HHS Public Access

Author manuscript

J Nat Prod. Author manuscript; available in PMC 2019 September 28.

Published in final edited form as:

J Nat Prod. 2018 September 28; 81(9): 2057–2068. doi:10.1021/acs.jnatprod.8b00394.

Complete Genome of *Micromonospora* sp. Strain B006 Reveals Biosynthetic Potential of a Lake Michigan Actinomycete

Jana Braesel[†], **Camila M. Crnkovic**[†,‡], **Kevin J. Kunstman**[§], **Stefan J. Green**[§], **Mark Maienschein-Cline**[⊥], **Jimmy Orjala**[†], **Brian T. Murphy**[†], and **Alessandra S. Eustáquio**[†,*]

[†]Department of Medicinal Chemistry and Pharmacognosy and Center for Biomolecular Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60607, USA

[‡]CAPES Foundation, Ministry of Education of Brazil, Brasília, Federal District 70040-020, Brazil

[§]DNA Services Facility, University of Illinois at Chicago, Chicago, IL 60612, USA

[⊥]Core for Research Informatics, University of Illinois at Chicago, Chicago, IL 60615, USA

Abstract

Actinomycete bacteria isolated from freshwater environments are an unexplored source of natural products. Here we report the complete genome of the Great Lakes-derived *Micromonospora* sp. strain B006, revealing its potential for natural product biosynthesis. The 7-megabase pair chromosome of strain B006 was sequenced using Illumina and Oxford Nanopore technologies followed by Sanger sequencing to close remaining gaps. All identified biosynthetic gene clusters (BGCs) were manually curated. Five known BGCs were identified encoding desferrioxamine, alkyl-*O*−dihydrogeranyl-methoxyhydroquinone, a spore pigment, sioxanthin, and diazepinomicin, which is currently in phase II clinical trials to treat Phelan-McDermid syndrome and co-morbid epilepsy. We report here that strain B006 is indeed a producer of diazepinomicin, and at yields higher than previously reported. Moreover, eleven of the sixteen identified BGCs are orphan, eight of which were transcriptionally active under the culture condition tested. Orphan BGCs include an enediyne polyketide synthase and an uncharacteristically large, 36-module polyketide synthasenonribosomal peptide synthetase BGC. We developed a genetics system for *Micromonospora* sp. B006 that shall contribute to deorphaning BGCs in the future. This study is one of the few attempts to report the biosynthetic capacity of a freshwater-derived actinomycete and highlights this resource as a potential reservoir for new natural products.

GRAPHICAL ABSTRACT

[*]**Corresponding Author:** Tel: +1-3124137082. ase@uic.edu (Alessandra S. Eustáquio).

Author Contributions

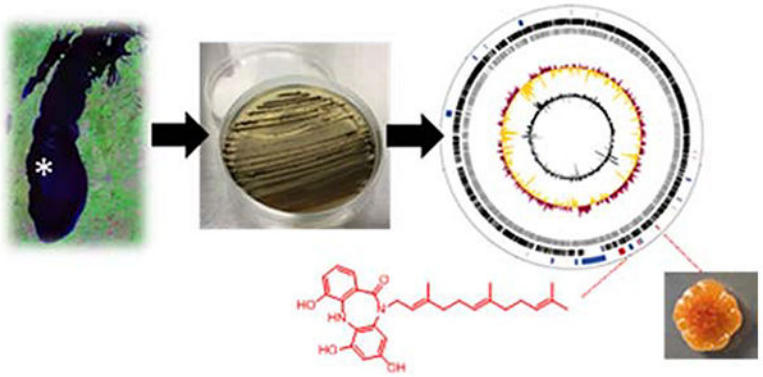
JB performed genome analysis, annotation of biosynthetic gene clusters, cultivation, HPLC analysis, transcript analysis and gene inactivation. CMC isolated diazepinomicin and performed MS and NMR analyses. KJK sequenced the genome of strain B006. MMC performed genome assembly. ASE and BTM conceptualized the project. ASE, SG, and JO oversaw different aspects of the project. ASE and JB wrote the paper with comments from all authors. All authors have given approval to the final version of the manuscript.

Supporting Information.

The following files are available free of charge.

Tables S1 to S5. Three BGCs identified only by PRISM (S1), predicted functions of ORFs in clusters 13 and 10 (S2 and S3), chemicals (S4) and oligonucleotide primers used during this study (S5). (PDF)

Figures S1 to S4. Multi-locus species tree (S1), known BGCs in *Micromonospora* sp. B006 (S2), phylogenetic analysis of enediyne PKSs (S3) and phylogenetic analysis of RifK homologs (S4). (PDF)



Environmental bacteria remain an invaluable resource for the discovery of natural product-derived drugs.[1] Actinomycete bacteria in particular have been a prolific source of structurally and functionally diverse natural products. Yet, only a handful of specialized metabolites from freshwater-derived actinomycetes have been reported in peer-reviewed literature to date, including 16-methyl-6(*Z*)-heptadecenoic acid and 16-methyl-8(*Z*)-heptadecenoic acid from *Micrococcus* sp., resistomycin from *Streptomyces* sp. B033, octalactin B from *Streptomyces* sp. B025 and the diazaquinomycins H and J from *Micromonospora* sp. strain B026 and B006.

Micromonospora spp. are widely distributed in diverse environments including soil, marine sediments, rocks, nitrogen fixing nodules from leguminous and actinorhizal plants, and freshwater sediments. Although more than 700 antibiotics have been isolated from *Micromonospora* species, only a limited number of studies have focused on the analysis of biosynthetic gene clusters (BGCs) from strains of this genus.[2–5]

The prospect of mining microbial genomes for new compounds has re-invigorated the field of natural products research. However, translating bioinformatics predictions from BGC sequence data into isolated and structurally characterized natural products remains a major bottleneck of genome mining. Some of the major challenges include 1) incomplete knowledge of natural product biosynthesis, on which bioinformatics predictions rely; 2) lack of metabolite production under a given culture condition; and 3) production of metabolites in low titers that hamper compound isolation and characterization. Thus, connecting natural products to BGCs is crucial not only to understand natural product biosynthesis, but also to enable future, productive genome mining approaches. Genetic engineering approaches can help connect BGCs to their natural products. Having a genetics system allows the generation of knock-out mutants which can then aid in natural product identification through comparative metabolite analysis of mutant and wild-type strains. In addition, genetic engineering of the native producer can contribute to increasing transcription of BGCs by for example promoter exchange, which can then lead to higher expression and increased yields.

Here we report the complete genome of *Micromonospora* sp. strain B006, which was isolated from Lake Michigan sediment. All identified BGCs were manually curated, including five BGCs encoding known compounds. Amongst the known compounds, diazepinomicin was of particular interest to us. Diazepinomicin (alternative names:

ECO-4601, TLN-4601, BU-4664L, and AMO-01) was isolated in 1996 from *Micromonospora* sp. M990–6, and twice more in 2004 from *Micromonospora* sp. strain DPJ12 and strain 046Eco-11, respectively. Due to its inhibition of the Ras-MAPK (mitogen-activated protein kinase) signaling pathway and ability to cross the blood-brain barrier and bind benzodiazepine receptors, diazepinomicin was evaluated in a phase II clinical trial to treat glioblastoma. Although the drug was generally well tolerated, the study was terminated due to lack of efficacy. However, a new phase II clinical trial, this time to treat Phelan-McDermid syndrome and co-morbid epilepsy, was recently initiated and is expected to be completed in 2019 (clinicaltrials.gov entry NCT03493607).

Moreover, eleven orphan BGCs were identified in the genome of strain B006, putatively encoding nonribosomal peptides, type I polyketides, hybrid nonribosomal peptide-polyketides, ribosomally synthesized and post-translationally modified peptides (RiPPs), and an enediyne. We performed transcript analysis to determine the proportion of orphan BGCs that are transcriptionally active under our culture conditions. Finally, we established a genetics system for this strain which will aid in assigning function to orphan BGCs in the future.

RESULTS AND DISCUSSION

Complete genome sequenced and assembled into one circular chromosome.

The genome of the freshwater-derived *Micromonospora* sp. strain B006 was initially sequenced with short-read Illumina sequencing and assembled with the software package Spades, resulting in an assembly consisting of 191 contigs. Initial analysis with the antibiotics & secondary metabolite analysis shell (antiSMASH) revealed that the predicted BGCs, especially nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) operons, were incomplete and located on different contigs. NRPSs and PKSs possess a highly-conserved modular architecture, which is reflected by highly repetitive sequences at the nucleotide level. These repeats are often longer than the 150-base pair (bp) read length obtained by Illumina sequencing, making it very difficult to correctly assemble the PKS and NRPS gene clusters.

Long-read sequencing technologies are being used to overcome the short insert size of Illumina libraries. Data generated with the most established long-read technology, Pacific Biosciences (PacBio), has been used in combination with short-read Illumina to assemble actinobacterial genomes into single contigs. The most recent long-read sequencing technology – Oxford Nanopore – is a compelling alternative to PacBio because of the low price of the MinION instrument, its portability, lower DNA input requirements, and potentially much longer sequence reads. Since its distribution starting in 2014, Oxford Nanopore is still in rapid development stage and progress is continuously being made regarding throughput and error rate. Giordano *et al.* recently compared Illumina MiSeq, PacBio and Oxford Nanopore platforms to sequence yeast genomes. The authors reported that *de novo* assemblies using long-read technologies alone could reach accuracies of 99% for PacBio and 98% for Oxford Nanopore, whereas accuracy of >99.9% can be achieved by hybrid assembly of long reads from either PacBio or Oxford Nanopore in conjunction with Illumina reads.

Therefore, we decided to re-sequence *Micromonospora* sp. B006 using Oxford Nanopore. Although an assembly using only nanopore data was not successful, likely due to the low coverage obtained, a joint assembly of Illumina and Nanopore reads using Spades resulted in seven contigs, representing a great improvement over the previous Illumina-only assembly (Table 1).

A comparison of the obtained seven contigs to a reference genome, *Micromonospora* sp. L5 (NCBI RefSeq: NC_014815), which was originally isolated from *Casuarina equisetifolia* root nodules, made it possible to predict the order of the seven contigs. Subsequently, the gaps were closed by polymerase chain reaction (PCR) and Sanger Sequencing leading to one circular chromosome (Figure 1).

The complete genome of *Micromonospora* sp. B006 contains a single chromosome composed of 7,040,153 bp, with no plasmids, and an average G+C content of 72.8% (Table 2). It is similar in size and G+C content to other *Micromonospora* genomes. A total of 6,251 protein-coding genes were identified by rapid annotation using subsystem technology (RAST) using ClassicRAST for coding sequence prediction.[−] rRNA operons and tRNA encoding genes were predicted using RAST, RNAmmer and tRNAscan (Table 2). A multi-locus species tree was generated using autoMLST (http://automlst.ziemertlab.com), which showed that strain B006 is most closely related to *M. marina* DSM 45555, *M. tulbaghiae* DSM 45142 and *M. aurantiaca* ATCC 27029, which belong to the group I taxon of *Micromonospora* (Figure S1).

Sixteen biosynthetic gene clusters predicted.

Genome analysis using antiSMASH, the bacteriocin genome mining tool BAGEL3, basic local alignment search tool (BLAST) and manual curation revealed 16 BGCs putatively encoding terpenes, lantipeptides, a thiopeptide, nonribosomal peptides, polyketides, and several hybrid compounds (Figure 1, Table 3), which is in range of what has been reported for other *Micromonospora* (from 7 up to 48 BGCs). We also re-analyzed the genome using the prediction informatics for secondary metabolomes (PRISM) tool to confirm the biosynthetic gene clusters. PRISM identified 13 BGCs in total, representing all biosynthetic classes compatible with its code, and including ten of the 16 BGCs identified by antiSMASH: nonribosomal peptide, type I and type II polyketide, and RiPP BGCs (clusters 3, 4, 5, 6, 10, 11, 13, 14, 15, and 16), whereas terpene, type III polyketide, non-NRPS siderophore, and diazepinomycin BGCs (clusters 1, 2, 7, 8, 9, and 12) were detected only by antiSMASH. In addition, three putative BGCs were identified only by PRISM. Analysis of these three BGCs indicate involvement in fatty acid biosynthesis and/or primary metabolism (Table S1).

Micromonospora sp. B006 dedicates approximately 8% of its genome to natural product biosynthesis.

Five biosynthetic gene clusters encode known natural products.

Bioinformatic analysis revealed five BGCs had a high level of similarity with previously annotated clusters: alkyl-*O*–dihydrogeranyl-methoxyhydroquinone BGC of *Actinoplanes missouriensis* (cluster 2, type III PKS-terpene), spore pigment cluster of *Salinispora tropica*

CNB-440 (cluster 5 type II PKS), desferrioxamine BGC of *Streptomyces coelicolor* A3(2)[9] (cluster 7, non-NRPS siderophore), and diazepinomicin BGC of *Micromonospora* sp. 046Eco11 (cluster 9, other) (Figure 1, Figure S2 and Table 3). In addition, *Micromonospora* sp. B006 produces a characteristic orange pigment during vegetative growth. One of the predicted terpene BGCs (cluster 1a) consisted of five genes (*MicB006_0242* to *MicB006_0246*) and showed high similarity to the carotenoid sioxanthin gene cluster *terp2* of *Salinispora tropica* CNB-440. The carotenoid biosynthesis genes are not confined to a single cluster. Four regions of the *Salinispora tropica* CNB-440 genome contribute to the biosynthesis of sioxanthin. The second gene cluster and the two independent genes involved in sioxanthin biosynthesis were identified using the tblastn function of BLAST which searches translated nucleotide databases using a protein query: *MicB006_3499* to *MicB006_3501* (*terp1*), *MicB006_2343* (*crtY*, lycopene cyclase gene), and *MicB006_1836* (*crtU*, β-carotene desaturase gene) (Figure 1, Figure S2A, and Table 3).

Micromonospora sp. B006 produces black spores as the culture ages. Although the precise structure of most spore pigments is unknown, their biosynthesis is reported to be catalyzed by type II PKS enzymes. Bioinformatics analysis revealed a 9.4-kilobase (kb) type II PKS gene cluster (cluster 5, Table 3), and the ten predicted genes were highly similar to a type II PKS gene cluster of *Salinispora tropica* CNB-440 that encodes a spore pigment (Figure S2B).

Cluster 7 shows sequence similarity to a gene set that encodes known desferrioxamine pathway biosynthetic enzymes (Figure S2C).[9] Desferrioxamines are low-molecular-weight, iron-chelating compounds. These NRPS-independent siderophores have been isolated from various actinomycetes, including *Streptomyces*, *Salinispora*, *Norcadia*, and *Micromonospora* species.[9]

Furthermore, a predicted type III PKS operon (cluster 2) showed high similarity (76 to 92%) to an alkyl-*O*−dihydrogeranyl-methoxyhydroquinone BGC of the actinomycete *Actinoplanes missouriensis* (Figure S2D).

Finally, bioinformatics analysis of the *Micromonospora* sp. B006 genome also revealed a BGC (cluster 9) that showed high homology and synteny with the diazepinomicin gene cluster from *Micromonospora* sp. 046Eco-11 (Figure S2E). The dibenzodiazepinone core of diazepinomicin is very rare among natural products and to date has only been reported to be produced by *Micromonospora* strains.[9, 9] Recently discovered BGCs in *Streptomyces griseoflavus* Tü4000 and in *Streptomyces* sp. WT3 showed high similarity to the diazepinomicin cluster, indicating that similar compounds are expected to be produced by *Streptomyces* strains, although experimental data linking the BGCs identified in *Streptomyces* to the produced compounds is missing.

Micromonospora sp. B006 produces diazepinomicin.

To confirm *Micromonospora* sp. B006 as a diazepinomicin producer, the strain was cultivated in HI medium, Bennett's medium, and modified Bennett's medium with activated Diaion® HP-20 resin and extracted with methanol as previously reported.[9] The compound was identified by comparison of spectroscopic data with those reported in the literature

(Figure 2). The UV absorption maxima (λ_{max} (MeOH) at 224 nm, 244 nm, and 294 nm) as well as the high resolution mass of 463.256 $[M + H]^+$ and the obtained fragmentation pattern corresponded with those reported for diazepinomicin (Figure 2).[9] 1H NMR data was also in agreement with the literature (see Experimental Section). Previously reported yields of pure diazepinomicin varied from 1.2 mg/L to 3.2 mg/L.[9,10] We obtained isolated yields of 19 mg/L from a culture grown in modified Bennett's medium which gave the highest titers among the media tested. HI medium was the second best medium (91% of modified Bennett's medium titers) and Bennett's medium, the third (70% of modified Bennett's medium titers).

Eleven biosynthetic gene clusters are orphan.

Eleven BGCs were identified in the genome of strain B006 for which the encoded compounds remain to be determined. Among these, an oligosaccharide-terpene cluster (cluster 8) contains six genes that show high similarity (68 to 74%) to brasilicardin biosynthetic genes, *bra1* to *bra5* and *bra11* from *Nocardia brasiliensis* IFM 0406 (Figure 3A). The corresponding *bra* genes found in B006 encode enzymes involved in the biosynthesis of the diterpene backbone of the immunosuppressant brasilicardin A. Homologs for the proteins Bra6 to Bra10 that are responsible for the L-rhamnose, N-acetylglucosamine, amino acid, and 3-hydroxybenzoate moieties of brasilicardin A, were not found. The predicted BGC instead contains genes encoding putative glycosyltransferases, an aminotransferase and enzymes involved in fatty acid biosynthesis. Thus, cluster 8 is predicted to encode a brasilicardin analogue containing the diterpene core but with distinct substitutions.

Moreover, a 56-kb enediyne PKS gene cluster (cluster 13) was found in the *Micromonospora* sp. B006 genome (Table 3, Figure 3B). Enediynes are among the most cytotoxic compounds reported, and only 13 members of this family have been structurally characterized to date.[7] Recent analyses of microbial genomes revealed that enediyne BGCs are relatively rare among bacteria.[7] A recent survey of an actinomycete collection containing 3,400 strains identified 81 hits for enediynes PKSs in *Streptomyces*, *Amycolatopsis*, *Kitasatospora*, *Nocardiopsis* and *Micromonospora*. Enediyne chemistry remains underexplored, as the identified gene clusters are non-redundant and highly diverse.[7] Comparison of known enediyne BGCs revealed that only a set of five genes is conserved in all clusters, encoding the enediyne PKS, a thioesterase and three unknown proteins commonly referred to as SgcE, SgcE10, and SgcE3/E4/E5, respectively, based on the C-1027 BGC nomenclature.[9] Cluster 13 contains the minimal enediyne PKS cassette as MicB006_4939 (SgcE), MicB006_4940 to MicB006_4942 (SgcE3/E4/E5) and MicB006_4938 (SgcE10) (Table S2).

Despite their remarkable structural diversity, enediynes can be classified into two subcategories, nine- or ten-membered enediyne cores. Phylogenetic analysis of the *Micromonospora* sp. B006 enediyne PKS revealed that it is putatively involved in the biosynthesis of a ten-membered enediyne as the B006 PKS gene *MicB006_4939* is most closely related to CalE8 from the calicheamicin cluster (59% identity at aa level, Figure S3). The hypothesis that cluster 13 encodes a ten-membered enediyne is also supported by the identification of a CalR3 homolog (MicB006_4950, 66% identity at aa level) and a CalT4

homolog (MicB006_4922, 43% identity at aa level), which appear to be 10-membered specific. The gene *calR3* encodes a putative protein kinase, while *calT4* is associated with membrane transport. Other genes identified in cluster 13 include five transcriptional regulators, five transporters, five genes encoding enzymes involved in sugar biosynthesis, three methyltransferases, one polyketide cyclase, one dehydrogenase, one α/β hydrolase fold enzyme, and one cytochrome P450. The identified enediyne cluster was also found in *Micromonospora* sp. TSRI0369. The genome neighborhood of cluster 13 appears to be distinct from enediyne BGCs that are linked to known structures, suggesting that this orphan cluster may code for the biosynthesis of a novel enediyne.

In addition to the two pathways putatively encoding novel analogues of known scaffolds, *Micromonospora* sp. B006 harbors three lantipeptide/thiopeptide operons, one terpene-bacteriocin cluster, two NRPS clusters, and three NRPS-PKS hybrid clusters (Table 3, Figure 4, Figure 5). Since these BGCs showed no sequence similarity to characterized clusters, they remain orphan.

Particularly remarkable is cluster 10, a 215.8-kb NRPS-PKS hybrid BGC consisting of 36 modules and 161 domains in total (Figure 5, Table S3). It contains genes encoding two monomodular NRPSs, 13 PKSs with a total of 34 modules, and a single thioesterase (TE) domain. 3-amino-5-hydroxybenzoic acid (AHBA) is the predicted substrate of the CoA ligase domain encoded by PKS gene *MicB006_3061*. Because in the case of rifamycin biosynthesis AHBA activation has been shown to be independent of coenzyme A, we referred to this domain as adenylation (A) domain instead of CoA ligase (Figure 5). In any case, AHBA is the likely starter unit for biosynthesis of the polyketide-nonribosomal peptide encoded in cluster 10. Therefore, we carried out a tblastn analysis using eight *rif* genes encoding RifG to RifN from *Amycolatopsis mediterranei* S699 as protein queries, which are involved in AHBA formation during rifamycin biosynthesis.[9] Six genes encoding Rif homologs were identified in cluster 10: *MicB006_3062* (RifG, amino-3-dehydroquinic acid synthase), *MicB006_3069* (RifJ, 5-deoxy-5-amino-3-dehydroquinate dehydratase), *MicB006_3068* (RifK, AHBA synthase), *MicB006_3067* (RifL, oxidoreductase), *MicB006_3066* (RifM, phosphatase), and *MicB006_3071* (RifN, kinase). In addition, two genes encoding RifH (amino-3-deoxy-D-arabinoheptulosonate 7-phosphate synthase) homologs were identified outside the predicted borders of cluster 10: *MicB006_3510* and *MicB006_2892*. The latter one is part of the diazepinomicin biosynthetic gene cluster. No RifI homolog was found. The gene *rifI* encodes an aminoshikimate dehydrogenase that has been shown to be nonessential for AHBA biosynthesis.

AHBA is the precursor to a large group of specialized metabolites produced mostly by actinomycetes, including the family of ansamycins, saliniketals, and the family of mitomycins. These natural products exhibit a number of biological and pharmacological activities, such as antibacterial, anticancer, lipoxygenase inhibitory and antioxidant activities. AHBA synthase proteins such as RifK can be divided into naphthalenic and benzenic subgroups. Phylogenetic analysis revealed that the *Micromonospora* sp. B006 AHBA synthase (MicB006_3068) forms a subclade in the benzenic AHBA synthase clade (Figure S4).

Loading of the PKS with AHBA would be followed by 32 extensions (rather than 33 as module 19 seems to be inactive) before polyketide chain release is catalyzed by the TE domain predicted in PKS gene *MicB006_3044*. The role of the two monomodular NRPS genes is less clear. Conceivably, a dipeptide could be condensed with the polyketide product and released by the C_2 domain in module 35. It was shown that C domains can use a free substrate as the acceptor nucleophile to couple with a carrier protein-tethered acyl donor for ester bond formation, *e.g.*, SgcC5 (C-1027), or amide bond formation, *e.g.*, VibH (vibriobactin).

Gene clusters with similarity to cluster 10 were also found in *Micromonospora* sp. L5, isolated from nitrogen fixing root nodules of *Casuarina equisetifolia*, and *Micromonospora aurantiaca* ATCC27029, which is closely related to the strain L5. In contrast to *Micromonospora aurantiaca* and to strain B006, *Micromonospora* sp. L5 contains a shorter version of the PKS gene *MicB066_3055*. An alignment of the three proteins revealed that modules 5 and 6 of the B006 cluster 10 are missing in the reported genome sequence of *Micromonospora* sp. L5. The fact that such a relatively large portion of the genome (~3%) is dedicated to making one compound (or one compound class) that appears to be conserved in various *Micromonospora* strains, suggests that the encoded natural product is of critical importance to its host.

We recently reported that *Micromonospora* sp. B006 produces diazaquinomycins H and J. None of the eleven orphan BGCs seem to match the biosynthesis of these compounds. The identification of the diazaquinomycin BGC is ongoing and will be reported elsewhere.

The majority of orphan BGCs are transcriptionally active.

We examined the transcription of selected genes within all eleven orphan BGCs under growth conditions routinely used in our laboratories. Semi-quantitative reverse transcription-PCR analyses indicated that eight (clusters 3, 4, 10, 12, 13, 14, 15, and 16) of the eleven orphan BGCs are transcribed to various extents under the culture conditions used, whereas transcription could not be detected for the remaining three BGCs (clusters 6, 8, and 11) (Figure 6).

Genetics system for strain B006 established.

A genetics system would be helpful to explore the function of the predicted orphan gene clusters in the native host. We first performed an antibiotic sensitivity test in order to identify suitable selection markers. Strain B006 proved to be sensitive to apramycin at concentrations of ≥25 mg/L. Therefore, the widely used *aac(3)IV* apramycin resistance gene (apr^R) was chosen for mutant selection. In order to establish a genetics system, we decided to target the only type II PKS cluster present in strain B006, the predicted spore pigment biosynthesis operon (cluster 5, Figure S2B). Plasmid pJB002EL was constructed to disrupt gene *MicB006_2548* encoding a putative ketosynthase alpha domain (KSα) via a single crossover (Figure 7A). pJB002EL is based on vector pKC1132 previously used in *Micromonospora*, and relying on the apramycin sensitivity of strain B006.

pJB002EL was transferred into *Micromonospora* sp. B006 from either *E. coli* S17–1, a methylation-proficient donor strain, or *E. coli* ET12567/pUZ8002, a methylation-deficient

donor strain. The conjugation procedures were based on the protocol for *Streptomyces* described by Kieser and colleagues. Exconjugants were obtained using 10^8 spores which were heat shocked without 2×YT medium and plated on A1 agar. Under these conditions the conjugation efficiency for both *E. coli* strains tested was 2×10^{-7} exconjugants per recipient spore. In contrast, conjugation using mycelium from *Micromonospora* sp. B006 liquid cultures did not yield any exconjugants. Apramycin-resistant exconjugants were confirmed by PCR (Figure 7B). Four, out of eight tested, apramycin-resistant clones carried a disrupted version of the KSα gene *MicB006_2548* based on PCR analysis. As expected, the KSα mutants showed loss of spore pigmentation compared to wild-type *Micromonospora* sp. B006, confirming that cluster 5 encodes the spore pigment (Figure 7C) and validating methods for gene inactivation in strain B006, which shall aid future exploration of orphan biosynthetic gene clusters.

Conclusions.

In contrast to their terrestrial and marine counterparts, freshwater actinomycetes are an underutilized resource of natural products. The complete genome of the Lake Michigan-derived *Micromonospora* sp. strain B006 (Figure 1) was sequenced with Illumina Technology and Oxford Nanopore MinION, the latter of which is a promising new sequencing technology with applications in scaffolding genome sequences from short reads and resolving repeat sequences as found in NRPS and PKS genes (Tables 1 and 2).

Bioinformatic analysis revealed that strain B006 dedicates approximately 8% of its genome to natural product biosynthesis. Five of the predicted 16 BGCs could be assigned to known products (Figure 1, Table 3, and Figure S2). As predicted from its genome, we showed that strain B006 can produce diazepinomicin. Isolated yields were approximately six-fold higher than previously reported for other *Micromonospora* spp.[9, 10] so that strain B006 may serve as an alternative source of this clinically relevant compound (Figure 2). The remaining eleven BGCs remain orphan (Figures 3–5). It is often assumed that orphan BGCs are transcriptionally silent under laboratory growth conditions. Transcript analyses indicated that the majority of the orphan BGCs of strain B006 are transcriptionally active to various extents under the growth conditions tested, whereas only three appear transcriptionally silent (Figure 6), which is in agreement with a recent report by Amos *et al.* showing that over half of 46 investigated BGCs were transcriptionally active. Nevertheless, it is unknown if the level of transcription observed would be sufficient for compound detection. Two orphan BGCs are particularly notable: the enedyine cluster 13 (Figure 3) due to the high cytotoxicity of most compounds biosynthesized by enedyine PKS enzymes, and the uncommonly large PKS-NRPS cluster 10 that appears to encode a polyene with an AHBA starter unit (Figure 5). Finally, we established a genetics system that can be used for gene inactivation (Figure 7) or promoter replacements which will facilitate future exploration of orphan BGCs. The genome of *Micromonospora* sp. B006 showcases the rich biosynthetic capabilities of a freshwater actinomycete, setting the stage for its future exploration towards natural product discovery.

EXPERIMENTAL SECTION

General Experimental Procedures.

All chemicals were acquired from Sigma-Aldrich, Alfa Aesar, VWR, and Fisher Scientific (Table S4). Solvents were of high performance liquid chromatography (HPLC) grade or higher. Restriction enzymes were purchased from New England Biolabs. Oligonucleotide primers were synthesized by Sigma-Aldrich. Molecular biology procedures were carried out according to the manufactures' instructions (Ambion, New England Biolabs, Fisher Scientific, Promega, Sigma-Aldrich, Qiagen, Zymo Research, Illumina, Oxford Nanopore) or as described below.

Genome Sequencing.

Micromonospora sp. B006 was isolated from Lake Michigan sediment. High molecular weight genomic DNA was isolated from strain B006 using the Blood & Cell Culture DNA Midi Kit (Qiagen). Genomic DNA was prepared for shotgun metagenome sequencing using a Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA). The library was sequenced on an Illumina NextSeq500 instrument, employing paired-end 150 base reads. In addition, sequencing libraries were generated from genomic DNA using the Oxford Nanopore 1D library protocol, according to the manufacturer's instructions. Sequencing was performed using a v9.4 SpotON MinION flow cell and employing an Oxford Nanopore MinION instrument. De novo assembly was performed using the Spades assembler on both raw Illumina and Nanopore reads, with multiple k-mers specified as "-k 31,51,71,91". Coverage levels were assessed by mapping raw Illumina reads back to the contigs with bowtie2 and computing the coverage as the number of reads aligning per contig times the length of each read divided by the length of the contig.

Gaps between contigs were closed by PCR using primer pairs oJB27/28, oJB31/32, oJB33/34, oJB35/36, oJB37/103, oJB68/69, and oJB74/75 (Table S5). The reaction (50 μL) with primer pair oJB68/69 consisted of 0.2 mM of each dNTP, 1× Q5 High GC Enhancer, 0.5 μM of each primer, and 0.02 U/μL Q5 High-Fidelity DNA Polymerase (New England Biolabs) in Q5 reaction buffer supplied with the enzyme. The following thermal cycling conditions were used: 30 s at 98 °C; 35 cycles of 98 °C for 10 s; and 72 °C for 80 s; and a terminal hold for 2 min at 72 °C. The other reactions (50 μL) consisted of 1.5–2.0 mM MgCl2, 0.2 mM each dNTP, 3% dimethyl sulfoxide (DMSO), 0.5 μM (each) primer, and 0.02 U/μL Phusion High-Fidelity DNA Polymerase (Thermo Scientific) in the buffer supplied with the enzyme and using the following thermal cycling parameters (except for primer pair oJB74/75): 30 s at 98 °C; 35 cycles of 98 °C for 10 s, and 72 °C for 75 s; and a terminal extension for 5 min at 72 °C. The thermocycling conditions for primer pair oJB74/75 were as follows: 30 s at 98 °C; 35 cycles of 98 °C for 10 s, 62 °C for 20 s, and 72 °C for 60 s; and a terminal extension for 5 min at 72 °C.

The complete genome has been deposited at GenBank under accession no. CP030865.

Bioinformatics analysis.

The genome was annotated using RAST.[−] Ribosomal RNA genes and transfer RNA were detected by using RAST, the RNAmmer 1.2 software and tRNAscan-SE 2.0. Biosynthetic gene clusters and encoded compounds were roughly predicted with antiSMASH, BAGEL3, and PRISM3 and further annotated using BLAST followed by manual curation.

Cultivation conditions.

Micromonospora sp. B006 was routinely cultivated at 30 °C on ISP2 medium (0.4% yeast extract, 1% malt extract, 0.4% dextrose, pH 7.3, and 2% agar for solid medium). To isolate DNA for genome sequencing, the wild-type strain was cultivated in a 1:1 mixture of TSB medium (3% tryptic soy broth) and YEME medium (0.3% yeast extract, 0.5% peptone, 0.3% malt extract, 1% glucose, 34% sucrose, 5 mM MgCl2) for two days at 30 °C and 200 rpm. A defrosted cryo stock of the wild-type strain was used for inoculation. Cryo stocks were prepared with mycelium from three- to five-day old ISP2 liquid cultures by adding glycerol to 20% [v/v] final concentration followed by storage at −80 °C. For transcriptional analysis, 50 mL A1 liquid medium in filtered Lake Michigan water (0.5 L of filtered Lake Michigan water, 0.5 L DI water, 10 g starch, 4 g yeast extract, 2 g peptone, 1 g calcium carbonate, 100 mg potassium bromide, and 40 mg iron sulfate per liter A1 medium) was inoculated with 400 µL of a *Micromonospora* sp. B006 wild-type strain cryo stock. The pre-culture was incubated for five days at 30 °C and 200 rpm. Subsequently, 5% [v/v] of this pre-culture were used to inoculate 50 mL A1 liquid medium in a 250 mL Erlenmeyer flask containing a stainless steel spring. The main culture was incubated for nine days at 20 °C and 210 rpm.

For diazepinomicin production, ISP2 liquid medium was inoculated with 400 µL of a *Micromonospora* sp. B006 wild-type cryo stock, and the culture was incubated at 30 °C and 200 rpm for five days. This seed culture (2% [v/v]) was used to inoculate either Bennett's medium (0.1% yeast extract, 0.077% beef extract, 0.2% N-Z-Amine A, 1% dextrose in tap water, pH 7.3), modified Bennet's medium (0.1% yeast extract, 0.1% beef extract, 0.2% N-Z-Amine A, 1.5% glucose in tap water, pH 7.3), or HI medium (3% glycerol, 2% potato dextrin, 0.25% Bactopeptone, 0.83% yeast extract, 0.1% CaCO3 in tap water, pH 7.0 adjusted before the addition of CaCO3) in 250-mL Erlenmeyer flasks containing 50 mL of medium. Each medium was supplemented with 5% activated Diaion® HP-20 resin at the time of inoculation. The HP-20 resin was activated prior to use by soaking in methanol and then rinsing thoroughly with distilled water. The cultures were incubated at 30 °C, 200 rpm for seven days.

For sporulation phenotyping, wild-type *Micromonospora* sp. B006 and single-crossover mutants were inoculated on solid A1 medium (2% w/v agar) in filtered Lake Michigan water and incubated for 14 to 21 days at 30 °C.

To obtain spores for conjugation, *Micromonospora* sp. B006 was cultivated at 30 °C on ISP2 solid medium for three to four weeks.

E. coli strains were cultivated in LB medium supplemented with the appropriate antibiotics. The following antibiotics were used as selection markers: apramycin (final concentration: 50

μg/mL), kanamycin (50 μg/mL), chloramphenicol (25 μg/mL), and nalidixic acid (25 μg/mL).

Metabolite analysis.

After seven days incubation, the pH of the diazepinomicin production fermentations was adjusted to 4 with acetic acid. Subsequently, the cultures were harvested by centrifugation. The supernatant was decanted and the cell/resin pellet was extracted three times, each with 20 mL methanol. After removing solvent under reduced pressure, the crude extracts were dissolved in methanol for HPLC (final concentration of 100 mg/mL) and mass spectrometry (MS) analyses (final concentration of 0.2 mg/mL).

HPLC analysis was performed on an Agilent 1260 Infinity system equipped with a Kinetex® C 18 column (150 × 4.6 mm, 5 μm particle size, 100 Å pore size, Phenomenex®). Solvent A was 0.1% [v/v] trifluoroacetic acid (TFA) in water, and solvent B was acetonitrile. The solvent gradient was: initial hold at 50% B for 2 min, linear gradient from 50% to 100% B within 10 min, and held for 5 min, at a flow rate of 1.0 mL/min. The detection wave length range was 200 – 600 nm; chromatograms were extracted at λ = 240 nm.

High-resolution and tandem MS spectra were recorded on a Shimadzu LC MS instrument in positive and negative mode using electrospray ionization and equipped with a Kinetex® C18 RPLC column (50 × 2.1 mm, 1.7 μm particle size, 100 Å pore size, Phenomenex®). Solvent A was 0.1% [v/v] formic acid in water, and solvent B was 0.1% [v/v] formic acid in acetonitrile. The solvent gradient was: linear gradient from 20% to 100% B in 7 min, at a flow rate of 0.5 mL/min. The detected mass range was *m*/*z* 150 – 3000. Prior to LC high resolution MS analysis, extracts were subjected to a cleanup procedure using Extract Clean C18 cartridges (Agela Technologies) and methanol.

Isolation of diazepinomicin and NMR analysis.

Diazepinomicin was isolated from a 50 mL *Micromonospora* sp. B006 culture (modified Bennett's medium with 5% activated Diaion® HP-20 resin). The culture was extracted with methanol as described above. Diazepinomicin isolation was carried out by semipreparative HPLC using an Agilent Microsorb 60–8 C_{18} Dynamax column (250 × 10 mm), flow rate of 4 mL/min, and a gradient program of water (solvent A) and acetonitrile (solvent B) set to 50% B for 5 min, 50% to 85% B for 25 min, 85% to 100% B for 5 min, and hold at 100% B for additional 10 min. Fractions were dried *in vacuo* to yield 1.17 mg of diazepinomicin (81% purity at 240 nm). To stabilize diazepinomicin, the sample was spiked with 0.06 mg TFA. Although the sample was spiked with TFA to stabilize it, we started to observe degradation of diazepinomicin in less than 24 h after isolation. Thus, the NMR spectrum and the quality control HPLC chromatogram to determine purity were obtained immediately after isolation. The ^{1}H NMR spectrum was acquired on a Bruker 800 MHz spectrometer. NMR chemical shifts were referenced to residual solvent peak (DMSO-d_6 δH 2.50).

Diazepinomycin:

UV (MeOH) λ_{max} 224, 244, 294 nm; ^{1}H NMR (DMSO-d_6, 800 MHz) δ 10.02 (1H, s), 9.93 (1H, s), 9.02 (1H, s), 7.06 (1H, dd, *J* = 7.9, 1.0), 6.83 (1H, dd, *J* = 7.8, 1.1), 6.73 (1H, s),

6.70 (1H, t, *J* = 7.8), 6.18 (1H, d, *J* = 2.4), 6.16 (1H, d, *J* = 2.4), 5.25 (1H, dd, *J* = 7.4, 5.3), 5.06 (1H, m), 5.04 (1H, m), 4.39 (2H, d, *J* = 6.0), 2.02 (2H, m), 1.99 (2H, m), 1.97 (2H, m), 1.91 (2H, m), 1.65 (3H, s), 1.61 (3H, s), 1.55 (3H, s), 1.52 (3H, s); High resolution MS *m/z* 463.256 $[M+H]^+$ (calcd. for C28H35N2O4, 463.2591 $[M+H]^+$).

Transcript analysis.

Micromonospora sp. B006 wild-type was cultivated in 50 mL A1 liquid medium for nine days. Two-mL samples were taken every 24 h. The cells were collected by centrifugation and stored at −20 °C. Total RNA was isolated from samples collected on days 3, 5, 7, and 9 using the RiboPure™-Bacteria Kit (Ambion). Cell lysis was obtained by homogenization using a Mini-Beadbeater™ (Biospec products) (5 cycles of 30 s). DNaseI treatment was carried out for 4 h using TURBO DNA-*free*™ Kit (Ambion). The absence of genomic DNA in the samples was confirmed by PCR using 16S rRNA primers oJB184 and oJB185 (Table S5). The 20 μL reactions consisted of 0.2 mM of each dNTP, 3% DMSO, 0.25 μM of each primer, and 1.25 U DreamTaq DNA Polymerase (Thermo Scientific) in reaction buffer supplied with the enzyme. Thermocycling parameters were initial denaturation for 2 min at 95 °C; amplification: 30 cycles (95 °C for 30 s, 60 °C for 30 s, 72 °C for 60 s); and terminal hold for 5 min at 72 °C. cDNA synthesis was carried out with 250 ng RNA by using the SuperScript ® IV First-Strand cDNA Synthesis System (Invitrogen). A portion of the first-strand reaction (1 μL) was used as template in subsequent PCRs. The 20 μL reactions to amplify cDNAs of selected genes of all orphan biosynthetic gene clusters (except cluster 8, see below) consisted of 0.2 mM of each dNTP, 1× Q5 High GC Enhancer, 0.5 μM of each primer, and 0.02 U/μL Q5 High-Fidelity DNA Polymerase (New England Biolabs) in Q5 reaction buffer supplied with the enzyme. The following thermal cycling conditions were used: 30 s at 98 °C; 30 cycles of 98 °C for 10 s, 72 °C (except for primer pairs oJB82/83 and oJB124/125: 68.9 °C instead) for 20 s, and 72 °C for 45 s; and a terminal hold for 2 min at 72 °C. The cDNA of *MicB006_2812* (cluster 8) was amplified with JumpStart™ REDTaq® ReadyMix™, 3% DMSO, and 0.4 μM of each primer, in a total volume of 25 μL. Thermocycling parameters were initial denaturation for 2 min at 94 °C; amplification: 35 cycles (94 °C for 30 s, 58.7 °C for 30 s, 72 °C for 60 s); and terminal hold for 5 min at 72 °C. RNA before reverse transcription was used as negative control.

Antibiotic sensitivity test.

Approximately 10^6 spores of *Micromonospora* sp. B006 were spread on ISP2 agar containing the following final concentrations of apramycin: 0 μg/mL, 25 μg/mL, 50 μg/mL, 100 μg/mL, and 200 μg/mL. The plates were incubated for seven days at 30 °C, after which time growth was observed only on plates without apramycin. Although strain B006 appeared sensitive to apramycin at 25 μg/mL, we chose to use the more commonly employed concentration of 50 μg/mL for the selection of mutants. pKC1132 carries the *aac(3)IV* apramycin resistance gene, and is a common vector used for gene disruption and gene replacement experiments in actinomycete bacteria. Therefore, pKC1132 was chosen for gene inactivation experiments as described below. For *Micromonospora* that are resistant to apramycin, other antibiotics could be tested such as thiostrepton.

Plasmid construction for gene inactivation in *Micromonospora* sp. B006.

Genomic DNA was isolated from *Micromonospora* sp. B006 using the GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich). Primers oJB3 and oJB4 (Table S5) were used to amplify a 617 bp fragment of gene *MicB006_2548* (Figure 7A), which encodes a putative KSα domain. The PCR reaction (50 μL) consisted of 1.5 mM MgCl2, 0.2 mM each dNTP, 3% DMSO, 0.5 μM each primer, and 0.02 U/μL Phusion High-Fidelity DNA Polymerase (Thermo Scientific) in Phusion HF buffer supplied with the enzyme and using the following thermal cycling parameters: 30 s at 98 °C; 35 cycles of 98 °C for 10 s, and 72 °C for 40 s; and a terminal extension for 5 min at 72 °C. The primers introduced the restrictions sites *Hind*III and *Bam*HI into the PCR product, which was cloned into the same sites of *E. coli* vector pKC1132 to generate plasmid pJB002EL.

Gene inactivation in *Micromonospora* sp. B006.

The KSα gene *MicB006_2548* was inactivated by homologous recombination via a single-crossover after introducing pJB002EL into *Micromonospora* sp. B006. Plasmid pJB002EL was transferred into *Micromonospora* sp. B006 by conjugation from *E. coli* S17–1 and ET12567/pUZ8002, respectively, as described by Kieser *et al*. with some modifications. Approximately 10^8 spores (in 20% glycerol) were defrosted on ice, centrifuged at 13,000 rpm for 1 min and the supernatant was removed. No 2×YT broth was added to the spores before heat shock. *E. coli* was prepared according to Kieser *et al.* and conjugation mixtures were plated on A1 agar prepared with filtered Lake Michigan water. After 16 hours incubation at 30 °C, the conjugation plates were overlaid with nalidixic acid (final concentration: 25 μg/mL) and apramycin (final concentration: 50 μg/mL). Apramycin-resistant colonies were streaked on A1 medium plates containing 50 μg/mL apramycin and 25 μg/mL nalidixic acid until pure cultures were obtained. Gene disruption was confirmed by PCR with primer pairs oJB20/oJB22 (Table S5) in reactions containing 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 3% DMSO, 0.2 μM (each) primer, and 1.25 U DreamTaq Green DNA polymerase (ThermoFisher) in the buffer supplied with the enzyme, in a total volume of 25 μL. Thermocycling conditions were initial denaturation: 2 min at 95 °C; amplification: 35 cycles (95 °C for 30 s, 56 °C for 30 s, 72 °C for 75 s); terminal extension: 5 min at 72 °C.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Financial support for this work was provided by the National Center for Advancing Translational Sciences, National Institutes of Health (NIH), under grant KL2TR002002 (to ASE), by NIH under grant UL1TR002003 (to MMC), by CAPES/Science without Borders BEX 13055-13-5 (to CMM), and by startup funds from the Department of Medicinal Chemistry and Pharmacognosy and the Center for Biomolecular Sciences, University of Illinois at Chicago (to ASE). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

(1). Newman DJ; Cragg GM J. Nat. Prod 2016, 79, 629–661. [PubMed: 26852623]

(2). Pye CR; Bertin MJ; Lokey RS; Gerwick WH; Linington RG Proc. Natl. Acad. Sci. U.S.A 2017, 114, 5601–5606. [PubMed: 28461474]

(3). Genilloud O Nat. Prod. Rep 2017, 34, 1203–1232. [PubMed: 28820533]

(4). Carballeira NM; Pagan M; Shalabi F; Nechev JT; Lahtchev K; Ivanova A; Stefanov KJ Nat. Prod 2000, 63, 1573–1575.

(5). Carlson S; Tanouye U; Omarsdottir S; Murphy BT J. Nat. Prod 2015, 78, 381–387. [PubMed: 25537064]

(6). Shaikh AF; Elfeki M; Landolfa S; Tanouye U; Green SJ; Murphy BT Nat. Prod. Sci 2015, 21, 261–267.

(7). Mullowney MW; Hwang CH; Newsome AG; Wei X; Tanouye U; Wan B; Carlson S; Barranis NJ; Ó hAinmhire E; Chen WL; Krishnamoorthy K; White J; Blair R; Lee H; Burdette JE; Rathod PK; Parish T; Cho S; Franzblau SG; Murphy BT ACS Infect. Dis 2015, 1, 168–174. [PubMed: 26594660]

(8). Barka EA; Vatsa P; Sanchez L; Gaveau-Vaillant N; Jacquard C; Meier-Kolthoff JP; Klenk HP; Clement C; Ouhdouch Y; van Wezel GP Microbiol. Mol. Biol. Rev 2016, 80, 1–43. [PubMed: 26609051]

(9). Boumehira AZ; El-Enshasy HA; Hacène H; Elsayed EA; Aziz R; Park EY Biotechnol. Bioproc. E 2016, 21, 199–223.

(10). Trujillo ME; Bacigalupe R; Pujic P; Igarashi Y; Benito P; Riesco R; Medigue C; Normand P PLoS One 2014, 9, e108522. [PubMed: 25268993]

(11). Talukdar M; Das D; Bora C; Bora TC; Deka Boruah HP; Singh AK Gene 2016, 594, 97–107. [PubMed: 27609432]

(12). Carro L; Nouioui I; Sangal V; Meier-Kolthoff JP; Trujillo ME; Montero-Calasanz MDC; Sahin N; Smith DL; Kim KE; Peluso P; Deshpande S; Woyke T; Shapiro N; Kyrpides NC; Klenk HP; Goker M; Goodfellow M Sci. Rep 2018, 8, 525. [PubMed: 29323202]

(13). Machado H; Tuttle RN; Jensen PR Curr. Opin. Microbiol 2017, 39, 136–142. [PubMed: 29175703]

(14). Okumura H; Kobaru S US Patent 5,541,181 A, 1996.

(15). Igarashi Y; Miyanaga S; Onaka H; Takeshita M; Furumai TJ Antibiot 2005, 58, 350–352.

(16). Charan RD; Schlingmann G; Janso J; Bernan V; Feng X; Carter GT J. Nat. Prod 2004, 67, 1431–1433. [PubMed: 15332871]

(17). Bachmann BO; McAlpine JB; Zazopoulos E; Farnet CM; Piraee M WO 2004/065591 A1, 2004.

(18). Mason WP; Belanger K; Nicholas G; Vallieres I; Mathieu D; Kavan P; Desjardins A; Omuro A; Reymond DJ Neurooncol 2012, 107, 343–349.

(19). 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 J. Comput. Biol 2012, 19, 455–477. [PubMed: 22506599]

(20). Blin K; Wolf T; Chevrette MG; Lu X; Schwalen CJ; Kautsar SA; Suarez Duran HG; de Los Santos ELC; Kim HU; Nave M; Dickschat JS; Mitchell DA; Shelest E; Breitling R; Takano E; Lee SY; Weber T; Medema MH Nucleic Acids Res 2017, 45, W36–W41. [PubMed: 28460038]

(21). Gomez-Escribano JP; Alt S; Bibb MJ Mar. Drugs 2016, 14, E78. [PubMed: 27089350]

(22). Giordano F; Aigrain L; Quail MA; Coupland P; Bonfield JK; Davies RM; Tischler G; Jackson DK; Keane TM; Li J; Yue JX; Liti G; Durbin R; Ning Z Sci. Rep 2017, 7, 3935. [PubMed: 28638050]

(23). Langmead B; Salzberg SL Nat. Methods 2012, 9, 357–359. [PubMed: 22388286]

(24). Hirsch AM; Alvarado J; Bruce D; Chertkov O; De Hoff PL; Detter JC; Fujishige NA; Goodwin LA; Han J; Han S; Ivanova N; Land ML; Lum MR; Milani-Nejad N; Nolan M; Pati A; Pitluck S; Tran SS; Woyke T; Valdes M Genome Announc 2013, 1, e00759–00713. [PubMed: 24072863]

(25). Carver T; Thomson N; Bleasby A; Berriman M; Parkhill J Bioinformatics 2009, 25, 119–120. [PubMed: 18990721]

(26). Aziz RK; Bartels D; Best AA; DeJongh M; Disz T; Edwards RA; Formsma K; Gerdes S; Glass EM; Kubal M; Meyer F; Olsen GJ; Olson R; Osterman AL; Overbeek RA; McNeil LK;

Paarmann D; Paczian T; Parrello B; Pusch GD; Reich C; Stevens R; Vassieva O; Vonstein V; Wilke A; Zagnitko O BMC Genomics 2008, 9, 75. [PubMed: 18261238]

(27). Overbeek R; Olson R; Pusch GD; Olsen GJ; Davis JJ; Disz T; Edwards RA; Gerdes S; Parrello B; Shukla M; Vonstein V; Wattam AR; Xia F; Stevens R Nucleic Acids Res 2014, 42, D206–214. [PubMed: 24293654]

(28). Brettin T; Davis JJ; Disz T; Edwards RA; Gerdes S; Olsen GJ; Olson R; Overbeek R; Parrello B; Pusch GD; Shukla M; Thomason JA, 3rd; Stevens R; Vonstein V; Wattam AR; Xia F Sci. Rep 2015, 5, 8365. [PubMed: 25666585]

(29). Lagesen K; Hallin P; Rodland EA; Staerfeldt HH; Rognes T; Ussery DW Nucleic Acids Res 2007, 35, 3100–3108. [PubMed: 17452365]

(30). Lowe TM; Chan PP Nucleic Acids Res 2016, 44, W54–57. [PubMed: 27174935]

(31). Nguyen LT; Schmidt HA; von Haeseler A; Minh BQ Mol. Biol. Evol 2015, 32, 268–274. [PubMed: 25371430]

(32). van Heel AJ; de Jong A; Montalban-Lopez M; Kok J; Kuipers OP Nucleic Acids Res 2013, 41, W448–453. [PubMed: 23677608]

(33). Altschul SF; Gish W; Miller W; Myers EW; Lipman DJ J. Mol. Biol 1990, 215, 403–410. [PubMed: 2231712]

(34). Skinnider MA; Merwin NJ; Johnston CW; Magarvey NA Nucleic Acids Res 2017, 45, W49–W54. [PubMed: 28460067]

(35). Awakawa T; Fujita N; Hayakawa M; Ohnishi Y; Horinouchi S Chem. Biochem 2011, 12, 439–448.

(36). Kersten RD; Lane AL; Nett M; Richter TK; Duggan BM; Dorrestein PC; Moore BS Chem. Biochem 2013, 14, 955–962.

(37). Barona-Gomez F; Wong U; Giannakopulos AE; Derrick PJ; Challis GL J. Am. Chem. Soc 2004, 126, 16282–16283. [PubMed: 15600304]

(38). Bunet R; Brock A; Rexer HU; Takano E FEMS Microbiol. Lett 2006, 262, 57–64. [PubMed: 16907739]

(39). McAlpine JB; Banskota AH; Charan RD; Schlingmann G; Zazopoulos E; Piraee M; Janso J; Bernan VS; Aouidate M; Farnet CM; Feng X; Zhao Z; Carter GT J. Nat. Prod 2008, 71, 1585–1590. [PubMed: 18722414]

(40). Richter TK; Hughes CC; Moore BS Environ. Microbiol 2015, 17, 2158–2171. [PubMed: 25329237]

(41). Lee TS; Khosla C; Tang YJ Antibiot 2005, 58, 663–666.

(42). Bruns H; Crusemann M; Letzel AC; Alanjary M; McInerney JO; Jensen PR; Schulz S; Moore BS; Ziemert N ISME J 2018, 12, 320–329. [PubMed: 28809850]

(43). Bonitz T; Zubeil F; Grond S; Heide L PLoS One 2013, 8, e85707. [PubMed: 24376894]

(44). Abdelmohsen UR; Szesny M; Othman EM; Schirmeister T; Grond S; Stopper H; Hentschel U Mar. Drugs 2012, 10, 2208–2221. [PubMed: 23170078]

(45). Hayashi Y; Matsuura N; Toshima H; Itoh N; Ishikawa J; Mikami Y; Dairi TJ Antibiot 2008, 61, 164–174.

(46). Liu W; Nonaka K; Nie L; Zhang J; Christenson SD; Bae J; Van Lanen SG; Zazopoulos E; Farnet CM; Yang CF; Shen B Chem. Biol 2005, 12, 293–302. [PubMed: 15797213]

(47). Liu W; Christenson SD; Standage S; Shen B Science 2002, 297, 1170–1173. [PubMed: 12183628]

(48). Van Lanen SG; Oh TJ; Liu W; Wendt-Pienkowski E; Shen BJ Am. Chem. Soc 2007, 129, 13082–13094.

(49). Lohman JR; Huang SX; Horsman GP; Dilfer PE; Huang T; Chen Y; Wendt-Pienkowski E; Shen B Mol. Biosyst 2013, 9, 478–491. [PubMed: 23360970]

(50). McGlinchey RP; Nett M; Moore BS J. Am. Chem. Soc 2008, 130, 2406–2407. [PubMed: 18232689]

(51). Lane AL; Nam SJ; Fukuda T; Yamanaka K; Kauffman CA; Jensen PR; Fenical W; Moore BS J. Am. Chem. Soc 2013, 135, 4171–4174. [PubMed: 23458364]

(52). Ahlert J; Shepard E; Lomovskaya N; Zazopoulos E; Staffa A; Bachmann BO; Huang K; Fonstein L; Czisny A; Whitwam RE; Farnet CM; Thorson JS Science 2002, 297, 1173–1176. [PubMed: 12183629]

(53). Gao Q; Thorson JS FEMS Microbiol. Lett 2008, 282, 105–114. [PubMed: 18328078]

(54). Liu W; Ahlert J; Gao Q; Wendt-Pienkowski E; Shen B; Thorson JS Proc. Natl. Acad. Sci. U.S.A 2003, 100, 11959–11963. [PubMed: 14528002]

(55). Yan X; Ge H; Huang T; Hindra; Yang D; Teng Q; Crnovcic I; Li X; Rudolf JD; Lohman JR; Gansemans Y; Zhu X; Huang Y; Zhao LX; Jiang Y; Van Nieuwerburgh F; Rader C; Duan Y; Shen B MBio 2016, 7, e02104–02116. [PubMed: 27999165]

(56). Yan X; Chen J-J; Adhikari A; Yang D; Crnovcic I; Wang N; Chang C-Y; Rader C; Shen B Org. Lett 2017, 19, 6192–6195. [PubMed: 29086572]

(57). Rudolf JD; Yan X; Shen BJ Ind. Microbiol. Biotechnol 2016, 43, 261–276.

(58). Shen B; Hindra Yan, Huang X, Ge T, Yang H, Teng D, Rudolf Q, Lohman JD, J. R. Bioorg. Med. Chem. Lett 2015, 25, 9–15. [PubMed: 25434000]

(59). Kang Q; Shen Y; Bai L Nat. Prod. Rep 2012, 29, 243–263. [PubMed: 22193711]

(60). Floss HG; Yu TW; Arakawa KJ Antibiot 2011, 64, 35–44.

(61). Bevitt DJ; Cortes J; Haydock SF; Leadlay PF Eur. J. Biochem 1992, 204, 39–49. [PubMed: 1740151]

(62). Kakavas SJ; Katz L; Stassi DJ Bacteriol 1997, 179, 7515–7522.

(63). Reid R; Piagentini M; Rodriguez E; Ashley G; Viswanathan N; Carney J; Santi DV; Hutchinson CR; McDaniel R Biochemistry 2003, 42, 72–79. [PubMed: 12515540]

(64). Valenzano CR; You YO; Garg A; Keatinge-Clay A; Khosla C; Cane DE J. Am. Chem. Soc 2010, 132, 14697–14699. [PubMed: 20925342]

(65). Zheng J; Keatinge-Clay AT J. Mol. Biol 2011, 410, 105–117. [PubMed: 21570406]

(66). Keatinge-Clay AT Nat. Prod. Rep 2012, 29, 1050–1073. [PubMed: 22858605]

(67). Lin S; Van Lanen SG; Shen B Proc. Natl. Acad. Sci. U.S.A 2009, 106, 4183–4188. [PubMed: 19246381]

(68). Lin S; Huang T; Horsman GP; Huang SX; Guo X; Shen B Org. Lett 2012, 14, 2300–2303. [PubMed: 22519717]

(69). Keating TA; Marshall CG; Walsh CT Biochemistry 2000, 39, 15513–15521. [PubMed: 11112537]

(70). Wang HX; Chen YY; Ge L; Fang TT; Meng J; Liu Z; Fang XY; Ni S; Lin C; Wu YY; Wang ML; Shi NN; He HG; Hong K; Shen YM J. Appl. Microbiol 2013, 115, 77–85. [PubMed: 23594089]

(71). Kieser T; Bibb MJ; Buttner MJ; Chater KF; Hopwood DA, Practical Streptomyces Genetics. The John Innes Foundation: Norwich, England, 2000.

(72). Bierman M; Logan R; O'Brien K; Seno ET; Rao RN; Schoner BE Gene 1992, 116, 43–49. [PubMed: 1628843]

(73). Erb A; Luzhetskyy A; Hardter U; Bechthold A Chembiochem 2009, 10, 1392–1401. [PubMed: 19384899]

(74). Bentley SD; Chater KF; Cerdeno-Tarraga AM; Challis GL; Thomson NR; James KD; Harris DE; Quail MA; Kieser H; Harper D; Bateman A; Brown S; Chandra G; Chen CW; Collins M; Cronin A; Fraser A; Goble A; Hidalgo J; Hornsby T; Howarth S; Huang CH; Kieser T; Larke L; Murphy L; Oliver K; O'Neil S; Rabbinowitsch E; Rajandream MA; Rutherford K; Rutter S; Seeger K; Saunders D; Sharp S; Squares R; Squares S; Taylor K; Warren T; Wietzorrek A; Woodward J; Barrell BG; Parkhill J; Hopwood DA Nature 2002, 417, 141–147. [PubMed: 12000953]

(75). Amos GCA; Awakawa T; Tuttle RN; Letzel AC; Kim MC; Kudo Y; Fenical W; Moore BS; Jensen PR Proc. Natl. Acad. Sci. U.S.A 2017, 114, E11121–E11130. [PubMed: 29229817]

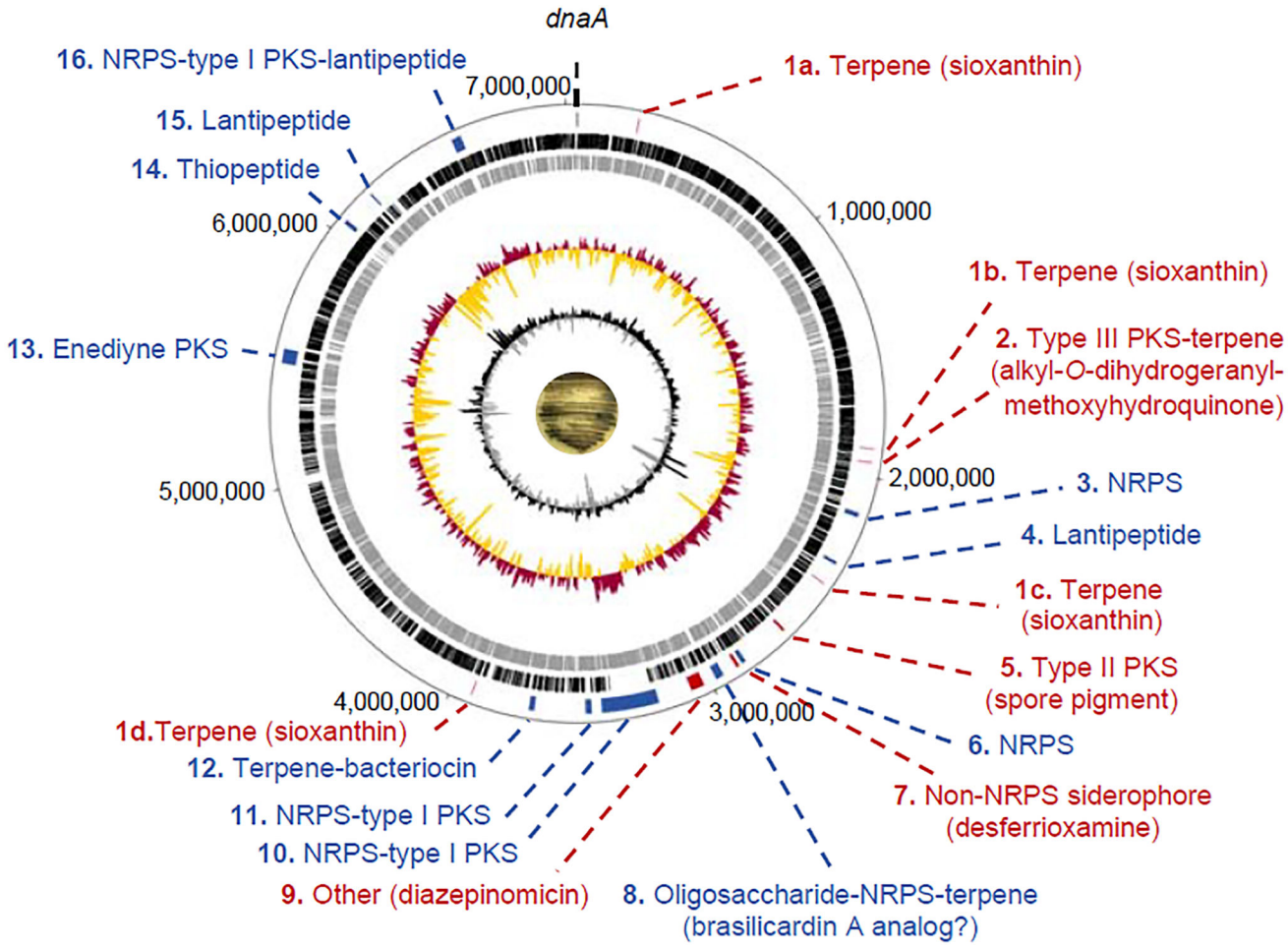


Figure 1. Chromosome map of *Micromonospora* sp. strain B006.

Lane 1 (from the outside in) shows the distribution of biosynthetic gene clusters (BGCs). Known BGCs are shown in red, whereas orphan BGCs are shown in blue (see Table 3 for further details). Lanes 2 and 3 show predicted open reading frames (ORFs) on the leading (black) and lagging (gray) strands, respectively. Lanes 4 and 5 depict a normalized plot of guanosine + cytosine (G+C) content (yellow/purple) and a normalized plot of G+C skew (black/grey), respectively. This representation is oriented to *dnaA*. The outer circle depicts the size in base pairs. The map was generated using DNAPlotter.

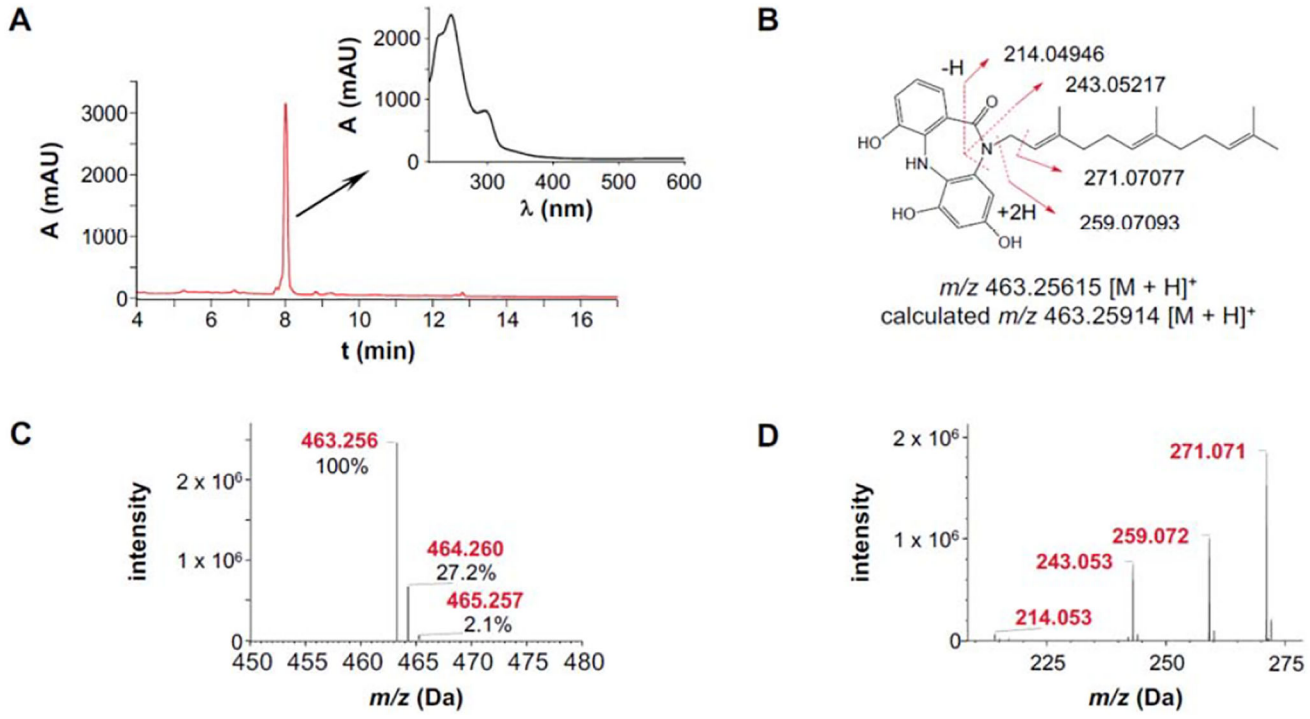


Figure 2. Diazepinomicin production by *Micromonospora* sp. B006.

(A) High performance liquid chromatography analysis of crude extract from a culture grown in modified Bennett's medium, with detection at 240 nm and showing the UV-Visible spectrum. (B) Structure of diazepinomicin and exact mass and fragments reported by Charan and colleagues. (C) *m/z* for $[M + H]^+$ and (D) fragmentation pattern obtained by Ion Trap Time of Flight Mass Spectrometry (positive mode).

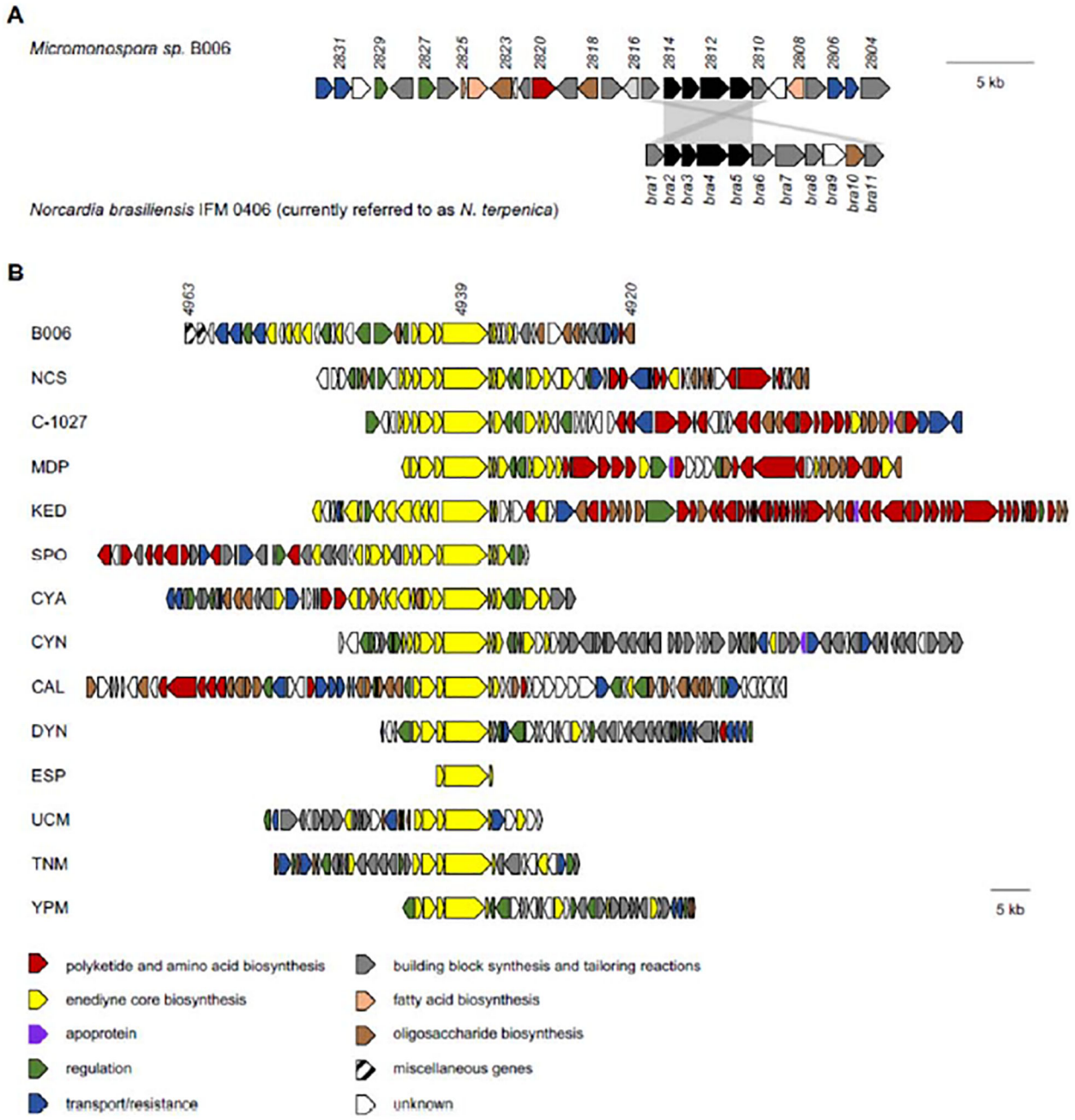


Figure 3. Two orphan gene clusters putatively encoding a brasilicardin A analogue and an enediyne, respectively.

(A) Comparison of cluster 8 with the brasilicardin A gene cluster in *Norcadia brasiliensis* IFM 0406. Homologous genes are connected by gray areas. (B) Comparison of enediyne PKS cluster 13 with previously reported enediyne gene clusters. NCS, neocarzinostatin from *Streptomyces carzinostaticus*. C-1027, C-1027 from *Streptomyces globisporus*. MDP, maduropeptin from *Actinomadura madurae*. KED, kedarcidin from *Streptoalloteichus* sp. SPO, sporolides from *Salinispora tropica*. CYA, cyanosporasides from *Salinispora pacifica* CNS-143. CYN, cyanosporasides from *Streptomyces* sp. CNT-179. CAL, calicheamicins from *Micromonospora echinospora*. DYN, dynemycin from *Micromonospora chersina*. ESP

(partial), esperamicins from *Actinomadura verrucasospora*. UCM, uncialamycin from *Streptomyces uncialis*. TNM, tiancimycin from *Streptomyces* sp. CB03234. YPM, yangpumicin from *Micromonospora yangpuensis* DSM 45577.

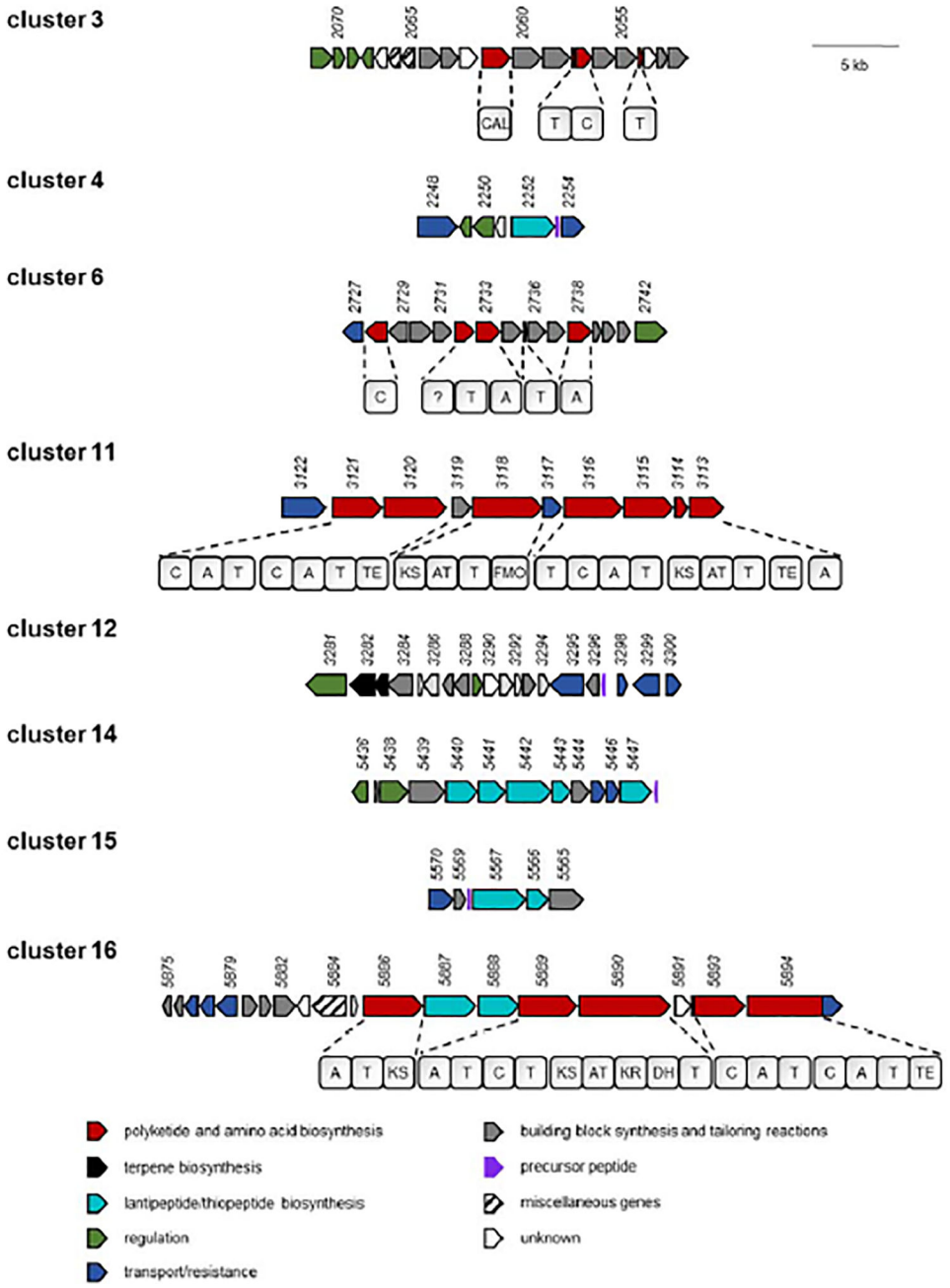


Figure 4. NRPS-PKS and RiPP orphan gene clusters identified in *Micromonospora* sp. B006.

Clusters 3, 6, 11, and 16 are associated with thiotemplate-based assembly (PKS and/or NRPS). The schematic representation of PKS and NRPS domain architecture is shown below the genes. Clusters 4, 14, and 15 are associated with RiPP biosynthesis. Cluster 12 is associated with terpene-RiPP biosynthesis. Domain key: A, adenylation; AT, acyltransferase; C, condensation; CAL, Co-A ligase; DH, dehydratase; FMO, Flavin-dependent monooxygenase; KR, ketoreductase; KS, ketosynthase; T, thiolation; TE, thioesterase.

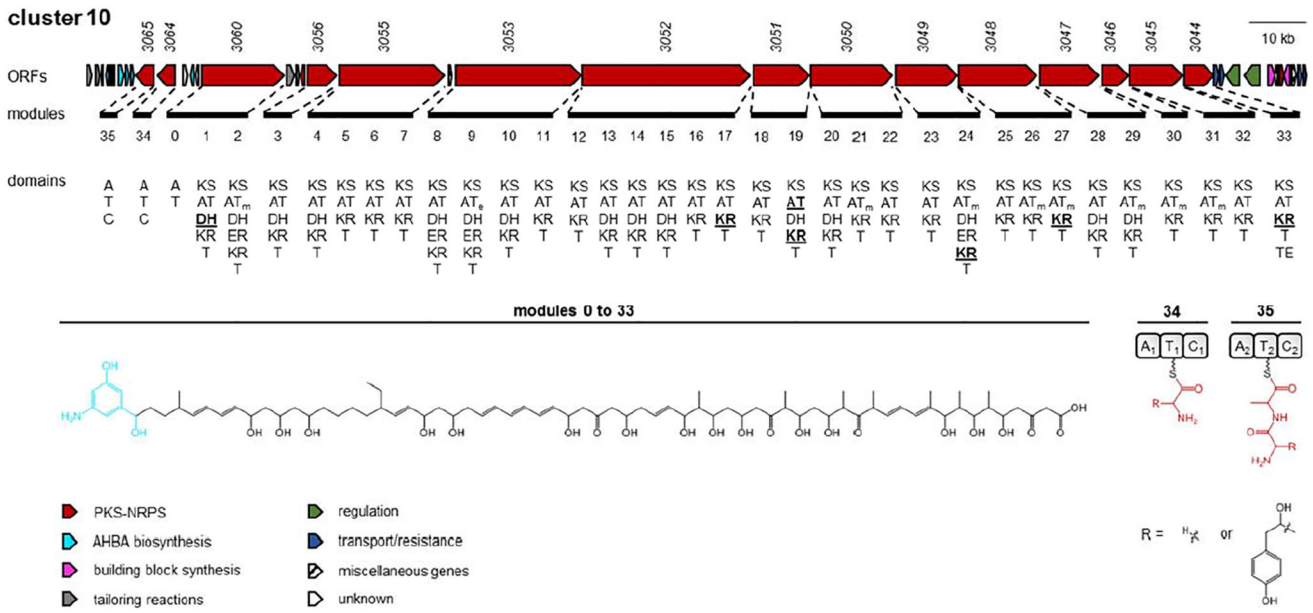


Figure 5. Structure prediction of natural product encoded in an exceptionally large NRPSPKS orphan gene cluster identified in *Micromonospora* sp. B006.

The 215.8-kb cluster 10 contains 42 ORFs including two monomodular NRPS genes, and 13 type I PKS genes, totaling 36 modules. Domain nomenclature is as follows. A, adenylation; T, thiolation; C, condensation; KS, ketosynthase; AT, acyltransferase predicted to load malonyl-CoA; ATm, AT predicted to load methylmalonyl-CoA; ATe, AT predicted to load ethylmalonyl-CoA; DH, dehydratase; KR, ketoreductase; ER, enoylreductase; TE, thioesterase. Underlined domains are predicted to be inactive.[–] Note that reductive domain redundancy seems to be prevalent in some AHBA-loading PKSs such as in rifamycin biosynthesis. The gene cluster also contains six genes encoding enzymes involved in 3-amino-5-hydroxybenzoic acid (AHBA) biosynthesis as identified through BLAST analysis (shown in light blue). Further modifications by tailoring enzymes encoded in the biosynthetic gene cluster (shown in dark gray), such as two glycosyl transferases, a flavin-dependent monooxygenase, a methyltransferase, and a metallophosphoesterase, are not shown.

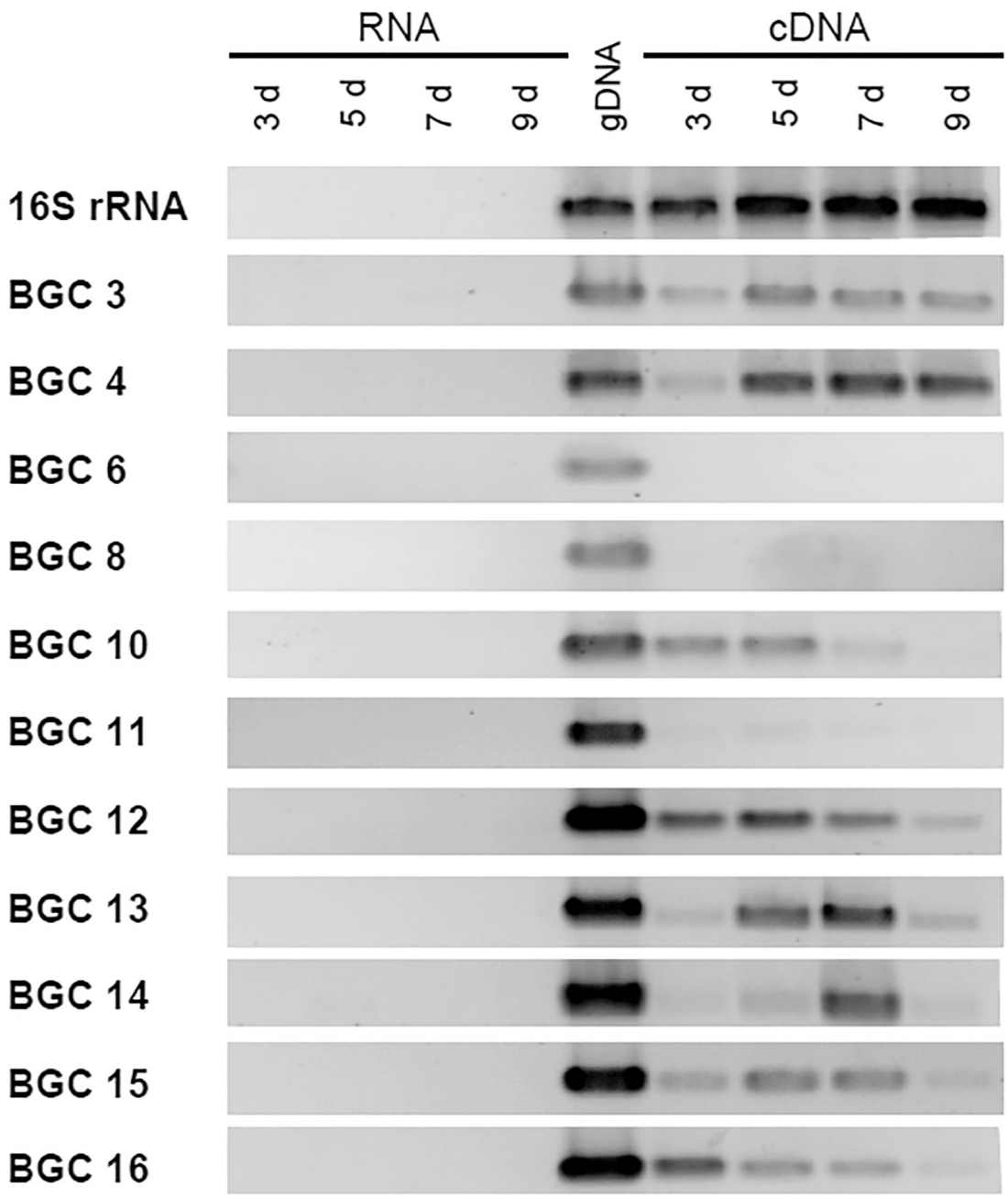


Figure 6. Transcriptional analysis of orphan biosynthetic gene clusters.

Micromonospora sp. B006 was grown in A1 liquid medium for nine days total. The expression of one representative gene per gene cluster was analyzed by reverse transcription-PCR (BGC 3, *MicB006_2061*; BGC 4, *MicB006_2552*; BCG 6, *MicB006_2733*; BGC 8, *MicB006_2812*; BGC 10, *MicB006_3059*; BGC 11, *MicB006_3117*; BGC 12, *MicB006_3283*; BGC 13, *MicB006_4942*; BGC 14, *MicB006_5442*; BGC 15, *MicB006_5566*; BGC 16, *MicB006_5887*) at four different time points (3, 5, 7, and 9 days). The house-keeping 16S rRNA gene was used as a positive control. RNA preparations before cDNA synthesis were used as template in negative control reactions.

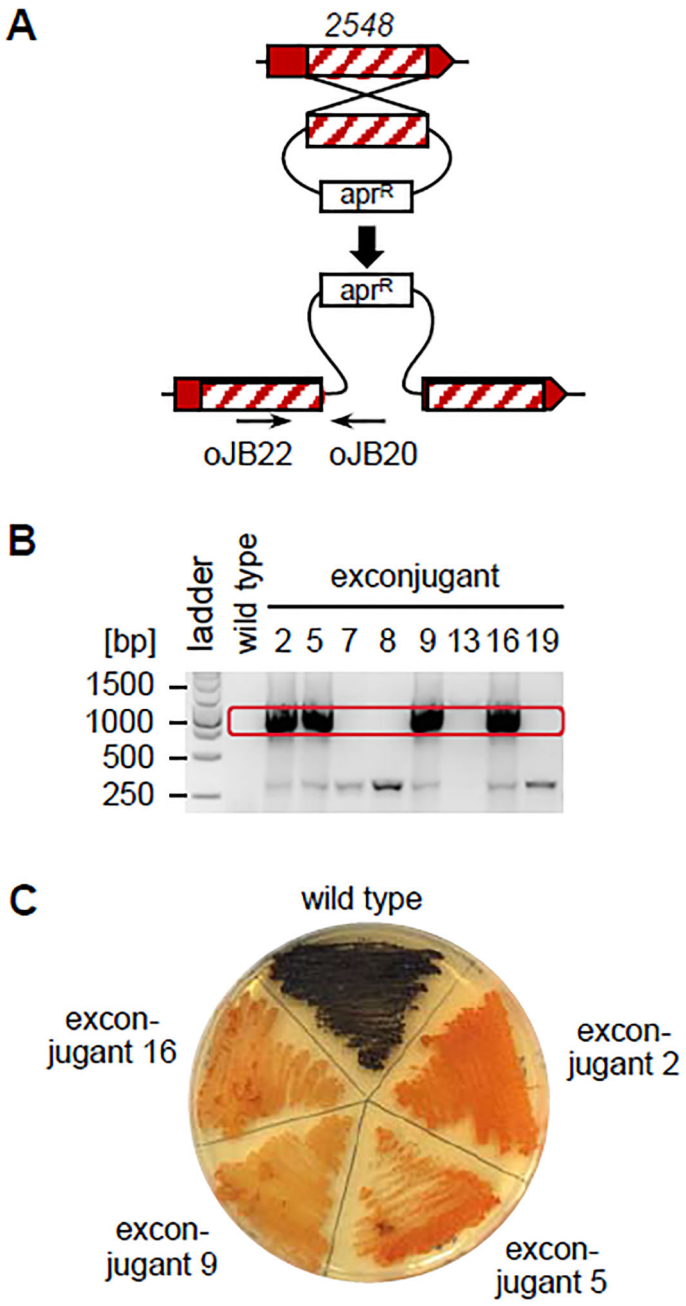


Figure 7. Inactivation of the type II PKS gene cluster encoding the spore pigment.

(A) Schematic presentation of the insertional inactivation of KSα gene *MicB006_2548* via a single crossover. The location of primers (oJB22 and oJB20) used to confirm mutants is indicated by inverted arrows. apr^R, apramycin resistance marker *aac3(IV)*. (B) Analysis of the exconjugants by PCR (expected fragment length for single crossover mutants: 1,105 bp). Genomic DNA from the wild-type strain was used as negative control. (C) Sporulation phenotyping of wild-type and single crossover mutants grown on A1 agar.

Table 1.

Summary statistics for assemblies from Illumina only and Illumina/Oxford Nanopore sequencing of the *Micromonospora* sp. B006 genome.

	Illumina only assembly	Illumina/Nanopore assembly
number of contigs	71	7
total length (bp)	6,987,783	7,037,487
shortest contig	1,045	46,834
largest contig	505,149	5,050,772
average coverage[a]	87.2	77.6
lowest coverage	69.5	71.5
highest coverage	226	84.9
N50 (bp)	208,576	5,050,772
N75 (bp)	123,779	678,460
N90 (bp)	71,577	402,230

[a]Coverage is average reads per base, calculated from an alignment of reads against the contigs using bowtie2 (Illumina data only). Average coverage is the unweighted average of the coverage of each contig separately.

Table 2.

General genome features of *Micromonospora* sp. B006.

Feature	
number of base pairs	7,040,153
contigs	1
G + C content (%)	72.8
protein coding genes	6,251
RNA genes	
rRNA genes	9
5S rRNA	3
16S rRNA	3
23S rRNA	3
tRNA genes	52

Table 3.

Biosynthetic gene clusters identified in the genome of *Micromonospora* sp. B006.

No.	Cluster location MicB006_	Cluster size [kb]	Type	Actual or predicted product[a]
1	0242 to 0246 1836 2343 3499 to 3501	11.3	terpene	sioxanthin
2	1884 to 1887	4.0	type III PKS-terpene	alkyl-*O*−dihydrogeranyl-methoxyhydroquinone
3	2051 to 2071	21.8	NRPS	
4	2248 to 2254	9.5	lantipeptide class III	
5	2547 to 2556	9.4	type II PKS	**spore pigment**
6	2727 to 2742	18.9	NRPS	
7	2753 to 2760	10.3	non-NRPS siderophore	desferrioxamine
8	2804 to 2832	32.4	oligosaccharide-NRPS-terpene	brasilicardin A analog
9	2883 to 2924	51.8	other	**diazepinomicin**
10	3033 to 3074	215.8	NRPS-type I PKS	
11	3113 to 3122	25.7	NRPS-type I PKS	
12	3281 to 3300	21.8	terpene-bacteriocin	
13	4920 to 4963	57.0	enediyne PKS	enediyne
14	5436 to 5448	17.8	thiopeptide	
15	5565 to 5570	9.0	lantipeptide class I	
16	5875 to 5894	39.3	NRPS-PKS-lantipeptide	

[a]Observed products are highlighted in bold.

BGCs were identified using antiSMASH and BLAST followed by manual curation.