Microbial Biotechnology

WILEY

MICROBIAL BIOTECHNOLOGY Open Access

RESEARCH ARTICLE OPEN ACCESS

Enabling Access to Novel Bacterial Biosynthetic Potential From ONT Draft Genomic Data

Marco A. Campos-Magaña[1,2] | Vitor A. P. Martins dos Santos[1,3] | Luis Garcia-Morales[3]

[1]Dept. Bioprocess Engineering, Wageningen University and Research, Wageningen, the Netherlands | [2]Dept. Systems and Synthetic Biology, Wageningen University and Research, Wageningen, the Netherlands | [3]Lifeglimmer GmbH, Berlin, Germany

Correspondence: Vitor A. P. Martins dos Santos (vitor.martinsdossantos@wur.nl) | Luis Garcia-Morales (luis.garcia.morales@merck.com)

Received: 30 April 2024 | **Revised:** 19 December 2024 | **Accepted:** 25 January 2025

Funding: This study was supported by Consejo Nacional de Ciencia y Tecnología. BIOS project, grant agreement no. 101070281.

Keywords: bioinformatics | microbial genomics | natural products | secondary metabolites | synthetic biology

ABSTRACT

Natural products comprise a wide diversity of compounds with a range of biological activities, including antibiotics, anti-inflammatory and anti-tumoral molecules. However, we can only access a small portion of these compounds due to various technical difficulties. We have herein developed a novel and efficient approach for accessing biosynthetic gene clusters (BGCs) that encode natural products from soil bacteria. The pipeline uses a combination of long-read sequencing, antiSMASH for BGC identification and Transformation-associated recombination (TAR) for cloning the BGCs. We hypothesized that a genome assembly using Oxford Nanopore Technology (ONT) sequencing could facilitate the detection of large BGCs at a relatively fast and low-cost DNA sequencing. Despite the relative low accuracy and sequence mistakes due to high GC content and sequence repetitions frequently found in BGC containing bacteria, we demonstrate that ONT long-read sequencing and antiSMASH are effective for identifying novel BGCs and enabling TAR cloning to isolate the BGC in a desired vector. We applied this pipeline on a previously non-sequenced myxobacteria *Aetherobacter fasciculatus* SBSr002. Our approach enabled us to clone a previously unknown BGC into a genome engineering-ready vector, illustrating the capabilities of this powerful and cost-effective strategy.

1 | Introduction

Microbial natural products are secondary metabolites with many diverse cellular and ecological functions. They have been exploited for a variety of therapeutical and chemical industry applications. Natural sources of these valuable compounds are usually found in soil and marine bacteria. However, culturing some of these bacteria in standard laboratory conditions is often challenging given the complex growth requirements and specific conditions needed, which complicates interrogating these sources for bioprospection of natural compounds and heterologous expression (Wang, Nielsen, et al. 2021). An important step in natural product research for drug discovery is the identification of genes responsible for the biosynthesis of secondary metabolites. These genes generally form clusters in the genome (hence the name Biosynthetic Gene Clusters (BGCs)), spanning from 10 kb up to hundreds of kilobases (kbs). BGCs encode a variety of proteins with diverse functions including product biosynthesis, regulation, transport and resistance to the final product, among others. The progress in genome sequencing technologies and genome mining of BGCs has facilitated the identification of secondary metabolites. However, there is still a limited number of genomes available from bacteria rich in secondary metabolites, such as myxo- and actinobacteria (Baltz 2021). These non-pathogenic microorganisms possess extremely large genomes in comparison to most other bacteria. As an example, the myxobacterium *Sorangium celullosum* has a genome of 14.7 Mb (Han et al. 2013), as compared to *E.coli* genome of about 5 Mb (Rode et al. 1999). This immense genome encodes numerous unique

biosynthetic pathways, featuring at least 35 recognised BGCs (Yue et al. 2023). Apart from large genomes, other factors discouraging researchers to undertaking complex sequencing projects include a reduced number of teams working natural products, strains not available in depositories and their slow growth. Myxobacteria are gram-negative fruiting bacteria that can be found in different environments, including soil, plant debris and marine settings. They exhibit complex social behaviour, such as cell movement by gliding, biofilm formation, fruiting bodies formation and microbial predation (Whitworth et al. 2021). Many of them are capable of producing antimicrobial metabolites that might play an ecological role, but whose specific function in nature remains unknown. The range of secondary metabolites produced by myxobacteria include alkaloids, terpenes, polyketides, ribosomally synthetized and post-translationally modified peptides (RiPPs), as well as nonribosomal peptides and their hybrids conjugates, such as polyketide-nonribosomal peptides, which are produced by specialised synthases (Pal et al. 2021). These synthases are at the heart of the biosynthetic machinery responsible for the production of these chemically diverse specialised metabolites (Hemmerling and Piel 2022). In addition, their core biosynthetic enzymes are often highly conserved, modular and characterised by high amino acid sequence similarity (Scherlach and Hertweck 2021). However, they use different building blocks and thus, they can produce a wide variety of chemical structures. This surprising chemical diversity is also attributed to the decoration enzymes expressed along with the main synthetases. These tailoring enzymes often allow the introduction of unique chemical groups to the final product (Lee et al. 2020).

Heterologous expression of BGCs in common laboratory microorganisms have been proposed as an alternative to engineering the complex and poorly genetically amenable myxobacteria (Huo et al. 2019; Eckert et al. 2023; Gao et al. 2023; Seyfert et al. 2023; Wang et al. 2023). However, to enable this, a number of bottlenecks needs to be overcome. These include: (i) sequencing of large, high GC content and repetitive genomes; (ii) BGCs identification, (iii) cloning of the large genomic fragments; (iv) refactoring or insertion of appropriate gene expression elements of the target species; (v) transfer into relevant target expression hosts; (vi) heterologous production; (vii) isolation and; (viii) characterisation of the novel compound. Recent advances have provided the basis to tackle some of these hurdles. For instance, methods have been developed for identification of BGCs (antiSMASH, Blin et al. 2023), CRISPR-based technologies have immensely facilitated the refactoring of large gene clusters (Montiel et al. 2015; Kim et al. 2019), and methodologies such as chassis-independent recombinase-assisted genome engineering or CRAGE (Wang et al. 2019) allow the easy transfer to a multitude of commonly used bacterial hosts. While some current methods used for heterologous expression, such as *de novo* DNA synthesis, need highly accurate whole-genome sequence of the BGCs (Waldron et al. 2001; Miao et al. 2005; Lee et al. 2020), other methods such as TAR cloning (as used in this work) or CATCH can be used with draft genomic sequences as they only need small oligonucleotide sequences for PCR primer or gRNA design, respectively. Draft genomes are nowadays easy and inexpensive to obtain using Oxford Nanopore Technology (ONT). In scalability terms, sequencing a number of very large and highly repetitive genomes with high GC content in a highly accurate way, which involves short- and long-read sequencing, is costly.

TAR cloning has also been used as an alternative to de novo DNA synthesis to capture natural product gene clusters (Kim et al. 2019; Zhang et al. 2019; Kenshole et al. 2021; Wang, Zheng, et al. 2021) in a more cost-effective way. We propose that this technology could also be used to precisely clone BGCs found in relatively poorly-accurate Oxford Nanopore draft genomes in vectors, despite the limitations of this sequencing technology and the difficulties at sequencing these genomes. To demonstrate the capabilities of this pipeline (Figure 1), we used ONT sequencing to obtain a draft genome of the unsequenced myxobacteria *Aetherobacter fasciculatus* SBSr002. We were able to identify 27 BGCs and selected a biosynthetic cluster for a novel lanthipeptide—a class of RiPPs that remains barely explored in myxobacteria and whose known representatives encode for antimicrobial agents—for TAR cloning into a CRAGE-ready vector. Once biosynthesis genes are captured using yeast-based TAR cloning, it could be possible to use multiple technologies for rapid and efficient BGC refactoring, such as CRISPR-Cas9 (Kang et al. 2016; Kim et al. 2018; Gu et al. 2023) and transfer them to a plethora of common heterologous producing bacteria by CRAGE (Wang et al. 2019). We focus this study on the scalability of BGCs discovery and cloning of one uncharacterized

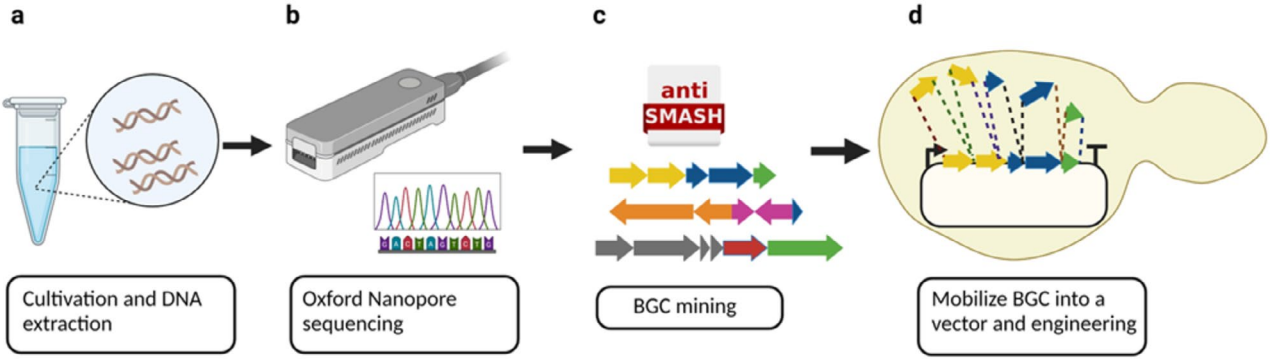

FIGURE 1 | Pipeline for the discovery and accession of novel BGCs from myxobacteria. (a) Cultivation of bacterial strain and DNA extraction. (b) Oxford Nanopore sequencing of the DNA extracted. (c) Identification of BGCs through genome mining tools such as antiSMASH. (d) Mobilisation of the BGC to a vector using TAR in yeast and refactoring for the insertion of more suitable expression regulatory elements.

lanthipeptide BGC from ONT draft genomic data. Our results show a > 99% similarity between the ONT sequencing result of the uncharacterized BGC and sequence validation with Sanger sequencing. Such a streamlined, rapid, and low-cost approach applied to the large collections of non-sequenced bacteria will enable exploration of BGCs in microbes. Specifically, it will enable the identification and capture of novel biosynthetic pathways, one of the initial steps to interrogate these biosynthesis genes for the discovery of novel bacterial secondary metabolites, which is important for accelerating drug development that our society increasingly needs to tackle problematics such as the antimicrobial resistance crisis.

2 | Methods

Aetherobacter fasciculatus SBSr002 was originally obtained from Leibniz-Institute DSMZ-German Collection of Microorganisms and Cell Cultures as live cultures on solid media. Subsequently, all strains were grown at 30°C in either VY/2 media, as liquid (200–250 rpm, vented cap Greiner tubes) or solid cultures. These strains were subcultured (at least) every 3 weeks due to their slow growth and the liquid medium was changed weekly with fresh medium.

2.1 | Isolation of Genomic DNA From Myxobacteria and Whole-Genome Sequencing

Genomic DNA was isolated from cultures of *A. fasciculatus* SBSr002, using the GenEluteTM Bacterial Genomic DNA Kit for Gram-negative (Sigma-Aldrich) and following the manufacturer instructions.

Agentcourt AMPure XP beads (Beckman Coulter) were used to concentrate and purify 100 μL of genomic DNA down to 10 μL, with 180 μL of beads (1.8× proportion). Preparation of the library was performed using the Rapid Barcoding Kit (SQK-RBK004, Oxford Nanopore Technologies) following the manufacturer instructions. The resulting DNA library was sequenced on the MinION Flow Cell Type R9 (Oxford Nanopore Technologies).

2.2 | Genome Assembly

Fast5 raw data files generated by the ONT MinION were basecalled, demultiplexed and trimmed using the data processing toolkit Guppy version 4.4.1. Flye genome assembler (https://github.com/mikolmogorov/Flye) (Kolmogorov et al. 2019) was used to assemble the ONT data into contigs. Medaka polisher was used to polish or correct base-calling errors in the sequences generated by ONT (https://github.com/nanoporetech/medaka). We uploaded our draft genome assembly or contigs to antiSMASH 7.1(https://github.com/antismash/antismash/releases) (Blin et al. 2023) with a relaxed strictness of detection to facilitate uncovering RIPPs. This version of antiSMASH includes a RRE-containing detection rule that enables the identification of potentially novel RiPPs clusters. In addition, all extra features during the analysis were selected on. The flowchart of the computational pipeline can be found in Figure S1. We used checkM (https://github.com/Ecogenomics/CheckM) (Parks et al. 2015) and BUSCO (https://gitlab.com/ezlab/busco) (Simão et al. 2015) to assess the quality and completeness of the assembled genome, respectively. Individual biosynthetic genes sequences identified by antiSMASH were uploaded to Pfam (Mistry et al. 2020) and BLAST (Camacho et al. 2009) to predict the function of the closest homologue based on similarity between sequences. The assembled genome was annotated using Bakta v1.9.3 (https://github.com/oschwengers/bakta) (Schwengers et al. 2021) and COG classifier tool (https://github.com/moshi4/COGclassifier) was used for the classification of gene functions into categories (Shimoyama, 2022).

2.3 | TAR Cloning of a Novel *A. fasciculatus* BGC in a Shuttle Vector

In vivo BGC assembly in *S. cerevisiae* W303 was performed using 5-kb amplicons approximately with 60 bp homology ends (Kouprina and Larionov 2008; Yamanaka et al. 2014). One microgram of linearized pCC1FOS-oriT (this is an in-house derivative of pCC1FOS single copy vector from EPICENTRE Biotechnologies in which we introduced the oriT to facilitate conjugation) by PCR amplification was cotransformed with 300 ng of the CEN/ARS-GmR cassette amplified from the plasmid pRS314, and 300 ng of CRE-LEU2 cassette, LEU2 amplified from the plasmid pRS315 and CRE sequence retrieved from iGEM parts (http://parts.igem.org/Main_Page) with specific overhangs to capture the *A. fasciculatus* BGC 5-kb amplicons. *S. cerevisiae* W303 (Table S1) transformants were screened using colony PCR with the screening primers listed in Table S2. *S. cerevisiae* W303 total DNA was extracted using the protocol from (Kouprina and Larionov 2008), and transformed into TransforMax EPI300 *E. coli* cells. Isolated plasmids from *E. coli* EPI300 were verified using whole plasmid sequencing (Plasmidsaurus inc.). The final vector contained the Cre recombinase and the lox sites flanking the Gentamycin resistance gene and the BGC of interest for genome integration in the heterologous host. Furthermore, Sanger sequencing (Macrogen inc.) was used to verify the plasmid sequence and compare the complete sequence of Lanthipeptide class-l BGC (amplicons 1–5) obtained from the genomic DNA of *A. fasciculatus* SBSr002 with the ONT sequence. Sequence of primers for Sanger sequencing of the complete lanthipeptide BGC are found in the Table S2.

3 | Results

3.1 | Genome Analysis

Myxobacteria are widely recognised for their large genome size and capacity to encode diverse BGCs secondary metabolites. In this research, the genome of *Aetherobacter fasciculatus SBSr002* was sequenced by ONT, assembled using Flye and polished with Medaka software resulting in a total length of 13.5 Mbp with 69.27% GC (Table 1). This represents one of the largest bacterial genomes sequenced to date (Table 2). The bacterial genome has been deposited in European Nucleotide Archive under the accession PRJEB72099. Bakta annotation tool (Schwengers et al. 2021) predicted a total of 18,656 genes that consist of 16,041 hypothetical proteins and 103 RNA genes (82 tRNAs, 1 transfer-messenger RNA, 9 rRNAs, 7 non-coding RNAs) (Table 3). In

addition, a genome completeness of 72.20% by checkM and complete BUSCOs of 40% were obtained (Tables S3 and S4).

3.2 | Identification of Biosynthetic Gene Clusters

The draft genome of *A. fasciculatus* SBSr002 was mined for putative BGCs using antiSMASH. This software uses a rule-based approach to predict different biosynthetic pathways associated to secondary metabolites production (Blin et al. 2023). In total, 27 BGCs were identified in the draft genome (Table 4). The number of BGCs detected here suggests this myxobacteria is a rich source of secondary metabolites. Analysis by antiSMASH allowed us to identify genes responsible for the biosynthesis of different natural product classes including one lanthipeptide-class-I, four RiPP-like, six terpenes, two hserlactona, two NRPS-like, three NRPS, 1 hybrid NRPS-RiPP-like, one arylpolyene, one phenazine, two RRE-containing, one T3PKS, one hglE-KS terpene, and one indole (Table 4). From the antiSMASH list, 23

TABLE 1 | Quality assessment of the assembled genome of *A. fasciculatus* SBSr002 obtained from Flye.

Attribute	Flye	Medaka polishing
Total length (Mbp)	13.5	13.5
N50 (nt)	3,121,864	3,125,783
Largest contig (nt)	3,648,973	3,653,705
Number of contigs	18	17
GC%	69.2	69.3
Number of BGCs	27	27
Coding density	79.2	80.6

TABLE 3 | Bakta analysis for annotation of the *A. fasciculatus* SBSr002 genome.

Feature type	Flye	Medaka_flye
tRNAs	82	82
tmRNAs	1	1
rRNAs	9	9
ncRNAs	6	7
ncRNA regions	3	3
CRISPR arrays	1	0
Total CDSs	20,581	18,656
pseudogenes	237	240
hypotheticals	18,523	16,041
signal peptides	0	0
sORFs	0	0
gaps	0	0
oriCs	0	0
oriVs	0	0
oriTs	0	0

TABLE 2 | Comparison of myxobacteria species.

Myxobacteria	Total length (Mb)	Genome assembly	GC (%)	BGCs	Reference
Minicystis rosea DSM 24000	16	Complete	69	40	Yue et al. (2023)
Sorangium cellulosum So0157–2	14.8	Complete	72.1	35	Han et al. (2013)
Sorangium cellulosum Soce836	14.6	Complete	72	42	Yue et al. (2023)
Sorangium cellulosum Soce26	14.56	Complete	71.7	36	Yue et al. (2023)
Aetherobacter fasciculatus SBSr002	13.5	Draft	69.3	27	In this work
Pyxidicoccus sp. SCPEA002	13.2	Complete	69.6	32	(Yue et al. 2023)
Sorangium cellulosum So ce56	13	Complete	71.4	35	Schneiker et al. (2007)
Myxococcus eversor AM011	11.6	Complete	68.9	—	Whitworth et al. (2021)
Stigmatella aurantica DW4/3–1	10.3	Complete	67.4	—	Huntley et al. (2010)
Corallococcus coralloides DSM 2259	10.1	Complete			Huntley et al. (2012)
Myxococcus vastator AM001	9.8	Complete	68.8	—	Whitworth et al. (2021)
Stigmatella erecta	9.8	—			Kaiser et al. (2010)
Myxococcus hansupus	9.5	Complete			Sharma et al. (2016)
Haliangium ochraceum SMP-2	9.4	Complete	69.5	—	Ivanova et al. (2010)
Myxococcus xanthus DK1622	9.1	Complete	68.9	22	Goldman et al. (2006)
Myxococcus macrosporus DSM 14697	8.9	Complete	70.6	—	Treuner-Lange et al. (2017)

TABLE 4 | Identified Biosynthetic Gene Clusters using antiSMASH.

Type	Contig	Most similar known cluster	Similarity (%)
lanthipeptide-class-i	1		
RiPP-like	1	ambruticin	5
hserlactone	1		
terpene	12	carotenoid	100
NRPS-like	12	crocacin	52
hglE-KS	2	eicosapentaenoic acid-like compound	9
arylpolyene	2		
RiPP-like	2		
terpene	2	geosmin	100
NRPS	2	vioprolide	33
phenazine	2		
NRPS-like	3		
RRE-containing	6		
terpene	6		
terpene	6	geosmin	100
hglE-KS, terpene	6	sarasinoside	18
hserlactone	6		
terpene	6		
LAP, RRE-containing	6		
T3PKS	6	toyoncin	18
RiPP-like	7		
NRPS	7	heinamide	36
NRPS, NRPS-like, RiPP-like	7	vacidobactin A/vacidobactin B	41
indole	8		
NRPS	8	vioprolide	33
terpene	8		
RiPP-like	9		

predicted BGCs were novel clusters. Three terpenes BGCs were 100% similar to known clusters including geosmin and a carotenoid. A BLAST search indicated that the hglE-KS gene cluster detected in contig 2 matched 99.2% to the polyunsaturated fatty acids (PUFAs) cluster (Gemperlein et al. 2014) from *A. fasciculatus* SBSr002 (GenBank: KF977699.1). This BGC served as a positive control for this analysis since it has been deposited and showed a high similarity between these two sequences, demonstrating a high accuracy to detect this type of secondary metabolites using only ONT.

As a comparative analysis, we used other bioinformatic tools to detect BGCs in the draft genome assembly obtained by ONT. However, the number of predicted BGCs varied substantially depending on the tool. For example, NP.searcher (Li et al. 2009) only found three NRPS, four mixed NRPS/PKS, and four mevalonate terpenoid mva genes. PRISM (Skinnider et al. 2017) identified two mixed modular NRPS/PKS, two lanthipeptide class ll clusters, five polyketides, and two nonribosomal peptides. Regarding specialised tools for RiPPs, BAGEL (Van Heel et al. 2018) predicted six different ribosomal peptides including three sactipeptides, two lanthipeptide class-ll, and one bacteriocin. On the other hand, RiPPMiner-Genome (Agrawal et al. 2021) detected one Linear azole(in)e-containing peptides (LAPS) and three lanthipeptides.

We chose a novel identified Lanthipeptide class-1 cluster (Figure 2C) to assess the feasibility of using an Oxford Nanopore draft genome assembly to predict BGCs and capture new biosynthetic pathways by using TAR assembly in yeast.

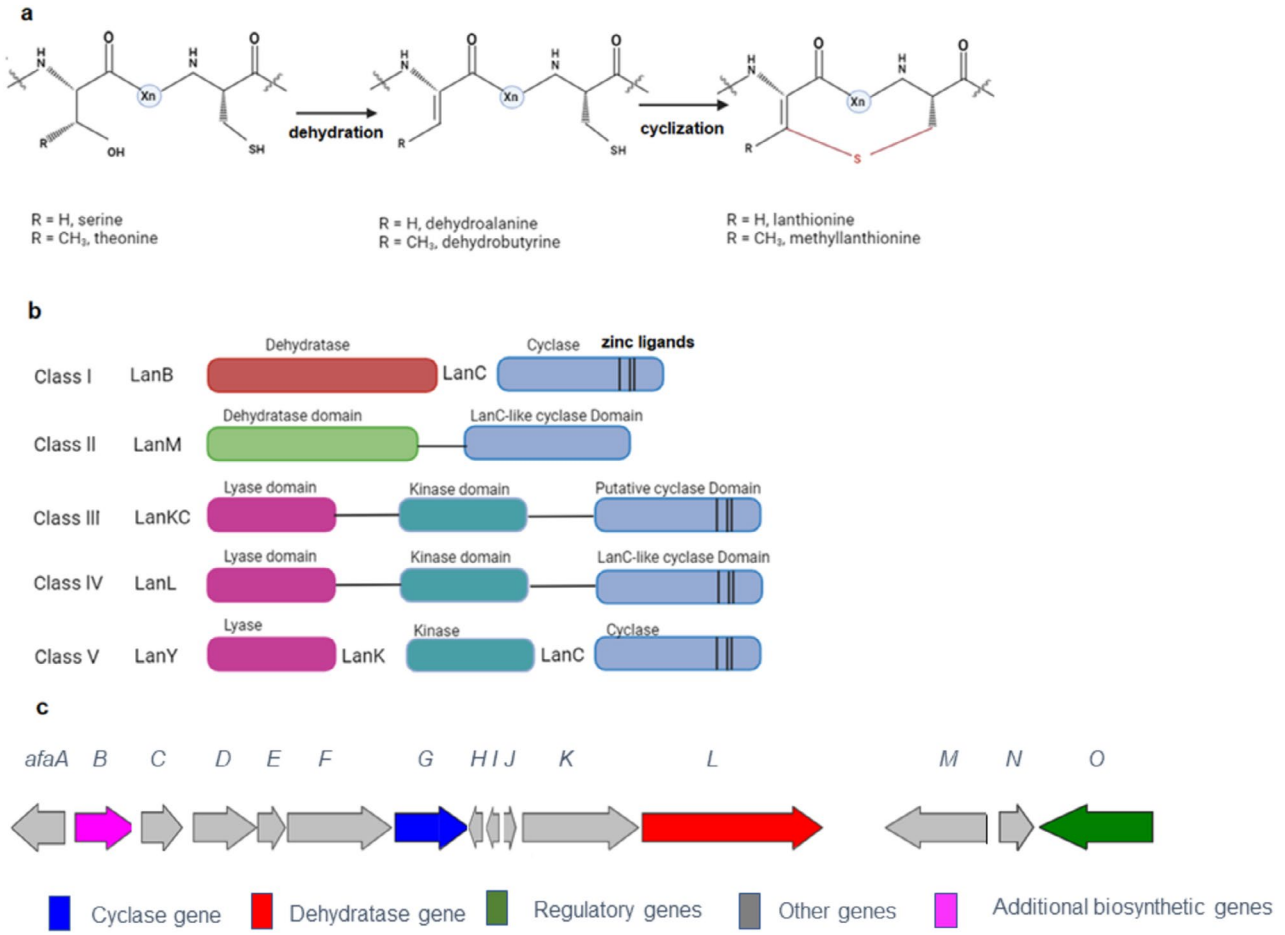

FIGURE 2 | Biosynthesis of lanthipeptides. (a) Biosynthesis of Lanthipeptides begins with the dehydration of Ser/Thr residues in the starting peptide that results in dehydroalanine or dehydrobutyrine followed by the generation of a thioether cross-link. (b) Domain organisation of the five common types of lanthipeptide synthetases. (c) BGC of the novel Lanthipeptide class l detected in *A. fasciculatus* SBSr002. *Afa A-O* genes are indicated in the BGC.

3.3 | Lanthipeptide Biosynthetic Gene Cluster Analysis

We selected a novel lanthipeptide class-1 BGC for cloning with 0% similarity to other known clusters. We named this BGC *afa*, as the initials of *A. fasciculatus*. This BGC can be found on GenBank: contig 1 NZ_CAWUQY010000001.1 at bps 1,423,624–1,449,748 bps. The production of lanthipeptides involves a two-step chemical reaction in which the dehydration of Ser/Thr residues within the peptide substrate results in the formation of an α, β-unsaturated dehydroamino acid. Then, conjugate addition onto the dehydroamino acid is catalysed by cyclases (Figure 2a). Lanthipeptides can be classified into five different classes depending on their domain organisation. Representative biosynthetic enzymes for lanthipeptides class-l (Figure 2b) were identified in the *afa* BGC (Cyclase and dehydratase domains, Figure 2c). This 26 kb-lanthipeptide BGC harbours 15 putative genes that includes main biosynthetic lanthipeptide class-l genes (*afaG* and *afaL*, Figure 2c). We predicted one Lanthipeptide dehydratase (*afaL*) of 978 a.a. (Table 5). It consists of a N-terminal domain of, responsible for the dehydration of Ser/Thr residues. This glutamylation domain (pfam accession PF04738) catalyses a transesterification reaction from glutamyl-tRNA^{Glu} to Ser/Thr. This lanthipeptide dehydratase also consists of a C-terminal elimination domain (PF14028), involved in the glutamate-elimination or β-elimination. Besides these two lanthipeptide dehydratase domains, we identified a Lanthionine synthetase (*afaG*) C-like protein (pfam accession PF05147) which is a cyclase enzyme of the lanthionine synthetase (Table 5). This protein controls cyclization of class I lanthipeptides by utilising a zinc ion to reduce the pKa of the Cys sulfur allowing the conjugate addition of the thiolate onto the dehydroamino acid (Pei et al. 2022).

We predicted two accessory post-translational modification (PTM) enzymes within this *afa* biosynthetic cluster using antiSMASH, BLAST and pfam. One M4 family metallopeptidase (*afaM*) and one carboxipeptidase (*afaF*). We also detected other presumably biosynthetic enzymes involved in the Lanthipeptide biosynthesis such as one F420-dependent NADP oxidoreductase (*afaB*, Table 5). Moreover, we predicted one sigma 54-interacting transcriptional regulator, one putative CapK protein, three hypothetical proteins, and four other putative genes for which no significant similarity was found. antiSMASH identified ten other different BGCs with similar biosynthetic genes to those found in the lanthipeptide BGC from *A. fasciculatus* SBSr002 (Table S6). The highest similarity was assigned to paenicidin A, paenicidin B and Geobacillin l, RiPP type secondary metabolites.

TABLE 5 | Predicted Function of ORFs in the lanthipeptide class-l BGC Locus of *A. fasciculatus* SBSr002.

Identifier	antiSMASH	Pfam	NCBI	% ID (NCBI)
AfaA	Other	—	Hypothetical protein	40.2
AfaB	Biosynthetic-additional	F420-dependent NADP reductase	NADP-dependent F420 reductase	89.7
AfaC	Other	—	No significant similarity found	—
AfaD	Other	—	Hypothetical protein	90.9
AfaE	Other	—	SUMF1/EgtB/PvdO family nonheme iron enzyme	28.0
AfaF	Other	—	Carboxypeptidase regulatory-like domain-containing protein family	57.1
AfaG	Biosynthetic	Lanthionine synthetase	Lanthionine synthetase LanC family protein	39.0
AfaH	Other	—	No significant similarity found	—
AfaI	Other	—	No significant similarity found	—
AfaJ	Other	—	Hypothetical protein	63.4
AfaK	Other	Capsular Polysaccharide Biosynthesis Protein CapK	Phenylacetate coA-ligase	69.8
AfaL	Biosynthetic	Lantibiotic dehydratase	Lanthionine biosynthesis protein LanB/ lantibiotic dehydratase	69.1/34.0
AfaM	Other	M4 family metallopeptidase	—	—
AfaN	Other	—	No significant similarity found	—
AfaO	Regulatory genes	Sigma factor 54 interaction domain	Sigma 54-interacting transcriptional regulator	52.7

Note: ORFS captured by TAR are named *AfaA-O*. Predicted function of the closest homologue using Pfam and BLAST search. The percentage of similarity is shown in % ID column for BLAST.

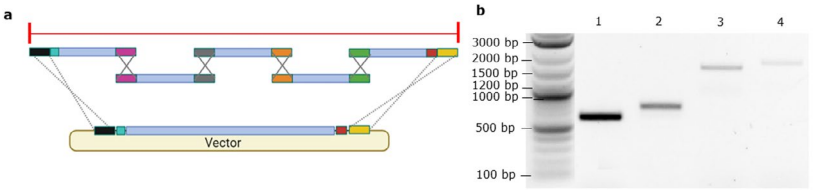

FIGURE 3 | TAR cloning of a new lanthipeptide BGC. (a) Lanthipeptide Biosynthetic Gene Cluster assembly of five lanthipeptide 5 kb-PCR fragments in a yeast shuttle vector using TAR cloning in yeast. (b) Gel electrophoresis of colony PCR for verification of assembled BGC fragments in the shuttle vector. Band in line 1, 2, 3, and 4 shows successful recombination of 60-bp homology ends from lanthipeptide 5 kb-amplicons marked in pink, grey, orange and green, respectively. The region sequenced using Sanger sequencing is marked in a red line in 3a. Sequencing alignment of complete lanthipeptide class-l BGC between ONT and Sanger sequencing can be found in Figure S6.

3.4 | TAR Cloning in Yeast

We directly captured this novel lanthipeptide BGC (*afa* genes) from *A. fasciculatus* SBSr002 using TAR cloning. We managed to amplify the 5 lanthipeptide BGC fragments of approximately 5 kb with 60-bp homology ends by using as reference the ONT sequencing result (Figure 3a). Colony PCR indicated the correct recombination between the different amplicons with 60-bp homology ends (Figure 3b). The final size of the assembled region was 25,722 bp. The number of positive colonies was 18%. Moreover, this vector already contained the Cre/lox sites for integration in the host chromosome.

3.5 | Pipeline Validation by Sanger Sequencing

After extracting the shuttle vector from *S. cerevisiae* W303 and amplifying it in *E. coli* EPI300 cells, we demonstrated that the complete lanthipeptide BGC was assembled as intended by whole plasmid sequencing using ONT (plasmidsaurus inc.). Furthermore, we performed Sanger sequencing of the entire gene cluster assembled in the vector. The alignment of Sanger sequencing with the original ONT sequencing showed 26 mismatches and 93 gaps/insertions from a total of 26 kb (99.5% identity, Figure S6), demonstrating that ONT sequencing has achieved high levels of accuracy. To ensure that these

mismatches are not mutations generated during the PCR amplifications and assembly process, we also performed Sanger sequencing on the genomic DNA and only 2 mismatches were found, all producing silent mutations. In all, the sequencing precision given by ONT allowed to obtain sufficient data to accurately assemble the whole gene cluster in a plasmid vector. The Sanger polished sequence of the BGC is available at ENA accession PRJEB72099.

4 | Discussion

Novel bioactive compounds are increasingly needed, for example, to counteract the current antimicrobial resistance crisis, which has been exacerbated by the lack of novel compounds beyond those already identified over the last decades by high-throughput screening (which are essentially the "lower hanging-fruit" ones). Access to sequencing data from rich producers of natural products such as myxobacteria enables us to identify novel BGCs that could be potentially interrogated for drug discovery. In fact, genome sequences from myxobacteria have been deposited in databases since 2006, when the genome of the reference myxobacterium, *Myxococcus xanthus* DK1622, was made accessible to the public. Since then, the number of released genomes has increased and by 2021, a total of 163 myxobacterial genomes were already available for further analysis (Whitworth et al. 2021). However, despite the advances in sequencing technologies, it is thus far only possible to access 313 myxobacterial genome assemblies in GenBank and 575 genomes from uncultured myxobacteria (metagenome-assembled genomes—MAGs) at the time of this publication. The in-house myxobacterial strain collection of the Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) contains more than 10,000 unique strains. This showcases that only a small fraction of isolated and cultivable myxobacterial genomes have been sequenced and deposited.

In this work, we report the draft genome sequence of the myxobacterium *A. fasciculatus* SBSr002 obtained by ONT long-read sequencing. A genome completeness of 72.20% by checkM and complete BUSCOs of 40% (Tables S3 and S4) were sufficient to identify new BGCs. With the draft sequence alone, we could identify 27 BGCs encoding biosynthetic genes involved in the synthesis of secondary metabolites. The number of BGCs detected here highlights this myxobacteria as a rich source of secondary metabolites with the potential for drug discovery. This bacterium contains one of the largest microbial genomes identified so far with a total length of 13.5 Mbp, similar to the myxobacteria *Minicystis rosea* DSM 24000 and *Sorangium cellulosum* So0157–2 (Table 2), suggesting that myxobacterial species have some of the largest known bacterial genomes. Even before genome sequencing became available, it had already been determined that myxobacteria typically have large genomes characterised by a significantly high (~70%) GC content (Whitworth et al. 2021). Generating fully assembled genomes for a large number of microbes would be costly and can be sometimes restrictive when analysing a large number of genomes. Alternatively, draft genomes can be generated for bioprospection and cloning of new BGCs. Draft genomes are often regarded as intermediate steps towards fully assembled genomes. However, our results show that important information can be obtained already from draft genomes such as the presence of new interesting BGCs. It also shows that the sequence quality is sufficient for primer design for different applications, such as TAR cloning. According to a cluster of orthologs analysis (COG), functions related to regulation, such as signal transduction mechanisms that encode regulatory proteins exemplified by serine/threonine/histidine kinases, form the largest percentage of proteins associated to this genome and they might be correlated with the complex regulatory networks controlling social lifestyle and secondary metabolism of myxobacteria (Figure S4). In fact, this had already been evidenced in other large myxobacterial genomes, for example, *S. cellulosum*. A COG analysis also revealed that the main functional category in *S. cellulosum* genome encodes regulatory proteins (Schneiker et al. 2007).

The long-read sequencing on *A. fasciculatus* resulted in 17 contigs that could not be fully closed. However, a N50 of 3.1 was sufficient to find BGCs using antiSMASH 7.1 (Blin et al. 2023). Accordingly, antiSMASH revealed 27 BGCs and 10 different classes of secondary metabolites in just this strain. Specifically, we detected six NRPSs, one PKSs, six terpenes, four RIPPs-like (other unspecified RiPP), one RRE-containing and one lanthipeptide-class-I cluster. We have investigated the feasibility of predicting BGCs for further downstream analysis including cloning and reconstruction of BGCs for heterologous expression by capturing a novel BGC into a yeast vector. We chose a novel BGC encoding the enzymatic machinery for the biosynthesis of a class-I lanthipeptide with no similarity to other known BGCs. Lanthipeptides are a class of RiPPs that contain thioether cross-links or Lanthionines catalysed by post-translational modification (PTMs) enzymes (Pei et al. 2022). Typically, class l lanthipeptides include one lanthipeptide dehydratase (LanB) and one lanthipeptide cyclase (LanC). Herein, we detected genes, *afaL* and *afaG*, encoding enzymes that belong to the specific domain architecture of LanB and LanC, respectively (Table 5). In general, these non-canonical BGCs have the potential to yield secondary metabolites with a variety of biological functions including antimicrobials, antivirals, anticancer and analgesic properties (Li et al. 2021). The most well known class-l lanthipeptide is nisin, a valuable natural product with potent antimicrobial properties widely used in the food industry (Delves-Broughton 1996). Another important RiPP is the ω-conotoxine, a neurotoxin isolated from the venom of sea cone snails and used to treat chronic pain (Schmidtko et al. 2010).

We also report the presence of multiple putative PTMs, and other accessory enzymes involved in the lanthipeptide biosynthesis. We predicted one lanthipeptide synthetase (*afaG*), one lanthipeptide dehydratase (*afaL*), one M4 metallopeptidase (*afaM*) and one carboxypeptidase (*afaF*) presumably involved in RiPPs biosynthesis (Table 5). Usually, one or two LanBs are required for lanthipeptide biosynthesis (Ortega et al. 2016; Reiners et al. 2017). We also performed protein similarity searches using NCBI Conserved Domain Architecture Retrieval Tool (Geer et al. 2002), and we found conserved domains for LanB and LanC.

The presence of multiple peptidases suggests that after the initial degradation of the N-terminal leader peptide by a protease for

lanthipeptide maturation, a number of amino acids are additionally removed by another peptidase to reach the final modified peptide (Chen et al. 2019; van der Donk W and Eslami 2023). Specifically, it was predicted a protease (*afa*M) from an unprecedented class of M4 family metalloproteases probably involved in leader peptide removal of class-l lanthipeptides. We also identified one putative carboxypeptidase (*afaF*), just upstream the lanthionine synthetase, probably responsible for the degradation of single C-terminal amino acids from the leader peptide.

Strikingly, a putative NADP oxidoreductase coenzyme F420-dependent (*afaB*) was detected in the lanthipeptide BGC (Table 5). This additional biosynthetic enzyme suggests oxidative steps to convert peptidic amino acids to the final modified peptide. This unusual biosynthetic feature is associated with important oxidoreductive reactions in the primary metabolism. Although there are a few examples in which coenzyme F420 has demonstrated a role in secondary metabolism (Greening et al. 2016; Bashiri 2022) including the reduction of this cofactor in oxytetracycline biosynthesis (Wang et al. 2013) and introduction of a D-amino acid to generate the mature lanthipeptide (Xu et al. 2020), there are no reports to date of this rare biosynthetic feature in myxobacteria for secondary metabolites biosynthesis.

Previous research has shown that receptor-histidine kinases and transcriptional response regulators are involved in the lanthipeptide production (Klein et al. 1993; Engelke et al. 1994; Ra et al. 1996; Altena et al. 2000; Upton et al. 2001). Typically, these components work similar to quorum sensing in which the presence of lanthipeptides trigger transcriptional activation of lanthipeptide biosynthetic genes. We found that *afaO* is a predicted ORF for a transcriptional regulator.

antiSMASH revealed (Table S6) that paenicidin A, paenicidin B and Geobacillin l are RiPPs type secondary metabolites with the highest similarity to the *A. fasciculatus* lanthipeptide BGC. These lanthipeptides have been investigated due to their potential applications in food preservation and the treatment of drug-resistant bacterial infections (Garg et al. 2012; Baindara et al. 2016; Belkum et al. 2016).

Although there might be some ONT-sequenced BGCs that would still require error correction with short-reads to be cloned, we were able to capture the aforementioned novel lanthipeptide class-l BGC in a yeast-bacteria shuttle vector for further analysis. The whole plasmid sequencing of the lanthipeptide vector showed sequence differences from the draft genome but indicated that the *A. fasciculatus* SBSr002 quality of the assembly from ONT was enough for capturing the 26 kb lanthipeptide class-l BGC. As validation method, we used Sanger sequencing to sequence the complete lanthipeptide class-l BGC using *A. fasciculatus* SBSr002 genomic DNA (Amplicons 1–5 covering the approximately 26 kb). Alignment of this sequence confirmed a 99.5% similarity with the ONT lanthipeptide class-l BGC DNA sequence (Figure S6). Sanger sequencing revealed ONT sequencing errors leading to a mistake in gene annotation. However, this analysis showcases the robustness of our pipeline to detect and clone BGCs even when there are sequencing errors. Designed primers for lanthipeptide class-l BGC cloning based on the ONT draft genome assembly successfully amplified all 5 kb DNA fragments. While provided primers have a correct sequence, it is not required to have an accurate BGC sequence because amplification only depends on the primer sequences. This is an example in which PCR-based DNA synthesis can be used for BGC cloning, but in other cases such as, metagenomic DNA, synthetic DNA might be more convenient. However, any errors coming from the original sequencing will be copied to the synthetic DNA constructs.

We were able to identify different classes of putative BGCs by mining the draft genome of *A. fasciculatus SBSr002*. With the aim to discover new RiPPs in myxobacteria, we selected a lanthipeptide class-l BGC that had not been previously identified and consisted of several novel PTM enzymes. These tailoring enzymes can modify the precursor peptide to generate new chemical structures with diverse biological activities. Importantly, exploration of novel PTM enzymes will enable adapting these catalytic units as synthetic biology tools for producing new-to-nature peptides. In the last decade, a vast variety of PTM enzymes have been discovered that are capable of installing unusual chemical structures including heterocyclic and/or macrocycles RiPPs (Schramma et al. 2015; Alexandru-Crivac et al. 2017). In addition, other tailoring enzymes have been reported to introduce rare halogen atoms (Nguyen et al. 2021; Wang et al. 2023), D-amino-acids (Shang et al. 2019; Korneli et al. 2021), β-amino-acids (Cabrele et al. 2014; Acedo et al. 2019; Scott et al. 2022), and β-hydroxylated amino-acids (An et al. 2018) in the core peptide. PTM enzymes with a wider substrate tolerance have been observed (Van Heel et al. 2016; Li et al. 2019; Arias-Orozco et al. 2022). This suggests that these modification enzymes could facilitate peptide engineering (Fu et al. 2023).

Myxobacteria have proven to be a valuable source of numerous natural products. Although there has been an increase in the use of third-generation sequencing technologies for microbial genomics during the last years, there are still a large diversity of unsequenced bacteria with secondary metabolite biosynthetic gene clusters to be discovered. Progress in sequencing technologies in terms of accessibility and increased accuracy will greatly advance this goal. We demonstrated that a draft genome assembly obtained using only ONT sequencing can be used for the discovery of BGCs and allowed us to successfully clone a novel lanthipeptide class-l BGC. As ONT is gradually increasing its accuracy, more applications, such as BGC discovery, are becoming more attainable by this technology. Although PacBio technology can generate more accurate data than ONT, the price and scalability of ONT make it more accessible to research groups and provide data of sufficient quality for BGC cloning as demonstrated in our work. Other alternatives to discover BGCs include hybrid ONT/Illumina approaches to generate a high-quality whole genome sequence (Jørgensen et al. 2024). Although our pipeline contributes to the discovery and capture of BGCs through a combination of relatively fast and more cost-effective sequencing, genome mining, yeast-based TAR cloning, other challenges in the natural product discovery are often found including BGC refactoring, heterologous expression, precursors availability and isolating the final product of a BGC. All these advances will positively impact the revitalization of the discovery of novel secondary metabolites for drug development amid the antimicrobial resistance crisis and emergence of new diseases.

Author Contributions

Marco A. Campos-Magaña: investigation, funding acquisition, writing – original draft, writing – review and editing, validation, formal analysis, resources, conceptualization, data curation, project administration, software, methodology, visualization. **Vitor A. P. Martins dos Santos:** conceptualization, writing – review and editing, project administration, resources, supervision, formal analysis, funding acquisition, writing – original draft, methodology. **Luis Garcia-Morales:** conceptualization, methodology, investigation, validation, funding acquisition, writing – original draft, writing – review and editing, formal analysis, project administration, resources, supervision, software.

Acknowledgements

This work was funded by CONACYT and BIOS project grant agreement no. 101070281. We thank Rob Smith for his support in editing the manuscript. We also thank Bart Nijsee for his assistance with genome analysis.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that supports the findings of this study are available in the Supporting Information of this article.

References

Acedo, J. Z., I. R. Bothwell, L. An, et al. 2019. "O-Methyltransferase-Mediated Incorporation of a β-Amino Acid in Lanthipeptides." *Journal of the American Chemical Society* 141, no. 42: 16790–16801. https://doi.org/10.1021/jacs.9b07396.

Agrawal, P., S. Amir, S. Deepak, D. Barua, and D. Mohanty. 2021. "RiPPMiner-Genome: A Web Resource for Automated Prediction of Crosslinked Chemical Structures of RiPPs by Genome Mining." *Journal of Molecular Biology* 433, no. 11: 166887. https://doi.org/10.1016/j.jmb.2021.166887.

Alexandru-Crivac, C. N., C. Umeobika, N. Leikoski, et al. 2017. "Cyclic Peptide Production Using a Macrocyclase With Enhanced Substrate Promiscuity and Relaxed Recognition Determinants." *Chemical Communications* 53: 10656–10659. https://doi.org/10.1039/c7cc05913b.

Altena, K., A. Guder, C. Cramer, and G. Bierbaum. 2000. "Biosynthesis of the Lantibiotic Mersacidin: Organization of a Type B Lantibiotic Gene Cluster." *Applied and Environmental Microbiology* 66, no. 6: 2565–2571. https://doi.org/10.1128/AEM.66.6.2565-2571.2000.

An, L., D. P. Cogan, C. D. Navo, et al. 2018. "Substrate-Assisted Enzymatic Formation of Lysinoalanine in Duramycin." *Nature Chemical Biology* 14, no. 10: 928–933. https://doi.org/10.1038/s41589-018-0122-4.

Arias-Orozco, P., Y. Yi, F. Ruijne, R. Cebrián, and O. P. Kuipers. 2022. "Investigating the Specificity of the Dehydration and Cyclization Reactions in Engineered Lanthipeptides by Synechococcal SyncM." *ACS Synthetic Biology* 12, no. 1: 164–177. https://doi.org/10.1021/acssynbio.2c00455.

Baindara, P., V. Chaudhry, G. Mittal, et al. 2016. "Characterization of the Antimicrobial Peptide Penisin, a Class Ia Novel Lantibiotic From Paenibacillus Sp. Strain A3." *Antimicrobial Agents and Chemotherapy* 60, no. 1: 580–591. https://doi.org/10.1128/AAC.01813-15.

Baltz, R. H. 2021. "Genome Mining for Drug Discovery: Progress at the Front End." *Journal of Industrial Microbiology and Biotechnology* 48: kuab044. https://doi.org/10.1093/jimb/kuab044.

Bashiri, G. 2022. "Cofactor F420, an Emerging Redox Power in Biosynthesis of Secondary Metabolites." *Biochemical Society Transactions* 50, no. 1: 253–267. https://doi.org/10.1042/BST20211286.

Van Belkum, M. J., C. T. Lohans, and J. C. Vederas. 2016. "Draft Genome Sequences of *Paenibacillus polymyxa* NRRL B-30509 and *Paenibacillus terrae* NRRL B-30644, Strains From a Poultry Environment That Produce Tridecaptin A and Paenicidins." *Genome Announcements* 3, no. 2: 15. https://doi.org/10.1128/genomeA.00372-15.

Blin, K., S. Shaw, H. E. Augustijn, et al. 2023. "AntiSMASH 7.0: New and Improved Predictions for Detection, Regulation, Chemical Structures and Visualisation." *Nucleic Acids Research* 51, no. 1: W46–W50. https://doi.org/10.1093/nar/gkad344.

Cabrele, C., T. A. Martinek, O. Reiser, and Ł. Berlicki. 2014. "Peptides Containing β-Amino Acid Patterns: Challenges and Successes in Medicinal Chemistry." *Journal of Medicinal Chemistry* 57, no. 23: 9718–9739. https://doi.org/10.1021/jm5010896.

Camacho, C., G. Coulouris, V. Avagyan, et al. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10: 421. https://doi.org/10.1186/1471-2105-10-421.

Chen, S., B. Xu, E. Chen, et al. 2019. "Zn-Dependent Bifunctional Proteases Are Responsible for Leader Peptide Processing of Class III Lanthipeptides." *Proceedings of the National Academy of Sciences of the United States of America* 116, no. 7: 2533–2538. https://doi.org/10.1073/pnas.1815594116.

Delves-Broughton, J. 1996. "Applications of the Bacteriocin, Nisin." *Antonie Van Leeuwenhoek* 69, no. 2: 193–202. https://doi.org/10.1007/BF00399424.

Eckert, N., Y. Rebets, L. Horbal, et al. 2023. "Discovery and Overproduction of Novel Highly Bioactive Pamamycins Through Transcriptional Engineering of the Biosynthetic Gene Cluster." *Microbial Cell Factories* 22, no. 1: 233. https://doi.org/10.1186/s12934-023-02231-x.

Engelke, G., Z. Gutowski-Eckel, P. Kiesau, K. Siegers, M. Hammelmann, and K. D. Entian. 1994. "Regulation of Nisin Biosynthesis and Immunity in *Lactococcus lactis* 6F3." *Applied and Environmental Microbiology* 60, no. 3: 814–825. https://doi.org/10.1128/aem.60.3.814-825.1994.

Fu, Y., Y. Xu, F. Ruijne, and O. P. Kuipers. 2023. "Engineering Lanthipeptides by Introducing a Large Variety of RiPP Modifications to Obtain New-To-Nature Bioactive Peptides." *FEMS Microbiology Reviews* 47: fuad017. https://doi.org/10.1093/femsre/fuad017.

Gao, Y., C. Walt, C. D. Bader, and R. Müller. 2023. "Genome-Guided Discovery of the Myxobacterial Thiolactone-Containing Sorangibactins." *ACS Chemical Biology* 18, no. 4: 924–932. https://doi.org/10.1021/acschembio.3c00063.

Garg, N., W. Tang, Y. Goto, S. K. Nair, and W. van der Donk. 2012. "Lantibiotics From *Geobacillus thermodenitrificans*." *Proceedings of the National Academy of Sciences of the United States of America* 109, no. 14: 5241–5246. https://doi.org/10.1073/pnas.1116815109.

Geer, L. Y., M. Domrachev, D. J. Lipman, and S. H. Bryant. 2002. "CDART: Protein Homology by Domain Architecture." *Genome Research* 12, no. 10: 1619–1623. https://doi.org/10.1101/gr.278202.

Gemperlein, K., A. Rachid, R. Garcia, S. Wenzel, and R. Müller. 2014. "Polyunsaturated Fatty Acid Biosynthesis in Myxobacteria: Different PUFA Synthases and Their Product Diversity." *Chemical Science* 5, no. 5: 1733. https://doi.org/10.1039/c3sc53163e.

Goldman, B. S., W. C. Nierman, D. Kaiser, et al. 2006. "Evolution of Sensory Complexity Recorded in a Myxobacterial Genome." *Proceedings of the National Academy of Sciences* 103, no. 41: 15200–15205. https://doi.org/10.1073/pnas.0607335103.

Greening, C., F. H. Ahmed, A. E. Mohamed, et al. 2016. "Physiology, Biochemistry, and Applications of F_{420}- and F_0- Dependent Redox Reactions." *Microbiology and Molecular Biology Reviews* 80, no. 2: 451–493. https://doi.org/10.1128/mmbr.00070-15.

Gu, B., D. G. Kim, D. K. Kim, M. Kim, H. U. Kim, and M. K. Oh. 2023. "Heterologous Overproduction of Oviedomycin by Refactoring Biosynthetic Gene Cluster and Metabolic Engineering of Host Strain *Streptomyces coelicolor*." *Microbial Cell Factories* 22, no. 1: 212. https://doi.org/10.1186/s12934-023-02218-8.

Han, K., Z. F. Li, R. Peng, et al. 2013. "Extraordinary Expansion of a *Sorangium cellulosum* Genome From an Alkaline Milieu." *Scientific Reports* 3: srep02101. https://doi.org/10.1038/srep02101.

Hemmerling, F., and J. Piel. 2022. "Strategies to access biosynthetic novelty in bacterial genomes for drug discovery." *Nature Reviews Drug Discovery* 21, no. 5: 359–378. https://doi.org/10.1038/s41573-022-00414-6.

Huntley, S., N. Hamann, S. Wegener-Feldbrügge, et al. 2010. "Comparative Genomic Analysis of Fruiting Body Formation in Myxococcales." *Molecular Biology and Evolution* 28, no. 2: 1083–1097. https://doi.org/10.1093/molbev/msq292.

Huntley, S., Y. Zhang, A. Treuner-Lange, S. Kneip, C. W. Sensen, and L. Søgaard-Andersen. 2012. "Complete Genome Sequence of the Fruiting Myxobacterium *Corallococcus coralloides* DSM 2259." *Journal of Bacteriology* 194, no. 11: 3012–3013. https://doi.org/10.1128/JB.00397-12.

Huo, L., J. J. Hug, C. Fu, X. Bian, Y. Zhang, and R. Müller. 2019. "Heterologous Expression of Bacterial Natural Product Biosynthetic Pathways." *Natural Product Reports* 36, no. 10: 1412–1436. https://doi.org/10.1039/c8np00091c.

Ivanova, N., C. Daum, E. Lang, et al. 2010. "Complete Genome Sequence of *Haliangium ochraceum* Type Strain (SMP-2 T)." *Standards in Genomic Sciences* 2, no. 1: 96–106. https://doi.org/10.4056/sigs.69.1277.

Jørgensen, T. S., O. S. Mohite, E. B. Sterndorff, et al. 2024. "A Treasure Trove of 1034 Actinomycete Genomes." *Nucleic Acids Research* 52, no. 13: 7487–7503. https://doi.org/10.1093/nar/gkae523.

Kaiser, D., M. Robinson, and L. Kroos. 2010. "Myxobacteria, Polarity, and Multicellular Morphogenesis." *Cold Spring Harbor Perspectives in Biology* 2, no. 8: a000380. https://doi.org/10.1101/cshperspect.a000380.

Kang, H. S., Z. Charlop-Powers, and S. F. Brady. 2016. "Multiplexed CRISPR/Cas9- and TAR-Mediated Promoter Engineering of Natural Product Biosynthetic Gene Clusters in Yeast." *ACS Synthetic Biology* 5, no. 9: 1002–1010. https://doi.org/10.1021/acssynbio.6b00080.

Kenshole, E., M. Herisse, M. Michael, and S. J. Pidot. 2021. "Natural Product Discovery Through Microbial Genome Mining." *Current Opinion in Chemical Biology* 60: 47–54. https://doi.org/10.1016/j.cbpa.2020.07.010.

Kim, H., C. H. Ji, H. W. Je, J. P. Kim, and H. S. Kang. 2019. "MpCRISTAR: Multiple Plasmid Approach for CRISPR/Cas9 and TAR-Mediated Multiplexed Refactoring of Natural Product Biosynthetic Gene Clusters." *ACS Synthetic Biology* 9, no. 1: 175–180. https://doi.org/10.1021/acssynbio.9b00382.

Kim, S. H., W. Lu, M. K. Ahmadi, D. Montiel, M. A. Ternei, and S. F. Brady. 2018. "Atolypenes, Tricyclic Bacterial Sesterterpenes Discovered Using a Multiplexed In Vitro Cas9-TAR Gene Cluster Refactoring Approach." *ACS Synthetic Biology* 8, no. 1: 109–118. https://doi.org/10.1021/acssynbio.8b00361.

Klein, C., C. Kaletta, and K. D. Entian. 1993. "Biosynthesis of the Lantibiotic Subtilin Is Regulated by a Histidine Kinase/Response Regulator System." *Applied and Environmental Microbiology* 59, no. 1: 296–303. https://doi.org/10.1128/aem.59.1.296-303.1993.

Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner. 2019. "Assembly of Long, Error-Prone Reads Using Repeat Graphs." *Nature Biotechnology* 37, no. 5: 540–546. https://doi.org/10.1038/s41587-019-0072-8.

Korneli, M., S. W. Fuchs, K. Felder, C. Ernst, L. V. Zinsli, and J. Piel. 2021. "Promiscuous Installation of d-Amino Acids in Gene-Encoded Peptides." *ACS Synthetic Biology* 10, no. 2: 236–242. https://doi.org/10.1021/acssynbio.0c00470.

Kouprina, N., and V. Larionov. 2008. "Selective Isolation of Genomic Loci From Complex Genomes by Transformation-Associated Recombination Cloning in the Yeast *Saccharomyces cerevisiae*." *Nature Protocols* 3: 371–377. https://doi.org/10.1038/nprot.2008.5.

Lee, N., S. Hwang, J. Kim, S. Cho, B. Palsson, and B. K. Cho. 2020. "Mini Review: Genome Mining Approaches for the Identification of Secondary Metabolite Biosynthetic Gene Clusters in *Streptomyces*." *Computational and Structural Biotechnology Journal* 18: 1548–1556. https://doi.org/10.1016/j.csbj.2020.06.024.

Li, C., K. Alam, Y. Zhao, et al. 2021. "Mining and Biosynthesis of Bioactive Lanthipeptides From Microorganisms." *Frontiers in Bioengineering and Biotechnology* 9: 2466. https://doi.org/10.3389/fbioe.2021.692466.

Li, M. H. T., P. M. U. Ung, J. Zajkowski, S. Garneau-Tsodikova, and D. H. Sherman. 2009. "Automated Genome Mining for Natural Products." *BMC Bioinformatics* 10: 185. https://doi.org/10.1186/1471-2105-10-185.

Li, Q., M. Montalban-Lopez, and O. P. Kuipers. 2019. "Feasability of Introducing a Thioether Ring in Vasopressin by nisBTC Co-Expression in *Lactococcus lactis*." *Frontiers in Microbiology* 10: 1508. https://doi.org/10.3389/fmicb.2019.01508.

Miao, V., M. F. Coëffet-LeGal, P. Brian, et al. 2005. "Daptomycin Biosynthesis in *Streptomyces roseosporus*: Cloning and Analysis of the Gene Cluster and Revision of Peptide Stereochemistry." *Microbiology* 151, no. 5: 1507–1523. https://doi.org/10.1099/mic.0.27757-0.

Mistry, J., S. Chuguransky, L. Williams, et al. 2020. "Pfam: The Protein Families Database in 2021." *Nucleic Acids Research* 49, no. D1: D412–D419. https://doi.org/10.1093/nar/gkaa913.

Montiel, D., H. S. Kang, F. Y. Chang, Z. Charlop-Powers, and S. F. Brady. 2015. "Yeast Homologous Recombination-Based Promoter Engineering for the Activation of Silent Natural Product Biosynthetic Gene Clusters." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 29: 8953–8958. https://doi.org/10.1073/pnas.1507606112.

Nguyen, N. A., Z. Lin, I. Mohanty, N. Garg, E. W. Schmidt, and V. Agarwal. 2021. "An Obligate Peptidyl Brominase Underlies the Discovery of Highly Distributed Biosynthetic Gene Clusters in Marine Sponge Microbiomes." *Journal of the American Chemical Society* 143, no. 27: 10221–10231. https://doi.org/10.1021/jacs.1c03474.

Ortega, M. A., Y. Hao, M. C. Walker, et al. 2016. "Structure and tRNA Specificity of MibB, a Lantibiotic Dehydratase From Actinobacteria Involved in NAI-107 Biosynthesis." *Cell Chemical Biology* 23, no. 3: 370–380. https://doi.org/10.1016/j.chembiol.2015.11.017.

Pal, S., G. Sharma, and S. Subramanian. 2021. "Complete Genome Sequence and Identification of Polyunsaturated Fatty Acid Biosynthesis Genes of the Myxobacterium Minicystis Rosea DSM 24000T." *BMC Genomics* 22, no. 1: 655. https://doi.org/10.1186/s12864-021-07955-x.

Parks, D. H., M. Imelfort, C. T. Skennerton, P. Hugenholtz, and G. W. Tyson. 2015. "CheckM: Assessing the Quality of Microbial Genomes Recovered From Isolates, Single Cells, and Metagenomes." *Genome Research* 25, no. 7: 1043–1055. https://doi.org/10.1101/gr.186072.114.

Pei, Z. F., L. Zhu, R. Sarksian, W. A. van der Donk, and S. K. Nair. 2022. "Class v Lanthipeptide Cyclase Directs the Biosynthesis of a Stapled Peptide Natural Product." *Journal of the American Chemical Society* 144, no. 38: 17549–17557. https://doi.org/10.1021/jacs.2c06808.

Ra, S. R., M. Qiao, T. Immonen, I. Pujana, and P. E. J. Saris. 1996. "Genes Responsible for Nisin Synthesis, Regulation and Immunity Form a Regulon of Two Operons and Are Induced by Nisin in Lactoccocus Lactis N8." *Microbiology* 142, no. 5: 1281–1288. https://doi.org/10.1099/13500872-142-5-1281.

Reiners, J., A. Abts, R. Clemens, S. H. J. Smits, and L. Schmitt. 2017. "Stoichiometry and Structure of a Lantibiotic Maturation Complex." *Scientific Reports* 7: srep42163. https://doi.org/10.1038/srep42163.

Rode, C. K., L. J. Melkerson-Watson, A. T. Johnson, and C. A. Bloch. 1999. "Type-Specific Contributions to Chromosome Size Differences

in." *Infection and Immunity* 67, no. 1: 230–236. https://doi.org/10.1128/iai.67.1.230-236.1999.

Scherlach, K., and C. Hertweck. 2021. "Mining and Unearthing Hidden Biosynthetic Potential." *Nature Communications* 12: 3864. https://doi.org/10.1038/s41467-021-24133-5.

Schmidtko, A., J. Lötsch, R. Freynhagen, and G. Geisslinger. 2010. "Ziconotide for Treatment of Severe Chronic Pain." *Lancet* 375, no. 9725: 1569–1577. https://doi.org/10.1016/S0140-6736(10)60354-6.

Schneiker, S., O. Perlova, O. Kaiser, et al. 2007. "Complete Genome Sequence of the Myxobacterium *Sorangium cellulosum*." *Nature Biotechnology* 25, no. 11: 1281–1289. https://doi.org/10.1038/nbt1354.

Schramma, K. R., L. B. Bushin, and M. R. Seyedsayamdost. 2015. "Structure and Biosynthesis of a Macrocyclic Peptide Containing an Unprecedented Lysine-To-Tryptophan Crosslink." *Nature Chemistry* 7, no. 5: 431–437. https://doi.org/10.1038/nchem.2237.

Schwengers, O., L. Jelonek, M. A. Dieckmann, S. Beyvers, J. Blom, and A. Goesmann. 2021. "Bakta: Rapid and Standardized Annotation of Bacterial Genomes via Alignment-Free Sequence Identification." *Microbial Genomics* 7, no. 11: 685. https://doi.org/10.1099/MGEN.0.000685.

Scott, T. A., M. Verest, J. Farnung, et al. 2022. "Widespread Microbial Utilization of Ribosomal β-Amino Acid-Containing Peptides and Proteins." *Chemistry* 8, no. 10: 2659–2677. https://doi.org/10.1016/j.chempr.2022.09.017.

Seyfert, C. E., A. V. Müller, D. J. Walsh, et al. 2023. "New Genetically Engineered Derivatives of Antibacterial Darobactins Underpin Their Potential for Antibiotic Development." *Journal of Medicinal Chemistry* 66, no. 23: 16330–16341. https://doi.org/10.1021/acs.jmedchem.3c01660.

Shang, Z., J. M. Winter, C. A. Kauffman, I. Yang, and W. Fenical. 2019. "Salinipeptins: Integrated Genomic and Chemical Approaches Reveal Unusual d -Amino Acid-Containing Ribosomally Synthesized and Post-Translationally Modified Peptides (RiPPs) From a Great Salt Lake *Streptomyces* Sp." *ACS Chemical Biology* 14, no. 3: 415–425. https://doi.org/10.1021/acschembio.8b01058.

Sharma, G., T. Narwani, and S. Subramanian. 2016. "Complete Genome Sequence and Comparative Genomics of a Novel Myxobacterium Myxococcus Hansupus." *PLoS One* 11, no. 2: e0148593. https://doi.org/10.1371/journal.pone.0148593.

Shimoyama, Y. (2022). "COGclassifier: A tool for classifying prokaryote protein sequences into COG functional category [Computer software]." https://github.com/moshi4/COGclassifier.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness With Single-Copy Orthologs." *Bioinformatics* 31, no. 19: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Skinnider, M. A., N. J. Merwin, C. W. Johnston, and N. A. Magarvey. 2017. "PRISM 3: Expanded Prediction of Natural Product Chemical Structures From Microbial Genomes." *Nucleic Acids Research* 45, no. W1: W49–W54. https://doi.org/10.1093/nar/gkx320.

Treuner-Lange, A., M. Bruckskotten, O. Rupp, A. Goesmann, and L. Søgaard-Andersen. 2017. "Complete Genome Sequence of the Fruiting Myxobacterium *Myxococcus macrosporus* Strain DSM 14697, Generated by PacBio Sequencing." *Genome Announcements* 5, no. 40: 1127. https://doi.org/10.1128/genomeA.01127-17.

Upton, M., J. R. Tagg, P. Wescombe, and H. F. Jenkinson. 2001. "Intra- and Interspecies Signaling Between *Streptococcus salivarius* and *Streptococcus pyogenes* Mediated by SalA and SalA1 Lantibiotic Peptides." *Journal of Bacteriology* 183, no. 13: 3931–3938. https://doi.org/10.1128/JB.183.13.3931-3938.2001.

van der Donk W, A., and S. M. Eslami. 2023. "Proteases Involved in Leader Peptide Removal During RiPP Biosynthesis." *ACS Bio & Med Chem Au* 4, no. 1: 20–36.

Van Heel, A. J., A. de Jong, C. Song, J. H. Viel, J. Kok, and O. P. Kuipers. 2018. "BAGEL4: A User-Friendly Web Server to Thoroughly Mine RiPPs and Bacteriocins." *Nucleic Acids Research* 46, no. W1: W278–W281. https://doi.org/10.1093/nar/gky383.

Van Heel, A. J., T. G. Kloosterman, M. Montalban-Lopez, et al. 2016. "Discovery, Production and Modification of Five Novel Lantibiotics Using the Promiscuous Nisin Modification Machinery." *ACS Synthetic Biology* 5, no. 10: 1146–1154. https://doi.org/10.1021/acssynbio.6b00033.

Waldron, C., P. Matsushima, P. R. Rosteck, et al. 2001. "Cloning and Analysis of the Spinosad Biosynthetic Gene Cluster of *Saccharopolyspora spinosa*11The DNA Sequence Reported Here Was Deposited in GenBank Under the Accession Number AY007564." *Chemistry & Biology* 8, no. 5: 487–499. https://doi.org/10.1016/S1074-5521(01)00029-1.

Wang, G., Z. Zhao, J. Ke, et al. 2019. "CRAGE Enables Rapid Activation of Biosynthetic Gene Clusters in Undomesticated Bacteria." *Nature Microbiology* 4, no. 12: 2498–2510. https://doi.org/10.1038/s41564-019-0573-8.

Wang, J., J. Nielsen, and Z. Liu. 2021. "Synthetic Biology Advanced Natural Product Discovery." *Metabolites* 11, no. 11: 785. https://doi.org/10.3390/metabo11110785.

Wang, P., G. Bashiri, X. Gao, M. R. Sawaya, and Y. Tang. 2013. "Uncovering the Enzymes That Catalyze the Final Steps in Oxytetracycline Biosynthesis." *Journal of the American Chemical Society* 135, no. 19: 7138–7141. https://doi.org/10.1021/ja403516u.

Wang, W., G. Zheng, and Y. Lu. 2021. "Recent Advances in Strategies for the Cloning of Natural Product Biosynthetic Gene Clusters." *Frontiers in Bioengineering and Biotechnology* 9: 692797. https://doi.org/10.3389/fbioe.2021.692797.

Wang, X., X. Chen, Z. J. Wang, et al. 2023. "Discovery and Characterization of a Myxobacterial Lanthipeptide With Unique Biosynthetic Features and Anti-Inflammatory Activity." *Journal of the American Chemical Society* 145, no. 30: 16924–16937. https://doi.org/10.1021/jacs.3c06014.

Whitworth, D. E., N. Sydney, and E. J. Radford. 2021. "Myxobacterial Genomics and Post-Genomics: A Review of Genome Biology, Genome Sequences and Related 'Omics Studies." *Microorganisms* 9, no. 10: 2143. https://doi.org/10.3390/microorganisms9102143.

Xu, M., F. Zhang, Z. Cheng, et al. 2020. "Functional Genome Mining Reveals a Class V Lanthipeptide Containing a d-Amino Acid Introduced by an F420H2-Dependent Reductase." *Angewandte Chemie International Edition* 59, no. 41: 18029–18035. https://doi.org/10.1002/anie.202008035.

Yamanaka, K., K. A. Reynolds, R. D. Kersten, et al. 2014. "Direct Cloning and Refactoring of a Silent Lipopeptide Biosynthetic Gene Cluster Yields the Antibiotic Taromycin A." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 5: 1957–1962. https://doi.org/10.1073/pnas.1319584111.

Yue, X., D. Sheng, L. Zhuo, and Y. Z. Li. 2023. "Genetic Manipulation and Tools in Myxobacteria for the Exploitation of Secondary Metabolism." *Engineering Microbiology* 3, no. 2: 100075. https://doi.org/10.1016/j.engmic.2023.100075.

Zhang, J. J., K. Yamanaka, X. Tang, and B. S. Moore. 2019. "Direct Cloning and Heterologous Expression of Natural Product Biosynthetic Gene Clusters by Transformation-Associated Recombination." *Methods in Enzymology* 621: 87–110. https://doi.org/10.1016/bs.mie.2019.02.026.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.