fermentation

MDPI

Article

Profile of PKS and NRPS Gene Clusters in the Genome of *Streptomyces cellostaticus* NBRC 12849^T

Hisayuki Komaki * and Tomohiko Tamura

Biological Resource Center, National Institute of Technology and Evaluation (NBRC), Chiba 292-0818, Japan; tamura-tomohiko@nite.go.jp

* Correspondence: komaki-hisayuki@nite.go.jp

Abstract: Polyketides and nonribosomal peptides are major secondary metabolites in members of the genus *Streptomyces*. *Streptomyces cellostaticus* is a validly recognized species and the type strain produces cellostatin. However, little is known about whether it has the potential to produce diverse polyketides and nonribosomal peptides. Here, we sequenced the whole genome of *S. cellostaticus* NBRC 12849^T and surveyed polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) gene clusters in the genome. The genome encoded 12 PKS, one NRPS and eight hybrid PKS/NRPS gene clusters. Among the 21 gene clusters, products of 10 gene clusters were annotated to be an annimycin congener, fuelimycins, lankamycin, streptovaricin, spore pigment, flaviolin, foxicin, blasticidin, lankacidin and an incarnatapeptine congener via our bioinformatic analysis. Although the other clusters were orphan and their products were unknown, five of them were predicted to be compounds derived from two independent diketides, a tridecaketide, a triketide and a tetraketide with a cysteine residue, respectively. These results suggest that *S. cellostaticus* is a source of diverse polyketides and hybrid polyketide/nonribosomal peptides, including unknown and new secondary metabolites.

Keywords: blasticidin; genome; lankacidin; lankamycin; nonribosomal peptide; polyketide; secondary metabolite; *Streptomyces*

Citation: Komaki, H.; Tamura, T. Profile of PKS and NRPS Gene Clusters in the Genome of *Streptomyces cellostaticus* NBRC 12849^T. *Fermentation* **2023**, *9*, 924. https://doi.org/10.3390/fermentation9110924

Academic Editors: Evgenya V. Prazdnova and Yihua Chen

Received: 10 July 2023
Revised: 25 September 2023
Accepted: 11 October 2023
Published: 24 October 2023

1. Introduction

Members of the genus *Streptomyces* are well-known as a rich source of pharmaceutically useful bioactive secondary metabolites [1]. Polyketides and nonribosomal peptides are major secondary metabolites in this genus because half to three quarters of secondary metabolite-biosynthetic gene clusters in each streptomycetal genome generally include polyketide synthase (PKS) and/or nonribosomal peptide synthetase (NRPS) genes. Each *Streptomyces* strain harbors a few dozen biosynthetic gene clusters for polyketides and nonribosomal peptides in its genome [2]. Which PKS and NRPS gene clusters are present is, in principle, related to which species the strain belongs to [3]. Hence, the profile of these gene clusters in each species is useful information for searching for new secondary metabolites. Recently, genome mining has often been employed to find new compounds [4], for which the profile of PKS and NRPS gene clusters provide a useful information.

PKSs are responsible for synthesizing polyketide chains from acyl-CoA molecules as the building blocks. NRPSs synthesize peptide chains from amino acids in a similar manner. The diversity of these chemical backbones is derived from differences in the chain lengths and the complex of various building blocks, in addition to modifications such as reduction, cyclization, methylation, and epimerization. Type-I PKSs and NRPSs are large modular enzymes with multiple catalytic domains. As chain elongations by type-I PKSs and NRPSs are based on the co-linearity rule of assembly lines [5], the chemical structures of the chains are determined by the domain organizations. In contrast, type-II and type-III PKSs iteratively catalyze polyketide chain elongations. Type-II PKSs are composed of three small enzymes, called minimal PKSs, which are ketosynthase (KS)α, KSβ (chain length

factor) and acyl carrier protein (ACP), and they are involved in the synthesis of aromatic compounds. Type-III PKSs are standalone enzymes with a KS domain and responsible for the synthesis of chalcone-like phenolic compounds. PKS and NRPS gene clusters are often surveyed to evaluate the potential of strains as a source for structurally diverse and pharmaceutically useful secondary metabolites.

Streptomyces cellostaticus was proposed by Hamada in 1958. The type strain produces cellostatin, an antifungal substance, but its chemical structure has not been elucidated yet [6]. No other compound has been reported from this species, although strains in an *S. cellostaticus* clade were reported to produce bafilomycins [7]. To investigate whether *S. cellostaticus* can be expected as a promising source for diverse polyketides and nonribosomal peptides, we conducted whole genome sequencing for the type strain of this species and analyzed its PKS and NRPS gene clusters in the genome.

2. Materials and Methods

2.1. Streptomyces cellostaticus NBRC 12849^T Growth Conditions

Streptomyces cellostaticus NBRC 12849^T was distributed from the NBRC Culture Collection. It was cultured in 15 mL YG medium (1% yeast extract, 1% glucose, pH 7.0) in three test tubes (diameter, 18 mm; length, 16.5 cm; 5 mL/tube) with shaking (260 strokes/min) at 28 °C using an incubation shaker TC-500R (Takasaki Kagaku Kiki Co., Ltd., Saitama, Japan) for two days to prepare genomic DNA whereas it was cultured on a modified A3M agar medium (0.5% glucose, 2% glycerol, 2% soluble starch, 1.5% Pharmamedia (Traders Protein, Memphis, TN, USA), 0.3% yeast extract, pH 7.0, 1.5% agar) for two weeks at 28 °C for production of blasticidin A.

2.2. Phylogenetic Analysis

The 16S rRNA gene sequences were retrieved from the GenBank or EzBioCloud server. Closely related strains were searched by using "16S-based ID" in the EzBioCloud server [8]. Phylogenetic trees were reconstructed using ClustalX [9] based on the neighbor-joining method created by Saiou and Nei [10]. A multilocus sequence analysis (MLSA) was conducted using concatenated *atpD*, *gyrB*, *recA*, *rpoB* and *trpB* gene sequences of 2479 nucleotides.

2.3. DNA Extraction

The cells cultured in 15 mL YG medium were harvested via centrifugation (5255× *g*, 10 min, 4 °C) and preserved at −20 °C until genomic DNA preparation. Genomic DNA of *S. cellostaticus* NBRC 12849^T for whole genome sequencing was prepared using phenol treatment [11].

2.4. Whole Genome Sequencing

The whole genome was sequenced by Kazusa DNA Research Institute using PacBio [12]. Library was prepared using the BluePippin system (Sage Science, Beverly, MA, USA) with a SMRTbell Template Prep Kit 1.0, a SMRTbell Damage Repair Kit and AMPure (Pacific Biosciences, Kanagawa, Japan). Sequencing was conducted using the Sequel system with Sequel SMRT cell 1M versions 2 and 3, Sequel Sequencing Kits 2.1 and 3.0, a Sequel Binding Kit 2.0 and a Sequel Binding and Internal Ctrl Kit 3.0 (Pacific Biosciences). The resulting reads were assembled using SMRT Link version 6.0 (Pacific Biosciences). DFAST [13] was used for annotation. The draft genome sequence has been deposited under the accession number BNDU01000001–BNDU01000013.

2.5. Annotation of PKS and NRPS Gene Clusters

PKS and NRPS gene clusters in the genome were searched for using antiSMASH [14] and manually annotated in the same manner of our previous report [12]. Orthologous gene clusters were searched for using BLASTp on the NCBI website. Putative products of the

gene clusters were predicted according to the domain organizations [5] and/or orthologs searched for using BLAST.

2.6. Detection of Blasticidin A Produced by S. cellostaticus NBRC 12849^T

The cells cultured on A3M agar medium were collected with agar in a vessel and two times the volume of acetone was added to it. The mixture was kept at 4 °C overnight for extraction. Ten microliters of the supernatant were analyzed using an UHPLC system coupled with a mass spectrometer (LC-MS) (UltiMate 3000 UHPLC coupled with Q Exactive, Thermo Fisher Scientific K.K., Tokyo, Japan). Acquity UPLC BEH C18 1.7 µm (2.1 × 5 mm) (Nihon Waters K.K., Tokyo, Japan) was used as a reverse phase column for separation in the system. Water (solvent A) and acetonitrile (solvent B), both containing 0.1% (*v*/*v*) formic acid, were used as the mobile phase in the following linear gradient program: 5% B for 0.5 min, 5% B to 85% B in 5 min, 85% B to 100% B in 0.5 min, 100% B for 2 min. The flow rate was set to 0.6 mL/min and the column oven temperature was set at 40 °C. Compounds in the eluate were detected in the ESI positive- and negative-ion modes with a spray voltage of 3.5 kV and a capillary temperature of 300 °C. Flow rates of nitrogen sheath gas and auxiliary gas were set at 55 and 15 arbitrary units, respectively. A full MS scan was performed in the range of 150–2000 (*m*/*z*) at a resolution of 70,000. Data were acquired with Xcalibur 2.0 software (Thermo Fisher Scientific K.K., Waltham, MA, USA). Purified blasticidin A was kindly distributed by Prof. Shohei Sakuda, Teikyo University, Japan, and solved in methanol at a concentration of 10 µg/mL. The 1 µL was analyzed as a standard using LC-MS.

3. Results

3.1. Phylogenetic Position of S. cellostaticus

In the 16S rRNA gene sequence analysis, *S. cellostaticus* NBRC 12849^T showed 99.7%, 99.5%, 99.5%, 99.2% and 99.0% similarities to *Streptomyces griseochromogenes* ATCC 14511^T, *Streptomyces yokosukanensis* DSM 40224^T, *Streptomyces cinnabarigriseus* JS360^T, *Streptomyces achromogenes* NBRC 12735^T and *Streptomyces griseoruber* NRRL B-1818^T, respectively. Sequence similarities to the other species were less than 99.0%. In the phylogenetic tree based on 16S rRNA gene sequences, *S. cellostaticus* formed a clade with *S. griseochromogenes* and *S. yokosukanensis* as shown in Figure 1. These two species were phylogenetic neighbors of *S. cellostaticus*.

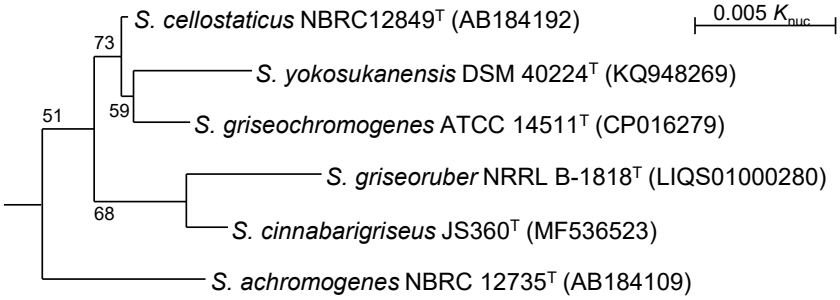

Figure 1. Phylogenetic tree based on 16S rRNA gene sequences. Numbers on the branches are the confidence limits estimated through bootstrap analysis with 1000 replicates. Values above 50% are indicated at branching points. *Kitasatospora viridis* DSM 44826^T (VIWT01000001) was used as an outgroup (not shown) to show the root.

In the MLSA, as *S. cellostaticus* formed a clade with *S. griseochromogenes* (Figure 2), *S. griseochromogenes* was phylogenetically the closest to *S. cellostaticus*.

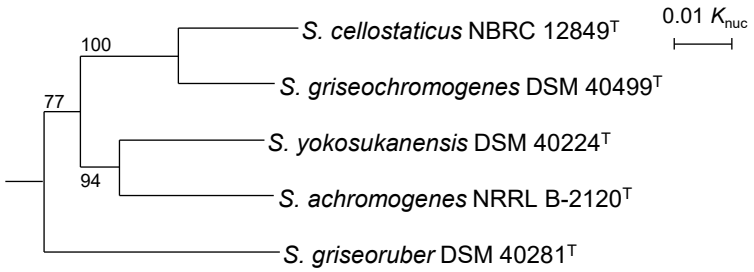

Figure 2. Phylogenetic tree based on MLSA using five housekeeping gene sequences. Numbers on the branches are the confidence limits estimated through bootstrap analysis with 1000 replicates. Values above 50% are indicated at the branching points. *Streptomyces albus* NBRC 13014^T were used as an outgroup (not shown) to show the root. Nucleotides accession numbers of *atpD, gyrB, recA, rpoB* and *trpB* are as follows: *Streptomyces achromogenes* NRRL B-2120^T, JODT01000015, JODT01000026, JODT01000013, JODT01000001 and JODT01000006; *Streptomyces albus* NBRC 13014^T, BBQG01000033, BBQG01000007, BBQG01000035, BBQG01000035 and BBQG01000017; *S. cellostaticus* NBRC 12849^T, BNDU01000006; *Streptomyces griseochromogenes* DSM 40499^T, JAGGLP010000014, JAGGLP010000007, JAGGLP010000030, JAGGLP010000004 and JAGGLP010000003; *Streptomyces griseoruber* DSM 40281^T, LMWW01000074, LMWW01000034, LMWW01000078, LMWW01000068 and LMWW01000008; *Streptomyces yokosukanensis* DSM 40224^T, LMWN01000038, LMWN01000001, LMWN01000048, LMWN01000007 and LMWN01000032.

3.2. PKS and NRPS Gene Clusters in S. cellostaticus NBRC 12849^T

We sequenced the whole genome of *Streptomyces cellostaticus* NBRC 12849^T using PacBio and consequently obtained a draft genome sequence composed of 13 contig sequences (Table 1). The genome size and G + C content were 9.9 Mbp and 70.9%, respectively.

Table 1. Draft genome sequence of *S. cellostaticus* NBRC 12849^T.

Contig	Accession No.	Length (bp)	G + C Content
sequence01	BNDU01000001.1	15,464	69.9%
sequence02	BNDU01000002.1	520,187	70.9%
sequence03	BNDU01000003.1	651,433	71.4%
sequence04	BNDU01000004.1	1,755,540	71.3%
sequence05	BNDU01000005.1	142,300	72.0%
sequence06	BNDU01000006.1	6,225,385	70.8%
sequence07	BNDU01000007.1	60,471	70.7%
sequence08	BNDU01000008.1	211,191	69.5%
sequence09	BNDU01000009.1	59,991	70.5%
sequence10	BNDU01000010.1	101,974	72.9%
sequence11	BNDU01000011.1	15,410	68.2%
sequence12	BNDU01000012.1	123,546	69.9%
sequence13	BNDU01000013.1	20,918	70.1%
	Total	9,903,810	70.9%

The genome encoded ten type-I PKS gene clusters (*t1pks*), one type-II PKS gene cluster (*t2pks*), one type-III PKS gene cluster (*t3pks*), one NRPS gene cluster (*nrps*) and eight hybrid PKS/NRPS gene clusters (*pks/nrps*) as listed in Table S1. These numbers were compared with those of PKS and NRPS gene clusters in the *Streptomyces* strains that we sequenced the whole genomes of in previous studies such as cited ones [12,15,16] (Figure 3). *S. cellostaticus* NBRC 12849^T has the fewest NRPS gene clusters (pink bar). However, this strain is more prevalent in the total number of PKS, NRPS and hybrid PKS/NRPS gene clusters. Especially, the number of hybrid PKS/NRPS gene clusters was the highest (yellow bar) among our whole genome-sequenced *Streptomyces* strains.

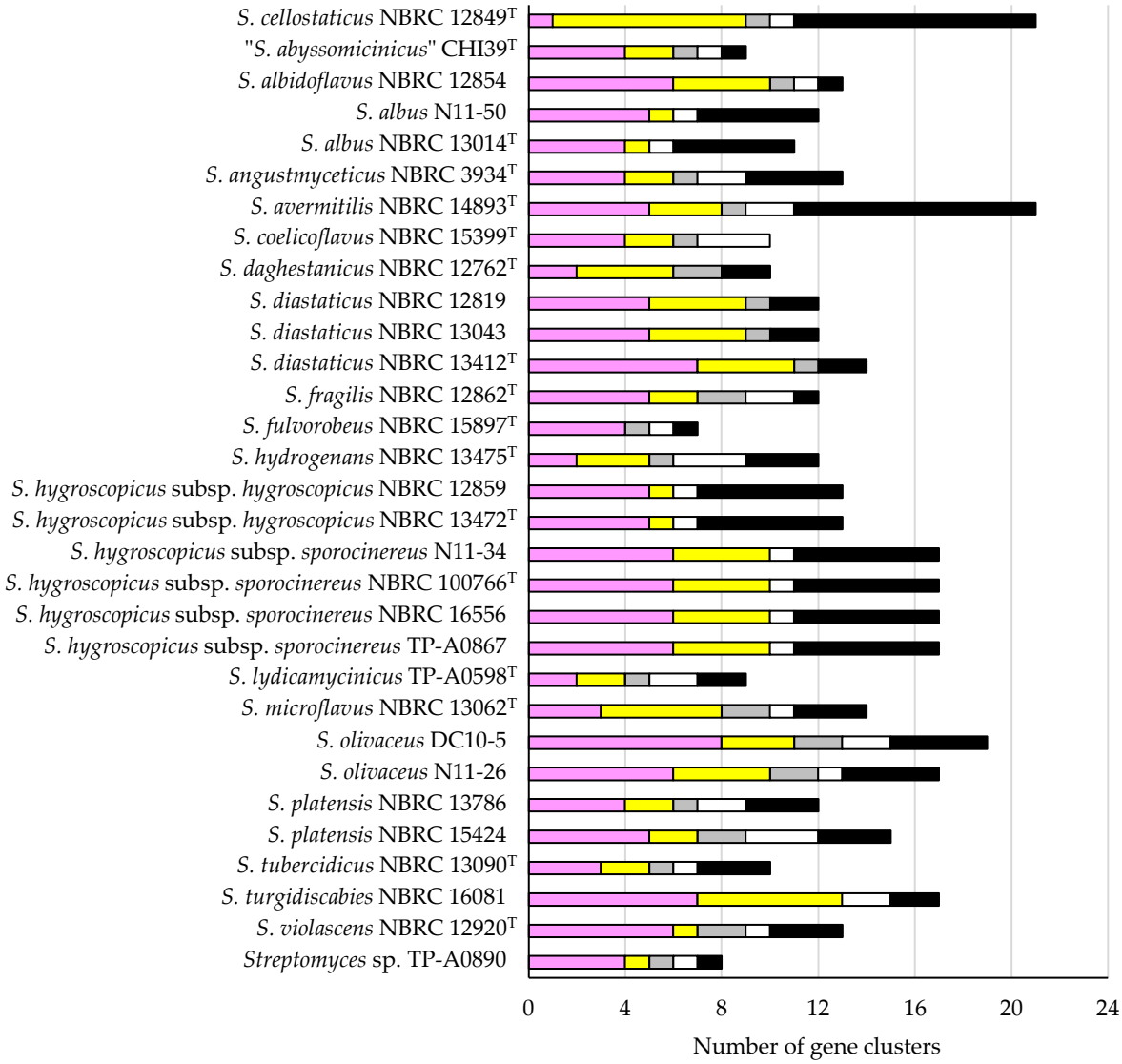

Figure 3. Number of PKS and NRPS gene clusters in the genome of whole genome-sequenced *Streptomyces* strains. Pink, NRPS gene cluster; yellow, hybrid PKS/NRPS gene cluster; gray, type-III PKS gene cluster; white, type-II PKS gene cluster; black, type-I PKS gene cluster.

3.3. Annotation of Each Gene Cluster in S. cellostaticus NBRC 12849^T

Cluster 1 was a type-I PKS gene cluster (*t1pks*), where two PKSs are encoded as shown in Figure 4i. They did not show high sequence similarities with published PKSs, suggesting that this cluster is orphan, whose product is unknown. According to the domain organization, the polyketide chain of this cluster was predicted to be diketide as shown in Figure 5a, although the function of a ketosynthase (KS) domain present at the terminal of Scel_01590 is unclear. The final product was predicted to be a compound derived from this diketide chain.

Figure 4. PKS and NRPS gene clusters of *S. cellostaticus* NBRC 12849^T. Clusters 5, 7, 10 and 12 are not included here because they are common *t2pks*, *t3pks*, or a cluster with only a single type-I PKS gene. Clusters shown in blue are orthologs, which were used to speculate the ORFs in the gap, shown by gray dashed lines, which could not be sequenced in this study. ORFs that do not encode PKS or NRPS are shown in gray.

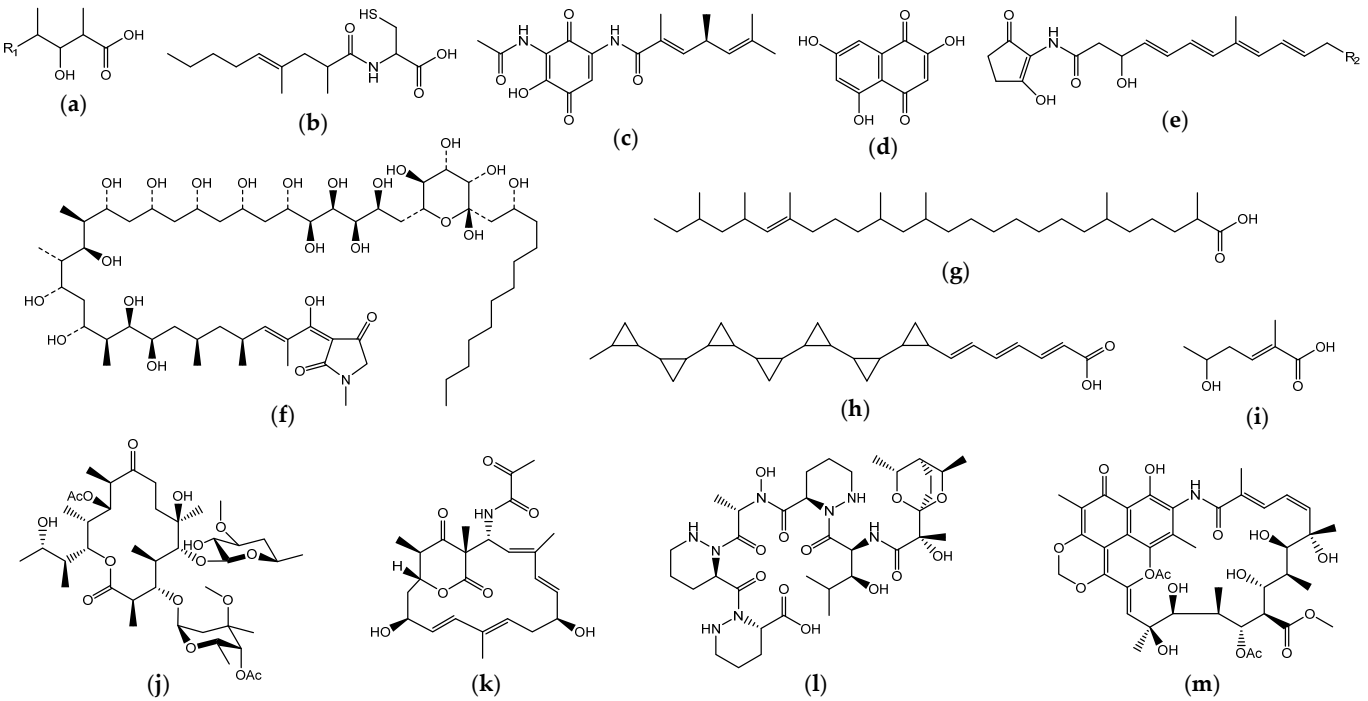

Figure 5. Chemical structures of putative products. (**a**) Diketide backbone derived from *t1pks-1* (R_1 = moiety derived from a starter molecule); (**b**) tetraketide backbone with a cysteine residue derived from *pks/nrps-1*; (**c**) foxicin A; (**d**) flaviolin; (**e**) annimycin (R_2 = Me) and its congener derived from *t1pks-3* (R_2 = H); (**f**) blasticidin A; (**g**) polyketide backbone derived from *t1pks-5*; (**h**) fuelimycin A; (**i**) triketide backbone derived from *t1pks-8*; (**j**) lankamycin; (**k**) lankacidin C; (**l**) incarnatapeptin A; (**m**) streptovarcin A.

Cluster 2 was a hybrid PKS/NRPS gene cluster (*pks/nrps*), which was observed at the terminals of contigs sequence02 and sequence12 (Figure 4ii). This suggests that its sequence was not completely determined by the presence of the gap between the two contigs in the draft genome sequence. The closest orthologs of Scel_04715 and Scel_87535 were a PKS (AYN42590) and a hybrid PKS/NRPS (AYN42588) of *Streptomyces dangxiongensis* Z022^T, respectively. In the genome sequence of *S. dangxiongensis* Z022^T, AYN42588 and AYN42590 form an orphan *pks/nrps* composed of two PKSs (AYN42590, AYN42589) and one hybrid PKS/NRPS (AYN42588), whose domain organizations were KS/AT_{mm}/ACP-KS/AT_m/DH/ER/KR/ACP, KS/AT_{mm} and DH/KR/ACP-KS/AT_{mm}/DH/ACP-C/A_{Cys}/T-TE, respectively. Although the gap could not be sequenced here, the domain organization of Cluster 2 was the same as that of the *pks/nrps* (AYN42588 to AYN42590) of *S. dangxiongensis* Z022^T (Figure 4ii). Hence, the domain organization of Cluster 2 was assumed to be identical to that of the *pks/nrps*. Based on the domain organization, we predicted that Cluster 2 synthesizes the tetraketide chain with a cysteine residue as shown in Figure 5b, from which the final product will yield.

Cluster 3 was a *pks/nrps* (Figure 4iii) and identified as a biosynthetic gene cluster (BGC) for foxicin (Figure 5c) based on the similarity to the *fox* gene cluster [15].

Cluster 4 was a *pks/nrps*, which is composed of six monomeric PKSs and two NRPSs with a single or two modules (Figure 4iv). Although its ortholog was present in *S. griseochromogenes* ATCC 14511^T, Cluster 4 was orphan because it did not show a high similarity to the BGCs whose products have been identified. We could not predict its product because it lacked condensation (C) and acyltransferase (AT) domains, and its modules for the assembly of the skeleton were unclear.

Cluster 5 was a common type-III PKS gene cluster (*t3pks*) and identified as a flaviolin-BGC [16], according to a high sequence similarity.

Cluster 6 was a *t1pks* composed of three PKS genes (Figure 4v). The PKSs formed two modules (KS/AT_m/ACP/KR, KS/AT_{mx}/ACP), but the position of the ketoreductase (KR) domain in Scel_76540 was unusual. The product of Cluster 6 would be derived from a diketide composed of one malonyl-CoA and one methoxymalonyl-CoA.

Cluster 7 was a common type-II PKS gene cluster (*t2pks*) (Table S1) for spore pigment [17].

Cluster 8 was a *t1pks* encoding two PKSs (Figure 4vi). Although the PKSs showed high sequence similarities to those in the annimycin-BGC [18], the substrates of their first AT domains differed, which were malonyl-CoA and methylmalonyl-CoA in Scel_36700 and Ann4 of annimycin-BGC, respectively. Therefore, the product of Cluster 8 was predicted to be an annimycin congener as shown in Figure 5e (R_2 = H).

Cluster 9 was a large *pks*/*nrps*, found in contigs sequence06, sequence07 and sequence03 (black in Figure 4vii) although whole the sequence was not determined here due to the two gaps between sequence06 and sequence07 and between sequence07 and sequence03 in the draft genome sequence (gray dot lines in Figure 4vii). Although this cluster was orphan, we annotated it as a BGC for blasticidin A [19] by comparison with its orthologous cluster found in *Streptomyces griseochromogenes* ATCC 14511^T (blue in Figure 4vii), as stated in the next subsection.

Cluster 10 was a *t1pks* encoding one PKS whose domains were KS/AT/DH/KR/ACP (Table S1). We could not predict its product because it was orphan and included only a single module.

Cluster 11 was a *t1pks*, where six PKSs were observed in the terminals of contigs sequence03, sequence13 and sequence04. The gaps between these contigs could not be sequenced in the draft genome sequence (Figure 4viii). The closest orthologs of PKSs in this cluster were those of *Kitasatospora viridis* DSM 44826^T, whose accession numbers and locus tags are TWF73701 to TWF73704 (Table S1) and FHX73_15317 to FHX73_15330 (blue in Figure 4viii), respectively. According to the similarity, we speculated that domain organizations of Scel_09990, Scel_88580 and Scel_10010, encoded at the terminals of contigs sequence03, sequence13 and sequence04, respectively, were KS/AT_m/DH/ER/KR/ACP, KS/AT_{mm}/DH/ER/KR/ACP-KS/AT_{mm}/DH/ER/KR/ACP-KS/AT_m/DH/ER/KR/ACP and KS/AT_m/DH/ER/KR/ACP-KS/AT_m/DH/ER/KR/ACP-KS/AT_{mm}/DH/KR/ACP-KS/AT_{mm}/DH/ER/KR/ACP-KS/AT_{mm}/DH/ER/KR/ACP (parts that could not be sequenced in this study and not shown in Table S1 are underlined). Based on these domain organizations, its product was predicted to be a compound derived from a tridecaketide backbone as shown in Figure 5g.

Cluster 12 was a *t1pks* encoding one PKS whose domain organization is KS/AT_m/ACP-KR (Table S1). As this cluster was orphan, the product was still unknown. We could not predict the polyketide chain biosynthesized by the PKS (Scel_10630) because this PKS did not harbor multiple modules.

Cluster 13 was a *t1pks* encoding two PKSs (Figure 4ix), showing high similarity to the BGC of fuelimycins [20] (Table S1). Hence, its product was predicted to be fuelimycins.

Cluster 14 was a *pks*/*nrps* (Figure 4x), which is orphan. Although its backbone seemed to be derived from a starter molecule, a valine residue and malonyl-CoA(s), the product was unknown.

Cluster 15 was the only NRPS gene cluster without any PKS genes in this strain (Figure 4xi). But it is unclear whether this cluster is responsible for the synthesis of nonribosomal peptides because the module is incomplete due to the lack of condensation (C) domains.

Cluster 16 was a *t1pks* encoding two PKSs (Figure 4xii). According to its domain organization, the product will be derived from the triketide skeleton shown in Figure 5i.

Clusters 17 and 18 (Figure 4viii,iv) were identified as BGCs for lankamycin and lankacidin (Figure 5j,k) [21], respectively, according to their high sequence similarities (Table S1) and identical domain organizations.

Cluster 19 was a *pks/nrps* found at the terminals of contigs sequence05 and sequence08. Four PKSs such as Scel_26590, Scel_26595, Scel_84485 and Scel_84490 are independently annotated as described in Table S1. However, the DNA sequence of Scel_26595 was completely included in that of Scel_84490. Similarly, the sequence of Scel_84485 was comprised in that of Scel_26590. Hence, these ORFs are not independent; rather, the four ORFs should be considered as two ORFs by removing Scel_26595 and Scel_84485 (black in Figure 4xv). Thus, the domain organization from Scel_26590 to Scel_84490 is C/A_{Leu}/T/E-C/A_{Leu}/T and C/A_{Thr}/T-TE, like that of *Streptomyces barringtoniae* JA03^T (blue in Figure 4xv). The ORFs in this cluster showed high sequence similarities to those of the incarnatapeptin-BGC in *Streptomyces incarnatus* NRRL 8089^T [22]. However, these domain organizations were not completely identical (Cluster 19, A, T, C/A/T/E-C/A_{Ala}/T, C/A_{Leu}/T/E-C/A_{Leu}/T, C/A_{Thr}/T-TE, KS/AT_{mm}/ACP, KS/AT_m/DH/KR/ACP, KS/AT_{mm}/DH/ER/KR/ACP, KS/AT/ACP and C/A/T; incarnatapeptin-BGC, A, T, C/A/T/E-C/A_{Ala}/T, C/A_{Leu}/T/E-C/A_{Leu}/T, C/A_{Thr}/T-TE, C/A_{Ala}/T-TE, KS/AT/ACP, KS/AT_{em}/DH/KR/ACP, KS/AT_m/DH/ER/KR/ACP, KS/AT_{mm}/ACP and C/A/T) as underlined. The differences may suggest that the product of Cluster 19 is a congener of incarnatapeptin.

Cluster 20 was a *t1pks* present at the terminals of contigs sequence09 and sequence10 in the draft genome sequence. The gap between Scel_86815 and Scel_87375 could not be sequenced using PacBio (Figure 4xvi). PKSs in this cluster showed high sequence similarities to those of the *stv* gene cluster [23] (Table S1). Both Scel_86815 and Scel_87375 showed high sequence similarities to StvB. The domain organization of StvB is KS/AT_{mm}/DH/KR/ACP-KS/AT_{mm}/DH/KR/ACP, whereas those of Scel_86815 and Scel_87375 are KS/AT_{mm}/... and .../KR/ACP-KS/AT_{mm}/DH/KR/ACP (three dots are a gap), respectively. Accordingly, it can be considered that Scel_86815 and Scel_87375 are a single PKS and the gap between them encodes a dehydratase (DH) domain. Thus, Cluster 20 was considered as a streptovaricin-BGC [23].

Cluster 21 was an orphan *pks/nrps* encoding NRPSs, type-II PKSs and a type-I PKS (Figure 4xvii). The product is unclear.

3.4. Identification of Blasticidin-Biosynthetic Gene Clusters

Although the sequence of Cluster 9 was not completely determined in the genome analysis, its ortholog was found in the genome of *Streptomyces griseochromogenes* ATCC 14511^T using a BLAST search (Figure 4vii). The orthologous cluster well resembled Cluster 9 and its complete sequence is available. This cluster encoded 10 PKSs and two NRPSs whose domain organizations were as follows: AVL59_10090, KS/AT_{mm}/KR/ACP-KS/AT_{mm}/KR/ACP-KS/AT_m/KR/ACP; AVL59_10095, KS/AT_{mm}/KR/ACP-KS/AT_{mm}/KR/ACP; AVL59_10100, KS/AT_m/KR/ACP-KS/AT_m/KR/ACP-KS/AT_m/KR/ACP; AVL59_RS54015, KS/AT_m/KR/ACP-KS/AT_{mm}/DH/KR/ACP-KS/AT_{mm}/KR/ACP; AVL59_10110, KS/AT_m/KR/ACP-KS/AT_{mm}/DH/KR/ACP-KS/AT_{mm}/KR; AVL59_10115, ACP-KS/AT_{mm}/DH/ER/KR/ACP; AVL59_RS54035, KS/AT_{mm}/DH/ER/KR/ACP-KS/AT_{mm}/DH/KR/ACP-KS/AT_{mm}/ACP; AVL59_10130, C/A/T; AVL59_10140, ACP; AVL59_RS55825, KS/AT_{mm}/ACP-KS/AT_m/DH; AVL59_10155, ER/KR/ACP-KS/AT_m/DH/ER/KR/ACP-KS/AT_m/DH/ER/KR/ACP; AVL59_10160, KS/AT_m/DH/ER/KR/ACP-KS/AT_m/DH/ER/KR/ACP; AVL59_RS10165, C/A_{Gly}/T. The underlined parts were gaps in the draft genome sequence of *S. cellostaticus* NBRC 12849^T (three dots in Table S1, gray dashed lines in Figure 4vii). The number of PKS modules was 24, suggesting that the product includes a C_{48}-derived polyketide backbone. Blasticidin A has been reported as a product of *S. griseochromogenes* IFO 13413^T, which is a compound derived from a C_{48}-polyketide chain and a tetramic acid moiety [19]. We examined whether the domain organization in the cluster fits the chemical structure of blasticidin A. As shown in Figure 6, the domain organization accounted well for the chemical structure. Therefore, we annotated this gene cluster to be responsible for the synthesis of blasticidin A, and here, we propose the putative biosynthetic pathway.

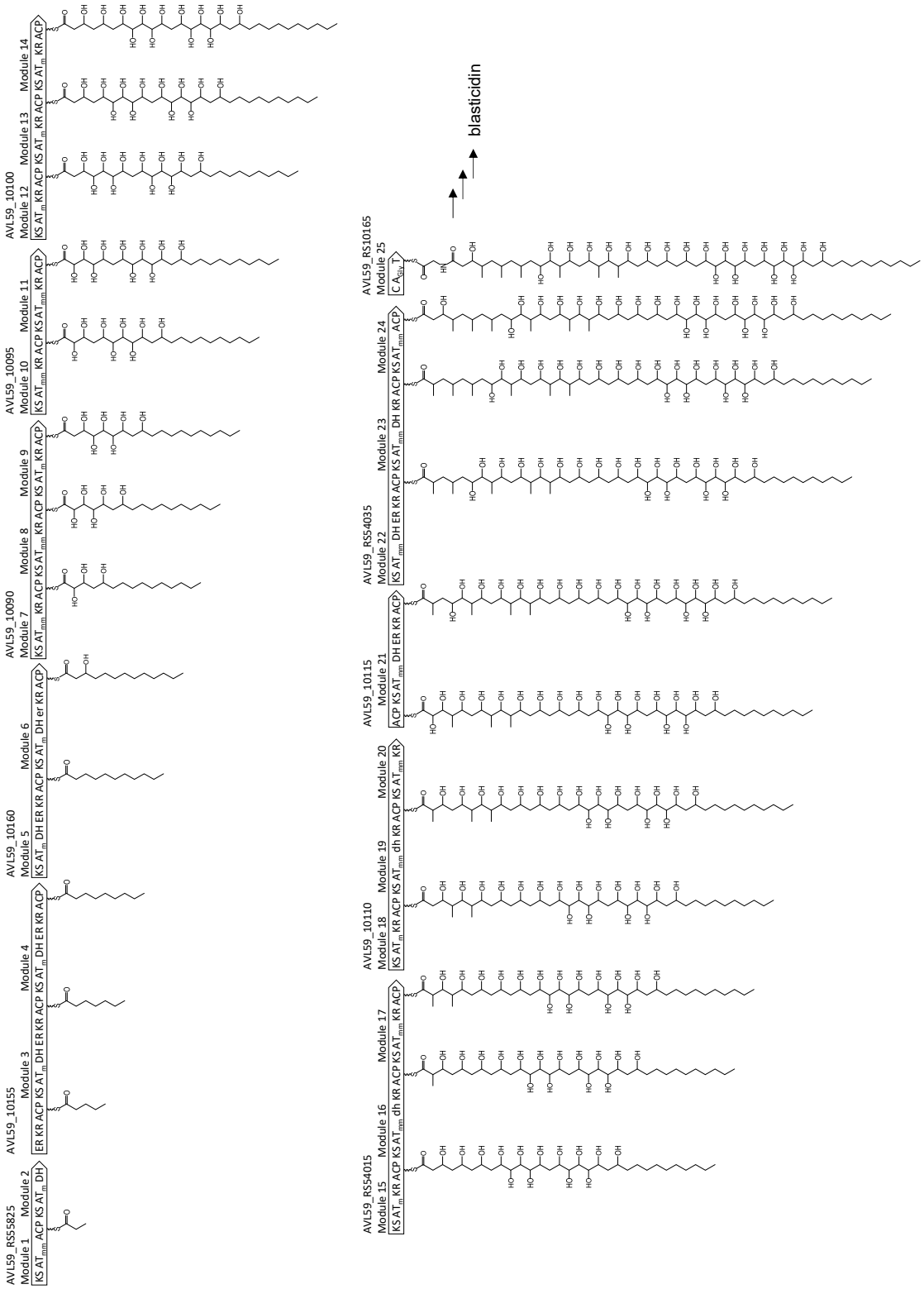

Figure 6. Biosynthetic pathway of blasticidin. The chemical structure of blasticidin A is shown in Figure 5f. er, inactive ER domain; dh, inactive DH domain; the other abbreviations are the same as in Table S1. Accession numbers of AVL59_10090, AVL59_10095, AVL59_10100, AVL59_RS54015, AVL59_10110, AVL59_10115, AVL59_RS54035, AVL59_10130, AVL59_10140, AVL59_RS55825, AVL59_10155, AVL59_10160 and AVL59_RS10165 are ANP49912, ANP49913, ANP49914, WP_269466063, ANP49915, ANP49916, WP_167549295, ANP49919, ANP49921, WP_261340673, ANP49924, ANP49925 and ANP49926, respectively.

According to the actual chemical structure of blasticidin, the ER domain in Module 6 and the DH domains in Module 16 and Module 19 were annotated as inactive. Although the substrate of the AT domain in Module 24 was predicted to be methylmalonyl-CoA by antiSMASH, it was malonyl-CoA in practice. Similarly, the substrates of the AT domains in Module 7, Module 8, Module 10, Module 11 and Module 20 were predicted to be methylmalonyl-CoA, but they were likely hydroxymalonyl-loading extender units such as hydroxymalonyl-ACP [24]. This is supported by the phylogenetic relationship of A domain amino acid sequences. These specific A domains were phylogenetically discriminated from common A domains for malonyl-CoA and methylmalonyl-CoA (Figure 7).

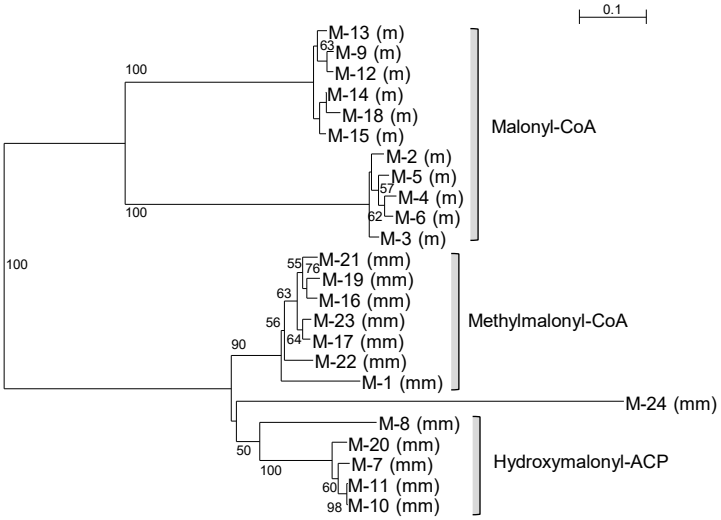

Figure 7. Phylogenetic tree based on amino acid sequences of adenylation (A) domains in the PKSs of blasticidin-BGC. Substrates predicted by antiSMASH are shown in parentheses. m, malonyl-CoA; mm, methylmalonyl-CoA. A-domain sequences of Modules 1 to 24 are shown as M-1 to M-24.

Although it was reported that *S. griseochromogenes* IFO 13413^{T} (= ATCC 14511^{T}) produces blasticidins [19,25], production by *S. cellostaticus* NBRC 12849^{T} has not been reported. We cultured this strain and analyzed the products in the culture extract using LC-MS. Positive and negative ion peaks at the retention time of 4.43 min. showed m/z values of 1168.72 and 1184.71, respectively. The time and values matched those in the LC-MS analysis of an authentic blasticidin A standard sample. Since the mass of blasticidin A is 1185.72 (M), the m/z values of 1168.72 and 1184.71 observed here were annotated as $[M+H-H_2O]^+$ and $[M-H]^-$, respectively. As any other gene clusters that can synthesize blasticidin A were not observed in the genome of *S. cellostaticus* NBRC 12849^{T}, Cluster 9 was annotated to be a blasticidin-BGC.

BLAST searches revealed that blasticidin-BGCs were specifically present in *S. griseochromogenes* and *S. cellostaticus* among whole genome sequence-published strains. Similar gene clusters were observed in the genomes of *Streptomyces chrestomyceticus* NBRC 13444^{T} (BHZC01000001.1) and *Streptomyces paromomycinus* NBRC 15454^{T} (BHZD01000001.1), but their domain organizations were not completely identical to those of blasticidin-BGCs (data not shown). Type strains of *S. chrestomyceticus* and *S. paromomycinus* may produce blasticidin congener(s).

4. Discussion

S. cellostaticus NBRC 12849^{T} harbored 21 PKS and NRPS gene clusters. The number was the highest among our whole genome-sequenced *Streptomyces* strains (Figure 3). The putative products of *S. cellostaticus* NBRC 12849^{T} are structurally diverse (Figure 5). Among the 21 clusters, Clusters 3, 5, 7, 9, 13, 17, 18 and 20 were predicted to synthesize known compounds such as foxicin, flaviolin, spore pigment, blasticidin A, fuelimycins, lankamycin, lankacidin and streptovaricin. In contrast, products of Clusters 8 and 19

were predicted to be not known compounds but new congeners of annimycin and incarnatapeptin, respectively (Table S1). Although flaviolin and spore pigments are common secondary metabolites in the genus *Streptomyces*, foxicin, blasticidin A, fuelimycins, lankamycin and lankacidin, streptovaricin, annimycin and incarnatapeptin have been reported as products of not *S. cellostaticus* but *Streptomyces diastatochromogenes* Tü 6028, *S. griseochromogenes* IFO 13413^T, *Streptomyces albireticuli* NRRL B-1670^T, *S. rochei* 7434AN4, *Streptomyces spectabilis* CCTCC M2017417, *Streptomyces calvus* ATCC 13382^T and *Streptomyces* sp. RJA2928, respectively [15,18–23]. The other eleven gene clusters were orphan, whose products are unknown. Our analysis suggested that the products of Clusters 1, 2, 6, 11, 14 and 21 are derived from a diketide (Figure 5a), a tetraketide with a cysteine residue (Figure 5b), a diketide composed of one malonyl-CoA and one methoxymalonyl-CoA, a tridecaketide (Figure 5g), a skeleton including a starter molecule, a valine residue and malonyl-CoA(s), and a triketide (Figure 5i), respectively. These di- to tetra-ketides differed from each other in their chemical structures. Taken together, these results suggest that *S. cellostaticus* NBRC 12849^T has a higher potential for producing diverse secondary metabolites such as polyketide- and hybrid polyketide/nonribosomal peptide-compounds, which include unknown and putative novel compounds as revealed in this study.

As shown in Table S1, PKS and NRPS gene clusters orthologous to those of *S. cellostaticus* NBRC 12849^T were often observed in *S. griseochromogenes* and *S. yokosukanensis*, which are phylogenetic neighbors of *S. cellostaticus*. However, the origins of the closest orthologs were not only these two species but also the other species or genera. This suggests that many PKS and NRPS gene clusters were horizontally transferred from/to *S. cellostaticus* NBRC 12849^T. BGCs for lankamycin and lankacidin exist not on chromosomal DNA but on a plasmid in *S. rochei* 7434AN4 [22] and *S. phaeogriseichromatogenes* DSM 40710^T, which has been reclassified as *S. murinus* [26]. Similarly, incarnatapeptin-BGC is also encoded on a plasmid in *S. incarnatus* JA03^T [27]. Therefore, it is speculated that these three clusters were transferred by HGT between phylogenetically distant species via plasmid(s). As the BGCs of these compounds are present on the contig sequence05 in *S. cellostaticus* NBRC 12849^T, this contig may not be chromosomal but a plasmid.

Nowadays, reports of whole genome sequences of many *Streptomyces* strains are being published. Although bioinformatic tools for analyzing secondary metabolite-BGCs (smBGCs), such as antiSMASH, are available, many of the smBGCs in the genomes are still orphan and their products are not well-annotated because this needs manual, careful and extensive analysis. In this study, we identified an orphan large *pks/nrps* in *S. cellostaticus* NBRC 12849^T and *S. griseochromogenes* ATCC 14511^T as the blasticidin-BGC. Blasticidins are compounds discovered more than 20 years ago, but the BGC has not been elucidated. This is also the first report on the biosynthetic pathway of blasticidins. It is essential to predict and then identify the products of orphan BGCs for the development of genome mining.

Pessotti et al. reported that their isolates belonging to the *S. cellostaticus* clade produced bafilomycin [7]. However, no bafilomycin-BGC was observed in the genome of *S. cellostaticus* NBRC 12849^T. In the report by Pessotti et al., it is unclear whether the strains are *S. cellostaticus* because the authors did not identify their isolates at the species level in comparison to the type strains. The isolates may not have been *S. cellostaticus*, although they were included in the *S. cellostaticus* clade. To deepen our knowledge on the relationship between taxonomical species and the produced secondary metabolites, correct classification of the strains at species level is essential.

During this study, whole genome sequences of *S. cellostaticus* DSM 40189^T were published by Ruckert et al. in GenBank under the accession numbers LMWL01000001-LMWL01000201. NBRC 12849^T and DSM 40189^T are the same strain derived from ISP 5089^T. The draft genome sequence determined in this study is composed of 13 contig sequences, whereas that of DSM 40189^T includes 200 contig sequences. Consequently, we were able to determine the number of PKS and NRPS gene clusters, since these genes were less fragmentated in the draft genome sequence. The draft genome sequence as well as

the data on the PKS and NRPS gene clusters in this study provide useful information for further studies.

5. Conclusions

The genome size and G + C content of *S. cellostaticus* NBRC 12849^T were 9.9 Mbp and 70.9%, respectively. This strain harbors 10 type-I PKS gene clusters, one type-II PKS gene cluster, one type-III PKS gene cluster, one NRPS gene cluster and eight hybrid PKS/NRPS gene clusters. The products were predicted using a bioinformatic analysis as follows: those of the 10 type-I PKS gene clusters are an annimycin congener (Figure 5e), fuelimycins (Figure 5h), lankamycin (Figure 5j), streptovaricin (Figure 5m), four compounds whose backbones are a diketide (Figure 5a), a diketide, a triketide (Figure 5i) and a tridecaketide (Figure 5g), and two unknown compounds whose structures were not predicted, respectively. That of the type-II PKS gene cluster is a spore pigment; that of the type-III PKS gene cluster is flaviolin (Figure 5d); that of the NRPS gene cluster is unknown; those of the eight hybrid PKS/NRPS gene clusters are foxicin (Figure 5c), blasticidin A (Figure 5f), lankacidin (Figure 5k), a congener of incarnatapeptin (Figure 5l), a compound whose backbone is tetraketide with a cysteine residue (Figure 5b) and three unknown compounds whose hybrid polyketide/nonribosomal peptide skeletons were not predicted, respectively. The results suggest the diversity of possible polyketide- and nonribosomal peptide-compounds of *S. cellostaticus*. This is the first report on the profile and putative products of PKS and NRPS gene clusters in *S. cellostaticus* and identification of the blasticidin-biosynthetic gene cluster. These results provide useful information for genome mining as well as for deepening our knowledge on the relationship between taxonomical species and possible secondary metabolites.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/fermentation9110924/s1, Table S1: PKSs and NRPSs in the gene clusters of *S. cellostaticus* NBRC 12849^T.

Author Contributions: Conceptualization, H.K.; methodology, T.T.; investigation, H.K.; resources, T.T.; data curation, H.K.; writing—original draft preparation, H.K.; writing—review and editing, T.T.; supervision, H.K.; project administration, T.T.; funding acquisition, T.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by a commissioned project from the Japan Patent Office.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The whole genome shotgun project of *S. cellostaticus* NBRC 12849^T has been deposited at GenBank under the accession numbers BNDU01000001-BNDU01000013. The BioProject and BioSample accession numbers are PRJDB9780 and SAMD00227995, respectively.

Acknowledgments: We are grateful to Shohei Sakuda for providing us with purified blasticidin A. We thank Shinpei Ino and Takahiro Matsuyama for genome DNA preparation, and to Aya Uohara for depositing the whole genome sequence to DDBJ.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Berdy, J. Bioactive microbial metabolites. *J. Antibiot.* **2005**, *58*, 1–26. [CrossRef] [PubMed]
2. Nett, M.; Ikeda, H.; Moore, B.S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **2009**, *26*, 1362–1384. [CrossRef] [PubMed]
3. Komaki, H. Recent progress of reclassification of the genus *Streptomyces*. *Microorganisms* **2023**, *11*, 831. [CrossRef] [PubMed]
4. Lee, N.; Hwang, S.; Kim, J.; Cho, S.; Palsson, B.; Cho, B. Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1548–1556. [CrossRef]
5. Fischbach, M.A.; Walsh, C.T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106*, 3468–3496. [CrossRef]

6. Hamada, S. A study of a new antitumor substance, cellostatin. I. On the isolation and some properties of cellostatin. *Tohoku J. Exp. Med.* **1958**, *67*, 173–179. [CrossRef]
7. Pessotti, R.C.; Hansen, B.L.; Reaso, J.N.; Ceja-Navarro, J.A.; El-Hifnawi, L.; Brodie, E.L.; Traxler, M.F. Multiple lineages of *Streptomyces* produce antimicrobials within passalid beetle galleries across eastern North America. *eLife* **2021**, *10*, e65091. [CrossRef]
8. Yoon, S.; Ha, S.; Kwon, S.; Lim, J.; Kim, Y.; Seo, H.; Chun, J. Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 1613–1617. [CrossRef]
9. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef]
10. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.
11. Saito, H.; Miura, K. Preparation of transforming deoxyribonucleic acid by phenol treatment. *Biochim. Biophys. Acta* **1963**, *72*, 619–629. [CrossRef]
12. Komaki, H.; Igarashi, Y.; Tamura, T. Taxonogenomic analysis of marine-derived *Streptomyces* sp. N11-50 and the profile of NRPS and PKS gene clusters. *Hydrobiology* **2023**, *2*, 382–394. [CrossRef]
13. Tanizawa, Y.; Fujisawa, T.; Nakamura, Y. DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* **2018**, *34*, 1037–1039. [CrossRef] [PubMed]
14. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. [CrossRef] [PubMed]
15. Greule, A.; Marolt, M.; Deubel, D.; Peintner, I.; Zhang, S.; Jessen-Trefzer, C.; De Ford, C.; Burschel, S.; Li, S.M.; Friedrich, T.; et al. Wide distribution of foxicin biosynthetic gene clusters in *Streptomyces* strains—An unusual secondary metabolite with various properties. *Front. Microbiol.* **2017**, *8*, 221. [CrossRef]
16. Bentley, S.D.; Chater, K.F.; Cerdeno-Tarraga, A.M.; Challis, G.L.; Thomson, N.R.; James, K.D.; Harris, D.E.; Quail, M.A.; Kieser, H.; Harper, D.; et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **2002**, *417*, 141–147. [CrossRef]
17. Komaki, H.; Tamura, T.; Igarashi, Y. Classification and secondary metabolite-biosynthetic gene clusters of marine *Streptomyces* strains including a lobophorin- and divergolide-producer. *Hydrobiology* **2023**, *2*, 151–161. [CrossRef]
18. Kalan, L.; Gessner, A.; Thaker, M.N.; Waglechner, N.; Zhu, X.; Szawiola, A.; Bechthold, A.; Wright, G.D.; Zechel, D.L. A cryptic polyene biosynthetic gene cluster in *Streptomyces calvus* is expressed upon complementation with a functional *bldA* gene. *Chem. Biol.* **2013**, *20*, 1214–1224. [CrossRef]
19. Sakuda, S.; Ono, M.; Ikeda, H.; Nakamura, T.; Inagaki, Y.; Kawachi, R.; Nakayama, J.; Suzuki, A.; Isogai, A.; Nagasawa, H. Blasticidin A as an inhibitor of aflatoxin production by *Aspergillus parasiticus*. *J. Antibiot.* **2000**, *53*, 1265–1271. [CrossRef]
20. Cruz-Morales, P.; Yin, K.; Landera, A.; Cort, J.R.; Young, R.P.; Kyle, J.E.; Bertrand, R.; Iavarone, A.T.; Acharya, S.; Cowan, A.; et al. Biosynthesis of polycyclopropanated high energy biofuels. *Joule* **2022**, *6*, 1590–1605. [CrossRef]
21. Arakawa, K. Genetic and biochemical analysis of the antibiotic biosynthetic gene clusters on the *Streptomyces* linear plasmid. *Biosci. Biotechnol. Biochem.* **2014**, *78*, 183–189. [CrossRef] [PubMed]
22. Morgan, K.D.; Williams, D.E.; Patrick, B.O.; Remigy, M.; Banuelos, C.A.; Sadar, M.D.; Ryan, K.S.; Andersen, R.J. Incarnatapeptins A and B, nonribosomal peptides discovered using genome mining and $^{1}H/^{15}N$ HSQC-TOCSY. *Org. Lett.* **2020**, *22*, 4053–4057. [CrossRef] [PubMed]
23. Liu, Y.; Chen, X.; Li, Z.; Xu, W.; Tao, W.; Wu, J.; Yang, J.; Deng, Z.; Sun, Y. Functional analysis of cytochrome P450s involved in streptovaricin biosynthesis and generation of anti-MRSA analogues. *ACS Chem. Biol.* **2017**, *12*, 2589–2597. [CrossRef] [PubMed]
24. Chan, Y.A.; Boyne, M.T., 2nd; Podevels, A.M.; Klimowicz, A.K.; Handelsman, J.; Kelleher, N.L.; Thomas, M.G. Hydroxymalonyl-acyl carrier protein (ACP) and aminomalonyl-ACP are two additional type I polyketide synthase extender units. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 14349–14354. [CrossRef] [PubMed]
25. Sakuda, S.; Ikeda, H.; Nakamura, T.; Kawachi, R.; Kondo, T.; Ono, M.; Sakurada, M.; Inagaki, H.; Ito, R.; Nagasawa, H. Blasticidin A derivatives with highly specific inhibitory activity toward aflatoxin production in *Aspergillus parasiticus*. *J. Antibiot.* **2000**, *53*, 1378–1384. [CrossRef]
26. Komaki, H. Reclassification of *Streptomyces costaricanus* and *Streptomyces phaeogriseichromatogenes* as later heterotypic synonyms of *Streptomyces murinus*. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 004638. [CrossRef]
27. Oshima, K.; Hattori, M.; Shimizu, H.; Fukuda, K.; Nemoto, M.; Inagaki, K.; Tamura, T. Draft genome sequence of *Streptomyces incarnatus* NRRL8089, which produces the nucleoside antibiotic sinefungin. *Genome Announc.* **2015**, *3*, e00715-15. [CrossRef]