethanol gradients of 10%, 30%, 50%, 70%, 80%, 90% and finally absolute ethanol. Finally, the cells were drop-cast, desiccated and coated with gold:palladium (20:80); FESEM was performed using a Zeiss SUPRA55VP (Carl Zeiss AG, Germany).

Genomic DNA extraction

A modified version of the existing phenol–chloroform procedure was used to extract genomic DNA (gDNA) from the growing culture [9, 10]. Briefly, the cells were pelleted for 5 min at 6,000 r.p.m. A lysis buffer (100 µl of 50 mM Tris-HCl, 100 µl of 20 mM EDTA, 100 µl of 400 mM NaCl, 100 µl of 750 mM sucrose and 10 µl of 10% SDS; Merck, India) was added to the resulting pellet, and it was then incubated for 30 min at 50 °C in a water bath. Proteinase K (10 mg ml⁻¹; 5 µl) (HiMedia Laboratories, India) was added, and the mixture was incubated for 3–4 h at 55 °C. Furthermore, 10 µl of lysozyme (10 mg ml⁻¹; Merck, India) was added, and the mixture was incubated for 1–2 h at 37 °C. To separate the DNA into the aqueous phase, phenol and chloroform (Merck, India) were added in a 1:1 ratio; the mixture was gently stirred for 5 min, left to stand for 15 min, and then centrifuged for 12 min at 12,000 r.p.m. Subsequently, 60 µl of 3 M sodium acetate (Merck, India) and 750 µl of 100% ethanol (Merck, Germany) were added for overnight incubation; gDNA was precipitated at –20 °C. The centrifugation was undertaken at 12,000 r.p.m. for 12 min to precipitate gDNA and was subsequently dissolved in 20 µl of Tris-HCl (10 mM, pH 8.0). The extracted gDNA was visualized by gel electrophoresis (1% agarose gel) using ethidium bromide (EtBr) fluorescent dye (Merck, India). The NanoDrop 2000c (Thermo Fisher Scientific, United States of America) was used to quantify the yield of gDNA.

Genotypic identification

Initially, the 16S rRNA amplicon was obtained from the gDNA of strain H39_3 using eubacterial primers (Fc27 and Rc1492; [11]). The PCR reaction consisted of 0.25 µl of DNA DreamTaq polymerase (5 U µl⁻¹; Thermo Fisher Scientific, USA), 2.5 µl of 10× DreamTaq buffer, 2.5 µl of 2 mM deoxynucleoside triphosphates (dNTPs), 2 µl of 25 mM magnesium chloride (MgCl₂), 0.5 µl of each 25 µM primer, 0.25 µl of DNA template, 0.5 µl of BSA (1 mg ml⁻¹) and nuclease-free water to make a final volume of 25 µl. The PCR conditions were as follows: initial denaturation at 95 °C for 10 min, 30 cycles of 95 °C for 1 min, 55 °C for 1 min, 72 °C for 3 min and final extension at 72 °C for 20 min. The PCR product was purified using the

Qiagen Gel Purification Kit (Qiagen, United States of America) following the manufacturer's instructions. The purified amplicon was sequenced using Sanger sequencing chemistry with the BigDye Terminator Cycle Sequencing Kit (Thermo Fisher Scientific, United States of America) on an ABI Prism 3500 Genetic Analyzer (Thermo Fisher Scientific, USA). BioEdit (v7.7.1.0) was used to verify the quality of the obtained raw chromatograms. Sequence identification was performed using the BLASTn algorithm against the GenBank/DDBJ/ENA databases (INSDC DNA databases). For constructing the 16S rRNA phylogeny, six near-full-length 16S rRNA sequences representing all species of the genus *Rosellomorea* were downloaded from nucleotide databases (GenBank/ENA/DDBJ), along with 12 additional sequences representative of the Bacillaceae family. These 18 sequences, along with the outgroup *Arthrobacter agilis*, were subsequently aligned with the MUSCLE algorithm [12] in MEGA version 11 [13]. Before constructing the phylogenetic tree, the best model was selected using jModelTest [14]. Based on the best model with the lowest Bayesian information criterion (BIC) value substitution model (TIM3+I+G), a maximum-likelihood tree was constructed in IQTree [15]. The robustness of the clades in the phylogenetic tree was determined using bootstrap analysis based on 1,000 replicates [16].

For whole-genome sequencing, extracted gDNA was quantified using a Qubit 3.0 fluorometer with the 1× dsDNA HS (High Sensitivity) Assay kit (Thermo Fisher Scientific, United States of America). The Native Barcoding Kit (SQK-NBD114.24, Oxford Nanopore Technologies, United Kingdom) was used to sequence the whole genome. Using long-read sequencing chemistry, the entire genome library was sequenced on the MinION platform (Oxford Nanopore Technologies, United Kingdom). Quality check was performed on the obtained raw reads (Q>8). Adapters were eliminated using Porechop (v0.2.0), and a further quality report of the data was obtained using NanoPlot [17]. Flye (v2.9.3; [18]) was used to assemble the *de novo* genome. The programme CheckM [19] was used to measure genome completeness and contamination. QUAST (v5.2.0 [20]) was used to evaluate assembled genome quality. Prokka (v1.14.6 [21]) was used to annotate the draft genome [21]. Proksee was used to generate and visualize the circular map of the entire genome [22].

Genome data processing

The Type (Strain) Genome Server (TYGS) (<https://tygs.dsmz.de>) was used to conduct whole-genome sequence-based phylogeny [23]. TYGS uses the Genome BLAST Distance Phylogeny (GBDP) method to measure genomic distances. Based on inter-genomic distances, the evolutionary tree was inferred with branch support from 100 replicates using FASTME 2.1.4 [24]. The branch lengths were scaled in terms of the GBDP distance formula d5. The tree was rooted at the midpoint.

MiGA was used to perform taxonomic classification, essential gene presence analysis and average amino acid identity (AAI) (<https://disc-genomics.uibk.ac.at/miga> [25]). AntiSMASH was used to identify biosynthetic gene clusters (BGCs) that encode natural products, also known as secondary metabolites [26]. The average nucleotide identity (ANI) between genomes was assessed using the EZBioCloud ANI calculator (Chalita et al., 2024) [27], which implements the Orthologous Average Nucleotide Identity (OrthoANIu) algorithm [28] and the Fast alignment-free computation of whole-genome average nucleotide identity (FastANI) algorithm, an alignment-free MinHash-based method [29]. Digital DNA–DNA hybridization (dDDH) values were calculated using the Genome Genome Distance Calculator (GGDC) [30]. Antimicrobial resistance genes were identified and annotated using the resistance gene identifier (RGI) module of the Comprehensive Antibiotic Resistance Database (CARD) [31].

The GhostKOALA-KEGG Mapper was used to infer functional pathways [32, 33]. Carbohydrate active enzymes (CAZymes) were annotated using the dbCAN3 server [34]. *In silico* phenotyping was performed using Traitair (<https://github.com/hzi-bifo/traitair>) [35].

RESULTS

The bright-field microscopy revealed that strain H39_3 is a Gram-positive, rod-shaped bacterium, and it was further verified based on FESEM (Fig. 1). The sequenced 16S rRNA data of strain H39_3 shared the highest identity with *Rosellomorea marisflavi* (query cover: 100%; percent identity: 100%) on the basis of blastn analysis, supporting the taxonomic affiliation of strain H39_3 to the genus *Rosellomorea*. In the 16S rRNA-based phylogenetic tree, strain H39_3 formed a distinct monophyletic clade with *Rosellomorea marisflavi* (Fig. S1, available in the online Supplementary Material). The assembled draft genome of strain H39_3 consisted of three contigs, and the size was 4,461,134 bp (Fig. 2). The G+C content of the genome was 48.64 mol%. The identified closest relative of strain H39_3 was *R. marisflavi*. The dDDH, orthoANI and AAI values of strain H39_3 exhibited the highest values of 46.1%, 92.23% and 98.7%, respectively, with *R. marisflavi*. The whole-genome-based phylogeny also confirmed the closest affiliation with *R. marisflavi*. The basic genome features and associated quality statistics are detailed in Table 1. The genome completeness was ~95%, and contamination was 0.75%, reflecting an acceptable benchmark for a high-quality draft genome [36]. The predicted number of genes present in the genome was 6,061, with 5,913 CDSs (Fig. S2, available in the online Supplementary Material). The dDDH computed by GGDC between the whole genome of strain H39_3 and its closest relative was 46.1%, which is less than the accepted cutoff value of 70% for species delineation (Table 2). Additionally, ANI value of the strain with *R. marisflavi* was found to be 92.23% based on OrthoANI and 92.41% based on FastANI. Both the values were significantly

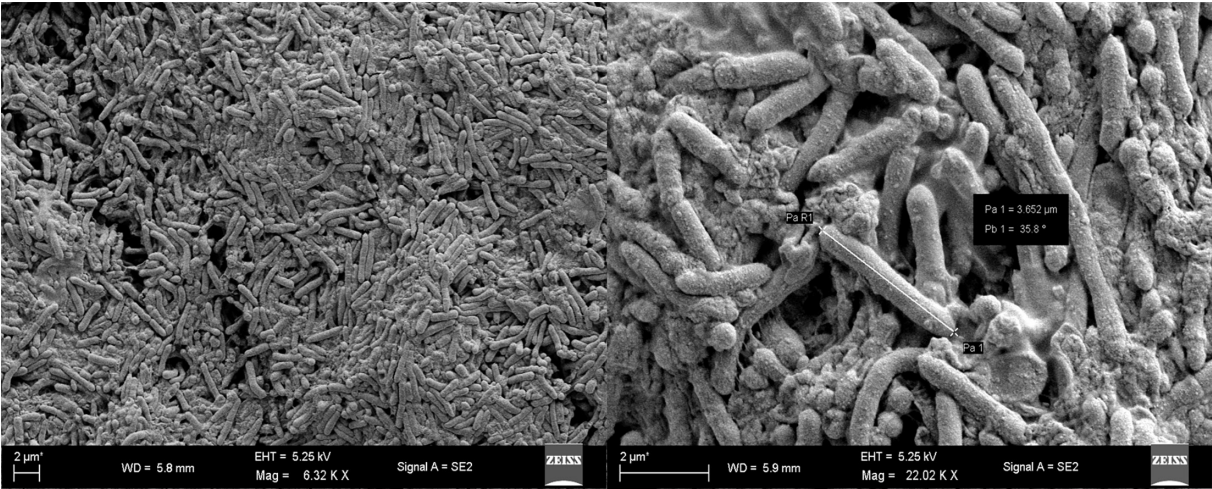


Fig. 1. FESEM image of *Rossellomorea* sp. strain H39_3. A scale bar is shown at the bottom left of the image. Mag: magnification, EHT: Electron High Tension, WD: working distance.

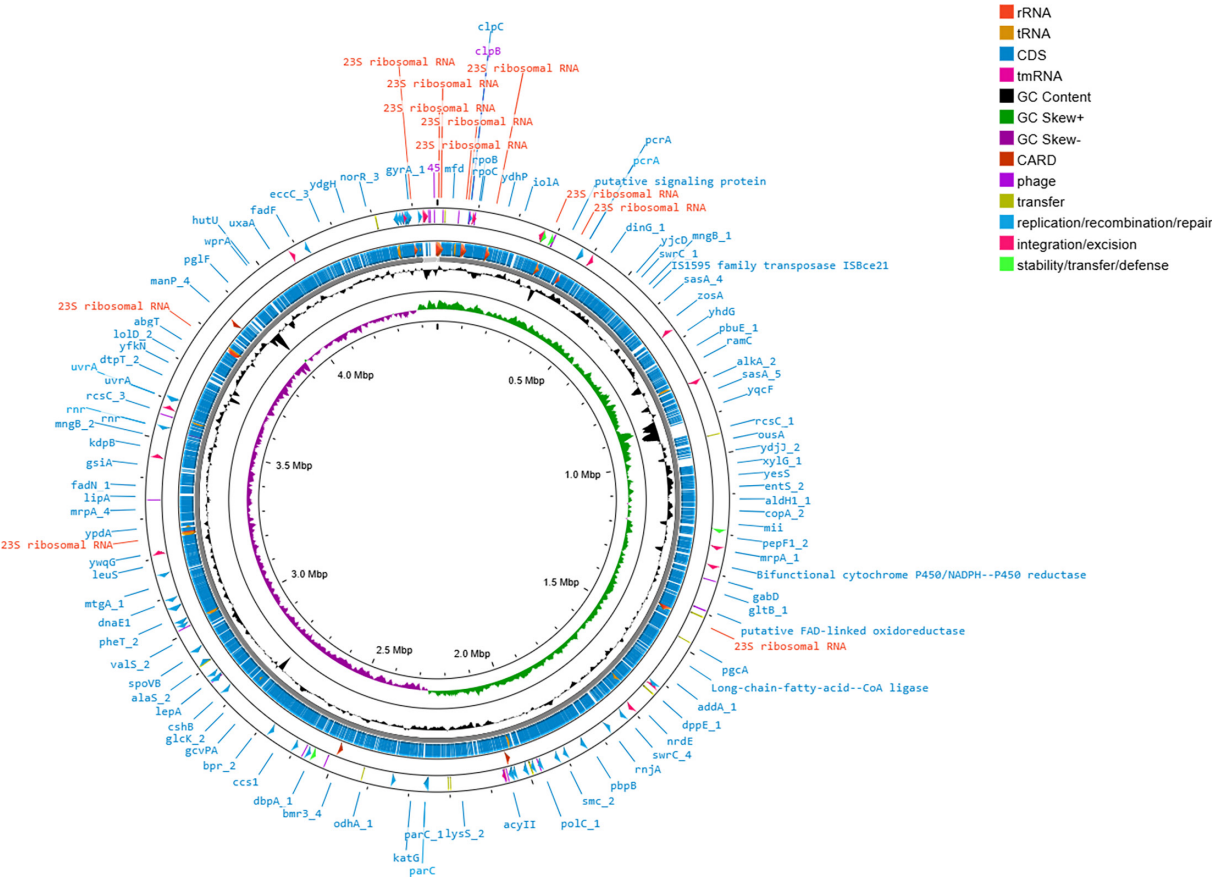


Fig. 2. Map showing a draft genome circular map of the strain H39_3, assembled into three contigs (representation only); markings on the outermost ring show identified genes based on Prokka annotation, including CDSs, rRNA, tRNA and tmRNA. The middle ring illustrates G+C content as deviations relative to the genome average. The innermost ring shows G+C skew, with positive values highlighted in green and negative values in purple. tmRNA: transfer-messenger RNA.

Table 1. Comparison of the genome of *Rosellomorea* sp. strain H39_3 with genomes of all reported species representing this genus

Reference genome	DDH (%)	ANI (%)	AAI (%)	GGDC
<i>R. marisflavi</i>	46.1	92.23	98.7	0.0814
<i>R. oryzaecorticis</i>	21	72.47		0.2092
<i>R. arthrocnemi</i>	19.7	72.6		0.2233
<i>R. aquimaris</i>	19.4	71.59	80.2	0.226
<i>R. vietnamensis</i>	19.3	73.14	79.9	0.2277
<i>R. pakistanensis</i>	21.2	69.22		0.2068

below the acceptable threshold for ANI (95–96%) [28, 37]. These results strongly suggest that strain H39_3 is a potential novel species, as per proposed standards for species delineation based on whole-genome-based relatedness metrics [37–39]. Despite this, 16S rRNA showed 100% identity with *R. marisflavi*. This highlights that although 16S rRNA is a valuable marker, it can fail to capture genome-wide divergence and may be insufficient as a sole criterion for species-level identification in this genus [40–42]. However, formal species description is awaited, following the completion of robust chemotaxonomic and phenotypic characterizations. The value of AAI was 98.7% with *R. marisflavi*, reflecting conserved protein-coding genes, though the lower ANI ($\approx 92\%$) signals broader genomic divergence in noncoding regions. Based on the GBDP tree, the isolate formed a separate monophyletic clade with *Rosellomorea marisflavi*. The tree branch was well supported by a bootstrap value of 100 (Fig. 3).

The genome annotation of strain H39_3 revealed the presence of motility-related genes. For example, genes encoding flagellar motor-mounted switch complex proteins (fliG, fliN, fliM, fliY), flagellar basal body proteins (flgCEFG), hook-basal body (fliE) and flagellar biosynthesis proteins and regulators (fliRO and flhF) were identified from the draft genome. The CAZymes involved in complex carbohydrate degradation, including glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and synthesis, were present in the genome [43]. The studied genome possesses genes from 15 different families of GHs, six different families of glycosyltransferase and six different families of CEs. Based on dbCAN analysis, there is a notably higher occurrence of CAZyme families predicted to act on pectin, including pectate lyase enzyme from the PL9 family, CE8 (pectin methylesterase), CE912 (pectin acetyltransferase), GH28 (polygalacturonases) and GH105 (unsaturated glucuronyl/galacturonylhydrolase). The genome has the highest number of hits (68) corresponding to GH13 family enzymes, which are major α -amylases mainly involved in catalysing the hydrolysis of polysaccharides such as glycogen and starch [8, 44]. The substrate utilization annotation with respect to CAZymes gene families revealed the *in silico* ability of this bacterium to break down chitin, pectin, hemicellulose, inulin, sucrose and starch, among others. *In silico* phenotyping by Traitair indicated that the organism is motile and Gram-negative. The isolate produces enzymes including oxidase, catalase and lipase and is susceptible to bile. It can convert nitrate to nitrite and nitrite to gas. The isolate can use tartrate, trehalose, cellobiose, acetate and mucate. The isolate can grow on ordinary blood agar, MacConkey agar and in the presence of high NaCl (6.5%) concentration, based on *in silico* phenotyping. The results of *in silico* phenotyping are summarized in Fig. S3 (available in the online Supplementary Material). The *in silico* analysis predicted a Gram-negative phenotype; however, microscopic observations confirmed that the strain is Gram-positive. Such mismatches can happen, as the accuracy of *in silico* phenotype predictions is affected by factors like genome completeness and can be less reliable for certain or less-studied, environmentally diverse taxa [35]. Thus, these observed features need to be validated by robust experimental approaches.

Table 2. Draft genome statistics of *Rosellomorea* sp. strain H39_3

Genome feature	No. or length
Size	4,461,134 bp
No. of contigs	3
No. of large contigs (>1,000 bp)	3
Contig N50	4,414,898
Contig N90	4,414,898
L50	1
L90	1
G+C (mol%)	48.64

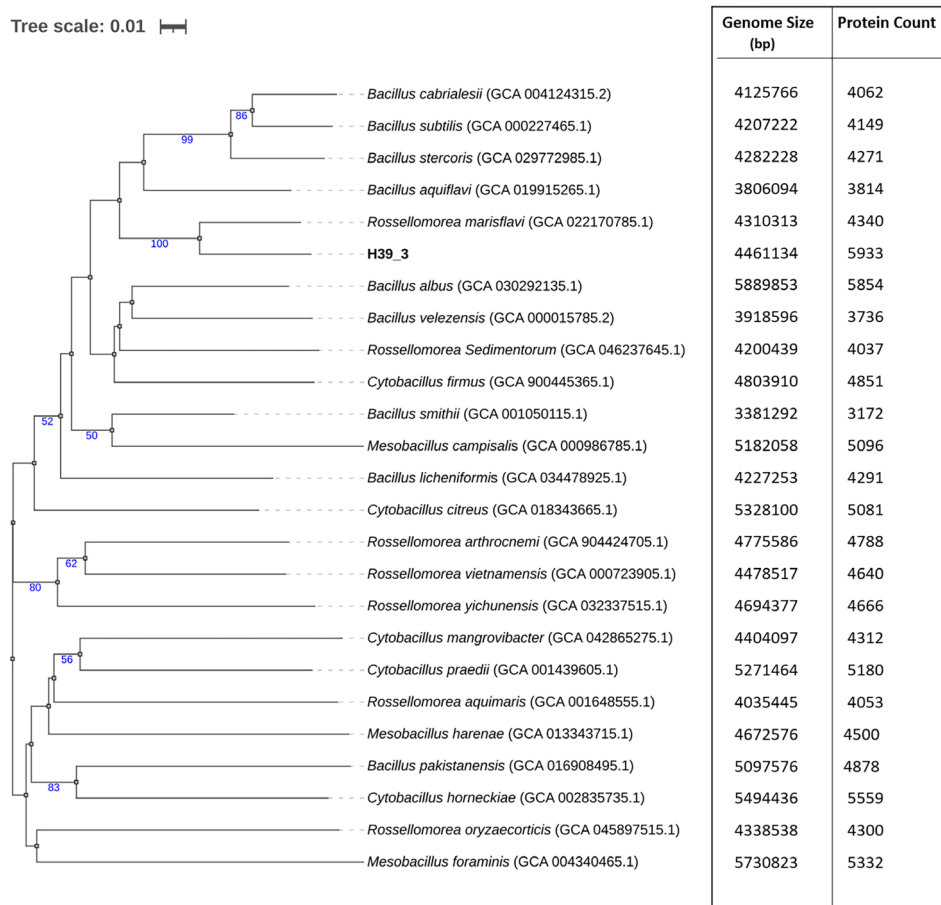


Fig. 3. Phylogram showing the phylogenetic relationship of strain *Rossellomorea* sp. H39_3 with other known species of *Rossellomorea* and selected representative within the family Bacillaceae. The phylogeny was inferred using the GBDP method implemented on the TYGS server. Distances were calculated from whole-genome sequences, and the tree was constructed using FastME v2.1.6.1. Branch lengths are scaled in terms of the GBDP distance formula d5, indicating genome-wide dissimilarity. Numbers (in blue) at nodes represent GBDP pseudo-bootstrap support values >50%, derived from 100 replications. The adjacent table on the right-hand side shows the size of genome and protein counts of the respective strains presented in the tree.

The BGCs identified by AntiSMASH included Thiopeptide-Linear azole (in)e-containing peptides (LAP), Lanthipeptide-Class IV, Terpene and Type III polyketide synthases (T3PKS). The antibiotic resistome of the isolate contains vanW, vanY and qacG genes, as predicted using the RGI database (strict hits only). The genome also showed the presence of the *arsC* (arsenate reductase) gene, encoding the ArsC protein, indicating the potential ability to reduce arsenate to arsenite. Additionally, a gene encoding the arsenical pump membrane protein YdfA was identified, suggesting the capability of the strain to mobilize or export arsenite from the cytoplasm. The genome also encodes numerous genes, including *cysN*, *cysC* and *sqr*, which are known to be involved in sulphur metabolism.

The functional pathways inferred by GhostKOALA-KEGG Mapper showed that *Rossellomorea* sp. strain H39_3 contains the nitrate assimilation gene *nirA*, linked to nitrogen metabolism.

Overall, this newly isolated bacterium strain encompasses a suite of functional traits indicative of multifaceted ecological significance. In particular, the presence of an array of genes linked to the metabolism of carbon, nitrogen and sulphur reflects the genomic potential for application towards the bioremediation of contaminated flowing aquatic ecosystems, such as the River Hindon. The genome-based insights are valuable for establishing a baseline for functional validation and to conduct further studies aimed at practical applications in the field of bioremediation.

Funding information

This study was partly supported by SwarnaJayanti Fellowship of Science and Engineering Research Board (DST/SJF/E and ASA-01/2017-18) awarded to P.B.

Acknowledgements

The authors would like to thank the members of the Integrative Taxonomy and Microbial Ecology Research Group (ITMERG) of IISER Kolkata for their help with the fieldwork.

Author contributions

C.K. undertook field sampling, performed the experimental work and drafted the initial version of the manuscript. N.S. contributed to experimental work and edited the manuscript. A.G. contributed to the conceptualization, validation of analysis and edited the manuscript. P.B. contributed to conceptualization, undertook funding acquisition and project administration, participated in field work, supervised the work and drafted and edited the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

- Gupta RS, Patel S, Saini N, Chen S. Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the *Subtilis* and *Cereus* clades of species. *Int J Syst Evol Microbiol* 2020;70:5753–5798.
- Navarro-Torre S, Carro L, Igual JM, Montero-Calasanz MDC. *Rosellomorea arthrocnemi* sp. nov., a novel plant growth-promoting bacterium used in heavy metal polluted soils as a phytoremediation tool. *Int J Syst Evol Microbiol* 2021;71:005015.
- Dastager SG, Mawlankar R, Tang SK, Srinivasan K, Ramana VV, et al. *Bacillus enclensis* sp. nov., isolated from sediment sample. *Antonie van Leeuwenhoek* 2014;105:199–206.
- Yoon JH, Kim IG, Kang KH, Oh TK, Park YH. *Bacillus marisflavi* sp. nov. and *Bacillus aquimaris* sp. nov., isolated from sea water of a tidal flat of the Yellow Sea in Korea. *Int J Syst Evol Microbiol* 2003;53:1297–1303.
- Yin QJ, Tang HZ, Zhu FC, Liu X, Xing YZ, et al. Complete genome of *Rosellomorea* sp. DA94, an agarolytic and orange-pigmented bacterium isolated from mangrove sediment of the South China Sea. *Mar Genomics* 2023;71:101059.
- Noguchi H, Uchino M, Shida O, Takano K, Nakamura LK, et al. *Bacillus vietnamensis* sp. nov., a moderately halotolerant, aerobic, endospore-forming bacterium isolated from Vietnamese fish sauce. *Int J Syst Evol Microbiol* 2004;54:2117–2120.
- Hong SW, Kwon S-W, Kim S-J, Kim SY, Kim JJ, et al. *Bacillus oryzae* sp. nov., a moderately halophilic bacterium isolated from rice husks. *Int J Syst Evol Microbiol* 2014;64:2786–2791.
- Mareček F, Terrapon N, Janeček Š. Two newly established and mutually related subfamilies GH13_48 and GH13_49 of the α -amylase family GH13. *Appl Microbiol Biotechnol* 2024;108:415.
- Boström KH, Simu K, Hagström Å, Riemann L. Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. *Limnol Ocean Methods* 2004;2:365–373.
- Ghosh A, Bhadury P. Insights into bacterioplankton community structure from Sundarbans mangrove ecoregion using Sanger and Illumina MiSeq sequencing approaches: a comparative analysis. *Genom Data* 2017;11:39–42.
- Lane DJ. 16S/23S rRNA Sequencing. In: Stackebrandt E and Goodfellow M (eds). *Nucleic Acid Techniques in Bacterial Systematics*. New York: John Wiley & Sons Ltd; 1991. pp. 115–175.
- Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004;5:113.
- Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol* 2021;38:3022–3027.
- Posada D. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 2008;25:1253–1256.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;39:783–791.
- De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 2023;39:btad311.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–1055.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
- Grant JR, Enns E, Marinier E, Mandal A, Herman EK, et al. Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res* 2023;51:W484–W492.
- Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 2019;10:2182.
- Lefort V, Desper R, Gascuel O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol* 2015;32:2798–2800.
- Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, et al. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res* 2018;46:W282–W288.
- Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res* 2023;51:W46–W50.
- Chalita M, Kim YO, Park S, Oh H-S, Cho JH, et al. EzBioCloud: a genome-driven database and platform for microbiome identification and discovery. *Int J Syst Evol Microbiol* 2024;74:006421.
- Chun J, Oren A, Ventosa A, Christensen H, Arahall DR, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 2018;68:461–466.
- Jain C, Diltthey A, Koren S, Aluru S, Phillippy AM. A fast approximate algorithm for mapping long reads to large reference databases. *J Comput Biol* 2018;25:766–779.
- Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 2022;50:D801–D807.
- Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2023;51:D690–D699.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28:27–30.
- Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;428:726–731.

34. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, *et al.* dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res* 2023;51:W115–W121.
35. Weimann A, Mooren K, Frank J, Pope PB, Bremges A, *et al.* From genomes to phenotypes: traitar, the microbial trait analyzer. *mSystems* 2016;1:e00101–16.
36. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017;35:725–731.
37. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* 2009;106:19126–19131.
38. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* 2005;102:2567–2572.
39. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;14:60.
40. Bartoš O, Chmel M, Swierczková I. The overlooked evolutionary dynamics of 16S rRNA revises its role as the “gold standard” for bacterial species identification. *Sci Rep* 2024;14:9067.
41. Chan JZ-M, Halachev MR, Loman NJ, Constantinidou C, Pallen MJ. Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. *BMC Microbiol* 2012;12:302.
42. Fox GE, Wisotzkey JD, Jurtshuk P Jr. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* 1992;42:166–170.
43. Huang L, Zhang H, Wu P, Entwistle S, Li X, *et al.* dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res* 2018;46:D516–D521.
44. MacGregor EA. Alpha-amylase structure and activity. *J Protein Chem* 1988;7:399–415.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org