Unveiling the Biosynthetic Gene Cluster Potential of *Cytobacillus horneckiae*: From Genome Mining to Heterologous Expression Design

PPMI SITH ITB 2026 KK Genetika dan Bioteknologi Molekuler

Popi Septiani[1], Feraliana[1], Grace Ronauli Lusia Tambunan[1], Ummi Salamah[1], Azzania Fibriani[1,2], Adi Pancoro[1,2], Tati Kristianti[1,2], Hendra Munandar[3], Taofik Jasalesmana[3], Mutiara Rachmat Putri[4], Fahrurrozi[3], Husna Nugrahapraja[1,2], Dadang Sumardi[1,2], Karlia Meitha[1,2], Sony Suhandono[1,2], Muhammad Hamzah Syaifullah Azmi[1], Fenny Martha Dwivany[1,2], I Dewa Made Kresna[1], Ernawati Arifin Giri-Rahman[1,2*]

[1] School of Life Sciences and Technology, Institut Teknologi Bandung,
[2] Bioscience and Biotechnology Research Center, Institut Teknologi Bandung
[3] Marine and Terrestrial Bioindustry, National Research and Innovation Agency (BRIN)
[4] Faculty of Earth Sciences and Technology, Institut Teknologi Bandung

Abstract

Cytobacillus horneckiae remains genomically underexplored, and its biosynthetic potential is largely unknown. Whole genome sequencing of strain G7GM, isolated from the halophilic lake of Gili Meno, North Lombok, Indonesia, yielded a high-quality 5.77 Mb assembly of three contigs with a GC content of 37.46% and 155× depth. The genome showed substantial divergence from reference genomes, with a highest ANI of 98.34% and 71 unique orthologous genes. antiSMASH identified a lasso peptide biosynthetic gene cluster with medium similarity to paeninodin, containing *padeC, padeA, padeK, padeB1, padeB2,* and *psmN*, along with modified leader peptide. This BGC cluster was cloned into the pET-32b vector by Gibson Assembly and transformed into *E. coli* BL21, and insertion was confirmed by PCR. These results provide a genomic foundation and an expression platform for characterizing a putative lasso peptide from G7GM.

Keywords: biosynthetic gene cluster, *C. horneckiae*, genome mining, lasso peptide, whole-genome sequencing.

Introduction

The halophilic lake of Gili Meno, North Lombok, Indonesia, harbors halotolerant microorganisms with demonstrated biosynthetic potential. However, exploration has remained limited to conventional cultivation, even though microbial secondary metabolites are a major source of bioactive compounds for medicine and biotechnology.

Whole-genome sequencing (WGS) enables systematic exploration of microbial genomes and uncovers biosynthetic gene clusters (BGCs) that are often silent or cryptic under laboratory conditions. Genome mining has become an important strategy for discovering bioactive compounds from extreme environments. *C. horneckiae* lacks comprehensive genomic characterization, and its secondary metabolite repertoire remains largely unresolved. Here, we performed WGS and de novo assembly of *C. horneckiae* G7GM, identified a lasso peptide BGC similar to the paeninodin cluster using antiSMASH, and cloned it into pET-32b (+) for expression in *E. coli* BL21.

1 Sampling & isolation (Gili Meno lake)
2 Whole-genome sequencing & read QC
3 De novo assembly & QUAST evaluation
4 ANI, TYGS & OrthoVenn3 comparison
5 antiSMASH 7.0 mining → lasso BGC
6 Primer design, PCR & Gibson Assembly
7 BL21 transformation & PCR verification

Figure 1. Overview of the study workflow.

Research Method

Sampling & sequencing. Strain G7GM was isolated from halophilic lake water (4.7 m depth) at Gili Meno (North Lombok, Indonesia). WGS produced 24,137 reads (300 Mb); quality was assessed with Phred scores (Q5–Q15) and low-quality bases were trimmed.

Assembly & comparison. Reads were assembled de novo and evaluated with QUAST; read mapping gave depth and coverage at ≥1×, ≥5× and ≥10×. ANI was calculated with OrthoANI, phylogenomics with TYGS and orthologous genes with OrthoVenn3.

Genome mining. antiSMASH 7.0 (default parameters, focus on RiPP clusters) was used to identify the lasso peptide BGC and its core genes (*padeA, padeB1, padeB2, padeC*).

Cloning. Primers with 5′ overlaps and template-specific 3′ regions amplified the pET-32b (+) backbone (~5,300 bp) and the BGC insert (~5,005 bp), checked by agarose gel electrophoresis. Fragments were joined by one-step isothermal Gibson Assembly and transformed by heat shock into chemically competent *E. coli* BL21 (CCMB80 method); transformants were selected on LB agar with ampicillin and recombinant plasmids were verified by PCR with insert-specific primers.

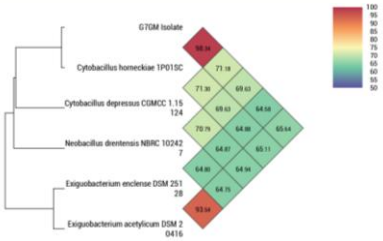


Figure 2. Average nucleotide identity (OrthoANI) between *C. horneckiae* G7GM and related reference genomes.

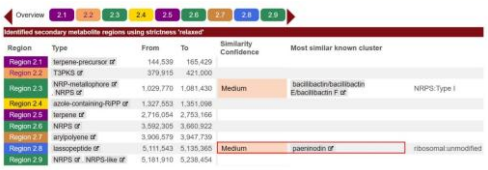


Figure 3. Biosynthetic Gene Cluster (BGC) from C. horneckiae G7GM strain

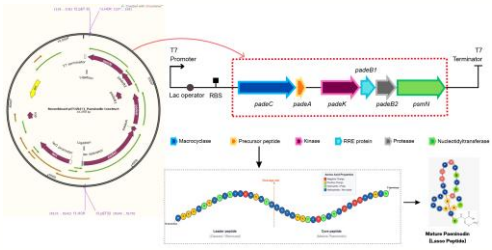


Figure 4. Recombinant pET-32b (+) construct carrying the paeninodin-like BGC and the primers designed for Gibson Assembly and proposed structure of the mature Paeninodin.

Discussion & Result

Genome quality. Whole-genome sequencing of *C. horneckiae* G7GM yielded 24,137 reads (300 Mb), 85.9% of which exceeded Q15. Assembly produced three contigs totaling 5.77 Mb (GC 37.46%, N50 5.64 Mb, L50 = 1) with 155× average depth and 100% coverage at ≥10×. These metrics indicate data quality sufficient for reliable genome assembly.

Novelty. Only 84.8% of the genome aligned to reference genomes, with numerous mismatches and misassemblies. The highest ANI was 98.34% (*Fig. 2*), and TYGS phylogenomic places G7GM closest to strain 1PO1SC. Both share 1,499 orthologous genes, of which 71 are unique to G7GM and 15 to 1PO1SC, suggesting genomic novelty.

Lasso peptide BGC. antiSMASH 7.0 identified a cluster with medium similarity to paeninodin (*Fig. 3*), containing *padeC* (macrolactam synthase), *padeA* (precursor peptide), *padeK* (kinase), *padeB1/B2* (leader recognition and proteolysis) and *psmN* (nucleotidyltransferase) *(Fig. 4). padeB1*, *padeB2* and *padeC* are highly conserved, whereas *padeA* and intergenic regions vary: the leader motif MKKQYSKPSLEVL is changed to MKKHWQKPTLEVL and the QTM motif to MTM, suggesting a structural variant of paeninodin. *psmN* is atypical for a paeninodin-like cluster. In the pseudomycoidin cluster of *Bacillus pseudomycoides*, it attaches hexose groups to a phosphorylated serine at the C-terminus of the core peptide; the G7GM peptide may therefore be both glycosylated and phosphorylated.

Recombinant construct. Colony PCR of ampicillin-resistant *E. coli* BL21 transformants confirmed the ~5,005 bp insert (*Fig. 5*), establishing a platform for heterologous expression.

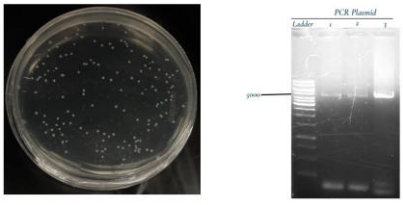


Figure 5. (A) Ampicillin-resistant E. coli BL21 colonies after transformation; (B) colony PCR showing the ~5,005 bp insert.

Conclusion

WGS and de novo assembly of *C. horneckiae* G7GM gave a high-quality genome that diverges from available references, indicating possible genomic novelty. A putative lasso peptide BGC with conserved core genes and precursor variation was identified, and cloned into pET-32b (+) and verified by PCR in *E. coli* BL21. This provides a platform for heterologous expression and LC-MS characterization in future work.