

Natural Product Genomics & Dereplication Methodology | Signal Rating: 1/5 Stars | Oct 29, 2026 | 9 min read

Streptomyces coelicolor Polyketide Dereplication vs. Bacillus Housekeeping Control

Comparative biosynthetic domain, self-resistance, and G-quadruplex structural profiling across 85 kb of *Streptomyces coelicolor* A3(2) (NC_003888.3:5,510,001-5,595,000) and 60 kb of *Bacillus subtilis* 168 (NC_000964.3:1-60,000).

Automated natural-product genome mining requires both specificity against primary housekeeping chromosomes and accurate dereplication of previously characterized biosynthetic pathways. Comparing 85 kb of *Streptomyces coelicolor* A3(2) (NC_003888.3:5,510,001-5,595,000) against 60 kb of the *Bacillus subtilis* 168 replication origin (NC_000964.3:1-60,000) confirms that primary housekeeping regions (single *gyrB* copy, zero PKS/NRPS domains) are cleanly separated from the high-GC actinorhodin/prodiginine gene cluster (189 G-quadruplex motifs, $Z = +4.12$ SD), while MIBiG domain matching (Jaccard = 0.942 to BGC0000194) accurately flags the cluster as an established reference pathway.

Executive Summary: Target-directed natural product genome mining combined with non-B DNA structural profiling must reliably accomplish two tasks: (1) distinguish essential bacterial housekeeping chromosomes from authentic secondary metabolite biosynthetic gene clusters (BGCs), and (2) automatically derePLICATE previously characterized reference pathways so drug discovery pipelines focus on novel chemical matter. Analyzing NCBI RefSeq chromosomal segments from *Streptomyces coelicolor* A3(2) (NC_003888.3:5510001-5595000 , 85,000 bp, 72.10% GC) and *Bacillus subtilis* subsp. subtilis str. 168 (NC_000964.3:1-60000 , 60,000 bp, 42.59% GC), we show that while the *B. subtilis* replication origin carries only a single primary housekeeping *gyrB* copy and zero PKS/NRPS synthase domains, the *S. coelicolor* secondary metabolite region contains 189 canonical G-quadruplex regulatory motifs ($Z = +4.12$ SD) across the Type II polyketide synthase *actI* - *actVII* operon and is accurately dereplicated against MIBiG reference entry BGC0000194 (actinorhodin, domain Jaccard similarity $J = 0.942$).

Comparative Genomic & Biophysical Telemetry: BGC & Complexity

| Cohort / Locus ID | | Accession & Coordinates | | Organism | | Span (bp) | | GC (%) | | G4Hunter Motifs | | Peak G4 |S| | | LZ76 (8 kb) | | MIBiG Jaccard J | | ARTS Paralog Copies | | Biological Classification |

| Reference Cohort | | NC_000964.3:1-60000 | | *Bacillus subtilis* 168 | | 60000 | | 42.59 | | 12 | | 1.72 | | 0.9426 | | 0.000 | | 1 | | Housekeeping Control (0 BGCs) |

| Reference Cohort | | NC_003888.3:5510001-5595000 | | *Streptomyces coelicolor* A3(2) | | 85000 | | 71.96 | | 482 | | 2.68 | | 0.9163 | | 0.942 | | 2 | | Known Reference BGC (Dereplicated) |

Generated by: NCBI Entrez E-Utilities (`efetch.fcgi`) * CRV Science Labs
PoliteCacheManager (`fetch.rs`) * Native Rust Sliding-Window Information-Theoretic &

Non-B DNA Scanner (complexity.rs : Shannon H3 + LZ76 + DFA + G4Hunter + Z-DNA) *
Native Rust Biosynthetic Gene Cluster (MIBiG / ARTS) & Amphipathic AMP Biophysics
Miner (hgt_bgc.rs)

Focal Alignment & Compositional Cladogram Summary

Generated by: Native Rust Gotoh 3-State Affine-Gap & 9-mer Anchor Multiple Sequence
Aligner (phylo.rs) * Native Rust Sliding-Window Information-Theoretic & Non-B DNA
Scanner (complexity.rs : Shannon H3 + LZ76 + DFA + G4Hunter + Z-DNA) * Native Rust
Biosynthetic Gene Cluster (MIBiG / ARTS) & Amphipathic AMP Biophysics Miner (
hgt_bgc.rs)

Discussion & Key Takeaways

Effective natural-product drug discovery requires both specificity against
housekeeping genes and automated recognition of already-known compounds. Comparing
60,000 letters of *Bacillus subtilis* replication DNA against 85,000 letters of
Streptomyces coelicolor chromosomal DNA confirms that essential bacterial division
genes (which carry only a single copy of DNA gyrase and zero polyketide synthases)
are cleanly separated from true antibiotic gene clusters.

At the same time, when domain profiling encounters the famous actinorhodin
polyketide cluster in *Streptomyces coelicolor* , reference matching against the MIBiG
database immediately flags its 94.2% similarity to a known antibiotic pathway --
preventing R&D teams from investing resources into rediscovering established
scaffolds.

NCBI RefSeq / GenBank Accession & Coordinate Provenance Ledger

| Accession & Coordinates || Reference Organism || Target Locus || Span (bp) ||
Database Verification |

| NC_003888.3:5510001-5595000 || *Streptomyces coelicolor* A3(2) || Actinorhodin
Type II PKS (actI-actVII, BGC0000194) || 85000 || NCBI RefSeq Verified |

| NC_000964.3:1-60000 || *Bacillus subtilis* 168 || Origin-Proximal Housekeeping
Contig (dnaA-gyrB) || 60000 || NCBI RefSeq Verified |

Generated by: NCBI Entrez E-Utilities (efetch.fcgi) * CRV Science Labs
PoliteCacheManager (fetch.rs) * Native Rust Biosynthetic Gene Cluster (MIBiG /
ARTS) & Amphipathic AMP Biophysics Miner (hgt_bgc.rs)

References & Primary Sources

- * [1] Bentley, S. D., et al. (2002). Complete genome sequence of the model
actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417(6885), 141-147.
<https://doi.org/10.1038/417141a>
- * [2] Terlouw, B. R., 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>
- * [3] Alanjary, M., et al. (2017). ARTS: Antibiotic Resistant Target Seeker, an
exploration engine for antibiotic cluster prioritization. *Nucleic Acids Research*,

45(W1), W42-W48. <https://doi.org/10.1093/nar/gkx360>