Integrated genomics and proteomics analysis of *Paenibacillus peoriae* IBSD35 and insights into its antimicrobial characteristics

Ng Ngashangva1, Pulok K. Mukherjee1, Chandradev Sharma1, Mohan C. Kalita2 & Indira Sarangthem1*

Antimicrobial resistance has been developing fast and incurring a loss of human life, and there is a need for new antimicrobial agents. Naturally occurring antimicrobial peptides offer the characteristics to counter AMR because the resistance development is low or no resistance. Antimicrobial peptides from *Paenibacillus peoriae* IBSD35 cell-free supernatant were salted out and purified using chromatography and characterized with liquid chromatography–tandem-mass spectrometry. The extract has shown a high and broad spectrum of antimicrobial activity. Combining the strain IBSD35 genome sequence with its proteomic data enabled the prediction of biosynthetic gene clusters by connecting the peptide from LC–MS/MS data to the gene that encode. Antimicrobial peptide databases offered a platform for the effective search, prediction, and design of AMPs and expanded the studies on their isolation, structure elucidation, biological evaluation, and pathway engineering. The genome-based taxonomy and comparisons have shown that *P. peoriae* IBSD35 is closely related to *Paenibacillus peoriae* FSL J3-0120. *P. peoriae* IBSD35 harbored endophytic trait genes and nonribosomal peptide synthases biosynthetic gene clusters. The comparative genomics revealed evolutionary insights and facilitated the discovery of novel SMs using proteomics from the extract of *P. peoriae* IBSD35. It will increase the potential to find novel bio-molecules to counter AMR.

The discovery of penicillin has heralded the modern era of drugs and saved millions of lives. However, it was short-lived and antimicrobial resistance (AMR) has been developing fast and incurring a loss of human life1,2. Given this, there is a continuous need for new antimicrobial agents. Prokaryotes perform a vital function for organisms and ecosystems and provide by-products for human and veterinary medicine, agriculture, and manufacturing. They are well-known to mediate microbe-host and microbe-microbe interactions for their survival competition. Consequently, they produced diverse bio-molecules vital for the structure and proper function of living cells or organisms3. At the same time, they caused infectious diseases and extended the AMR through the environment, reproduction, and gene transfer4.

The first functionalized peptide-producing endophyte was isolated from *Kennedia nigricans*5. Afterward, *Bacillus amyloliquefaciens* sp. was isolated from *Ophiopogon japonicas* which afforded the discovery of antitumor exo-polysaccharides6. We presumed that the plant-microbes association might have affected each other in their respective biochemical properties7. This interaction might enable microbe existence as an endophyte through absorbing or secreting biomolecules8. Furthermore, the discovery of anticancer taxol and taxane production by *Taxomyces andreanae* has stimulated research interest in bacterial and fungal endophytes9. The under-study *P. peoriae* IB3D35 (or strain IB3D35) was isolated from the stem of *Millettia pachycarpa* in our previous experiment10–14, and it has shown antimicrobial activity from its cell-free supernatant (CFS)15. Therefore, we investigated the source genome's antimicrobial characteristics, extracellular antimicrobial compounds, and the potential to counter AMR.

---

1A National Institute of Department of Biotechnology, Institute of Bioresources and Sustainable Development (IBSD), Govt. of India, Takyelpat, Imphal, Manipur 795001, India. 2Department of Biotechnology, Gauhati University, Jalukbari, Guwahati, Assam 781014, India. *email: sidevil@yahoo.co.in
Naturally occurring antimicrobial peptides (AMPs) offer the characters to counter AMR. They are present in every phylum and kingdom and are known for antibacterial, anticancer, antiviral, antifungal, antiparasite, effectors of innate immunity, and antibiotic adjuvant characteristics. The resistance development of AMPs by gene mutation is less expected as mature peptides are created from a non-descript sequence of pro-peptides and lack a definite recognition site. They have multiple attacking target sites in contrast to common antibiotics by gene mutation.

A comprehensive genomics analysis, advanced proteomics techniques, AMP databases, and structural models offer a platform for effective search, prediction, drug development, and elaboration of AMPs. The mass spectrometry (MS) technique can identify AMP sequence and mass from a mixture sample and subsequently disclose their primary, secondary, and tertiary structure. The liquid chromatography-tandem-mass spectrometry (LC–MS/MS) proteomics, and AMP databases have broadly impacted AMP biology and the development of peptide-based drugs. Moreover, genome mining has enabled the prediction of a biosynthetic gene cluster (BGC), linking the proteins and their peptides from LC–MS/MS data and connecting these natural products to the genes that encode them. Proteomics has enabled genomic data to guide the discovery of new molecules to be applied in biomedical research.

Therefore, the present study analyzes the characteristics of the Paenibacillus peoriae IBSD35 draft genome and the proteomics of its extracellular AMPs. The understudy genome and its close strain comparisons illustrate its endophytic and antimicrobial traits. It provided an insight into the potential of novel AMPs to counter AMR.

Results

Genomic deoxyribonucleic acid (DNA) estimation and annotation. Paenibacillus peoriae IBSD35 inoculated on Luria Bertani (LB) agar plate was visible after 12 h of incubation at 38 °C. It is an endospore-forming and neutralophilic bacterium based on Integrated Microbial Genomes Annotation Pipeline (IMG AP) v4.16.6. The genomic deoxyribonucleic acid (gDNA) assessed with 0.8% agarose was estimated to be more than 50-kilo base pairs (kbp), and the concentration was 188.63 µg µl⁻¹ (Supplementary Fig. S1). Paenibacillus peoriae IBSD35 has a linear DNA topology with a total length of 5,862,582 bp. Its closest species reference genome is Paenibacillus peoriae KCTC 3763 with a symmetrical identity of 62.7495% and a symmetrical identity of 89.4884% with Paenibacillus FSL J3-0120 in National Center for Biotechnology Information (NCBI) neighbor genome report.

Spectrum of antimicrobial activity of the Paenibacillus peoriae IBSD35 inoculum. The CFS of Paenibacillus peoriae IBSD35 inoculum retained antimicrobial activity against Gram-positive Staphylococcus aureus American Type Culture Collection (ATCC) 25923. The crude extract has shown antimicrobial activity against the Gram-negative Escherichia coli ATCC 25922 and fungal pathogen Candida albicans ATCC 10231, indicating it's broad-spectrum of antimicrobial activity. The variation of antimicrobial activity at different stages of purification and concentration suggested the complex nature of biomolecules.

Optimum growth conditions for antimicrobial metabolites production. The Paenibacillus peoriae IBSD35 inoculum's optimum optical density (OD) was measured to be 0.65 at λ600 nm observed on the 6th day, coinciding with the exponential growth phase in the LB medium. The optimum growth was at pH 6.8–7. Paenibacillus peoriae IBSD35 is a mesophile bacterium that grows best in moderate temperatures observed at 38 °C (Fig. 2A). The supernatant antimicrobial activity was determined as an arbitrary unit (AU) per milliliter which is the reciprocal of the highest dilution (2ⁿ) that resulted in the inhibition of Salmonella typhimurium ATCC 14028. The AU was measured to be 1600 AU ml⁻¹.

Salting out and antimicrobial activity retention test of dialysate. The fraction (F4) eluted out with the 700 milimolar (mM) sodium chloride (NaCl) at the flow rate of 2 ml min⁻¹ has shown antimicrobial activity against S. aureus ATCC 25923. The fraction OD was measured to be 1.147 and 2.803 at wavelength (λ)214 and λ280, respectively (Table S1). The fraction (F4) was collected, lyophilized into powder form, and stored at – 20
°C. It dissolved in distilled water. It was dialyzed with 12.4 kilo dalton (kDa) molecular weight cut-off (MWCO) dialysis tube (Sigma Aldrich, Germany) to eliminate the larger protein complexes. The dialysate has retained antimicrobial activity against *S. aureus* ATCC 25923.

Reverse phase-high performance liquid chromatography (RP-HPLC) and estimation of protein. The protein mixture gradient elution was at an isocratic flow rate of 2.5 ml min⁻¹. The P4 (4.165) retention time was 4.165 min with a purity index of 0.86310, and the P5 (4.567) retention time was 4.567 min with a purity index of 0.95597 (Table S2). The RP-HPLC peaks P4 and P5 are sufficiently separated (Fig. 2B), and their absorption at λ205 and λ214 indicated the presence of amino acids and peptide bonds with other impurities. The absorbance at λ280 confirmed the presence of proteins. The P4 fractions had exhibited higher antimicrobial activity than P5. The purification fold of RP-HPLC was 151.72 starting from the initial crude extract (Table S1). The loss of bioactive metabolites from the harvesting of the cells ranged from 0 to 7.7%. The loss of cells after extraction ranged from 0.6 to 2.3%.

Effect of temperature, detergents, inorganic solvents, pH variation, and degradative enzymes on antimicrobial activity. A 0.2 µm filter (Avixa) sterilized supernatant treated at different temperatures retained the antimicrobial activity against *S. aureus* ATCC 25923. Autoclaving at 121 °C for 15 min and treatment with Tween 20, Tween 80, and sodium dodecyl sulfate (SDS) have slightly increased the antimicrobial activity (from 10 to 14 mm zone of inhibition diameter). It might be due to the disaggregation and exposure of peptide bonds. The inorganic solvents, pH variation, and degradative enzymes treated at a 1:1 ratio with the filtrate did not affect the antimicrobial activity significantly. The antimicrobial activity of the CFS is almost comparable to the ampicillin at 1 mg ml⁻¹ (Table S3). Moreover, the sample persisted in low-temperature freezer storage (−20 °C) without any preservative for 1 year.

Mass spectrometry and proteomics analysis of peptide mixtures. The *P. peoriae* IBSD35 proteins searched with Sequest, Mascot, and PEAKS from the uninterpreted experimental LC–MS/MS database confirmed the prediction of AMPs (Supplementary LC–MS Experiment data)³². The number of sequence protein entries predicted in the *P. peoriae* IBSD35 Uniprot database is 15 protein groups. Different peptide sequences have detected during the process (Table 1). An uncharacterized protein was predicted from the experimental LC–MS/MS with accession no. A0A2S6NUT6. The Uniprot search has shown that it is 94.87% similar to A0A509FZN2_PAEP0 of *Paenibacillus polymyxa* (A0A509FZN2) (Bit score-296, E-value-3.6 e⁻⁸⁹)³³. Combining the genome mining and proteomics analysis has shown that the gene (ID 2816869382) is 97% similar to the gene (ID 2556935736) of *Paenibacillus polymyxa* CR1 (NC_023037) (Bit score-286; 1e⁻⁹⁵) and it has 134 amino acid residues.

An AMP predicted from the protein lists with accession no. A0A2S6NUT6 is designated peoriaerin IBSD3 based on its source name (Table 2). Its sequence is AAGIQAQAGFGLSDSIQGTGKQKCSFCK. The N-terminal of the sequence considered is alanine (Ala). It has 28 amino acid residues with a molecular weight (MW) of 2802.17 Da (Supplementary Fig. S2). Its sequence backbone has a large portion of helix and residues embedded in the membrane. It suggested peoriaerin IBSD3 might insert the membrane for its antimicrobial activity (Fig. 3). The peptide sequence alignment indicated that it is 38.23% similar to fermentin³⁰. The three-dimensional (3-D) model of peoriaerin IBSD35 was generated using an automated protein structure de novo modeling and deposited at ModelArchive with accession no. 10.5452ma-id7l4 (Fig. 3).
The antimi-
crobial action had observed after 1 h of incubation. The minimum inhibition concentration (MIC) of peoriaerin IBSD35 was determined to be 0.04 µg µl −1. Scanning electron microscopy (SEM) imaging indicated that the pathogen colonies were slowly disappearing. It correlated with the decreased colony count from more than 400 to 4 in 2 h (Fig. 4). The OD at λ600 nm was 1.409 indicating the cell growth and debris turbidity. There was no

**Table 1.** The peptide sequences, masses, m/z, PTM, and their similarity percentage with the known peptide. The AMP sequences of the *P. peoriae* IBSD35 which have putative BGCs in the genome were generated using LC–MS/MS. m/z mass to a charge ratio, PTM post translational modification, AMP antimicrobial peptide, % percentage.

| Peptide          | m/z    | Mass  | PTM  | AMP             | Similarity (%) |
|------------------|--------|-------|------|-----------------|----------------|
| ACATEYTAK        | 479.32 | 956.42| N    | Fusaricidin D   | 40             |
| MFPPVYVK         | 540.31 | 1078.58| N    | Bacteriocin NS-2| 35.71          |
| MTNYTSK          | 473.30 | 944.42| N    | Fusaricidin D   | 37.5           |
| AC(+57.02)TVDVHK | 465.30 | 928.44| Y    | Lipopeptide     | 37.5           |
| AGTLPEPLAQDGTVYR | 850.49 | 1698.89| N    | Lassomycin      | 41.17          |
| ATAGGVNLG(+15.99)SAQTAEGLAM-NECAMLTR | 886.08 | 2655.21| Y    | Macedovicin, lantibiotic | 36.66 |
| KCTLEVHK         | 479.32 | 956.51| N    | Colistin [Polymyxin E1 and E2] | 33.33 |
| K(+42.01)ANPLSGK | 479.32 | 956.52| Y    | Duramycin C (Lantibiotic, type B, class 1 bacteriocin) | 36.84 |
| MESEDHIslPYNTHVSRSTTVT-5NLSHTYTLTFTESQK | 982.46 | 4685.13| Y    | Macedovicin, lantibiotic | 36.66 |
| AAGIQAGFGLSDSPIQTGKQKCSFCK | 2858.40 | 2802.17| C    | Fermentin | 38.23 |

**Table 2.** Protein group accession list from LC–MS data and the peptide sequence generated from MS-ion search.

| Accession no. | Description | Peptide | Modification | Charge | MH + [Da] | ΔM [ppm] | RT [min] | # Missed cleavages |
|---------------|-------------|---------|--------------|--------|-----------|----------|----------|-------------------|
| A0A2S6NUT6    | Uncharacterized protein | AAGIQAGFGLSDSPIQTGKQKCSFCK | C27(Carbamidomethyl) | 5      | 2858.40070 | 2.86     | 39.06    | 2                 |

**Figure 3.** A 3-D model of peoriaerin IBSD35 predicted using de novo modeling. The amino acids composition of peptide is AAGIQAGFGLSDSPIQTGKQKCSFCK. Sequence contains high Gly content. It is indicated in black color. Glycine is integral to the formation of alpha-helices in secondary protein structure due to its compact form. The purple color indicated the helix structure property. It included the green color which intersects with the purple color. The green color indicated the residues embedded in membrane. The light orange color indicated the Trp which is the rings in the residues. The red color indicated the Ala. The whole structure represented the backbone and side chain of the AMP except the Gly indicated in black color.

**Peoraerin IBSD35 mode of action imaging through scanning electron microscopy.** The antimicrobial action had observed after 1 h of incubation. The minimum inhibition concentration (MIC) of peoriaerin IBSD35 was determined to be 0.04 µg µl −1. Scanning electron microscopy (SEM) imaging indicated that the pathogen colonies were slowly disappearing. It correlated with the decreased colony count from more than 400 to 4 in 2 h (Fig. 4). The OD at A600 nm was 1.409 indicating the cell growth and debris turbidity. There was no
evidence of DNA and protein leakage. It showed that the AMP might kill \textit{S. aureus} ATCC 25923 by penetrating the cell cytoplasm and periplasm.

**Draft genome sequencing, assembly and annotations of \textit{P. peoriae} IBSD35.** The nucleic acid and protein ratio (A260/A280) was estimated to be 1.9 (Fig. 5A). The draft genome with 250 bp-end reads produced a total number of raw reads (R1 + R2) of 15,497,242 bp\textsuperscript{34}. Total bp are 3,874,320,000 bp with 46.01 guanidine-cytosine (GC) % and a read length of 250. The clean reads produced a total length of 5,862,582 bp with 58.0× fold coverage. The \textit{de-novo} whole genome assembled with ABySS v.2.0\textsuperscript{35} has produced 66 contigs totaling 5,862,582 bp with N50 of 400,732 and L50 of 6, 42 scaffolds and a GC content of 45.6% (Table S4). The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v.4.4 has identified 5,245 genes, 4,983 coding sequences (CDSs), 7, 1, 3 (5S, 16S, 23S) rRNAs (ribosomal ribonucleic acids), 2, 1 (5S, 16S) complete rRNAs, 7, 1, 3 (5S, 16S, 23S) partial rRNAs, 81 tRNAs (transfer ribonucleic acids), and 4 nc RNAs (noncoding RNAs) (Table S4). CRISPR/csd (clustered regularly interspaced short palindromic repeats) is involved in the prokaryotic defense system\textsuperscript{36}. Out of two 16S rRNA bacterial short sequence units (SSU), one has 1,560 bp, and the other has 45 bp based on IMG AP v.4.16.6.

**Global overview of \textit{P. peoriae} IBSD35 genome.** The maximum number of protein-coding genes (4,983) is related to the genetic information process (47%), environment information processing (45%), carbohydrate metabolism protein family (42%), and signaling process (31%) (Supplementary Fig. S3)\textsuperscript{37,38}. Membrane and cell contributed the highest percentage of genes in the cellular component category, while the catalytic activity, binding, and transportation account for the highest number in molecular function. The metabolic and cellular processes have the highest percentage within the biological processes category (Fig. 5B).

\textit{P. peoriae} IBSD35 harbored cytochrome bd ubiquinol oxidase [EC: 1.18.6.1]\textsuperscript{39}, phosphotransferase system (PTS), and sugar-specific component. They might be utilizing

\begin{equation}
\text{Ubiquinol} \left(8 + 2.5 \textbf{H} + 0.5 \textbf{O}_2 \right) \Leftrightarrow \text{Ubiquinol-8} + \textbf{H}_2\textbf{O} + 2.5 \textbf{H}.
\end{equation}
extra-cellular trehalose, cellobiose, and maltose. It revealed the *P. peoriae* IBSD35 auxotrophic lifestyle (www.kegg.jp/kegg/kegg1.html)\(^42,43\). The Joint Genome Institute-Integrated Microbial Genomes and Microbiomes System (JGI-IMG/M) analysis revealed a putative horizontally transferred genes count of 10 (0.19%) with the best hits to genes from *Proteobacteria* and *Clostridia*. The trans-membrane count is 1392 (26.28%) and might serve as a gateway to transport specific substances across the membrane\(^28,29\). The signal peptide is 303 (5.72%) of the total CDS identified. It might prompt the translocation of a cell protein secretion\(^44,45\).

*Paenibacillus peoriae* IBSD35 harbored *NisK–NisR* biosynthesis sensor histidine kinase [EC: 2.7.13.3], a two-component signal transduction system (TCS), known to affect the biosynthesis of AMP\(^46,47\). It also harbored sensor histidine kinase and response regulators, two *DesK–DesR* of the NarL family, four *CheA–CheY* of a chemo-taxis family, two *Res–ResD*, and two *degS–degU* for motility, regulation, and interactions. Flagellar proteins (fls) in the cellular chaperones might cause microbe motility and signal transduction. A flagellin/flagellar hook-associated protein (*flhC*) includes glycerol kinase and molecular chaperone (HtpG) that might involve in plant–microbe interaction\(^48\).

*P. peoriae* IBSD35 genome also harbored biodegradation capability proteins for xenobiotic compounds\(^49\).

**Comparison of genomes characteristics.** The selected *Paenibacillus* species have shown varying numbers of DNA scaffolds ranging from 1 to 56. *Paenibacillus* sp. JDR-2 has the largest genome size (7,184,930 bp), and *B. subtilis subtilis* AG1839 has the smallest genome size (4,193,640 bp) among the selected genomes (Table 3). The dot plot homology revealed a low level of *P. peoriae* IBSD35 synteny with *Paenibacillus peoriae* KCTC 3763 but a higher synteny with *Paenibacillus peoriae* HS311, *P. polymyxa* SC2, and *P. polymyxa* E681 (Supplementary Fig. S4). The predicted numbers of nonribosomal peptide synthase (NRPS) BGCs ranged from 0 to 18 among the eleven genomes indicating their diverse and rich SM natural products (Table 3). *P. peoriae* IBSD35 harbored 18 NRPS (72% of the total 25 BGCs) and encompassed the highest number of NRPS BGCs among the selected genomes comparisons.

**Antibiotics and other secondary metabolites (SMs) gene cluster of *P. peoriae* IBSD35.** The CFS of *P. peoriae* IBSD35 has shown a high and broad-spectrum antimicrobial activity against *S. aureus* ATCC 25923, *E. coli* ATCC 25922, *C. albicans* ATCC 10231, *Klebsiella pneumoniae* ATCC 4352, and *Salmonella typhimurium* ATCC 14028 (Fig. 1A–E). The activity was comparable with ampicillin and nisin at a much lower dose. Antibiotics and Secondary Metabolite Analysis Shell (AntiSMASH) v5 using Integrated Microbial Genomes Atlas of Biosynthetic gene clusters (IMG-ABC) web server tools predicted *P. peoriae* IBSD35 harbored 25 BGCs\(^60,61\). LC–MS sequencing of RP-HPLC purified sample have predicted 15 protein lists (Table 3).

*Paenibacillus peoriae* IBSD35 harbored resistance genes for beta-lactam, vancomycin, cationic antimicrobial peptide (CAMP), platinum drug, and antifolate drug\(^62–65\). The Rapid Annotation of microbial genomes using Sub-systems Technology (RAST) tool predicted that the *P. peoriae* IBSD35 genome harbored 29 compounds resistant...
**Table 3.** Comparative genome features. 1. *P. peoriae* KCTC 3763 (NZ_AGFX01000000) 2. *P. polymyxa* E681 (NZ_CP048793) 3. *P. peoriae* IBSD35 (NZ_PTJM01000000) 4. *P. polymyxa* SC2 (NC_014622) 5. *P. peoriae* HS311 (NZ_CP011512) 6. *P. vortex* V453 (NZ_ADIH31000000) 7. *B. subtillis* subtilis AG1839 (NZ_CP008698) 8. *S. coelicolor* A3(2) (AL648882) 9. *P. terrae* HPL-003 (NC_016641) 10. *Paenibacillus* sp. JDR-2 (NC_012914), and 11. *P. polymyxa* CR1 (NC_023037).

To antibiotics and toxic compounds as compared to the metabolic reconstruction of *P. polymyxa* E681. They shared 23 compounds, while one is unique in the *P. peoriae* IBSD35 and nine are unique in *P. polymyxa* E681.

*Paenibacillus peoriae* IBSD35 harboring 18 NRPS indicated that nonribosomal peptide compounds are the most abundant SmS of our understudy strain IBSD35 genome (Fig. 6A). The genome encompassed three gene clusters- one Trans-acyltransferase polyketide synthase (Trans AT-PKS) and two clusters of Trans AT-PKS. Two BGCs are associated with Butyrolactone, and one each for Type III polyketide synthase (T3PKS), bacteriocin, phosphonate, lanthipeptide, NRPS-like, polyketide synthases-like, and lasso peptide. At least 12 BGCs start from a position, and 2 BGCs code rhizomides A, B, and C as predicted by the Antibiotics and Secondary Metabolite Analysis Shell-Minimum Information about a Biosynthetic Gene Cluster (AntiSMASH-MIBiG) comparison (Table 4).

**Orthologous gene clusters of *P. peoriae* IBSD35 genome.** OrthoVenn analysis has shown the selected *Paenibacillus* species (*P. peoriae* HS311, *Paenibacillus* sp. JDR-2, *P. polymyxa* E681, *P. polymyxa* SC2, *P. terrae* HPL-003, and *P. peoriae* IBSD35 formed 5615 clusters, 5296 orthologous clusters (at least contains two genes), and 2125 single-copy gene clusters (E-value = 1e-2 and inflation value = 1.5) (Fig. 6B). The OrthoVenn diagram has shown that all 6 *Paenibacillus* species have shared 2241 gene clusters suggesting their conservation in the lineage. Additionally, it has 6-clusters specific to *P. peoriae* IBSD35 that might be lineage-specific gene clusters. The single-copy genes have suggested they maintained single-copy status through evolutionary time after the divergence of these species.

**Central carbon metabolism (CCM) of *P. peoriae* IBSD35 genome.** The metabolite precursors of *P. peoriae* IBSD35 cell biomass included gluconeogenesis (GG), citrate cycle (TCA), pentose phosphate pathway (PPP) metabolites, and pyruvate metabolism (Table 5).

**Genome-scale phylogeny analysis.** *Paenibacillus peoriae* IBSD35 had a digital DNA-DNA hybridization (dDDH) of Genome-To-Genome Distance Calculator (GGDC) formula 2 (d4) 76.1, confidence intervals (C.I.) d4 of [73.1–78.9], and a difference G + C% of 0.16 with *P. peoriae* HS311 indicating their close species relation. *P. peoriae* IBSD35 has a dDDH of d4 86.2, C.I. d4 of [83.6–88.5], and a difference G + C% of 0.00 with *P. peoriae* FSL J3-0120 which indicated their very close relationship. Additionally, *P. polymyxa* CR1 and *P. peoriae* IBSD35 exhibited a close relationship with d4 75.6, C.I. [72.6–78.4] and a difference G + C% of 0.02 (Fig. 6C). The strain IBSD35 exhibited OrthoANI values of 96.37 with *P. peoriae* FSL H8-0551 which is above the species boundary value of average nucleotide identity (ANI) = 95–96%. The result indicated that the G + C% difference computed from the query and subject genome sequences varied by no more than 1% and proved they are within species.

**Commensal lifestyle of *P. peoriae* IBSD35.** The genome was predicted to harbor sugar-specific component phenotypes, thus possibly utilizing the host plant trehalose, L-arabinoside, sucrose, and maltose. It harbored...
Tat A (twin-arginine translocation) and Sec (general secretory pathway) protein secretion systems for transpor-
tation across the cytoplasmic membrane. Most secretion systems genes related to virulence and pathogenic-
ity, such as types I, II, III, V, and VI secretion systems, were absent in the genome, which supported its non-pathogenicity characters while handling.

Nitrogen fixation characters of *P. peoriae* IBS35. It harbored putative genes encoding Mo-nitro-
genase MoFe protein (EC 1.18.6.1), multi-functional system (MFS) transporter, nitrate/nitrite transporter, and ammonium transporter involved in the uptake of nitrate and nitrite (Supplementary Figs. S5, S6). The nitrogen fixation genes include *nar*G, *nar*H, *nar*I, *nar*J, *nar*W, *nir*A, *nir*B, *nas*A, *hcp*, *gdh*A, *isc*U, *nif*B, *nif*D, *nif*E, *nif*K, *nif*N, *nif*E, *nif*H, *nif*V, *nif*X, *nif*Z and other putative nitrogen-fixing proteins plus a hypothetical protein that lacks an assigned function were identified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway data-
base analysis which indicated its nitrogen-fixation nature (Table S5) (http://www.kegg.jp/kegg/kegg1.html). However, the nitrogen fixation regulator (*nif*L/A) was not visible in the nitrogen metabolism pathway. *P. peoriae* IBS35 might absorb some of the nitrate sources from the host plant.

Plant-associated endophyte characters of *P. peoriae* IBS35. The proteins involved in the flagel-
lum assembly gene ontology (GO): 0044780, flagellum-dependent cell motility (GO: 0071973), and chemotaxis
Table 4. JGI-ABC using AntiSMASH v 5.0 predicted the putative BGCs coding for antibiotics and secondary metabolites. The clusters include NRPS cluster such as transatpks, bacillaene, misakinolide, and nosperin biosynthetic; other unspecified ribosomally synthesized and post-translationally modified peptide product (RiPP) cluster; Transatpks-NRPS, and Lassopeptide cluster. (AntiSMASH cluster numbers, region, similarity % and similar known cluster are shown, – indicates which do not have similar cluster in the database).

Table 5. The main precursor metabolites identified from the P. peoriae IBSD35 genome using KEGG pathway database analysis.

(GO: 0006935) identified from in silico genome analysis of P. peoriae IBSD35 confirmed the bacterium motility traits. It harbored TCS CheA/CheY, methyl-accepting chemoreceptor proteins (mcp, K03406), an adaptor protein CheW (K03408), and flagellar motor switch protein Flim (K02416). They are the core of bacterial chemotaxis signaling genes. They are known to involve in colonization and interaction with host plants. The analysis also revealed cellular motility proteins for the resources transport and substrate-binding protein of ribose (rbsB, K10439). These traits are required for mobility and thriving in heterogeneous and highly competitive environments. Endoglucanase [EC: 3.2.1.4] enzyme was visible in the understudy genome, and it is necessary for spreading to other intracellular tissues to enter the endophytic life stage (www.kegg.jp/kegg/kegg1.html).
Discussion

Antibiotics have revolutionized the treatment of tropical infections, communicable diseases, food-borne, and other poverty-related infectious diseases. However, the efficacy is short-lived and turned into a resilient clinical problem eliciting innumerable public health crises and causing economic burden \(^\text{2,4}\). Therefore, there is a continuous need for a new antimicrobial agent with high specificity, efficacy, efficiency, and anti-resistant to tackle the AMR challenge \(^\text{23,44}\). AMPs have prospects because of their antibacterial, anticancer, antiviral, antifungal, anti-aging, and anti-parasite, regulating cell proliferation, extracellular matrix production, cellular immune responses, and adjuvant characteristics \(^\text{13,59-62}\). Lately, many important SMs have reported from endophytes of medicinal plants from less explored sources which offered unprecedented opportunities to find novel antimicrobial agents \(^\text{56}\). Moreover, the availability of ever-increasing genome sequences, a comprehensive proteomics database, and advanced bioassay techniques offer a platform for the dereplication of old compounds, effective search, prediction, and design of novel AMP to counter AMR \(^\text{28,20}\).

*Paenibacillus peoriae* IBSD35 from the stock solution has retained antimicrobial activity against *S. aureus* ATCC 25923, indicating its inherent antimicrobial potential \(^\text{86}\). Endospore forming traits may help its endophytic association with the host plant \(^\text{87}\). Its neutralophilic characters might help neutralize its cytoplasm relative to the plant to meet the pH challenges. It might be through direct active uptake or efflux of protons from the cells or tissues \(^\text{88}\). These indirectly suggested host-plant physiology as well. The optimum growth condition of the strain IBSD35 recorded at 38°C, and pH 6.8–7 in LB medium indicated the mesophilic traits. The cultivable characters confirmed *P. peoriae* IBSD35’s ability to live as a free-living bacterium in laboratory conditions.

An antimicrobial protein mixture precipitated from LB fermentation broth with 70% ammonium sulfate suggests that at least one or more of the putative 25 BGCs of *P. peoriae* IBSD35 might express antimicrobial metabolites. The 700 mM NaCl eluent from DEAE-C column chromatography has retained activity against *S. aureus* ATCC 25923. The RP-HPLC peaks P4 and P5, absorption at λ205 and A214, suggested the presence of amino acids and peptide bonds with other impurities. The absorbance at λ280 confirmed the presence of proteins \(^\text{89,90}\). Antimicrobial protein total yield was low at 0.05%, which can be enhanced using an advanced purification system or by up-regulating its BGC expression. MIC of peoriaerin IBSD35 at 0.04 µg µl\(^{-1}\) indicating its high specific activity and potency against *S. aureus* ATCC 25923 compared with ampicillin and nisin at 1 mg ml\(^{-1}\).

The purified AMP has shown antimicrobial activity against *S. aureus* ATCC 25923, *E. coli* ATCC 25922, *C. albicans* ATCC 10231, *K. pneumoniae* ATCC 4352, and *S. typhimurium* ATCC 14028 connoted its broad-spectrum antimicrobial character (Tables S6, S7). The sample has retained activity after prolonged storage, and its resistance to degradative enzymes indicated its potential for drug development against AMR. Antimicrobial activity was altering at different stages of purification. It might be due to the disintegration of proteins. The treatment of SDS and high-temperature slightly increased the activity. It might be from the disintegration or aggregation of the proteins and other biomolecule complexes (Table S3). The sample can withstand autoclaving at 121°C for 15 min presenting a potential antimicrobial agent and food preservatives \(^\text{77,21,79}\).

The peptide sequences generated using LC–MS/MS coupled to a quadrupole time-of-flight (Q-TOF) had aligned with *P. peoriae* IBSD35 proteome using proteomics tools \(^\text{1,2,22,23,38}\). The sequence alignment using the Antimicrobial Peptide Database (APD3) tool indicated that peoriaerin IBSD35 is 38.23% similar to fermentin from *Lactobacillus fermentum* \(^\text{20,22}\). Peoriaerin IBSD35 is a trans–membrane helix with an extended strand of 10.71%, beta-turn of 21.43%, and a random coil of 64.00% (Fig. 3, Supplementary Fig. S2). It has a high percentage of helix and residues embedded in the membrane, suggesting at least some part of the sequence may embed in the hydrophobic region of the membrane \(^\text{10}\). Peoriaerin IBSD35 has a total net charge of +2 and a theoretical pI of 8.86. These properties indicated an antimicrobial peptide nature. The aliphatic index was 56.07, and the grand average of the hydropathicity index (GRAVY) was -0.132 \(^\text{25}\). Antimicrobial action was visible after 1 h of incubation. The slower kinetics of killing suggested that peoriaerin IBSD35 might penetrate the cell membrane and transport it inside the cell cytoplasm to cause the effect \(^\text{13}\). However, the leakage of DNA and proteins was lacking.

The draft genome sequencing and analysis of *P. peoriae* IBSD35 provided insights into its endophytic characters, antimicrobial traits, and interaction with the host plant \(^\text{10,11,17}\). The characteristics contributing to strain IBSD35 epiphytic fitness are iron acquisition, capacity to colonize plants, and stress-tolerant. Additional functional genes suggested *P. peoriae* IBSD35’s close association with the environment through nitrogen fixation \(^\text{14,37,40}\). However, there are about 980 (18.50%) protein-coding genes without assigning any function, as per IMG AP v.4.16.6 (Table S4), and their relationship to molecular networks of the database is unknown \(^\text{28,29,49}\). Possibly, they may be endophytic specific, which needs further study to ascertain.

The genes encoding flagellins and flagellar motor proteins are visible in *P. peoriae* IBSD35. They helped bacterial motility in the plant-associated lifestyle \(^\text{46,71}\). The flagellin proteins (Flg) also facilitated the endophyte to pass through the first line of the plant immune defense system. Moreover, the presence of flagellin (flc), glycerol kinase [EC: 2.7.1.30], elongation factor (Tu), and molecular chaperone (htpG) are crucial for plant–pathogen interaction indicating *P. peoriae* IBSD35 close association with the host plant \(^\text{72}\). Generally, motility enabled bacteria to move and colonize plants and systematically spread within the plant, promoting the endophytic lifestyle. The TCS genes represented a primary means by which an organism sensed and responded to stress and changing physiological environments \(^\text{45,71}\). The TCS generally helped the bacterial endospore DNA against heat, UV-radiation, and enzymic degradation \(^\text{45}\). They are responsible for sporation septum, engulfment, spore morphogenesis, and germination in harsh environments. The genome ATP-binding cassette (ABC) transporter proteins are other plausible candidates for genes mediating the export of small molecules across the cells \(^\text{41}\). The presence of siderophores and...
genes conferring tolerance to oxidative stress in the strain IBSD35 genome supported the proposed importance of oxidative stress tolerance to fitness in symbiotic association. It is consistent with *P. peariae* IBSD35 endophytic lifestyle. However, there was no evidence of the biosynthesis of phytoxins or cellulases.

The KEGG pathway database analysis (www.kegg.jp/kegg/kegg1.html) indicated that carbohydrate and amino acid metabolism accounted for a large proportion of the *P. peariae* IBSD35 genome's metabolic processes, and its CCM network analysis included GG, TCA, PPP metabolites, and pyruvate metabolism. Most organisms have shared a common core to their metabolic networks. The CCM pathways convert sugars into metabolic precursors and provide the entire cell biomass, including the SMs. *P. peariae* IBSD35 harbored a collection of metabolic and drug efflux systems and bioremediation gene clusters which indicated a need to protect against toxic concentrations of metabolites or metabolic analogs. It has shown the strain's capability for bioremediation, survival defense against oxidative stress, and production of antimicrobial compounds and toxins. A repertoire of NRPS (72%) and their expressions might neutralize the competing microbes and antagonistic biomolecules. Gene resistance to β-lactam, cationic AMPs, and antifolate with the antibiotic BGCs in the strain IBSD35 genome supported that the clusters might involve in antibiotic production, which is vital to avoid self-toxicity from its own antibiotics production.

The SMs gene clusters predicted from the *P. peariae* IBSD35 genome showed low similarity with other BGCs. It exhibited 100% similarity for fusaricidin and polymyxin BGCs in the MiBiG cluster database (Table 4). A repertoire of NRPS SMs might help microbes in existence and communication, and their elucidation will facilitate the finding of novel bio-molecules. They might also provide a competitive advantage in the battle for resources and adaptive mechanisms for survival in host plant cells or tissues. The different number of BGCs in the analysis of the genomes reflected their habitats, genetics, and functional differences. Thus, a difference in genomes' SM is a difference in environment and resource competition. Generally, two closely related strains share most of their genes and will not be unique to the individual species49. The in-silico identification of PKS, NRPS, or SM gene clusters revealed considerable differences between the closely related species supporting that *P. peariae* IBSD35 is a unique species slightly different from free-living or symbiotic rhizobium bacteria.

The terpenoids and polyketides biosynthetic are the backbones of other SMs. Pathwhiz has been used to illustrate the type II polyketide biosynthesis pathway of the *P. peariae* IBSD35 genome predicting its proteins, enzymes, SMs, and other products. Dihydrokalafungin, oxytetracycline, tetracycline, nogalamivine, ellomycin, and tetranymin F are predicted to be derived from simple carbon building blocks of the pathway (Supplementary Fig. S7). The KEGG pathway analysis (www.kegg.jp/kegg/kegg1.html) indicated that terpenoids have derived from 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate (MEP/DOXP) pathway in *P. peariae* IBSD35. The non-mevalonate Pathway of strain IBSD35 has the potential to produce isoprenoids, antibiotics, and anticancer alkaloids. However, the expressions of SM clusters might involve in antibiotic production, which is vital to avoid self-toxicity from its own antibiotics production.

The *P. peariae* IBSD35 genome harbored the putative genes of the nitrogen cycle. Nitrogen fixation by nitrate to ammonia and other biologically reduced forms might incorporate into amino acids and other vital compounds through nitrogen metabolism. Nitrogen fixation is energetically costly, and therefore, it is tightly regulated at the transcriptional level by sophisticated regulatory networks that respond to multiple environmental cues. Generally, two closely related strains share most of their genes and will not be unique to the individual strains. The in-silico identification of PKS, NRPS, or SM gene clusters revealed considerable differences between the closely related species supporting that *P. peariae* IBSD35 is a unique species slightly different from free-living or symbiotic rhizobium bacteria.

The Type (Strain) Genome Server annotation (TYGS) and PGAP annotation predicted that the strain IBSD35 has a close species relationship with other *P. peariae* strains. Phylogenetically, *P. peariae* IBSD35 is closely related to *P. peariae* FSL J3-0120. However, the protein Basic Local Alignment Search Tool (BLAST) indicated *P. peariae* IBSD35 close similarity with *P. polymyxa* SC2 (E-value: 0.0, Score: 2273, Ident.: 98.9%). Therefore, the understudy genome is within the species of *P. peariae* and related to the *P. polymyxa* species. The genomic data quality and analysis need further improvement as some of the selected genomes are draft and limited for web server analysis. The variation in genome size suggested that different strains of the same species from different habitats potentially produced their own unique NRPS products. Several gene clusters are common in selected *Paenibacillus* species suggesting their conservation in the lineage. *Paenibacillus* has only been designated a separate species since 1993. Some lineage-specific clusters and the predicted unique clusters in our understudy genome could be involved in SMs biosynthesis.

The unique genes or 2125 single-copy genes from the six *Paenibacillus* genomes might encode some of the functions in the cell, as they are found only in one organism. These genes might be involved in environmental adaptation, specific competition with other species, or speciation. The synteny dot plot indicated *P. peariae* IBSD35 resemblance to its genus. There is also evidence of gene sharing (Mitogen-activated protein kinase (MAPK) pathway from plant, nitrogen fixation gencluster) and horizontal transfer of genes.

Given the *P. peariae* IBSD35 potential applications in agriculture, medicine, industries, and bio-remedy, its genome sequencing and its extracellular metabolite analysis are crucial to the basic and applied research to find novel antimicrobial agents. Genome mining has enabled the prediction of the putative BGC region in the *P. peariae* IBSD35 by linking a protein group from LC–MS/MS data and connecting these natural products to the genes. The genome information and the proteomics analysis provided here will further enhance the study of environmental information processing and its antimicrobial pathways. It will also facilitate the understanding of evolutionary relationships among *Paenibacillus* species. Moreover, the less explored ecological niches provide new sources of an unprecedented opportunity to find novel microbes and metabolites to counter AMR to our already exhausted sources of bio-molecules.
Conclusions

*Paenibacillus peoriae* IBSD35 is an endophytic bacterium. It is a mesophile, endophyte forming, and neutralophile bacterium. It has a linear DNA topology with a total length of 5,862,582 bp. It harbored 25 SMs BGCs, out of which 18 BGCs are NRPs. The *P. peoriae* IBSD35 closest species reference genome is *P. peoriae* KCTC 3763 with a symmetrical identity of 62.7495% and a symmetrical identity of 89.4884% with *P. peoriae* FSL J3-0120. The CFS of *P. peoriae* IBSD35 exhibited a broad spectrum of antimicrobial activity against *S. aureus* ATCC 25923, *E. coli* ATCC 25922, *C. albicans* ATCC 10231, *K. pneumoniae* ATCC 4352, and *S. typhimurium* ATCC 14028.

*P. peoriae* IBSD35 inoculum’s optimum growth pH was observed at 6.8–7. It is a mesophile bacterium that grows best at 38 °C. The supernatant antimicrobial activity arbitrary unit was 1600 AU ml−1. The putative extracellular AMPs were purified using salting out and chromatography. A 15 proteins list and AMP sequences have predicted from LC–MS coupled to Q-TOF. An AMP from the protein lists with accession no. A0A2S6NUT6 was selected. It has been given the name peoriaerin IBSD35, based on its source name. The amino acid sequence of peoriaerin IBSD35 is AAGIQAQAGFGLSDSIQGTGKQKSCKF with a MW of 2802.17 Da. Its sequence backbone has a large portion of helix and residues embedded in the membrane. It exhibited high antimicrobial activity against *S. aureus* ATCC 25923 at range of 0.04 µg µl−1.

*Paenibacillus peoriae* IBSD35 exhibited a close phylogenetic relationship with *P. peoriae* FSL J3-0120 and *P. polymyxa* CR1. The genome analysis indicated that it harbored nitrogen fixation genes and other putative nitrogen-fixing proteins but without a nitrogen fixation regulator (nifL/A). Its genome has a repertoire of 18 NRPS BGCs. The type II polyketide biosynthesis pathway revealed that terpenoids might be exclusively through (MEP/DOXP) pathway in *P. peoriae* IBSD35. Combining the genomics, mass spectrometry, and proteomics results obtained from *P. peoriae* IBSD35 through its extracellular biomolecules will enhance our search for AMPs that are stable, specific, and effective to counter AMR. The potential for applying comparative genomics to get evolutionary insights and discover novel AMPs from less explored regions will increase the probability of finding novel bio-molecules.

Materials and methods

**Plant ethics statements.** *Millettia pachycarpa* plant was used to isolate an endophyte, *Paenibacillus peoriae* IBSD35 which is used in this study. It is a commonly found wild species in Ukhrul, Manipur, India. Experimental research and field studies on plants including the collection of plant is complied with National guidelines and legislation. *Millettia pachycarpa* is not included in the list of IUCN Species at Risk of Extinction.

The stem of *Millettia pachycarpa* plant was collected from Phalee village (25.143524 N & 94.28334 E, 1533.14 M), Ukhrul, Manipur, India. Permissions to collect plant material were obtained from Phalee BMC by payment of fee as per NBA guidelines for collecting plant from area falls within the territorial jurisdiction under sub-section (3) of section 41 of the Act (Biological Diversity Act, 2002).

**P. peoriae** IBSD35 culture and its morphology and genomic characters. An endophytic bacterium *P. peoriae* IBSD35 isolated from the stem of *M. pachycarpa* in our previous experiment was picked in our present context9–9. It was sub-cultured from 15% phosphate buffer saline (PBS) glycerol stock preserved in the Institute of Bioresources and Sustainable Development-Microbial Resources Division (IBSD-MRC) repository (Accession No. MRC-75001) by thawing for 5 min at room temperature (27 °C)99. A 100 µl of stock culture was pipetted using a micropipette and spread on a 25 ml LB agar plate. It was incubated at 38 °C to check the contamination and colony purity84. A single colony was streaked and sub-cultured repeatedly to maintain its purity and colony morphology.

Screening of antimicrobial activity. A 200 ml of *P. peoriae* IBSD35 inoculums was harvested by centrifugation (Sorvall ST 16R Centrifuge) at 3200×g for 15 min at 4 °C, and the supernatant was collected in an Eppendorf tube88. The gDNA was extracted from an overnight grown culture using the cetyl trimethyl ammonium bromide (CTAB) method40. The quality and quantity of the gDNA were determined by A260/A280 ratio using UV-spectrophotometer (Nanodrop, Shimadzu Biotech Bio Spec-nano) and assessed in a 0.8% agarose gel electrophoresis. The genomic and molecular characters of the genome were analyzed by draft genome sequencing using Illumina HiSeq 2500 platform97.

Optimum growth pH and temperature for antimicrobial production. 2 l of the production medium (pH 7 ± 2) was inoculated with an overnight grown culture of *P. peoriae* IBSD35 and incubated in a shaker incubator at 80 rpm (Eppendorf innova R42 Rotary shaker). Its optimum pH, OD, and temperature concerning antimicrobial activity were determined by aseptically aspirating out the inoculum from the fermentation broth86. The production medium was harvested by centrifugation at 3200×g for 15 min at 4 °C. The supernatant...
antimicrobial activity was tested by serial dilution, and its arbitrary unit was calculated against *S. aureus* ATCC 25923. To rule out any false-positive inhibition caused by acid production and catalytic enzyme, 0.6% of calcium carbonate and a unit (10 mg/ml) of 0.01% catalase were added to the production medium. Toluene was added to avoid contamination from other microbes.

Precipitation and purification of the antimicrobial protein mixture. The 6-day-old inoculum of 2 l of CFS was harvested from the 6-day-old inoculums by centrifugation at 3200×g for 15 min at 4 °C. Its pH was adjusted to 3 with 2 N hydrochloric acids (HCl) to kill the experimental strain and re-adjusted to the original harvesting state (pH 6.8) after 4 h by adding 2 N sodium hydroxide (NaOH). The supernatant pH was re-adjusted to the original harvesting state (pH 6.8) by adding 2 N NaOH. The CFS was filtered through a 0.2 μm filter (Avixa) and precipitated in 70% ammonium sulfate (NH₄)₂SO₄) followed by centrifugation (3200×g, 35 min at 4 °C). The precipitated protein from centrifugation was collected. The resulting pellet was re-suspended in 30 ml distilled water and dialyzed using PBS in a dialysis membrane (12.4 kDa cut-off) with intermittent buffer changing.

The pelleted was re-dissolved in 30 ml of distilled water and loaded in diethylaminoethyl cellulose (DEAE-C (Sigma Aldrich). A weakly basic resin, DEAE-cellulose was loaded for anion exchange column chromatography. Sample was eluted out using 0.1, 0.2, 0.3, 0.4, 0.5 and 0.7 M or 100–700 mM NaCl gradient elution and the OD were recorded at selected wavelength of λ205, λ214, and λ280. The fractions were pooled together and lyophilized into powder form and tested for retention of antimicrobial activity according to the CLSI assay. Fractions showing antimicrobial activity were pooled together for lyophilization and further purified with RP-HPLC.

RP-HPLC purification of protein mixture. The partially purified sample was again purified using ZORBAX 300 SB-C-18 semi-preparative (9.4×250 mm) 5 microns. The column was irrigated with 0.1% TFA in 20% ACN for 45 min at an isocratic flow rate. The sample (150 μl per experiment) was loaded on the HPLC (UFLC Shimadzu) with semi-preparative Agilent ZORBAX 300SB reverse-phase C-18 of 5 μm and 9.4×250 mm column and run with 0.1% ion-pairing reagent, trifluoroacetic acid (TFA) in an isocratic system at a flow rate of 2.5 ml min⁻¹. Eluents were monitored at a wavelength of 205, 214, and 280 using a UV-spectrometer (Eppendorf, Biospectrometer) and collected manually. The fractions which exhibited higher antimicrobial activity were pooled together and collected separately for further analysis. The total protein concentration and activity, specificity, yield percentage, and purification fold of the extract compared to the initial starting crude extract were calculated using UV-Spectroscopy. The UV-spectrophotometric method for antimicrobial protein concentration was calculated as: concentration (mg/ml) = (1.55× A280 (0.76× A260) × dilution factor.

The effect of external factors on sample antimicrobial activity. The sample was evaluated for sensitivity to degradative enzymes, detergents, organic solvents, temperature, and pH changes. The retention of antimicrobial activity after the treatment was qualitatively determined through cut-well agar bioassay against the reference pathogen, *S. aureus* ATCC 25923 as described earlier. For the pH stability test, the sample dissolved in dist. water was pre-adjusted to pH 2.0–8.9, followed by incubation at 38 °C for 12 h. The sample temperature sensitivity was tested by exposing to temperatures ranging from 20 to 100 °C at the interval of 10 °C from 2–3 h, and additionally at 121 °C for 15 min by autoclaving.

The samples were treated with different degradative enzymes at the concentration of 1 mg ml⁻¹ for 2–3 h at room temperature to test the stability against enzymes. The organic solvent and test sample were mixed in a 1:1 ratio and incubated at 27 °C to evaporate all the organic solvent, and the samples were assessed for retention of antimicrobial activity. The samples were treated with detergents in a 1:1 ratio and checked for retention of antimicrobial activity. Negative control consists of PBS (pH 7) and the untreated sample solution was used as a positive control.

Mass spectrometric analysis of peptide mixture. The experiment was performed using EASY-nLC 1000 system (Thermo Fisher Scientific) coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific) equipped with a nano-electrospray ion source. 1.0 μg of the RP-HPLC purified peptide mixture was resolved using a 15 cm PicoFrit column (360 μm outer diameter, 75 μm inner diameter, 10 μm tip) filled with 1.9 μm of C18-resin (Dr. Maier, Germany). The peptides were loaded with buffer A and eluted with a 0–40% gradient of buffer B (95% ACN, 0.1% formic acid) at a flow rate of 300 nl/min for 90 min. The source heater was set at 300 °C to turn the sample into a vapor for analysis. The UPLC-separated samples were then analyzed by electrospray ionization at 4.5 kV and the peptides were dissociated with collision induced dissociation (CID) for MS/MS spectra. MS data was acquired using a data-dependent top10 method dynamically choosing the most abundant precursor ions from the survey scan.

One sample was processed and 1 RAW file generated was analyzed with a Proteome Discoverer against *P. peoriae* IBSD35 Uniprot reference proteome database. For the Sequest search, the precursor and fragment mass tolerances were set at 10 ppm and 0.5 Da, respectively. Trypsin was used to generate peptides, carbamidomethyl on cysteine as fixed modification, and oxidation of methionine, and N-terminal acetylation was considered as variable modifications for database search. Both peptide spectrum match (PSM) and protein false discovery rate (FDR) were set to 0.01 FDR.

Physiochemical properties of the protein and AMPs were analyzed using the ExPASy-Protparam tool (http://web.expasy.org/protparam) and APD3 Antimicrobial Peptide Calculator. The secondary structure was...
analyzed using SOPMA. The 3-D conformation structure of the peptide was predicted using PEP-FOLD 3 de novo modeling and deposited in ModelArchive. The model was built based on the target-template alignment using ProMod3.

**Peoraerin IBSD35 mode of action imaging through scanning electron microscopy.** The AMP mode of action against the test pathogens was checked by the absorbance at λ260 and λ280 with UV-spectrometry. A direct visualization of peptide-induced S. aureus cellular damage and morphological changes were evaluated by SEM. The samples were prepared by incubating S. aureus ATCC 29523 with the test sample as described earlier. S. aureus were treated with antimicrobial agent at 1× MIC for 2–3 h and incubated at 30 °C±0.2. The inoculants were harvested by centrifugation as described earlier. The pellet was washed with 0.1 M PBS (pH 7.2) and fixed in 2.5% glutaraldehyde and 2% paraformaldehyde made in PBS. The fixed sample was further processed for coating and imaging. The pathogen without any treatment served as the experimental control.

**Genome sequencing and assembly.** The gDNA integrity was confirmed using a 0.8% agarose gel electrophoresis. The draft genome sequencing was performed through Illumina HiSeq 2500 system. The cleaned reads were subjected to the Kmer genie to predict the optimal k value and assembly size. De novo assembly was performed using AbYSS v.2.0. The predicted genes were matched with bacterial proteins from the Uniprot database using the BLASTX program. Genes that could establish homology relationships (E-value ≤ 10−5 and similarity score ≥ 40%) were retained in the annotation pipeline for further annotation while the others remain unannotated. Comparative genomic analysis. The close strains of *P. peoriae* IBSD35 (NZ_PTM01000000) were selected from the NCBI database (*P. peoriae* KCTC 3763, *P. polymyxa* E681(NZ_CP048793), *P. polymyxa* SC2(NC_014622), *P. vortex* V453 (NZ_ADH01000000), *P. terrae* HPL-003(NC_016641), *Paenibacillus* sp. JDR-2(NC_012914), *P. polymyxa* CR1(NC_023037), *B. subtilis* subtilis AG1839 (NZ_CP008698), *S. coelicolor* A3(2) (AL645882), and *P. peoriae* HS311 (NZ_CP011512)). *Bacillus subtilis* subtilis AG1839 is selected because *Paenibacillus* represents a new genus from *Bacillus*. *S. coelicolor* A3(2) is the model organism for SM BGC studies. *P. peoriae* HS311 represents the completely sequenced genomes of *Paenibacillus* and serves as the reference for the natural variation within the genus. *P. peoriae* KCTC 3763 represents the closest species reference genome with a symmetrical identity of 62.7495%.

**Comparative genomic analysis.** OrthoVenn was used for Venn diagram drawing which is based on the protein sequences clusters (E-value = 1e−2 and inflation value = 1.5). Web Gene Ontology Annotation Plot (WEGO) was used to com-reference genome with a symmetrical identity of 62.7495%.

**Genome annotation and bioinformatics analysis.** JGI-IMG Annotation Pipeline v.4.16.6 were used for gene annotation and the NCBI PGAP were used for gene annotation. The gene functional prediction was performed by BLAST against the databases.

**Comparative genomic analysis.** The close strains of *P. peoriae* IBSD35 (NZ_PTM01000000) were selected from the NCBI database (*P. peoriae* KCTC 3763, *P. polymyxa* E681(NZ_CP048793), *P. polymyxa* SC2(NC_014622), *P. vortex* V453 (NZ_ADH01000000), *P. terrae* HPL-003(NC_016641), *Paenibacillus* sp. JDR-2(NC_012914), *P. polymyxa* CR1(NC_023037), *B. subtilis* subtilis AG1839 (NZ_CP008698), *S. coelicolor* A3(2) (AL645882), and *P. peoriae* HS311 (NZ_CP011512)). *Bacillus subtilis* subtilis AG1839 is selected because *Paenibacillus* represents a new genus from *Bacillus*. *S. coelicolor* A3(2) is the model organism for SM BGC studies. *P. peoriae* HS311 represents the completely sequenced genomes of *Paenibacillus* and serves as the reference for the natural variation within the genus. *P. peoriae* KCTC 3763 represents the closest species reference genome with a symmetrical identity of 62.7495%.

**Genome sequencing and assembly.** The gDNA integrity was confirmed using a 0.8% agarose gel electrophoresis. The draft genome sequencing was performed through Illumina HiSeq 2500 system. The cleaned reads were subjected to the Kmer genie to predict the optimal k value and assembly size. De novo assembly was performed using AbYSS v.2.0. The predicted genes were matched with bacterial proteins from the Uniprot database using the BLASTX program. Genes that could establish homology relationships (E-value ≤ 10−5 and similarity score ≥ 40%) were retained in the annotation pipeline for further annotation while the others remain unannotated.

**Comparative genomic analysis.** OrthoVenn was used for Venn diagram drawing which is based on the protein sequences clusters (E-value = 1e−2 and inflation value = 1.5). Web Gene Ontology Annotation Plot (WEGO) was used to compare the provided gene datasets and plot the distribution of GO annotation into a histogram. SMs biosynthesis gene clusters were analyzed with AntiSMASH V.5 and JGI-IMG/ABC, and the type II polyketide biosynthesis pathway was drawn using KEGG and PathWhiz tool. All comparative analyses were performed using IMG/M and RAST tools.

**Genome-based taxonomy and phylogenetic tree.** Genome-based taxonomy was determined using TYGS which infers genome-scale phylogenies and estimates for species and subspecies boundaries and automatically determined closest type genome sequences. The pair-wise digital DNA:DNA hybridization (dDDH) values between user genomes and the selected types-strain genomes were performed. The dDDH values are provided along with their confidence intervals (C.I.) and difference of G + C% using the Genome BLAST Distance Phylogeny (GBDP) formula. The average nucleotide identity with the closely related species was determined using the OAT tool. The phylogenetic tree was inferred with FastME 2.1.6.1 from GBDP distances calculated from genome sequences. The AMP model was built based on the target-template alignment using ProMod.

**Data availability**

The datasets generated or analyzed during the current study are given in the table form with the name, accession number, and the database or web link.

| Name                           | Accession number | Database/web link                  |
|--------------------------------|------------------|------------------------------------|
| *Paenibacillus peoriae* IBSD35 | BioProject: PRJNA224116 | NCBI                               |
| Unknown protein from LC–MS experiment | A0A2S6NUT6 | Uniprot                             |
| Peoraerin IBSD35              | ma-id74(DOI 10.5452ma-id74) | ModelArchive                       |
| Nitrogen metabolism (map00910) | Ref: 210413      | https://www.kegg.jp/kegg/           |
| Biosynthesis of type II polyketide backbone | PW122264 | https://smpdb.ca/pathways/ pathways/PW122264 |

ma-id74(DOI 10.5452ma-id74) will be activated when we submitted the accepted paper to ModelArchive. Ref: 210413—A Permission document from KEGG is provided separately.
References

1. Fischbach, M. A. & Walsh, C. T. Antibiotics for emerging pathogens. Science 325, 1089–1093. https://doi.org/10.1126/science.1176667 (2009).

2. World Health Organization. Antimicrobial Resistance: Global Report on Surveillance (World Health Organization, 2014).

3. Gouguillas, C., Clark, J. M. & Shaw, L. J. The role of soil microbes in the global carbon cycle: Tracking the below-ground microbial processing of plant-derived carbon for manipulating carbon dynamics in agricultural systems. J. Soil Sci. Agric. 94, 2362–2371. https://doi.org/10.1002/psa.6577 (2014).

4. European Centre for Disease Prevention and Control. Surveillance of Antimicrobial Resistance in Europe 2018 (ECDC, 2019).

5. Castillo, U. F. et al. Munumbincins, wide spectrum antibiotics produced by Streptomyces munumbi, endophytic on Kennedia nigricans. Microbiology 148, 2675–2685. https://doi.org/10.1099/00221287-148-9-2675 (2002).

6. Chen, Y. T. et al. Antitumor activity of bacterial exopolysaccharides from the endophyte Bacillus amyloliquefaciens sp. isolated from Ophiopogon japonicus. Oncol. Lett. 5, 1787–1792. https://doi.org/10.3892/ol.2013.1284 (2013).

7. Rosenbluth, M. & Martinez-Bonero, E. Bacterial endophytes and their interactions with hosts. Mol. Plant Microbe Interact. 19, 827–837. https://doi.org/10.1099/MPMI-19-0827 (2006).

8. Reinhold-Hurek, B. & Hurek, T. Living inside plants: Bacterial endophytes. Curr. Opin. Plant Biol. 14, 435–443. https://doi.org/10.1016/j.pbi.2011.04.004 (2011).

9. Gudlaugsson, S. et al. Polyphasic reassessment of the genus Paenibacillus, reclassification of Bacillus lausius (Nakamura 1984) as Paenibacillus lausius comb. nov. and of Bacillus porieae (Montefusco et al. 1993) as Paenibacillus porieae comb. nov., and emended descriptions of P. lausius and of P. porieae. Int. J. Syst. Bacteriol. 46, 988–1003. https://doi.org/10.1099/00207713-46-4-988 (1996).

10. Grady, E. N., MacDonald, J., Liu, L., Richman, A. & Yuan, Z. C. Current knowledge and perspectives of Paenibacillus. A review. Microb. Cell Fact. 15, 203. https://doi.org/10.1186/s12934-016-0603-7 (2016).

11. Ngasangva, N., Indira, S. D. & Kalita, M. C. Screening of endophytes from traditionally used medicinal plants of Manipur for their antimicrobial activity: An impact towards future drug discovery. Int. J. Sci. Res. Biol. Sci. 6, 39–47. https://doi.org/10.26043/jirsbs/v6is19 (2019).

12. Singhal, A. K., Sharma, R. P., Baruah, J. N., Govindan, S. V. & Herz, W. Rotenoids from roots of Ophiopogon japonicus. Acta Pharmacol. Sin. 21, 949–951. https://doi.org/10.1038/s41393-9242282103-91 (2003).

13. Srivastava, R. C. Traditional knowledge of Nyishi (Dafla) tribe of Arunachal Pradesh. Confl. Res. 36, 3551–3567 (1999).

14. Baindara, P. et al. Characterization of the antimicrobial peptide penisin, a class Ia novel lantibiotic from Paenibacillus sp. strain A3. Antimicrob. Agents Chemother. 60, 580–591. https://doi.org/10.1128/AAC.01813-15 (2015).

15. Epand, R. M. & Vogel, H. J. The diversity of antimicrobial peptides and their mechanisms of action. Biochim. Biophys. Acta 1462, 11–28. https://doi.org/10.1016/S0005-2736(99)00198-4 (1999).

16. Wang, W. G., Liang, X. & Wang, Z. APD3: The antimicrobial peptide database as a tool for research and education. Nucleic Acids Res. 44, D1087–D1093. https://doi.org/10.1093/nar/gkv1278 (2016).

17. Pavlova, A. S. et al. Identification of antimicrobial peptides from novel Lactobacillus fermentum strain. Protein J. 39, 73–84. https://doi.org/10.1007/s10930-019-09789-8 (2020).

18. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567 (1999).

19. Ma, B. et al. PEAKS: powerful software for MS/MS peptide de novo sequencing. Rapid Commun. Mass Spectrom. 20, 2337–2342. https://doi.org/10.1002/rcm.1196 (2003).

20. Seidler, I., Zinn, N., Boehm, M. E. & Lehmann, W. D. De novo sequencing of peptides by MS/MS. Proteomics 10, 634–649. https://doi.org/10.1002/pmic.200900459 (2010).

21. Gimerancio, P. et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158, 412–421. https://doi.org/10.1016/j.cell.2014.06.034 (2014).

22. Gasteiger, E. et al. Protein identification and analysis tools on the ExPaSy Server. In The Proteomics Protocols Handbook (ed. Walker, J. M.) (Humana Press, 2005).

23. Lammable, A. et al. PEP-FOLD3: Faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res. 41, W449–W454. https://doi.org/10.1093/nar/gkw329 (2016).

24. Schwede, T. et al. Outcome of a workshop on applications of protein models in biomedical research. Structure 17, 151–159. https://doi.org/10.1016/j.str.2008.12.014 (2009).

25. Mukherjee, S. et al. Genomes Online Database (GOLD) vol 8: Overview and updates. Nucleic Acids Res. 49, D723–D733. https://doi.org/10.1093/nar/gkaa983 (2021).

26. Markowitz, V. M. et al. The integrated microbial genomes (IMG) system in 2007: Data content and analysis tool extensions. Nucleic Acids Res. 36, 846. https://doi.org/10.1093/nar/gkm846 (2008).

27. Tatusova, T. et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 43, 169–177. https://doi.org/10.1093/nar/gkw569 (2016).

28. Papagianni, M., Avramidis, N., Filioussis, G., Dasiou, D. & Ambrosiadis, I. Determination of bacteriocin activity with bioassays and of Paenibacillus lautus sp. strain. Protein J. 34, 1787–1792. https://doi.org/10.1007/s12934-016-0603-7 (2016).

29. Junqueira, M. et al. Protein identification pipeline for the homology-driven proteomics. J. Proteom. 71, 346–356. https://doi.org/10.1016/j.jprot.2008.07.003 (2008).

30. Apweiler, R. et al. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 1, D115–D119. https://doi.org/10.1093/nar/gkl131 (2004).

31. Korostin, D. et al. Comparative analysis of novel MGISEQ-2000 sequencing platform vs Illumina HiSeq 2500 for whole-genome sequencing. PLoS ONE 15, e0230301. https://doi.org/10.1371/journal.pone.0230301 (2020).

32. Jackman, S. D. et al. ABYSS 2.0: Resource-efficient assembly of large genomes using a bloom filter. Gen. Res. 27, 768. https://doi.org/10.1101/gr.214346.116 (2017).

33. Makarova, K. S. et al. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. 9, 467–477. https://doi.org/10.1038/nrmicro2577 (2011).

34. Overbeek, R. et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42, D206–D214. https://doi.org/10.1093/nar/gkt1226 (2014).

35. Jia, Y. et al. WEGO: A web tool for plotting GO annotations. Nucleic Acids Res. 34, W293–W297. https://doi.org/10.1093/nar/gkl031 (2006).
40. Wang, L. Nitrogen regulation in bacteria and archaea. Annu. Rev. Microbiol. 61, 349–377. https://doi.org/10.1146/annurev.micro.61.080706.093409 (2007).

44. Lewis, V. G., Ween, M. P., & McDevitt, C. A. The role of ATP-binding cassette transporters in bacterial pathogenicity. Sci. Rep. 4, 249, 919–942. https://doi.org/10.1038/srep00360-8 (2014).

47. Hathout, Y., Setlow, B., Cabrera-Martinez, R. M., Fenselau, C., & Setlow, P. Small, acid-soluble proteins as biomarkers in mass spectrometry of Bacillus spores. Appl. Environ. Microbiol. 69, 1100–1107. https://doi.org/10.1128/AEM.69.2.1100-1107.2003 (2003).

48. Falke, J. J., Bass, R. B., Butler, S. L., Chervitz, S. A., & Danielson, M. A. The two-component signaling pathway of bacterial chemotaxis: A molecular view of signal transduction by receptors, kinases, and adaptation enzymes. Annu. Rev. Cell Dev. Biol. 13, 457–512. https://doi.org/10.1146/annurev.cellbio.13.1.457 (1997).

50. Blin, K. AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. Nat. Chem. Biol. 47, W81–W87. https://doi.org/10.1038/nchembio.2310 (2019).

51. Palarangan, K. IMG-ABC vol 5.0: An update to the IMG/Atlas of biosynthetic gene clusters knowledgebase. Nucleic Acids Res. 48, D422–D430. https://doi.org/10.1093/nar/gkz932 (2020).

52. Demain, A. L. & Fang, A. The natural functions of secondary metabolites. Adv. Biochem. Eng. Biotechnol. 69, 1–39. https://doi.org/10.1007/3-540-44964-7_1 (2000).

55. Osbourn, A. Secondary metabolic gene clusters: Evolutionary toolkits for chemical innovation. Nat. Rev. Mol. Cell Biol. 19, 455–476. https://doi.org/10.1038/nrm3327 (2018).

57. Noor, E., Eden, E., Milo, R., & Alon, U. Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. Mol. Cell 39, 663–664. https://doi.org/10.1016/j.molcel.2010.08.034 (2010).

60. Green, E. R. & Mecsas, J. Bacterial secretion systems: An overview. J. Bacteriol. 186, 1100–1107. https://doi.org/10.1128/JB.186.3.1100-1107.2004 (2004).

61. Koretke, K. K., Lupas, A. N., Warren, P. V., Rosenberg, M. & Brown, J. R. Evolution of two-component signal transduction. Annu. Rev. Genet. 41, 1–7 (2007).

69. Liu, S., Takala, T. M., Wan, X., Reunanen, J., & Saris, P. E. J. Cell-mediated killing of L. monocytogenes by leucocin C producing Millettia pachycarpa. PLoS Genet. 8, e1002541. https://doi.org/10.1371/journal.pgen.1002541 (2012).

72. Muir, R. E. & Gober, J. W. Regulation of late flagellar gene transcription and cell division by flagellum assembly in Caulobacter crescentus. Mol. Microbiol. 110, 117–130. https://doi.org/10.1111/j.1365-2958.2001.02506.x (2001).

75. Choi, S. K. The molecular architecture of the two-component sensor of two-component signaling system. Trends Genet. 11, 302–309. https://doi.org/10.1016/S0168-9525(00)01945-8 (2000).

80. Koo, C. et al. Isolavonoids with antiestrogenic activity from Pseudomonas aeruginosa. J. Nat. Prod. 69, 10774. https://doi.org/10.1021/acs.jnatprod.9b00479 (2019).

101. Simonian, M. D. Spectrophotometric determination of protein concentration. Curr. Protoc. Toxicol. 21, 1–7 (2004).

121. Liu, S., Takala, T. M., Wan, X., Reunanen, J., & Saris, P. E. J. Cell-mediated killing of L. monocytogenes by leucocin C producing Escherichia coli. Microbiol. Res. 168, 300–304. https://doi.org/10.1016/j.micres.2012.11.011 (2013).

135. Ryzhykova, Y. L., et al. Molecular model of a sensor of two-component signalling system. Sci. Rep. 11, 10774. https://doi.org/10.1038/s41598-021-89613-6 (2021).

153. Thung, T. T., Tyler, B. M. & Setubal, J. C. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiol. 5(Suppl 1), S2. https://doi.org/10.1186/1471-2180-5-S1-S2 (2005).

172. Choi, S. K. et al. Identification and functional analysis of the fusaricidin biosynthetic gene of Pseudomonas syringae. Biochem. Biophys. Res. Commun. 365, 89–95. https://doi.org/10.1016/j.bbrc.2007.07.086 (2007).

187. Wang, Y., Coleman-Derr, D., Cheng, G. & Gu, Y. Q. OrthoVenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 43, W78–W84. https://doi.org/10.1093/nar/gkv487 (2015).

193. Gorka, J. et al. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91. https://doi.org/10.1099/ijs.0.04484-3 (2007).

208. Pattnaik, B. & Lindberg, P. Terpenoids and their biosynthesis in cyanobacteria. Life. 5, 269–293. https://doi.org/10.3390/life5010026 (2015).
for the Scanning Electron Microscope. The draft genome sequencing was performed at Agri Genome Labs (P) Ltd, Kerala.

Author contributions
N.N. performed the experiment and wrote the main manuscript and prepared all the figures and tables. C.K.S. performed the RP-HPLC. P.K.M., N.N., S.I., and M.C.K. initiated the ideas and proofread the manuscript. The IBSD manuscript no. is IBSD/MS/2020/01/094.

Funding
This work was supported by Department of Biotechnology of the Ministry of Science and Technology, Government of India.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-23613-y.

Correspondence and requests for materials should be addressed to I.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022