

Drug Discovery & Functional Genomics | Signal Rating: 1/5 Stars | Oct 28, 2026 | 9 min read

Cryptic Biosynthetic Gene Cluster & AMP Miner Baseline: MIBiG Dereplication and Embedded Self-Resistance GyrB Duplication De-Orphan a Marine Streptomyces NRPS-PKS & sORF Peptide Pair

Running the CRV Science Labs Rust bio-suite bgc-mine pipeline filters known polyketide rediscoveries while pairing an orphan 46.5 kb hybrid NRPS-T1PKS cluster with its embedded Asn426/His136 self-resistant gyrB target paralog and profiling a 31-aa amphipathic antimicrobial peptide.

1. Abstract & Target-First Natural Product Mining Strategy

Marine actinomycetes, soil myxobacteria, and filamentous fungi encode thousands of cryptic biosynthetic gene clusters (BGCs) -- including Non-Ribosomal Peptide Synthetases (NRPS), Polyketide Synthases (PKS), and ribosomally synthesized and post-translationally modified peptides (RiPPs) -- that remain transcriptionally silent in standard laboratory monocultures [1]. Historically, natural product discovery suffered from two major bottlenecks: (i) high rediscovery rates of known scaffolds (such as streptothricin, erythromycin, or actinomycin), and (ii) unknown molecular mechanisms of action (MoA) for newly isolated secondary metabolites.

Pipeline 4 (bio-suite bgc-mine) resolves both bottlenecks computationally before wet-lab heterologous expression. First, it combines short-ORF calling (Prodigal-gv tuned for sORFs < 100 amino acids) with rule-based and CRF/Transformer BGC detection (antiSMASH 7 [1] and GECCO [2]), filtering out known clusters that share $\geq 80\%$ domain-architecture Jaccard similarity against the MIBiG v3.1 reference database via BiG-SCAPE [3]. Second, it applies the Antibiotic Resistant Target Seeker (ARTS) evolutionary principle [4]: an organism synthesizing a lethal toxin must encode a mutated, resistant paralog of its own essential housekeeping target gene inside the BGC to prevent suicide. Third, for short RiPP/AMP candidates, it computes the 11-residue helical-wheel Eisenberg amphipathic hydrophobic moment (u_H) [5], net charge at pH 7.4, and Boman protein-binding index alongside Macrel [6] and ESMFold structural folding.

2. MIBiG Dereplication & Prioritized Natural Product Dossiers

Across the marine sediment metagenome-assembled genome benchmark (Streptomyces sp. SCSIO-64), Step 1 detected three candidate clusters. Step 2 automatically dereplicated and discarded Cluster-01 after computing a 92.4% domain Jaccard similarity to known Erythromycin A (BGC0000055), while retaining two orphan scaffolds that passed the < 0.80 novelty gate: BGC-NRPS-SCSIO64-02 (Jaccard = 0.342) and BGC-RIPP-SCSIO64-03 (Jaccard = 0.185). Click Copy for Excel (TSV) to export the BGC dereplication and target-mining summary:

| Cluster ID || Contig Region || Length (kb) || BGC Structural Class || Nearest MIBiG Hit || MIBiG Jaccard Sim || Novelty Gate (<0.80) || Embedded Target / Hook || MedChem Priority Score |

| BGC-T1PKS-SCSIO64-01 || NODE_1:24,100..78,500 || 54.40 || Type-I Modular Polyketide (T1PKS) || BGC0000055 (Erythromycin A) || 0.9240 || FILTERED (Known) |
| ermE 23S rRNA methyltransferase || 14.2 |

| BGC-NRPS-SCSIO64-02 || NODE_4:118,400..164,920 || 46.52 || Hybrid NRPS / Type-I PKS || BGC0001555 (Cinnabaramide A) || 0.3420 || PASSED (Novel) || gyrB2 (Asp426Asn / Arg136His) || 94.6 |

| BGC-RIPP-SCSIO64-03 || NODE_11:42,100..51,850 || 9.75 || Class II sORF RiPP / Cationic AMP || BGC0000581 (Microcin J25) || 0.1850 || PASSED (Novel) ||
SCSIO64_sORF_1942 (31-aa AMP) || 88.4 |

3. Self-Resistance Target De-Orphaning (GyrB2) & sORF AMP Biophysics

Within BGC-NRPS-SCSIO64-02 (NODE_4:118,400..164,920 , encoding core modules A(Pip)-PCP-C , A(dhb)-PCP-E , KS-AT(mmal)-KR-ACP , and C-A(Orn)-PCP-TE), the ARTS scanner identified a duplicated copy of the essential housekeeping gene gyrB (DNA Gyrase Subunit B, TIGR01059). Whereas the canonical chromosomal housekeeping copy resides at SCSIO64_00142 , the BGC-embedded paralog SCSIO64_02189 exhibits 19.8% amino-acid divergence and carries two classic ATP-binding pocket resistance substitutions -- Asp426Asn and Arg136His -- directly predicting that the hybrid depsipeptide-polyketide product acts as a competitive GyrB ATPase inhibitor.

In parallel, biophysical profiling of the 31-aa cryptic peptide SCSIO64_sORF_1942 inside BGC-RIPP-SCSIO64-03 confirms an amphipathic cationic alpha-helical hairpin architecture capable of Gram-negative outer-membrane disruption:

| sORF ID || Peptide Sequence (1-Letter IUPAC) || Length (aa) || Net Charge (pH 7.4) || Eisenberg Moment μ_H || Boman Index (kcal/mol) || ESMFold pLDDT ||
Membrane Disruption Prob |

| SCSIO64_sORF_1942 || GIGKFLHSAKKFGKAFVGEIMNSKWRKVGRK || 31 || 8.1 || 0.6862 |
| 0.6194 || 89.4 || 0.8511 |

4. Medicinal Chemistry Outlook & 1* Signal Rating Assessment

Pairing domain-architecture dereplication with embedded self-resistance target mining turns genome mining from a blind isolation exercise into a mechanism-guided drug discovery workflow. A medicinal chemistry team prioritizing BGC-NRPS-SCSIO64-02 knows prior to fermentation or total synthesis that the encoded metabolite targets the bacterial GyrB ATPase lid.

Signal Rating Assessment (1* Baseline / Review Note): This report is rated 1* because it evaluates the deterministic reference benchmark dataset included with bio-suite bgc-mine (n = 3 clusters in a single actinomycete reference assembly) without wet-lab LC-MS/MS metabolomic confirmation, heterologous expression in *Streptomyces coelicolor* , or minimum inhibitory concentration (MIC) bioassays.

References & Primary Sources

* [1] Blin, K., Shaw, S., Augustijn, H. E., et al. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research*, 51(W1), W46-W50.
<https://doi.org/10.1093/nar/gkad344>

- * [2] Carroll, L. M., Larralde, M., Fleck, J. S., Ponnudurai, R., Milanese, A., Cappio, E., & Zeller, G. (2021). Accurate de novo identification of biosynthetic gene clusters with GECCO. *bioRxiv / EMBL*. <https://doi.org/10.1101/2021.05.03.442509>
- * [3] Terlouw, B. R., Blin, K., Navarro-Muoz, J. C., et al. (2023). MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research*, 51(D1), D603-D610. <https://doi.org/10.1093/nar/gkac1049>
- * [4] Mungan, M. D., Alanjary, M., Blin, K., Weber, T., Medema, M. H., & Ziemert, N. (2020). ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. *Nucleic Acids Research*, 48(W1), W546-W552. <https://doi.org/10.1093/nar/gkaa374>
- * [5] Eisenberg, D., Weiss, R. M., & Terwilliger, T. C. (1982). The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature*, 299(5881), 371-374. <https://doi.org/10.1038/299371a0>
- * [6] Santos-Jnior, C. D., Pan, S., Zhao, X.-M., & Coelho, L. P. (2020). Macrel: antimicrobial peptide screening in genomes and metagenomes. *PeerJ*, 8, e10555. <https://doi.org/10.7717/peerj.10555>