The Journal of Antibiotics (2018) 71:495–505
https://doi.org/10.1038/s41429-017-0023-x

JARA Japan Antibiotics Research Association
The Society for Actinomycetes Japan

ARTICLE

The chejuenolide biosynthetic gene cluster harboring an iterative *trans*-AT PKS system in *Hahella chejuensis* strain MB-1084

Bee Gek Ng[1] · Jae Woo Han[2] · Dong Wan Lee[3] · Gyung Ja Choi[2] · Beom Seok Kim[1,3]

Received: 1 August 2017 / Revised: 21 November 2017 / Accepted: 25 December 2017 / Published online: 26 February 2018

Abstract

Hahella chejuensis MB-1084 is a Gram-negative marine bacterial strain that produces unusual 17-membered carbocyclic tetraenes, chejuenolide A and B. Two fosmid clones responsible for chejuenolide production were identified from the genomic DNA library of the MB-1084 strain. Systematic inactivation of the open reading frames (ORFs) in the sequenced region defines the boundaries of the chejuenolide (*che*) biosynthetic gene cluster (24.9 kbp) that encodes one non-ribosomal peptide synthase (NRPS)-polyketide synthase (PKS) hybrid protein, three modular PKSs, two PKS domains, and an amine oxidase homolog. Based on the results, we found that the *che* PKSs have non-canonical features such as *trans*-AT system and insufficient number of KS domains (five KS domains) for chejuenolide production (requires eight rounds of Claisen condensation reaction). Heterologous expression of the *che* PKSs in the *E. coli* BAP1 strain provides strong evidence of the iterative characteristic of the modular PKSs. Additionally, the phylogenetic relatedness of the KS domains of *che* PKSs and other *trans*-AT PKSs was analyzed to propose a possible pathway for chejuenolide biosynthesis.

Introduction

Polyketides are a class of microbial secondary metabolites that possess immense structural diversity and broad biological applicability as antimicrobial, antiparasitic, and therapeutic agents [1]. Despite the structural complexity of polyketides, their biosynthesis follows a simple and common assembly mechanism using multiple Claisen condensations of short-chain carboxylic acids, such as acetate and propionate. The core structure of polyketides is synthesized by polyketide synthases (PKSs). Diverse PKS systems that have discovered to date are mainly classified into three types: non-iterative type I PKSs, iterative type II PKSs, and acyl carrier protein (ACP)-independent type III PKSs [2].

Type I polyketides, mostly macrolides, are synthesized by the large modular enzymes known as type I PKSs. Each module non-iteratively catalyzes one round of chain extension to generate the β-keto acyl intermediate, which consists of acyltransferase (AT) domain that specifically selects the appropriate building block for chain initiation/extension and acylates the ACP domain, ACP domain that covalently tethers the growing polyketide chain via the phosphopantetheine arm, and ketosynthase (KS) domain that catalyzes the two-carbon extension of the polyketide chain by Claisen condensation. The nascent β-keto intermediate is exposed to a combinatorial reduction catalyzed by the ketoreductase (KR), dehydratase (DH), and enoyl reductase domains in the module. Finally, the fully processed polyketide chain is liberated from the ACP domain by the thioesterase (TE) domain to form a free carboxylic acid or a lactone ring. The length of a polyketide chain is dependent on the number of modules participating in the polyketide synthesis, and the fate of the β-keto structure is determined by the combination of the reductive domains. The strong collinearity between the genetic architecture of the domains/modules and the chemical structure of the

Bee Gek Ng and Jae Woo Han contributed equally to this work

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41429-017-0023-x) contains supplementary material, which is available to authorized users

✉ Beom Seok Kim
biskim@korea.ac.kr

[1] Department of Biotechnology, Korea University Graduate School, Seoul, Korea

[2] Korea Research Institute of Chemical Technology, Center for Eco-Friendly New Materials, Daejeon, Seoul, Korea

[3] Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul, Korea

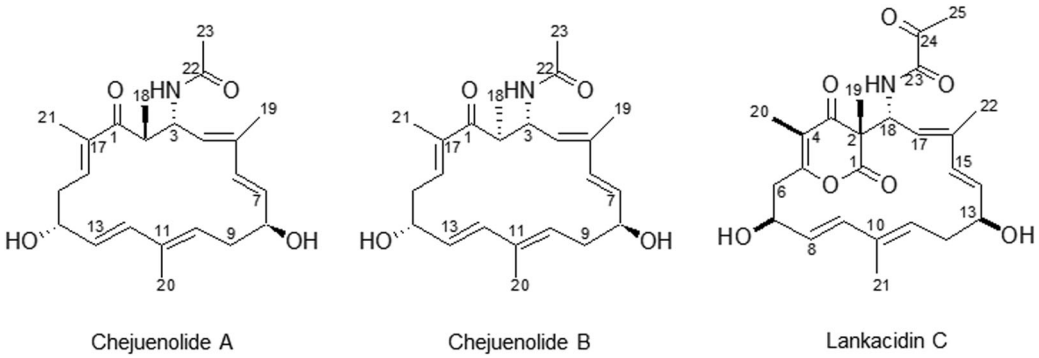

Fig. 1 The structures of chejuenolides and lankacidin C

eventual product portends the possibility of generating a designed polyketide by modifying the genetic organization of PKSs [3, 4].

In recent years, several PKSs deviating from the collinearity rule were discovered in bacteria isolated from diverse environments [5–7]. Some non-canonical type I PKSs have domains or modules that are skipped and apparently not used for the synthesis of the corresponding product [8]. Some domains (AT and DH) are not in modular arrangement and also iteratively used during the polyketide synthesis. The AT-less PKSs was first reported in pederin biosynthetic gene cluster where the function of missing AT domains are possibly provided by the two products of discrete genes *pedC* and *pedD* [9]. The *trans*-AT PKS system was then confirmed in the leinamycin (*lnm*) biosynthesis, where LmnG interacts with and loads all ACP domains of the *lnm* PKSs in vitro [10]. The standalone DH (*lkcB*) domain discovered in the lankacidin (*lkc*) PKSs is reported to provide DH activity in *trans* on multiple modules of the *lkc* PKSs [11]. Of the non-canonical examples, the most interesting phenomenon is the discrepancy between the number of modules in a PKS gene cluster and the number of chain-extension reactions required for the completion of the corresponding polyketide biosynthesis. Such discrepancies are found in the biosynthetic gene clusters of stigmatellin [12], aureothin [13], borrelidin [14], lankacidin [15], neoaureothin [16], etnangien [17], crocacin [18], thiolactomycin [19], and azalomycin [20]. Unless any PKS-related enzymes outside the gene cluster are involved, one or more modules of the PKSs should be utilized iteratively, which is beyond the definition of modular and non-iterative type I PKSs [5].

Chejuenolide A and B, identified from the metabolites of marine bacterial strain MB-1084 (*Hahella chejuensis*) are reported to exhibit inhibitory effect on the activity of protein tyrosine phosphatase 1B [21], which is considered to be a potential target for the therapy of type 2 diabetes and obesity [22]. Chejuenolides are the first example of the 17-membered carbocyclic tetraenes discovered in Gram-negative bacteria, which possess macrocyclic ring structure that is different from most macrolide polyketides having lactone ring. Another well-known 17-membered carbocyclic antibiotic, lankacidin has similar structure to chejuenolides except a δ-lactone ring and a pyruvyl group connected to C-18 via a nitrogen atom (Fig. 1). Given the unusual chemical structure and biological activity of chejuenolides, it is of interest to elucidate the molecular basis of their biosynthesis.

Here, we report the complete sequence and functional assignment of the chejuenolide biosynthetic gene cluster of the *Hahella chejuensis* strain MB-1084 and provide evidence for iterative work by the chejuenolide PKSs (*che* PKSs).

Materials and methods

Bacterial strain, plasmids and growth conditions

The *H. chejuensis* strain MB-1084 was used to construct various *che* gene disruption mutants. All the mutants constructed and the plasmids used in this study are listed in Supplementary Table S1. *E. coli* ET12567/pUZ8002 and DH5α were used for routine sub-cloning and plasmid preparation. *H. chejuensis* strains were grown in Zobell's medium (peptone 5 g, yeast extract 1 g, $FeSO_4.7H_2O$ 0.01 g, and NaCl 30 g, pH 7.2 in 1 l of distilled water) at 28 °C for a week. All *E. coli* strains were grown in Luria–Bertani (LB) medium (NaCl 5 g, tryptone 10 g, and yeast extract 5 g in 1 l of distilled water) supplemented with appropriate antibiotics when necessary at 37 °C. DNA manipulations of *H. chejuensis* and *E. coli* were performed according to standard procedures [23, 24].

Identification of a KS domain involved in the chejuenolide biosynthesis

DNA fragments homologous to the KS domains of type I PKSs were amplified using two degenerative KS primers,

KSLF (5′-GTS CCS GTS CCG TGS GYS TCS A-3′) and KSLR (5′-CCS CAG SAG CGC STS YTS CTS GA-3′), from the genomic DNA of the MB-1084 strain. Two amplified fragments were respectively cloned into pVIK112 at *EcoR*I restriction site to be used for site-directed disruption in the genome. The plasmids (pVIKA2, harboring 629 bp fragment and pVIKB1, harboring 665 bp) were introduced into *E. coli S17-1 λpir* and conjugally transferred to *H. chejuensis* strain as follows. An *E. coli* transformant and *H. chejuensis* MB-1084 strain were grown overnight in LB liquid media supplemented with 50 μg μl^{-1} kanamycin at 37 °C and Zobell's liquid media at 28 °C, respectively. After removal of kanamycin in the *E.coli* culture by washing with LB liquid media twice, the same amounts of donor and recipient cells were mixed and spread on LB agar plate and incubated overnight at 37 °C. The cultured cells were recovered and spread on Zobell's agar media supplemented with kanamycin and ampicillin (50 and 100 μg μl^{-1}, respectively) and incubated overnight at 28 °C. A single crossover mutant was obtained from the kanamycin selection and verified by polymerase chain reaction (PCR) analysis.

Genomic library construction and screening

A genomic DNA library of *H. chejuensis* was constructed using CopyControl pCC2FOS from Epicentre Biotechnologies (Madison, WI, USA) by random shearing of genomic DNA into approximately 40 kb according to manufacturer's instruction. The library was screened using two primer pairs designed from the KS sequence involved in the chejuenolides biosynthesis identified from site-directed disruption: Che_KSF, 5′-TGCGCCGTCCCATTTATCC-3′; Che_KSR, 5′-GGTTTTCCGCCACGCTTTCAA-3′ and Che_KSFP3, 5'-ATGGCGCCGAATCCCTGCTC-3′; Che_KSRP3, 5′-GTGCTGGTGGCGACGGACTG-3′. Two fosmid clones harboring the KS sequences were designated as pBG6E11 and pBG19A6.

DNA sequencing and analysis

Fosmids pBG6E11 and pBG19A6 were fully sequenced by Macrogen Co (Seoul, Korea). Contigs were assembled using DNASTAR-SeqMan and individual ORFs were identified and assigned with the assistance of DNASTAR-SeqBuilder and BLAST analyses.

Construction of gene disruption mutants

Gene disruption mutants were obtained by a PCR-targeted gene replacement system according to the standard protocol [25]. Briefly, an apramycin resistance gene *aac(3)IV/oriT* cassette was amplified with the primers having 5′-overhang and 3′-overhang of 39 nucleotides homologous to the target gene. (Table S2). The amplified apramycin cassette was introduced into *E. coli* EPI300-$T1^R$ containing either pBG6E11 or pBG19A6 where the target gene was replaced by λ-Red-mediated recombination. The resulting plasmid harboring the recombinant DNA was then transferred into *E.coli* ET12567/pUZ8002 for conjugal transfer to *H. chejuensis*. Double crossover mutants were selected on Zobell's agar supplemented with 50 mg ml^{-1} of apramycin and verified by PCR using verification primer sets as listed in Table S2. The difference of the amplified fragment sizes between the native gene and the disrupted gene allowed us to confirm that the target gene was correctly replaced by the apramycin cassette.

Construction of point mutation strains

Point mutations were performed using Enzynomics EZchange™ site-directed mutagenesis kit according to manufacturer's instructions. The gene fragments to be mutated were amplified by PCR using the primer sets as listed in Table S2. Mutation was introduced into the target region by the mutagenesis primers (Table S2). Mutated gene fragments were subcloned into a suicide vector, pKNG101. The resulting plasmid was then transferred into *E. coli* S17-1 λpir and subjected to conjugal transfer into *H. chejuensis*. Single crossover mutants were selected by streptomycin resistance, followed by the induction of double crossover event to replace the target region with the mutated sequence by the addition of 5% sucrose. Finally, double crossover mutants that are sensitive to streptomycin were selected.

Complementation study

A disruption mutant was complemented by the plasmid-based expression of the native gene using pPROBE-OT [26]. Arabinose inducible promoter and *araC* regulator gene were cloned into pPROBE-OT prior to the cloning of the native gene. The expression plasmid was then transferred to *E. coli* DH5α and introduced into the disruption mutant by a triparental mating method using a helper strain, pRK2013/DH5α. The expression of the native gene was induced by the addition of 0.2% L-arabinose prior to incubation at 28 °C for 7 days.

Heterologous expression of cheA-G in E. coli BAP1

PrimeSTAR HS DNA polymerase (Takara, Japan) was used to amplify *cheA-G* by primer sets as summarized in Supplementary Table S2 and separately cloned into pET21 and pET28a. *CheA-D* was cloned into pET21a and *cheE-G* was cloned into pET28a and expressed in *E. coli* BAP1 [27]. A

single colony of *E.* coli strain containing *cheA-G* was cultured into 10 ml LB media overnight at 37 °C. Then 1 ml of the seed culture was transferred into 200 ml LB media and cultured until OD_{600} 0.7 prior to the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to induce the expression of *cheA-G*. Incubation was continued for 48 h at 22 °C and extracted with 1:1 (v/v) ethyl acetate. The ethyl acetate extract was concentrated in vacuo and dissolved in MeOH. Mass spectrometry of the extract was recorded on a quadrupole time-of-flight tandem mass spectrometer (Waters, MA) using the electrospray ionization mass spectroscopy (ESI-MS) method. The MS–MS analysis was recorded on LTQ XL mass spectrometer (Thermo fisher scientific Inc.) employing ESI-MS method.

Analysis of chejuenolides production

The analysis of chejuenolides production was carried out as follow. A colony of *H. chejuensis* strain was inoculated in 10 ml of Zobell's medium and cultured overnight at 28 °C. The 1 ml seed culture was transferred to 100 ml Zobell's media and cultured at 28 °C. After 7 days, the supernatant was extracted with 1: 1 (v/v) ethyl acetate. The extract was then concentrated in vacuo and dissolved in 1 ml of methanol. High-performance liquid chromatography (HPLC) analysis was carried out on an Pursuit XRs C-18 column (5 µm, 250 × 4.6 mm, Varian, CA) using Varian HPLC system and was developed using a linear gradient solvent system from 20 to 80% acetonitrile in water containing 0.1% formic acid for 20 min, followed by an isocratic elution with 80% acetonitrile in water containing 0.1% formic acid for 10 min at a flow rate of 1 ml min^{-1} and UV detection at 254 nm.

Feeding study of ^{13}C-labeled sodium acetates

For the feeding studies of 13C-labeled sodium acetates, [1-^{13}C], [2-^{13}C] and [1-2-^{13}C]sodium acetate (20 mg ml^{-1}) was respectively mixed with the same volume of unlabeled sodium acetate (20 mg ml^{-1}). The acetate solution (2 ml) was added into *H. chejuensis* MB-1084 culture (200 ml of Zobell's broth) at 24 and 48 h post inoculation. After additional 5-day culture, the supernatant was extracted with 1: 1 (v/v) ethyl acetate. The ethyl acetate extract was concentrated in vacuo and subjected to semi-preparative Varian Prostar 210 HPLC System (Varian Inc. Palo Alto, CA, USA) for the purification of chejuenolide A. The ^{13}C NMR spectra (125 MHz) of chejuenolide A were recorded in CD_3OD by broad-band proton decoupling on a Varian 500 NMR spectrometer (Varian Inc., USA) at room temperature. The data were processed by MestReNova version 6.0.2–5475 software (Mestrelab research, Santiago de Compostela, Spain). The relative enrichment of ^{13}C-labeled acetates were calculated as follows. Enrichment in each carbon of chejuenolide was calculated as the resonance intensity of the enriched sample minus the resonance intensity of the natural-abundance sample divide by the resonance intensity of the natural-abundance sample.

Phylogenetic analysis of KS domains

Amino acid sequences of 72 KS domains of known *trans*-AT PKSs were retrieved from the GenBank database and aligned using a combination of manual and clustal omega. The construction of the phylogenetic tree was performed using MEGA 7 software employing a neighbor-joining algorithm, and KS domains from cis-AT systems was used as the outgroup. Bootstrap analysis was performed with 1000 replicate sequences.

Result and discussion

Identification of the chejuenolide biosynthetic gene cluster

KS homologous gene fragments (629 and 665 bp) were amplified from the genomic DNA of *H. chejuensis* MB-1084 by using a degenerative KS primer set (KSLF and KSLR). The 629 bp fragment shows high DNA sequence similarity to a PKS gene of *Hahella chejuensis* KCTC 2396 (accession number ABC30194.1, 95% identity), whereas the 665 bp PCR product exhibits significant DNA sequence similarity to a PKS gene of *Hahella chejuensis* KCTC 2396 (accession number ABC30237.1, 93% identity). The fragments were cloned into a disruption vector pVIK112 to give pVIKA2 (harboring 629 bp fragment) and pVIKB1 (harboring 665 bp fragment), and used for site-directed integration into the chromosome of the MB-1084 strain. The disruption mutant carrying pVIKB1, strain HC-B1 completely lost the ability to produce chejuenolides while the disruption mutation carrying pVIKA2 (strain HC-A2) did not significantly affect the titer of chejuenolides (Figure S1). The result shows the involvement of the 665 bp gene fragment in the biosynthesis of chejuenolides. Two primer sets (Che_KSF and Che_KSR; Che_KSFP3 and Che_KSRP3) were designed based on the sequence of the DNA fragment and used to screen the genomic DNA library of the MB-1084 strain. The PCR-based screening yielded two positive clones, designated as pBG6E11 and pBG19A6. Shortgun sequencing and primer walking revealed an overlapping region of 26 open reading frames (ORFs) spanning 51.4 kbp. Among them, seven ORFs are of PKS gene homologs. The other ORFs are of transporter, regulator, and tailoring gene homologs and of unknown function (Fig. 2). The functions of the ORFs were proposed

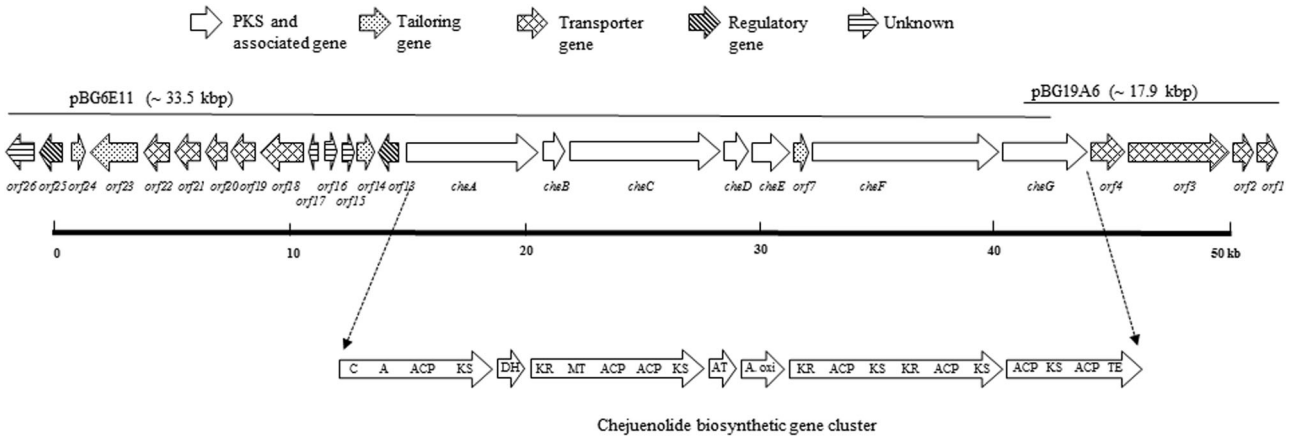

Fig. 2 DNA region of the overlapping fosmids pBG6E11 and pBG19A6 that encompasses the chejuenolides biosynthetic gene cluster. The chejuenolides biosynthetic genes were named in the leftward direction from *cheA* to *cheG*. Arrows indicate the transcription direction of genes. The deduced functions of each gene products are summarized in Table 1

by comparing the deduced amino acid sequences with proteins of known function in the NCBI database (Table 1). The GenBank accession number of the sequence is designated as HE664023.

The gene cluster required for chejuenolides biosynthesis was defined by sequential inactivation of the genes residing at the distal ends of the PKS genes. Deletion of *orf13* (strain HC13), *orf14* (strain HC14), *orf4* (strain HC04), and *orf3* (strain HC03) displayed insignificant effect on chejuenolides production. (Figure S2). These results defined the boundaries of the chejuenolides biosynthetic gene cluster that consists of *cheA* at the left boundary and *cheG* at the right boundary that span 24.9 kbp (Fig. 2).

Genes encoding the chejuenolide PKSs (che PKSs)

The *cheA* encodes a fusion protein of non-ribosomal peptide synthase (NRPS) and PKS domains. The NRPS region located at the N-terminal of CheA harbors condensation (C) and adenylation (A) domains. The C domain has a consensus sequence of HxxxDG (124–129 amino acid position) that is essential for catalytic function [28], and the A domain has a potential motif 235D, 236I, 239 L, 278Q, 299 L, 301 G, 322 M, 330I, 331 W, 517 K that has substrate specificity to glycine [29, 30]. The PKS domains following the A domain are ACP and KS, which contain the 4′-phosphopantetheine attachment site in a signature motif Gx(D/H)S(L/I) (1010–1013 amino acid position) [31] and catalytic Cys residue in TACSSS motif [32], respectively.

CheC is a protein of 1911 amino acids containing a KR, a methyltransferase (MT), two tandemly aligned ACPs, and a KS domain. The MT domain is homologous to S-adenosylmethionine-dependent MTs containing ExGxG motif (628–632 amino acid position) characteristic to C-MT [33]. Although the two tandemly aligned ACP domains share a relatively low sequence similarity (49% amino acid identity), they both harbor the Ser residue attachment site [31]. CheF encodes the KR, ACP, KS, KR, ACP, and KS domains, which consists of 2346 amino acid residues. The two KR domains showed 29% sequence similarity to each other, which is quite low; however, both domains contain the GGxGxxG motif, the NADPH binding site that is critical for the ketoreduction reaction [34]. CheG encodes a polypeptide of 1067 amino acids consisting of ACP, KS, ACP, and TE domains. The TE domain has the conserved active site of GxSxG (908–912 amino acid position) [35].

Interestingly, the *che* PKSs do not possess cognate AT domains in the modules. A discrete AT (CheD) shows significant sequence homology (40–54% identity) to the *trans*-ATs reported. CheD is predicted to take the role of AT as it contains the GHSxG motif (90–94 amino acid position) conserved in functional AT domains and also the substrate binding motif of GAFH (193–196 amino acid position) specific to malonyl-CoA [36, 37]. To confirm the involvement of CheD in chejuenolides biosynthesis, a point mutation mutant, HCD ($CheD^{G193Y.F195S}$) was constructed and examined for chejuenolide production. HPLC analysis of the culture extract of the HCD strain revealed that chejuenolide production was completely abolished. The complementation of *cheD* (strain HCDC) restored chejuenolide production (Fig. 3), indicating that CheD plays a role in chejuenolides biosynthesis, most probably in providing the missing AT activity. [1-^{13}C]sodium acetate feeding experiment showed enrichment of the stable isotope at C-1, C-6, C-8, C-10, C-12, C-14, and C-16 (Table S3) relative to the natural abundance of chejuenolide A, indicating that all of these carbons are supplied by malonyl-CoA. This result also shows

Table 1 Deduced function of ORFs in the chejuenolides biosynthetic gene clusters

Gene	DNA size (bp)	Protein homolog (accession number, identity/similarity)
orf3	3870	Transmembrane 2 protein (*Microscilla marina* WP_002692690.1, 52/66)
orf4	1296	Major facilitator superfamily permease (*Serratia symbiotica*, WP_006708725.1, 91/93)
cheG	3210	Type I PKS (KS,ACP, TE) (LkcG, *Streptomyces rochei* subsp. volubilis, ADN64227.1, 48/61)
cheF	7038	Type I PKS (KR, ACP, KS, KR, ACP, KS) (LkcF, *Streptomyces rochei*, NP_851435.1, 49/61)
orf7	555	Putative isochorismatase (LkcH, *Streptomyces rochei*, ADN64226.1, 61/74)
cheE	1419	Amine oxidase (LkcE *Streptomyces rochei*, NP_851436.1, 63/80)
cheD	882	Type I PKS (AT) LkcD *Streptomyces rochei*, NP_851437.1, 54/68)
cheC	5736	Type I PKS (KR, MT, ACP, ACP, KS) (LkcC, *Streptomyces rochei*, NP_851438.1, 54/67)
cheB	858	Dehydratase (LkcB, *Streptomyces rochei* subsp. Volubilis, ADN64232.1, 50/64)
cheA	5043	Type I PKS (A, ACP, KS) (LkcA, *Streptomyces rochei* subsp. Volubilis, ADN64233.1, 49/62)
orf13	765	Putative transcriptional regulator (*Vibrio parahaemolyticus*, WP_069544430.1, 74/89)
orf14	630	Putative haloacid dehalogenase (*Halomonas* sp. KHS3, WP_041158864.1, 67/83)

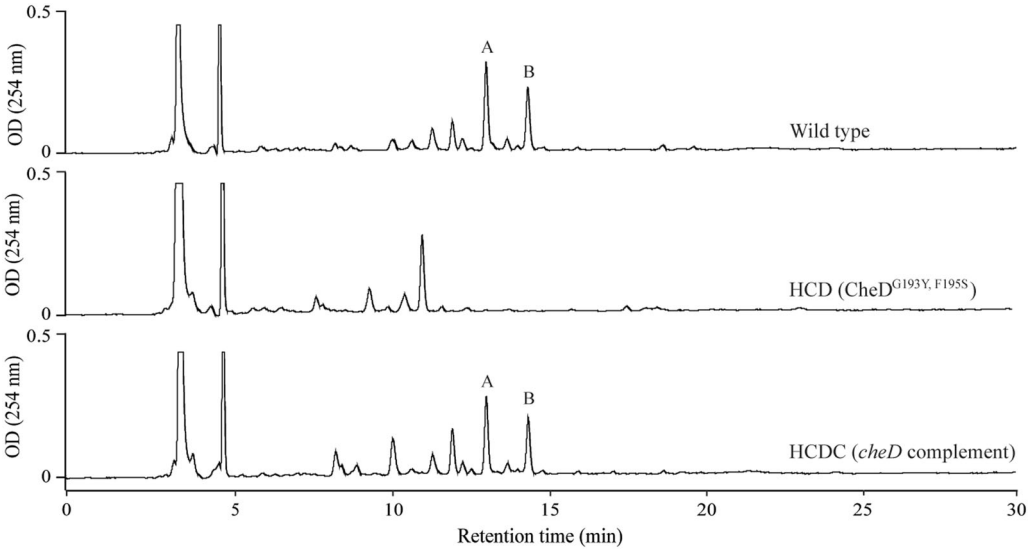

Fig. 3 HPLC analyses of the culture extracts of *Hahella chejuensis* HCD strain (CheD$^{G193Y, F195S}$) and the complemented strain. The plasmid pKNG-cheD carrying the mutated fragment of *cheD* with 1 kb homologous at both ends was used for site-directed integration into the chromosome of the MB-1084 strain. Complementation of *cheD* was performed by introducing an expression vector of *cheD* to the HCD strain. Ethyl acetate extract was prepared from 7-days old cultures. A and B indicate chejuenolide A (retention time 12.6 min) and B (retention time 14.8 min), respectively

that the four methyl groups at C-18, C-19, C-20, and C-21 are not of methylmalonyl-CoA and possibly incorporated by an S-adenosylmethionine-dependent MT domain found in CheC.

Two genes, o*rf7* (isochorismatase) and *cheE* (amine oxidase) encoding non-PKS proteins located within the contiguous gene cluster (between *cheE* and *cheF*) were respectively inactivated to examine their roles in

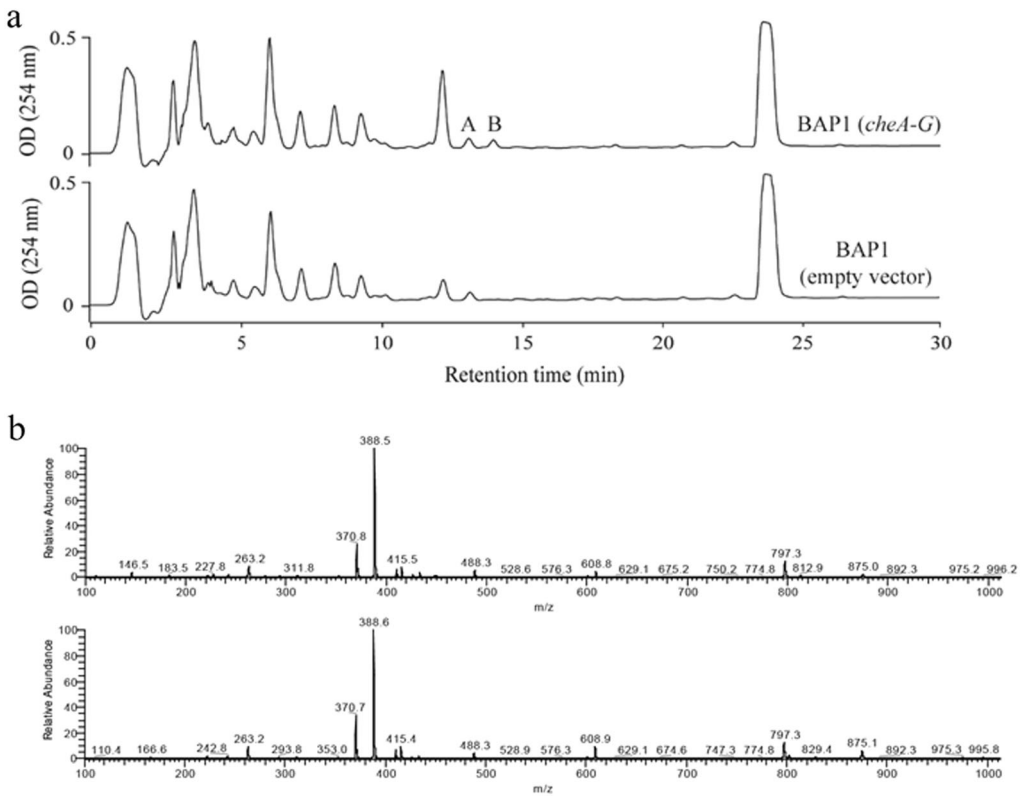

Fig. 4 HPLC analysis and mass spectrometry of the metabolites of BAP1(*cheA-G*) strain harboring the entire chejuenolide gene cluster (*cheA-G*). Expression of *cheA-G* was induced with 0.1 mM IPTG for 48 h at 22 °C. The broth culture was extracted with ethyl acetate prior to the HPLC analyses. **a** HPLC analysis of BAP1(*cheA-G*) and BAP1 (empty vector). **b** Mass spectrum of compound A and B

chejuenolides biosynthesis. The HC07 strain (Δ*orf7*) displayed no changes in chejuenolide production, whereas the HCE strain (*cheE* disruption mutant) did not produce chejuenolides (Figure S3), suggesting that *cheE* is essential for the biosynthesis of chejuenolides but not *orf7*. The *cheE* gene shows 63% identity with the *lkcE* of lankacidin which are reported to be involved in the cyclization of lankacidin [11]. In the study, Arakawa et al. suggested that LkcE catalyzes an amide intermediate of lankacidin (LC-KA05) to be an imide intermediate (at C-18) of which protonated form accepts the nucleophilic attack of an enolate ion (C-2) to give the 17-membered carbocyclic structure.

Feeding experiments revealed that [2-^{13}C]acetate feeding enriches eight carbons (C-2, C-5, C-7, C-9, C-11, C-13, C-15 and C-17) and [1-2-^{13}C]acetate feeding results in seven pairs of carbon-carbon coupling signals instead of eight pairs (Table S3). These results clearly shows that the C-1 of an acetate unit was removed. Since the C-2 of chejuenolide originates from the C-2 of acetate where the C-1 was removed, we suggest that a chejuenolide intermediate is first cyclized to form δ-lactone ring structure by CheE like in the lankacidin biosynthesis, then decarboxylation occurs to make the carbocyclic skeleton without δ-lactone ring of chejuenolides. The stereoisomer production of chejuenolide at C-18 may possibly originate from the decarboxylation reaction.

Heterologous expression of the che biosynthetic gene cluster

The hybrid NRPS/PKS (CheA) and the three *che* PKS (CheC, CheF and CheG) have less number of modules than the number of Claisen condensation reactions required for the chejuenolide biosynthesis. We postulate that one or more modules of the *che* PKS may iteratively participate in the chain elongation reaction to produce chejuenolides. To support this proposition, we must exclude the possibility of the involvement of other PKS homologous genes residing outside of the defined gene cluster. The entire *che* PKS (*cheA-G*) was cloned into two plasmids and heterologously expressed in *E. coli* BAP1. The *E. coli* BAP1(*cheA-G*) strain expressing *cheA-G* by IPTG induction generated two peaks in HPLC analyses that were not found in the culture of the BAP1(empty vector) strain containing pET21 and pET28a. Mass spectrometry of compound A and B produced by BAP1(*cheA-G*) revealed a molecular ion peak at *m/z* 388.5 $(M + H)^+$, which is in accordance with the mass of chejuenolides (Fig. 4) [21]. Compound A and B were

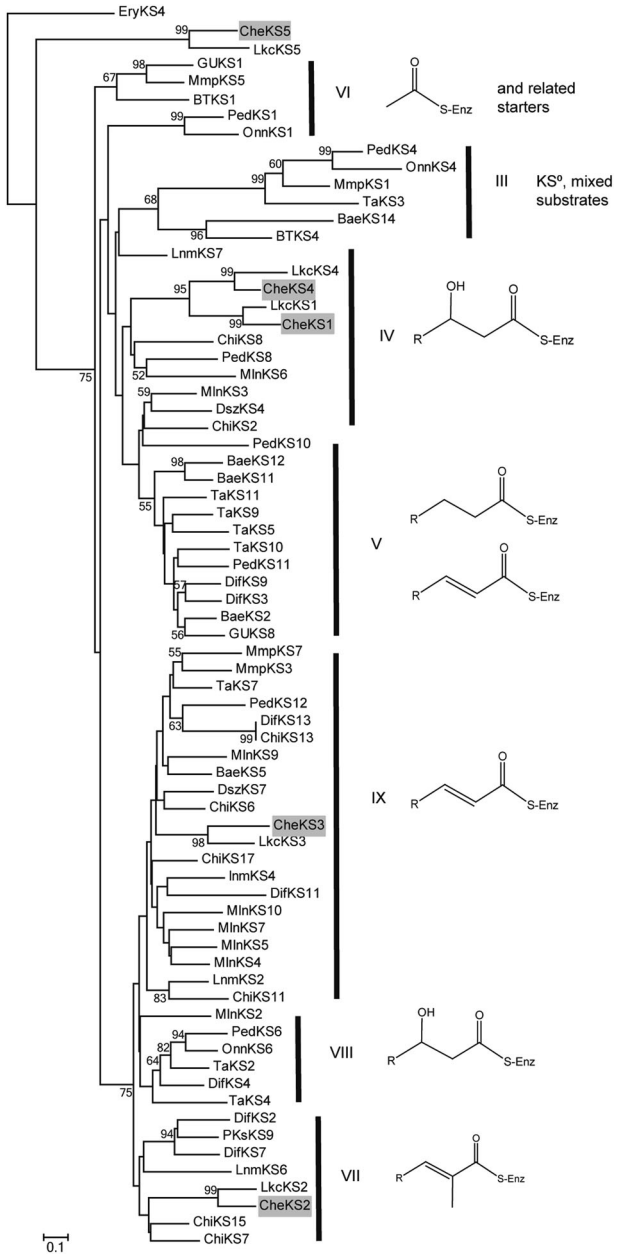

Fig. 5 Phylogram of 72 KS domains of *trans*-AT PKSs using neighbor-joining (NJ) algorithm. The bootstrap values from 50% are indicated at the nodes. KS numbering refers to the position in the gene cluster; for example, OnnKS4 is the fourth onnamide KS domain. The roman numbers refer to clade types with specific substrates. KS° represents the non-elongating KSs lacking the HGTG histidine. Mmp, mupirocin; Ta, myxivirescin; Mln, macrolactin; Lkc, lankacidin; Che, chejuenolide; Ped, pederin; Chi, chivosazol; Dif, difficidin; Lnm, leinamycin; Pks, Bacillaene; Onn, onamide; BT, uncharacterized PKSs from *B. thailandensis;* Dsz, disorazol; GU, uncharacterized PKSs from *Geobacter uraniumreducens*

purified from the culture of BAP1(*cheA*-G) and subjected to MS-MS analyses, which showed the identical mass fragmentation patterns to chejuenolide A and B isolated from MB-1084 (Figure S4). 1H-NMR spectra of compound A and B were identical to those of chejuenolide A and B, respectively (Table S4) [21]. These results clearly show that *cheA-G* is sufficient for chejuenolide biosynthesis and most importantly, this result strongly suggests that the five KS domains within the *che* PKS catalyze eight rounds of chain elongation to produce the chejuenolide skeleton.

Proposed biosynthetic pathway of chejuenolides

Although the heterologous expression study provided strong support for iterative *che* PKSs, we could not determine whether the module functions iteratively. We reasoned that the tandemly aligned ACPs of CheC could be a signature of iterative action of the module, where one ACP tethers a growing polyketide chain intermediate while the other ACP may be primed with an extension unit of malonyl-CoA. To test this hypothesis, point mutations of the active sites of CheC-ACP1 and CheC-ACP2 were carried out. Point mutations on either one of the ACP active sites, CheC-ACP1^{S869A} (strain M1) and CheC-ACP2^{S974A} (strain M2), did not affect chejuenolides production, while double mutations on the active sites of both ACPs (strain M12) caused the loss of chejuenolides production (Figure S5a), suggesting that either ACP is sufficient for chejuenolides biosynthesis. Since point mutation of the tandemly aligned ACPs does not provide any clues for the iterative functioning of the CheC module, we next mutated the KR domains to generate unreduced chejuenolides derivatives that indicate whether the module functioned iteratively. All of the mutant strains, HCKR (CheC-KR^{Y360F}), HCF1 (CheF-KR1^{Y342F}), and HCF2 (CheF-KR2^{Y1527F}), are abolished in the chejuenolides production (Figure S5b). Although we could not obtain unreduced intermediates from the KR mutations, these results indicate that all of the KR domains in the modules of Che PKS are functionally involved in chejuenolides biosynthesis.

Recent studies on *trans*-AT PKSs revealed that KS domains have preferences on incoming acyl-intermediates and that the preference of a given KS domain is predictable by phylogenetic analyses of their amino acid sequences [38]. Piel and his colleagues classified the KS domains of various *trans*-AT PKSs into clades I–XVI [39] and showed a close relationship between the observed KS selectivity and the structures predicted by the phylogenetic analysis. For example, clade II, IV, and VIII are assigned for β-hydroxylated, clade IX, XI, and XII for olefinic, and clade VII for α-methylated olefinic acyl-intermediates [40]. Analyses of *els* [38] and *sor* [41] PKSs have reinforced the utility of this approach in predicting the intermediate structures. Phylogenetic analysis of the five KS domains in the chejuenolides biosynthetic gene cluster showed that CheKS1 (CheA) and CheKS4 (KS2 of CheF) fall into clade IV, specific for

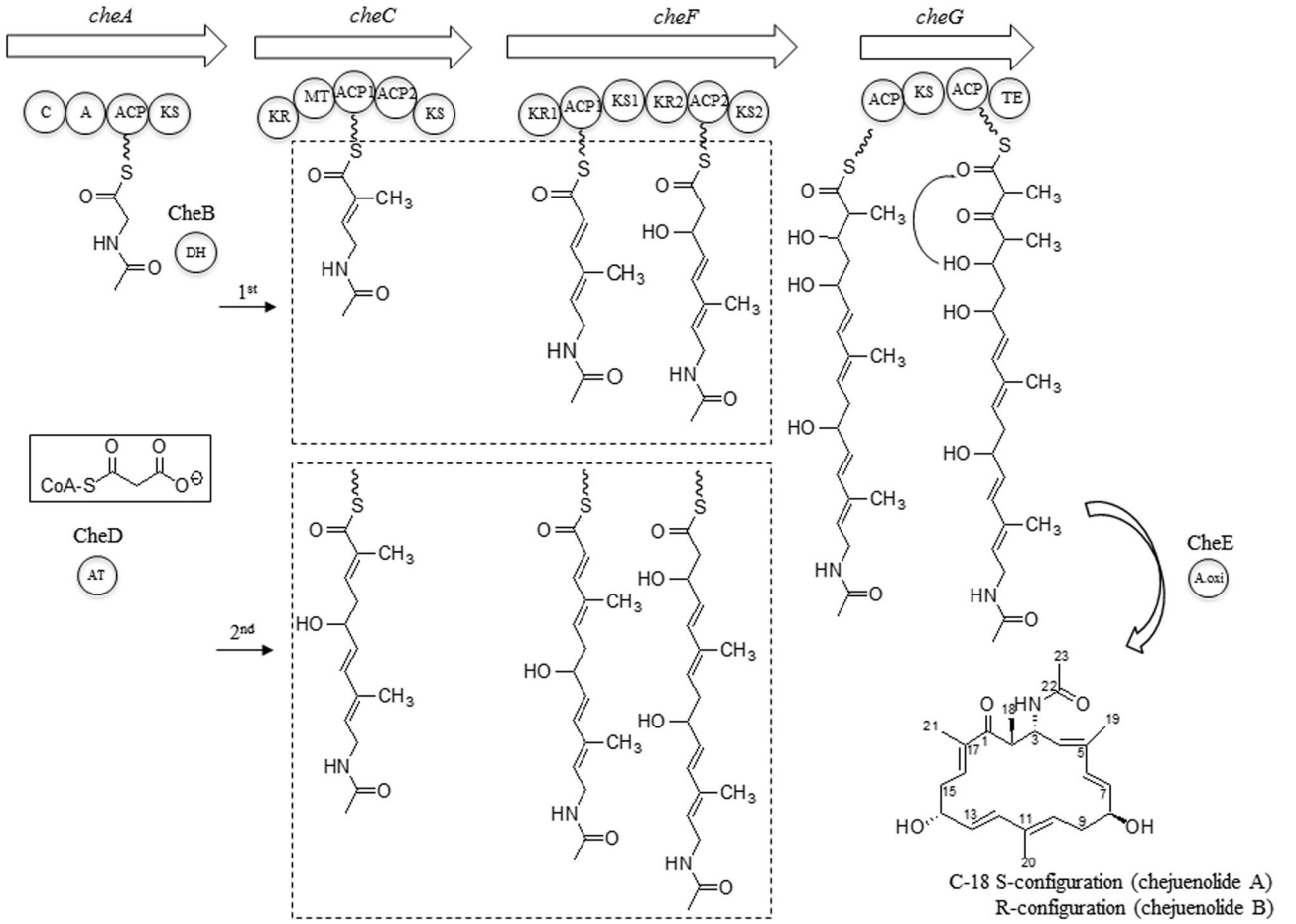

Fig. 6 A proposed biosynthetic route of chejuenolides and modular organization of the chejuenolide PKSs. The box represents the predicted iterative use of the modules. A, adenylation; ACP, acyl carrier protein; AT, acyltransferase; KR, ketoreductase; DH, dehydratase; KS, ketosynthase; MT, methyltransferase; TE, thioesterase

β-hydroxylated substrates. CheKS2 (CheC) is included in clade VII for α-methylated olefinic substrates, and CheKS3 (KS1 of CheF) is in clade IX for olefinic substrates (Fig. 5). CheKS5 (CheG) showed a high accuracy (99%) with the last KS domain of the lankacidin PKSs that were included in clade III for mixed substrates. The phylogenetic analysis proposed a reaction sequence of CheA–CheC–CheF–CheC–CheF–CheG in which CheC (contains one module) and CheF (contains two modules) coordinately repeat Claisen condensations. The proposition is in accordance with the structure of chejuenolides, which contain repeats of the α-methylated olefinic, olefinic, and β-hydroxylated carbons, suggesting a repeated use of three modules in a row (Fig. 6).

The *che* PKSs have similar modular arrangement to the *lkc* PKSs producing lankacidin C. The carbocyclic skeleton of lankacidin C is also similar to that of chejuenolides excluding the *δ*-lactone ring. Previous study on the lankacidin biosynthesis proposed an iterative work of LkcC (four times), which have not been supported by evidences [15]. In the proposition, the KS domain of LkcC should have a relaxed substrate specificity to accept three different acyl intermediates like α-methylated olefinic, olefinic, and β-hydroxylated substrates. Recent studies on the role/specificity of KS domain in *trans*-AT PKS systems revealed the gatekeeper function of KS domain to discriminate the incoming acyl-intermediates, and also a close correspondence between the observed KS selectivity and the one predicted by phylogenetic analyses. Based on the results of phylogenetic analysis of *che*PKS KS domains, we proposed that the three modules of CheC and CheF consecutively catalyze three Claisen condensation reactions, which are repeated again. The multimodular iteration proposed in this study is unusual in the modular PKSs but it is not unprecedented. In the biosynthesis of mannosyl-beta-1-phosphomycoketide, two modules catalyze five rounds of alternative condensations of methylmalonyl and malonyl units, which results in the mycoketide structure containing

branching at every alternate ketide unit [42]. More biochemical evidence of the selectivity of the KS domains and the iteration mechanism of *che* PKSs is required to validate this unusual biosynthetic pathway.

Acknowledgements This work was supported by a Grant of Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2017R1D1A1B03027996). We sincerely thank Professor Piel Jörn from Department of Biology, ETH Zürich for the advice on phylogenetic analysis, Professor Blaine Pfeifer from Department of Chemical and Biological Engineering, State University of New York for providing the *E. coli* BAP1 strain and Professor Kim Kun Soo from Department of Life Sciences, Sogang University for providing plasmid pKNG101.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. Nat Prod Rep. 2001;18:380–416.
2. Weissman KJ. Introduction to polyketide biosynthesis. Methods Enzymol. 2009;459:3–16.
3. McDaniel R, et al. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. Proc Natl Acad Sci USA. 1999;96:1846–51.
4. Chan YA, Podevels AM, Kevany BM, Thomas MG. Biosynthesis of polyketide synthase extender units. Nat Prod Rep. 2009;26:90–114.
5. Chen H, Du L. Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. Appl Microbiol Biotechnol. 2016;100:541–57.
6. Shen B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol. 2003;7:285–95.
7. Moss SJ, Martin CJ, Wilkinson B. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. Nat Prod Rep. 2004;21:575–93.
8. Tang G-L, Cheng Y-Q, Shen B. Polyketide chain skipping mechanism in the biosynthesis of the hybrid nonribosomal peptide—polyketide antitumor antibiotic leinamycin in Streptomyces a troolivaceus S-140⊥. J Nat Prod. 2006;69:387–93.
9. Piel J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. Proc Natl Acad Sci USA. 2002;99:14002–7.
10. Tang G-L, Cheng Y-Q, Shen B. Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. Chem Biol. 2004;11:33–45.
11. Arakawa K, Sugino F, Kodama K, Ishii T, Kinashi H. Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in Streptomyces rochei. Chem Biol. 2005;12:249–56.
12. Gaitatzis N, et al. The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. J Biol Chem. 2002; 277:13082–90.
13. He J, Hertweck C. Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. Chem Biol. 2003;10:1225–32.
14. Olano C, et al. Biosynthesis of the angiogenesis inhibitor borrelidin by Streptomyces parvulus Tü4055: cluster analysis and assignment of functions. Chem Biol. 2004;11:87–97.
15. Tatsuno S, Arakawa K, Kinashi H. Analysis of modular-iterative mixed biosynthesis of lankacidin by heterologous expression and gene fusion. J Antibiot. 2007;60:700–8.
16. Traitcheva N, Jenke-Kodama H, He J, Dittmann E, Hertweck C. Non-colinear polyketide biosynthesis in the Aureothin and Neoaureothin pathways: an evolutionary perspective. ChemBioChem. 2007;8:1841–9.
17. Menche D, et al. Stereochemical determination and complex biosynthetic assembly of etnangien, a highly potent RNA polymerase inhibitor from the myxobacterium Sorangium cellulosum. J Am Chem Soc. 2008;130:14234–43.
18. Müller S, et al. Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. Chem Biol. 2014;21:855–65.
19. Tao W, et al. A genomics-led approach to deciphering the mechanism of thiotetronate antibiotic biosynthesis. Chem Sci. 2016;7:376–85.
20. Hong H, et al. Evidence for an iterative module in chain elongation on the azalomycin polyketide synthase. Beilstein J Org Chem. 2016;12:2164–72.
21. Choi Y-H, et al. Chejuenolides A and B, new macrocyclic tetraenes from the marine bacterium Hahella chejuensis. Tetrahedron Lett. 2008;49:7128–31.
22. Sohn JH, Lee Y-R, Lee D-S, Kim Y-C, Oh H. PTP1B inhibitory secondary metabolites from marine-derived fungal strains Penicillium spp. and Eurotium sp. J Microbiol Biotechnol. 2013;23:1206–11.
23. Kieser, T. Practical streptomyces genetics. (John Innes Foundation, Norwich, UK, 2000).
24. Sambrook, J, Fritsch, EF & Maniatis, T. Molecular cloning. Vol. 2. (Cold spring harbor laboratory press, New York, 1989).
25. Gust B, Kieser T, Chater K. PCR targeting system in Streptomyces coelicolor A3 (2). Norwich: The John Innes Foundation; 2002.
26. Miller WG, Leveau JH, Lindow SE. Improved gfp and inaZ broad-host-range promoter-probe vectors. Mol Plant-Microbe Interact. 2000;13:1243–50.
27. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science. 2001;291:1790–2.
28. Marahiel MA, Stachelhaus T, Mootz HD. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem Rev. 1997;97:2651–74.
29. Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol. 1999;6:493–505.
30. Du L, Sánchez C, Chen M, Edwards DJ, Shen B. The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. Chem Biol. 2000;7:623–42.
31. Mofid MR, Finking R, Marahiel MA. Recognition of hybrid peptidyl carrier proteins/acyl carrier proteins in nonribosomal peptide synthetase modules by the 4′-phophopantetheinyl transferases AcpS and Sfp. J Biol Chem. 2002;277:17023–31.
32. Zhang Y-M, Hurlbert J, White SW, Rock CO. Roles of the active site water, histidine 303, and phenylalanine 396 in the catalytic mechanism of the elongation condensing enzyme of Streptococcus pneumoniae. J Biol Chem. 2006;281: 17390–9.
33. Ansari MZ, Sharma J, Gokhale RS, Mohanty D. In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. BMC Bioinforma. 2008;9:1.

34. Donadio S, Katz L. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in Saccharopolyspora erythraea. Gene. 1992; 111:51–60.
35. Tsai SC, et al. Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. Proc Natl Acad Sci USA. 2001;98:14808–13.
36. Reeves CD, et al. Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. Biochemistry. 2001;40: 15464–70.
37. Haydock SF, et al. Divergent sequence motifs correlated with the substrate specificity of (methyl) malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases. FEBS Lett. 1995;374:246–8.
38. Teta R, et al. Genome mining reveals trans-AT polyketide synthase directed antibiotic biosynthesis in the bacterial phylum Bacteroidetes. ChemBioChem. 2010;11:2506–12.
39. Nguyen T, et al. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat Biotechnol. 2008;26:225–33.
40. Jenner M, et al. Substrate specificity in ketosynthase domains from trans-AT polyketide synthases. Angew Chem Int Ed Engl. 2013;52:1143–7.
41. Irschik H, et al. Analysis of the Sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by trans-AT PKS. ChemBioChem. 2010;11:1840–9.
42. Chopra T, et al. Novel intermolecular iterative mechanism for biosynthesis of mycoketide catalyzed by a bimodular polyketide synthase. PLoS Biol. 2008;6:e163.