COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

Review

Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*

Namil Lee [a], Soonkyu Hwang [a], Jihun Kim [a], Suhyung Cho [a], Bernhard Palsson [b,c,d], Byung-Kwan Cho [a,e,f,*]

[a] *Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea*
[b] *Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA*
[c] *Department of Pediatrics, University of California San Diego, La Jolla, CA 92093, USA*
[d] *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby 2800, Denmark*
[e] *Innovative Biomaterials Research Center, KI for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea*
[f] *Intelligent Synthetic Biology Center, Daejeon 34141, Republic of Korea*

ARTICLE INFO

ABSTRACT

Streptomyces are a large and valuable resource of bioactive and complex secondary metabolites, many of which have important clinical applications. With the advances in high throughput genome sequencing methods, various *in silico* genome mining strategies have been developed and applied to the mapping of the *Streptomyces* genome. These studies have revealed that *Streptomyces* possess an even more significant number of uncharacterized silent secondary metabolite biosynthetic gene clusters (smBGCs) than previously estimated. Linking smBGCs to their encoded products has played a critical role in the discovery of novel secondary metabolites, as well as, knowledge-based engineering of smBGCs to produce altered products. In this mini review, we discuss recent progress in *Streptomyces* genome sequencing and the application of genome mining approaches to identify and characterize smBGCs. Furthermore, we discuss several challenges that need to be overcome to accelerate the genome mining process and ultimately support the discovery of novel bioactive compounds.

Contents

---

* Corresponding author at: Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea.
  E-mail address: bcho@kaist.ac.kr (B.-K. Cho).

1. Introduction

Streptomyces species are filamentous Gram-positive bacteria found in the soil and a member of the largest genus of Actinobacteria. They are well-known for their ability to produce a wide array of bioactive secondary metabolites, which have a number of antiviral, antifungal, anticancer, immunosuppressive, and antibiotic functions. The large number of secondary metabolites produced by these bacteria allow them to compete in diverse microbial communities and survive in various habitats including soils, rivers, lakes, and marine ecosystems [1]. Since the first report of *Streptomyces* ability to produce antibiotics in the 1940s, a significant number of novel antibiotics have been characterized by screening the antimicrobial activity of soil *Streptomyces* against the target pathogens. Most of the currently available antibiotic classes were discovered in and produced from *Streptomyces* species isolated between 1940 and 1962. However, after two decades of success using traditional biochemical screening approaches, since cultured soil microorganism constitutes less than 0.1% of the total of soil microorganisms, the rediscovery rate of known species and compounds has continuously increased and reached 99% in the late 1980s, with no new classes of antibiotics being approved since [2,3]. Meanwhile, with the rapid emergence of broad-spectrum antibiotic resistance, it is increasingly important that we isolate novel classes of antimicrobial compounds. This search for new bioactive products has reinvigorated the field of *Streptomyces* research [4].

One bottleneck in the traditional screening methods is that *Streptomyces* often downregulates or inactivates secondary metabolite production under axenic laboratory culture conditions. Secondary metabolites are produced from the multi enzyme complexes encoded by the secondary metabolite biosynthetic gene clusters (smBGCs). smBGCs generally contain whole pathways that facilitate precursor biosynthesis, assembly, modification, resistance, and regulation of their product. The expression of these clusters is tightly controlled by complex regulatory networks governed by biotic and abiotic stresses found in the bacteria's natural habitat [5]. Therefore, only a small fraction of secondary metabolites can be produced under laboratory culture conditions, especially when we do not know the precise environmental stimuli needed to induce their synthesis. To fully realize the biosynthetic potential of *Streptomyces*, it is necessary to develop tools to identify all of the smBGCs, including those that are silenced under laboratory conditions, in their entirety encoded in the *Streptomyces* genomes.

With the recent advances in DNA sequencing technology, the number of fully sequenced *Streptomyces* genomes has increased exponentially [6,7]. As a result, it has become increasingly necessary to develop a suite of bioinformatics tools that can be used to annotate and mine these genomes. Several bioinformatics tools have been developed, including BAGEL [8], ClustScan [9], CLUSEAN [10], NP.searcher [11], PRISM [12], and antiSMASH [13], to identify smBGCs within the genome, with most of these technologies relying on the highly conserved sequences within the smBGCs to map their location. Genome mining approaches have revealed that each *Streptomyces* species possesses about 30 smBGCs, including many clusters whose products are not yet identified. These findings have supported the hypothesis that the biosynthetic potential of *Streptomyces* has been underestimated [14]. Genome mining approaches enable prediction of smBGCs from *Streptomyces* genome data quickly and easily, but characterizing these predicted smBGCs still requires extensive laboratory work, including the activation of silenced smBGCs, purification of the final products, and determination of their chemical structure. Therefore, accelerating the process linking the product with their corresponding smBGCs is of paramount importance in the effort to advance our practical under-standing of the secondary metabolite biosynthetic pathways of these bacteria.

This mini review focuses on the genome mining approaches for smBGC identification from *Streptomyces* genome data and their utility in discovering novel bioactive compounds. We briefly describe the current status of the *Streptomyces* genome sequencing projects and the importance of high-quality genomic data for smBGC identification. Next, we introduce several *in silico* genome mining tools that have been developed for this purpose and describe the characterization process of several key examples. Finally, we highlight future challenges that need to be overcome for the efficient discovery of novel secondary metabolites from *Streptomyces*.

2. Current status of the *Streptomyces* genome sequencing projects

2.1. Features of Streptomyces genomes

Unlike other bacteria, *Streptomyces* have large linear chromosomes with high G + C content. The origin of replication (oriC) is usually located at the center of the linear chromosome and terminal inverted repeats sequences (TIRs), which covalently bind with the terminal proteins, are found at each end [15]. Interestingly, the lengths and sequences of the TIRs are highly variable between species, and the number of TIR iterations does not correlate with the size of the genome [16]. The most distinct feature of the *Streptomyces* genome is the high degree of chromosomal instability, which leads to frequent spontaneous deletions and rearrangements, especially at the ends of the chromosome. For example, about 0.5% of the germinating spores of *Streptomyces lividans* undergo large deletions removing up to 25% of the genome (~2 Mbp) under laboratory-culture conditions [17]. As a result, essential genes related to cell maintenance, including transcription, translation, and DNA replication, are located in the "core" region of the chromosome. In contrast, conditionally adaptive genes, especially those related to the secondary metabolism, are usually located within the "arm" regions of the chromosome [18]. This chromosomal plasticity results in a high degree of variation in the smBGCs, which could have been acquired via prevalent gene duplications and horizontal gene transfer with other *Streptomyces* species or bacteria [19].

2.2. Currently available Streptomyces genomes

Since the whole genome sequences of *Streptomyces coelicolor* A3 (2) and *Streptomyces avermitilis* were completed by shotgun sequencing in 2003 [20,21], a number of *Streptomyces* genome sequences has been reported. Next Generation Sequencing (NGS) has revolutionized the field and enabled a drastic increase in the number of reported genomes for *Streptomyces* since 2013 (Fig. 1A). According to the RefSeq database, a total of 1,749 *Streptomyces* genomes had been deposited as of the 6th of February 2020, and more than 73% of the genomes were sequenced by NGS techniques, such as Illumina, PacBio, 454, and MinION. The 1,749 *Streptomyces* genomes composed of 867 contig level (i.e., genomes include only contigs), 646 scaffold level (i.e., genomes include scaffolds and contigs), 36 chromosome level (i.e., genomes include chromosomes, scaffolds, and contigs), and 200 complete genomes (Fig. 1A) [6]. Considering that the 36 chromosome level genomes were assembled using ambiguous (N) bases, and that three of the complete genomes contained ambiguous (N) bases, high-quality *Streptomyces* genomes comprise only about 11% of the total *Streptomyces* genomes available. The length of the 236
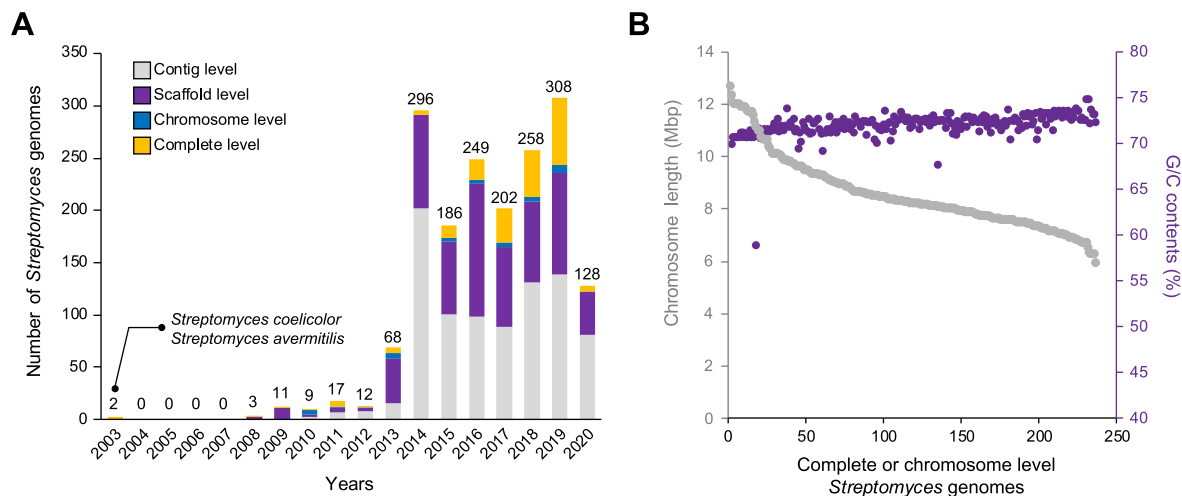
Fig. 1. Current status of *Streptomyces* genome sequences. (A) Annual number of deposited *Streptomyces* genomes in RefSeq database as of the 6th of Feb 2020. (B) Chromosome length and G + C content of 236 scaffold and complete level *Streptomyces* genomes.

scaffold and complete chromosomes ranged from 5.9 to 12.7 Mbp and an average G + C content of 71.7% (**B**). Large genome size and unusually high G + C content are the representative features of the *Streptomyces* genome as mentioned above. Interestingly, the shorter the chromosome length, the higher the G + C content observed (Fig. 1**B**). This is probably because G + C content in the "core" region is highly conserved between various species, while the "arms" are less conserved and contain relatively low G + C content.

2.3. Importance of high-quality genome sequences for genome mining

SmBGC prediction using contig-level genome is usually inappropriate because genes in an smBGC are often predicted to be scattered through several contigs. As described above, approximately 90% of the reported *Streptomyces* genomes are incomplete, containing varying degrees of contig or ambiguous sequence contributions. Securing high-quality *Streptomyces* genome sequences is challenging as a result of the low fidelity of current sequencing techniques when dealing with high G + C and repetitive sequences [7]. Furthermore, because of its linear chromosome, it is difficult to evaluate the completeness of the genome assembly when compared to other bacteria with circular chromosomes. Completeness of the genome is typically quantified using Benchmarking Universal Single-Copy Orthologs (BUSCO), which measures the number of copies of single copy genes in the sequence data and provides a quantitative assessment of genome assembly and gene sets [22]. In 2016, genome completeness of 653 *Streptomyces* genomes was analyzed using BUSCO, which revealed that about 36% of *Streptomyces* genomes have poor completeness [23]. Given that the *Streptomyces* BUSCO markers used in this analysis included only 40 genes, and the fact that the number of single copy genes in the *Streptomyces* genome now sits at 352, even the reported completeness of the *Streptomyces* genomes needs to be reassessed.

In addition to the completeness of genome assembly, quality of genome sequence (i.e., quality of bases) is also important for determining smBGC, in the aspect of accurate coding sequence (CDS) prediction. Especially as most of the smBGCs are composed of long core biosynthetic genes (>5 kb) containing repetitive sequences, thus, inaccurate genome sequence often results in frameshift errors during the prediction of CDSs within the smBGCs. For instance, the genome sequence of *Streptomyces clavuligerus* ATCC 27064, which produces β-lactam class antibiotic clavulanic acid,

has been determined, but the quality of the reported genomes was poor and contained a large number of ambiguous (N) sequences [24,25]. Recently, a high-quality genome sequence for *S. clavuligerus* was obtained using PacBio and Illumina sequencing methods, revealing that 2,184 genes out of a total of 7,163 genes were miss- or not-predicted in the previous low-quality genome sequences, including 47 genes encoded in smBGCs. The accurate CDS prediction often improves the functional annotation of genes. For example, in a low-quality genome, CRV15_02370, which located in terpene BGC, was annotated as unknown lipoprotein. Meanwhile, in the high-quality genome, the exact sequence of tandem ambiguous (N) sequences located at the upstream region of CRV15_02370 was determined, resulting in correction and re-annotation of the CRV15_02370 as 1-hydroxy-2-methyl-2-butenyl 4-diphosphatereductase [26]. As in the case of *S. clavuligerus*, applying both the PacBio sequencing method generates long reads of several kb, and the Illumina sequencing method, which has a low error rate, could be the solution to obtaining high-quality *Streptomyces* genomes [6,26]. In addition, Oxford nanopore sequencing method, which has been dominating the long-read sequencing platform with PacBio sequencing method, is more cost-effective and provides even longer reads (current record of 2.3 Mbp) than PacBio sequencing method [27]. Thus, Oxford nanopore sequencing method is expected to be an attractive alternative to PacBio sequencing method for securing high-quality *Streptomyces* genomes.

Nevertheless, the functional annotation of high-quality *Streptomyces* genome still yields a considerable amount of hypothetical proteins, due to the limited number of experimentally validated genes in the database. Indeed, about 24% of total *S. clavuligerus* genes and 25% of smBGC encoded genes were annotated as unknown genes. Even worse, the currently automated gene annotation pipelines utilize incorrect annotation existing in previous genomes to annotate new genomes because the public annotation database does not update any corrected annotation errors, neglecting the spread of misinformed functional role of the gene [28]. These incomplete functional annotations have been hampered the accurate genome mining of smBGCs and mechanistic understanding of secondary metabolite biosynthesis. Frequent and efficient update of the smBGC database with the support of individual functional genomic studies would mitigate these problems.

3. Genome mining for smBGCs

3.1. Classical approaches for the identification of smBGCs

The traditional method for identifying smBGCs in *Streptomyces* relies on the identification of the secondary metabolites using chemistry-based methods, like mass spectrometry and NMR, and then isolating the corresponding biosynthetic genes by randomized gene deletion or mutagenesis, followed by screening for non-producing clones [29–31]. This tedious method was improved after the identification of the conserved regions in the smBGCs, which could be used to screen for unknown smBGCs. Secondary metabolites have tremendous structural diversity, but biosynthetic machineries, including assembling and tailoring enzymes, for secondary metabolites belong to the same highly conserved enzyme families [32]. Especially, polyketide synthases (PKS) producing polyketides (PK) and non-ribosomal peptide synthetases (NRPS) producing non-ribosomally synthesized peptides (NRP) which are assembled by the core enzymes of large multi-modular complexes consisting of highly conserved domains and sequences. Designing probes based on these conserved sequences and screening for smBGCs using Southern blots was a popular and widely used approach for several decades. One example of this approach is the identification of the aminocoumarin antibiotic clorobiocin BGC discovered by screening the cosmid library of *Streptomyces roseochromogenes* using two heterologous probes designed against the sequence of the novobiocin BGC [33].

3.2. In silico tools for genome mining of smBGCs

Development of *in silico* nucleotide or amino acid sequence alignment tools, such as BLAST, Diamond, and HMMer, enabled researchers to mine for novel smBGCs in databases and genome sequences using a conserved sequence without the time-consuming processes of performing a Southern blot. The first microbial natural-product biosynthetic loci database for *in silico* genome mining of smBGCs was DECIPHER, which is a proprietary database constructed by Ecopia Biosciences Inc. [34]. Since then, various free databases and tools for smBGC prediction have been developed, including BAGEL [8], ClustScan [9], CLUSEAN [10], and NP.searcher [11]. Many of these tools have already been comprehensively reviewed, and the recently released web portal called "Secondary Metabolite Bioinformatics Portal" provides a description of and manual for each of these mining software and databases [35]. However, most of these tools are limited to the discovery of specific classes of secondary metabolites, including PKS and NRPS.

PRISM and antiSMASH are representative *in silico* tools for predicting various types of smBGCs [12,13]. These tools predict smBGC types by employing a sequence alignment-based profile in a Hidden Markov Model (HMM) of genes that are specific for certain types of smBGCs. For example, antiSMASH identifies smBGCs based on the highly conserved core biosynthetic enzymes and evaluates the results using a set of manually curated BGC cluster rules, followed by discarding false positives using negative models (e.g., fatty acid synthases are homologous to PKSs). The latest version, PRISM version 3, can identify 22 different types of smBGCs, and antiSMASH version 5 can predict up to 52 different types of smBGCs. Both tools are user-friendly web applications, which provide rapid gene annotation when bacterial genomes are submitted in FASTA format, making them popular tools in current mining studies. These are the most used genome mining tools, but these rule-based tools are restricted to detect similar smBGCs to known pathways. Accordingly, recently, smBGC mining tools that utilize machine learning strategies like ClusterFinder and DeepBGC, have been developed to allow the identification of unknown smBGCs [36,37]. However, current machine learning based genome mining tools have a much higher false-positive rate than the rule-based tools. Moreover, these tools are trained with the set of known clusters (e.g., MIBiG database) or a set of clusters predicted by one of the rule-based tools (e.g., antiSMASH); thus, it is still challenging to detect completely novel smBGCs.

SmBGC mining of *Streptomyces* genomes using these *in silico* tools confirmed that the genetic potential of *Streptomyces* to produce secondary metabolites has been under-estimated. According to the genome-wide study of Actinobacteria, the genomes of each *Streptomyces* species possesses about 40 smBGCs [14,38]. Considering that *Streptomyces* is the largest genus of Actinobacteria (approximately 700 valid species at present) and that the smBGCs of each *Streptomyces* are highly different, *Streptomyces* are inestimable resources for the discovery of novel bioactive compounds. In addition, a recent genome mining study of the 1,110 publicly available *Streptomyces* genomes suggested the importance of genome mining at the strain level as it increases the likelihood that researchers discover useful derivatives of known secondary metabolites and expands the diversity of recognized secondary metabolites used in new mining approaches [14].

3.3. Characterizing smBGCs identified by genome mining

Although genome mining approaches showcase the full biosynthetic potential of *Streptomyces*, it is worthless without linking the predicted smBGCs to their product. In this section, we describe several examples of genome mining approaches, which connect various metabolites with their corresponding smBGCs using (i) reverse (metabolites to genes) or (ii) forward (genes to metabolites) approaches. The reverse approach allows researchers to determine the BGCs of known secondary metabolites, and forward approach identifies the products of novel smBGCs (Fig. 2).

In the pre-genomic era, especially in the golden age of antibiotic discovery (1950 to 1960), plenty of *Streptomyces* species were isolated from the environment and screened for antimicrobial activity. However, after isolation, only antimicrobial compounds were identified via chemistry-based methods, and in most cases, the corresponding smBGCs were not determined as a result of the lack of information and relevant technologies, including DNA sequencing method [39,40]. In recent years, advances in genome mining tools have allowed researchers to adopt a reverse approach to determining the BGCs of known secondary metabolites produced from *Streptomyces*. These efforts have enabled us to identify and elucidate the biosynthetic pathways of various important secondary metabolites much faster and more efficiently than conventional randomized mutagenesis-based methods (Table 1). For example, anthracimycin, a macrolide antibiotic that exhibits antibacterial activity against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococci [41], was isolated from *Streptomyces* sp. T676 in 1995, but its BGC could not be determined at the time [42]. Recently, the genome sequence of *Streptomyces* sp. T676 was captured, and two type I modular PKS gene clusters were identified by genome mining using antiSMASH. Through additional bioinformatics analysis, one PKS gene cluster was identified as the candidate pathway for the production of anthracimycin, and heterologous expression of this BGC in *S. coelicolor* resulted in the production of anthracimycin [42]. SmBGC information obtained from reverse approaches has expanded smBGC databases, increasing the accuracy of the genome mining tools and the number of predictable smBGC types.

Accumulated *Streptomyces* genome sequences and advanced genome mining tools provide opportunities for the forward approach to smBGC identification, which allows researchers to identify the novel smBGCs from the genome, then identify the pro-
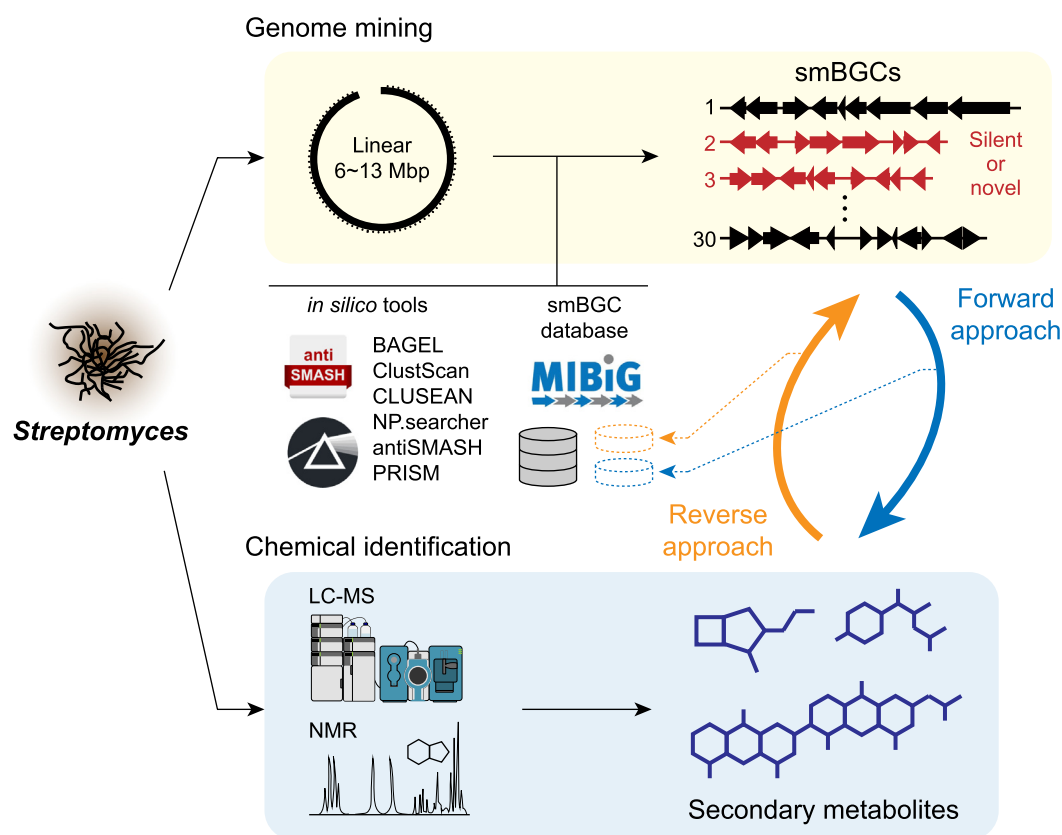
Fig. 2. Overview of genome mining approaches to identify smBGCs in *Streptomyces*. Minimum Information about a Biosynthetic Gene cluster (MIBiG) is repository for secondary metabolite biosynthetic gene clusters.

Table 1
Selected examples of the reverse approach in smBGC genome mining from *Streptomyces*.

Strains	Genome mining methods	Compound name	Year	Ref.
Streptomyces chromofuscus ATCC 49982	PKS gene search	Herboxidiene	2012	[54]
Streptomyces netropsis CGMCC 4.1650	BLAST	Pyrroleamides	2014	[55]
Streptomyces sp. T676	antiSMASH	Anthracimycin	2015	[42]
Streptomyces paulus NRRL 8115	antiSMASH	Paulomycin	2015	[60]
Streptomyces olivaceus strain FXJ7.023	antiSMASH	Lobophorin	2016	[56]
Streptomyces sp. MSC090213JE08	antiSMASH	Ishigamide	2016	[57]
Streptomyces leeuwenhoekii DSM 42122	antiSMASH	Chaxamycin	2016	[58]
Streptomyces sp. CNR-698	BLASTP	Ammosamides	2016	[59]
Stretpomyces anulatus 3533-SV4	RiPP gene search	Telomestatin	2017	[61]
Streptomyces lydicus A02	BLASTP and antiSMASH	Natamycin	2017	[62]
Streptomyces sp. MP131-18	antiSMASH	Lynamicins and spiroindimicins	2017	[63]
Streptomyces sp. SD85	antiSMASH	Sceliphrolactam	2018	[64]
Streptomyces sp. strain fd1-xmd	antiSMASH	Streptothricin and tunicamycin	2018	[65]
Streptomyces koyangensis SCSIO 5802	antiSMASH	Neoabyssomicin and abyssomicin	2018	[66]
Streptomyces olivaceus FXJ8.012	BLAST	Mycemycin	2018	[67]
Streptomyces sp. ATCC 14903	antiSMASH and BLAST	Actinonin	2018	[68]
Streptomyces aureofaciens ATCC 31442	antiSMASH	Triacsins	2018	[69]
Streptomyces lunaelactis MM109ᵀ	antiSMASH	Ferroverdins and bagremycins	2019	[70]
Streptomyces nigrescens HEK616	BLASTP	Streptoaminals	2019	[71]
Streptomyces sp. Tu 4128	antiSMASH	Bagremycin	2019	[72]
Streptomyces caniferus CA-271066	antiSMASH	Caniferolides	2019	[73]
Streptomyces sp. S816	antiSMASH	Pentamycin	2019	[74]
Streptomyces humidus CA-100629	antiSMASH	Humidimycin	2020	[75]
Streptomyces cacaoi subsp. cacaoi NBRC 12748 T	antiSMASH and NRPSsp	Pentaminomycin	2020	[76]

duct of this smBGC (Table 2). Curacozole is the first sequential oxazole/methyloxazole/thiazole ring-containing macrocyclic peptide compound identified using a genome mining based approach. Genome mining of *Streptomyces curacoi* isolated a new precursor peptide gene for ribosomally synthesized and post-translationally modified peptides (RiPPs). Purifying and determining the structure

of the product of this RiPP BGC using ESI-MS and NMR resulted in the discovery of new cytotoxic compound, curacozole [43]. In the case of curacozole, the structural prediction of RiPPs from the genomic data is comparatively easier than that of other secondary metabolites, because the entire sequence of the core peptides translated from the nucleotide sequence is generally retained in

Table 2
Selected examples of the forward approach in smBGC genome mining from *Streptomyces*.

Strains	Genome mining methods	Compound name	Year	Ref.
Streptomyces coelicolor M145	NRPS gene search	Coelichelin	2005	[77]
Streptomyces coelicolor M145	Type III PKS gene search	Germicidin	2006	[78]
Streptomyces venezuelae ATCC 10712	Lanthipeptides gene search	Venezuelin	2010	[79]
Streptomyces ambofaciens ATCC 23877	SEARCHPKS and SEARCHGTr	Stambomycins	2011	[80]
Streptomyces coeruleorubidus	BLASTP	Pacidamycin	2011	[81]
Streptomyces sp. W007	BLASTP	Angucyclinone antibiotics	2012	[82]
Streptomyces peucetius ATCC 27952	NRPS gene search	Siderophore	2013	[83]
Streptomyces sp. SANK 60404	BLASTP	Cembrane	2013	[84]
Streptomyces viridochromogenes DSM 40736	RiPPquest	Informatipeptin	2014	[85]
Streptomyces collinus Tü 365	antiSMASH	Streptocolin	2015	[86]
Streptomyces leeuwenhoekii strain C58	antiSMASH	Chaxapeptin	2015	[87]
Streptomyces chartreusis AN1542	BLASTP	Complestatin	2016	[88]
Streptomyces venezuelae ATCC 10712	BLASTP	Venemycin	2016	[89]
Streptomyces kebangsaanensis	antiSMASH	Phenazine antibiotic	2017	[90]
Streptomyces atratus SCSIO ZH16	antiSMASH	Ilamycins	2017	[91]
Streptomyces argillaceus ATCC 12956	antiSMASH	Argimycins P	2017	[92]
Streptomyces lavendulae FRI-5	antiSMASH	New diol-containing polyketide	2017	[93]
Streptomyces sp. YIM 130001	antiSMASH	Thiopeptide Antibiotic	2018	[94]
Streptomyces avermitilis KA-320	PKS gene search	Phthoxazolin A	2018	[95]
Streptomyces actuosus ATCC 25421	antiSMASH	Avermipeptin Analogue	2018	[96]
Streptomyces sp. YIM 130001	antiSMASH	Geninthiocin B	2018	[94]
Streptomyces sp. DUT11	antiSMASH and BLAST	Tunicamycin	2018	[97]
Streptomyces curacoi NBRC 12761 [T]	antiSMASH and BLAST	Curacozole (cytotoxic peptide)	2019	[43]
Streptomyces albus subsp. Chlorinus NRRL B-24108	antiSMASH	Nybomycin	2018	[98]
Streptomyces isolatess ICC1 and ICC4	antiSMASH	2′,5′–dimethoxyflavone and nordentatin	2019	[99]
Streptomyces hawaiiensis NRRL 15010	antiSMASH and BLAST	Acyldepsipeptide (ADEP)	2019	[100]
Streptomyces atratus SCSIO ZH16	antiSMASH	Atratumycin	2019	[101]
Streptomyces leeuwenhoekii C34[T]	antiSMASH	Leepeptin	2019	[102]
Streptomyces olivaceus SCSIO T05	antiSMASH	Lobophorin CR4	2019	[103]
Streptomyces sp. Tu6314	antiSMASH	Streptoketides	2020	[104]

the final product. To overcome the low productivity of curacozole and allow its robust purification, *S. curacoi* was treated with rifampicin to induce mutations, one of which occurred within the RNA polymerase β subunit, which facilitated an increased production of secondary metabolites. Thus, successful forward approaches for smBGC genome mining require two things; (i) there needs to be a predictable draft structure of the final product and (ii) the novel smBGCs need to be expressed at a high enough level to produce detectable quantities of the secondary metabolite.

There are several computational methods to predict the putative products of smBGCs, which use databases of experimentally characterized smBGCs as a reference, especially for PKSs and NRPSs. These methods use the basic rules of structure prediction which consider the substrate specificity of the catalytic domains of PKSs and the NRPSs modules to construct the backbone structure of the product, which is followed by the identification of tailoring domains to estimate further modifications or cyclization of the compounds and these results are mapped back to the database to give the user an idea of the secondary metabolite produced by their unknown smBGC. Comprehensive genome mining tools, antiSMASH and PRISM, also provide the chemical structure predictions of putative products from unknown smBGCs [44]. The accuracy of chemistry prediction is dependent on the algorithm and the database used to predict the catalytic domains of the enzyme and the substrate specificity of the domains. When PRISM version 1 was released, it was the unique tool capable of predicting the chemical structure of type II PKs, and the chemistry prediction accuracy for NRPs and type I PKs was also much higher than antiSMASH version 3.0 or NP.searcher [45]. After further improvement, PRISM version 3 became available for chemical structure prediction of products arising from non-modular biosynthetic paradigms, including RiPPs, aminocoumarins, antimetabolites, bisindoles, and phosphonate-containing natural product [12]. AntiSMASH also improved chemistry prediction when updated to version 4.0, but it provides conservative structure prediction compared to PRISM, which generates a wide range of combinatorial libraries of predicted structures by considering the uncertainty of tailoring sites [46]. Although the chemistry prediction accuracy of the most recent versions of PRISM version 4 and antiSMASH version 5.0 has never been compared, it is appropriate to use both tools according to the user's research purposes. Despite the aforementioned advances in chemistry prediction, lack of information on tailoring enzymes and frequent assignment of nearby smBGCs as hybrid smBGCs still require further experimental validation of the chemistry prediction.

To fulfill the second requirement for forward mapping approaches, several other technologies were integrated into the genome mining approach to increase secondary metabolite production or activate silent smBGCs. Since, most smBGCs of *Streptomyces* are silent under laboratory-culture condition, altering the expression level of smBGCs to produce enough amount of secondary metabolites have to come before linking the secondary metabolites to the corresponding smBGCs. This method relies on the treatment of cultures with elicitors or mutagens to increase the expression of smBGCs as in the case of curacozole discovery. Genome engineering is also a suitable method for inducing silent smBGCs, for example, one study used CRISPR-Cas9 to introduce constitutive promoters to silent novel smBGCs loci forcing the production of unique metabolites which were then evaluated by NMR [47]. Since smBGCs consist of dozens of genes, to efficiently activate the entire cluster, most of the studies have engineered the expression level of global or cluster-specific regulatory genes. Genome engineering of *Streptomyces* for the characterization of silent smBGCs has strengthened with the development of synthetic biology tools for *Streptomyces* [48]. However, genome engineering is not always applicable as a result of the difficulty in manipulating the genome of these bacteria and their slow growth rates. Heterologous expression of silent smBGCs in other *Streptomyces* is also a suitable alternative [49]. To enable this, there has been a significant amount of efforts put into the construction of a *Streptomyces*

chassis strain, which has a reduced chemical diversity as a result of the removal of its endogenous smBGCs, meaning that it can be used as a heterologous expression host for novel smBGCs characterization with reduced confounding effects [50].

Forward experimentation is significantly more challenging than the reverse method when it comes to chemical characterization of secondary metabolites. Notably, the existence of a large number of completely unknown genes, which may encode enzymes catalyzing the product tailoring steps, prevent the accuracy of predictions for the forward approach. As the smBGC database constantly expands along with the accumulation of individual functional genomics experiments, the forward approach will continue to evolve and has the most potential for the isolation and identification of novel bioactive compounds from *Streptomyces*.

4. Summary and outlook

In this mini review, we discussed the current status of *Streptomyces* genome sequencing data and *in silico* genome mining tools for smBGCs prediction. Technical advances in DNA sequencing and the rapid development of *in silico* genome mining tools demonstrate that the biosynthetic potential of *Streptomyces* has been vastly underestimated. We went on to discuss the fact that mining of smBGCs from the *Streptomyces* genome and characterization of their corresponding products using forward and reverse approaches are feasible and illustrated this with several examples. Reverse approaches link known secondary metabolites to their corresponding smBGCs and expand the current smBGC database pools enhancing the accuracy and versatility of *in silico* genome mining tools. In contrast, forward approaches enable the discovery of novel bioactive compounds from the *Streptomyces*, securing new drug candidates. The important lesson from the genome mining examples is that major bottlenecks in this process are limitation on detecting poorly characterized classes of smBGCs and determining final products of detected smBGCs. Several challenges must be overcome to enable the efficient discovery of novel secondary metabolites from *Streptomyces*. For accurate *in silico* structure predictions of putative products from smBGCs, mechanistic understanding of secondary metabolite biosynthesis based on accumulated knowledge is still lacking. Simultaneously, the induction of silent smBGCs to experimentally validate the structure of final products remains difficult, which means that there needs to be a focus on the development of synthetic biology tools for genome engineering and the construction of a *Streptomyces* chassis strain to facilitate heterologous expression.

The final use of *Streptomyces* smBGC information obtained from genome mining approaches will be the knowledge-based repurposing of smBGCs to produce derivatives of original products or non-natural compounds to improve human health and industry. Recently, several groups undertook the construction of a new assembly line for the production of fuels and synthetic industrial compounds facilitated by the rearrangement of PKS and NRPs modules [51,52]. In addition, ClusterCAD, an *in silico* toolkit for designing novel PKS assembly lines, has been developed and applied in several retro-biosynthesis studies [53]. If genome mining and characterization of smBGCs' products are repeated in a positive feedback cycle, it could ultimately be used to design and generate synthetic BGCs for the production of novel bioactive compounds.

CRediT authorship contribution statement

Namil Lee: Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing. **Soonkyu Hwang:** Formal analysis. **Jihun Kim:** Formal analysis. **Suhyung Cho:** Writing - orig-

inal draft. **Bernhard Palsson:** Writing - original draft. **Byung-Kwan Cho:** Conceptualization, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Author contributions

B.-K.C. conceived and supervised the study. N.L., S.H., and J.K. performed the analysis. N.L., S.C., B.P., and B.-K.C. wrote the manuscript.

References

[1] O'Brien J, Wright GD. An ecological perspective of microbial secondary metabolism. Curr Opin Biotechnol 2011;22:552-8. .

[2] Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 1998;5:R245–9. https://doi.org/10.1016/s1074-5521(98)90108-9.

[3] Shore CK, Coukell A. Roadmap for antibiotic discovery. Nat Microbiol 2016;1:16083. https://doi.org/10.1038/nmicrobiol.2016.83.

[4] McClure NS, Day T. A theoretical examination of the relative importance of evolution management and drug development for managing resistance. Proc Biol Sci 2014;281. https://doi.org/10.1098/rspb.2014.1861.

[5] Craney A, Ahmed S, Nodwell J. Towards a new science of secondary metabolism. J Antibiot (Tokyo) 2013;66:387–400. https://doi.org/10.1038/ja.2013.25.

[6] Lee N, Kim W, Hwang S, Lee Y, Cho S, Palsson B, et al. Thirty complete *Streptomyces* genome sequences for mining novel secondary metabolite biosynthetic gene clusters. Sci Data 2020;7:55. https://doi.org/10.1038/s41597-020-0395-9.

[7] Harrison J, Studholme DJ. Recently published *Streptomyces* genome sequences. Microb Biotechnol 2014;7:373–80. https://doi.org/10.1111/1751-7915.12143.

[8] de Jong A, van Hijum SA, Bijlsma JJ, Kok J, Kuipers OP. BAGEL: a web-based bacteriocin genome mining tool. Nucleic Acids Res 2006;34:W273–9. https://doi.org/10.1093/nar/gkl237.

[9] Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and *in silico* prediction of novel chemical structures. Nucleic Acids Res 2008;36:6882–92. https://doi.org/10.1093/nar/gkn685.

[10] Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson DH, et al. CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. J Biotechnol 2009;140:13–7. https://doi.org/10.1016/j.jbiotec.2009.01.007.

[11] Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH. Automated genome mining for natural products. BMC Bioinf 2009;10:185. https://doi.org/10.1186/1471-2105-10-185.

[12] Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. Nucleic Acids Res 2017;45:W49–54. https://doi.org/10.1093/nar/gkx320.

[13] Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 2019;47:W81–7. https://doi.org/10.1093/nar/gkz310.

[14] Belknap KC, Park CJ, Barth BM, Andam CP. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. Sci Rep 2020;10:2003. https://doi.org/10.1038/s41598-020-58904-9.

[15] Lin YS, Kieser HM, Hopwood DA, Chen CW. The chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol Microbiol 1994;14:1103. https://doi.org/10.1111/j.1365-2958.1994.tb01342.x.

[16] Choulet F, Gallois A, Aigle B, Mangenot S, Gerbaud C, Truong C, et al. Intraspecific variability of the terminal inverted repeats of the linear chromosome of *Streptomyces ambofaciens*. J Bacteriol 2006;188:6599–610. https://doi.org/10.1128/JB.00734-06.

[17] Dyson P, Schrempf H. Genetic instability and DNA amplification in *Streptomyces lividans* 66. J Bacteriol 1987;169:4796–803. https://doi.org/10.1128/jb.169.10.4796-4803.1987.

[18] Kim JN, Kim Y, Jeong Y, Roe JH, Kim BG, Cho BK. Comparative genomics reveals the core and accessory genomes of *Streptomyces* species. J Microbiol Biotechnol 2015;25:1599–605. https://doi.org/10.4014/jmb.1504.04008.

[19] Zhou Z, Gu J, Li YQ, Wang Y. Genome plasticity and systems evolution in *Streptomyces*. BMC Bioinf 2012;13(Suppl 10):S8. https://doi.org/10.1186/1471-2105-13-S10-S8.

[20] Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 2002;417:141–7. https://doi.org/10.1038/417141a.

[21] Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, et al. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 2003;21:526–31. https://doi.org/10.1038/nbt820.

[22] Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015;31:3210–2. https://doi.org/10.1093/bioinformatics/btv351.

[23] Studholme DJ. Genome update. Let the consumer beware: *Streptomyces* genome sequence quality. Microb Biotechnol 2016;9:3–7. https://doi.org/10.1111/1751-7915.12344.

[24] Medema MH, Trefzer A, Kovalchuk A, van den Berg M, Muller U, Heijne W, et al. The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biol Evol 2010;2:212–24. https://doi.org/10.1093/gbe/evq013.

[25] Song JY, Jeong H, Yu DS, Fischbach MA, Park HS, Kim JJ, et al. Draft genome sequence of *Streptomyces clavuligerus* NRRL 3585, a producer of diverse secondary metabolites. J Bacteriol 2010;192:6317–8. https://doi.org/10.1128/JB.00859-10.

[26] Hwang S, Lee N, Jeong Y, Lee Y, Kim W, Cho S, et al. Primary transcriptome and translatome analysis determines transcriptional and translational regulatory elements encoded in the *Streptomyces clavuligerus* genome. Nucleic Acids Res 2019;47:6114–29. https://doi.org/10.1093/nar/gkz471.

[27] Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 2020;21:30. https://doi.org/10.1186/s13059-020-1935-5.

[28] Salzberg SL. Next-generation genome annotation: we still struggle to get it right. Genome Biol 2019;20:92. https://doi.org/10.1186/s13059-019-1715-2.

[29] Rudd BA, Hopwood DA. Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). J Gen Microbiol 1979;114:35–43. https://doi.org/10.1099/00221287-114-1-35.

[30] Malpartida F, Hopwood DA. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. Nature 1984;309:462–4. https://doi.org/10.1038/309462a0.

[31] Ikeda H, Kotaki H, Omura S. Genetic studies of avermectin biosynthesis in *Streptomyces avermitilis*. J Bacteriol 1987;169:5615–21. https://doi.org/10.1128/jb.169.12.5615-5621.1987.

[32] Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes - a review. Nat Prod Rep 2016;33:988–1005. https://doi.org/10.1039/c6np00025h.

[33] Pojer F, Li SM, Heide L. Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: new insights into the biosynthesis of aminocoumarin antibiotics. Microbiology 2002;148:3901–11. https://doi.org/10.1099/00221287-148-12-3901.

[34] Zazopoulos E, Huang K, Staffa A, Liu W, Bachmann BO, Nonaka K, et al. A genomics-guided approach for discovering and expressing cryptic metabolic pathways. Nat Biotechnol 2003;21:187–90. https://doi.org/10.1038/nbt784.

[35] Weber T, Kim HU. The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. Synth Syst Biotechnol 2016;1:69–79. https://doi.org/10.1016/j.synbio.2015.12.002.

[36] Hannigan GD, Prihoda D, Palicka A, Soukup J, Klempir O, Rampula L, et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. Nucleic Acids Res 2019;47:. https://doi.org/10.1093/nar/gkz654 e110.

[37] Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 2014;158:412–21. https://doi.org/10.1016/j.cell.2014.06.034.

[38] Doroghazi JR, Metcalf WW. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. BMC Genomics 2013;14:611. https://doi.org/10.1186/1471-2164-14-611.

[39] Waksman SA, Reilly HC, Johnstone DB. Isolation of streptomycin-producing strains of *Streptomyces griseus*. J Bacteriol 1946;52:393–7.

[40] Ehrlich J, Bartz QR, Smith RM, Joslyn DA, Burkholder PR. Chloromycetin, a new antibiotic from a soil Actinomycete. Science 1947;106:417. https://doi.org/10.1126/science.106.2757.417.

[41] Jang KH, Nam SJ, Locke JB, Kauffman CA, Beatty DS, Paul LA, et al. Anthracimycin, a potent anthrax antibiotic from a marine-derived actinomycete. Angew Chem Int Ed Engl 2013;52:7822–4. https://doi.org/10.1002/anie.201302749.

[42] Alt S, Wilkinson B. Biosynthesis of the novel macrolide antibiotic anthracimycin. ACS Chem Biol 2015;10:2468–79. https://doi.org/10.1021/acschembio.5b00525.

[43] Kaweewan I, Komaki H, Hemmi H, Hoshino K, Hosaka T, Isokawa G, et al. Isolation and structure determination of a new cytotoxic peptide, curacozole, from *Streptomyces curacoi* based on genome mining. J Antibiot (Tokyo) 2019;72:1–7. https://doi.org/10.1038/s41429-018-0105-4.

[44] Khater S, Anand S, Mohanty D. *In silico* methods for linking genes and secondary metabolites: the way forward. Synth Syst Biotechnol 2016;1:80–8. https://doi.org/10.1016/j.synbio.2016.03.001.

[45] Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster AL, et al. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). Nucleic Acids Res 2015;43:9645–62. https://doi.org/10.1093/nar/gkv1012.

[46] Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res 2017;45:W36–41. https://doi.org/10.1093/nar/gkx319.

[47] Zhang MM, Wong FT, Wang Y, Luo S, Lim YH, Heng E, et al. CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. Nat Chem Biol 2017. https://doi.org/10.1038/nchembio.2341.

[48] Lee N, Hwang S, Lee Y, Cho S, Palsson B, Cho BK. Synthetic biology tools for novel secondary metabolite discovery in *Streptomyces*. J Microbiol Biotechnol 2019;29:667–86. https://doi.org/10.4014/jmb.1904.04015.

[49] Nah HJ, Pyeon HR, Kang SH, Choi SS, Kim ES. Cloning and heterologous expression of a large-sized natural product biosynthetic gene cluster in *Streptomyces* Species. Front Microbiol 2017;8:394. https://doi.org/10.3389/fmicb.2017.00394.

[50] Bu QT, Yu P, Wang J, Li ZY, Chen XA, Mao XM, et al. Rational construction of genome-reduced and high-efficient industrial *Streptomyces* chassis based on multiple comparative genomic approaches. Microb Cell Fact 2019;18:16. https://doi.org/10.1186/s12934-019-1055-7.

[51] Liu Q, Wu K, Cheng Y, Lu L, Xiao E, Zhang Y, et al. Engineering an iterative polyketide pathway in *Escherichia coli* results in single-form alkene and alkane overproduction. Metab Eng 2015;28:82–90. https://doi.org/10.1016/j.ymben.2014.12.004.

[52] Yuzawa S, Mirsiaghi M, Jocic R, Fujii T, Masson F, Benites VT, et al. Short-chain ketone production by engineered polyketide synthases in *Streptomyces albus*. Nat Commun 2018;9:4569. https://doi.org/10.1038/s41467-018-07040-0.

[53] Eng CH, Backman TWH, Bailey CB, Magnan C, Garcia Martin H, Katz L, et al. ClusterCAD: a computational platform for type I modular polyketide synthase design. Nucleic Acids Res 2018;46:D509–15. https://doi.org/10.1093/nar/gkx893.

[54] Shao L, Zi J, Zeng J, Zhan J. Identification of the herboxidiene biosynthetic gene cluster in *Streptomyces chromofuscus* ATCC 49982. Appl Environ Microbiol 2012;78:2034–8. https://doi.org/10.1128/AEM.06904-11.

[55] Hao C, Huang S, Deng Z, Zhao C, Yu Y. Mining of the pyrrolamide antibiotics analogs in *Streptomyces netropsis* reveals the amidohydrolase-dependent "iterative strategy" underlying the pyrrole polymerization. PLoS ONE 2014;9:. https://doi.org/10.1371/journal.pone.0099077 e99077.

[56] Yue C, Niu J, Liu N, Lu Y, Liu M, Li Y. Cloning and identification of the lobophorin biosynthetic gene cluster from marine *Streptomyces olivaceus* strain FXJ7.023. Pak J Pharm Sci 2016;29:287–93.

[57] Du D, Katsuyama Y, Onaka H, Fujie M, Satoh N, Shin-Ya K, et al. Production of a novel amide-containing polyene by activating a cryptic biosynthetic gene cluster in *Streptomyces* sp. MSC090213JE08. ChemBioChem 2016;17:1464–71. https://doi.org/10.1002/cbic.201600167.

[58] Castro JF, Razmilic V, Gomez-Escribano JP, Andrews B, Asenjo JA, Bibb MJ. Identification and heterologous expression of the chaxamycin biosynthesis gene cluster from *Streptomyces leeuwenhoekii*. Appl Environ Microbiol 2015;81:5820–31. https://doi.org/10.1128/AEM.01039-15.

[59] Jordan PA, Moore BS. Biosynthetic pathway connects cryptic ribosomally synthesized posttranslationally modified peptide genes with pyrroloquinoline alkaloids. Cell Chem Biol 2016;23:1504–14. https://doi.org/10.1016/j.chembiol.2016.10.009.

[60] Li J, Xie Z, Wang M, Ai G, Chen Y. Identification and analysis of the paulomycin biosynthetic gene cluster and titer improvement of the paulomycins in *Streptomyces paulus* NRRL 8115. PLoS ONE 2015;10:. https://doi.org/10.1371/journal.pone.0120542 e0120542.

[61] Amagai K, Ikeda H, Hashimoto J, Kozone I, Izumikawa M, Kudo F, et al. Identification of a gene cluster for telomestatin biosynthesis and heterologous expression using a specific promoter in a clean host. Sci Rep 2017;7:3382. https://doi.org/10.1038/s41598-017-03308-5.

[62] Wu H, Liu W, Shi L, Si K, Liu T, Dong D, et al. Comparative genomic and regulatory analyses of natamycin production of *Streptomyces lydicus* A02. Sci Rep 2017;7:9114. https://doi.org/10.1038/s41598-017-09532-3.

[63] Paulus Constanze, Rebets Yuriy, Tokovenko Bogdan, Nadmid Suvd, Terekhova Larisa P, Myronovskyi Maksym, Zotchev Sergey B, Rückert Christian, Braig Simone, Zahler Stefan, Kalinowski Jörn, Luzhetskyy Andriy. New natural

products identified by combined genomics-metabolomics profiling of marine *Streptomyces* sp. MP131-18. Sci Rep 2017;7(1). https://doi.org/10.1038/srep42382.

[64] Low ZJ, Pang LM, Ding Y, Cheang QW, Le Mai Hoang K, Thi Tran H, et al. Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a *Streptomyces* strain isolated from mangrove sediment. Sci Rep 2018;8:1594. https://doi.org/10.1038/s41598-018-20018-8.

[65] Yu Y, Tang B, Dai R, Zhang B, Chen L, Yang H, et al. Identification of the streptothricin and tunicamycin biosynthetic gene clusters by genome mining in *Streptomyces* sp. strain fd1-xmd. Appl Microbiol Biotechnol 2018;102:2621–33. https://doi.org/10.1007/s00253-018-8748-4.

[66] Tu J, Li S, Chen J, Song Y, Fu S, Ju J, et al. Characterization and heterologous expression of the neoabyssomicin/abyssomicin biosynthetic gene cluster from *Streptomyces koyangensis* SCSIO 5802. Microb Cell Fact 2018;17:28. https://doi.org/10.1186/s12934-018-0875-1.

[67] Song F, Liu N, Liu M, Chen Y, Huang Y. Identification and characterization of mycemycin biosynthetic gene clusters in *Streptomyces olivaceus* FXJ8.012 and *Streptomyces* sp. FXJ1.235. Mar Drugs 2018;16. https://doi.org/10.3390/md16030098.

[68] Wolf F, Leipoldt F, Kulik A, Wibberg D, Kalinowski J, Kaysser L. Characterization of the actinonin biosynthetic gene cluster. ChemBioChem 2018. https://doi.org/10.1002/cbic.201800116.

[69] Twigg FF, Cai W, Huang W, Liu J, Sato M, Perez TJ, et al. Identifying the biosynthetic gene cluster for triacsins with an N-hydroxytriazene moiety. ChemBioChem 2019;20:1145–9. https://doi.org/10.1002/cbic.201800762.

[70] Martinet L, Naome A, Deflandre B, Maciejewska M, Tellatin D, Tenconi E, et al. A single biosynthetic gene cluster is responsible for the production of bagremycin antibiotics and ferroverdin iron chelators. mBio 2019;10. .

[71] Ozaki T, Sugiyama R, Shimomura M, Nishimura S, Asamizu S, Katsuyama Y, et al. Identification of the common biosynthetic gene cluster for both antimicrobial streptoaminals and antifungal 5-alkyl-1,2,3,4-tetrahydroquinolines. Org Biomol Chem 2019;17:2370–8. https://doi.org/10.1039/c8ob02846j.

[72] Ye J, Zhu Y, Hou B, Wu H, Zhang H. Characterization of the bagremycin biosynthetic gene cluster in *Streptomyces* sp. Tu 4128. Biosci Biotechnol Biochem 2019;83:482–9. https://doi.org/10.1080/09168451.2018.1553605.

[73] Perez-Victoria I, Oves-Costales D, Lacret R, Martin J, Sanchez-Hidalgo M, Diaz C, et al. Structure elucidation and biosynthetic gene cluster analysis of caniferolides A-D, new bioactive 36-membered macrolides from the marine-derived *Streptomyces caniferus* CA-271066. Org Biomol Chem 2019;17:2954–71. https://doi.org/10.1039/c8ob03115k.

[74] Zhou S, Song L, Masschelein J, Sumang FAM, Papa IA, Zulaybar TO, et al. Pentamycin biosynthesis in Philippine *Streptomyces* sp. S816: Cytochrome P450-catalyzed installation of the C-14 hydroxyl group. ACS Chem Biol 2019;14:1305–9. https://doi.org/10.1021/acschembio.9b00270.

[75] Sanchez-Hidalgo M, Martin J, Genilloud O. Identification and heterologous expression of the biosynthetic gene cluster encoding the lasso peptide humidimycin, a caspofungin activity potentiator. Antibiotics (Basel) 2020;9. https://doi.org/10.3390/antibiotics9020067.

[76] Kaweewan I, Hemmi H, Komaki H, Kodani S. Isolation and structure determination of a new antibacterial peptide pentaminomycin C from *Streptomyces cacaoi* subsp. cacaoi. J Antibiot (Tokyo) 2020;73:224–9. https://doi.org/10.1038/s41429-019-0272-y.

[77] Lautru S, Deeth RJ, Bailey LM, Challis GL. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. Nat Chem Biol 2005;1:265–9. https://doi.org/10.1038/nchembio731.

[78] Song L, Barona-Gomez F, Corre C, Xiang L, Udwary DW, Austin MB, et al. Type III polyketide synthase beta-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by *Streptomyces coelicolor* genome mining. J Am Chem Soc 2006;128:14754–5. https://doi.org/10.1021/ja065247w.

[79] Goto Y, Li B, Claesen J, Shi Y, Bibb MJ, van der Donk WA. Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. PLoS Biol 2010;8:. https://doi.org/10.1371/journal.pbio.1000339e1000339.

[80] Laureti L, Song L, Huang S, Corre C, Leblond P, Challis GL, et al. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. Proc Natl Acad Sci U S A 2011;108:6258–63. https://doi.org/10.1073/pnas.1019077108.

[81] Rackham EJ, Gruschow S, Goss RJ. Revealing the first uridyl peptide antibiotic biosynthetic gene cluster and probing pacidamycin biosynthesis. Bioeng Bugs 2011;2:218–21. https://doi.org/10.4161/bbug.2.4.15877.

[82] Zhang H, Wang H, Wang Y, Cui H, Xie Z, Pu Y, et al. Genomic sequence-based discovery of novel angucyclinone antibiotics from marine *Streptomyces* sp. W007. FEMS Microbiol Lett 2012;332:105–12. https://doi.org/10.1111/j.1574-6968.2012.02582.x.

[83] Park HM, Kim BG, Chang D, Malla S, Joo HS, Kim EJ, et al. Genome-based cryptic gene discovery and functional identification of NRPS siderophore peptide in *Streptomyces peucetius*. Appl Microbiol Biotechnol 2013;97:1213–22. https://doi.org/10.1007/s00253-012-4268-9.

[84] Meguro A, Tomita T, Nishiyama M, Kuzuyama T. Identification and characterization of bacterial diterpene cyclases that synthesize the cembrane skeleton. ChemBioChem 2013;14:316–21. https://doi.org/10.1002/cbic.201200651.

[85] Mohimani H, Kersten RD, Liu WT, Wang M, Purvine SO, Wu S, et al. Automated genome mining of ribosomal peptide natural products. ACS Chem Biol 2014;9:1545–51. https://doi.org/10.1021/cb500199h.

[86] Iftime D, Jasyk M, Kulik A, Imhoff JF, Stegmann E, Wohlleben W, et al. Streptocollin, a Type IV lanthipeptide produced by *Streptomyces collinus* Tu 365. ChemBioChem 2015;16:2615–23. https://doi.org/10.1002/cbic.201500377.

[87] Elsayed SS, Trusch F, Deng H, Raab A, Prokes I, Busarakam K, et al. Chaxapeptin, a lasso peptide from extremotolerant *Streptomyces leeuwenhoekii* strain C58 from the hyperarid atacama desert. J Org Chem 2015;80:10252–60. https://doi.org/10.1021/acs.joc.5b01878.

[88] Park OK, Choi HY, Kim GW, Kim WG. Generation of new complestatin analogues by heterologous expression of the complestatin biosynthetic gene cluster from *Streptomyces chartreusis* AN1542. ChemBioChem 2016;17:1725–31. https://doi.org/10.1002/cbic.201600241.

[89] Thanapipatsiri A, Gomez-Escribano JP, Song L, Bibb MJ, Al-Bassam M, Chandra G, et al. Discovery of unusual biaryl polyketides by activation of a silent *Streptomyces venezuelae* biosynthetic gene cluster. ChemBioChem 2016;17:2189–98. https://doi.org/10.1002/cbic.201600396.

[90] Remali J, Sarmin NM, Ng CL, Tiong JJL, Aizat WM, Keong LK, et al. Genomic characterization of a new endophytic *Streptomyces kebangsaanensis* identifies biosynthetic pathway gene clusters for novel phenazine antibiotic production. PeerJ 2017;5:. https://doi.org/10.7717/peerj.3738e3738.

[91] Ma J, Huang H, Xie Y, Liu Z, Zhao J, Zhang C, et al. Biosynthesis of ilamycins featuring unusual building blocks and engineered production of enhanced anti-tuberculosis agents. Nat Commun 2017;8:391. https://doi.org/10.1038/s41467-017-00419-5.

[92] Ye S, Molloy B, Brana AF, Zabala D, Olano C, Cortes J, et al. Identification by genome mining of a type I polyketide gene cluster from *Streptomyces argillaceus* involved in the biosynthesis of pyridine and piperidine alkaloids argimycins P. Front Microbiol 2017;8:194. https://doi.org/10.3389/fmicb.2017.00194.

[93] Pait IGU, Kitani S, Roslan FW, Ulanova D, Arai M, Ikeda H, et al. Discovery of a new diol-containing polyketide by heterologous expression of a silent biosynthetic gene cluster from *Streptomyces lavendulae* FRI-5. J Ind Microbiol Biotechnol 2018;45:77–87. https://doi.org/10.1007/s10295-017-1997-x.

[94] Schneider O, Simic N, Aachmann FL, Ruckert C, Kristiansen KA, Kalinowski J, et al. Genome mining of *Streptomyces* sp. YIM 130001 isolated from lichen affords new thiopeptide antibiotic. Front Microbiol 2018;9:3139. https://doi.org/10.3389/fmicb.2018.03139.

[95] Suroto DA, Kitani S, Arai M, Ikeda H, Nihira T. Characterization of the biosynthetic gene cluster for cryptic phthoxazolin A in *Streptomyces avermitilis*. PLoS ONE 2018;13:. https://doi.org/10.1371/journal.pone.0190973e0190973.

[96] Liu W, Sun F, Hu Y. Genome mining-mediated discovery of a new avermipeptin analogue in *Streptomyces actuosus* ATCC 25421. ChemistryOpen 2018;7:558–61. https://doi.org/10.1002/open.201800130.

[97] Xu XN, Chen LY, Chen C, Tang YJ, Bai FW, Su C, et al. Genome mining of the marine Actinomycete *Streptomyces* sp. DUT11 and discovery of tunicamycins as anti-complement agents. Front Microbiol 2018;9:;1318. https://doi.org/10.3389/fmicb.2018.01318.

[98] Rodriguez Estevez M, Myronovskyi M, Gummerlich N, Nadmid S, Luzhetskyy A. Heterologous expression of the nybomycin gene cluster from the marine strain *Streptomyces albus* subsp. chlorinus NRRL B-24108. Mar Drugs 2018;16. https://doi.org/10.3390/md16110435.

[99] Gosse JT, Ghosh S, Sproule A, Overy D, Cheeptham N, Boddy CN. Whole genome sequencing and metabolomic study of cave *Streptomyces* Isolates ICC1 and ICC4. Front Microbiol 2019;10:1020. https://doi.org/10.3389/fmicb.2019.01020.

[100] Thomy D, Culp E, Adamek M, Cheng EY, Ziemert N, Wright GD, et al. The ADEP biosynthetic gene cluster in *Streptomyces hawaiiensis* NRRL 15010 reveals an accessory clpP gene as a novel antibiotic resistance factor. Appl Environ Microbiol 2019;85. https://doi.org/10.1128/AEM.01292-19.

[101] Sun C, Yang Z, Zhang C, Liu Z, He J, Liu Q, et al. Genome mining of *Streptomyces atratus* SCSIO ZH16: Discovery of atratumycin and identification of its biosynthetic gene cluster. Org Lett 2019;21:1453–7. https://doi.org/10.1021/acs.orglett.9b00208.

[102] Gomez-Escribano JP, Castro JF, Razmilic V, Jarmusch SA, Saalbach G, Ebel R, et al. Heterologous expression of a cryptic gene cluster from *Streptomyces leeuwenhoekii* C34(T) yields a novel lasso peptide, leepeptin. Appl Environ Microbiol 2019;85. https://doi.org/10.1128/AEM.01752-19.

[103] Zhang C, Ding W, Qin X, Ju J. Genome sequencing of *Streptomyces olivaceus* SCSIO T05 and activated production of lobophorin CR4 via metabolic engineering and genome mining. Mar Drugs 2019;17. https://doi.org/10.3390/md17100593.

[104] Qian Z, Bruhn T, D'Agostino PM, Herrmann A, Haslbeck M, Antal N, et al. Discovery of the streptoketides by direct cloning and rapid heterologous expression of a cryptic PKS II gene cluster from *Streptomyces* sp. Tu 6314. J Org Chem 2020;85:664–73. https://doi.org/10.1021/acs.joc.9b02741.