

Whole-genome sequencing and annotation of probiotic *Bacillus amyloliquefaciens* COCAU_P1

Tanmoy Gon Choudhury,¹ Md Idrish Raja Khan,^{1,2} Partha Sarathi Tripathy,³ Suraj Kumar,¹ Ratan Kumar Saha,^{1,4} Dibyendu Kamilya⁵

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT *Bacillus amyloliquefaciens* COCAU_P1, an autochthonous probiotic from *Labeo rohita*, was genome sequenced (Illumina NovaSeq; GenBank [JBHEQH000000000](#)). Its 4.09 Mb genome encodes 4,351 genes linked to stress tolerance, nutrient utilization, antimicrobial defense, enzymes, and immunomodulation. Phylogenomics and functional evidence confirm its safety and probiotic potential for aquaculture applications in aquaculture systems.

KEYWORDS probiotics

Bacillus amyloliquefaciens COCAU_P1 is an autochthonous probiotic strain isolated from the intestine of *Labeo rohita* (Rohu) (1). The strain displays strong antagonistic activity against major fish pathogens, tolerance to a wide pH range (2–9), and high bile concentrations (up to 10%) and exhibits high mucus adhesion, hydrophobicity, and aggregation abilities. It is non-hemolytic and produces proteinase, amylase, lipase, and cellulase. Genetic screening confirmed multiple probiotic-associated marker genes, and feeding trials in *L. rohita* demonstrated enhanced immune responses and resistance to *Aeromonas hydrophila* (2). These combined attributes support COCAU_P1 as a potent and safe probiotic candidate for aquaculture (3).

A single purified colony was transported to the sequencing laboratory in tryptone soy agar plate. The bacteria was grown in tryptone soy broth at 37°C under agitation. Genomic DNA was extracted using the phenol–chloroform method and quality-checked using a NanoDrop spectrophotometer (4). Whole-genome sequencing was performed on the Illumina NovaSeq 6000 platform using 2 × 150 bp paired-end reads. Raw reads were quality-checked and trimmed using FastQC v0.12.0 and Trimmomatic v0.40 before downstream analysis. High-quality sequencing reads were assembled using SPAdes v4.2.0, and the assembled genome was annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP, v6.8). Taxonomic assignment of the assembled genome was performed using GTDB-Tk v2.6.1. Assembly quality assessment using QUAST v5.3.0 showed a high-quality 4.09 Mb genome with 46.94% guanine-cytosine (GC) content and 109× coverage. BUSCO v6.0.0 and CheckM v1.2.4 estimated ~99% completeness with <1% contamination, indicating a complete, low-contamination assembly. Functional categorization of probiotic-associated genes was carried out using the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (5). Whole-genome phylogenetic analysis was performed with GTTree (v1.8.13) (6), and the resulting phylogeny was visualized using iTOL (v6) (7) (Fig. 1). The assembled genome is 4,090,354 bp and contains 4,351 genes, including 4,118 protein-coding sequences. Eighty-five RNA genes were identified (70 tRNAs, 5 ncRNAs, and multiple copies of 5S, 16S, and 23S rRNAs). A total of 148 pseudogenes were predicted (Table 1).

Editor Silvia T. Cardona, University of Manitob, Winnipeg, Manitoba, Canada

Address correspondence to Tanmoy Gon Choudhury, tanmoygc@gmail.com.

The authors declare no conflict of interest.

Received 22 December 2025

Accepted 8 April 2026

Published 14 May 2026

Copyright © 2026 Gon Choudhury et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

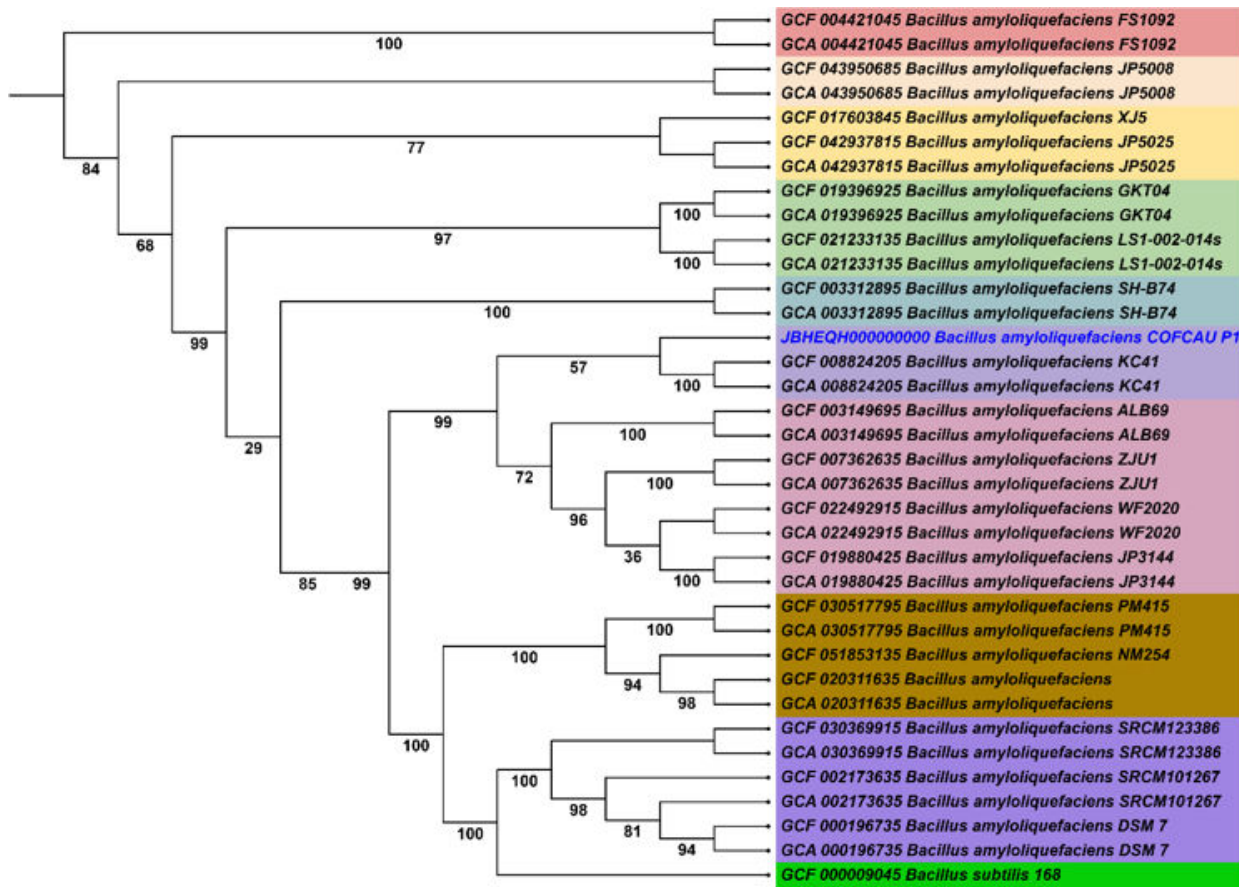


FIG 1 Phylogenomic tree contains the isolated *B. amyloliquefaciens* COCAU_P1 from this fish and other strains from NCBI.

TABLE 1 Genomic and probiotic functional features of *Bacillus amyloliquefaciens* COCAU_P1

Feature category	Description
Sequencing technology	Illumina NextSeq
Assembly method	NCBI PGAP (best-placed reference protein set + GeneMarkS-2+)
Genome coverage	High-depth Illumina coverage
Genome length (bp)	4,090,354
GC content (%)	46.940,674
Total genes	4,351
Protein-coding genes	4,118
RNA genes	85
tRNA genes	70
rRNA genes	5S × 3, 16S × 4, 23S × 3
ncRNAs	5
Pseudogenes	148
BioProject accession	PRJNA1158341
BioSample accession	SAMN43535920

ACKNOWLEDGMENTS

All authors sincerely acknowledge Central Agricultural University, Imphal, Manipur, India for providing the necessary facilities to carry out this study.

AUTHOR AFFILIATIONS

¹Department of Aquatic Health and Environment, College of Fisheries, Central Agricultural University (Imphal), Lembucherra, Tripura, India

²Department of Aquatic Animal Health Management, College of Fisheries Science, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India

³College of Fisheries, Rani Lakshmi Bai Central Agricultural University, Jhansi, India

⁴Techno India University, Tripura, India, Agartala, Tripura

⁵Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India

AUTHOR ORCIDS

Tanmoy Gon Choudhury  <http://orcid.org/0000-0002-0208-6691>

Partha Sarathi Tripathy  <http://orcid.org/0000-0003-2068-5484>

AUTHOR CONTRIBUTIONS

Tanmoy Gon Choudhury, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review and editing | Md Idrish Raja Khan, Conceptualization, Formal analysis, Methodology | Partha Sarathi Tripathy, Formal analysis, Software | Suraj Kumar, Data curation, Software | Ratan Kumar Saha, Supervision, Writing – review and editing | Dibendu Kamilya, Project administration, Supervision, Writing – review and editing

DATA AVAILABILITY

The WGS sequence of *Bacillus amyloliquefaciens* COCAU_P1 can be accessed from NCBI GenBank [JBHEQH0000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JBHEQH0000000000). The project is deposited under BioProject [PRJNA1158341](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1158341) and BioSample [SAMN43535920](https://www.ncbi.nlm.nih.gov/biosample/SAMN43535920). The raw sequence data are available in the Sequence Read Archive (SRA) database under accession number [SRR36553005](https://www.ncbi.nlm.nih.gov/sra/SRR36553005).

REFERENCES

1. Khan MIR, Kamilya D, Choudhury TG, Tripathy PS, Rathore G. 2021. Deciphering the probiotic potential of *Bacillus amyloliquefaciens* COCAU_P1 isolated from the intestine of *Labeo rohita* through *in vitro* and genetic assessment. *Probiotics Antimicrob Proteins* 13:1572–1584. <https://doi.org/10.1007/s12602-021-09788-2>
2. Khan MIR, Kamilya D, Choudhury TG, Rathore G. 2022. Dietary administration of a host-gut derived probiotic *Bacillus amyloliquefaciens* COCAU_P1 modulates immune-biochemical response, immune-related gene expression, and resistance of *Labeo rohita* to *Aeromonas hydrophila* infection. *Aquaculture* 546:737390. <https://doi.org/10.1016/j.aquaculture.2021.737390>
3. Gon Choudhury T, Roy S, Kumar S, Rath D, Malla S, Tripathy PS, Kamilya D, Parhi J, Saha RK, Patel AB. 2025. Freeze-drying of *Bacillus amyloliquefaciens* COCAU_P1 probiotics: process optimization, and evaluation of its immunoprotective role in *Labeo rohita*. *Aquacult Int* 33:243. <https://doi.org/10.1007/s10499-025-01922-z>
4. Sambrook J, Russell DW. 2001. *Molecular cloning a laboratory manual*. 3rd ed. Vol. 1. Cold Spring Harbor Laboratory Press, New York.
5. Olson RD, Assaf R, Bretin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich EM, Kenyon RW, et al. 2023. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res* 51:D678–D689. <https://doi.org/10.1093/nar/gkac1003>
6. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35:4162–4164. <https://doi.org/10.1093/bioinformatics/btz188>
7. Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 52:W78–W82. <https://doi.org/10.1093/nar/gkae268>