JOURNAL OF BACTERIOLOGY, June 2006, p. 4024–4036 Vol. 188, No. 11
0021-9193/06/$08.00+0 doi:10.1128/JB.00052-06

Structural and Functional Characterization of Three Polyketide Synthase Gene Clusters in *Bacillus amyloliquefaciens* FZB 42

Xiao-Hua Chen,[1] Joachim Vater,[2] Jörn Piel,[3] Peter Franke,[4] Romy Scholz,[1] Kathrin Schneider,[5] Alexandra Koumoutsi,[1] Gabriele Hitzeroth,[2] Nicolas Grammel,[2] Axel W. Strittmatter,[6] Gerhard Gottschalk,[6] Roderich D. Süssmuth,[5] and Rainer Borriss[1]*

Institut für Biologie, AG Bakteriengenetik, Humboldt-Universität Berlin, Berlin,[1] Institut für Chemie, AG Biochemie und Molekulare Biologie, Technische Universität Berlin, Berlin,[2] Kekulé-Institut für Organische Chemie und Biochemie, Universität Bonn, Bonn,[3] Institut für Chemie und Biochemie, Freie Universität Berlin, Berlin,[4] Institut für Chemie/Biologische Chemie, Technische Universität Berlin, Berlin,[5] and Institut für Mikrobiologie und Genetik, Laboratorium für Genomanalyse, Georg-August-Universität Göttingen,[6] Göttingen, Germany

Received 12 January 2006/Accepted 28 February 2006

Although bacterial polyketides are of considerable biomedical interest, the molecular biology of polyketide biosynthesis in *Bacillus* spp., one of the richest bacterial sources of bioactive natural products, remains largely unexplored. Here we assign for the first time complete polyketide synthase (PKS) gene clusters to *Bacillus* antibiotics. Three giant modular PKS systems of the *trans*-acyltransferase type were identified in *Bacillus amyloliquefaciens* FZB 42. One of them, *pks1*, is an ortholog of the *pksX* operon with a previously unknown function in the sequenced model strain *Bacillus subtilis* 168, while the *pks2* and *pks3* clusters are novel gene clusters. Cassette mutagenesis combined with advanced mass spectrometric techniques such as matrix-assisted laser desorption ionization–time of flight mass spectrometry and liquid chromatography-electrospray ionization mass spectrometry revealed that the *pks1* (*bae*) and *pks3* (*dif*) gene clusters encode the biosynthesis of the polyene antibiotics bacillaene and difficidin or oxydifficidin, respectively. In addition, *B. subtilis* OKB105 (*pheA sfp*^{0}), a transformant of the *B. subtilis* 168 derivative JH642, was shown to produce bacillaene, demonstrating that the *pksX* gene cluster directs the synthesis of that polyketide.

Environmental *Bacillus amyloliquefaciens* strain FZB 42 is distinguished from the domesticated model organism *Bacillus subtilis* 168 (23) by several features important for rhizosphere competence particularly by its abilities to suppress competitive organisms present in the plant rhizosphere (17, 21) and to promote plant growth (16). In a previous contribution (20), we have reported that *B. amyloliquefaciens* FZB 42 is a producer of three families of lipopeptides, surfactins, bacillomycins D, and fengycins, which are well-known secondary metabolites with mainly antifungal activity. They are also produced by numerous *B. subtilis* strains (48). Furthermore, three giant gene clusters containing genes with homology to polyketide synthase (PKS) genes of modular organization were identified but not assigned functional roles. Mutants of FZB 42 deficient in the synthesis of cyclic lipopeptides were unable to suppress phytopathogenic fungi but still retained their antibacterial potency.

Polyketides belong to a large family of secondary metabolites that include many bioactive compounds with antibacterial, immunosuppressive, antitumor, or other physiologically relevant bioactivities. Their biosynthesis is accomplished by stepwise decarboxylative Claisen condensations between the extender unit and the growing polyketide chain, generating enzyme-bound β-ketoacyl intermediates. Before a subsequent round of chain extension, a variable set of modifying enzymes can locally introduce structural variety. Similar to the nonribosomal synthesis of peptides, the PKS multienzyme system uses acyl carrier proteins (ACPs) that are posttranslationally modified with the 4′-phosphopantetheine prosthetic group to channel the growing polyketide intermediate during elongation processes (3). Type I PKSs are modularly organized giant synthases, each module of which usually contains a β-ketoacyl synthase (KS), an acyltransferase (AT), and the ACP as essential and basic domains that may be complemented by a variable set of additional domains. The order of the modules dictates the sequence of biosynthetic events, and the generally observed colinearity between PKS structure and biosynthetic steps was shown to permit combinatorial manipulation of type I PKSs in order to generate novel compounds (7).

Although the model strain *B. subtilis* 168 has been shown to contain a large PKS cluster designated the *pksX* system (1), it is unable to synthesize polyketides because of a mutation in the *sfp* gene (27–29) encoding 4′-phosphopantetheine transferase. Sfp not only phosphopantetheinylates the peptidyl carrier proteins of surfactin and other lipopeptides but displays broad substrate specificity for other carrier proteins as well, including ACPs of PKSs (27). Production of polyketide-like compounds with antibacterial activity by wild-type isolates of *B. subtilis* has been described previously (15). The polyene antibiotics difficidin and oxydifficidin are highly unsaturated 22-member macrolides with a rare phosphate group (51). Another antibiotic, bacillaene, was demonstrated to be a conjugated hexaene with a linear structure (33), but its chemical structure is still un-

* Corresponding author. Mailing address: Institut für Biologie, Humboldt Universität Berlin, Chausseestrasse 115, D-10115 Berlin, Germany. Phone: 49-30-2093-8137. Fax: 49-30-2093-8127. E-mail: rainer.borriss@rz.hu-berlin.de.

TABLE 1. Strains and plasmids used in this study

Strain or plasmid	Description	Source and/or reference
B. subtilis		
JH642	*trpC2 pheA1 sfp*^{0}	27
OKB105	*pheA1 sfp*$^+$, surfactin producer, JH642 transformed with DNA of ATCC 21332	27
ATCC 39320	Wild type, difficidin producer	53
B. amyloliquefaciens		
FZB 42	Wild type, producer of lipopeptides and polyketides	17
AK2	Δ*fenA*::*cat*	20
CH1	*dsrfAA*::*ermAM*, deficient in surfactin synthesis	20
CH2	Δ*fenA*::*cat* Δ*srfAA*::*ermAM*, deficient in fengycin and surfactin synthesis	TF[a] CH1 → AK2
CH3	Δ*sfp*::*ermAM*, deficient in lipopeptides and polyketides	This work
CH4	Δ*yczE*::*ermAM*, reduced synthesis of polyketides	X.-H. Chen, unpublished
CH6	*dpks1KS1*::*cat*, no bacillaene synthesis	pPKS1cat → FZB 42, this work
CH7	*dpks2KS1*::*cat*, no synthesis of unknown polyketide 2	pPKS2cat → FZB 42, this work
CH8	*dpks3KS1*::*ermAM*, no synthesis of difficidin	pPKS3erm → FZB 42, this work
CH11	*dpks1KS1*::*cat* Δ*pks3KS1*::*ermAM*, no synthesis of bacillaene or difficidin	CH6 → CH8, this work
CH12	*dpks2KS1*::*cat* Δ*pks3KS1*::*ermAM*, no synthesis of unknown polyketide 2 or difficidin	CH8 → CH7, this work
CH13	*dpks2KS1*::*neo*, no synthesis of unknown polyketide 2	pECE73 → CH7, this work
CH14	*dpks1KS1*::*cat* Δ*pks2KS1*::*neo*, no synthesis of unknown polyketide 2 or bacillaene	CH13 → CH6, this work
Plasmids		
pMX39	*ermAM*, *E. coli-Bacillus* shuttle plasmid based on pBR322 and pDB101	5
pDG364	*cat*, *Bacillus* integration vector	8
pECE73	*cat* → *neo* exchange vector	44
pPKS1cat	*Bacillus* integration vector containing Cm^r cassette flanked by *pks1KS1* sequences	This work
pPKS2cat	*Bacillus* integration vector containing Cm^r cassette flanked by *pks2KS1* sequences	This work
pPKS3erm	*Bacillus* integration vector containing Cm^r cassette flanked by *pks3KS1* sequences	This work

[a] TF, transformant.

known. Since the discovery of the *pksX* system in the genome of *B. subtilis* 168, its function has been subject to ample speculation, with difficidin commonly being discussed as its metabolic product (42). So far, no genetic structure could be functionally assigned to the biosynthesis of this or any other polyketide in bacilli, except for the zwittermicin system, in which a part of the genes contributing to its biosynthesis have recently been uncovered (12).

Here we characterize the giant modular PKS gene clusters *pks1* and *pks3*, which are responsible for the biosynthesis of the antibiotics bacillaene and difficidin or oxydifficidin in *B. amyloliquefaciens*. Another gene cluster, *pks2*, is involved in the synthesis of an additional but not yet characterized polyketide that displayed weak antibacterial activity in our assays. Notably, all three PKS systems exhibit the unusual *trans*-AT architecture described recently; i.e., all PKS modules lack an AT domain and are complemented by ATs encoded on isolated genes (34). The clusters display a high degree of homology, suggesting that they might have evolved from a single ancestral operon.

MATERIALS AND METHODS

Strains, growth conditions, and DNA transformation. *B. amyloliquefaciens* strain FZB 42 was described previously (17). The indicator strains used for bioassays were *Bacillus megaterium* (laboratory stock) and *Erwinia carotovora* (B. Krebs, FZB Biotech, Berlin, Germany). Difficidin or oxydifficidin producer strain *B. subtilis* ATCC 39320 (51) was used for comparison of matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectra. The strains used in this study are summarized in Table 1. Bacteria were cultivated routinely on Luria broth (LB) medium solidified with 1.5% agar. For polyketide production and MS analysis, the bacteria were grown in Landy medium (24). The media and buffer used for DNA transformation of *Bacillus* cells were prepared according to the method of Kunst and Rapoport (22). Competent cells were prepared as previously described (20).

Mass spectrometric (MS) analysis. (i) MALDI-TOF MS. Polyketides of *B. amyloliquefaciens* FZB 42 were identified in extracts of lyophilized culture filtrates and XAD7 extracts by MALDI-TOF MS. Mass spectra were recorded with a Bruker Daltonik Reflex MALDI-TOF instrument containing a 337-nm nitrogen laser for desorption and ionization. One hundred to 200 single scans were accumulated for every spectrum. 2,5-Dihydroxbenzoic acid was used as the matrix. For MS analysis, 1- to 2-μl portions of extracts were mixed with an equal volume of matrix solution, spotted onto the target, and air dried.

Positive-ion detection and reflector mode were used. The acceleration and reflector voltages were 20 and 23.4 kV in pulsed ion extraction mode. A molecular mass gate of 300 Da improved the measurements by filtering out most matrix ions. Monoisotopic mass numbers were obtained.

(ii) High-performance liquid chromatography-electrospray ionization (HPLC-ESI) MS. HPLC-ESI MS was performed on a QTRAP 2000 system (Applied Biosystems, Darmstadt, Germany) coupled with an Agilent 1100 HPLC system (Agilent, Waldbronn, Germany). Aliquots of acetonitrile (ACN)-water extracts of the culture filtrates of wild-type and mutant strains were fractionated by reversed-phase HPLC on a Merck LiChroCART C_{18} 5-μm column (125 by 4 mm) at a flow rate of 1.5 ml/min with a gradient of 0% ACN plus 0.1% formic acid to 100% ACN plus 0.1% formic acid in 10 min. The eluent was split to a flow rate of 300 μl/min and fed online into the spray chamber. Every sample was measured in the negative and positive modes, and mass spectra were acquired in an *m*/*z* range of 300 to 800 at a scan rate of 1,000 atomic mass units/s.

Construction of mutants defective in polyketide synthesis. A summary of all of the mutant strains and plasmids used in this study is given in Table 1. To generate *sfp* mutant CH3 (Δ*sfp*::*ermAM*), a 2.1-kb PCR fragment was amplified with primers sfp-1 (5′-TCAACGTGTCCAACGTCAAG) and sfp2-2 (5′-AGTGATT AAGGATTTGGCGAAC) and cloned into pGEM-T. Plasmid pCH3 was obtained by insertion of *ermAM* isolated from plasmid pMX39, a derivative of pDB101 (5), into the central *sfp* gene region as previously described (20). Mutant CH3 was obtained after transformation of FZB 42 by the linearized plasmid.

Mutant CH6 (Δ*pks1KS1*::*cat*) was obtained as follows. With primers pks1Up (5′-AAAGGAGCGGAGTGCAACATC) and pks1Dw (5′-TGAGATGATGCC GTCCTCTTC), a 3.3-kb fragment containing the first KS domain of the *pks1* gene cluster was amplified by PCR and cloned into vector pGEM-T. A central

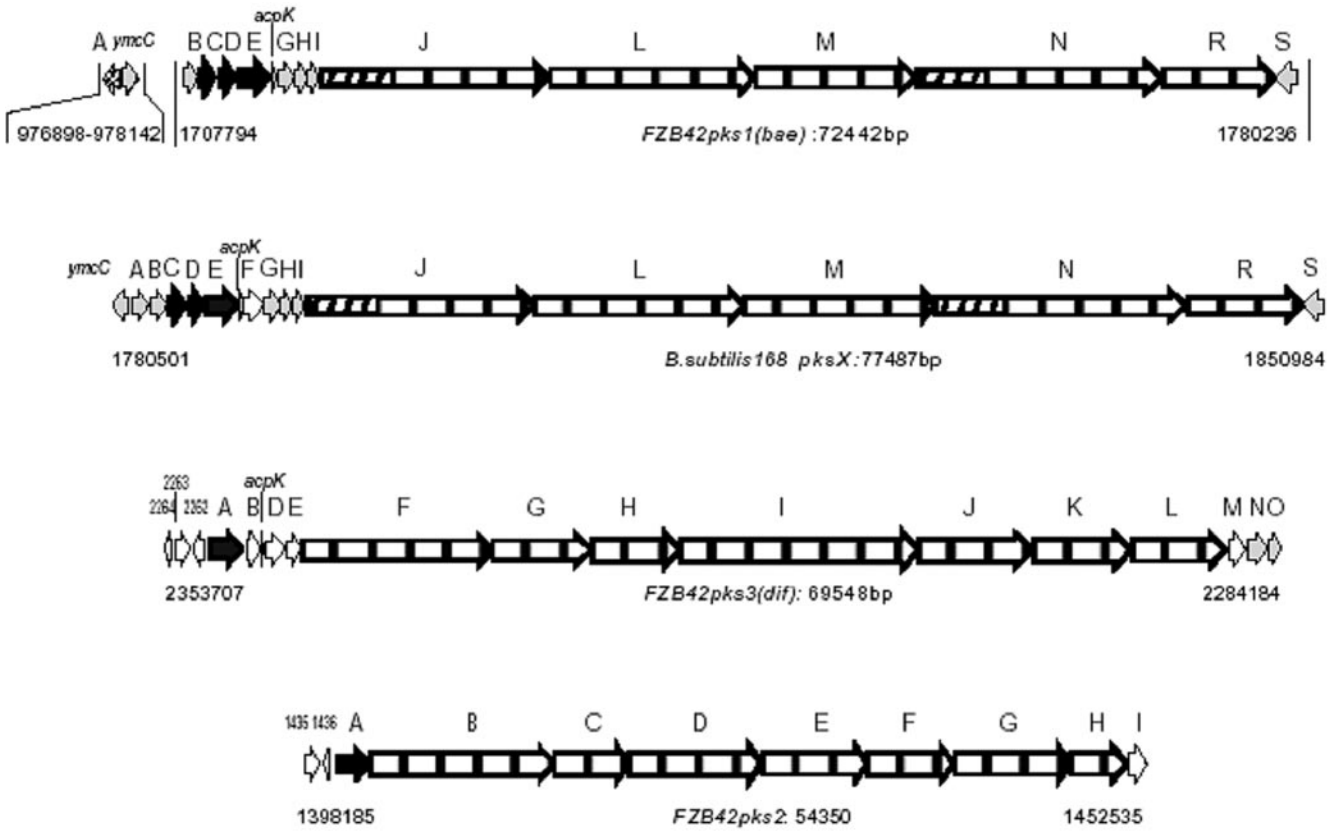

FIG. 1. Organization of the gene clusters involved in polyketide synthesis in *B. amyloliquefaciens* FZB 42 (*pks1*, *pks2*, and *pks3*) and *B. subtilis* 168 (*pksX*). Discrete ATs are indicated by filled arrows, and modular PKS are indicated by bold arrows. The NRPS portions occurring in hybrid NRPS-PKS enzymes are shaded. Further details of the gene products are presented in Table 2.

1.4-kb fragment of the insert was removed by digestion with EcoRI and HpaI and replaced with a chloramphenicol resistance cassette, yielding pCH6. The ScaI-linearized plasmid was recombined via double homologous recombination into competent *B. amyloliquefaciens* FZB 42 cells, yielding mutant CH6.

Mutant CH7 (Δ*pks2KS1*::*cat*) was obtained as follows. A 3.0-kb PCR fragment containing the first KS domain of *pks2* was amplified with primers pks2Up (5′-AGCTCATTGACAGCGATGCTGC) and pks2Dw (5′-ATGATCGCCGCT TCCTCATCAG) and cloned into vector pGEMT. An internal 1.3-kb fragment was removed by EcoRI-KpnI digestion and replaced with the 1.3-kb *cat* cassette. This plasmid was linearized with ScaI and transformed into FZB 42 to gain *pks2* knockout mutant CH7.

Mutant CH8 (Δ*pks3KS1*::*ermAM*) was obtained as follows. The *pks3* gene cluster was disrupted by insertion of an Em^r gene cassette. A 2.2-kb fragment containing the first KS domain of the biosynthesis gene cluster was amplified with primers pks3Up (5′-ACATTGACCGCATCCTCAATCTG) and pks3Dw (5′-TCAGCTGCTGTTCGGAATGTG), cloned into vector pGEM-T, and then digested with EcoRI and Eco72I, which removed a central 376-bp fragment. An Em^r gene cassette (1.9 kb) derived from pMX39 was cloned into the deleted construct, yielding pCH8. The *pks3* mutant CH8 was obtained after transformation of competent FZB 42 cells with ApaI-linearized pCH8.

Strain CH11 (Δ*pks1KS1*::*cat* Δ*pks3KS1*::*ermAM*), impaired in the synthesis of bacillaene and difficidin, was generated by transformation of strain CH8 with chromosomal DNA isolated from strain CH6.

Strain CH12 (Δ*pks2KS1*::*cat* Δ*pks3KS1*::*ermAM*), impaired in the synthesis of difficidin and with a knockout mutation of the *pks2* gene cluster, was generated by transformation of strain CH7 with chromosomal DNA isolated from strain CH8.

Strain CH14 (Δ*pks1KS1*::*cat* Δ*pks2KS1*::*neo*), impaired in the synthesis of bacillaene and with a knockout mutation of the *pks2* gene cluster, was generated by transformation of strain CH6 with chromosomal DNA isolated from strain CH13. In order to avoid identical antibiotic resistances in the double-mutant strain, the antibiotic resistance of strain CH8 was previously changed to neomycin resistance by means of marker exchange plasmid pECE73 (44), yielding strain CH13 (Table 1).

All mutants were selected from media containing the appropriate antibiotics at the concentrations recommended by Cutting and van der Horn (8) and confirmed by Southern hybridization and colony PCR with appropriate primers.

Bioautography. Bioautographs were prepared as previously described (15). In short, supernatants from bacteria grown in Landy medium were loaded onto an XAD7 (Sigma) resin column, washed, and eluted with methanol. After drying, the samples obtained were redissolved in methanol, spotted onto silica gel 60 F254 thin-layer chromatography (TLC) aluminum sheets (20 by 20 cm; Merck, Darmstadt, Germany), and separated by TLC chromatography (chloroform-methanol-water at 65:25:4, vol/vol). Afterwards, strips of the TLC plate were placed for 2 h at room temperature on the surface of a top agar layer containing the indicator strain. The agar plates were then incubated overnight at 37°C. Zones of inhibition documented the positions of antibiotics separated by TLC. The experiments were repeated at least three times.

Analysis of domain structure. The program package SEARCHPKS of the modular polyketide synthase database available at http://linux1.nii.res.in/~pksdb/DBASE/page.html and BLAST comparison with annotated domains of similar PKSs were used to detect conserved active-site motifs.

Construction of phylogenetic trees. The amino acid sequences of discrete *Bacillus* ATs and of the KS domains extracted from modular *Bacillus* PKSs were used for BLASTP comparison in order to detect their closest orthologs. Sequences were aligned by the ClustalW program (47) accessible at http://www.ebi.ac.uk/clustalw/. A distance matrix was calculated from this alignment by protein distance matrix calculation (PROTDIST program), and the matrix was then transformed into a tree by the NEIGHBOR program. In order to verify the accuracy of the tree, multiple data sets were generated with the SEQBOOT program using 1,000 bootstrap replicates. A tree was built from each replicate with the PROTDIST program, and then bootstrap values were computed with the CONSENSE program. The phylogenetic tree was visualized with the TreeView32 program (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). The programs used

to construct the phylogenetic tree were obtained from the PHYLIP package, v.3.65 (13), which is accessible at http://evolution.gs.washington.edu/phylip.html.

Evaluation of oligonucleotide usage (OU) pattern. Computational analysis to determine local OU variance (OUV) was performed as described by Reva and Tummler (39).

Nucleotide sequence accession numbers. The GenBank accession numbers for gene clusters *pks1* (*bae*), *pks2*, and *pks3* (*dif*) are AJ634060.2, AJ634061.2, and AJ634062.2, respectively.

RESULTS AND DISCUSSION

Discovery of three large *trans*-AT PKS gene clusters in *B. amyloliquefaciens*. Three PKS gene clusters together covering 196,340 bp (Fig. 1) were found at sites around 1.4 Mbp (*pks2*), 1.7 Mbp (*pks1*), and 2.3 Mbp (*pks3*) distant clockwise from the replication origin of the *B. amyloliquefaciens* genome (20), which is 3,916 kb in size (X.-H. Chen and R. Borriss, unpublished data). All three gene clusters contain the giant open reading frames typical of type I PKS systems. In addition, the *pks1* gene cluster contains two hybrid PKS-nonribosomal peptide synthetase (NRPS) genes. The organization of gene cluster *pks1* is very similar to that of the *pksX* gene cluster of *B. subtilis* 168 (23), but a *pksA*-like sequence corresponding to a putative transcriptional regulator is missing within the *pks1* gene cluster. However, we detected a homolog of *pksA* adjacent to a β-ketoacyl-ACP synthase (KAS) in a region 0.97 Mbp distant from the FZB 42 replication origin (Table 2). The question of whether this probable regulator, a member of the TetR family, does regulate in *trans* expression of the *pks1* gene cluster remains open to investigation. Gene clusters *pks2* and *pks3* contain upstream regulatory elements resembling transcriptional regulators. Notably, the *pks3* gene cluster is preceded by two genes encoding a putative LysR-like transcriptional regulator and an antiterminator protein (Table 2). None of the three clusters encodes PKS modules harboring AT domains that are a feature of most other known modular PKSs. Instead, discrete AT genes were detected upstream of the megasynthase open reading frames (Table 2). This architecture, which has been recently described as *trans*-AT PKS (34), is also known from a smaller number of other systems, such as the pederin and onnamide PKSs from bacterial endosymbionts of beetles (34) and sponges (36), respectively; the leinamycin (*lnm*) PKS from *Streptomyces atroolivaceus* (7); the mupirocin PKS from *Pseudomonas fluorescens* (11); the myxovirescin (TA) PKS from *Myxococcus xanthus* (30); and others (4, 19, 40). Experimental evidence presented for the LnmG AT suggested that the discrete ATs act iteratively by loading malonyl coenzyme A (CoA) onto all PKS modules during polyketide synthesis (7). The unexpected presence of three members of this small PKS family in the *B. amyloliquefaciens* genome indicates that *trans*-AT PKSs are much more common than previously anticipated.

Characterization of the polyketides produced by *B. amyloliquefaciens* strain FZB 42 by MALDI-TOF MS and HPLC-ESI MS. The DNA sequence data imply that strain FZB 42 has the biosynthetic machinery for the production of at least three different kinds of polyketides. Polyketide formation by FZB 42 was investigated by MALDI-TOF MS and HPLC-ESI MS, which yielded complementary results. MALDI-TOF MS measurements were performed with culture filtrates without further purification. Polyketides of *B. amyloliquefaciens* FZB 42 were investigated at three growth times. Cells were harvested at 12, 24, and 48 h. By means of MALDI-TOF MS, the highest polyketide yield was detected after cultivation for 12 h. Figure 2A shows MALDI-TOF mass spectra of wild-type strain FZB 42. Polyketide mass peaks were found in an *m/z* range of 450 to 650. These signals disappeared in culture filtrates of mutant CH3 (Δ*sfp*::*ermAM*), which is a derivative of FZB 42 with a disrupted *sfp* gene (Fig. 2B). We demonstrated that this mutant is deficient in polyketide production because in vivo Sfp protein functions as a 4′-phosphopantetheinyl transferase in the conversion of the apo into the holo form of both NRPSs and PKSs, as already verified in vitro (27).

In the MALDI-TOF mass spectra, several groups of related mass signals were observed that represent protonated, as well as mono- and dialkali, ion adducts (data not shown). The corresponding compounds were identified by comparison with MS data available for the polyketides of *B. subtilis* ATCC 39320 and *B. subtilis* 1A/3 (15, 51) and attributed to the polyene antibiotics bacillaene and difficidin (33, 51). Two biosynthetic variants of bacillaene were observed with molecular masses differing by 2 mass units ($[M + H]^+ = 581.3$ and 583.3). Apparently, in the latter compound one double bond of the polyene system of bacillaene is saturated. A characteristic feature in the MALDI-TOF mass spectra of this polyketide is the appearance of dehydrated species with $[M + H - 18]^+ = 563.3$ and 565.3 originating from the parent ions by the loss of H_2O.

Difficidin and its oxidized form oxydifficidin appeared mainly as their alkali ion adducts, whereas the intensity of the mass peaks of their protonated forms at *m/z* 545.3 and 561.3 are very low. Oxydifficidin bears a hydroxyl group at position 5 of the difficidin ring system (51) showing mass signals 16 mass units higher than those of the main metabolite. In particular, peaks were observed at *m/z* 583.3, 605.3, 621.3, and 637.3 which were attributed to mono- and dialkali ion adducts of these compounds (data not shown). Wilson et al. (51) showed that difficidin is phosphorylated. In agreement, alkali adducts of the dephosphorylated species were detected at *m/z* 487.4 and 503.3 (difficidin), as well as at *m/z* 503.4 and 519.4 (oxydifficidin). Presumably, the phosphate group is responsible for attachment of multiple alkali ions. This pattern of polyketide signals was the same as that found for strain ATCC 39320, which has been used by Wilson et al. (51) for isolation and structure analysis of the polyene antibiotic difficidin or oxydifficidin.

For HPLC-ESI MS analysis, lyophilization products of culture filtrates of wild-type *B. amyloliquefaciens* FZB 42 and the mutants derived therefrom were extracted with acetonitrile–0.1% formic acid. Extracts were submitted to analytical HPLC coupled with an ESI-QTrap mass spectrometer. The data obtained from these HPLC-ESI MS runs are summarized in Table 3. In positive mode, bacillaene variants with molecular masses of $[M + H]^+ = 581.5$ and 583.5, as well as their dehydrated species ($[M + H - 18]^+ = 563.5$ and 565.5), were found at retention times of 7.61 and 7.71 min, respectively. In addition to MS characterization, UV-visible light spectra were obtained by diode array detection, which clearly confirmed the identities of bacillaenes compared to UV-visible light spectra published previously (33). Difficidin and oxydifficidin could only be detected by ESI MS in their deprotonated forms ($[M - H]^- = 543.4$ and 559.3) in the negative mode at

TABLE 2. Organization of the three polyketide biosynthesis gene clusters

Gene	Product size (no. of amino acids)	Proposed function[a]	Closest homolog (protein, origin)	Protein similarity/ identity (%)
Distant from *pks* gene clusters				
baeA	221	TetR transcription regulator	PksA, BSU17080	66/84
ymcC	183	Hypothetical protein	YmcC, BSU17070	48/61
Gene cluster *pks1* (*bae*), involved in bacillaene biosynthesis (AJ634060)				
baeB	225	3-Hydroxy-3-methylglutaryl-CoA synthase	PksB, BSU17090	64/79
baeC	289	Malonyl-CoA-ACP transacylase	PksC, BSU17100	68/82
baeD	324	Malonyl-CoA-ACP transacylase	PksD, BSU17110	52/65
baeE	746	Malonyl-CoA-ACP transacylase/ FMN-dependent oxidoreductase	PksE, BSU17120	69/80
baeF	82	ACP	AcpK, BSU17130	69/86
baeG	420	3-Hydroxy-3-methylglutaryl-CoA synthase	PksG, BSU17150	83/91
baeH	257	Enoyl-CoA hydratase	PksH, BSU17160	60/74
baeI	249	Enoyl-CoA hydratase	PksI, BSU17170	75/87
baeJ	4,982	NRPS/PKS domains: CL, PCP, C, A, PCP/KS, DH KR, ACP, ACP, KS, KR, ACP, (KS)[b]	PksJ, BSU17180	61/74
baeL	4,475	PKS domains: DH, ACP, KS, KR, ACP, KS, ACP, ACP, KS, KR, ACP, (KS)	PksK, BSU17190	60/73
baeM	3,511	PKS domains: DH, ACP, (KS), KR, ACP, KS, KR, ACP	PksM, BSU17200	60/74
baeN	5,433	NRPS/PKS domains: C, A, PCP, KS, DH, KR, ACP, KS, DH, KR, ACP, KS, DH, KR	PksN, BSU17210	61/74
baeR	2,482	PKS domains: MT, ACP, (KS), ACP, KS, ACP, TE	PksR, BSU17220	56/71
baeS	415	P450-like hydroxylase	PksS, BSU17230	73/83
Gene cluster *pks2*, involved in biosynthesis of putative polyketide 2 (AJ634061)				
1435[c]	212	Putative transcription factor	YkyA, BSU14570	32/52
1436	55	Unknown protein, phage related		
pks2A	768	Malonyl-CoA-ACP transacylase/ FMN-dependent oxidoreductase	PksE, BSU17120	47/66
pks2B	4,086	PKS domains: KS, KR, ACP, KS, DH, KR, ACP, ACP, KS, DH	AAV97877 OnnI, *Theonella swinhoei*-symbiotic bacterium	34/52
pks2C	1,590	PKS domains: KR, ACP, KS, DH	PksL, BSU17190	43/60
pks2D	2,912	PKS domains: KR, ACP, KS, KR, ACP, KS	YP_111013 Pks, *B. pseudomallei*	39/53
pks2E	2,334	PKS domains: DH, KR, ACP, KS, KR, ACP, KS	PksJ, BSU17180	34/52
pks2F	1,903	PKS domains: DH, ACP, ACP, KR, KS	PksL, BSU17190	50/64
pks2G	2,460	PKS domains: KR, ACP, KS, KR, ACP, KS	PksL, BSU17190	42/59
pks2H	1,287	PKS domains: DH, ACP, ACP, KR, TE	PksN, BSU17210	31/54
pks2I	375	PBP-related β-lactamase	CAB49062	32/50
Gene cluster *pks3* (*dif*), involved in difficidin-oxydifficidin biosynthesis (AJ634062)				
2264	107	Unknown protein		
2263	292	LysR transcription regulator	AAP10132, *B. cereus*	35/54
2262	176	Transcription antiterminator	AAM23992, *Thermoanaerobacter*	32/52
difA	791	Malonyl-CoA-ACP transacylase/ FMN-dependent oxidoreductase	PksE, BSU17120	58/75
difB	326	Hypothetical bacterial protein kinase	ZP_00829240, *Yersinia frederiksenii*	30/50
difC	90	Putative ACP	AAM12933 Macp15, *P. fluorescens*	47/62
difD	454	Acyl-CoA synthetase (AMP forming/ AMP acid ligase II)	AAM12931 MupQ, *P. fluorescens*	43/62
difE	245	3-Oxoacyl-ACP reductase	AAM12932 MupS *P. fluorescens*	42/64

Continued on following page

TABLE 2—*Continued*

Gene	Product size (no. of amino acids)	Proposed function[a]	Closest homolog (protein, origin)	Protein similarity/ identity (%)
difD	454	Acyl-CoA synthetase (AMP forming/ AMP acid ligase II)	AAM12931 MupQ, *P. fluorescens*	43/62
difE	245	3-Oxoacyl-ACP reductase	AAM12932 MupS *P. fluorescens*	42/64
difF	4,196	PKS domains: ACP, KS, DH, KR, MT, ACP, ACP, KS, DH, KR, ACP, KS	PksN, BSU17210	44/61
difG	2,098	PKS domains: (DH), ACP, KS, KR, ACP, (KS)	PksJ, BSU17180	46/60
difH	1,917	PKS domains: DH, ACP, KS, DH, KR	PksL, BSU17190	58/72
difI	5,204	PKS domains: MT, ACP, ACP, KS, DH, KR, ACP, KS, DH, KR, ACP, KS, KR, ACP, KS	AAS47562 PedH, *P. fuscipes*-symbiotic bacterium	38/55
difJ	2,572	PKS domains: DH, ACP, KR, KS, KR, ACP, (KS)	PksJ, BSU17180	37/55
difK	2,050	PKS domains: DH, DH, KS, DH, KR	PksN, BSU17210	41/58
difL	2,071	PKS domains: MT, ACP, ER, KS, ACP, ACP, TE	AAM12913 MmpIV, *P. fluorescens*	49/62
difM	384	P450 monooxygenase	YP_077930 P450, *B. licheniformis*	38/59
difN	415	3-Hydroxy-3-methylglutaryl-CoA synthase	PksG, BSU17150	72/84
difO	248	Enoyl-CoA hydratase	PksI, BSU17170	56/72

[a] Abbreviations: A, adenylation; C, condensation; CL, acyl-CoA ligase with unknown specificity; ER, enoyl reductase; MT, methyltransferase; PCP, peptidyl carrier protein; TE, thioesterase; PBP, penicillin-binding protein; FMN, flavin adenine dinucleotide.
[b] KS and DH domains presumably nonfunctional are in parentheses.
[c] Gene number according to the *B. amyloliquefaciens* genome sequencing project.

retention times of 10.91 and 8.54 min, respectively. The MS data obtained by MALDI-TOF MS (Fig. 2) and HPLC-ESI MS (Table 3) and the bioautographs performed with enriched culture filtrates of wild-type FZB 42 (Fig. 3) indicate that difficidin or oxydifficidin represents the main antibacterial compound of this strain.

Assignment of biological functions to polyketide gene clusters on the basis of gene disruption. Disruptions of all three *pks* target genes were carried out by insertion of an antibiotic resistance cassette via homologous recombination (see Materials and Methods). Successful inactivation of gene clusters was proven by MS analysis and bioautography of mutants bearing gene replacement mutations in the respective first KS domains of *pks1* (CH6 [Δ*pks1KS1*::*cat*]), *pks2* (CH7 [Δ*pks2KS1*::*cat*]), and *pks3* (CH8 [Δ*pks3KS1*::*ermAM*]). Extracted ion chromatograms of HPLC-ESI MS runs (Table 3 and data not shown) and MALDI-TOF mass spectra (data not shown) from *B. amyloliquefaciens* FZB 42 and the respective mutant strains, including double-knockout mutants CH11 (Δ*pks1KS1*::*cat* Δ*pks3KS1*::*ermAM*), CH12 (Δ*pks2KS1*::*cat* Δ*pks3KS1*::*ermAM*), and CH14 (Δ*pks1KS1*::*cat* Δ*pks2KS1*::*neo*), yielded complementary results. Mutant CH6 (Δ*pks1KS1*::*cat*) harbored a mutation of the *pks1* system, and as a consequence the bacillaene-specific molecular mass peaks ($[M + H]^+ = 581.3$ and 583.3 and $[M + H - 18]^+ = 563.3$ and 565.3) of the dehydrated species are absent in the MALDI-TOF mass spectrum compared to the wild-type mass spectra. In contrast, difficidin or oxydifficidin biosynthesis remains unaffected, which is shown by signals in HPLC-ESI MS in negative mode ($[M - H]^- = 543.4$ and 559.3). MALDI-TOF mass spectra yield $m/z = 605.3$ and 621.3, apparently indicating adducts with two alkali ions. Mutant CH7 (Δ*pks2KS1*::*cat*), with a gene disruption in the *pks2* cluster, showed the same pattern of bacillaene- and difficidin- or oxydifficidin-related molecular masses as found previously for the wild-type strain. However, we did not identify any molecular mass that could be related to a putative *pks2* product; for mutant CH8 (Δ*pks3KS1*::*ermAM*), only bacillaene-specific molecular masses were detected. CH11 (Δ*pks1KS1*::*cat* Δ*pks3KS1*::*ermAM*) is a double mutant with gene disruptions in both the *pks1* and *pks3* clusters. For this strain, no molecular masses were identified that would indicate the presence of polyketides. Both the bacillaene- and difficidin- or oxydifficidin-related peaks were absent. In CH12 (Δ*pks2KS1*::*cat* Δ*pks3KS1*::*ermAM*), *pks2* and *pks3* were inactivated and only the molecular masses characteristic for bacillaene were observed. Finally, CH14 (Δ*pks1KS1*::*cat* Δ*pks2KS1*::*neo*), bearing mutations in the *pks1* and *pks2* gene clusters exclusively, showed the difficidin- or oxydifficidin-specific molecular mass peaks (data not shown).

On the basis of these results, also confirmed by HPLC-ESI MS (Table 3), the *pks* regions in the genome of *B. amyloliquefaciens* FZB 42 were assigned as follows. *pks3* codes for the production of difficidin or oxydifficidin and is designated the *dif* gene cluster, while *pks1* is responsible for the biosynthesis of bacillaene and is designated the *bae* gene cluster. Since *pks1* is an ortholog of the *pksX* system in *B. subtilis* 168 (more details are given below), it was challenging to confirm unambiguously that bacillaene is the natural product of this silent gene cluster. To this end, *B. subtilis* OKB105 (*pheA* sfp^+), known as a potent producer of the lipopeptide surfactin (27), was tested for the ability to produce bacillaene. Bioautographs performed with *B. subtilis* JH642 (*pheA trpC2* sfp^0) and its derivative OKB105 demonstrated that the presence of a functional *sfp* gene enabled *B. subtilis* to produce that polyketide (Fig. 3). The function of *pks2* in *B. amyloliquefaciens* is less clear insofar as no *pks2*-specific compound was detected by HPLC-ESI MS or

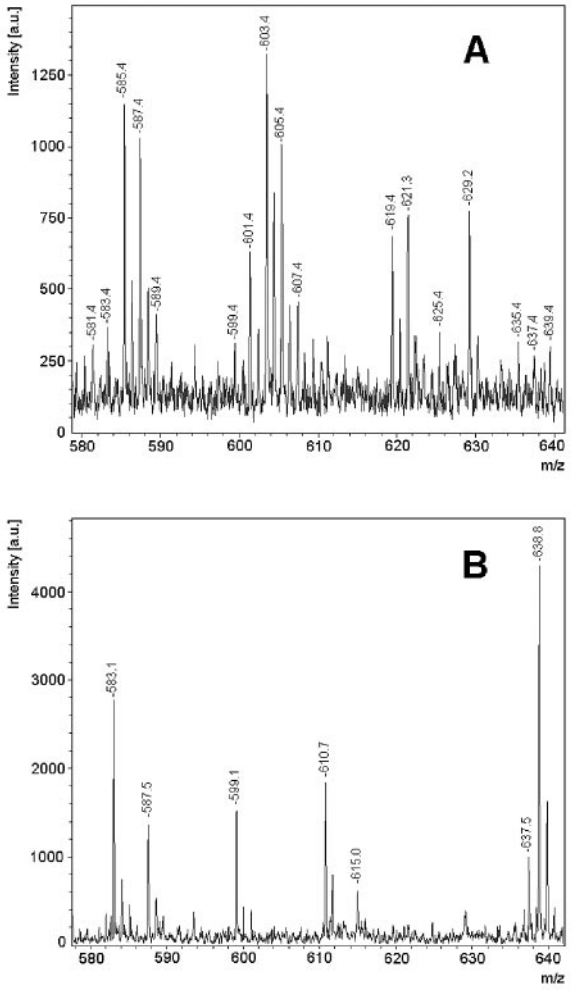

FIG. 2. MALDI-TOF mass spectra of polyketide compounds produced by FZB 42 and mutant CH3 after 12 h of cultivation in Landy medium (24; see Materials and Methods). For MS analysis, 1- to 2 μl portions of lyophilized culture filtrates were used. Experimental details are described in Materials and Methods. (A) MALDI-TOF mass spectra of polyketides from *B. amyloliquefaciens* FZB 42 culture filtrates in an *m/z* range of 450 to 650. (B) MALDI-TOF mass spectrum of culture filtrates from Sfp$^-$ mutant CH3 (Δ*sfp*::*ermAM*), which is deficient in polyketide production. a.u., arbitrary units.

MALDI-TOF MS. However, in bioautographs (Fig. 3) we observed the disappearance in the *pks2* mutant CH7 and CH12 supernatants of a weak antibiotic activity, suggesting that this gene cluster might be responsible for the synthesis of an unknown polyketide (tentatively named PK2) that is weakly expressed and/or has low bioactivity.

Remarkably, the growth of phytopathogenic bacteria such as *E. carotovora* was strongly suppressed by difficidins and bacillaene as the most active agents to suppress bacterial growth. As mentioned above, *sfp* mutation blocked nearly the entire antibacterial activity. However, bioautographs demonstrated that an antibacterial compound with an unknown structure, putatively produced by conventional ribosomal synthesis, remained present in lipopeptide- and polyketide-deficient mutants (Fig. 3).

Analysis of PKS architecture. (i) Modular organization of *Bacillus* PKSs is of unusual architecture. In most of the PKS systems studied so far, the modular architecture strictly reflects the order and structure of building blocks present in the polyketide chain. This feature has been described as the colinearity rule and is the underlying paradigm that should allow the rational redesign of PKS architectures in order to generate novel polyketides (18, 37). In recent years, however, a growing number of PKSs have been discovered that do not follow the classic biosynthetic rules (50). Domain analysis of the *B. amyloliquefaciens* PKSs reveals a situation similar to those systems described previously. In none of the three systems, for example, do the deduced PKS proteins feature the textbook architecture consisting of an N-terminal KS domain and a C-terminal ACP domain. Instead, modules with incomplete domain sets occupy the protein termini, suggesting that single modules are encoded on two distinct proteins. An example is the C-terminal KS residing on BaeL that presumably interacts with the remaining part of the module at the N terminus of BaeM in *trans*, yielding a complete PKS module (Table 2). A similar domain organization has recently been reported for the *lnm* biosynthetic gene cluster (45). A second notable feature is that several proteins exhibit unusual domain arrangements, such as ketoreductase (KR) or enoylreductase domains located behind ACP domains (DifJ and -L and Pks2F, -G, and -H) or tandem dehydratase (DH) domains preceding a KS domain (DifK).

The *pks1* (*bae*) gene cluster encodes a hybrid NRPS-PKS system. Only one of the *B. amyloliquefaciens* PKSs resembles

TABLE 3. Molecular masses of polyketide products of *B. amyloliquefaciens* FZB 42 and mutant strains detected by LC-ESI mass spectrometry[a]

Polyketide, retention time (min)	FZB 42	CH6 Δ*pks1KS1*::*cat*	CH7 Δ*pks2KS1*::*cat*	CH8 Δ*pks3KS1*::*ermAM*	CH12 Δ*pks2KS1*::*cat* Δ*pks3KS1*::*ermAM*	CH14 Δ*pks1KS1*::*cat* Δ*pks2KS1*::*neo*
Bacillaene A $[M + H]^+$, 7.61	581.5		581.5	581.5	581.5	
Bacillaene A $[M - H_2O + H]^+$, 7.61	563.5		563.5	563.5	563.5	
Bacillaene A $[M + Na]^+$, 7.61	603.5		603.5	603.5	603.5	
Bacillaene A $[M - H]^-$, 7.61	579.3		579.3	579.3	579.3	
Bacillaene B $[M + H]^+$, 7.71	583.5		583.5	583.5	583.5	
Bacillaene B $[M - H_2O + H]^+$, 7.71	565.5		565.5	565.5	565.5	
Bacillaene B $[M + Na]^+$, 7.71	605.5		605.5	605.5	605.5	
Bacillaene B $[M - H]^-$, 7.71	581.3		581.3	581.3	581.3	
Difficidin $[M - H]^-$, 10.91	543.4	543.4	543.4			543.4
Oxydifficidin $[M - H]^-$, 8.54	559.3	559.3	559.3			559.3

[a] Culture filtrates of strains CH3 (Δ*sfp*::*ermAM*) and CH11 (Δ*pks1KS1*::*cat* Δ*pks3KS1*::*ermAM*) did not contain molecular masses $[M - H] = 579.3$, $[M - H]^- = 543.4$, and $[M - H]^- = 559.3$, respectively. Molecular masses in daltons are shown.

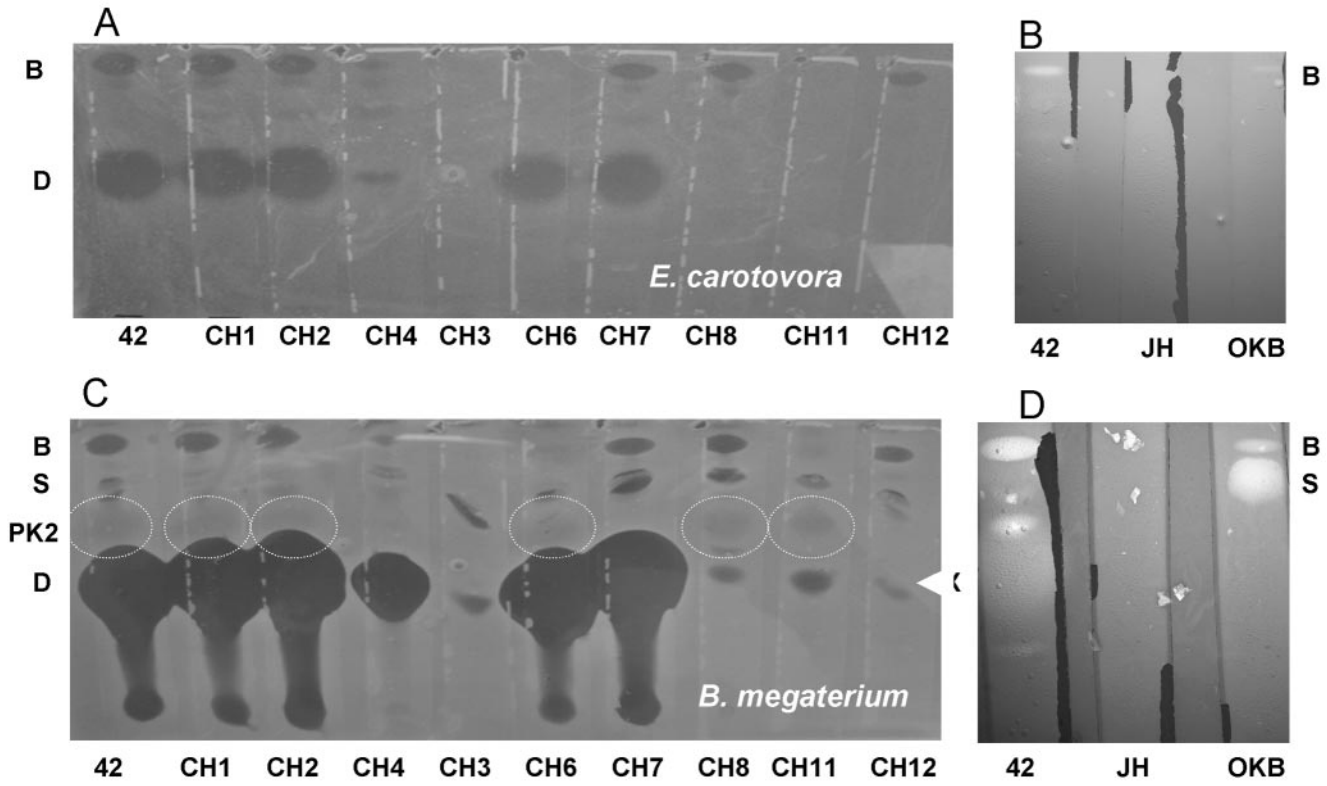

FIG. 3. Survey of antibacterially acting metabolites through bioautography. Supernatants prepared from strain FZB 42 and mutant strains (A, C) and *B. subtilis* JH642 (*sfp*^{0}) and OKB105 (*sfp*$^+$) (B, D) grown for 12 h in Landy medium (24) were separated by TLC and sandwiched with indicator strains *E. carotovora* (top) and *B. megaterium* (bottom) as described in Materials and Methods. Inhibition zones with no or reduced growth are indicative of antibiotic activities. Abbreviations: B, bacillaene synthesized by gene cluster *pks1* (*bae*); S, surfactin synthesized by the *srf* gene cluster (20); PK2, unknown polyketide synthesized by gene cluster *pks2*; D, difficidin or oxydifficidin synthesized by gene cluster *pks3* (*dif*); X, unknown metabolite with antibacterial activity not dependent on phosphopantetheine transfer catalyzed by the *sfp* gene product. The weak antibacterial activity exerted by gene cluster *pks2* is indicated by circles. Strain CH1 (Δ*srfAA*::*ermAM*) is deficient in surfactin synthesis (20). Strain CH2 (Δ*fen*::Cm^r Δ*srfAA*::*ermAM*) was obtained by transformation of strain AK2 (20) with chromosomal DNA isolated from strain CH1. Strain CH4 harbors a knockout mutation of the *yczE* gene (23) and will be described elsewhere. The other mutant strains used for bioautography were CH3 (Δ*sfp*::*ermAM*), CH6 (Δ*pks1KS1*::*cat*), CH7 (Δ*pks2KS1*::*cat*), CH8 (Δ*pks3KS1*::*ermAM*), CH11 (Δ*pks1KS1*::*cat* Δ*pks3KS1*::*ermAM*), CH12 (Δ*pks1KS1*::*cat* Δ*pks3KS1*::*ermAM*), JH (JH642 *pheA trpC2 sfp*^{0}), and OKB (OKB105 *pheA sfp*^{0}).

systems identified earlier. The *bae* gene cluster exhibits striking architectural and sequence similarity to the *pksX* gene cluster of *B. subtilis* 168, which clearly indicates that it is its ortholog (Fig. 1). The core region of both gene clusters differs from that of gene clusters *pks2* and *dif* by the occurrence of two hybrid NRPS-PKS orthologs (*pksJ*/*baeJ* and *pksN*/*baeN*, Table 2). The N-terminal part of PksJ and its ortholog BaeJ consists of a loading module for an unknown starter unit, while the subsequent module displays features typical of NRPS with consecutive condensation, amino acyl adenylation, and peptidyl carrier protein domains. Downstream of the NRPS module are two *trans*-AT PKS modules. A second hybrid NRPS-PKS protein, PksN (BaeN), contains in its N-terminal part a complete NRPS-type module, followed by a total of three further PKS modules in which the last one is incomplete, consisting only of the KS and KR domains.

Although with bacillaene for the first time a secondary metabolite can be attributed to this gene cluster, the sparse available chemical data do not allow clear assignment of gene functions to specific synthesis steps of a molecule with an as-yet-unknown structure. It has been reported that bacillaene is a conjugated hexaene with a molecular formula of $C_{35}H_{48}O_7$, as deduced from MS data (33). However, the presence of two NRPS modules in the *bae* cluster is not in accordance with this formula, since the incorporation of amino acids would require the presence of at least two nitrogen atoms. In two independent studies, Stachelhaus, Challis, and coworkers (6, 43) have identified amino acid residues within the adenylation domain of NRPSs that allow deduction of the substrate specificity of NRPS modules. Therefore, in order to predict the amino acids incorporated into the bacillaene structure, the nonribosomal code was extracted from the two *bae* adenylation domains (Table 4). For the first NRPS module of BaeJ, the analysis gave 100% identity with the glycine-specific code (31, 46) while no similarity to known codes was found for the other BaeN module. These data give rise to the assumption that bacillaene should contain at least one glycine residue.

Model for difficidin or oxydifficidin synthesis. Similar to the *bae* genes, the *dif* system exhibits an unusual architecture that sets it apart from most other PKSs. Nevertheless, a plausible biosynthetic pathway can be proposed for difficidin which accounts for the presence of most of the features of the *dif* system

TABLE 4. Specificity-conferring code of A domains residing in hybrid NPRS/KS of the *pks1* (*bae*) and *pksX* biosynthesis gene clusters

Module (reference[s])	Residue[a] at position:										Code
	235	236	239	278	299	301	322	330	331	517	
Gly consensus (9, 43)	**D**	**I**	**L**	**Q**	**L**	**G**	L/M	**I**	**W**	**K**	Gly
BaeJ-A	**D**	**I**	**L**	**Q**	**L**	**G**	M	**I**	**W**	**K**	Gly
PksJ-A (23)	**D**	**I**	**L**	**Q**	**L**	**G**	M	**I**	**W**	**K**	Gly
Ta1-A (31)	**D**	**I**	**L**	**Q**	**L**	**G**	M	**I**	**W**	**K**	Gly
OnnI-A (36)	**D**	**I**	**L**	**Q**	**L**	**G**	L	**I**	**W**	**K**	Gly
PedF-A (34)	**D**	**I**	**L**	**Q**	**L**	**G**	L	**I**	**W**	**K**	Gly
BaeN-A	**D**	V	S	N	M	A	I	**I**	Y	**K**	ND[b]
PksN-A (23)	**D**	V	S	N	M	A	I	**I**	Y	**K**	ND

[a] Bold letters indicate strictly conserved residues.
[b] ND, unknown.

(Fig. 4). The starting point would be a C_3 precursor, such as pyruvate, which is converted to an ACP-bound acrylyl moiety by reduction (perhaps catalyzed by DifE), dehydration, and transfer by the acyl-CoA ligase DifD to the N-terminal loading ACP of DifF. The incorporation of pyruvate and other α-oxygenated three-carbon starter units into polyketides is not unprecedented (14, 26); however, to our knowledge an acrylate starter moiety has not yet been reported. The subsequent biosynthetic steps that extend the starter unit can be largely rationalized by the colinearity rule, with several notable deviations. A number of modules lack domains that would be predicted from the polyketide structure. Those are a KR domain in module 3, two DH domains in modules 4 and 9, and two enoylreductase domains in modules 2 and 8. It is possible that these activities are provided by other proteins acting in *trans*. Such a complementation has also been suggested for other polyketides, such as chalcomycin (49) and epothilone (46). Thus, in order to explain the presence of a double bond in epothilone, Tang and coworkers proposed a mechanism consisting of module skipping, use of a DH domain of a downstream module, and subsequent iterative use of the same module for the next round of elongation (46). However, in difficidin biosynthesis the architecture of most downstream modules would not allow such a scenario. Further studies are necessary to determine whether elements within the *dif* system are responsible for complementation or whether the activities are provided from somewhere else in the genome, for example, the *bae* or *pks2* system. Another deviation from the colinearity rule is the presence of two superfluous modules, module 5, jointly encoded by *difG* and *difH*, and module 11, a product of *difJ* and *difK*, which are obviously skipped during biosynthesis. Closer

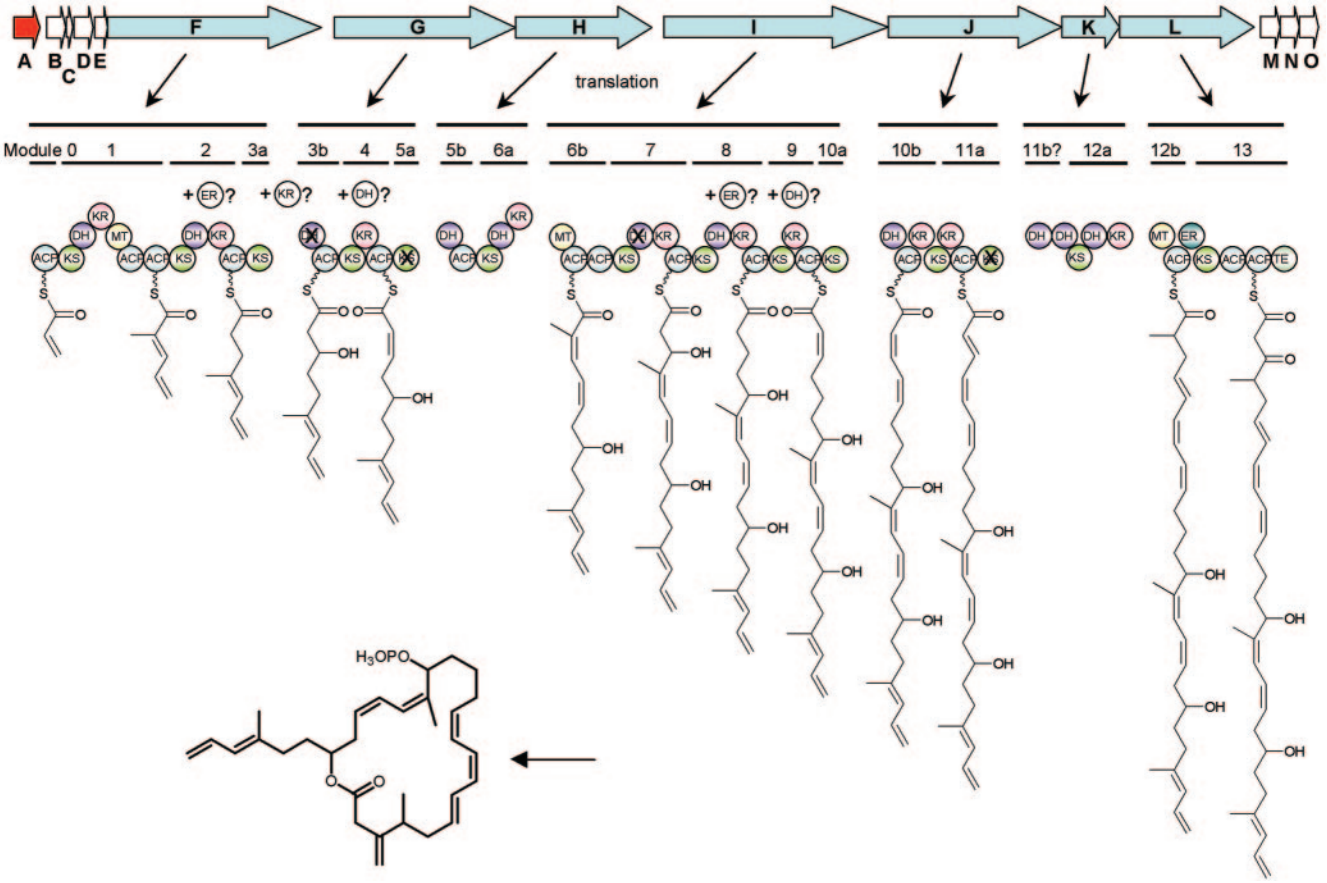

FIG. 4. Model for difficidin biosynthesis. Domains predicted to be inactive are crossed out. Some activities are predicted to be provided in *trans* by unknown enzymes, which are shown above some of the modules. Domain abbreviations not defined in the text: ER, enoylreductase; MT, methyltransferase.

inspection of the KS domains, which are an essential part of functional PKS modules, revealed that most of them contained the functional residues characterized as necessary for their catalytic activity, i.e., C163, D306, E309, H298, K328, and H333 (numbering according to *E. coli* type I β-ketoacyl-ACP synthase) (25, 52). However, the KSs of modules 5 and 11 both feature an H298 → N substitution and should therefore be inactive, since they would be unable to catalyze the decarboxylation of malonyl units. Similar mutations have been previously reported for the pederin, onnamide, and mupirocin systems and are also present in the *bae* and *pksX* clusters. In addition, we were unable to identify an ACP located inside module 11 of the *dif* PKS, indicating that this module might be a partially decayed evolutionary remnant.

An interesting feature of the difficidin structure is the occurrence of four double bonds with a Z configuration, a geometry that is very rare among polyketides (2, 32, 41, 49). It has been proposed that an aspartate residue in the KR domain plays a crucial role in defining the stereochemistry of the ketoreduction step, which in turn would influence the double-bond configuration after dehydration by the downstream DH domain (38). An alignment of *dif* KR domains revealed an agreement with this model for the synthetic assembly of most of the double bonds except for the KRs of DifJ. These KRs contain sequence features that would lead to a prediction of the respective opposite configurations. Similar deviations from the Asp rule have been observed previously for the phoslactomycins, which are products of a *cis*-AT PKS (32), suggesting the need for further refinement of the model.

The likely final steps in difficidin biosynthesis are the conversion of a keto function into an exomethylene group, subsequent cyclization by the thioesterase domain, and attachment of the rare phosphate moiety to the hydroxyl group at C-15. The formation of exomethylene or other carbon-containing functionalities from keto groups has been described for a limited number of other polyketides, such as pederin-type compounds (34, 36), myxovirescin (31), and the jamaicamides (10). It has been proposed that these groups are generated by joint action of a 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) and an enoyl-CoA hydratase (EH) homolog, as well as free-standing KS and ACP units, which are invariably found encoded within the corresponding gene clusters (10, 36). We identified proteins with similarity to HMGSs, EHs, and ACPs encoded by *difN*, *difO*, and *difC*, respectively, but no individual KS gene. A candidate for the likely final step in difficidin biosynthesis, phosphorylation, is DifB, which resembles (30% identity and 50% similarity) a hypothetical bacterial protein kinase (accession no. ZP_00829240) and contains typical kinase-like sequence profiles (smart00220, COG0661).

Phylogenetic analysis of *pks* gene clusters. While the *bae* gene cluster harbors genes that are clearly orthologs of the *B. subtilis* 168 *pksX* operon (Fig. 1), favoring the idea that this cluster is part of the *B. subtilis* or *B. amyloliquefaciens* core region, the origin of the additional *pks2* and *dif* gene clusters is less obvious. One possibility is that *pks2* and *bae* might have been acquired by horizontal gene transfer from other soil bacteria. Alternatively, *pks2* and *dif* might have evolved from an ancestral *pks* operon by several gene duplication events, or—more likely—by both events. Recently, it has been shown that the operon encoding pederin biosynthesis in the bacterial symbiont of *Paederus fuscipes* beetles is part of a genomic island characterized by several decayed insertion sequence elements (35). However, no genetic features typical of horizontally acquired DNA were identified in the *pks* regions of *B. amyloliquefaciens*. The G+C contents of *bae* (49.47%), *pks2* (43.39%), and *dif* (49.48%) were not clearly distinct from the average G+C content of the whole genome of 46.2% (Chen and Borriss, unpublished). Moreover, computational analysis of the local OUV (39) of the *pks* gene clusters did not reveal an island-specific OU architecture. Local deviations in the tetranucleotide pattern from the signature of the whole FZB 42 genome were only 2.47 for *bae* and 1.91 for *dif*. A slightly higher value was determined for *pks2* (3.18). Moreover, features indicating events of horizontal gene transfer such as remnants of phage-, insertion sequence element-, or transposase-like sequences were not detected, except for a phage-like sequence residing at the flanking region of the *pks2* gene cluster (Table 2).

Sequence alignments of the discrete ATs of all of the *pks* gene clusters revealed similarities to the probably functionally related PksC, PksD, and PksE ATs of *B. subtilis* 168 (23); the PedC and PedD ATs involved in pederin biosynthesis (34); the MmpIII AT1/AT2 of *P. fluorescens* involved in mupirocin biosynthesis (11); and the LnmG AT of *S. atroolivaceus*, an iteratively acting AT in leinamycin biosynthesis (7). A phylogenetic tree constructed for this group of discrete malonyl-CoA-specific ATs revealed that the hybrid AT-oxidoreductase (AT-OR) proteins present in all hitherto known *Bacillus* PKS gene clusters form, together with the mono domain ATs PksC (*pksX*) and BaeC, a distinct clade not including other members of the family of discrete ATs (Fig. 5A). Therefore, we hypothesize that the discrete AT-ORs have evolved from a PksE-like ancestor protein residing in ancient bacilli. Gene duplication events might have generated the paralogous mono domain ATs PksC and BaeC. In contrast, the third pair of discrete ATs, PksD and BaeD, might have evolved independently.

In order to analyze PKS evolution in more detail, we used a similar phylogenetic approach using all of the KS domains extracted from the *Bacillus* PKSs and hybrid NRPS-PKSs and their closest relatives identified by BLASTP comparison. In total, 100 KS sequences extracted from *trans*-AT PKS operons were included in this phylogenetic analysis. At least five clusters consisting of paralogous *Bacillus* KS domains were identified. Notably, a distinct clade (group 2) is formed by a group of KS domains located in hybrid *Bacillus* NRPS-PKS proteins adjacent to the NRPS modules. As expected, the shortest evolutionary distances were detected between the corresponding KS domains derived from *bae* and *pksX*, but several KSs derived from the *dif* operon were also involved in clades formed by related members of the *Bacillus pks* gene clusters (Fig. 5B). Especially striking was the close relatedness between domain KS-6 residing at DifH and the corresponding KS domains residing at PksL (*pksX*) and BaeL (76% identity, 87% similarity), for example, suggesting that at least parts of the bacillaene and difficidin biosynthesis gene clusters have the same origin. It is tempting to speculate that processes of evolution of novel *pks* operons were completed by several events of homologous recombination leading to gene duplications. Because of the presence of sequences with repeated homology located at distinct sites of the modularly organized PKSs, such events are considered to be very likely. Homologies of 67% identity and 81% similarity detected between domains KS-2 and KS-3 re-

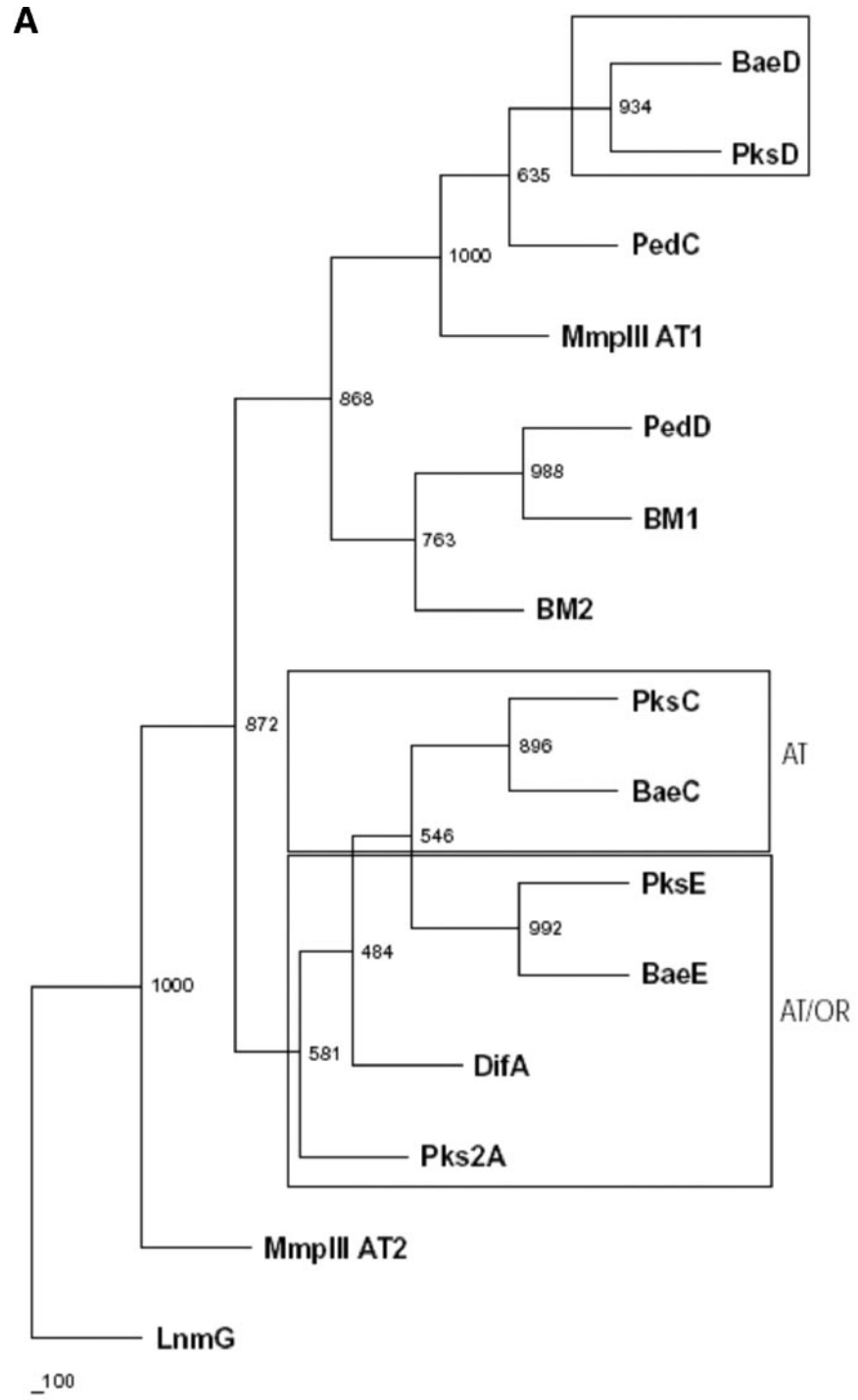

FIG. 5. Phylogenetic tree analyses of discrete ATs and integrated KS domains. Bootstrap values (n = 1,000) for branches are given. The scale bar represents the number of substitutions. (A) Phylogenetic tree analysis of discrete ATs occurring in bacterial *trans*-AT *pks* gene clusters. *Bacillus* ATs are boxed. AT, monodomain AT; AT/OR, bifunctional hybrid of AT and OR domains. (B) Phylogenetic tree analysis of *Bacillus* KS domains and their closest orthologs occurring in other modular AT-less type I PKSs. Groups 1 to 5 of related *Bacillus* KS domains are framed. Group 2 is exclusively formed by KS domains located in hybrid NRPS-PKS proteins.

siding at DifF with their counterparts KS-7 and KS-9 residing at DifI support such an assumption. Gene duplication events are particularly visible in gene cluster *pks2*, where many of the KS domains appeared as a separate branch of the phylogenetic tree and a high degree of similarity between KS domains residing at distinct proteins was detected. Similarity between the *Bacillus pks* gene clusters is also apparent in tailoring enzymes, e.g., between the adjacent HMGS- and EH-encoding genes

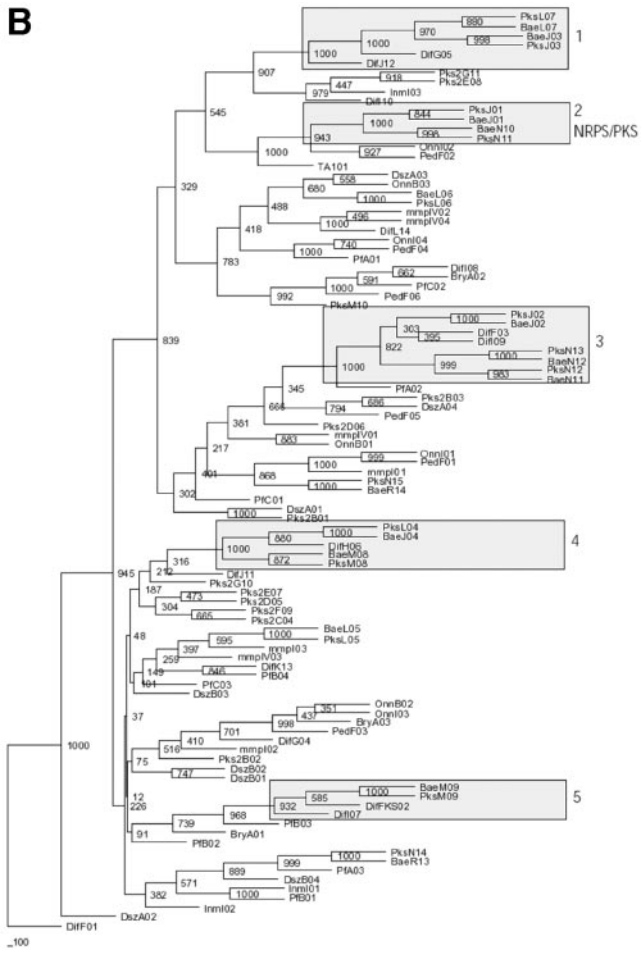

FIG. 5—*Continued.*

present in the *pksX* (*pksGHI*), *bae* (*baeGHI*), and *dif* (*difNO*) gene clusters (Table 2). While these genes are arranged colinearly within the *pksX* and *bae* gene cluster, they have been reshuffled to a position downstream from the PKS-encoding genes within the *dif* gene cluster (Fig. 1).

Taken together, these results support the idea that the *bae* and *dif* gene clusters might have the same origin but also several gene duplication events might have contributed to the evolution of the recent biosynthesis gene clusters. Lower similarity exists between *pks2* and the other *Bacillus* gene clusters. Despite its different G+C content, its different OUV, and the presence of an adjacent phage-like sequence, we cannot exclude the possibility that major parts of the *pks2* operon have been acquired by horizontal gene transfer. The greater evolutionary distances between the *pks2* ATs and KSs within the phylogenetic trees constructed from type I *trans*-AT PKS systems support this notion.

Conclusions. The antibacterial polyketides bacillaene and difficidin or oxydifficidin were identified in the culture filtrate of *B. amyloliquefaciens* FZB 42 by MALDI-TOF MS and HPLC-ESI MS. It has been previously suggested that biosynthesis of those polyketides should be accomplished by giant operons similar to that of the inactive *pksX* operon in *B. subtilis* 168. Here we present clear genetic evidence that the *pks1* and *pks3* gene clusters encoding modular type I PKSs of the *trans*-AT architectural group are responsible for the biosyntheses of bacillaene and difficidin, respectively, while the type I PKS of the *pks2* gene cluster seems to be involved in the synthesis of an as-yet-unknown polyketide with weak antibacterial activity. Bacillaene and difficidin are the main antibacterial secondary metabolites of environmental *B. amyloliquefaciens* FZB 42. We have defined for the first time the complete gene clusters involved in their synthesis and have presented a model for the synthesis of difficidin based on the structure of the *pks3* (*dif*) gene cluster. We suggest that the impressive biosynthetic capacity of strain FZB 42 to produce antagonistically acting polyketides and lipopeptides enables FZB 42 to cope successfully with competing organisms, e.g., phytopathogenic bacteria and fungi, within its natural environment. Structural similarity detected between gene clusters *pks1*(*bae*) and *pks3* (*dif*), which are involved in the synthesis of two different polyketides, should inspire future experiments in which swapping of domains occurring in the distinct but related gene clusters of *B. amyloliquefaciens* might allow study of their unusual enzymologies and generation of a variety of novel bioactive polyketides.

ACKNOWLEDGMENTS

This work was supported by funds from the competence network Genome Research on Bacteria (GenoMik), financed by the German Ministry for Education and Research (R.B. and G.G.), and by grants from the Lower Saxony Ministry of Science and Culture (G.G.). The work of K.S. and R.D.S. was supported by an Emmy-Noether fellowship from the DFG (SU 239/2).

An anonymous referee is thanked for the suggestion to investigate the bacillaene production of OKB105. We thank Mohammed Marahiel for supplying strain OKB105. Oleg Reva is thanked for introducing us to OUV analysis of genomic regions.

REFERENCES

1. **Albertini, A. M., T. Caramori, F. Scoffone, C. Scotti, and A. T. Galizzi.** 1995. Sequence around the 159° region of the *Bacillus subtilis* genome: the pksX locus spans 33.6 kb. Microbiology **141:**299–309.
2. **August, P. R., L. Tang, Y. J. Yoon, S. Ning, R. Muller, T. W. Yu, M. Taylor, D. Hoffmann, C. G. Kim, X. Zhang, C. R. Hutchinson, and H. G. Floss.** 1998. Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. Chem. Biol. **5:**69–79.
3. **Cane, D. E.** 1997. A special thematic issue on polyketide and nonribosomal polypeptide biosynthesis. Chem. Rev. **97:**2463–2706.
4. **Carvalho, R., R. Reid, N. Viswanathan, H. Gramajo, and B. Julien.** 2005. The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation. Gene **359:**91–97.
5. **Ceglowski, P., and J. C. Alonso.** 1994. Gene organization of the Streptococcus pyogenes plasmid pDB101: sequence analysis of the ORF eta-copS region. Gene **145:**33–39.
6. **Challis, G. L., J. Ravel, and C. A. Townsend.** 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. **7:**211–224.
7. **Cheng, Y.-Q., G.-L. Tang, and B. Shen.** 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proc. Natl. Acad. Sci. USA **100:**3149–3154.
8. **Cutting, S. M., and P. B. van der Horn.** 1990. Genetic analysis, p. 27–74. *In* C. R. Harwood and S. M. Cutting (ed.), Molecular biological methods for *Bacillus*. Wiley-Interscience, Chichester, United Kingdom.
9. **Du, L., C. Sanchez, M. Chen, D. J. Edwards, and B. Shen.** 2000. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC 15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. Chem. Biol. **7:**623–642.
10. **Edwards, D. J., B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts, and W. H. Gerwick.** 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. Chem. Biol. **11:**817–833.
11. **El-Sayed, A. K., J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson, and C. M. Thomas.** 2003. Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. Chem. Biol. **10:**419–430.

12. **Emmert, E. A. B., A. K. Klimowicz, M. G. Thomas, and J. Handelsman.** 2004. Genetics of zwittermicin A production by *Bacillus cereus*. Appl. Environ. Microbiol. **70:**104–113.
13. **Felsenstein, J.** 1989. PHYLIP: phylogeny inference package. Cladistics **5:**164–166.
14. **Hildebrand, M., L. E. Waggoner, H. Liu, S. Sudek, S. Allen, C. Anderson, D. H. Sherman, and M. Haygood.** 2004. *bryA*: an unsual polyketide synthase gene from the uncultivated bacterial symbiont of the bryozoan *Bugula neritina*. Chem. Biol. **11:**1543–1552.
15. **Hofemeister, J., B. Conrad, B. Adler, B. Hofemeister, J. Feesche, N. Kucheryava, G. Steinborn, P. Franke, N. Grammel, A. Zwintscher, F. Leenders, G. Hitzeroth, and J. Vater.** 2004. Genetic analysis of the biosynthesis of nonribosomal peptide- and polyketide-like antibiotics, iron uptake and biofilm formation by *Bacillus subtilis* A1/3. Mol. Genet. Genomics **272:**363–378.
16. **Idriss, E. E. S., H. Bochow, H. Ross, and R. Borriss.** 2004. Use of *Bacillus subtilis* as biocontrol agent. 6. Phytohormone like action of culture filtrates prepared from plant growth promoting *Bacillus amyloliquefaciens* FZB24, FZB 42, FZB45 and *Bacillus subtilis* FZB37. J. Plant Dis. Prot. **111:**583–597.
17. **Idriss, E. E. S., O. Makarewicz, A. Farouk, K. Rosner, R. Greiner, H. Bochow, T. Richter, and R. Borriss.** 2002. Extracellular phytase activity of *Bacillus amyloliquefaciens* FZB 45 contributes to its plant-growth-promoting effect. Microbiology **148:**2097–2109.
18. **Katz, L., and S. Donadio.** 1993. Polyketide synthesis: prospects for hybrid antibiotics. Annu. Rev. Microbiol. **47:**875–912.
19. **Kopp, M., H. Irschik, S. Pradella, and R. Muller.** 2005. Production of the tubulin destabilizer disorazol in Sorangium cellulosum: biosynthetic machinery and regulatory genes. ChemBioChem **6:**1277–1286.
20. **Koumoutsi, A., X.-H. Chen, A. Henne, H. Liesegang, G. Hitzeroth, P. Franke, J. Vater, and R. Borriss.** 2004. Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB 42. J. Bacteriol. **186:** 1084–1096.
21. **Krebs, B., B. Höding, S. M. Kübart, A. Workie, H. Junge, G. Schmiedeknecht, P. Grosch, H. Bochow, and M. Heves.** 1998. Use of *Bacillus subtilis* as biocontrol agent. 1. Activities and characterization of *Bacillus subtilis* strains. J. Plant Dis. Prot. **105:**181–197.
22. **Kunst, F., and G. Rapoport.** 1995. Salt stress is an environmental signal affecting degradative enzyme synthesis in *Bacillus subtilis*. J. Bacteriol. **177:** 2403–2407.
23. **Kunst, F., N. Ogasawara, I. Moszer, et al.** 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature **390:**249–256.
24. **Landy, M., G. H. Warren, S. B. Roseman, and L. G. Colio.** 1948. Bacillomycin, an antibiotic from *Bacillus subtilis* active against pathogenic fungi. Proc. Soc. Exp. Biol. Med. **67:**539–541.
25. **McGuire, K. A., M. Siggaard-Andersen, M. G. Bangera, J. G. Olsen, and P. von Wettstein-Knowles.** 2001. β-Ketoacyl-acyl carrier protein synthase I of *Escherichia coli*: aspects of the condensation mechanism revealed by analysis of mutations in the active site pocket. Biochemistry **40:**9836–9845.
26. **Mochizuki, S., K. Hiratsu, M. Suwa, T. Ishii, F. Sugino, K. Yamada, and H. Kinashi.** 2003. The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. Mol. Microbiol. **48:**1501–1510.
27. **Mootz, H. D., R. Finking, and M. A. Marahiel.** 2001. 4′-Phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. J. Biol. Chem. **276:**37289–37298.
28. **Nakano, M. M., M. M. Marahiel, and P. Zuber.** 1988. Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *Bacillus subtilis*. J. Bacteriol. **170:**5662–5668.
29. **Nakano, M. M., N. Corbell, J. Besson, and P. Zuber.** 1992. Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in Bacillus subtilis. Mol. Gen. Genet. **232:** 313–321.
30. **Paitan, Y., G. Alon, E. Orr, E. Z. Ron, and E. Rosenberg.** 1999. The first gene in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase. J. Mol. Biol. **286:**465–474.
31. **Paitan, Y., E. Orr, E. Z. Ron, and E. Rosenberg.** 1999. Genetic and functional analysis of genes required for the post-modification of the polyketide antibiotic TA of *Myxococcus xanthus*. Microbiology **145:**3059–3067.
32. **Palaniappan, N., B. S. Kim, Y. Sekiyama, H. Osada, and K. A. Reynolds.** 2003. Enhancement and selective production of phoslactomycin B, a protein phosphatase IIa inhibitor, through identification and engineering of the corresponding biosynthetic gene cluster. J. Biol. Chem. **278:**35552–35557.
33. **Patel, P. S., S. Huang, S. Fisher, D. Pirnik, C. Aklonis, L. Dean, E. Meyers, P. Fernandes, and F. Mayerl.** 1995. Bacillaene, a novel inhibitor of prokaryotic protein synthesis produced by *Bacillus subtilis*: production, taxonomy, isolation, physico-chemical characterization and biological activity. J. Antibiot. **48:**997–1003.
34. **Piel, J.** 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. Proc. Natl. Acad. Sci. USA **99:**14002–14007.
35. **Piel, J., I. Höfer, and D. Hui.** 2004. Evidence for a symbiosis island involved in horizontal acquisition of pederin biosynthetic capabilities by the bacterial symbiont of *Paederus fuscipes* beetles. J. Bacteriol. **186:**1280–1286.
36. **Piel, J., D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, and S. Matsunaga.** 2004. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. Proc. Natl. Acad. Sci. USA **101:**16222–16227.
37. **Rawlings, B. J.** 2001. Type I polyketide biosynthesis in bacteria (part B). Nat. Prod. Rep. **18:**231–281.
38. **Reid, R., M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson, and R. McDaniel.** 2003. A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. Biochemistry **42:**72–79.
39. **Reva, O. N., and B. Tummler.** 2004. Global features of sequences of bacterial chromosomes, plasmids and phages revealed by analysis of oligonucleotide usage patterns. BMC Bioinformatics **5:**90. [Online.] doi:10.1186/1471-2105-5-90.
40. **Royer, M., L. Costet, E. Vivien, M. Bes, A. Cousin, A. Damais, I. Pieretti, A. Savin, S. Megessier, M. Viard, R. Frutos, D. W. Gabriel, and P. C. Rott.** 2004. Albicidin pathotoxin produced by *Xanthomonas albilineans* is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes. Mol. Plant-Microbe Interact. **17:**414–427.
41. **Schupp, T., C. Toupet, N. Engel, and S. Goff.** 1998. Cloning and sequence analysis of the putative rifamycin polyketide synthase gene cluster from *Amycolatopsis mediterranei*. FEMS Microbiol. Lett. **159:**201–207.
42. **Scotti, C., M. Piatti, A. Cuzzoni, P. Perani, A. Tognoni, G. Grandi, A. Galizzi, and A. M. Albertini.** 1993. A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases. Gene **130:**65–71.
43. **Stachelhaus, T., H. D. Mootz, and M. A. Marahiel.** 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. **6:**493–505.
44. **Steinmetz, M., and R. Richter.** 1994. Plasmids designed to alter the antibiotic resistance expressed by insertion mutations in *Bacillus subtilis*, through in vivo recombination. Gene **142:**79–83.
45. **Tang, G.-L., Y.-Q. Cheng, and B. Shen.** 2004. Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. Chem. Biol. **11:**33–45.
46. **Tang, L., S. L. Ward, L. Chung, J. R. Carney, Y. Li, R. Reid, and L. Katz.** 2004. Elucidating the mechanism of cis double bond formation in epothilone biosynthesis. J. Am. Chem. Soc. **126:**46–47.
47. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
48. **Vater, J., X. Gao, G. Hitzeroth, C. Wilde, and P. Franke.** 2003. "Whole cell"—matrix-assisted laser desorption ionization-time of flight mass spectrometry, an emerging technique for efficient screening of biocombinatorial libraries of natural compounds—present state of research. Comb. Chem. High Throughput Screen. **6:**557–567.
49. **Ward, S. L., Z. H. Hu, A. Schirmer, R. Reid, W. P. Revill, C. D. Reeves, O. V. Petrakovsky, S. D. Dong, and L. Katz.** 2004. Chalcomycin biosynthesis gene cluster from *Streptomyces bikiniensis*: novel features of an unusual ketolide produced through expression of the *chm* polyketide synthase in *Streptomyces fradiae*. Antimicrob. Agents Chemother. **48:**4703–4712.
50. **Wenzel, S. C., and R. Muller.** 2005. Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. Curr. Opin. Chem. Biol. **9:**447–458.
51. **Wilson, K. E., J. E. Flor, R. E. Schwartz, H. Joshua, J. L. Smith, B. A. Pelak, J. M. Liesch, and D. Hensens.** 1987. Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*. 2. Isolation and physicochemical characterization. J. Antibiot. **40:**1682–1691.
52. **Witkowski, A., A. K. Joshi, and S. Smith.** 2002. Mechanism of the beta-ketoacyl synthase reaction catalyzed by the animal fatty acid synthase. Biochemistry **41:**10877–10887.
53. **Zweerink, M. M., and A. Edison.** 1987. Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*. III. Mode of action of difficidin. J. Antibiot. **40:**1692–1697.