

Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites

Satoshi Ōmura^{*†}, Haruo Ikeda[‡], Jun Ishikawa[§], Akiharu Hanamoto^{*}, Chigusa Takahashi^{*}, Mayumi Shinose^{*}, Yoko Takahashi^{*}, Hiroshi Horikawa[¶], Hidekazu Nakazawa[¶], Tomomi Osonoe[¶], Hisashi Kikuchi[¶], Tadayoshi Shiba^{||}, Yoshiyuki Sakaki^{**††}, and Masahira Hattori^{††}

^{*}The Kitasato Institute for Life Sciences, Kitasato University, Tokyo 108-8642, Japan; [‡]School of Pharmaceutical Sciences, Kitasato University, Tokyo 108-8641, Japan; [§]National Institute of Infectious Diseases, Tokyo 162-8640, Japan; and [¶]National Institute of Technology and Evaluation, Tokyo 151-0066, Japan; ^{||}School of Sciences, Kitasato University, Kanagawa 228-8555, Japan; ^{**}Institute of Medical Science, University of Tokyo, 4-6-1, Shiohane-dai, Minato-ku, Tokyo 108-8639, Japan; and ^{††}The Institute of Physical and Chemical Research (Japan), Yokohama Institute, Genomic Sciences Center, Kanagawa 230-0045, Japan

Contributed by Satoshi Ōmura, August 16, 2001

Streptomyces avermitilis is a soil bacterium that carries out not only a complex morphological differentiation but also the production of secondary metabolites, one of which, avermectin, is commercially important in human and veterinary medicine. The major interest in this genus *Streptomyces* is the diversity of its production of secondary metabolites as an industrial microorganism. A major factor in its prominence as a producer of the variety of secondary metabolites is its possession of several metabolic pathways for biosynthesis. Here we report sequence analysis of *S. avermitilis*, covering 99% of its genome. At least 8.7 million base pairs exist in the linear chromosome; this is the largest bacterial genome sequence, and it provides insights into the intrinsic diversity of the production of the secondary metabolites of *Streptomyces*. Twenty-five kinds of secondary metabolite gene clusters were found in the genome of *S. avermitilis*. Four of them are concerned with the biosyntheses of melanin pigments, in which two clusters encode tyrosinase and its cofactor, another two encode an ochronotic pigment derived from homogentiginic acid, and another polyketide-derived melanin. The gene clusters for carotenoid and siderophore biosyntheses are composed of seven and five genes, respectively. There are eight kinds of gene clusters for type-I polyketide compound biosyntheses, and two clusters are involved in the biosyntheses of type-II polyketide-derived compounds. Furthermore, a polyketide synthase that resembles phloroglucinol synthase was detected. Eight clusters are involved in the biosyntheses of peptide compounds that are synthesized by nonribosomal peptide synthetases. These secondary metabolite clusters are widely located in the genome but half of them are near both ends of the genome. The total length of these clusters occupies about 6.4% of the genome.

Streptomyces is a genus of Gram-positive bacteria that grows in soil, marshes, and coastal marine habitats and forms filamentous mycelium-like eukaryote fungi. Morphological differentiation in *Streptomyces* involves the formation of a lawn of aerial hyphae on the colony surface that stands up into the air and differentiates into chains of spores (1). This process, unique among Gram-positive bacteria, requires the specialized coordination of metabolism and is more complex than other Gram-positive bacteria. The most interesting property of *Streptomyces* is its ability to produce secondary metabolites including antibiotics and bioactive compound (2) value in human and veterinary medicine, agriculture, and unique biochemical tools. Structural diversity is observed in these secondary metabolites that encompass not only antibacterial, antifungal, antiviral, and antitumor compounds, but also metabolites with immunosuppressant, antihypertensive, and antihypercholesterolemic properties. Thus, *Streptomyces* is a rich source of the secondary metabolites in which common intermediates in the cell (amino acids, sugars, fatty acids, terpenes, etc.) are condensed into more complex structures by defined biochemical pathways.

Characterization of chromosome ends of eight *Streptomyces* strains has revealed evidence of linear chromosomes, indicating that chromosomal linearity might be common in the streptomycetes (3). Most *Streptomyces* chromosomal DNA molecules are about 8-Mb-long, with terminal-inverted repeats and covalently bound terminal proteins supposedly at the 5' end. This size is unusually large for a bacterium, compared with well known microorganisms such as *Escherichia coli* and *Bacillus subtilis*. Streptomycetes have a higher G + C content (more than 70%) than nearly all other organisms. Thus, the *Streptomyces* chromosome is unique in its structure and size.

Here we describe the structure of the genome and properties of the sequence of *Streptomyces avermitilis*. We especially focus on the description of secondary metabolism in this microorganism.

Materials and Methods

Bacterial Strains. *S. avermitilis* ATCC31267 was used as the source of DNA for the physical map and genomic sequences. *E. coli* DH10B, DH5 α , and JM108 were used for the preparation of linking, shotgun, and cosmid libraries, respectively.

DNA Manipulation. The genomic DNA was prepared in agarose plugs. Agarose plugs containing intact genomic DNA of *S. avermitilis* were subjected to field-inverted gel electrophoresis by using a switching time of 3 sec for forward and 1 sec for reverse to remove two linear plasmids, SAP1 and SAP2. The agarose plugs were harvested and melted at 68°C for 15 min, and the chromosomal DNA was purified by phenol extraction.

For the construction of the shotgun library, the purified chromosomal DNA was cut to 1–2 kb by shearing force, using Hydro-Shear at the setting 6 and 20 passages. The sheared DNA fragments, 1–2 kb in length, were blunt-ended by using a DNA blunting kit (Takara, Kyoto, Japan). The blunt-ended DNA was ligated to the *Hinc*II site of pUC118, which had been treated previously with bacterial alkaline phosphatase, and ligation products were introduced into the *E. coli* DH5 α cell.

Cosmid and plasmid preparations, DNA restriction digestion, size fractionation, DNA fragment isolation, ligation reactions, lambda packaging, and gel electrophoresis were performed by

Abbreviations: ACP, acyl carrier protein; AT, acyl transferase; KS, β -ketoacyl-ACP synthase; NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AB070934–AB070957).

[†]To whom reprint requests should be addressed at: The Kitasato Institute, 9-1, Shirokane 5-chome, Minato-ku, Tokyo 108-8642, Japan. E-mail: omura-s@kitasato.or.jp.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

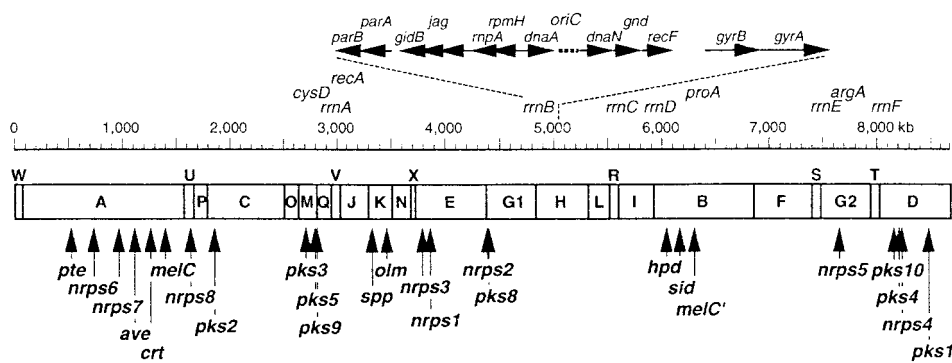


Fig. 1. Linear physical map of the chromosome of *S. avermitilis* ATCC31267 showing the position of known genes (*cysD*, *recA*, *proA*, and *argA*), six *rrn* operons, region around the replication origin, and secondary metabolite clusters. Vertical lines in boxes indicate recognition sites of the restriction enzyme *AseI*. All *rrn* regions have a unique *AseI* site between Q and V; G1 and H; R and I; I and B; S and G2; and T and D. Abbreviations of biosynthetic gene cluster symbols: *ave*, avermectin; *crt*, carotenoid; *hpd*, ochronotic pigment; *melC* and *melC'*, melanin; *nrps1–8*, peptide; *olm*, oligomycin; *pks1–10*, polyketide; *pte*, polyene macrolide; *sid*, siderophore; *spp*, spore pigment. Both *AseI*-W and -D contain telomere sequences.

standard procedures (4). *E. coli* transformation was performed by electroporation. The pUC118 was the routine cloning vector for the shotgun library, and pKU402 (5) and pKU310 were the cosmid vectors used for genomic DNA library construction and *AseI*-linking library construction, respectively.

DNA Sequencing and Assembly. The genome of *S. avermitilis* was basically sequenced by the whole genome shotgun sequencing method as described (6–8), and its sequencing relied on standard data-collection and sequence-assembly methods. The DNA fragment inserted into pUC118 was amplified by PCR, using M13 forward and reverse primers. The PCR fragments, treated with exonuclease I and shrimp alkaline phosphatase to eliminate excess primers and nucleotides in the PCR reaction mixture, were used for sequencing analysis as template DNA. The data were processed with the Phred/Phrap/Consed package (<http://www.phrap.org>) or the parallel-assembly SPS-Phrap package (Southwest Parallel Software, Albuquerque, NM) of base-calling, sequencing assembly, and finishing editing software. The data used dye-terminator chemistry to acquire usable shotgun-sequencing traces. The shotgun traces (177,631 forward and 9,033 reverse) provided about 10-fold coverage of the genome in high-quality base calls. The sequence traces (14,016) of both ends from 7,398 cosmid clones containing 40-kb inserts were also acquired. The inserts in these cosmids covered 99% of the genome. Hence, their end sequences provided a strong check on the validity of the final assembly. Furthermore, some cosmid clones containing repeat sequences, including rDNA loci, insertion sequences, and type-I polyketide synthase (PKS) genes, were determined to obtain the complete sequence of the insert.

Results and Discussion

(i) Physical map of *S. avermitilis*. At the beginning of sequencing the genome of *S. avermitilis*, the physical map of this organism was determined by using restriction enzyme *AseI*. The linking clones were isolated by insertion of the *AseI* fragment of a streptomycin/spectinomycin-resistant gene [*aad(3'')*] into the 5-kb genomic library. Each linking clone was determined by Southern hybridization of *AseI*-cut chromosome DNA with each insert of linking clone as a probe. Two hybridized bands were detected in almost all of the hybridization experiments, but some linking clones were hybridized with 11 *AseI* fragments that corresponded to *AseI*-B, -D, -G1, -G2, -H, -I, -J, -R, -S, -T, and -V, in which G1 and G2 fragments overlapped. This result indicated that the linking clones contained highly homologous sequences around the *AseI* site. The sequence linking the *AseI* site of these clones revealed that these clones contained the *rrn* operon, in which the *AseI* site was located in a 23S rDNA region. Cosmid clones containing an *rrn* operon were selected from the cosmid library, and the regions outside of the *rrn* operon were used for hybridization probes to prevent crosshybridization with the *rrn* region. The cosmid clones containing the *rrn* operons were classified into six groups, indicating that *S. avermitilis*

has six *rrn* operons in the genome. Ultimately, we determined the physical map of *S. avermitilis* by using results of linking patterns of 25 *AseI* segments and hybridization experiments, using PCR-derived amplified segments corresponding to *cysD*, *recA*, *oriC*, *proA*, and *argA* loci of *Streptomyces coelicolor* A3(2) (Fig. 1).

(ii) Sequencing, Assembly, and Structure of the Genome. We obtained about 200 contigs of more than 1 kb by assembling all of the data. The contigs could be turned into five valuable chains by using the linking information provided by cosmid-end sequences. Finally, the chains were ordered and oriented on the *AseI* physical map. The complete genome sequence is not yet determined. However, our sequence covers not only more than 99% of the genome but also gives us enough information for deducing the mechanisms of production of the secondary metabolites. Although the sequence of *S. avermitilis* genome has not been completely annotated (a few gaps remain in the assembled sequence data), we could recognize, however, most of the ORFs in the genome, because the gaps would contain less than a few hundred bp. The total ORFs in the genome was annotated to be at least 7,600, which is about 30 and 20% more than in the genomes of *Pseudomonas aeruginosa* (8) and the eukaryotic yeast *Saccharomyces cerevisiae* (9), respectively. Six *rrn* operons, in which each operon consists of the order 16S rDNA-23S rDNA-5S rDNA, were found. The genes coding transfer RNAs were estimated to number at least 65.

The linearity of chromosomes in *Streptomyces* was first discovered in *Streptomyces lividans* (10). Other *Streptomyces* chromosomes were later found to be linear structures with two terminal-inverted repeats at the both ends that were telomeres. Because these terminal sequences covalently bind proteins, these protein-DNA molecules bind to glass beads and are retarded during electrophoresis (3). The intact chromosomal DNA preparation without the treatment of proteinase K was cut by *AseI*, and the digests were subjected to pulse-field gel electrophoresis. Two *AseI* fragments, D and W, were not detected after electrophoresis, suggesting that these two fragments contain terminal ends, respectively. Two fragments of 0.9 and 4.5 kb were isolated from *Bam*HI-digested intact chromosome by a glass-binding procedure (3, 10) and were hybridized to *AseI*-W and -D, respectively. The linearity of the genome of *S. avermitilis* was also confirmed in terms of its assembly. There were no contigs at the outsides of both the *AseI*-W and -D fragments. In the terminal regions of *Streptomyces* chromosomes examined, there was a strong homology among the first 160 nucleotides in these sequences. These first 160 nucleotide sequences were searched for homology to shotgun sequence data of *S. avermitilis*, and two contigs were found to be highly homologous to the 160 nucleotides. The terminal sequence alignment of *S. avermitilis* and the other four *Streptomyces* (3) indicates that both terminal sequences of *S. avermitilis* share extensive homology to each other in the first 160 nucleotides. The sizes of the terminal-inverted repeats at both ends of chromosome range widely, from 24

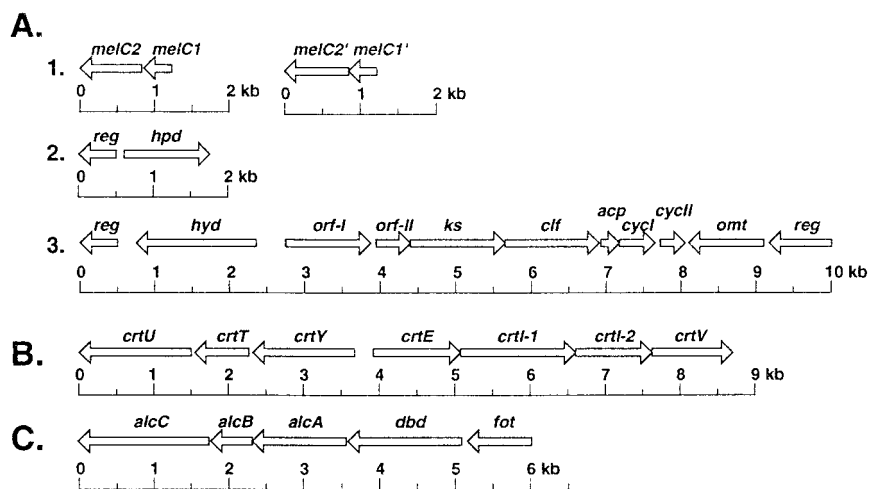


Fig. 2. Gene clusters for pigment and siderophore biosyntheses. Melanin pigment gene clusters are under *A*, carotenoid gene cluster is under *B*, and siderophore gene cluster is under *C*. Four melanin gene clusters are classified into three types: (i) melanin pigment formation involving tyrosinase, (ii) hydroxyphenylpyruvate dioxygenase, and (iii) type-II PKS. Abbreviations of gene symbols: *acp*, acyl carrier protein; *alcA*, monooxygenase; *alcB*, acetyltransferase; *alcC*, urease homolog; *clf*, chain-length factor; *crtE*, geranylgeranyl pyrophosphate synthase; *crtI*, phytoene synthetase; *crtT*, methyltransferase; *crtU*, β -carotene desaturase; *crtV*, methyltransferase; *crtY*, lycopene cyclase; *cyc*, cyclase; *dbd*, L-2,4-diaminobutyrate decarboxylase; *fot*, formyl transferase; *hpd*, 4-hydroxyphenylpyruvate dioxygenase; *hyd*, hydroxylase; *ks*, β -ketoacyl synthase; *melC1*, tyrosinase cofactor; *melC2*, tyrosinase; *omt*, O-methyltransferase; *reg*, regulatory protein.

to 550 kb in the *Streptomyces* chromosome (10–13), but the terminal-inverted repeats were found in the first 174 nucleotides, and long repeats such as in other *Streptomyces* chromosomes were not found in the genome of *S. avermitilis*.

As shown in Fig. 1, the replication origin (*oriC*) of *S. avermitilis* was located near the middle of the linear chromosome (precisely, *oriC* was shifted from the center of the chromosome to about 500 kb toward the right end), and the replication proceeds bidirectionally toward the telomeres. The gene organization within the replication origin region, where the gene order was *parB-parA-gidB-jag-orf-orf-rnpA-rpmH-dnaA-oriC-dnaN-gnd-recF-gyrB-gyrA*, is typical of bacteria possessing circular genomes (14).

(iii) Organization of Secondary Metabolite Clusters on the Chromosome. *S. avermitilis* has the highest proportion of predicted secondary metabolite gene clusters of all bacterial genomes sequenced. Analysis, using FRAMEPLOT (15), BLASTP (16), and HMMERPFAM (17), showed 25 clusters involving the biosynthesis of melanin, carotenoid, siderophore, polyketide, and peptide compounds (Fig. 1). The total lengths of these gene clusters were estimated to be about 560 kb. This analysis predicted that 6.43% of the *S. avermitilis* genome is occupied by genes concerned with the biosyntheses of secondary metabolites, a far higher proportion than has been found in other sequenced genomes. Almost none of these secondary metabolite clusters in *S. avermitilis* were located near the center of the chromosome and more than half were in the left hand from the *oriC*. Furthermore, about half of these clusters were also found near both ends of the chromosome. On the other hand, genes involved with primary metabolism, replication, transcription, and translation were located in a region about 6 Mb from *AseI-C* to *-T* fragments. These results indicate that some of the secondary metabolite clusters might have been horizontally transferred from donor microorganisms in the past. Furthermore, regions near both ends contain many transposase genes, indicating that transposases played an important evolutionary role in horizontal gene transfer and also in internal genetic rearrangements in the genome. Because some transposase genes were adjacent to secondary metabolite clusters, these transposases might have been involved in the transfer of these clusters.

(iv) Gene Clusters Involving Pigment and Siderophore Biosyntheses. *S. avermitilis* produces at least three kinds of melanin pigments; two are derived from tyrosine and one is an aromatic polyketide. The synthesis of the former involves tyrosinase, and the latter is synthesized from malonyl-CoA by a type-II PKS. These melanin pigments are produced on solid medium and then later accumulate in the spores. Another melanin is an ochronotic pigment that is

derived from homogentiginic acid and produced in both solid and liquid media (18). Two melanin gene clusters involving tyrosinase were found in the genome (Figs. 1 and 2). Both clusters were composed of two genes, tyrosinase cofactor (*MelC1*) and tyrosinase (*MelC2*), which have been found and sequenced in seven *Streptomyces* strains (GenBank accession nos. M11582, X95705, AB052940, AB022095, X95703, M11302, and AL356595). The alignment of amino acid sequences of these tyrosinases indicates that *MelC2* of *S. avermitilis* is similar to that of *Streptomyces galbus* (X95705). On the other hand, *MelC2'* was similar to that of *S. coelicolor*, which does not produce melanin. *MelC2'* probably does not function or its transcription level is too low. The genes involving melanin biosynthesis by the aromatic polyketide route have been found in most streptomycetes producing a spore pigment. The gene organization of aromatic polyketide melanin was quite similar to that in *S. coelicolor* (19). Another pigment gene cluster encodes the biosynthesis of a carotenoid, but the product synthesized by these genes has not yet been identified. Siderophores are involved in the transport of iron in bacteria. A gene cluster was found in the *S. avermitilis* that is presumably involved in the biosynthesis of desferrioximine derivatives, because most of the genes in the cluster are quite similar to those of *Bordetella bronchiseptica* and *Sinorhizobium meliloti* (GenBank accession nos. U61153 and AAK65921, respectively), which are responsible for desferrioxamine biosynthesis.

Gene Clusters Involving Polyketide Biosyntheses. Polyketides and the enzymes that makeup their carbon framework are ubiquitous components of microbial metabolism. *Streptomyces* and related bacteria are a rich source of structurally diverse polyketide natural products, which are derived from simple carboxylic acid precursors by a biosynthetic pathway closely analogous to the one that leads to long-chain fatty acids. There are two basic types of the PKS enzymes, iterative and modular, which are distinguished by both their architecture and reaction mechanism (20, 21). The type-I modular PKSs consist of relatively large multifunctional polypeptides commonly associated with the production of highly reduced metabolites such as macrolide antibiotics. In these PKSs, each catalytic domain is used only once during assembly of the product. On the other hand, iterative PKSs consist of both fungal type-I and bacterial type-II polypeptides, and each active domain often is used several times as the product is assembled. Bacterial type-II iterative PKSs are involved in the biosynthesis of aromatic polyketides.

S. avermitilis produces anthelmintic polyketide compounds, avermectins, which are the most important drugs for the treatment of endo- and ectoparasitic infections of livestock and humans (22). Eight clusters containing type-I PKS genes, including the avermectin biosynthetic gene cluster (23), were found in the *S. avermitilis*

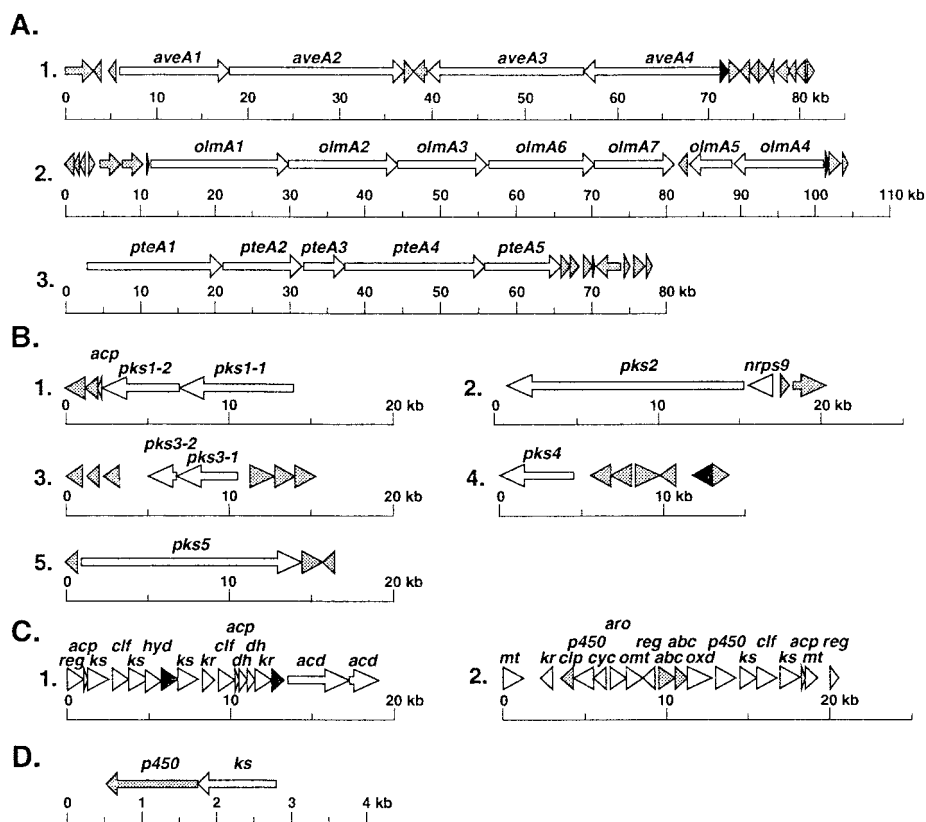


Fig. 3. Gene clusters for polyketide biosyntheses. Structures of metabolites assembled by type-I PKSs are characterized (A) and uncharacterized (B). Clusters contain type-II PKSs (C) and other types of PKSs (D). Open-boxed ORFs indicate type-I, -II, and other type PKS genes; shadowed-boxed ORFs are probably involved in the postpolyketide modification; and hatched-boxed ORFs would not be involved in the biosynthesis. *abc*, ABC transporter; *acd*, acyl-CoA dehydrogenase; *acp*, acyl carrier protein; *aro*, aromatase; *ave*, avermectin PKS; *clf*, chain-length factor; *clp*, ATP-dependent protease homolog; *cyc*, cyclase; *dh*, dehydratase; *hyd*, hydroxylase; *kr*, ketoreductase; *ks*, β -ketoacyl synthase; *mt*, methyltransferase; *p450*, cytochrome P450; *olm*, oligomycin PKS; *omt*, O-methyltransferase; *oxy*, oxidoreductase; *pte*, polyene macrolide PKS; *reg*, regulatory protein.

genome (Fig. 3). The deduced amino acid sequence of each PKS was analyzed by multiple-alignment, BLASTP, and HMMERPFAM search programs. Fundamentally, the modular PKS contains several catalytic domains in which the acyl-chain elongation involves acyl carrier protein (ACP), and β -ketoacyl-ACP synthase (KS) and acyltransferase (AT), and the reduction of the β -position is performed by β -ketoacyl-ACP reductase (KR), dehydratase, and enoylreductase.

Two of above clusters are involved in the biosyntheses of the macrocyclic lactone compounds, oligomycin and a polyene macrolide. The largest gene cluster (*olm*) consists of 7 genes encoding a PKS carrying 17 modules, including a loading module. These 17 modules contain 79 catalytic domains, but some are probably nonfunctional. Disruption of the *olmA4* region by transposition led to an oligomycin nonproducing phenotype, indicating that these seven PKSs catalyze the assembly of the polyketide backbone of oligomycin (24). On the other hand, there are five genes encoding PKS in the *pte* gene cluster (Fig. 3). These PKSs consist of 13 modules carrying 57 catalytic domains without nonfunctional domains. In consideration of the organization of domains in each module, 5 PKSs would yield a 26-membered pentaene compound.

Another five clusters were found to be type-I PKS genes (Table 1), but the putative metabolites formed from these gene products were not identified. Gene cluster *pks1* contained two kinds of PKS genes, but these genes were different from the general modular type-I PKSs. By BLASTP analysis, Pks1-1 and Pks1-2 have homology to KS and AT, but an ACP and other reduction domain(s) were not found. On the other hand, the HMMERPFAM search revealed that Pks1-1 possesses ACP and β -ketoacyl-ACP reductase (KR) domains but the homologies were at a low level. Surprisingly, the domain organization of Pks1-1 was different from that of other type-I PKSs. The ACP domain is normally located at the C terminus in type-I PKS, but the KR domain was found in the C terminus of Pks1-1. Furthermore, Pks1-2 did not possess ACP and other reduction domains by BLASTP and HMMERPFAM analyses, but a gene

encoding monofunctional ACP was adjacent to *pks1-2* downstream. Pks1-2 might be a new type of PKS and its catalytic form might be associated with a monofunctional ACP that is the gene product of the adjacent ORF. The common loading module, which consists of AT and ACP domains or nonfunctional KS, AT, and ACP domains, was not found in Pks2, but a gene encoding nonribosomal peptide synthetase (NRPS) was adjacent to upstream of *pks2*. The putative metabolite assembled by these gene products would be a derivative of a macrolactam. Pks3-2 and Pks7 were not found to contain KS and AT, which are fundamental domains for the chain elongation of polyketide synthesis, indicating that these clusters would be nonfunctional or another polypeptide(s) would be necessary for their catalytic reactions.

In contrast to reduced polyketides assembled by modular PKS,

Table 1. Deduced functions of ORFs in type-I polyketide biosynthetic gene clusters

Polypeptide	Module	Proposed function					
Pks1-1	Module 1	KS	AT			ACP [†]	KR [†]
Pks1-2	Module 2	KS	AT				
Pks2	Module 1	KS	AT	DH	KR	ACP	
	Module 2	KS	AT	DH	KR	ACP	TE
Pks3-1	Loading	KS*	AT			ACP	
Pks3-2	Module 1				KR	ACP	
Pks4	Module 1	KS	AT		KR	ACP	
Pks5	Loading	KS*	AT			ACP	
	Module 1	KS	AT	DH	KR	ACP	
	Module 2	KS	AT	DH	KR*	ACP	
Pks6	Module 1	KS	AT			ACP [†]	
Pks7	Module 1	KS				ACP [†]	

*Enzymatic activity is possibly nonfunctional.

[†]These domains were found by HMMERPFAM search and the E-values are relatively low. DH, dehydratase; TE, thioesterase.

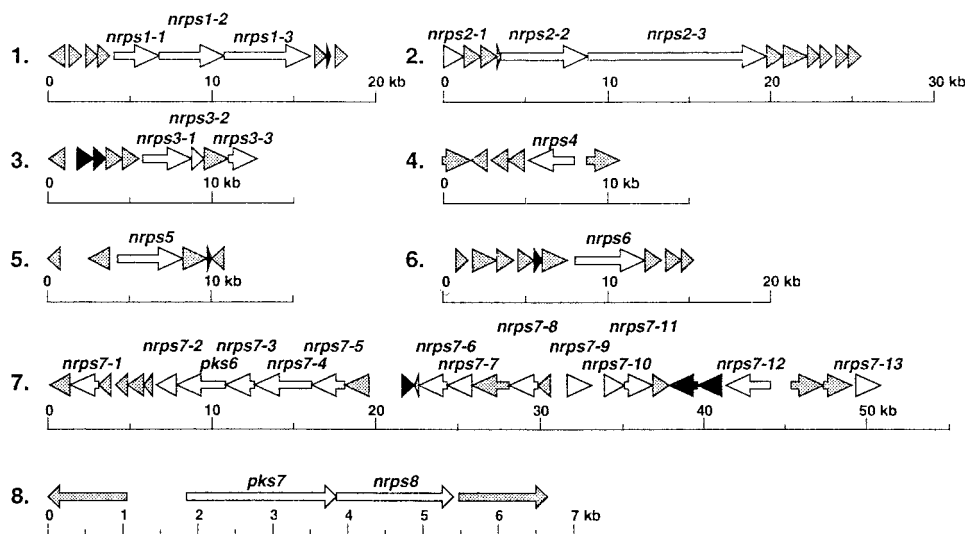


Fig. 4. Gene clusters for peptide biosynthesis. Open-boxed ORFs indicate NRPS genes, shadowed-boxed ORFs are probably involved in the postpeptide modification, and hatched-boxed ORFs would not be involved in biosynthesis.

type-II PKSs are composed of several usually monofunctional polypeptides that carry out the same action repeatedly and are involved in the synthesis of cyclic aromatic polyketides (20). There were three kinds of clusters containing type-II PKS genes, including polyketide pigment biosyntheses as described above for melanin

biosynthesis. Two of them would be involved in the synthesis of cyclic aromatic polyketides because they contain a minimal PKS unit (a monofunctional KS, chain-length factor, and ACP) and a dehydratase (aromatase and cyclase having dehydration activity). Surprisingly, the cluster of *pks8* (Fig. 3C-I) consisted of two pairs

Table 2. Prediction of adenylation domain specificity determining residues, amino acid substrates, and domain organization of NRPSs

Polypeptide	Residue in adenylation domain*								Substrate	Domain [†] organization
	235	236	239	278	299	301	322	330		
Nrps1-1	D	F	W	N	V	G	M	V	Threonine	C-A-T
Nrps1-2	D	A	W	L	L	G	A	V	Leucine	C-A-T-E
Nrps1-3	D	V	W	H	V	S	L	L	Serine	A-T
	D	G	T	L	T	A	E	V	Tyrosine	C-A-T
Nrps2-1	D	A	Q	E	L	A	V	L	Glutamine	A-T
Nrps2-2	D	A	W	L	Y	G	L	V	Leucine	C-A-T-E
	D	L	P	K	V	G	E	V	Asparagine	C-A-T
Nrps2-3	D	V	W	N	L	S	L	I	Serine	C-A-T-E
	D	L	P	K	V	G	E	V	Asparagine	C-A-T-E
	D	L	P	K	V	G	E	V	Asparagine	C-A-T-Te
Nrps3-1	D	M	E	L	L	G	L	I	Ornithine	C-A-T
Nrps3-2	ND	ND	ND	ND	ND	ND	ND	ND [†]	—	E-Te
Nrps3-3	D	V	W	H	V	S	L	V	Serine	A-T
Nrps4	D	L	T	K	L	G	E	V	Asparagine	A-T
Nrps5	D	V	Q	L	L	A	H	V	Proline	A-T
Nrps6	D	V	Q	L	I	A	H	V	Proline	C-A-T
Nrps7-1	D	F	E	T	T	A	A	V	Valine	A-T
Nrps7-2	D	A	K	D	L	G	V	V	Glutamate	A
Nrps7-3	D	F	Q	L	L	G	L	A	Pipecolate	A-T
Nrps7-4	D	A	F	W	L	G	G	T	Valine	A-T-C
Nrps7-5	D	A	Q	D	L	G	L	V	Glutamate	A-T
Nrps7-6	D	F	Q	L	V	G	V	A	Pipecolate	C-A-T
Nrps7-7	D	V	W	H	V	T	V	V	Serine	A-T
Nrps7-8	ND	ND	ND	ND	ND	ND	ND	ND [†]	—	T-C
Nrps7-9	ND	ND	ND	ND	ND	ND	ND	ND [†]	—	C
Nrps7-10	ND	ND	ND	ND	ND	ND	ND	ND [†]	—	C
Nrps7-11	D	L	Y	N	L	S	L	I	Cysteine	A-T
Nrps7-12	ND	ND	ND	ND	ND	ND	ND	ND [†]	—	T-C-T-Te
Nrps7-13	ND	ND	ND	ND	ND	ND	ND	ND [†]	—	T-C
Nrps8	D	L	V	F	G	L	G	I	Alanine	A
Nrps9	D	H	E	S	D	V	G	I	Cysteine	A

*According to GrsA numbering.

[†]Undetected consensus amino acid residue.

[‡]C, condensation domain; A, adenylation domain; T, thiolation domain; E, epimeration domain; Te, thioesterase domain; ND, not determined.

of minimal PKS units. On the other hand, the cluster of *pks9* (Fig. 3C-2) had one minimal PKS unit, a β -ketoacyl-ACP reductase, an aromatase, and a cyclase. The phylogenetic analysis from the results of alignments of the deduced amino acid sequences of type-II ketosynthase and chain-length factor indicates that the metabolite assembled by gene products from a cluster of *pks9* would be a decaketide, because both ketosynthase and chain-length factor have been classified into the group involved in the biosynthesis of decaketides.

Recently, a new type of PKS gene has been reported in the genomes of *Pseudomonas* (25) and *Streptomyces* (26) strains. Although they have homology to plant chalcone synthase, they could not use *p*-coumaroyl-CoA as substrate and their reactions are similar to type-II PKSs, which are used iteratively during chain elongation of polyketides. *Pks10* (Fig. 3D) has homology to these PKSs, suggesting that *Pks10* is involved in the synthesis of a tetraketide or pentaketide.

(v) Gene Clusters Involving Peptide Biosyntheses. Some microorganisms contain multifunctional complexes that build specific protein templates for a ribosomal-independent biosynthesis of low molecular weight peptides of diverse structure and a broad spectrum of biological activities. Although structurally diverse, NRPSs share a common mode of synthesis (27, 28). Peptide bond formation takes place on a multifunctional polypeptide (NRPS) on which amino acid substrates are first activated by ATP to the corresponding adenylate. The unstable adenylate is subsequently transferred to another site of the multifunctional polypeptide where it is bound as a thioester. Then thioesterified substrate amino acids are integrated into the peptide product through a step by step elongation by a series of transpeptidation reactions. Thus, the synthetic reaction of NRPS is similar to that of type-I PKS.

Eight clusters containing NRPS genes were found in the *S. avermitilis* genome (Figs. 1 and 4). Although screening for peptide products synthesized by NRPSs from cultures of *S. avermitilis* has not yet been carried out, *S. avermitilis* has the ability to produce peptide products. The adenylation domain of the NRPS selects the cognate amino acid from the pool of available substrates. Recent

studies have revealed that similarity between adenylation domains activating the same substrate is significantly high and there are defined general rules for the structural basis of substrate recognition by adenylation domains of NRPSs (29). The functional domains in each NRPS were searched for by HMMERPFAM analysis. The composition of domains in each module and conserved sequence motifs in adenylation domains are summarized in Table 2. Three clusters, *nrps1*, *nrps2*, and *nrps3*, contain three NRPS genes, respectively. It was assumed that the peptide products synthesized by these NRPSs were tetrapeptide, hexapeptide, and dipeptide, respectively. Because the *nrps6* cluster has a gene encoding a long chain fatty acid:CoA ligase and *Nrps6* contains a condensation domain, the product synthesized by these gene products would be acylated. Surprisingly, the *nrps7* cluster contains many genes encoding NRPSs with unusual architecture. In contrast to the common modular NRPSs consisting of multiple domains, *Nrps7-2*, *-7-8*, *-7-9*, *-7-10*, *-7-12*, and *-7-13* are discrete polypeptides homologous to individual domains of modular NRPSs. This type of unusual NRPSs has been found in the gene cluster for bleomycin biosynthesis (30).

Conclusion

We have found 25 kinds of secondary metabolite clusters by searching for homology to polypeptides of known function involved in secondary metabolism; it thus seems that *S. avermitilis* has at least 25 secondary metabolite clusters. There are many other uncharacterized genes involving secondary metabolism in this culture. For example, the volatile substance geosmin has been detected during the cultivation of *S. avermitilis*. Why do *Streptomyces* strains produce so many kinds of secondary metabolites including antibiotics and bioactive compounds? One of the answers is that *Streptomyces* strains have many gene clusters, which encode enzymes for many secondary metabolic pathways.

We thank K. Ohshima, H. Shimidzu, Y. Nakao, and N. Kushida for the technical assistance in the shotgun sequencing. We also thank Hitachi Instruments Service (Tokyo) for permission to use an Alpha server workstation.

- Waksman, S. A. & Henrici, A. T. (1943) *J. Bacteriol.* **46**, 337–341.
- Demain, A. L. (1999) *Appl. Microbiol. Biotechnol.* **52**, 455–463.
- Huang, C.-H., Lin, Y.-S., Yang, Y.-L., Huang, S.-W. & Chen, C. W. (1998) *Mol. Microbiol.* **28**, 905–916.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- Pang, C.-H., Shiiyama, M., Ikeda, H., Tanaka, H. & Ōmura, S. (1994) *Actinomycetologica* **8**, 21–25.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A. & Merrick, J. M. (1995) *Science* **269**, 496–512.
- Sakiyama, T., Takami, H., Ogasawara, N., Kuhara, S., Doga, K., Ohyama, A. & Horikoshi, K. (2000) *Biosci. Biotechnol. Biochem.* **64**, 670–673.
- Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrenner, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., et al. (2000) *Nature (London)* **406**, 959–964.
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., et al. (1996) *Science* **274**, 563–567.
- Lin, Y.-S., Kiser, H. M., Hopwood, D. A. & Chen, C. W. (1993) *Mol. Microbiol.* **10**, 923–933.
- Lezhava, A. L., Mizukami, T., Kajitani, T., Kameoka, D., Redenbach, M., Shinkawa, H., Nimi, O. & Kinashi, H. (1995) *J. Bacteriol.* **177**, 6492–6498.
- Leblond, P., Fischer, G., Francou, F. X., Berger, F., Guerieu, M. & Decaris, B. (1996) *Mol. Microbiol.* **19**, 261–271.
- Pandza, K., Pfalzer, G., Cullum, J. & Hranueli, D. (1997) *Microbiology* **143**, 1493–1501.
- Gal-Mor, O., Borovok, I., Av-Gay, Y., Cohen, G. & Aharonowitz, Y. (1998) *Gene* **217**, 83–90.
- Ishikawa, J. & Hotta, K. (1999) *FEMS Microbiol. Lett.* **174**, 251–253.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
- Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000) *Nucleic Acids Res.* **28**, 263–266.
- Denoya, C. D., Skinner, D. D. & Morgenstern, M. R. (1994) *J. Bacteriol.* **176**, 5312–5319.
- Davis, N. K. & Chater, K. F. (1990) *Mol. Microbiol.* **4**, 1679–1691.
- Hopwood, D. A. & Sherman, D. H. (1990) *Annu. Rev. Genet.* **24**, 37–66.
- Katz, L. & Donadio, S. (1993) *Annu. Rev. Microbiol.* **47**, 875–912.
- Burg, R. W., Miller, B. M., Baker, E. E., Birnbaum, J., Currie, S. A., Hartman, R., Kong, Y.-L., Monaghan, R. L., Olson, G., Putter, I., et al. (1979) *Antimicrob. Agents Chemother.* **15**, 361–367.
- Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. & Ōmura, S. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 9509–9514.
- Ikeda, H., Takada, Y., Pang, C.-H., Tanaka, H. & Ōmura, S. (1993) *J. Bacteriol.* **175**, 2077–2082.
- Bangera, M. G. & Thomashow, L. S. (1999) *J. Bacteriol.* **181**, 3155–3163.
- Funai, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y. & Horinouchi, S. (1999) *Nature (London)* **400**, 897–899.
- Lipmann, F. (1980) *Adv. Microbiol. Physiol.* **21**, 227–266.
- Zocher, R. & Keller, U. (1997) *Adv. Microbiol. Physiol.* **38**, 85–131.
- Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. (1999) *Chem. Biol.* **6**, 493–505.
- Du, L., Sanchez, S., Chen, M., Edwards, J. & Shen, B. (2000) *Chem. Biol.* **7**, 623–642.