Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative genomics of Paenibacillus secondary metabolism: unveiling the putative biosynthetic gene cluster for Paenialvins in Paenibacillus alvei strain 32

Drago Haas[1]
Fatoumata Tambadou[1]
Thibault Caradec[1]
Valérie Sopéna[1]
Sandrine Didelot[1]
Cyrille Barthélémy[1]
Valérie Thiéry[1]
Romain Chevrot[1]
romain.chevrot01@univ-lr.fr

Research Article

Keywords: Paenibacillus alvei, non-ribosomal peptide synthetase, paenialvin, polymyxin, methicillin-resistant Staphylococcus aureus (MRSA), cystic fibrosis, comparative genomics

Posted Date: June 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9942673/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Probiotics and Antimicrobial Proteins on September 4th, 2026. See the published version at https://doi.org/10.1007/s12602-026-11167-8.

Abstract

In this study, we used comparative genomics and culture-based methods to investigate Biosynthetic Gene Clusters (BGCs) responsible for the production of antimicrobial peptides. *Paenibacillus alvei* strain 32 was isolated from a cystic fibrosis sputum. Its genome was sequenced using Illumina, showing a size of 6,584,590 bp with 239 contigs assembled in 26 scaffolds, an average coverage of 243X, and 6,832 coding sequences. ANI analysis and *in silico* DNA-DNA hybridization showed its affiliation inside *Paenibacillus alvei*, with a clear separation from other related strains, leading us to propose a distinct species-level genomic clade (genomospecies) within this group. AntiSMASH analysis predicted 22 putative BGCs in the genome of strain 32. Its culture supernatant exhibited inhibitory activity against Gram-positive pathogens, including methicillin-resistant *Staphylococcus aureus* (MRSA), *Bacillus cereus*, and *Enterococcus faecalis*. By comparing *in silico* BGC predictions with activities described in the literature, we propose that strain 32 harbours a specific 110-kb cluster (cluster 6.2) with five non-ribosomal peptide synthetase (NRPS) genes. These synthetases are predicted to direct the assembly of a 16-amino acid backbone that correlates with the structure of paenialvins, which are known anti-MRSA molecules. This study describes the putative biosynthetic pathway of the paenialvins and explains structural variations, bringing useful data on *Paenibacillus* secondary metabolism for future antibiotic development.

Introduction

The discovery and medicinal use of antibiotics in the early 1940s significantly increased human life expectancy, and made people think, at least temporarily, that bacterial infections were a problem of the past [1]. However, the overuse and misuse of antibiotics have driven the rapid evolution of resistant pathogens, posing a major challenge for modern healthcare. Multi-drug resistant (MDR) bacteria are very dangerous in hospitals where they cause severe infections. Therefore, we need to find new antibacterial molecules to fight these resistant pathogens [1].

Historically, most clinical antibiotics have been discovered by screening microorganisms from diverse habitats, such as soil and marine ecosystems. The lungs of individuals with cystic fibrosis (CF) are another unique environment where intense microbial competition occurs. Because of CFTR protein deficiency, the resulting thick mucus promotes persistent bacterial infections and a highly complex microbiota. We previously used a screening method on sputum from a CF patient, leading to the isolation of a *Paenibacillus alvei* strain (named 32). This strain showed strong antimicrobial activity against *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Klebsiella pneumoniae* [2]. This isolate belongs to the *Paenibacillus* genus, which contains motile, Gram-positive, spore-forming bacteria found in diverse habitats [3, 4]. Separated from *Bacillus* in 1993, the genus is known for its rich secondary metabolism and its ability to produce many bioactive compounds [5, 6].

Paenibacilli are known for producing polymyxins, a class of antimicrobial peptides discovered between 1947 and 1948 [7, 8]. Polymyxin B and E (colistin) are efficient against Gram-negative bacteria, but their

clinical use was limited in the past. Due to their cationic nature attributed to five free amino groups on diaminobutyric acid (Dab) residues, they cause toxicity in kidneys and on the nervous system. Today polymyxins are used as last-resort treatments for severe MDR Gram-negative infections, especially in cystic fibrosis care [9]. However, sequencing projects show that the real potential of this genus is bigger than expected. A study of 284 high-quality *Paenibacillus* genomes predicted a total of 3,273 biosynthetic gene clusters (BGCs), meaning 11.25 BGCs per genome on average. Only 16.87% of these clusters had high similarity with known data in the MIBiG database, leaving the vast majority (83.13%) as unknown [10].

This reservoir of unknown BGCs can be a source of new molecules of interest. A good example is the paenialvins, a group of antibiotics recently characterized from the supernatant of *Paenibacillus alvei* DSM 29 [11–12]. This family has four peptide variants (A-D) with molecular masses between 1875 Da and 1923 Da, made of an 11-amino-acid linear chain and a 5-amino-acid lactone ring. These molecules kill methicillin-resistant *S. aureus* (MRSA) with low MIC values (0.8–3.2 µg/mL) as well as *Escherichia coli* and *P. aeruginosa*. However, the authors did not identify the BGC responsible for paenialvin bioproduction [11]. This aligns with genomic profiles of related *P. alvei* isolates, such as strain JR949, in which 58% of the BGCs (18 out of 31) display low (< 25%) or no homology to characterized pathways, confirming the high rate of metabolic novelty in this taxon [12].

In this study, we characterize the clinical isolate *Paenibacillus alvei* strain 32. We present its genome sequence and use phylogenomics and Average Nucleotide Identity (ANI) to determine its taxonomical position. Based on these data, we propose to reclassify several strains previously called *P. alvei* into a distinct species-level clade, representing a novel genomospecies named "*Paenibacillus rickenmortis*" sp. nov. We also tested the antimicrobial activity of its culture supernatant against 40 target pathogens and used a Principal Component Analysis (PCA) to compare these phenotypic profiles with reference strains and antibiotics. Finally, we mined the genome to analyze its BGCs and identified a specific non-ribosomal peptide synthetase (NRPS) cluster as the highly probable candidate responsible for paenialvin biosynthesis, elucidating the potential genetic basis for this novel class of antibiotics previously described in *P. alvei* DSM 29. These findings establish *P. alvei* 32 as a versatile producer of antimicrobial compounds and highlight the *Paenibacillus* genus as a continuing source of high-value molecules for combating infectious diseases.

Materials and Methods

Growth Conditions and Genomic DNA Preparation

Paenibacillus alvei strain 32 was cultured in 3 mL of Brain Heart Infusion (BHI) broth for 24 h at 37°C. Genomic DNA (gDNA) was extracted using the PureLink Genomic DNA Mini Kit (Invitrogen, USA). Lysis was performed with Proteinase K at 55°C on cells collected from the overnight culture following the Gram-negative protocol. RNA was digested for 2 min by adding RNase A. The sample was bound to the purification column using the binding buffer and ethanol. Columns were washed twice, and the gDNA

was eluted in two steps with 25 µL of water after a 5-min incubation at 60°C. The quantity and quality of the recovered gDNA were evaluated by loading 1 µL of the sample onto a 0.8% agarose gel in TAE buffer, followed by electrophoresis for 10 min at 100 V. The gel was stained with ethidium bromide and visualized under UV light. A total volume of 30 µL with a concentration of 417 ng/µL of gDNA was obtained using a Quantus fluorometer (Promega, USA) and used for sequencing.

Antimicrobial Tests

P. alvei 32, *P. alvei* DSM 29, and *Paenibacillus* strain B-LR were cultured in nutrient broth for 24 h at 30°C and 160 rpm in three independent biological replicates. Cultures were centrifuged at 14,000 g for 20 min at 4°C. Supernatants were collected, sterilized by heating at 100°C for 10 min to ensure protein denaturation while preserving heat-stable lipopeptides, and stored at -80°C [2–11]. Target strains were grown in nutrient broth until they reached the stationary phase. A 5µL aliquot of the target strain culture was seeded into 15 mL of nutrient agar plates. Sterile glass wells (4 mm x 8 mm, Polylabo) were placed on the agar surface. Each well was filled with either 100 µL of culture supernatant, 100 µL of 0.005% (w/v) polymyxin E reference solution containing 5µg of active compound (Sigma) to match the diffusion volume of the supernatants, or 100 µL of uncultured medium adjusted to the pH of the supernatant as a negative control. Inhibition zones (halos) were measured after overnight incubation at 30°C or 37°C, depending on the optimal growth temperature of the target strain. Specific antimicrobial activities were calculated as the mean of three halo diameters (cm) divided by the biomass concentration (OD_{600nm}) of the cultures.

Statistical Analysis

To visualize and compare the susceptibility patterns among the 40 targeted strains, a Principal Component Analysis (PCA) was performed. The data matrix consisted of the mean inhibition zone diameters obtained from independent experimental replicates. To account for differences in scale and ensure homoscedasticity, the quantitative diameter data were standardized using Z-score transformation ((value - mean) / standard deviation) prior to analysis. The PCA was executed using XLSTAT 2024.3 based on a correlation approach to project the phenotypic inhibition profiles of the supernatants and reference antibiotics onto orthogonal axes.

Genome Sequencing, Assembly, and Annotation

A unique dual indexing library was prepared using the Illumina DNA Prep kit. Automated cluster generation was performed in a reagent cartridge according to the manufacturer's guidelines (Illumina). Sequencing was conducted on an Illumina NovaSeq 6000 S4 XP (2 times 150-bp) platform platform (Eurofins Genomics, Germany). Adapters and low-quality bases were removed using Trimmomatic (v0.33) [13], discarding reads with an average quality score per base below 15 or a minimum length under 36 bp. *De novo* genome assembly was performed using SPAdes (v3.11.0) [14, 15]. Assembly

quality was evaluated with QUAST (v5.0.2) [16]. The genome sequence was annotated using the Microbial Genome Annotation and Analysis Platform (MAGE/MicroScope) [17].

Comparative Genomics

Protein sequences were aligned using the MUSCLE program. Phylogenetic trees were constructed with the Maximum Likelihood method using MEGA-X with 150 bootstrap replicates, the JTT matrix model, and uniform rates. Average Nucleotide Identity (ANI) values were calculated using the EzBioCloud online platform. Digital DNA-DNA hybridizations were determined using the Genome-to-Genome Distance Calculator (GGDC 3.0) with Formula 2 [18]. Secondary metabolism biosynthetic gene clusters (BGCs) were identified using the antiSMASH 8.0 platform with relaxed parameters [19]. Reference sequences for genes and proteins were retrieved from the NCBI database. The domain structures and substrate specificities of the NRPS enzymes were analyzed using NRPSsp and PKS/NRPS analysis tools available on the same website. Gene cluster synteny and alignments were visualized using the clinker pipeline [20].

Nucleotide Sequence Accession Numbers

The draft genome sequence project has been deposited at GenBank/ENA/DDBJ under the accession number JAOQAR000000000. The raw sequencing reads are deposited at GenBank under the accession number SAMN31029884.

Results

Genome of *Paenibacillus* strain 32 overview

Paenibacillus strain 32 was previously isolated from an expectorate sample of a cystic fibrosis patient and showed antibiotic activity against Gram-negative and Gram-positive pathogens [2]. The genome of strain 32 was sequenced by the NovaSeq 6000 S4 XP (2 times 150-bp) platform. A total of 5,550,110 paired-end reads with an average coverage depth of 243X were obtained. A genome of 6,582,090 bp with 268 contigs, a GC content of 45.9%, and an N50 of 169,582 bp was obtained. No plasmid was found among the genomic data.

A total of 6,140 CDS were annotated as well as 3 rRNA and 53 tRNA (Table 1). Using the very close *Paenibacillus alvei* DSM 29 genome as a reference, we reduced the number of contigs to 239, which were manually assembled into 26 scaffolds. This corrected and annotated version of the assembly was deposited in the GenBank/ENA/DDBJ databases under the accession number JAOQAR000000000.

Reclassification within the *P. alvei* group

Initial phylogenetic analysis based on the 16S rDNA sequence placed strain 32 within the *Paenibacillus alvei* clade. However, a higher-resolution phylogeny constructed from eight concatenated housekeeping

gene sequences (*dnaA*, *rpoB*, *gyrB*, *pyrH*, *recA*, *pgi*, *trpB*, and *sucC*) revealed a clear bifurcation within the species. In this multilocus tree, strain 32 formed a distinct clade with the *P. alvei* reference strain DSM 29, MP1, and JR949, while remaining phylogenetically distant from other strains historically classified as *P. alvei*, including B-LR, TS-15, and A6-6i-x (Figure 1a).

This division was strongly supported by whole-genome comparisons. Average Nucleotide Identity (ANI) values were above the 95% species threshold within each group but dropped to less than 85% when comparing strains between the two groups (Figure 1b). This inter-group value is significantly below the established cutoff for species delineation. These findings were further corroborated by the Genome-to-Genome Distance Calculator (GGDC), which simulates DNA-DNA hybridization. The GGDC analysis yielded a 98% probability that strains 32 and DSM 29 belong to the same species, whereas the probability of strain 32 belonging to the same species as strain B-LR was only 0.06%. Taken together, these phylogenetic data indicate that the strains currently classified as *P. alvei* represent at least two distinct species-level lineages. As the group containing the reference type strain DSM 29 retains the original designation, these findings highlight that strains B-LR, TS-15, and A6-6i-x form a distinct genomic group. For this highly divergent environmental clade, we propose the designation of a novel genomospecies, designated as the "*Paenibacillus rickenmortis group*".

Antimicrobial Activity Spectra of *P. alvei* Supernatants

The antimicrobial activities of supernatants from *P. alvei* 32, DSM 29, and B-LR were tested against a panel of 40 target strains, including clinical drug-resistant isolates, and compared to pure polymyxin E (5 µg) (Table 2). The antimicrobial profile of the B-LR supernatant closely resembled that of polymyxin E, consistent with its genomic potential and previous reports on its polymyxin production [20]. The *P. alvei* 32 supernatant was active against several Gram-negative bacteria, including *P. aeruginosa*, *P. stutzeri*, and *Klebsiella pneumoniae*, an activity that might be due to polymyxin A1/A2, as the corresponding BGC was identified in its genome. Notably, the *P. alvei* 32 supernatant also demonstrated broad activity against all tested Gram-positive bacteria and fungi. This included strong inhibition of *Staphylococcus aureus* strains, with zones exceeding 12 mm even for methicillin-resistant *S. aureus* (MRSA) clinical isolates. This distinct anti-Gram-positive profile was observed to be highly similar to that of the reference strain, *P. alvei* DSM 29.

Statistical Analysis of Activity Profiles

To evaluate these observations, a Principal Component Analysis (PCA) was performed on the standardized inhibition zone data. The resulting correlation model was robust, explaining 85.07% of the total variability (Figure 2). The PCA segregated the antimicrobial profiles into two distinct groups along orthogonal axes. The vectors for the B-LR supernatant and polymyxin E were closely clustered and aligned with the first principal component (F1, 56.92%). In contrast, the vectors for the *P. alvei* 32 and DSM 29 supernatants were grouped together along the second principal component (F2, 28.15%). This analysis provides clear statistical evidence that *P. alvei* 32 and DSM 29 share a correlated antimicrobial spectrum that is functionally distinct from the polymyxin E profile.

Clustering of Target Strains by Susceptibility

Furthermore, the PCA enabled the clustering of the 40 target strains into four distinct groups based on their susceptibility profiles. Group A consisted of strains with intrinsic resistance to polymyxins, such as *Serratia* sp., *Proteus* sp., and *Providencia* sp., which showed limited or no sensitivity to any of the tested supernatants. Group B included strains primarily sensitive to the B-LR supernatant and polymyxin E. This group was largely composed of *Pseudomonas* strains and other typical targets of polymyxin antibiotics. Group C comprised strains preferentially sensitive to the supernatants of *P. alvei* 32 and DSM 29. This group included all tested *Staphylococcus* and *Bacillus* species, as well as *Enterococcus faecalis* and *Micrococcus luteus*, highlighting the unique anti-Gram-positive activity produced by these two strains. Finally, Group D consisted of strains whose susceptibility was not specific to any treatment.

Biosynthetic gene cluster analysis and comparison

Genome mining using the antiSMASH 8.0 bioinformatics tool with relaxed parameters revealed a highly conserved secondary metabolite blueprint across the three analyzed genomes, identifying a total of 24 distinct biosynthetic gene clusters (BGCs) across the dataset (Table 3) [19]. Individually, strain 32, reference strain DSM 29, and strain B-LR harboured 22, 23, and 24 BGCs, respectively**.** While the vast majority of these loci are shared among the three strains, most could not be affiliated with any known compound and therefore represent uncharacterized pathways. Notably, the Polymyxin and Resorcinol-Lanthipeptide gene clusters were present in all strains with high sequence identity (>95%), suggesting a conserved core secondary metabolome. The pellasoren cluster was also conserved but appeared incomplete in strain B-LR. Strain B-LR displayed a distinct genomic profile compared to DSM 29 and *P. alvei* 32. B-LR harbored a unique cluster architecture, although its presence on contig edges and subsequent fragmentation limits definitive structural resolution. It is worth noting that even among those conserved clusters, some sequence variability exists. This is exemplified by the polymyxin BGC, in which the gene product PmxA produces the Leu-Leu-Dab-Dab peptide in *P. alvei* B-LR, and Leu-Thr-Dab-Dab in *P. alvei* 32.

Putative *P. alvei* 32 specific clusters

Among the predicted clusters, nine differ in completeness or sequence identity from those present in the DSM 29 draft genome assembly. Among them, cluster 6.2 (3,537,111 - 3,648,681 nt) is a 110 kb NRPS cluster with five NRPS enzymes (PAENIB_v1_40105 to PAENIB_v1_40109) that are bioinformatically predicted to assemble: (i) (Leu/Ile)-(Ala/Gly/Orn)-(Leu/Ile), (ii (Phe/Val)-(Leu/Ile/Val-Lys/Arg/Orn)-(Leu/Ile), (iii) (Leu/Ile)-(Val), (iv) (Lys/Orn)-(Leu/Ile)-(Thr), (v) (Leu/Ile)-(Ser/Thr)-(Lys/Orn)-(Lys/Orn), where "/" represents alternative amino acid predictions (Figure 3-5). The presence of several uncharacterized adenylation domain substrate codes could indicate that the cluster 6.2 NRPS incorporates unusual monomers that are not fully implemented in databases. We observe the presence of several genes upstream that could be involved in the synthesis of atypical substrates. Homologs of the *pdhA/B/C/D* (PAENIB_v1_40093-40096) pyruvate dehydrogenase, a thymidylate synthase (PAENIB_v1_40098), and a glutamate synthase *gltB* (PAENIB_v1_40100) are located upstream of the

NRPS cluster (Figure 3). It is possible that these genes are involved in the supply of rare amino acid monomers, an organization previously observed in the polymyxin cluster where genes supplying diaminobutyric acid (Dab) are located upstream of the core NRPS. Several transport genes were also identified in this cluster, including (PAENIB_v1_40127) to (PAENIB_v1_40129) and (PAENIB_v1_40111), annotated as amino acid ABC transporters, alongside (PAENIB_v1_40117), an MFS transporter. The four genes located directly upstream of the core NRPS (PAENIB_v1_40101 to PAENIB_v1_40104) are respectively annotated as FkbH-like, acyl-carrier protein, FkbH-like, and FabG-like reductase, which represent typical fatty acid tailoring/PKS functions. The three genes located downstream of the NRPS (PAENIB_v1_40110 to PAENIB_v1_40112) are respectively annotated as a thioesterase, an ABC transporter, and a hypothetical protein.

To further investigate the evolutionary history and genomic architecture of this locus, a comparative synteny analysis was performed across multiple *Paenibacillus* genomes using Clinker (Figure 4). This analysis revealed that cluster 6.2 exhibits a high degree of structural conservation with the corresponding BGC found in the reference strain DSM 29, as well as in the isolates B-LR, MP1, 4300, and JR949. However, high-resolution sequence alignment revealed a localized structural modification within the core biosynthetic machinery. In *P. alvei* 32, a combination of two closely spaced out-of-frame indels occurs: a single-nucleotide insertion followed downstream by a 4-base pair deletion. This net change of 3 base pairs (-3 bp) maintains the overall downstream triplet reading frame, preserves the single continuous open reading frame architecture. At the protein level, this compensatory mutation blueprint introduces a transient frameshift, resulting in a variable window of approximately 20 alternative amino acids (positions 1466–1490) specific to strain 32 (as well as MP1, 4300, and JR949) before the consensus sequence restores (Supplementary Data). Apart from this specific event, the entire cluster shares high identity across the compared strains.

To characterize the biosynthetic potential of cluster 6.2, the substrate specificity of its 16 adenylation domains was bioinformatically predicted based on their binding pocket chemical signatures (Figure 5 - Table 4). The *in silico* analysis predicted a consensus 16-amino acid peptide backbone: (Leu/Ile-Ala-Leu/Ile-Leu/Ile-Leu/Ile-Lys/Arg-Leu/Ile-Leu/Ile-Val-Lys-Leu/Ile-Thr-Leu/Ile-Ser-Lys-Lys). Interestingly, a detailed alignment of the catalytic pockets revealed a specific sequence polymorphism among the Leucine/Isoleucine-selecting units. While the vast majority of these domains feature the conventional D215, W236, L239 signature code, the fourth NRPS module unique to this assembly line possesses an alternative D235, A136, L239 motif (Table 4). This distinct residue configuration at position 236 marks module 4 as a unique structural block within the otherwise uniform core biosynthetic machinery.

Discussion

The genome sequence and annotation of the clinical isolate *Paenibacillus* strain 32 confirms, via phylogenomics and Average Nucleotide Identity (ANI) calculations, its assignment to the *Paenibacillus alvei* species boundary alongside reference strains DSM 29, MP1, and JR949. Conversely, the closely related strains B-LR, TS-15, and A6-6i-x exhibit genomic distances well below the species threshold (<

85% ANI), supporting their reclassification into a distinct lineage designated as the novel genomospecies, "the *Paenibacillus rickenmortis* group".

Genome mining via antiSMASH identified a total of 24 distinct biosynthetic gene clusters (BGCs) across the dataset, with strain 32, DSM 29, and B-LR harboring 22, 23, and 24 loci, respectively. The data show that these strains share the near-totality of their secondary metabolome, with the exception of a single specific BGC exclusive to the environmental isolate B-LR. Among the shared pathways, only three could be confidently assigned to known compounds: polymyxin, paenibacterin, and resorcinol. In contrast to the polymyxin E (colistin) forms typically described in the *P. alvei* literature, the locus in strain 32 is predicted to synthesize a polymyxin A variant [7, 21].

Non-ribosomal peptide synthetase (NRPS) and hybrid NRPS/PKS clusters constitute the majority of the secondary metabolome in strain 32. Several small clusters (four or fewer modules) could theoretically code for short specialized peptides like cyclodipeptides; however, their small size may also reflect assembly fragmentation caused by the highly repetitive sequences of NRPS domains. In contrast, the larger, structurally intact core clusters, such as 2.5 (polymyxin A), 7.1 (paenibacterin), and the uncharacterized cluster 6.2, suggest evolutionary maintenance driven by a selective advantage. In the highly competitive, polymicrobial environment of the cystic fibrosis lung from which strain 32 was isolated, the co-production of these multiple functional metabolites may act concurrently against opportunistic pathogens like *Pseudomonas aeruginosa* and MRSA [2].

Principal Component Analysis (PCA) of the *in vitro* antimicrobial profiles segregates the isolates into distinct groups: the supernatant of the environmental strain B-LR displays a narrow anti-Gram-negative profile clustering with pure polymyxin E controls, whereas *P. alvei* 32 and DSM 29 exhibit a broad, potent anti-Gram-positive effect against targets including MRSA. This functional dichotomy reinforces the taxonomic separation of the *P. rickenmortis* clade and demonstrates that the presence of homologous BGCs does not guarantee metabolic uniformity, as variations in regulatory networks can fundamentally alter the active exometabolome.

Cross-referencing genomic annotations with the literature allows us to propose a chemical identity for the uncharacterized locus 6.2. Meng et al. (2018) characterized the paenialvins A-D, a family of linear lipopeptides isolated from *P. alvei* DSM 29 with potent anti-Gram-positive activity that matches the spectrum observed for strain 32. Crucially, public database screening reveals that cluster 6.2, occurs across the majority of available *P. alvei* genomes on the NCBI platform. Much like the ubiquitous polymyxin BGC, this widespread distribution indicates that cluster 6.2 constitutes a foundational core genomic component of the species rather than an accessory or transient metabolic trait.

The bioinformatic prediction of the 16-module peptide backbone from cluster 6.2 strictly correlates with the chemical composition of the paenialvins. The single predicted divergence at module 4 (Phe/Val instead of Leu/Ile) is also present in the DSM 29 reference genome, indicating a software alignment bias rather than a true biological variation. Furthermore, the Marfey's configuration analysis established by

Meng et al. (2018) mirrors our *in silico* mapping of epimerization (E) domains [11]. This strict modular collinearity and stereochemical congruence firmly associate cluster 6.2 with paenialvin biosynthesis.

Except for a localized 20-amino-acid window in module 4 (positions 1466–1490), the cluster remains identical across all tested strains. Because the B-LR genome was resolved using long-read PacBio sequencing [22], this sequence profile rules out short-read assembly artifacts. The *in vitro* activity of both DSM 29 and strain 32 confirms the functional viability of both protein variants. Therefore, the lack of production in B-LR points to transcriptional silencing or specific regulatory checkpoints unique to this environmental isolate, whereas the clinical strain 32 appears primed for expression under standard laboratory conditions.

We also hypothesize a biosynthetic link between cluster 6.2 and depsipeptide L, a shorter peptide previously isolated from *P. alvei* 32 [2]. Their mass spectrometry fragmentation spectra share identical fragments at *m/z* 273 and 386 (Leu/Ile-Ala and Leu/Ile-Ala-Leu/Ile motifs), confirming a shared N-terminal initiation path [2, 11]. Cluster 6.2 encodes a continuous assembly line of five NRPS enzymes. At module 6, the catalytic pocket (D235-V236-E239) predicts a Lysine substrate, but the structural similarity to Arginine (found in the paenialvins) points to substrate promiscuity. Depsipeptide L might therefore represent a truncated co-product resulting from premature chain termination at module 6. This truncation correlates with culture conditions: the full-length paenialvin complex is secreted in nutrient-rich complex media (TSB, BHI), whereas depsipeptide L was historically purified from a modified synthetic medium (MOD 50) enriched with free amino acids [2], conditions known to induce kinetic bottlenecks during non-ribosomal peptide elongation [24–25]. Finally, the upstream presence of the genes PAENIB_v1_40101 to PAENIB_v1_40104 (FkbH-like proteins, ACP, and a FabG-like reductase) might provides the necessary enzymatic machinery to synthesize the N-terminal GHPD lipid moiety common to both molecules [2].

In the absence of direct genetic knockouts or heterologous expression, this strict modular collinearity and three-dimensional stereochemical matchmaking provide a robust theoretical framework. Experimental manipulation of these mega-loci (> 110 kb) in *Paenibacillus* is hindered by significant technical bottlenecks, including genomic instability during CRISPR-based editing and frequent failures in heterologous expression due to codon bias or the lack of compatible PPTases. Consequently, *in silico* synteny and collinearity comparisons represent essential alternative pipelines to map and prioritize these complex candidate clusters before labor-intensive genetic engineering is pursued.

In conclusion, this study characterizes the clinical isolate *P. alvei* 32 as a multi-producer of complex antimicrobials and establishes cluster 6.2 as the central candidate locus for paenialvin biosynthesis. Identifying sequence plasticity in module 4 and potential substrate promiscuity in module 6 provides a genetic rationale for structural variations and truncated metabolic products like depsipeptide L. Concurrently, resolving these whole-genome metrics refines the taxonomic boundaries of the *P. alvei* complex, opening new avenues to study the regulatory mechanisms that govern the active secondary metabolome of these micro-organisms.

Declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Ethics approval

The article does not contain any studies with human participants or animals performed by any of the authors.

Author Contribution

Drago Haas (DH): Investigation, Methodology, Data curation.Fatoumata Tambadou (FT): Investigation, Methodology, Validation.Thibault Caradec (TC): Investigation.Valérie Sopéna (VS): Investigation, Formal analysis.Sandrine Didelot (SD): Writing – review & editing, Validation.Cyrille Barthélémy (CB): Writing – review & editing, Validation.Valérie Thiéry (VT): Writing – review & editing, Resources.Romain Chevrot (RC): Conceptualization, Supervision, Project administration, Writing – original draft, Writing – review & editing.

Acknowledgement

The LABGeM (CEA/Genoscope & CNRS UMR8030), the France Génomique and French Bioinformatics Institute national infrastructures (funded as part of Investissement d'Avenir program managed by Agence Nationale pour la Recherche, contracts ANR-10-INBS-09, ANR-11-INBS-0013 and ANR-21-ESRE-0048) are acknowledged for support within the MicroScope annotation platform

Data Availability

The data supporting the findings of this study are openly available in public repositories. The draft genome sequence project for Paenibacillus alvei strain 32 has been deposited at GenBank/ENA/DDBJ under the accession number JAOQAR000000000. The raw sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number SAMN31029884. All other relevant datasets analyzed during the current study are available from the corresponding author upon reasonable request

References

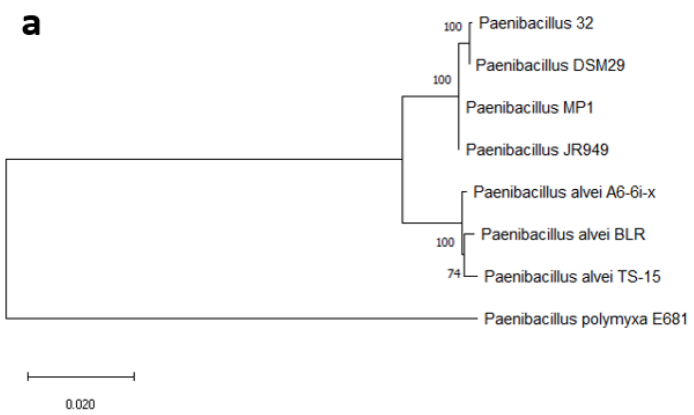
1. Hutchings MI, Truman AW, Wilkinson B (2019) Antibiotics: past, present and future. Curr Opin Microbiol 51:72–80. https://doi.org/10.1016/j.mib.2019.10.008
2. Chevrot R, Didelot S, Van den Bossche L, Tambadou F, Caradec T, Marchand P, Izquierdo E, Sopéna V, Caillon J, Barthélémy C, Van Schepdael A, Hoogmartens J, Rosenfeld E (2013) A Novel Depsipeptide Produced by *Paenibacillus alvei* 32 Isolated from a Cystic fibrosis Patient. Probiotics Antimicrob Proteins 5:18–25. https://doi.org/10.1007/s12602-012-9121-z
3. Ebeling J, Knispel H, Hertlein G, Fünfhaus A, Genersch E (2016) Biology of *Paenibacillus larvae*, a deadly pathogen of honey bee larvae. Appl Microbiol Biotechnol 100:7387–7395. https://doi.org/10.1007/s00253-016-7716-0
4. Ambrosini A, Sant'Anna FH, Heinzmann J, de Carvalho Fernandes G, Bach E, Passaglia LMP (2018) *Paenibacillus helianthi* sp. nov., a nitrogen fixing species isolated from the rhizosphere of *Helianthus annuus* L. Antonie Van Leeuwenhoek 111:2463–2471. https://doi.org/10.1007/s10482-018-1135-4
5. Bérdy J (2005) Bioactive microbial metabolites. J Antibiot 58:1–26. https://doi.org/10.1038/ja.2005.1
6. Cochrane SA, Vederas JC (2016) Lipopeptides from *Bacillus* and *Paenibacillus* spp.: A Gold Mine of Antibiotic Candidates. Med Res Rev 36:4–31. https://doi.org/10.1002/med.21321
7. Tambadou F, Caradec T, Gagez AL, Bonnet A, Sopéna V, Bridiau N, Thiéry V, Didelot S, Barthélémy C, Chevrot R (2015) Characterization of the colistin (polymyxin E1 and E2) biosynthetic gene cluster. Arch Microbiol 197:521–532. https://doi.org/10.1007/s00203-015-1084-5
8. Wilkinson S, Lowe LA (1964) Structures of polymyxin B2 and polymyxin E1. Nature 204:185. https://doi.org/10.1038/204185a0
9. Crass RL, Al Naimi T, Wen B, Souza E, Murray S, Pai MP, Jia S (2021) Pharmacokinetics of Polymyxin B in Hospitalized Adults with Cystic Fibrosis. Antimicrob Agents Chemother 65:e00792–e00721. https://doi.org/10.1128/aac.00792-21
10. Song L, Nuhamunada M, Weber T, Kovács AT (2025) Unlocking the biosynthetic potential of *Paenibacilli* through a genus-wide exploration of gene clusters for secondary metabolite production. bioRxiv. https://doi.org/10.1101/2025.01.22.634348
11. Meng J, Zhong Z, Qian PY (2018) Paenialvin A-D, four peptide antibiotics produced by *Paenibacillus alvei* DSM 29. J Antibiot 71:769–777. https://doi.org/10.1038/s41429-017-0001-3
12. Djukic M, Becker D, Poehlein A, Voget S, Daniel R (2012) Genome sequence of *Paenibacillus alvei* DSM 29, a secondary invader during European foulbrood outbreaks. J Bacteriol 194:6365. https://doi.org/10.1128/JB.01698-12
13. Pranav PS, Mahalakshmi B, Sivakumar R, Karthikeyan R, Rajendhran J (2021) Whole-Genome Sequence Analysis of *Paenibacillus alvei* JR949 Revealed Biosynthetic Gene Clusters Coding for Novel Antimicrobials. Curr Microbiol 78(4):1168–1176. https://doi.org/10.1007/s00284-021-02393-0
14. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

15. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021
16. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski A, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA (2013) Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 20:714–737. https://doi.org/10.1089/cmb.2013.0084
17. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A (2018) Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34:i142–i150. https://doi.org/10.1093/bioinformatics/bty266
18. Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, Rouy Z, Roche D, Salvignol G, Scarpelli C, Médigue C (2009) MicroScope: a platform for microbial genome annotation and comparative genomics. *Database* 2009:bap021. https://doi.org/10.1093/database/bap021
19. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M (2022) TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res 50:D801–D807. https://doi.org/10.1093/nar/gkab902
20. Blin K, Shaw S, Vader L, Szenei J, Reitz ZL, Augustijn HE, Cediel-Becerra JDD, de Crécy-Lagard V, Koetsier RA, Williams SE, Cruz-Morales P, Wongwas S, Segurado Luchsinger AE, Biermann F, Korenskaia A, Zdouc MM, Meijer D, Terlouw BR, van der Hooft JJJ, Ziemert N, Helfrich EJN, Masschelein J, Corre C, Chevrette MG, van Wezel GP, Medema MH, Weber T (2025) antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. Nucleic Acids Res 53:gkaf334. https://doi.org/10.1093/nar/gkaf334
21. Gilchrist CLM, Chooi YH (2021) clinker & clustermap.js: automatic generation of gene cluster comparison figures. Bioinformatics 37:2473–2475. https://doi.org/10.1093/bioinformatics/btab007
22. Tambadou F, Caradec T, Gagez AL, Bonnet A, Sopéna V, Bridiau N, Thiery V, Didelot S, Barthélémy C, Chevrot R (2015) Characterization of the colistin (polymyxin E1 and E2) biosynthetic gene cluster. Arch Microbiol 197:521–532. https://doi.org/10.1007/s00203-015-1084-5
23. Salsano C, Didelot S, Tambadou F, Caradec T, Sopéna V, Barthélémy C, Thiéry V, Chevrot R (2018) Complete Genome Sequence of the Polymyxin E (Colistin)-Producing *Paenibacillus* sp. Strain B-LR. Microbiol Resour Announc 7:e01013–e01018. https://doi.org/10.1128/MRA.01013-18
24. Huang E, Guo Y, Yousef AE (2014) Biosynthesis of the new broad-spectrum lipopeptide antibiotic paenibacterin in *Paenibacillus thiaminolyticus* OSY-SE. Res Microbiol 165:243–251. https://doi.org/10.1016/j.resmic.2014.02.002
25. Ginitanz S, Marahiel MA (2012) Modular enzymatic assembly lines: their structural organization and catalytic mechanisms. ChemBioChem 13:2163–2174. https://doi.org/10.1002/cbic.201200382
26. Strieker M, Tanović A, Marahiel MA (2010) Nonribosomal peptide synthetases: structures and dynamics. Curr Opin Struct Biol 20:234–240. https://doi.org/10.1016/j.sbi.2010.01.009

Tables

Tables 1 to 4 are available in the Supplementary Files section.

Figures



b

	BLR	A6-6i-x	TS-15	32	DSM29	JR949
BLR						
A6-6i-x	**95,58%**					
TS-15	**96,05%**	**95,46%**				
32	84,88%	84,61%	84,46%			
DSM29	85,14%	84,75%	84,83%	**99,94%**		
JR949	84,80%	84,64%	84,74%	**97,81%**	**97,70%**	
MP1	84,88%	84,60%	84,81%	**97,75%**	**97,69%**	**99,01%**

Figure 1

Phylogenomics and genomic boundaries of the *Paenibacillus alvei* group.

a) Multilocus phylogenetic tree of *Paenibacillus alvei* strains constructed from concatenated housekeeping gene sequences (*dnaA*, *rpoB*, *gyrB*, *pyrH*, *recA*, *pgi*, *trpB*, and *sucC*) using the Maximum Likelihood method. **b)** Heatmap matrix of Average Nucleotide Identity (ANI) percentages among compared *P. alvei* isolates. *Paenibacillus polymyxa* strain E681 is included as the phylogenetic outgroup.

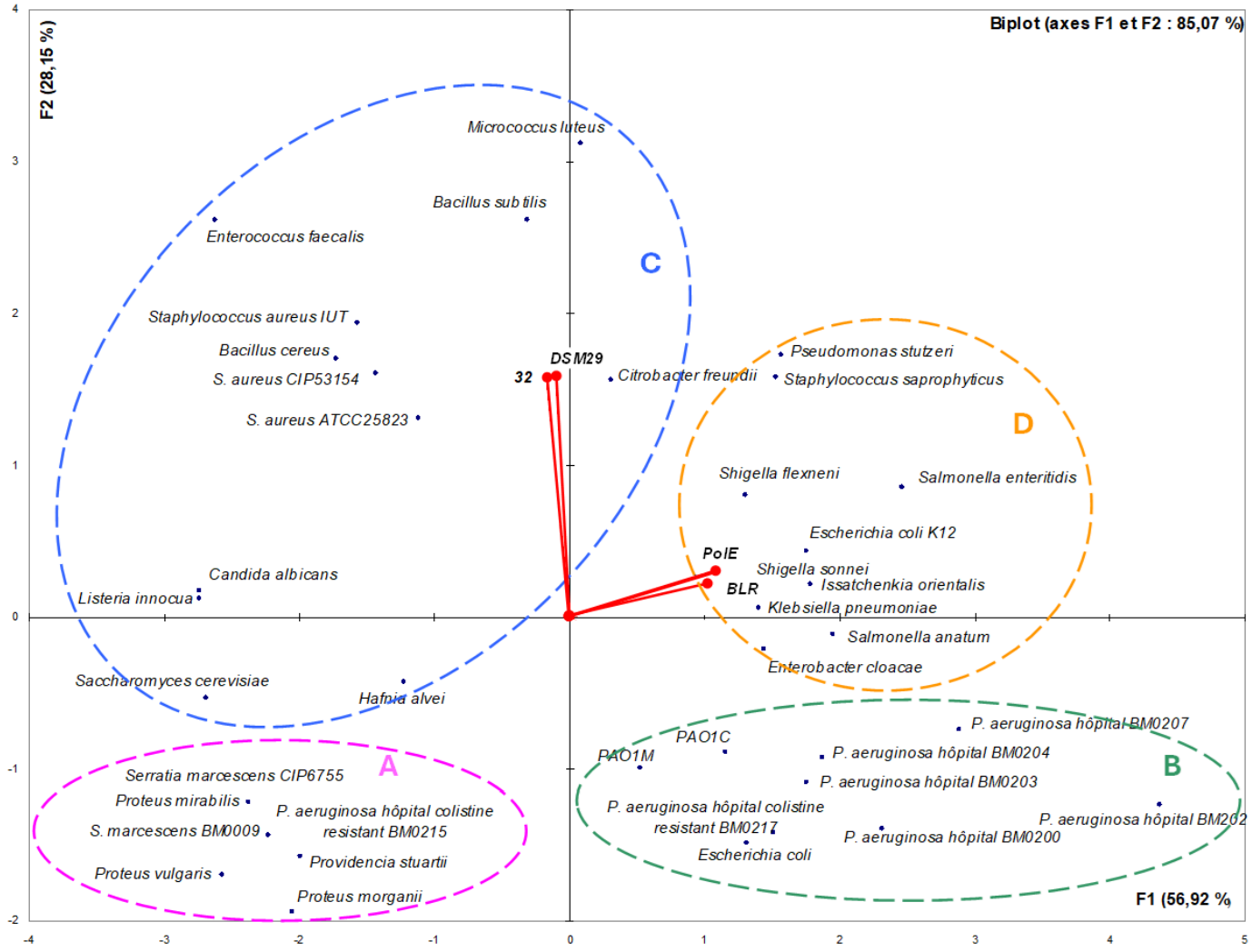


Figure 2

Principal Component Analysis (PCA) biplot of functional bioactivity spectra.

Two-dimensional biplot projection illustrating the mathematical superposition of the variable correlation circles (antimicrobial profiles of strain 32, DSM 29, and B-LR supernatants, alongside purified polymyxin E active controls) onto the individual susceptibility coordinate map of the 40 targeted indicator pathogens. Ellipses denote the distinct phenotypic clusters (Groups A–D) resolved across the primary F1 (56.92%) and orthogonal F2 (28.15%) components, explaining 85.07% of the cumulative dataset variance

.

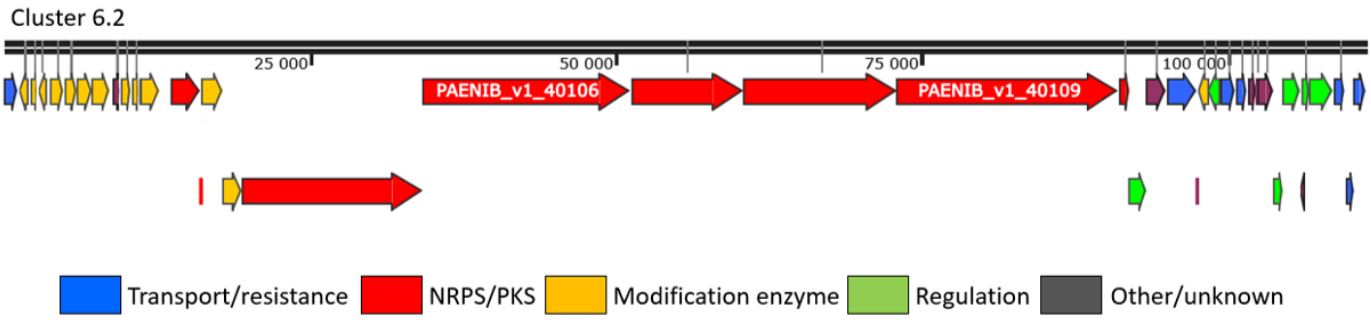


Figure 3

Representation of clusters 6.2 of *P. alvei* 32, identified by antiSMASH.

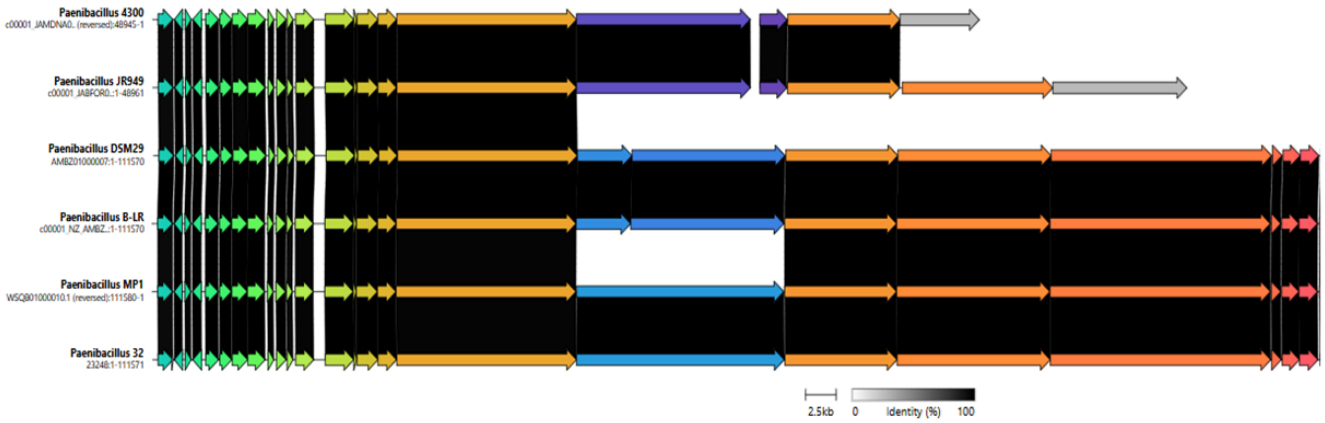


Figure 4

Multi-genome synteny tracks and structural architecture of the paenialvin candidate locus.

High-resolution chromosomal synteny alignment map of candidate cluster 6.2 across *Paenibacillus alvei* strains (DSM 29, 32) and relevant references (strains 4300, JR949, MP1, and B-LR) resolved via Clinker. Arrows define open reading frames (ORFs) with matching colors displaying orthologous core orientations. Connective grayscale shaded lanes link regions of nucleotide similarity based on continuous sequence identity values ($0\%\text{--}100\%$). Region truncations resolved at the boundaries of assemblies 4300 and JR949 map to short-read contig formatting drops rather than directional gene deletion.

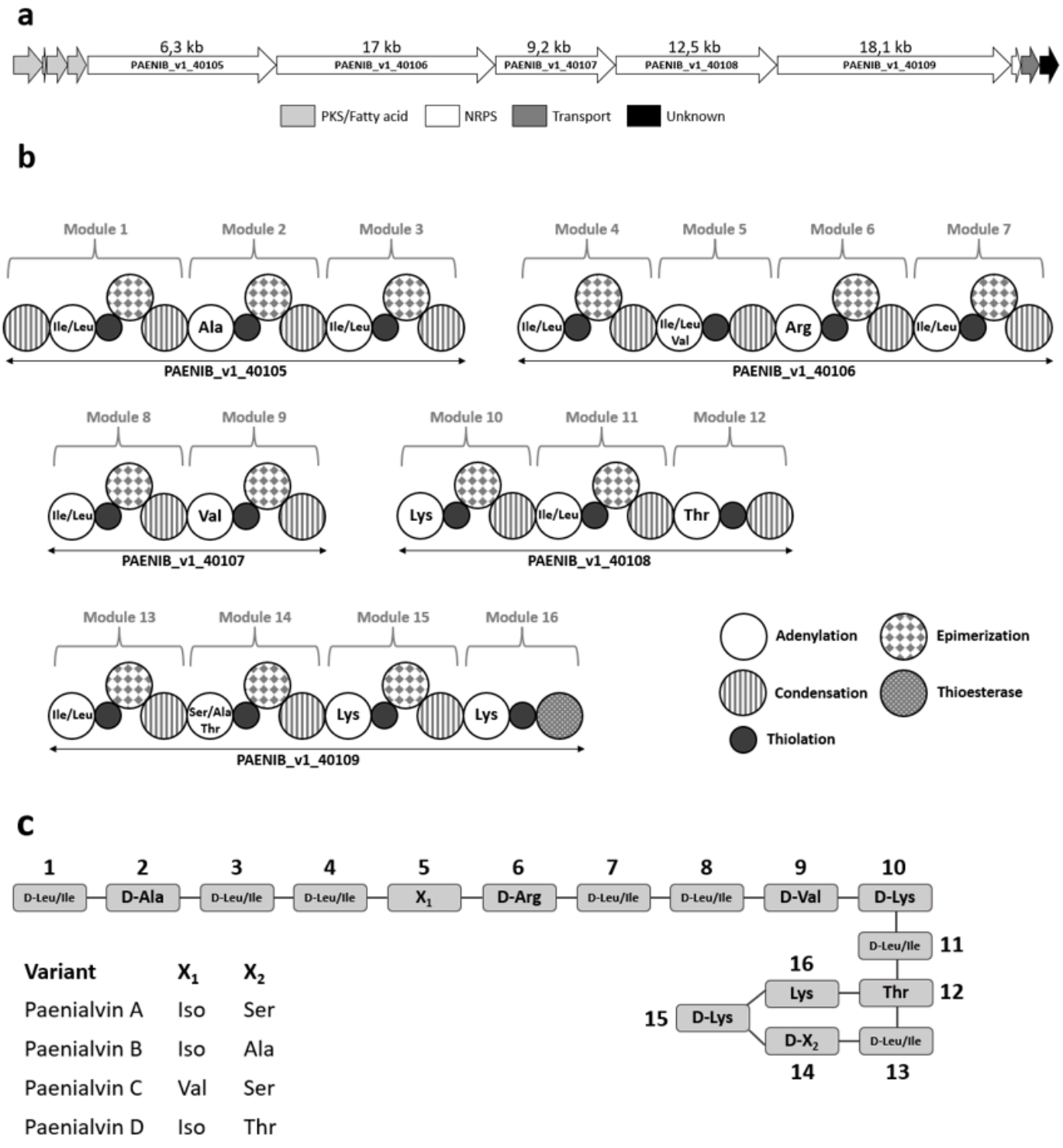


Figure 5

Computational mapping and biosynthetic logic of cluster 6.2.

a) Schematic physical map and functional block annotation of the 110-kb cluster 6.2 locus highlighting PKS/lipid tailoring lines (upstream), the core non-ribosomal peptide megasynthetase modules, and co-localized amino acid transporters. b) Linear modular schematic displaying domain arrangements

(Adenylation, Condensation, Thiolation, Epimerization, and Thioesterase modules) of the five functional multi-enzyme proteins (PAENIB_v1_40105 to PAENIB_v1_40109). c) Correlated theoretical structures of the target paenialvin peptide variants A–D highlighting modular congruence.

PAENIB_v1_40105

Adenylation domains

Key amino acid of binding pocket

Putative substrate

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx