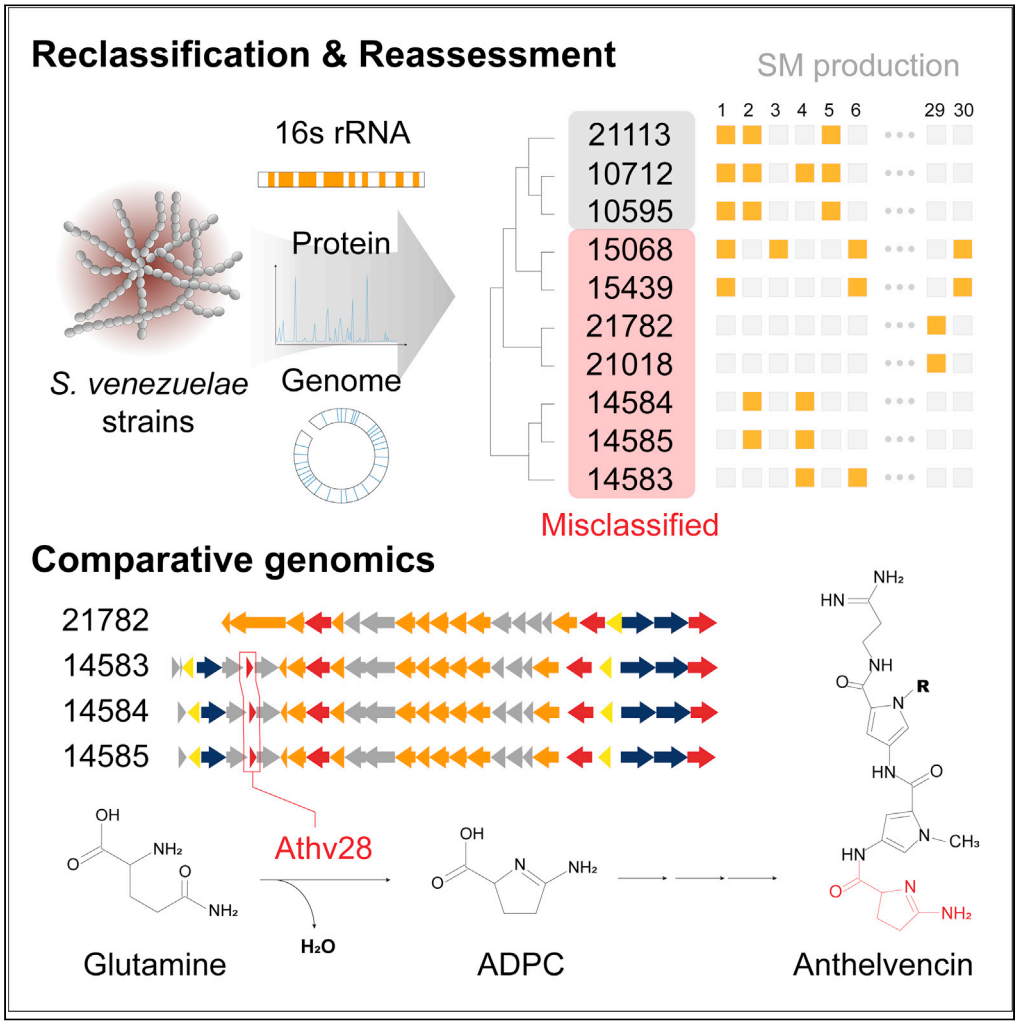
iScience

CellPress
OPEN ACCESS

Article

Re-classification of *Streptomyces venezuelae* strains and mining secondary metabolite biosynthetic gene clusters



Namil Lee, Mira Choi, Woori Kim, ..., Bernhard Palsson, Kyoung-Soon Jang, Byung-Kwan Cho

ksjang@kbsi.re.kr (K.-S.J.)
bcho@kaist.ac.kr (B.-K.C.)

Highlights

Seven out of ten *Streptomyces venezuelae* strains were misclassified

59secondary metabolites productions were identified from the ten strains

Athv28 converts glutamine to ADPC, the most important anthelvencin precursor

Lee et al., iScience *24*, 103410
December 17, 2021
https://doi.org/10.1016/j.isci.2021.103410

iScience

CellPress
OPEN ACCESS

Article

Re-classification of *Streptomyces venezuelae* strains and mining secondary metabolite biosynthetic gene clusters

Namil Lee,[1,2,8] Mira Choi,[6,8] Woori Kim,[1,2,8] Soonkyu Hwang,[1,2] Yongjae Lee,[1,2] Ji Hun Kim,[1,2] Gahyeon Kim,[1,2] Hyeseong Kim,[1,2] Suhyung Cho,[1,2] Sun Chang Kim,[1] Bernhard Palsson,[3,4,5] Kyoung-Soon Jang,[6,7,*] and Byung-Kwan Cho[1,2,9,*]

SUMMARY

***Streptomyces* species have attracted considerable interest as a reservoir of medically important secondary metabolites, which are even diverse and different between strains. Here, we reassess ten *Streptomyces venezuelae* strains by presenting the highly resolved classification, using 16S rRNA sequencing, MALDI-TOF MS protein profiling, and whole-genome sequencing. The results revealed that seven of the ten strains were misclassified as *S. venezuelae* species. Secondary metabolite biosynthetic gene cluster (smBGC) mining and targeted LC-MS/MS based metabolite screening of *S. venezuelae* and misclassified strains identified in total 59 secondary metabolites production. In addition, a comparison of pyrrolamide-type antibiotic BGCs of four misclassified strains, followed by functional genomics, revealed that *athv28* is critical in the synthesis of the anthelvencin precursor, 5-amino-3,4-dihydro-2H-pyrrole-2-carboxylate (ADPC). Our findings illustrate the importance of the accurate classification and better utilization of misclassified *Streptomyces* strains to discover smBGCs and their secondary metabolite products.**

INTRODUCTION

Streptomyces bacteria comprise the largest genus of the phylum Actinobacteria, which have attracted considerable interest because of their ability to produce various secondary metabolites, such as antibiotics, anticancer drugs, anthelmintic enzymes, antifungal agents, and herbicides (Embley and Stackebrandt, 1994; Berdy, 2005; Hopwood, 2007). Particularly, more than 70% of clinically relevant antibiotics are currently produced by *Streptomyces* (Law et al., 2019). On average, approximately 30 secondary metabolite biosynthetic gene clusters (smBGCs) are encoded in each *Streptomyces* genome, and most of these smBGCs are species-specific (Omura et al., 2001; Bentley et al., 2002). Moreover, recent genome mining studies have shown high chemical diversity even within closely related *Streptomyces* strains (Seipke, 2015; Antony-Babu et al., 2017; Park and Andam, 2019; Sottorff et al., 2019; Belknap et al., 2020; Kim et al., 2020), thereby emphasizing the importance of genome mining at the strain level.

Although morphological and physiological observations are still being used as a standard procedure for the classification of *Streptomyces* (Chandra and Chater, 2014; Li et al., 2016), the similar morphologies between *Streptomyces* species often lead to the misclassification of the *Streptomyces* genus (Kim et al., 2012). Although 16S rRNA sequence-based phylogeny has been considered as an alternative for genus-level identification, the resolution of 16S rRNA gene sequencing is limited for species or strain-level classification (Meklat et al., 2011; Kim et al., 2012). In particular, the heterogeneity of 16S rRNA sequences among the multiple copies in the genome and a low correlation with the DNA–DNA relatedness are drawbacks of the 16S rRNA gene sequencing method (Kim et al., 2012). To establish a robust phylogeny of *Streptomyces* at the species level, a standardized pipeline for classification with higher discriminatory power than that provided by the 16S rRNA sequence-based method is required.

An accurate classification of *Streptomyces* strains is crucial in two respects. First, taxonomically close strains or species are likely to possess similar smBGCs, conserved regulatory systems, and common precursors from a primary metabolism (Zakrzewski et al., 2012; Romero-Rodriguez et al., 2015; Xu et al., 2019).

[1]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

[2]Innovative Biomaterials Research Center, KI for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

[3]Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA

[4]Department of Pediatrics, University of California San Diego, La Jolla, CA 92093, USA

[5]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby 2800, Denmark

[6]Bio-Chemical Analysis Team, Korea Basic Science Institute, Cheongju 28119, Republic of Korea

[7]Division of Bio-Analytical Science, University of Science and Technology, Daejeon 34113, Republic of Korea

[8]These authors contributed equally

[9]Lead contact

*Correspondence: ksjang@kbsi.re.kr (K.-S.J.), bcho@kaist.ac.kr (B.-K.C.)

https://doi.org/10.1016/j.isci.2021.103410

Therefore, utilization of an accurate phylogenetic analysis is vital in comparative genomics to identify genetic factors that influence the secondary metabolites production and host selections for heterologous expressions (Cho et al., 2019; Myronovskyi and Luzhetskyy, 2019; Kim et al., 2020). Second, misclassified *Streptomyces* strains produce distinct pools of secondary metabolites (Doroghazi and Metcalf, 2013). Thus, a reassessment is needed to fully exploit the underestimated potential of misclassified strains with regards to new secondary metabolites discovery.

Among the many *Streptomyces* species, *S. venezuelae* has emerged as an efficient host for heterologous expression and developmental studies because of its fast growth rate, ease of genetic manipulation, lack of mycelial clumping, and sporulation in liquid media (Bush et al., 2015; Phelan et al., 2017). Among a total of 29 *S. venezuelae* strains that have been isolated to date, *S. venezuelae* ATCC 10712 has been extensively studied with regard to secondary metabolite production and developmental regulation. The strain has been noted to produce chloramphenicol for the treatment of bacterial infections (Vitayakritsirikul et al., 2016), and jadomycin exhibiting antimicrobial activities and breast cancer cytotoxicity (Kim et al., 2020). Also, *S. venezuelae* ATCC 15439 encodes the pikromycin BGC, which has been extensively studied because of its unique ability to produce diverse macrolides from one cluster (Xue et al., 1998). Unlike *S. venezuelae* ATCC 10712 and ATCC 15439 strains, the smBGCs and related characteristics of other *S. venezuelae* strains remain largely unknown. Indeed, several strains were classified as *S. venezuelae* species through morphology and carbon utilization characterization about 50 years ago (Probst et al., 1964; Meyers et al., 1972). Therefore, more accurate and reliable classifications are needed to better understand their physiology and ability to produce secondary metabolites.

In this study, we reassessed ten *S. venezuelae* strains along with 15 strains of other *Streptomyces* species. First, we determined the taxonomy of the strains using 16S rRNA sequence and MALDI-TOF MS profile-based phylogenetic analysis, resulting in different phylogenies in the classification of several strains. To further validate the classification, whole-genome alignment was conducted supporting that MALDI-TOF MS protein profile-based classification provides higher resolution than 16S rRNA-based classification for determining the phylogenetic distance between *Streptomyces* species. As a result, seven misclassified *S. venezuelae* strains were identified. Furthermore, using smBGC mining followed by LC-MS/MS-based metabolite screening, we identified 59 secondary metabolites produced by the three *S. venezuelae* and seven misclassified strains, including 45 previously not reported cases. Seven non-*S. venezuelae* strains produced highly distinct secondary metabolites compared to the three *S. venezuelae* strains. Specifically, the comparison of pyrrolamide-type antibiotic BGCs from four phylogenetically close non-*S. venezuelae* strains in tandem with functional genomics studies demonstrated the enzymatic reaction required for biosynthesizing the anthelvencin precursor, 5-amino-3,4-dihydro-2H-pyrrole-2-carboxylate (ADPC). The accurate classification and re-evaluation of the ten *Streptomyces* strains belonging to the *S. venezuelae* species expanded the genetic repositories for new secondary metabolite discovery and elucidated secondary metabolite biosynthetic processes by comparing closely related strains.

RESULTS

Phenotypic characteristics of *S. venezuelae* strains

To date, 29 *Streptomyces* have been categorized as *S. venezuelae* strains on the basis of the similarity in their morphology and carbon utilization (Ehrlich et al., 1948; Probst et al., 1965; Kim et al., 2020). In order to reassess this classification, we first compared phenotypic characteristics of ten *S. venezuelae* strains commercially available from ATCC (listed in Table 1). Hereafter, the strain names of *S. venezuelae* are denoted as their ATCC numbers. All strains showed distinct colony formations, and after 270 h growth on solid MYM media distinct pigment production and spore formation were observed (Figure 1A). Next, ten strains were cultured in MYM liquid media and doubling times were measured. As a result, their doubling times ranged from 54.69 min (ATCC 10712) to 70.52 min (ATCC 21782) at exponential growth phase (Figure 1B). These strains were then categorized into two groups based on the maximum cell growth, representing the difference in their ability to utilize maltose, the main carbon source of MYM media (Figure 1B). These observations demonstrate the distinct characteristics of these strains in morphological features and carbon utilization, thus requiring additional identification of the strains.

Phylogenetic analyses based on the 16S rRNA sequences

To investigate the phylogenetic relationship of the *S. venezuelae* strains, we determined their 16S rRNA sequences (Figure S1 and Table S1). Applying 97% identity of the 16S rRNA sequence as canonical

Table 1. The features of ten *S. venezuelae* strains used in this study

Strain number	Features	Origin	Reference
ATCC 10712	Chloramphenicol and jadomycin B producer	Soil from Carcas, Venezuela	(Ehrlich et al., 1948)
ATCC 15439	Methymycin and pikromycin producer	Florence, Italy	(Dutcher et al., 1956)
ATCC 15068	Methymycin producer	Oswego, New York, USA	(Dutcher et al., 1956)
ATCC 21113	Xylose isomerase and glucose isomerase producer	Unknown	(Iizuka et al., 1969)
ATCC 10595	Chloramphenicol producer	Soil from Urbana, Illinois, USA	(La Farina et al., 1996)
ATCC 21782	Antibiotic EM-98 producer	Soil from Florida, USA	(Meyers et al., 1972)
ATCC 21018	3-methylthiopropylamine and methionine decarboxylase producer	Soil	(Nakayama and Hagino, 1968)
ATCC 14583	Substilisin inhibitor and anthelvencin A producer	Soil	(Probst et al., 1964, 1965)
ATCC 14584	Anthelvencin A producer	Soil	(Probst et al., 1964, 1965)
ATCC 14585	Anthelvencin A producer	Soil	(Probst et al., 1964, 1965)

threshold for species delimitation (Stackebrandt and Goebel, 1994), all strains were classified as identical species (Figure 2A). However, the inaccuracy of the criterion has been recently brought forth, and a suggestion of 99% identity as a new threshold has been made (Edgar, 2018). According to the 99% threshold, only five strains (ATCC 10595, 10712, 15068, 15439, and 21113), containing two representative *S. venezuelae* type strains (ATCC 10712 and 15439), were classified as *S. venezuelae* species (Figure 2A). Among the remaining five strains, only ATCC 21018 and 21782 strains were the same species, whereas ATCC 14583, 14584, and 14585 strains were supposed to be different species.

Additional 16S rRNA sequence-based phylogenetic analysis of the 15 strains from other *Streptomyces* species (Table 2) further supported a misclassification of some of the strains. With the 99% threshold, four clades existed, and streptomycetes belonging to each clade could be regarded as identical species (Figure 2B). The first clade consists of five *S. venezuelae* strains, including two type strains. These strains were phylogenetically closer to *S. cinereoruber* and *S. filamentosus* than the remaining five *S. venezuelae* strains. Actually, according to the NCBI taxonomy database, *S. filamentosus* is a subspecies of *S. venezuelae* and is also called *S. venezuelae* subsp. *roseospori*, indicating the phylogenetically close relationship with *S. venezuelae*. Among the five remaining *S. venezuelae* strains, ATCC 21018 and 21782 strains were grouped in the second clade along with *S. subrutilus* and *S. vinaceus*, showing that these two strains are distinct from the *S. venezuelae* species. Moreover, the ATCC 14583 strain belonged to the third clade with *S. galilaeus*, *S. chartreusis*, and *S. alboniger*; ATCC 14584 and 14585 strains did not cluster with any other species. The fourth clade consisted of *S. coelicolor* and *S. lividans*, a representative example of closely related streptomycetes. Although they have been widely accepted as discrete species, they have to be classified as the same species according to the current taxonomic criteria. Thus, 16S rRNA sequence-based phylogenetic analysis unveiled misclassifications in *Streptomyces* taxonomy and, particularly, five out of ten *S. venezuelae* strains are likely to be distinct species and not *S. venezuelae*.

Phylogenetic analyses based on MALDI-TOF MS protein profiles

Although 16S rRNA sequence-based phylogenetic analysis is considered to be a suitable method for genus-level identification, low phylogenetic power at the species and strain level is a drawback (Kim et al., 2012). In particular, resolution problems at the species level have been found in several groups of bacteria, including *Actinomyces* (Janda and Abbott, 2007). In the past decades, MALDI-TOF MS-based whole-cell protein profiling has emerged as a powerful tool in microbiology because of its capability to identify and differentiate microorganisms (Claydon et al., 1996; Evason et al., 2001; Kostrzewa et al., 2013). To determine the phylogenetic distance between *S. venezuelae* strains based on their whole-cell protein profiles, we analyzed and compared the total protein profiles of 25 streptomycetes grown in ISP2 liquid media. The profiles were acquired in a mass range of m/z 2,000 to 12,000 using MALDI-TOF MS (Figure S2A). A series of peaks in the range of 4,000 to 6,000 m/z were abundantly presented in almost all of the protein profiles of the strains. To evaluate the reliability of this method, each mass spectrum was subjected to Biotyper 3 microbial identification. The MALDI-TOF MS spectra of 25 streptomycetes were completely identified by their corresponding main spectra profiles (MSPs) as a first hit with confident ID scores ($\geq$2.5), although some strains showed very similar mass spectral patterns (Table S2).

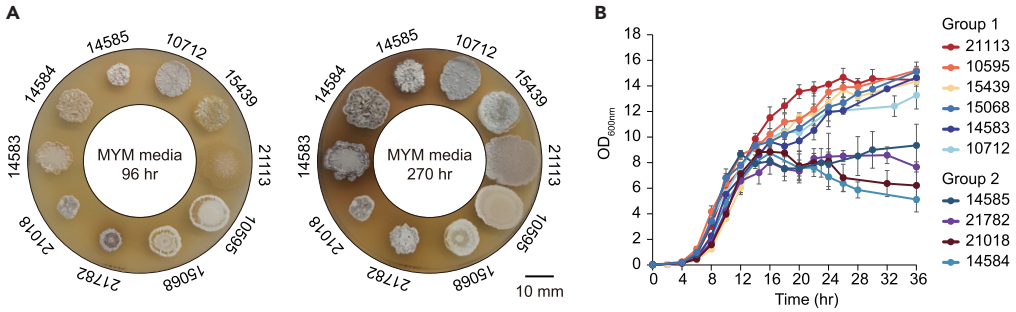


Figure 1. Phenotypic characteristics of ten *Streptomyces venezuelae* strains

(A) Colony proliferation shape and pigment production on solid MYM media.

(B) The growth profile of ten *S. venezuelae* strains in MYM liquid media. Data are represented as mean value and error bars indicate standard deviation (s.d.) of three biological replicates.

Next, a phylogenetic dendrogram of 25 streptomycetes was generated using principal component analysis (PCA) by calculating the Euclidean distance between MSPs of 25 streptomycetes (Figures 2C and S2B). According to the canonical threshold of ID score for species-level identification (≥2.0) (Mörtelmaier et al., 2019), five *S. venezuelae* strains (ATCC 21113, 10712, 10595, 15439, and 15068) were considered as identical species, consistent with the results of the 16S rRNA sequence-based classification. In the case of ATCC 21018 and 21782 strains, although they were not regarded as the same species, they were consistently clustered with *S. subrutilus* and *S. vinaceus*. Moreover, MALDI-TOF MS protein profile-based classification identified ATCC 14583, 14584, and 14585 strains as identical species (Table S2), whereas 16S rRNA sequence-based classification revealed that these three strains were distantly related. Additionally, this result conflicted with the ATCC 14583 strain being considered to be the same species as the other *Streptomyces* species based on the 16S rRNA sequence similarity (Figures 2B and 2C). In summary, MALDI-TOF MS protein profile-based classification determined that only five out of ten *S. venezuelae* strains are *S. venezuelae* species, which is consistent with 16S rRNA sequence-based classification results. However, phylogeny of the remaining five *S. venezuelae* strains was clearly different between the 16S rRNA sequence- and MALDI-TOF MS protein profile-based classifications.

Genome sequence-based phylogenomics analysis

Phylogenomics based on the whole-genome sequence comparison is the most robust taxonomic method and is considered to be the gold standard for species-level taxonomic circumscription (Lalucat et al., 2020). To determine the indubitable species circumscription of *S. venezuelae* species and evaluate the phylogeny based on 16S rRNA sequences and MALDI-TOF MS protein profiles, we compared the complete genome sequences of 25 streptomycetes. The full genome sequences of 21 streptomycetes used in this study were recently determined by our group using a hybrid strategy, exploiting both long-read and short-read sequencing methods (Lee et al., 2020), whereas the genome sequences of the remaining four streptomycetes were obtained from the NCBI database (Table 2). The MUMmer ultra-rapid aligning tool (ANIm) values between 25 genome sequences were calculated using PYANI v0.2.9 (Pritchard et al., 2016) because ANIm is considered to provide more accurate results than the BLAST algorithm (ANIb) when the genome sequences are highly similar (Richter and Rossello-Mora, 2009).

Clustering 25 streptomycetes based on the ANIm values determined rigorous species circumscription of ten *S. venezuelae* strains (Figure 2D). First, five *S. venezuelae* strains, which were determined to be identical species based on both 16S rRNA sequence-based and MALDI-TOF MS protein profile-based classification results, were clearly separated into G1 group (ATCC 10595, 10712, and 21113 strains) and G2 group (ATCC 15439 and 15068 strains). According to the common threshold of ANIm value for species delimitation (>95%) (Richter and Rossello-Mora, 2009), only strains in the G1 group were thought to be of the same species. In fact, highly distinct secondary metabolite production was reported for strains in G1 and G2 groups; strains in G1 group are known to produce chloramphenicol and jadomycin, whereas those in the G2 group produce pikromycin derivatives (Xue and Sherman, 2001; Kim et al., 2020). Collectively, only chloramphenicol producers (ATCC 10712, 21113, and 10595 strains) can be classified as *S. venezuelae* species. Although the ATCC 15439 strain has been widely accepted as a type strain of *S. venezuelae* species with the ATCC 10712 strain, it has to be classified as a different species. Note that the ATCC 15439 strain was originally

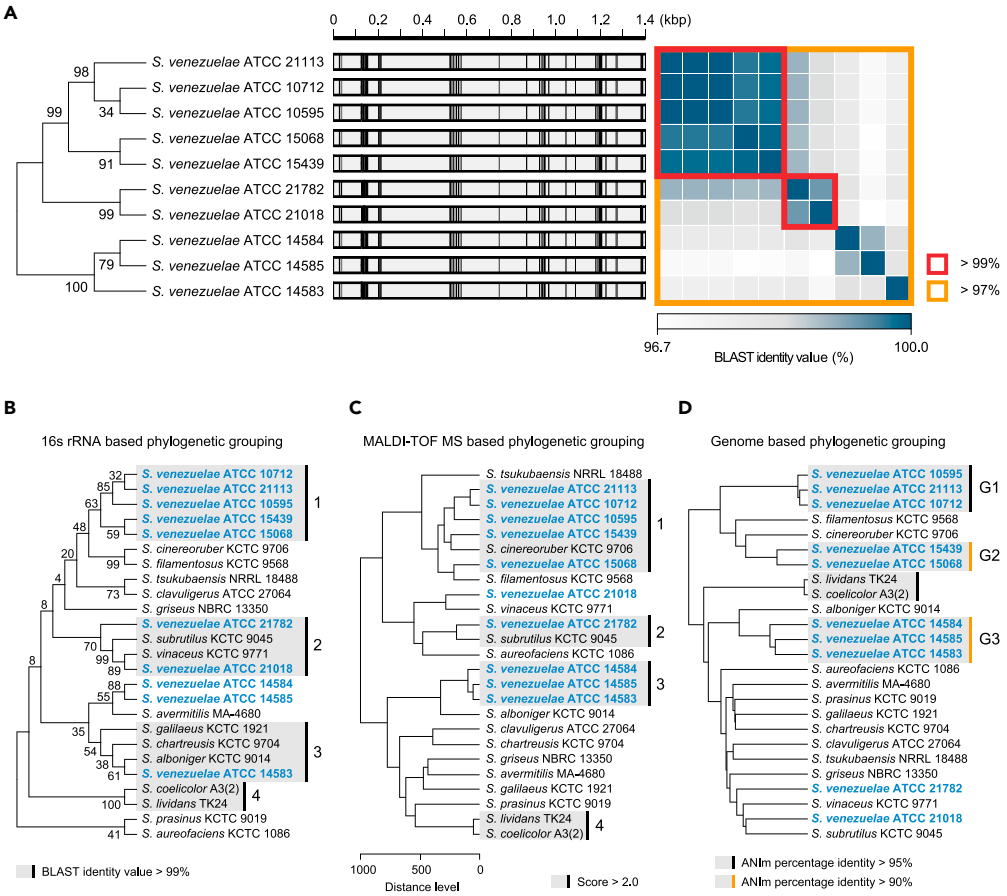


Figure 2. Phylogenetic analysis of ten *Streptomyces venezuelae* strains with other streptomycetes

(A) 16S rRNA sequence-based distance between ten *Streptomyces venezuelae* strains. Numbers at node indicate percentage of bootstrap support.

(B) 16S rRNA sequence-based phylogenetic analysis. Four clades were detected with the BLAST identity value of more than 99%. Numbers at node indicate percentage of bootstrap support.

(C) MALDI-TOF MS protein fingerprinting-based phylogenetic analysis. Four clades were detected with a Biotyper score value of more than 2.0.

(D) Phylogenomic analysis. Eight *S. venezuelae* strains were clustered into three clades, G1 with the ANIm percentage identity of more than 95% and G2 and G3 with the ANIm percentage identity of more than 90%. The remaining two strains were not clustered with any streptomycetes.

described as *S. venezuelae* in the NCBI database, but was recently renamed *S. gardneri*. Second, genome sequences of the strains in the G3 group were largely different from those in G1 and G2 groups (ANIb <78%), consistent with both 16S rRNA sequence-based and MALDI-TOF MS protein profile-based classifications. However, genome sequences of strains in G3 group were highly comparable (ANIb >90%), supporting the result that MALDI-TOF MS protein profile-based classification was more reliable than 16S rRNA sequence-based classification, which grouped ATCC 14584 and 14585 strains with *S. avermitilis* rather than the ATCC 14583 strain (Figures 2C and 2D). The ANIm value between strains in G3 group and *S. avermitilis* was less than 85%, representing the low discriminatory power of 16S rRNA sequence-based classification. Third, ATCC 21782 and 21018 strains were close to *S. vinaceus* and *S. subrutilus*, consistent with both 16S rRNA sequence- and MALDI-TOF MS protein profile-based classifications, but the ANIm value between the two strains was 87%, implying that they are distinct species.

Phylogenomic analysis confirmed that genomic similarity does not always guarantee phenotypic similarity. For example, the ANIm value between strains in the G3 group was more than 90% (Figure 2D), but the maximum cell mass of the ATCC 14583 strain in the MYM liquid culture condition was approximately twice as high as those of the other two strains belonging to the G3 group (ATCC 14584 and 14585) (Figure 1B).

Table 2. Accession number and information of 25 *Streptomyces* genomes used in this study

	Species	Accession number	Genome size
1	*Streptomyces coelicolor* A3(2)	AL645882.2	8,667,507
2	*Streptomyces lividans* TK24	CP009124.1	8,345,283
3	*Streptomyces griseus* NBRC 13350	AP009493.1	8,545,929
4	*Streptomyces avermitilis* MA-4680	BA000030.4	9,025,608
5	*Streptomyces clavuligerus* ATCC 27064	CP027858	6,748,591
6	*Streptomyces tsukubaensis* NRRL 18488	CP020700	7,963,742
7	*Streptomyces aureofaciens* KCTC 1086	CP023698	7,757,877
8	*Streptomyces galilaeus*KCTC 1921	CP023703	7,756,194
9	*Streptomyces subrutilus* KCTC 9045	CP023701	7,604,974
10	*Streptomyces prasinus* KCTC 9019	CP023697	7,647,592
11	*Streptomyces alboniger* KCTC 9014	CP023695	7,962,786
12	*Streptomyces cinereoruber* KCTC 9706	CP023693	7,516,652
13	*Streptomyces filamentosus* KCTC 9568	PDCL00000000	5,744,022 and 2,131,385
14	*Streptomyces vinaceus* KCTC 9771	CP023692	7,673,509
15	*Streptomyces chartreusis* KCTC 9704	CP023689	9,912,098
16	*Streptomyces venezuelae* ATCC 10712	CP029197	8,223,505
17	*Streptomyces venezuelae* ATCC 15439	CP059991	9,054,831
18	*Streptomyces venezuelae* ATCC 21018	CP029189	7,746,267
19	*Streptomyces venezuelae* ATCC 21782	CP029190	7,525,322
20	*Streptomyces venezuelae* ATCC 15068	CP029194	8,558,202
21	*Streptomyces venezuelae* ATCC 14583	CP029193	8,018,484
22	*Streptomyces venezuelae* ATCC 14584	CP029192	8,942,078
23	*Streptomyces venezuelae* ATCC 14585	CP029191	8,048,154
24	*Streptomyces venezuelae* ATCC 21113	CP029196	7,893,803
25	*Streptomyces venezuelae* ATCC 10595	CP029195	7,871,480

This result highlighted the inappropriateness of morphology-based classification for determining *Streptomyces* taxonomy. However, comparing phenotypes under more diverse culture conditions will help determine strain similarity. Explaining phenotypes via differences in genomes is challenging because a single nucleotide difference sometimes causes an extreme phenotypic change (Kim et al., 2020).

Next, the genomic contents of ten *Streptomyces* strains were comprehensively compared using pan-genome analysis, which identified the diversity in genomic components of the strains (Figures S3A and S3B). The functional classification of orthologous groups using the COG category revealed that genes related with cell growth and maintenance were highly conserved across the strains, whereas a wide distinction was observed in terms of secondary metabolite production, requiring further reassessment of the strains (Figure S3C). In addition, a large portion of genes related to cell motility, replication, recombination, and repair were strain-specific, supporting the distinguishable colony proliferation shape of strains (Figure 1A).

Mining smBGCs from the genomes

To evaluate the potential of the ten strains producing secondary metabolites, smBGCs in their genomes were predicted using antiSMASH (Blin et al., 2017). The results showed that these strains contain a total of 292 smBGCs. The ATCC 14584 strain contained the largest number of 29 smBGCs whereas the ATCC 21782 strain exhibited the smallest number of 17 smBGCs.

Distinguishing the boundary between nearby independent smBGCs is still challenging due to the presence of various hybrid type smBGCs (Cimermancic et al., 2014; Belknap et al., 2020). In this aspect, antiSMASH algorithm uses a term "region," which include hybrid smBGCs, but often contain two or more independent

smBGCs (Blin et al., 2019), requiring manual curation. For example, the venemycin BGC of the ATCC 10595 strain included thiazostatin/watasemycin BGC, whereas the thiazostatin/watasemycin BGC of ATCC 21113 and 10712 strains included venemycin BGCs; thus, the two smBGCs were manually separated (Figure S4A). The nenestatin BGC of the ATCC 14584 strain was included in the lomaiviticin BGC, whereas the mirubactin BGC of the ATCC 14583 strain was included in the netropsin BGC (Figures S4B and S4C).

Meanwhile, the flaviolin BGC of the ATCC 15439 strain, the alkylresorcinol BGC of the ATCC 15068 strain and the purincyclamide BGC of the ATCC 14585 strain were missing in the antiSMASH prediction. These smBGCs were identified using BLAST-based searching with the corresponding smBGC sequences of phylogenetically close strains (Figures S4D–S4F). Additionally, owing to the limitation of the MIBiG database, several well-characterized smBGCs of these strains were predicted as different smBGCs. For example, ATCC 10595, 21113, and 10712 strains are known to produce jadomycin, but the jadomycin BGC was annotated as rabelomycin or auricin BGCs. In addition, the pikromycin BGC of ATCC 15439 and 15068 strains was predicted to be rosamicin BGC (Figure 3A). Overall, 246 of 292 smBGCs were matched with the known smBGCs in the MIBiG database (Kautsar et al., 2020), which were 100 unique smBGCs among the ten strains (Table S3). To minimize the false-positive smBGCs predicted with the low similarity, we selected smBGCs based on their sequence similarities with >30% according to the jadomycin BGC in the ATCC 10595 strain, because the jadomycin production from this BGC was previously validated (Kim et al., 2020). Finally, a total of 48 unique smBGCs were identified, and their distribution was distinct among the genomes of the strains (Figure 3A).

Consistent with the genome sequence-based phylogenetic analysis, ATCC 10595, 21113, and 10712 strains in G1 group, ATCC 15439 and 15068 strains in G2 group, and ATCC 14584, 14585, and 14583 strains in G3 group shared most of the smBGCs, respectively (Figure 3B). Strains in G1 group that are thought to be *S. venezuelae* species shared almost the whole smBGCs, including the chloramphenicol and jadomycin BGCs (Figure 3A). Strains in G2 group shared 10 smBGCs, especially the pikromycin BGC (Figure 3B). Strains in G3 group possessed almost identical smBGCs and shared only three smBGCs (geosmin, hopene, and ectoine BGCs) with the strains in G1 and G2 groups, indicating that strains in G3 group produce completely different secondary metabolites compared to other *S. venezuelae* species (Figure 3B). A representative product of strains in G3 group is the pyrrolamide metabolite anthelvencin, of which a BGC was recently identified in the ATCC 14583 genome (Aubry et al., 2020). Indeed, all strains in G3 group contained the anthelvencin BGC (predicted as netropsin BGC) (Figure 3A). The remaining ATCC 21018 and 21782 strains contained various strain-specific smBGCs (Figure 3B). Next, we compared smBGCs predicted in the strains, and found that geosmin, hopene, and ectoine BGCs are highly conserved (100% similarity for geosmin BGCs, 61%–84% similarity for hopene BGCs, and 100% similarity for ectoine BGCs) (Figure 3C). Geosmin, which has an earthy odor, is speculated to repel predators or attract organisms which could disperse the spore, helping the growth of *Streptomyces* (Becher et al., 2020). In addition, it has been reported that hopene and ectoine BGCs are well-conserved throughout the *Streptomyces* species (Kim et al., 2015), because hopene and ectoine provide the cellular resistance against harsh conditions such as heat or acidity (Poralla et al., 2000; Bursy et al., 2008). Overall, smBGC mining of the genomes of the three *S. venezuelae* and seven misclassified strains enabled us to reassess their genetic potential to produce various secondary metabolites.

Examination of secondary metabolite production

We investigated the expected secondary metabolites of smBGCs encoded in each genome using targeted LC-MS/MS analysis. To increase the chemical diversity of the strains, they were cultured in two different culture media: (1) liquid MYM media, which is a conventional medium for chloramphenicol and jadomycin production of the ATCC 10712 strain (Kim et al., 2020); and (2) liquid MP5 media, which is a representative culture condition for the production of pyrrolamide-type antibiotics, including netropsin and anthelvencin (Juguet et al., 2009; Aubry et al., 2020). Since the carbon sources of the two-culture media are different (maltose and glycerol), the diversity of secondary metabolite production of the ten strains was expected to increase. After the ethyl acetate extraction of the cultures, the extracts were subjected to the LC-MS/MS analysis (Dataset S1). On average, six secondary metabolites were produced per strain, with nine compounds (highest number) from ATCC 10712 and 14583 strains, and one compound (lowest number) from the ATCC 21018 strain (Figure 3A). As expected, the similarity of detected smBGCs to the MIBiG database was higher, and the corresponding products were detected more often (Figure S5). The targeted LC-MS/MS screening method results were highly consistent with the previous studies on the ten strains

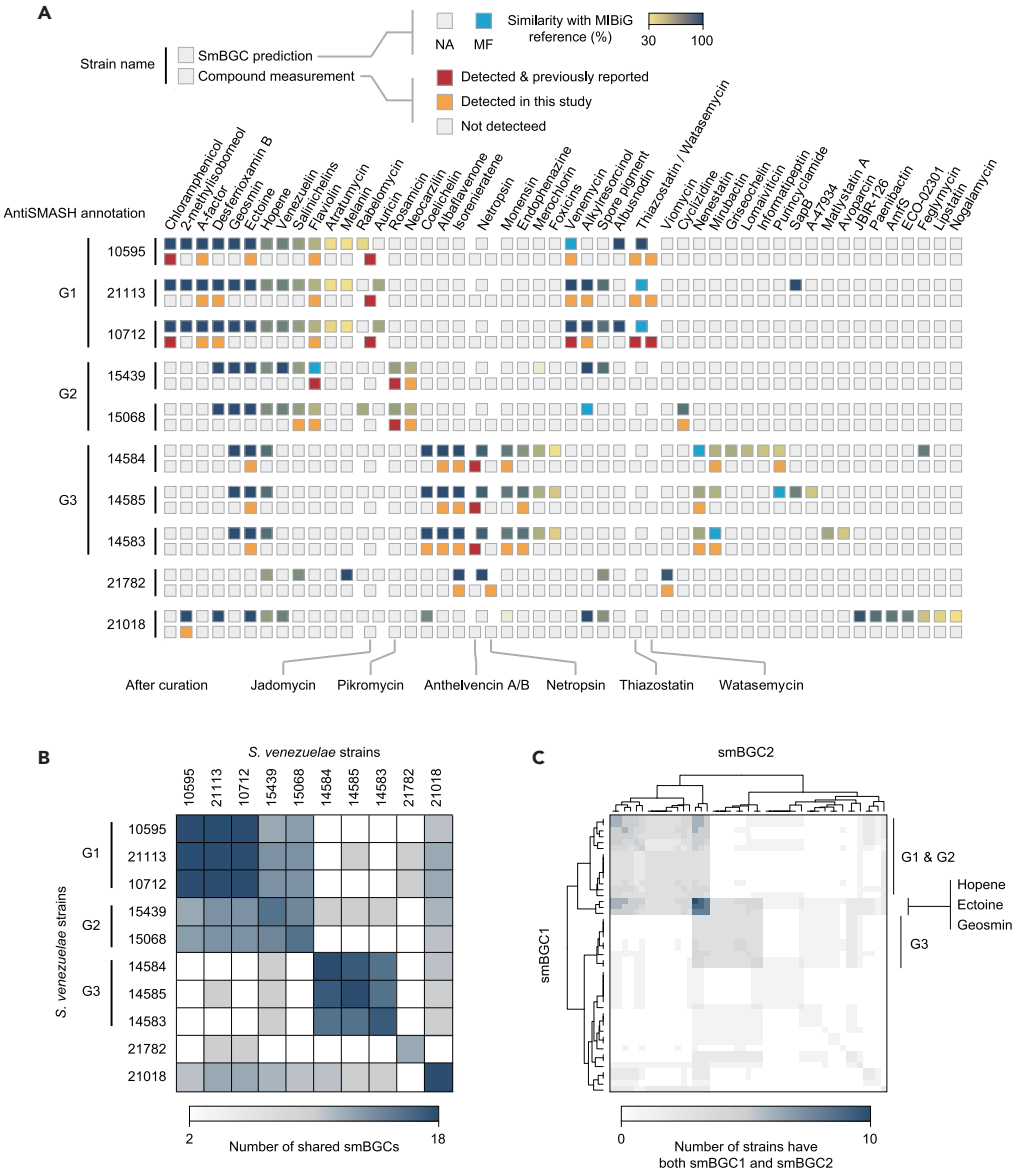


Figure 3. SmBGC mining and comparison of ten *Streptomyces venezuelae* strains

(A) Mining smBGCs and measuring secondary metabolite production using LC-MS/MS. Leftmost G1, G2, and G3 are clades based on genome alignment in Figure 2D. Second numbers indicate the ATCC number of *S. venezuelae* strains. Upper boxes represent existence of corresponding smBGC in the genome based on antiSMASH prediction. Names above the upper boxes are the putative product names predicted by antiSMASH. For upper boxes of each species, sky blue boxes indicate manually found (i.e., MF) and other colored boxes indicate similarity with known BGC in MiBiG database. Lower boxes represent whether corresponding secondary metabolites were produced and names below the boxes are curated products of corresponding smBGCs. Red boxes indicate that according secondary metabolites were produced and it is previously reported, and orange boxes represent secondary metabolites production identified first time in this study. Among the identified 59 secondary metabolite productions by ten strains (red and orange boxes), 45 cases (orange boxes) were firstly reported in this study. If two lower boxes are assigned to single upper box, it means that two secondary metabolites can be produced from the single smBGC. For example, thiazostatin/watasemycin BGC produces both thiazostatin and watasemycin. On the other hand, two upper boxes assigned to a single lower box means that two smBGCs produce the same secondary metabolites. Jadomycin producing BGC was predicted as rabelomycin BGC in *S. venezuelae* ATCC 10595 and 15068, whereas it was predicted as auricin BGC in ATCC 21113 and 10712.

(B) Number of shared smBGCs between *S. venezuelae* strains.

(C) Number of strains have both smBGC1 and smBGC2. SmBGC1 and smBGC2 indicates each row and column.

(Thanapipatsiri et al., 2016; Inahashi et al., 2017; Kim et al., 2020), and they further enabled the determination of unexplored secondary metabolite production in the ten strains.

Putative chloramphenicol production was detected in the strains in G1 group, except for the ATCC 21113 strain that contains a loss-of-function mutation in the *cmlI* gene of chloramphenicol BGC, disrupting the chloramphenicol production (Kim et al., 2020). In addition, consistent with the studies that focused on the ATCC 10712 strain, putative jadomycin (Kim et al., 2020), venemycin (Thanapipatsiri et al., 2016), thiazostatin, and watasemycin (Inahashi et al., 2017) production was observed in all strains in G1 group. In particular, putative flaviolin production was newly identified from all strains in G1 group, whereas flaviolin production was only reported in the ATCC 15439 strain in G2 group (Maharjan et al., 2010). Moreover, putative A-factor and alkylresorcinol were also detected in strains in G1 group, production of which has not yet been reported for this group. All strains in G2 group seem to produce flaviolin (Maharjan et al., 2010) and pikromycin (Xue and Sherman, 2001) as reported previously. In addition, putative neocarzilin production from ATCC 15439 and 15068 strains was newly observed, which is an antitumor chlorinated polyenone produced by *S. carzinostaticus* (Otsuka et al., 2004). Moreover, the ATCC 15068 strain produced cyclizidine, which has nonselective immunostimulatory properties (Huang et al., 2015). Strains in G3 group are known to produce only anthelvencin, but genome mining and targeted LC-MS/MS analysis revealed that the strains in G3 group produced various putative bioactive compounds, such as albaflavenone, isorenieratene, monensin, endophenazine, nenestatin, mirubactin, and purincyclamide. Among the remaining two strains, the ATCC 21782 strain produced netropsin, which belongs to the pyrrolamide-type antibiotic family, and viomycin with anti-tuberculosis activity. Moreover, the ATCC 21018 strain was the only one of ten strains to produce putative 2-methylisoborneol. In summary, genome mining and metabolite identification using targeted LC-MS/MS revealed that these ten strains can produce more diverse and distinct secondary metabolites than expected. Among the identified 59 secondary metabolite productions by ten strains (number of red and orange boxes in Figure 3A), 45 cases (number of orange boxes in Figure 3A) were reported for the first time in this study.

Pyrrolamide-type antibiotic production in ATCC 14583, 14584, 14585, and 21782 strains

Current studies on *S. venezuelae* strains have been highly focused on ATCC 10712 and 15439 strains, which are representative strains in G1 and G2 groups, respectively (Kim et al., 2020). However, only a few studies have been reported for strains in G3 group (ATCC 14583, 14584, and 14585) and the ATCC 21782 strain. The noteworthy characteristics of these four strains are their ability to produce pyrrolamide-type antibiotics, which bind to the minor groove of DNA in a sequence-specific manner and have the potential to be used as anticancer agents (Neidle, 2001). As described above, all strains in G3 group produced anthelvencin and the ATCC 21782 strain produced netropsin.

The anthelvencin BGCs of ATCC 14583, 14584, and 14585 strains varied in size from 59 to 105 kbp, but contained the same 30 kbp region, which was determined as the combined BGC of ectoine and NRPS-type BGC. Among the anthelvencin BGC-encoded genes, 20 had a homolog within netropsin or distamycin BGC and showed an amino acid sequence identity range between 51.69% and 85.26% with the corresponding genes (Figure 4A and Table 3). Based on sequence similarity, genes in anthelvencin BGC were named *athv* and numbered according to the number of homologous genes in netropsin and distamycin BGCs (Table 3). A notable difference between netropsin and anthelvencin BGCs was the absence of *cgc7* and *cgc18* homologs in anthelvencin BGCs (Figure 4A and Table 3). Both genes are involved in the synthesis of guanidinoacetate, one of the main precursors for netropsin assembly (Juguet et al., 2009). In addition, the *dst26* gene, which is involved in formylation during distamycin biosynthesis, did not exist in the anthelvencin BGC (Figure 4A and Table 3) (Vingadassalon et al., 2015). The absence of these genes might be the reason why ATCC 14583, 14584, and 14585 strains are unable to produce other pyrrolamide-type antibiotics, such as netropsin, distamycin, and disgocidine. In contrast, the netropsin BGC of the ATCC 21782 strain contained the identical order of genes as the netropsinBGC of *S. ambofaciens*, supporting the netropsin production from the ATCC 21782 strain.

Anthelvencin comprises two compounds, anthelvencin A and B, which are distinguished by the number of methylated nitrogens in pyrrole units (Figure 4B). We observed putative anthelvencin A and B production using high-resolution FT-ICR MS analysis. The peaks corresponding to the molecular weight of anthelvencin A and B were detected with retention times of 2.3 min and 4.1 min, respectively, in the extract ion chromatograms (Figure 4C). To confirm whether the predicted BGC is actually involved in anthelvencin production, three BGC-encoded genes of the ATCC 14585 strain were deleted using the CRISPR/Cas9

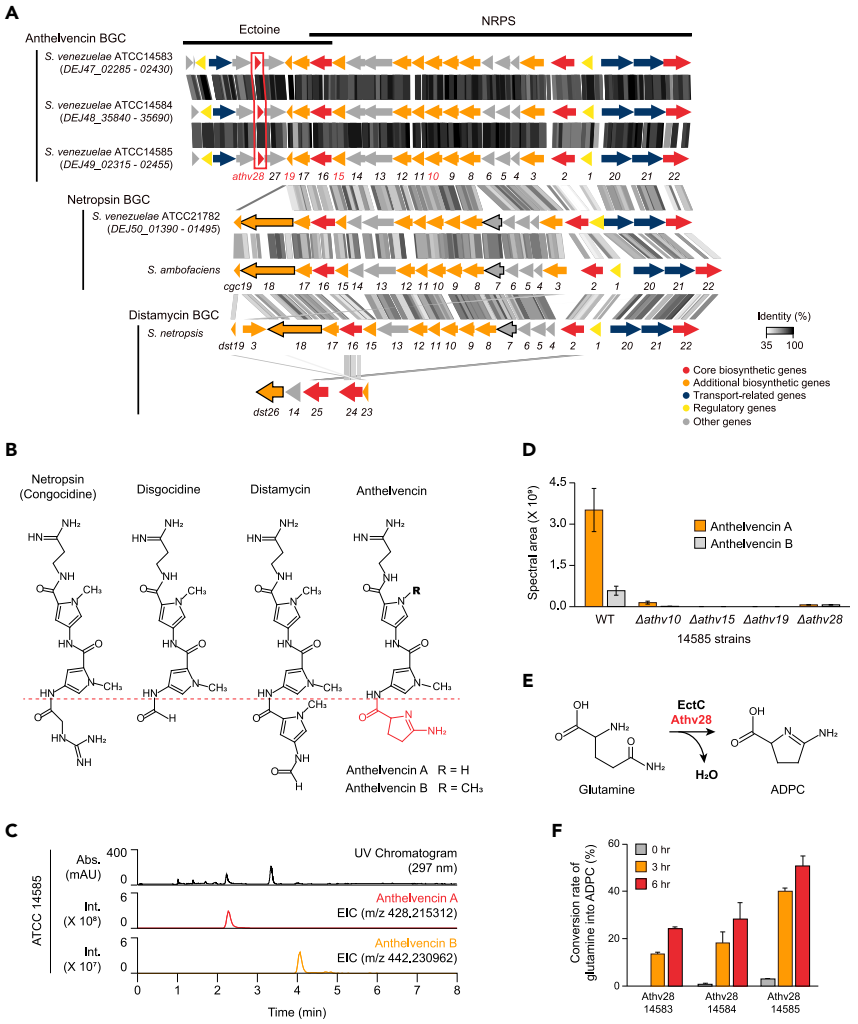


Figure 4. Pyrrolamide-type antibiotic production of *Streptomyces venezuelae* strains

(A) Comparison of anthelvencin BGC and netropsin BGC of *S. venezuelae* strains with netropsin BGC of *S. ambofaciens* and distamycin BGC of *S. netropsis*. Genes of the *S. venezuelae* ATCC 14585 strain marked in red are targets for deletion studies. NRPS, non-ribosomal peptide synthetase.

(B) Chemical structures of pyrrolamide-type antibiotics.

(C) FT-ICR-based anthelvencin A and B measurement.

(D) Abolition of the anthelvencin production in the *S. venezuelae* ATCC 14585 strain mutant with *athv10*, *athv15*, *athv19*, and *athv28* deletion. Data are represented as mean value and error bars indicate standard deviation (s.d.) of three biological replicates.

(E) Proposed reaction of Athv28 to produce ADPC from glutamine.

(F) *In vitro* conversion of glutamine to ADPC using purified Athv28 proteins of *S. venezuelae* ATCC 14583, 14584, and 14585 strains. Data are represented as mean value and error bars indicate standard deviation (s.d.) of three biological replicates.

mediated knockout system: (1) *athv10*, which is involved in precursor (4-acetamidopyrrole-2-carboxylate) synthesis; (2) *athv15*, methylating pyrrole groups; and (iii) *athv19*, involved in the anthelvencin assembly (see STAR Methods and Figure S6A). Anthelvencin A and B production of all three deletion strains was decreased to undetectable levels, supporting that the BGCs encoding these genes are associated with anthelvencin production (Figure 4D).

The notable structural difference between anthelvencin and other pyrrolamide-type antibiotics is the third connected unit during the assembly, which is predicted to be 5-amino-3,4-dihydro-2H-pyrrole-2-carboxylate (ADPC) for anthelvencin, not guanidinoacetate for netropsin or pyrrole-2-carboxamide for distamycin

Table 3. Proposed function of genes encoded in the anthelvencin BGC

Genes in anthelvencin BGC	Genes in netropsin BGC	Sequence identity (%)	Genes in distamycin BGC	Sequence identity (%)	Putative function	Role in biosynthesis
athv28	–	–	–	–	L-ectoine synthase	Precursor biosynthesis (ADPC)
athv27	–	–	–	–	Hypothetical protein	–
athv19	*Cg 19*	76.77	*dst19*	72.41	NRPS, PCP domain	Congocidine, distamycin, anthelvencin assembly
–	*cgc18*	–	*dst18*	–	NRPS, A-PCP-C domains	Congocidine assembly (guanidinoacetate)
athv17	*cgc17*	83.67	*dst17*	77.94	Alcohol dehydrogenase	Precursor biosynthesis
athv16	*cgc16*	68.31	*dst16*	62.7	NRPS, C domain	Congocidine, distamycin, anthelvencin assembly
athv15	*cgc15*	83.64	*dst15*	75.09	Methyltransferase	Methylation of pyrrole groups
athv14	*cgc14*	84.43	*dst14*	73.15	Amidohydrolase, deacetylase	Precursor biosynthesis (4-acetamidopyrrole-2-carboxylate)
athv13	*cgc13*	80.23	*dst13*	70.49	Glycoside hydrolase	Precursor biosynthesis
athv12	*cgc12*	79.68	*dst12*	73.96	Nucleotide-sugar aminotransferase	Precursor biosynthesis
athv11	*cgc11*	76.19	*dst11*	73.41	Sugar nucleotidyltransferase	Precursor biosynthesis
athv10	*cgc10*	81.44	*dst10*	71.87	Glycosyltransferase	Precursor biosynthesis (4-acetamidopyrrole-2-carboxylate)
athv9	*cgc9*	83.99	*dst9*	81.45	Nucleoside diphosphate sugar epimerase/ dehydrogenase	Precursor biosynthesis
athv8	*cgc8*	84.37	*dst8*	76.69	Nucleotide sugar dehydrogenase	Precursor biosynthesis
–	*cgc7*	–	*dst7*	–	Hypothetical protein	Precursor biosynthesis (guanidinoacetate)
athv6	*cgc6*	84.62	*dst6*	80.44	Creatinase	Precursor biosynthesis
athv5	*cgc5*	84.48	*dst5*	81.79	Dihydroorotate dehydrogenase	Precursor biosynthesis (3-aminopropionamidine)
athv4	*cgc4*	83.33	*dst4*	79.44	Nucleoside 2-deoxyribosyltransferase	Precursor biosynthesis
athv3	*cgc3*	77.18	*dst3*	59.69	aldehyde dehydrogenase	Precursor biosynthesis
athv2	*cgc2*	67.58	*dst2*	61.93	NRPS, C domain	Congocidine, distamycin, anthelvencin assembly
athv1	*cgc1*	68.47	*dst1*	76.44	Two-component response regulator	Regulation
athv20	*cgc20*	84.07	*dst20*	72.52	ABC transporter, transmembrane protein	Resistance
athv21	*cgc21*	85.26	*dst21*	76.35	ABC transporter, ATP-binding protein	Resistance
athv22	*cgc22*	73.54	*dst22*	61.24	Acyl-CoA synthase	Congocidine, distamycin, anthelvencin assembly
–	–	–	*dst26*	–	NRPS, formyltransferase domain	Distamycin assembly
–	–	–	*dst25*	–	NRPS, C domain, similar with *dst2*	Distamycin assembly

(Figure 4B). Indeed, ADPC was used as a precursor in the chemical synthesis of anthelvencin (Witt et al., 2011), and a recent study characterized the anthelvencin BGC from the ATCC 14583 genome sequence and showed that chemical complementation of ADPC to the *athv28* deleted ATCC 14583 stain restored anthelvencin production (Aubry et al., 2020). Athv28 was hypothesized to catalyze the intramolecular condensation of glutamine to ADPC (Figure 4E); however, this hypothesis has not yet been validated. We also demonstrated that anthelvencin was not produced when *athv28* was deleted from the ATCC 14585 strain, indicating that *athv28* plays an important role in anthelvencin assembly (Figure 4D). The amino acid sequence of Athv28 was highly similar to that of the EctC encoded in the ectoine BGC and structurally important metal binding sites, substrate interacting sites, and reaction coordinating sites were well-conserved (Figure S6B) (Czech et al., 2019). EctC is an ectoine synthase that catalyzes the intramolecular condensation of N-gamma-acetyl-L-2,4-diaminobutyrate (ADABA) into ectoine. Furthermore, EctC is known to accept more than one substrate and produce ADPC by cyclic condensation of glutamine as a side reaction (Witt et al., 2011). Based on these facts, we also speculated that ADPC, which is necessary for the production of anthelvencin, was produced by *athv28* (Figure 4E). In order to identify the involvement of Athv28 in ADPC biosynthesis *in vitro*, we purified Athv28 from ATCC 14583, 14584, and 14585 strains, respectively. *N*-terminal (His)10-MBP-TEV tagged Athv28 proteins were successfully overexpressed in *Escherichia coli* BL21 (DE3) RIPL strain and purified using Ni-NTA affinity chromatography, followed by treatment with tobacco etch virus protease to remove the His and MBP tags from the Athv28 (Figures S6C and S6D). Purified enzymes were subject to react with 10 mM glutamine and the amount of glutamine and ADPC was measured using FT-ICR according to the reaction times. As a result, all three Athv28 enzymes converted glutamine into ADPC, supporting that the ADPC precursor for anthelvencin is provided by Athv28 mediated cyclic condensation of glutamine (Figure 4F).

In summary, we compared the pyrrolamide-type antibiotic BGCs in four non-*S. venezuelae* strains and revealed that the ATCC 14583, 14584, and 14585 strains produce ADPC, the most important precursor of anthelvencin, using the *athv28* gene, thereby producing anthelvencin. However, because of the lack of homologs needed for assembling other pyrrolamide-type antibiotics, these strains could not produce netropsin, disgocidine, and distamycin. In contrast, ATCC 21782 was identified as a netropsin producer.

DISCUSSION

Streptomyces has the potential to synthesize more than 30 secondary metabolites per species, thereby providing an enormous resource for discovering unexplored bioactive compounds (Omura et al., 2001; Bentley et al., 2002). Furthermore, prevalent genetic exchanges within *Streptomyces* strains and/or species (McDonald and Currie, 2017) may lead to variations in secondary metabolite production, even at the strain level (Belknap et al., 2020). The current taxonomy of *Streptomyces* is notably problematic, because *Streptomyces* classifications rely on routine morphological and physiological characterizations, which often lead to incorrect classifications of species because of the phenotypic similarity between *Streptomyces* species (Kim et al., 2012). Owing to their inaccurate taxonomy, several *Streptomyces* strains are often undervalued as producers of identical secondary metabolites from already known close strains (Sottorff et al., 2019). In this respect, accurate classifications and evaluations of various streptomycetes used to produce secondary metabolites are necessary to expand repositories for new drug discovery (Sottorff et al., 2019). Moreover, determining phylogenetically close strains and comparing their smBGCs elucidates genetic factors involved in secondary metabolite biosynthesis (Cho et al., 2019; Kim et al., 2020). In addition, certain *Streptomyces* species are human or plant pathogens, such as *S. sudanensis* (Quintana et al., 2008) and *S. ipomoeae* (Lee et al., 2015); thus, an accurate classification pipeline is required to determine *Streptomyces* taxonomy.

In this study, we verified the taxonomy of ten *Streptomyces* strains belonging to the *S. venezuelae* species, which has garnered considerable attention as a model organism in studies on secondary metabolism and heterologous expression (Hong et al., 2004; Phelan et al., 2017). We have utilized three different phylogenetic analysis methods: (1) 16S rRNA sequence-based classification, which is the conventional method for bacterial phylogenetic analysis, (2) MALDI-TOF MS protein profile-based classification, which is a rapid and cost-effective method for identifying microorganisms, and (3) phylogenomic analysis, which is the most robust taxonomic method and the gold standard for species delineation. Most of the *S. venezuelae* strains were classified approximately 50 years ago with uncertain criteria, and comparative studies have not been conducted so far. Our revised taxonomy of the ten studied *Streptomyces* strains revealed that only ATCC 10595, 21113, and 10712 strains were determined to be *S. venezuelae*, whereas the remaining seven strains

were re-classified as distinct *Streptomyces* species. Specifically, the genome sequences of ATCC 15439 and 15068 strains were comparable, but distinct from that of the ATCC 10712 strain. Additionally, the genomes of ATCC 14583, 14584, and 14585 strains were highly similar, but were distant from the ATCC 10712 strain. ATCC 21782 and 21018 strains were distantly related to the other strains and should be considered as independent species. The results of phylogenomic analysis were more consistent with the MALDI-TOF MS protein profile-based phylogeny than the 16S rRNA sequence-based phylogeny. Compared to 16S rRNA and genome sequence-based phylogenetic analysis, the MALDI-TOF MS protein profile-based classification may be considered a reasonable alternative method for the rapid and cost-effective identification of existing *Streptomyces* and discovery of new *Streptomyces* species. For this purpose, the MALDI-TOF MS protein profile database of various *Streptomyces* species must be updated routinely.

A comparison of smBGCs predicted to exist in three *S. venezuelae* and seven misclassified strains, followed by an analysis of the putative products corresponding to the smBGCs using targeted LC-MS/MS, revealed the chemical diversity between these strains. Specifically, we observed the putative production of (1) flaviolin, A-factor, and alkylresorcinol in the ATCC 10595, 21113, and 10712 strains; (2) neocarzilin or cyclizidine in the ATCC 15439 and 15068 strains; (3) various secondary metabolites, including albaflavenone, isorenieratene, monensin, endophenazine, nenestatin, mirubactin, and purincyclamide in the ATCC 14583, 14584, and 14585 strains; (4) netropsin in the ATCC 21782 strain; and (5) 2-methylisoborneol in the ATCC 21018 strain for the first time. As all the studied strains showed a similar growth rate to the ATCC 10712 and ATCC 15439 strains (Figure 1B), which are known for rapid growth (Hong et al., 2004; Moore et al., 2021), they have a strong potential to be utilized as industrial strains producing corresponding secondary metabolites in the future.

In addition, comparative genomics of phylogenetically close non-*S. venezuelae* strains determine that some essential genes for synthesizing pyrrolamide-type antibiotics, such as netropsin, disgocidine, and distamycin, were absent in the pyrrolamide-type antibiotic BGCs of three strains (ATCC 14583, 14584, and 14585). Therefore, these strains produced only anthelvencins (Figure S6E). The production of putative anthelvencins in these strains was demonstrated using a targeted LC-MS/MS analysis (Figure 4C). Furthermore, we verified for the first time that the *athv28* gene in the anthelvencin BGCs of the ATCC 14583, 14584, and 14585 strains was responsible for converting glutamine to ADPC, which is an irreplaceable precursor in anthelvencin assembly. Pyrrolamide-type antibiotics can recognize and bind to DNA minor grooves with sequence specificity and could potentially be used as anti-infective or anti-tumor agents (Neidle, 2001). However, owing to the toxicity of these compounds, the synthesis of numerous pyrrolamide derivatives has attracted considerable attention (Vaijayanthi et al., 2012). Therefore, we believe that the *athv28* gene can be utilized in the design of unnatural pyrrolamide structure for developing special drugs or sequence-selective DNA probes.

Among the 243 smBGCs predicted from the ten *Streptomyces* strains, known as *S. venezuelae* species, and matched with the known smBGCs in the MiBiG database (Table S3), only 148 smBGCs (sequence similarities with >30%) (Figure 3A) were examined in this study. In addition, considering that most smBGCs are normally not expressed under laboratory culture conditions (Walsh and Fischbach, 2010), these ten *Streptomyces* strains are expected to produce more variable secondary metabolites than those measured in this study. *Streptomyces* is one of the largest genera with more than 1,000 species described in the NCBI taxonomy database. A reclassification and reassessment of the currently identified *Streptomyces* strains to fully evaluate their genetic potential will further detail the discovery of unexplored bioactive compounds.

Limitations of the study

59 secondary metabolites productions identified in this study are conditional on the culture media, extraction method, and instrument. In addition, considering that total 243 smBGCs were mined from the genomes of the ten *Streptomyces* strains, these strains are expected to produce more variable secondary metabolites than those measured in this study.

SIGNIFICANCE

Current *Streptomyces* taxonomy contains numerous errors, which hamper expanding the repository for secondary metabolite gene clusters. Here we have identified misclassification of *S. venezuelae* strains using 16S rRNA sequencing, MALDI-TOF MS protein profiling, and phylogenomics approach. Among the strains, only three strains can be classified as *S. venezuelae* species and genetic potentials of remaining seven

strains to produce secondary metabolite were highly distinct with *S. venezuelae* species. Especially, by comparing secondary metabolite gene clusters of the strains, we determined anthelvencin biosynthetic gene clusters in three strains and unexplored gene involved in anthelvencin precursor biosynthesis. Together, this study emphasizes that accurate and reliable classification methods are needed to better understand physiology of *Streptomyces* and their ability to produce biologically relevant secondary metabolites.

DATA AVAILABILITY

The whole datasets used in this study were downloaded from the NCBI dataset (Accession numbers are listed in STAR Methods).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial strains and culture media
- METHOD DETAILS
  - 16S rRNA sequence-based phylogenetic analysis
  - MALDI-TOF MS protein profile-based phylogenetic analysis
  - Genome sequencing, assembly, and annotation
  - Targeted LC-MS/MS analysis for putative secondary metabolite measurement
  - Determination of anthelvencin production using high-resolution FT-ICR MS
  - Disruption of anthelvencin BGC-encoded genes using CRISPR/Cas9 system
  - Expression and purification of Athv28
  - *In vitro* assay with Athv28 and measuring the ADPC
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103410.

ACKNOWLEDGMENTS

This work was supported by the Bio & Medical Technology Development Program (grant no. 2018M3A9F3079664 to B.-K.C. and NRF2021M3A9I5023245 to B.-K.C.) through the National Research Foundation (NRF) funded by the Ministry of Science and ICT (MSIT). This work was also supported by a grant from the Novo Nordisk Foundation (grant no. NNF10CC1016517 to B.P.) and KBSI grant (C030440 to K.-S.J.).

AUTHOR CONTRIBUTIONS

N.L., K.-S.J., and B.-K.C. designed the study. N.L., M.C., and W.K. performed the experiments. N.L., M.C., W.K., S.H., Y.L., J.H.K, G.K., H.K., S.C., K.-S.J., and B.-K.C. performed data analysis. N.L., W.K., S.H., S.C., B.P., K.-S.J., and B.-K.C. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 7, 2021
Revised: September 5, 2021
Accepted: November 4, 2021
Published: December 17, 2021

REFERENCES

Antony-Babu, S., Stien, D., Eparvier, V., Parrot, D., Tomasi, S., and Suzuki, M.T. (2017). Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S rRNA gene sequences. Sci. Rep. *7*, 11089. https://doi.org/10.1038/s41598-017-11363-1.

Aubry, C., Clerici, P., Gerbaud, C., Micouin, L., Pernodet, J.L., and Lautru, S. (2020). Revised structure of anthelvencin A and characterization of the anthelvencin biosynthetic gene cluster. ACS Chem. Biol. *15*, 945–951. https://doi.org/10.1021/acschembio.9b00960.

Becher, P.G., Verschut, V., Bibb, M.J., Bush, M.J., Molnár, B.P., Barane, E., Al-Bassam, M.M., Chandra, G., Song, L., Challis, G.L., et al. (2020). Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a soil arthropod to *Streptomyces* bacteria promoting spore dispersal. Nat. Microbiol. *5*, 821–829. https://doi.org/10.1038/s41564-020-0697-x.

Belknap, K.C., Park, C.J., Barth, B.M., and Andam, C.P. (2020). Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. Sci. Rep. *10*, 2003. https://doi.org/10.1038/s41598-020-58904-9.

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature *417*, 141–147. https://doi.org/10.1038/417141a.

Berdy, J. (2005). Bioactive microbial metabolites. J. Antibiot. (Tokyo) *58*, 1–26. https://doi.org/10.1038/ja.2005.1.

Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., Suarez Duran, H.G., de Los Santos, E.L.C., Kim, H.U., Nave, M., et al. (2017). antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res. *45*, W36–W41. https://doi.org/10.1093/nar/gkx319.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., and Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. *47*, W81–W87. https://doi.org/10.1093/nar/gkz310.

Bonfield, J.K., and Whitwham, A. (2010). Gap5–editing the billion fragment sequence assembly. Bioinformatics *26*, 1699–1703. https://doi.org/10.1093/bioinformatics/btq268.

Bursy, J., Kuhlmann, A.U., Pittelkow, M., Hartmann, H., Jebbar, M., Pierik, A.J., and Bremer, E. (2008). Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by *Streptomyces coelicolor* A3(2) in response to salt and heat stresses. Appl. Environ. Microbiol. *74*, 7286–7296. https://doi.org/10.1128/AEM.00768-08.

Bush, M.J., Tschowri, N., Schlimpert, S., Flardh, K., and Buttner, M.J. (2015). c-di-GMP signalling and the regulation of developmental transitions in streptomycetes. Nat. Rev. Microbiol. *13*, 749–760. https://doi.org/10.1038/nrmicro3546.

Chandra, G., and Chater, K.F. (2014). Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. FEMS Microbiol. Rev. *38*, 345–379. https://doi.org/10.1111/1574-6976.12047.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569. https://doi.org/10.1038/nmeth.2474.

Cho, H.S., Jo, J.C., Shin, C.H., Lee, N., Choi, J.S., Cho, B.K., Roe, J.H., Kim, C.W., Kwon, H.J., and Yoon, Y.J. (2019). Improved production of clavulanic acid by reverse engineering and overexpression of the regulatory genes in an industrial *Streptomyces clavuligerus* strain. J. Ind. Microbiol. Biotechnol. *46*, 1205–1215. https://doi.org/10.1007/s10295-019-02196-0.

Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Wieland Brown, L.C., Mavrommatis, K., Pati, A., Godfrey, P.A., Koehrsen, M., Clardy, J., et al. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell *158*, 412–421. https://doi.org/10.1016/j.cell.2014.06.034.

Claydon, M.A., Davey, S.N., Edwards-Jones, V., and Gordon, D.B. (1996). The rapid identification of intact microorganisms using mass spectrometry. Nat. Biotechnol. *14*, 1584–1586. https://doi.org/10.1038/nbt1196-1584.

Cobb, R.E., Wang, Y., and Zhao, H. (2015). High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. ACS Synth. Biol. *4*, 723–728. https://doi.org/10.1021/sb500351f.

Czech, L., Höppner, A., Kobus, S., Seubert, A., Riclea, R., Dickschat, J.S., Heider, J., Smits, S.H.J., and Bremer, E. (2019). Illuminating the catalytic core of ectoine synthase through structural and biochemical analysis. Sci. Rep. *9*, 364. https://doi.org/10.1038/s41598-018-36247-w.

Darling, A.C., Mau, B., Blattner, F.R., and Perna, N.T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. *14*, 1394–1403. https://doi.org/10.1101/gr.2289704.

Doroghazi, J.R., and Metcalf, W.W. (2013). Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. BMC Genom. *14*, 611. https://doi.org/10.1186/1471-2164-14-611.

Dutcher, J.D., Richard, D., Heuser, L.J., Pagano, J.F., and David, P. (1956). Methymycin. US2916483A.

Edgar, R.C. (2018). Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics *34*, 2371–2375. https://doi.org/10.1093/bioinformatics/bty113.

Ehrlich, J., Gottlieb, D., Burkholder, P.R., Anderson, L.E., and Pridham, T.G. (1948). *Streptomyces venezuelae*, n. sp., the source of chloromycetin. J. Bacteriol. *56*, 467–477. https://doi.org/10.1128/JB.56.4.467-477.1948.

Embley, T.M., and Stackebrandt, E. (1994). The molecular phylogeny and systematics of the actinomycetes. Annu. Rev. Microbiol. *48*, 257–289. https://doi.org/10.1146/annurev.mi.48.100194.001353.

Evason, D.J., Claydon, M.A., and Gordon, D.B. (2001). Exploring the limits of bacterial identification by intact cell-mass spectrometry. J. Am. Soc. Mass Spectrom. *12*, 49–54. https://doi.org/10.1016/S1044-0305(00)00192-6.

Haft, D.H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M.K., Gonzales, N.R., et al. (2018). RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res. *46*, D851–D860. https://doi.org/10.1093/nar/gkx1068.

Hong, J.S., Park, S.H., Choi, C.Y., Sohng, J.K., and Yoon, Y.J. (2004). New olivosyl derivatives of methymycin/pikromycin from an engineered strain of *Streptomyces venezuelae*. FEMS Microbiol. Lett. *238*, 391–399. https://doi.org/10.1016/j.femsle.2004.08.002.

Hopwood, D.A. (2007). *Streptomyces* in Nature and Medicine : The Antibiotic Makers (Oxford University Press).

Huang, W., Kim, S.J., Liu, J., and Zhang, W. (2015). Identification of the polyketide biosynthetic machinery for the indolizidine alkaloid cyclizidine. Org. Lett. *17*, 5344–5347. https://doi.org/10.1021/acs.orglett.5b02707.

Inahashi, Y., Zhou, S., Bibb, M.J., Song, L., Al-Bassam, M.M., Bibb, M.J., and Challis, G.L. (2017). Watasemycin biosynthesis in *Streptomyces venezuelae*: thiazoline C-methylation by a type B radical-SAM methylase homologue. Chem. Sci. *8*, 2823–2831. https://doi.org/10.1039/c6sc03533g.

Janda, J.M., and Abbott, S.L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J. Clin. Microbiol. *45*, 2761–2764. https://doi.org/10.1128/JCM.01228-07.

Jang, K.S., Park, M., Lee, J.Y., and Kim, J.S. (2017). Mass spectrometric identification of phenol-soluble modulins in the ATCC(R) 43300 standard strain of methicillin-resistant *Staphylococcus aureus* harboring two distinct phenotypes. Eur. J. Clin. Microbiol. Infect. Dis. *36*, 1151–1157. https://doi.org/10.1007/s10096-017-2902-2.

Juguet, M., Lautru, S., Francou, F.X., Nezbedová, S., Leblond, P., Gondry, M., and Pernodet, J.L. (2009). An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in *Streptomyces ambofaciens*. Chem. Biol. *16*, 421–431. https://doi.org/10.1016/j.chembiol.2009.03.010.

Kautsar, S.A., Blin, K., Shaw, S., Navarro-Muñoz, J.C., Terlouw, B.R., van der Hooft, J.J.J., van Santen, J.A., Tracanna, V., Suarez Duran, H.G., Pascal Andreu, V., et al. (2020). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. *48* (D1), D454–D458. https://doi.org/10.1093/nar/gkz882.

Kieser, T., Bibb, M., Buttner, M., Chater, K., and Hopwood, D. (2000). Practical Streptomyces Genetics (John Innes Foundation).

Kim, K.O., Shin, K.S., Kim, M.N., Shin, K.S., Labeda, D.P., Han, J.H., and Kim, S.B. (2012). Reassessment of the status of *Streptomyces setonii* and reclassification of *Streptomyces fimicarius* as a later synonym of *Streptomyces setonii* and *Streptomyces albovinaceus* as a later synonym of *Streptomyces globisporus* based on combined 16S rRNA/*gyrB* gene sequence analysis. Int. J. Syst. Evol. Microbiol. *62* (Pt 12), 2978–2985. https://doi.org/10.1099/ijs.0.040287-0.

Kim, J.N., Kim, Y., Jeong, Y., Roe, J.H., Kim, B.G., and Cho, B.K. (2015). Comparative genomics reveals the core and accessory genomes of *Streptomyces* species. J. Microbiol. Biotechnol. *25*, 1599–1605. https://doi.org/10.4014/jmb.1504.04008.

Kim, W., Lee, N., Hwang, S., Lee, Y., Kim, J., Cho, S., Palsson, B., and Cho, B.K. (2020). Comparative genomics determines strain-dependent secondary metabolite production in *Streptomyces venezuelae* strains. Biomolecules *10*, 864. https://doi.org/10.3390/biom10060864.

Kostrzewa, M., Sparbier, K., Maier, T., and Schubert, S. (2013). MALDI-TOF ms: an upcoming tool for rapid detection of antibiotic resistance in microorganisms. Proteomics Clin. Appl. *7*, 767–778. https://doi.org/10.1002/prca.201300042.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874. https://doi.org/10.1093/molbev/msw054.

La Farina, M., Stira, S., Mancuso, R., and Grisanti, C. (1996). Characterization of *Streptomyces venezuelae* ATCC 10595 rRNA gene clusters and cloning of *rrnA*. J. Bacteriol. *178*, 1480–1483. https://doi.org/10.1128/jb.178.5.1480-1483.1996.

Lalucat, J., Mulet, M., Gomila, M., and Garcia-Valdes, E. (2020). Genomics in bacterial taxonomy: impact on the genus *Pseudomonas*. Genes (Basel) *11*, 139. https://doi.org/10.3390/genes11020139.

Law, J.W., Chan, K.G., He, Y.W., Khan, T.M., Ab Mutalib, N.S., Goh, B.H., and Lee, L.H. (2019). Diversity of Streptomyces spp. from mangrove forest of Sarawak (Malaysia) and screening of their antioxidant and cytotoxic activities. Sci. Rep. *9*, 15262. https://doi.org/10.1038/s41598-019-51622-x.

Lee, M.Y., Kim, H.Y., Lee, S., Kim, J.G., Suh, J.W., and Lee, C.H. (2015). Metabolomics-based chemotaxonomic classification of *Streptomyces* spp. and its correlation with antibacterial activity. J. Microbiol. Biotechnol. *25*, 1265–1274. https://doi.org/10.4014/jmb.1503.03005.

Lee, H.S., Shin, J.H., Choi, M.J., Won, E.J., Kee, S.J., Kim, S.H., Shin, M.G., and Suh, S.P. (2017). Comparison of the Bruker Biotyper and VITEK MS Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass spectrometry systems using a formic acid extraction method to identify common and uncommon yeast isolates. Ann. Lab. Med. *37*, 223–230. https://doi.org/10.3343/alm.2017.37.3.223.

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

Li, Q., Chem, X., Jiang, Y., and Jiang, C. (2016). Morphological identification of actinobacteria. In Actinobacteria-basics and Biotechnological Applications (InTechOpen), pp. 59–86.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T. (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene *111*, 61–68.

Maharjan, S., Park, J.W., Yoon, Y.J., Lee, H.C., and Sohng, J.K. (2010). Metabolic engineering of *Streptomyces venezuelae* for malonyl-CoA biosynthesis to enhance heterologous production of polyketides. Biotechnol. Lett. *32*, 277–282. https://doi.org/10.1007/s10529-009-0152-9.

McDonald, B.R., and Currie, C.R. (2017). Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. mBio *8*, e00644–17. https://doi.org/10.1128/mBio.00644-17.

Meklat, A., Sabaou, N., Zitouni, A., Mathieu, F., and Lebrihi, A. (2011). Isolation, taxonomy, and antagonistic properties of halophilic actinomycetes in Saharan soils of Algeria. Appl. Environ. Microbiol. *77*, 6710–6714. https://doi.org/10.1128/AEM.00326-11.

Iizuka, H., Ayukawa, Y., Suekane, M., and Kanno, M. (1969). Production of Extracellular Glucose Isomerase by Streptomyces, US3622463A.

Meyers, E., Slusarchyk, D., and Liu, W. (1972). Antibiotic Em-98, US3853992A.

Moore, S.J., Lai, H.E., Chee, S.M., Toh, M., Coode, S., Chengan, K., Capel, P., Corre, C., de Los Santos, E.L., and Freemont, P.S. (2021). A *Streptomyces venezuelae* cell-free toolkit for synthetic biology. ACS Synth. Biol. *10*, 402–411. https://doi.org/10.1021/acssynbio.0c00581.

Mörtelmaier, C., Panda, S., Robertson, I., Krell, M., Christodoulou, M., Reichardt, N., and Mulder, I. (2019). Identification performance of MALDI-ToF-MS upon mono- and bi-microbial cultures is cell number and culture proportion dependent. Anal. Bioanal. Chem. *411*, 7027–7038. https://doi.org/10.1007/s00216-019-02080-x.

Myronovskyi, M., and Luzhetskyy, A. (2019). Heterologous production of small molecules in the optimized *Streptomyces* hosts. Nat. Prod. Rep. *36*, 1281–1294. https://doi.org/10.1039/c9np00023b.

Nakayama, K., and Hagino, H. (1968). Process for Producing Methionine Decarboxylase. US3579427A.

Neidle, S. (2001). DNA minor-groove recognition by small molecules. Nat. Prod. Rep. *18*, 291–309.

Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., et al. (2001). Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc. Natl. Acad. Sci. U S A *98*, 12215–12220. https://doi.org/10.1073/pnas.211433198.

Otsuka, M., Ichinose, K., Fujii, I., and Ebizuka, Y. (2004). Cloning, sequencing, and functional analysis of an iterative type I polyketide synthase gene cluster for biosynthesis of the antitumor chlorinated polyenoneneocarzilin in "*Streptomyces carzinostaticus*". Antimicrob. Agents Chemother. *48*, 3468–3476. https://doi.org/10.1128/AAC.48.9.3468-3476.2004.

Park, C.J., and Andam, C.P. (2019). Within-species genomic variation and variable patterns of recombination in the tetracycline producer *Streptomyces rimosus*. Front. Microbiol. *10*, 552. https://doi.org/10.3389/fmicb.2019.00552.

Phelan, R.M., Sachs, D., Petkiewicz, S.J., Barajas, J.F., Blake-Hedges, J.M., Thompson, M.G., Reider Apel, A., Rasor, B.J., Katz, L., and Keasling, J.D. (2017). Development of next generation synthetic biology tools for use in *Streptomyces venezuelae*. ACS Synth. Biol. *6*, 159–166. https://doi.org/10.1021/acssynbio.6b00202.

Poralla, K., Muth, G., and Hartner, T. (2000). Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). FEMS Microbiol. Lett. *189*, 93–95. https://doi.org/10.1111/j.1574-6968.2000.tb09212.x.

Pritchard, L., Glover, R., Humphris, S., Elphinstone, J., and Toth, I. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal. Methods *8*, 12–24. https://doi.org/10.1039/C5AY02550H.

Probst, G.W., Hoehn, M.M., and Mcguire, J.M. (1964). Anthelvencin and Process for the Production Thereof. US3467750A.

Probst, G.W., Hoehn, M.M., and Woods, B.L. (1965). Anthelvencins, new antibiotics with anthelmintic properties. Antimicrob. Agents Chemother. (Bethesda) *5*, 789–795.

Quintana, E.T., Wierzbicka, K., Mackiewicz, P., Osman, A., Fahal, A.H., Hamid, M.E., Zakrzewska-Czerwinska, J., Maldonado, L.A., and Goodfellow, M. (2008). *Streptomyces sudanensis* sp. nov., a new pathogen isolated from patients with actinomycetoma. Antonie Van Leeuwenhoek *93*, 305–313. https://doi.org/10.1007/s10482-007-9205-z.

Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U S A *106*, 19126–19131. https://doi.org/10.1073/pnas.0906412106.

Romero-Rodriguez, A., Robledo-Casados, I., and Sanchez, S. (2015). An overview on transcriptional regulators in *Streptomyces*. Biochim. Biophys. Acta *1849*, 1017–1039. https://doi.org/10.1016/j.bbagrm.2015.06.007.

Seipke, R.F. (2015). Strain-level diversity of secondary metabolism in *Streptomyces albus*. PLoS One *10*, e0116457. https://doi.org/10.1371/journal.pone.0116457.

Sottorff, I., Wiese, J., Lipfert, M., Preusske, N., Sonnichsen, F.D., and Imhoff, J.F. (2019). Different secondary metabolite profiles of phylogenetically almost identical *Streptomyces griseus* strains originating from geographically remote locations. Microorganisms *7*, 166. https://doi.org/10.3390/microorganisms7060166.

Stackebrandt, E., and Goebel, B.M. (1994). Taxonomic Note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. *44*, 846–849. https://doi.org/10.1099/00207713-44-4-846.

Thanapipatsiri, A., Gomez-Escribano, J.P., Song, L., Bibb, M.J., Al-Bassam, M., Chandra, G., Thamchaipenet, A., Challis, G.L., and Bibb, M.J. (2016). Discovery of unusual biaryl polyketides by activation of a silent *Streptomyces venezuelae* biosynthetic gene cluster. ChemBioChem *17*, 2189–2198. https://doi.org/10.1002/cbic.201600396.

Vaijayanthi, T., Bando, T., Pandian, G.N., and Sugiyama, H. (2012). Progress and prospects of pyrrole-imidazole polyamide-fluorophore conjugates as sequence-selective DNA probes. ChemBioChem *13*, 2170–2185. https://doi.org/10.1002/cbic.201200451.

Vingadassalon, A., Lorieux, F., Juguet, M., Le Goff, G., Gerbaud, C., Pernodet, J.L., and Lautru, S. (2015). Natural combinatorial biosynthesis involving two clusters for the synthesis of three pyrrolamides in *Streptomyces netropsis*. ACS Chem. Biol. *10*, 601–610. https://doi.org/10.1021/cb500652n.

Vitayakritsirikul, V., Jaemsaeng, R., Lohmaneeratana, K., Thanapipatsiri, A., Daduang, R., Chuawong, P., and Thamchaipenet, A. (2016). Improvement of chloramphenicol production in *Streptomyces venezuelae* ATCC 10712 by overexpression of the *aroB* and *aroK* genes catalysing steps in the shikimate pathway. Antonie Van Leeuwenhoek *109*, 379–388. https://doi.org/10.1007/s10482-015-0640-y.

Walsh, C.T., and Fischbach, M.A. (2010). Natural products version 2.0: connecting genes to molecules. J. Am. Chem. Soc. *132*, 2469–2493. https://doi.org/10.1021/ja909118a.

Witt, E.M., Davies, N.W., and Galinski, E.A. (2011). Unexpected property of ectoine synthase and its application for synthesis of the engineered compatible solute ADPC. Appl. Microbiol. Biotechnol. *91*, 113–122. https://doi.org/10.1007/s00253-011-3211-9.

Xu, L., Ye, K.X., Dai, W.H., Sun, C., Xu, L.H., and Han, B.N. (2019). Comparative genomic insights into secondary metabolism biosynthetic gene cluster distributions of marine *Streptomyces*. Mar. Drugs *17*, 498. https://doi.org/10.3390/md17090498.

Xue, Y., and Sherman, D.H. (2001). Biosynthesis and combinatorial biosynthesis of pikromycin-related macrolides in *Streptomyces venezuelae*. Metab. Eng. *3*, 15–26. https://doi.org/10.1006/mben.2000.0167.

Xue, Y., Zhao, L., Liu, H.W., and Sherman, D.H. (1998). A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. Proc. Natl. Acad. Sci. U S A *95*, 12111–12116.

Zakrzewski, P., Medema, M.H., Gevorgyan, A., Kierzek, A.M., Breitling, R., and Takano, E. (2012). MultiMetEval: comparative and multi-objective analysis of genome-scale metabolic models. PLoS One *7*, e51511. https://doi.org/10.1371/journal.pone.0051511.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial strains		
Escherichia coli strain ET12567/pUZ8002	Prof. Jung-Hye. Roe. Seoul National University, South Korea	N/A
Escherichia coli strain BL21 (DE3) RIPL	Agilent Technologies	Cat#230240
25 *Streptomyces* strains (see Tables 1 and 2)	American Type Culture Collection (ATCC), Korean Collection for Type Cultures (KCTC)	N/A
Chemicals, peptides, and recombinant proteins		
α-cyano-4-hydroxycinnamic acid	HCCA, Bruker Daltonics	Cat#NC1324132
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma-aldrich	Cat#I6758
TEV protease	NEB	Cat#P8112S
Glutamate	Sigma-aldrich	Cat#1294976
Critical commercial assays		
Wizard Genomic DNA Purification Kit	Promega	Cat#A1120
SMRTbell™ Template Prep Kit	Pacific Biosciences	Cat#100-259-100
TruSeq DNA-Free LT kit	Illumina	Cat#20015963
Lysozyme	Sigma-aldrich	Cat#L6876
Bruker Bacterial Test Standard	Bruker Daltonics	Cat#255343
APCI Calibration Solution	AB Sciex	Cat#5042914
Deposited data		
Streptomyces coelicolor A3(2) genome	NCBI	AL645882.2
Streptomyces lividans TK24 genome	NCBI	CP009124.1
Streptomyces griseus NBRC 13350 genome	NCBI	AP009493.1
Streptomyces avermitilis MA-4680 genome	NCBI	BA000030.4
Streptomyces clavuligerus ATCC 27064 genome	NCBI	CP027858
Streptomyces tsukubaensis NRRL 18488 genome	NCBI	CP020700
Streptomyces aureofaciens KCTC 1086 genome	NCBI	CP023698
*Streptomyces galilaeus*KCTC 1921 genome	NCBI	CP023703
Streptomyces subrutilus KCTC 9045 genome	NCBI	CP023701
Streptomyces prasinus KCTC 9019 genome	NCBI	CP023697
Streptomyces alboniger KCTC 9014 genome	NCBI	CP023695
Streptomyces cinereoruber KCTC 9706 genome	NCBI	CP023693
Streptomyces filamentosus KCTC 9568 genome	NCBI	PDCL00000000
Streptomyces vinaceus KCTC 9771 genome	NCBI	CP023692
Streptomyces chartreusis KCTC 9704 genome	NCBI	CP023689
Streptomyces venezuelae ATCC 10712 genome	NCBI	CP029197
Streptomyces venezuelae ATCC 15439 genome	NCBI	CP059991
Streptomyces venezuelae ATCC 21018 genome	NCBI	CP029189
Streptomyces venezuelae ATCC 21782 genome	NCBI	CP029190
Streptomyces venezuelae ATCC 15068 genome	NCBI	CP029194
Streptomyces venezuelae ATCC 14583 genome	NCBI	CP029193
Streptomyces venezuelae ATCC 14584 genome	NCBI	CP029192

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Streptomyces venezuelae ATCC 14585 genome	NCBI	CP029191
Streptomyces venezuelae ATCC 21113 genome	NCBI	CP029196
Streptomyces venezuelae ATCC 10595 genome	NCBI	CP029195
Oligonucleotides		
See Table S1	Integrated DNA Technologies, Macrogen	N/A
Recombinant DNA		
pJET 1.2 (CloneJET PCR Cloning Kit)	ThermoFisher Scientific	Cat#K1231
pCRISPomyces-2	Addgene	Cat#61737
pET28a	Addgene	Cat#69864-3
Software and algorithms		
MEGA 7.0.26	(Kumar et al., 2016)	https://www.megasoftware.net/
flexAnalysis 3.3	Bruker Daltonics	N/A
MALDI Biotyper 3 software	Bruker Daltonics	N/A
Hierarchical Genome Assembly Process workflow ver 2.3	(Chin et al., 2013)	https://www.pacb.com/products-and-services/analytical-software/smrt-analysis/
CLC Genomics Workbench	CLC Bio	https://www.qiagen.com/us/products/discovery-and-translational-research/next-generation-sequencing/informatics-and-data/analysis-and-visualization/clc-genomics-workbench/
MAUVE 2.4.0	(Darling et al., 2004)	http://darlinglab.org/mauve/download.html
GAP5	(Bonfield and Whitwham, 2010).	https://www.sanger.ac.uk/tool/gap5/
Analyst TF 1.6	AB Sciex	https://sciex.com/products/software/analyst-tf-software
PeakView 1.2	AB Sciex	https://sciex.com/products/software/peakview-software
ftmsControl 2.1	Bruker Daltonics	https://bruker-compass-ftmscontrol.software.informer.com/
HyStar 4.1 software	Bruker Daltonics	N/A
DataAnalysis 4.4 program	Bruker Daltonics	N/A
PYANI v0.2.9	(Pritchard et al., 2016)	https://github.com/widdowquinn/pyani/releases
antiSMASH 5.0	(Blin et al., 2019)	https://antismash.secondarymetabolites.org/#!/start
BLASTp	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Other		
UltrafleXtreme MALDI-TOF mass spectrometer equipped with a SmartBeam-II Na:YAG laser	Bruker Daltonics	N/A
Covaris	Covaris Inc.	N/A
PacBio SMRT cells with P6-C4-chemistry	Pacific Biosciences	N/A
Illumina HiSeq 2500	Illumina	N/A
Illumina MiSeq v.2	Illumina	N/A
TripleTOF 5600	AB Sciex	N/A
ACQUITY UPLC™ system	Waters	N/A
ACQUITY BEH Shield RP18 column	Waters	N/A
ACQUITY HSS T3 column	Waters	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
15 Tesla FT-ICR MS	Bruker Daltonics	N/A
Spectra/Por Float-A-Lyzer	Sigma-aldrich	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Byung-Kwan Cho (bcho@kaist.ac.kr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The genome sequence datasets used in this study are available at NCBI database (accession numbers are listed in Key resources table).
- This manuscript does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial strains and culture media

Ten *S. venezuelae* strains were purchased from the American Type Culture Collection (ATCC, Manassas, VA), whereas the other 15 streptomycetes used in this study were purchased from the Korean Collection for Type Cultures (KCTC, Daejeon, South Korea). For initial propagation, we first streaked freeze-dried cells on ISP2 agar medium (yeast extract 4 g, malt extract 10 g, glucose 4 g, agar 20 g, and distilled water 1 L). A single colony of each strain was inoculated into 50 mL ISP2 medium with 8 g of glass beads (3 $\pm$ 0.3 mm diameter) in 250 mL of a baffled flask, and grown at 30°C in a shaking incubator at 220 rpm. This culture condition was used for all liquid cultures in this study. Cell stocks were prepared with 50% glycerol and then stored at −80°C. For morphology check, cells were grown on MYM agar medium (yeast extract 4 g, malt extract 10 g, maltose 4 g, agar 20 g, and distilled water 1 L) at 30°C. For the genome sequencing, streptomycetes were grown in one of four different culture media, R5(–), 1 × sporulation, YEME, and MYM medium, which were generally used for *Streptomyces* culture (Lee et al., 2020). R5(–) medium is composed of 25 mM TES (pH 7.2), 103 g/L sucrose, 1% (w/v) glucose, 5 g/L yeast extract, 10.12 g/L $MgCl_2 \cdot 6H_2O$, 0.25 g/L K_2SO_4, 0.1 g/L casamino acids, 0.08 g/L ZnCl2, 0.4 mg/L $FeCl_3$, 0.02 mg/L $CuCl_2 \cdot 2H_2O$, 0.02 mg/L $MnCl_2 \cdot 4H_2O$, 0.02 mg/L $Na_2B_4O_7 \cdot 10H_2O$, and 0.02 mg/L $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$). 1 × sporulation medium is composed of 3.33 g/L glucose, 1 g/L yeast extract, 1 g/L beef extract, 2 g/L tryptose, and 0.006 g/L $FeSO_4 \cdot 7H_2O$). YEME medium is composed of 340 g/L sucrose, 10 g/L glucose, 3 g/L yeast extract, 5 g/L bacto peptone, and 3 g/L oxoid malt extract. MYM medium is composed of 4 g/L maltose, 4 g/L yeast extract, and 10 g/L malt extract.

METHOD DETAILS

16S rRNA sequence-based phylogenetic analysis

For phylogenetic analysis, 16S rRNA sequences of 25 *Streptomyces* were downloaded from the NCBI database (Table 2). 16S rRNA gene sequences of eight *S. venezuelae* strains out of 25 were verified using PCR and Sanger sequencing method (Figure S1). 16S rRNA regions were PCR amplified from the genomic DNA (gDNA) samples and cloned into the pJET 1.2 vector using the CloneJET PCR Cloning Kit (Thermo, Waltham, MA) following the manufacturer's instructions. Sanger sequencing of the amplified sequences was conducted using both pJET 1.2 universal forward and reverse primers (Table S1). All 16S rRNA sequences were aligned based on the CLUSTAL W algorithm using MEGA 7.0.26 (Kumar et al., 2016). A phylogenetic tree was constructed using the maximum-likelihood method with the Kimura 2-parameter model and 'Use all sites' mode. To evaluate the phylogeny, the bootstrap method was used with n = 1000.

MALDI-TOF MS protein profile-based phylogenetic analysis

For sample preparation, the conventional formic acid extraction method was used in this study (Lee et al., 2017). Briefly, 25 streptomycetes, cultivated in ISP2 media for 16 h at 37°C, were placed and pelleted in microtubes, and then resuspended in 25 μL of 70% (v/v) formic acid. One equivalent of acetonitrile was added to the suspension. After gentle mixing, the solution was centrifuged at 15,871 *g* for 2 min at 4°C, and the supernatant containing bacterial protein extracts was transferred to a clean vial for MALDI MS analysis. Acquisition of microbial protein profiles was performed on an ultrafleXtreme MALDI-TOF mass spectrometer equipped with a SmartBeam-II Na:YAG laser (355 nm, Bruker Daltonics, Billerica, MA) operating in linear and positive ion modes (Jang et al., 2017). Bacterial extracts were spotted on a MALDI target followed by α-cyano-4-hydroxycinnamic acid (HCCA, Bruker Daltonics) matrix solution (i.e., 10 mg/mL in 50% acetonitrile/47.5% water/2.5% trifluoroacetic acid (v/v/v), 1 μL). After air drying the samples at room temperature, the samples were subjected to MALDI MS analysis. The instrument was externally calibrated using a bacterial test standard (Bruker Daltonics) solution prior to sample analysis. Mass spectra of the colonies were acquired from 1000 laser shots per spot for each strain. The MS spectra of each strain were obtained using flexAnalysis 3.3 (Bruker Daltonics) in biological duplicates and technical triplicates, and then integrated into the main spectrum profiles (MSPs) using the MALDI Biotyper 3 software (Bruker Daltonics). A dendrogram of 25 *Streptomyces* strains was created from the MSPs using Euclidean and average. According to the guidelines of the vendor, a score of ≥ 2 depicts identification to the species level and an intermediate log score between < 2 and ≥ 1.7 for identification to the genus level. A low score of < 1.7 was regarded as unreliable for identification.

Genome sequencing, assembly, and annotation

Among the 25 streptomycetes genomes studied, 21 genomes were reported in our previous study (Lee et al., 2020), except *S. coelicolor* A3(2), *S. lividans* TK24, *S. griseus* NBRC 13350, and *S. avermitilis* MA-4680. For gDNA extraction, 25 mL of the streptomycetes cultures were harvested at exponential growth and washed twice with 10 mM EDTA. After lysozyme treatment (10 mg/mL) at 37°C for 45 min, gDNA was extracted using the Wizard Genomic DNA Purification Kit (Promega, Madison, NJ). Extracted gDNA was sequenced using a hybrid strategy exploiting both long-read and short-read genome sequencing methods. PacBio genome-seq libraries were prepared using SMRTbell™ as per the manufacturer's protocol and sequenced using SMRT cells with P6-C4-chemistry (DNA Sequencing Reagent 4.0; Pacific Biosciences, Menlo Park, CA). The PacBio sequencing reads were filtered and assembled using the Hierarchical Genome Assembly Process workflow (version 2.3; Pacific Biosciences). For Illumina short-read sequencing libraries, gDNA was fragmented into approximately 350 bp size using Covaris (Covaris Inc., Woburn, MA), followed by library construction using TruSeq DNA-Free LT kit (Illumina, La Jolla, CA). Twenty-one libraries were sequenced with the HiSeq 2500 (Illumina) with a 100 bp single-read running scale, except for *S. tsukubaensis*, which was sequenced with the MiSeq v.2 (Illumina) with a 50 bp single-read recipe. Illumina sequencing reads were assembled using the *de novo* assembly function of CLC Genomics Workbench (CLC Bio, Aarhus, Denmark) by default setting. Assembled Illumina contigs were aligned with the PacBio contigs using MAUVE 2.4.0 and linked PacBio contigs and/or extended 5′ and/or 3′ using GAP5 program (Staden package) (Bonfield and Whitwham, 2010). Twenty-one complete genomes were annotated by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 4.5 (Haft et al., 2018). Entire genome sequences are available at NCBI under the accession numbers listed in Key resources table.

Targeted LC-MS/MS analysis for putative secondary metabolite measurement

Secondary metabolite production in *S. venezuelae* strains was identified using a TripleTOF 5600 System, a hybrid triple quadrupole TOF mass spectrometer (AB Sciex, Framingham, MA) coupled with an ACQUITY UPLC™ system (Waters, Milford, MA). The extracts (2–5 μL of each) were separated on an ACQUITY BEH Shield RP18 column (1.7 μm, 2.1 × 100 mm) at a flow rate of 200 μL/min at 30°C. The gradient was established using mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile): 5% B for 2.5 min, 5–60% B for 10.5 min, ramping up to 95% B and holding for 4 min, and back to 5% B for 3 min for column conditioning. The eluent was introduced into the TripleTOF MS through a DuoSpray ion source with an electrospray potential of 4.5 kV for positive ion mode. The optimal conditions for targeted MS and MS/MS spectra were as follows: source temperature of 500°C, nebulizer gas (GS1 and GS2) of 50 psi, and declustering potential and collision energy of 60 V and 25 ± 15 V, respectively. The full scan and product ion scan ranges were set from 50 to 1,000 m/z. Mass calibration in MS and MS/MS modes was performed after every five injections with a set of standards (APCI Calibration Solution, AB

Sciex) to ensure accuracy during the LC-MS/MS analysis. Data acquisition and processing were managed using Analyst TF 1.6 and PeakView 1.2 software.

Determination of anthelvencin production using high-resolution FT-ICR MS

For determining anthelvencin production of ATCC 14583, 14584, and 14585 strains, a 15 Tesla FT-ICR MS (solariX XR^{TM} system, Bruker Daltonics) equipped with a standard ESI interface was used with the continuous accumulation of selected ions (CASI) function in positive ion mode (Q1 mass of m/z 400 and an isolation window of 350 Da). Chromatographic separation of anthelvencin was performed as a putative secondary metabolite. The eluent that came from the ACQUITY UPLC system was introduced into the 15T FT-ICR MS to acquire high-resolution MS and MS/MS spectra in positive ion mode. The MS parameters for ESI(+) FT-ICR MS were a drying gas of 6 L/min, a drying gas temperature of 220°C, a skimmer voltage of 45 V, a collision gas energy of −3.0 V, an accumulation time of 400 ms, a transient length of 2.73 s, a scan number of 1, and an acquisition size of 2 MB with a sine-bell apodization window function applied to the time-domain signal. External calibration was performed with linear regression using a 100 μg/mL NaTFA solution. Data acquisition was controlled using the ftmsControl 2.1 and HyStar 4.1 software (Bruker Daltonics), and data processing for the selection of molecular features was performed using the DataAnalysis 4.4 program (Bruker Daltonics).

Disruption of anthelvencin BGC-encoded genes using CRISPR/Cas9 system

Dual-guide RNA cassettes were designed and cloned into the BbsI site of the pCRISPomyces-2 vector following the established protocol (Cobb et al., 2015). For the homologous arm, 1 kb downstream and upstream region of each *athv10*, *athv15*, *athv19*, and *athv28* were PCR amplified from the gDNA of ATCC 14585 strain to have 30 bp overlap at the junction of fragments. The two fragments were linked by overlapping extension PCR and were cloned into the XbaI site of the pCRISPomyces-2 vector, followed by transformation into the methylation-deficient *Escherichia coli* strain ET12567/pUZ8002 (MacNeil et al., 1992). Demethylated vectors were introduced into the ATCC 14585 strain through intergeneric conjugation (Kieser et al., 2000). Several transformants were cultured at 30°C for 2 days on MYM solid media containing 50 μg/mL of apramycin. The transformants were moved to the MYM solid media without apramycin and cultured at 37°C for vector curing. Finally, gDNA of each transformant was isolated using a Wizard Genomic DNA Purification Kit (Promega), and gene deletion was verified using PCR. The sequences of all oligonucleotides used in this study are listed in the Table S1.

Expression and purification of Athv28

The genes encoding Athv28 of ATCC 14583, 14584, and 14585 strains were codon optimized for *E. coli* K strain and synthesized at Integrated DNA Technologies (IDT, Coralville, IA) (Table S1). The synthesized genes were cloned into the pET28a plasmid that was modified to tag the protein with N-terminal $(His)_{10}$-linked MBP and expressed in the *E. coli* BL21 (DE3) RIPL strain. The expression strains were cultured in 1 L LB medium at 37°C for 4 h after induction with 0.5 mM IPTG. Cells were harvested using centrifugation at 10,000 ×g at 4°C for 10 min and resuspended in a lysis buffer containing 20 mM Tris-HCl (pH 8.0) and 300 mM NaCl. After sonication on ice, cell lysates were collected using centrifugation at 10,000 ×g at 4°C for 10 min. Proteins were purified from the cell lysates using Ni-NTA affinity chromatography. To remove the N-terminal tag from the proteins, 1.5 mg TEV protease was added to purified proteins at 4°C for 16 h, followed by Ni-NTA affinity chromatography. Final proteins were dialyzed against lysis buffer using a Spectra/Por Float-A-Lyzer (3 mL, 8–10 kDa MWCO).

In vitro assay with Athv28 and measuring the ADPC

In vitro assays were performed in 100 μL volume containing 10 μg of purified Athv28 enzyme, 10 mM glutamine, 20 mM Tris-HCl (pH 8.0), and 300 mM NaCl. Ten microliters of reaction mixture were harvested at 0 h, 3 h, and 6 h of incubation time at 30°C and mixed with 100 μL of methanol to quench the reaction. After incubation at –80°C for 2 h, 100 μL of the supernatant was collected using centrifugation at 13,000 ×g at 4°C for 10 min and diluted with 200 μL distilled water. ADPC and glutamine levels were measured using ESI(+) FT-ICR MS and an ACQUITY UPLC^{TM} system with an ACQUITY HSS T3 column (1.8 μm, 2.1 × 100 mm, Waters) after injecting 2 μL of each sample. Chromatographic separation of ADPC and glutamate was performed as follows. The column was maintained at 25°C, and the mobile phase consisted of water containing 0.1% formic acid (solvent A) and acetonitrile containing 0.1% formic acid (solvent B). The solvent gradient started with 5% B for 0.5 min, followed by a linear increase of 5%–70% B from 0.5-5 min, ramped up

to 98% B, held at 98% B for 3 min, returned to 5% B, and finally held for 3.0 min at a flow rate of 200 μL/min. The effluent was monitored with 15T FT-ICR MS. The MS parameters of the FT-ICR MS were the same as those for the anthelvencin measurement, except for the CASI function (Q1 mass of m/z 160 and an isolation window of 100 Da) and a skimmer voltage of 15 V.

QUANTIFICATION AND STATISTICAL ANALYSIS

A phylogenetic tree of 16S rRNA sequences was constructed using the maximum-likelihood method with the Kimura 2-parameter model and 'Use all sites' mode. To evaluate the phylogeny, the bootstrap method was used with n = 1000. For phylogenetic analysis of genome sequences, the MUMmer ultra-rapid aligning tool (ANIm) values between 25 genome sequences were calculated using PYANI v0.2.9. The MALDI-TOF MS spectra were obtained using flexAnalysis 3.3 (Bruker Daltonics) in biological duplicates and technical triplicates, and then integrated into the main spectrum profiles (MSPs) using the MALDI Biotyper 3 software (Bruker Daltonics). A dendrogram of 25 *Streptomyces* strains was created from the MSPs using Euclidean and average. According to the guidelines of the vendor, a score of $\geq$ 2 depicts identification to the species level and an intermediate log score between < 2 and $\geq$ 1.7 for identification to the genus level. A low score of < 1.7 was regarded as unreliable for identification. For targeted LC-MS/MS analysis, mass calibration in MS and MS/MS modes was performed after every five injections with a set of standards (APCI Calibration Solution, AB Sciex) to ensure accuracy during the LC-MS/MS analysis. Data acquisition and processing were managed using Analyst TF 1.6 and PeakView 1.2 software. For anthelvencin measurement by FT-ICR MS, external calibration was performed with linear regression using a 100 μg/mL NaTFA solution. Data acquisition was controlled using the ftmsControl 2.1 and HyStar 4.1 software (Bruker Daltonics), and data processing for the selection of molecular features was performed using the DataAnalysis 4.4 program (Bruker Daltonics).