Genome mining and UHPLC–QTOF–MS/MS to identify the potential antimicrobial compounds and determine the specificity of biosynthetic gene clusters in Bacillus subtilis NCD-2

Zhenhe Su
Hebei Academy of Agriculture and Forestry Sciences
Xiuye Chen
Hebei Academy of Agriculture and Forestry Sciences
Xiaomeng Liu
Hebei Academy of Agriculture and Forestry Sciences
Qinggang Guo
Hebei Academy of Agriculture and Forestry Sciences
Shezeng Li
Hebei Academy of Agriculture and Forestry Sciences
Xiuyun Lu
Hebei Academy of Agriculture and Forestry Sciences
Xiaoyun Zhang
Hebei Academy of Agriculture and Forestry Sciences
PeiPei Wang
Hebei Academy of Agriculture and Forestry Sciences
Lihong Dong
Hebei Academy of Agriculture and Forestry Sciences
Weisong Zhao
Hebei Academy of Agriculture and Forestry Sciences
Ping Ma (pingma88@126.com)
Institute of Plant Protection

Research article

Keywords: Bacillus subtilis NCD-2, Genome mining, UHPLC–QTOF–MS/MS, Secondary metabolites, Fengycin

Posted Date: September 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-33091/v3

Version of Record: A version of this preprint was published on November 5th, 2020. See the published version at https://doi.org/10.1186/s12864-020-07160-2.

Abstract

Background: *Bacillus subtilis* strain NCD-2 is an excellent biocontrol agent against plant soil-borne diseases and shows broad-spectrum antifungal activities. This study aimed to explore some secondary metabolite biosynthetic gene clusters and related antimicrobial compounds in strain NCD-2. An integrative approach combing genome mining and structural identification technologies using ultra-high-performance liquid chromatography coupled to quadrupole time-of-flight tandem mass spectrometry (UHPLC-MS/MS), was adopted to interpret the chemical origins of metabolites with significant biological activities.

Results: Genome mining revealed nine gene clusters encoding secondary metabolites with predicted functions, including fengycin, surfactin, bacillaene, subtilosin, bacillibactin, bacilysin and three unknown products. Fengycin, surfactin, bacillaene and bacillibactin were successfully detected from the fermentation broth of strain NCD-2 by UHPLC-QTOF-MS/MS. The biosynthetic gene clusters of bacillaene, subtilosin, bacillibactin, and bacilysin showed 100% amino acid sequence identities with those in *B. velezensis* strain FZB42, whereas the identities of the surfactin and fengycin gene clusters were only 83% and 92%, respectively. Further comparison revealed that strain NCD-2 had lost the *fenC* and *fenD* genes in the fengycin biosynthetic operon. The biosynthetic enzyme-related gene *srfAB* for surfactin was divided into two parts. Bioinformatics analysis suggested that FenE in strain NCD-2 had a similar function to FenE and FenC in strain FZB42, and that FenA in strain NCD-2 had a similar function to FenA and FenD in strain FZB42. Five different kinds of fengycins, with 26 homologs, and surfactin, with 4 homologs, were detected from strain NCD-2. To the best of our knowledge, this is the first report of a non-typical gene cluster related to fengycin synthesis.

Conclusions: Our study revealed a number of gene clusters encoding antimicrobial compounds in the genome of strain NCD-2, including a fengycin synthetic gene cluster that might be unique by using genome mining and UHPLC–QTOF–MS/MS. The production of fengycin, surfactin, bacillaene and bacillibactin might explain the biological activities of strain NCD-2.

Background

The *Bacillus* genus has received considerable attention as a biological resource for the development of microbial pesticides, partly because some or most of its members form stress-resistant spores that do not harm the environment and are useful in pesticide production [1-3]. *Bacillus subtilis* and its closely related species are ubiquitous inhabitants of soil, and are widely recognized as powerful biocontrol agents against plant soil-borne diseases [4]. The mechanisms used by *B. subtilis* to suppress plant soil-borne diseases include competing with phytopathogens for nutrients and spatial sites, inducing systemic resistance in plants, and inhibiting pathogen growth by producing antimicrobial compounds [5]. The latter is a general characteristic of *B. subtilis* biocontrol agents and plays an important role in the suppression of plant diseases [6, 7]. *B. subtilis* produces more than two dozen antimicrobial compounds with amazing structural variety. Based on their biosynthetic pathways, the antimicrobial compounds are divided into small molecular compounds synthesized by the ribosomal pathway, such as bacteriocins, and peptide compounds synthesized by the non-ribosomal pathway, such as lipopeptides and polyketides [8]. Most antimicrobial compounds are secondary metabolites produced by biocontrol of *Bacillus* spp, and are not necessary for their growth and reproduction but lead to shifts of rhizospheric microbial functional subsystems and affect the availability of nutrients for the plant [9]. Secondary metabolites also function as essential chemical signals for the induction of cellular differentiation in the producing organism and for controlling its metabolism [10, 11].

The genes encoding the secondary metabolites commonly exist in clusters and encode enzyme complexes with multiple functions [12]. The polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) gene clusters have been well studied. The PKS pathway polyketides require at least three domains: an acyl transferase, a ketosynthase, and an acyl

carrier protein [13]. The NRPS synthetic pathways share a common multicarrier thiotemplate mechanism requiring the cooperation of three basic domains [14]. The adenylation domain selects the cognate amino acid and generates an enzymatically stabilized aminoacyl adenylate. The peptidyl carrier domain is equipped with a 4'-phosphopantetheine prosthetic group, to which the adenylated amino acid substrate is transferred and bonded by a thioester bond. The condensation domain catalyzes the formation of a new peptide bond [13]. The carbon skeleton of the metabolite is synthesized by the core PKS and NRPS enzymes, and is then modified to form the final product with the assistance of various modifying enzymes [15]. The bioactive secondary metabolites produced by the PKS/NRPS pathway in *B. subtilis* have received extensive studies, such as bacilysin [16], bacilysocin [17], surfactin [18], iturin A [19], fengycin [20], mycosubtilin [21], bacillomycins [8], and difficidin [16].

The conventional method for screening new active products is generally based on biological tests, which is time-consuming and often result in repeatedly screened out the same products [22]. Thus, a more rapid and effective screening strategy to detect secondary metabolites was required [23, 24]. Genome mining is a technology that uses modern bioinformatics to recognize specific functional genes or gene clusters from genome sequences [25]. With the rapid development of DNA sequencing technology and the decrease of sequencing cost, a large number of microbial genome sequences have been determined [26], which makes genome mining an accurate and efficient strategy to find new metabolites [25].

B. subtilis strain NCD-2 is a plant soil-borne disease-suppressive agent producing lipopeptides, fengycin, and surfactin [27]. Fengycin shows strong antifungal activity, and surfactin facilitates the root colonization. Both fengycin and surfactin play important roles in suppression of plant soil-borne diseases by strain NCD-2 [28]. The purpose of this study was to identify potential secondary metabolites in strain NCD-2, reveal the different gene clusters of the secondary metabolites between strain NCD-2 and the reference strain *B. velezensis* FZB42, and identify the potential secondary metabolites produced by strain NCD-2.

Results

Genomic features of strain NCD-2

A total of 501,671,500 paired-end reads and 5,016,715 clean single reads (412-bp library; paired-ends of 75 bp) were assembled using the software package Velvet [29]. The genome of *B. subtilis* NCD-2 contains 189 contigs (>133 bp; N90, 16,187) of 4,644,322 bp, with an average G+C content of 43.5%. The final assembled genome comprises 4,444 genes, including 4,329 protein-coding genes (418 signal peptide-coding genes), 83 tRNA genes for all 20 amino acids, 30 rRNA genes, and 2 clustered regularly interspaced short palindromic repeats (CRISPR) genes. A total of nine putative gene clusters responsible for antimicrobial metabolite biosynthesis were identified. These gene clusters included PKS and NRPS genes (Fig. 1).

The taxonomic status of strain NCD-2

At present, 272 *B. subtilis* genome sequences have been deposited in the GenBank database, including 113 whole- and 159 incomplete genome sequences. The genome sizes of the 272 *B. subtilis* strains range from 2.68 Mb to 5.35 Mb, and the GC contents range from 42.9% to 46.6%. These genome sequences were downloaded from the GenBank database, and their accession numbers were listed in Additional file 1, Table S1. To analyze the evolution of different *B. subtilis* strains, a phylogenetic tree was constructed based on the complete genome sequences. The 272 strains of *B. subtilis* were divided into four subspecies, *subtilis*, *inaquosorum*, *spizizenii*, and *stercoris* due to producing different bioactive secondary metabolites [30]. As shown in Fig. 2, strain NCD-2 (represented by the black bar) clustered with *B. subtilis* strain UD1022 and was closely related to *B. subtilis* strains XF-1, BAB-1, HJ5, SX01705, and BSD-2.

Secondary metabolite biosynthetic gene clusters in strain NCD-2

The secondary metabolite biosynthetic gene clusters in the genome of strain NCD-2 were predicted using antiSMASH [31]. In total, nine such clusters were identified (Table 1), including three NRPSs, two terpenes, one hybrid NRPS-TransAT PKS-Other KS, one type III polyketide, one sactipeptide-head to tail gene cluster, and a gene cluster with unknown function. The structural compositions of the gene clusters are shown in Fig. 3. These clusters were composed of core biosynthetic, additional biosynthetic, transport-related, regulatory, and other genes. Among them, clusters 3, 7, 8, and 9 had 100% amino acid sequence homologies with known gene clusters that synthesize bacillaene, bacillibactin, subtilosin, and bacilysin, respectively (Table 1). Gene cluster 1 showed 82% amino acid similarity with a surfactin synthetase gene cluster, and gene cluster 4 showed 93% amino acid similarity with a fengycin biosynthetic gene cluster in *B. velezensis* strain FZB42. However, gene clusters 2, 5, and 6 did not match any known gene clusters. Clusters 1 and 4 of strain NCD-2 were further compared with their counterparts in the model strain 168 and *B. subtilis* strains closely related to strain NCD-2 in the phylogenetic tree. The predicted fengycin biosynthetic gene cluster in strain NCD-2 contained three genes, *fenEAB*, while all the other strains contained five genes, *fenCDEAB* (Additional file 1, Fig. S1). SrfAB of surfactin was synthesized via typical transcription and translation of *srfAB* in the 11 strains. However, the same SrfAB was potentially assembled with Gms0366 and Gms0367 and then separately transcribed and translated by *gms0366* and *gms0367* in strain NCD-2 (Additional file 1, Fig. S2). Therefore, we hypothesized that the structures and functions of fengycin and surfactin from strain NCD-2 might be different from those in other *B. subtilis* strains.

Specificity of the surfactin and fengycin synthetase gene clusters in *B. subtilis* NCD-2

The surfactin biosynthetic gene cluster *gms0365-0368* in strain NCD-2 was analyzed using PRISM, and the core genes were selected for PKS/NRPS analysis. Gms0365 had an identical conserved domain, CATCATCATe, with SrfAA in strain FZB42, in which C, A, T, and Te represent the condensation, adenylation, thiolation, and thioesterase domains, respectively (Fig. 4a). Compared with SrfAB in strain FZB42, Gms0366 in strain NCD-2 lacked the T and E domains, but the amino acid residues for the binding pockets in Gms0366 were exactly the same as those of SrfAB. The residues of the different adenylation domains A6 and A2 from the enzymes Gms0365 and Gms0366, respectively, were exactly the same, and both bound the amino acid leucine. Gms0367 only had T and E domains, and no specific substrate-binding domain. The superposition of the Gms0367 and Gms0366 domains formed a complete SrfAB. The T domain was reversed between Gms0367 and Gms0368. Gms0368 contained CATe domains, in which the thioesterase domain releases linear peptide chains. The domains of Gms0368 were the same as those of SrfAC, but the binding pocket (DAF-LGCV) had one missing residue compared with that of strain FZB42 (DAFXLGCV).

The fengycin biosynthetic cluster in strain FZB42 contained five genes *fenCDEAB* (Fig. 4b). However, the same cluster in strain NCD-2 contained only three genes: *gms1961*, *gms1960*, and *gms1959* (Fig.4b). Gms1961 corresponded to FenE in strain FZB42 had conserved residues of A8 and A9, which bind amino acids Glu and Val, respectively (Fig. 4b). Gms1960 and Gms1959 had conserved amino acid sequences related to FenA and FenB in strain FZB42, respectively. Interestingly, no homologs of FenC and FenD were identified in the genome of strain NCD-2. Consequently, the amino acid sequences of FenC and FenD of strain FZB42 were compared with the strain NCD-2 proteome using BioEdit, and it was found that their most similar proteins were Gms1961 and Gms1960, respectively (Additional file 1, Tables. S2, S3). This finding led to the hypothesis that Gms1961 and Gms1960 performed the functions of FenC and FenD in strain NCD-2, respectively. Thus, Gms1961 and Gms1960 might both have dual functions in the synthesis of fengycin. Gms1961 in strain NCD-2 had the functions of FenE and FenC in strain FZB42, and Gms1960 had the functions of FenA and FenD. However, it should be pointed out that the FenD domain in strain NCD-2 varied greatly with that of FZB42, and we cannot rule out the possibility that other enzymes in NCD-2 might have similar functions as FenD.

PCR amplification using the primer set targeting the *fenE* and *dacC* genes produced the expected 1,032 bp fragment in the extremely large size of the target (16,555 bp) (Fig. 5a-b). Sequencing of

the 1,032 bp fragment and alignment with the gene locus *gms1959-1962* confirmed the lack of *fenC* and *fenD* homologs in this cluster (Fig. 5c-d). Compared to wild-type NCD-2, the in-frame deletion mutant of *gms1961* completely lost fengycin production (Fig. 6a-c).

We further compared the fengycin synthetase gene cluster of NCD-2 with other 11 corresponding clusters from *B. subtilis* strains closely related to strain NCD-2 (Additional file 1, Fig. S1). All the strains contained the fengycin biosynthetic gene cluster *fenCDEAB* (also *ppsABCDE*), except strain NCD-2, which contained *fenEAB*, suggesting that the fengycin biosynthetic gene cluster of strain NCD-2 is unique.

MS/MS of fengycin and surfactin in NCD-2

Fengycin was separated from the lipopeptide extract of strain NCD-2 using fast protein liquid chromatography (FPLC) (Additional file 1, Fig. S3). QTOF–MS/MS analysis revealed five fractions in the fengycin cluster (Fig. 7a–e), which had mass-to-charge ratio (*m/z*) values of 732.4, 746.4, 725.4, 739.4, and 767.4 (secondary MS), representing fengycin A, fengycin B, fengycin A2, fengycin B2, and fengycin C, respectively. The typical MS/MS spectra showed the distributions of key fragmentation ions (α and β), representing the linear N-terminal and the cyclic C-terminal segments, respectively, of diverse fengycin species (Fig. 7a–e and Additional file 1, Fig. S4a-b). The MS/MS spectrum of the fengycin ion at *m/z* 732.4 yielded two intense product ions at *m/z* 966.5 and 1,080.5, representing fengycin A (Fig. 5a), while the MS/MS spectrum of the fengycin ion at *m/z* 746.4 (Fig. 7b) yielded key product ions at *m/z* 994.5 and 1,108.6, representing fengycin B (Fig. 7b). The MS/MS spectrum of the fengycin ion at *m/z* 725.4 yielded two intense product ions at *m/z* 952.4 and 1,066.5, representing fengycin A2 (Fig. 7c), while the MS/MS spectrum of the fengycin ion at *m/z* 739.4 (Fig. 7d) yielded key product ions at *m/z* 980.5 and 1,094.5 representing fengycin B2 (Fig. 7d). The MS/MS spectrum of the fengycin ion at *m/z* 767.4 yielded two intense product ions at *m/z* 994.5/1,008.5 and 1,108.6/1,122.6 representing fengycin C (Fig. 7e). Five classes of fengycins were identified based on the key product ions of β-hydroxy fatty acid (β-OH FA) with chain lengths varying from C12 to C20 (Table 2, Figs. S5–S9). The MS/MS spectrum of the surfactin ion at *m/z* 1,008.7 yielded one intense product ion at *m/z* 685.5 (Fig. 7f and Additional file 1, Fig. S4c). Based on this key product ion, one class of compounds was identified: surfactins (*m/z* values of 994.6, 1,008.7, 1,022.7 and 1,036.7) with fatty acid chains varying from C11 to C15 (Fig. S10).

Detection of other antimicrobial active compounds in NCD-2

Besides of the fengycin and surfactin, other four antimicrobial compounds bacillaene, bacilysin, bacillibactin and subtilosin were also extracted from the fermentation broth of strain NCD-2 by using different extracting methods, respectively. However, only bacillaene and bacillibactin were detectable from the extracts by UHPLC-QTOF-MS (Fig. 8a, 8b).

Discussion

B. subtilis has the potential to produce two dozen antimicrobial substances, and 5%–8% of the *B. subtilis* genome contributes to the production of antimicrobial substances [32]. Some of these substances inhibit the growth of pathogens and the germination of spores. The lipopeptide mixture of *B. subtilis* C232 inhibits the formation of *Verticillium dahliae* microsclerotia [33], and the volatile compounds secreted by *B. subtilis* JA inhibit the conidial formation and mycelial growth of *Glomus etunicatum* [34].

However, certain antimicrobial compounds are synthesized only in respond to external stimulation or under special conditions, this made it difficult to harvest all antimicrobial compounds produced by a given *Bacillus* strain using traditional cultivation and extraction methods [22]. Genome mining allows the prediction of metabolites based on genome sequences, including both identified antimicrobial compounds and novel antibiotics that have not been

previously described. For example, the new NRPS antibiotic coelichelin was identified by genomic analysis from *Streptomyces coelicolor* [35]. Pseudomycoicidin in *Bacillus pseudomycoides* DSM 12442 was discovered by genome mining and through heterologous expression of its BGC in *Escherichia coli* [36].

Several lipopeptide antibiotics, including fengycin and surfactin have been identified in *B. subtilis* NCD-2 by using traditional cultivation and extraction methods [28]. Fengycin showed strong antifungal abilities against *V. dahliae* and *B. cinerea*. In this study, seven additional secondary metabolite gene clusters were found by genome mining, and some of them were identified using MS/MS. In total, *B. subtilis* NCD-2 had the potential to produce at least 9 kinds of secondary metabolites including surfactin, bacillaene, fengycin, bacillibactin, subtilosin, bacilysin, two terpenes, and one unknown product. Surfactin exhibits antibacterial, antiviral, antitumor and hemolytic action [37]. Bacillaene inhibits bacterial growth by inhibiting prokaryotic protein synthesis [38]. Fengycin shows specific antifungal activity against filamentous fungi [39]. Bacillibactin functions as a siderophore to compete for irons with environmental microbes especially under the iron deficiency conditions. *B. subtilis* expresses genes involved in the synthesis for bacillibactin to pirate other microbial iron.[40]. Subtilosin possesses antibacterial activity against a diverse range of bacteria [41]. Bacilysin exhibits antimicrobial activities against both bacteria and *Candida albicans* [42]. However, only fengycin, surfactin, bacillaene and bacillibactin were successfully detected from the extract of strain NCD-2 by UHPLC-MS/MS (Fig. 7, 8). Bacilysin and subtilosin remained undetectable. A likely reason for their undetectability is the low expression level of their biosynthetic gene clusters under the experimental conditions.

B. velezensis FZB42 is a model strain of plant beneficial rhizobacteria. Thirteen gene clusters involved in the non-ribosomal and ribosomal synthesis of secondary metabolites with putative antimicrobial action have been identified within the genome of strain FZB42, including fengycin. The mechanism of fengycin synthesis has been well studied in *B. velezensis* strain FZB42 [43]. *B. subtilis* 168 has the entire gene cluster for synthesizing fengycin, but it can not produce fengycin because of a deficiency of a native *sfp* gene [44]. The BGC repository MIBiG (Minimum Information about a Biosynthetic Gene cluster) includes only has one fengycin biosynthetic gene cluster from *B. velezensis* FZB42 [45, 46]. Therefore, the fengycin biosynthetic gene cluster of strain NCD-2 was compared with that of *B. velezensis* FZB42. Fengycin comprises a peptide ring circled by 10 amino acids with a fatty acid chain tail. The fengycin biosynthetic gene clusters in most producing strains consist of the genes *fenCDEAB* (38 kb), which encode 5 enzymes, among which FenC recognizes and carries glutamate and ornithine; FenD recognizes and carries tyrosine and threonine, FenE recognizes and carries glutamate and valine; FenA recognizes and carries proline, glutamine, and tyrosine; and FenB recognizes and carries isoleucine. FenCDEAB recognizes 10 amino acids and carries them to the β-OH FA chain to form fengycin [47-49]. However, NCD-2 only had *fenEAB*, but no *fenC* and *fenD*, compared with the typical cluster structure of *fenCDEAB* in strain FZB42 strain and other 10 *Bacillus* strains (Fig. 4b and Additional file 1, Fig. S1). To exclude the errors introduced by genome sequencing or assembly, the fragment between *fenE* and *dacC* was cloned and sequenced, and it was confirmed that *fenC* and *fenD* were lost in strain NCD-2 (Fig.5a-d). To identify the enzymes FenC and FenD in the NCD-2 genome, their amino acid sequences from FZB42 were selected to screen for homologs by scanning the local NCD-2 proteome using BioEdit. The protein Gms1961 in strain NCD-2 had the greatest similarity to FenC at an amino acid sequence level (Additional file 1, Table S2). The Gms1961 protein contained 2,550 amino acids, with a molecular weight of 287.50 kDa. The substrate bound by the adenylation domain of the Gms1961 protein was predicted (Additional file 1, Table S4). The adenylation A9 domain bound valine and N5-hydroxyornithine, with the latter being a transitional form of ornithine combined with the adenylation domain [50]. The UHPLC-QTOF-MS/MS results of the fengycins revealed the amino acid ornithine at position 2 in all of the examined structures (Fig. 7a–e), indicating the presence of a protein that transports ornithine in the NCD-2 strain. We thus hypothesized that Gms1961 functions as FenC and FenE. The same analysis was performed using the Gms1960 protein, which had the greatest similarity with FenD (Additional file 1, Table S3). However, the FenD domains in Gms1960 and FZB42 varied greatly. Therefore, it was hypothesized that Gms1960 or other enzymes function similarly to FenD.

Although lacking two important genes, *fenC* and *fenD*, strain NCD-2 was capable of producing 26 homologs of 5 types of fengycins (A, B, A2, B2, and C) (Additional file 1, Fig. S4). The amino acids at positions 6 and 10 in the cyclic peptide ring of fengycin determine its structural type. When the amino acid at position 6 was valine with isoleucine or valine at position 10, fengycin B or fengycin B2, respectively, was produced (Fig. 7a, b and Additional file 1, Fig. S4); however, when the amino acid at position 6 was alanine with isoleucine or valine at position 10, fengycin A or fengycin A2, respectively, was produced (Fig. 7c, d and Additional file 1, Fig. S4). When the amino acid at position 6 was isoleucine or leucine with valine at position 10, fengycin C was produced (Fig. 7e and Additional file 1, Fig. S4). Each fengycin type had different homologs according to the number of carbon atoms in its β-OH FA chain, and the molecular weights of each homologs differed by 14 (-CH2) [51]. Compared to the short-chain varieties, long-chain fatty acids increase the hydrophobic activities of lipopeptides, making them more likely to have membrane-bound antimicrobial effects [52]. A *B. circulans* strain produces four fengycin homologs, but only fengycins with C16 and C17 carbon atoms in their β-OH FA chains have antibacterial activity [53]. Among the 26 fengycin homologs produced by strain NCD-2, 14 fengycins had more than 16 carbon atoms in their β-OH FA chains, which might be the most important composition for antimicrobial function. The *B. siamensis* SCSIO 05746 strain produces a large number of fengycin homologs, including 19 homologs of fengycin B [54]. Using MS/MS analysis, the five fengycins produced by the NCD-2 strain were divided into 26 homologs (Fig. 7a–e and Additional file 1, Fig. S5-S9). Therefore, NCD-2 is currently the strain with the largest number of known fengycin homologs [55].

During microbial synthesis of secondary metabolites, such as lipopeptide, the relatively high energy-consuming process of protein synthesis takes priority [56]. Excessive energy consumption is not conducive to the normal growth of microbes, and, generally, microbes produce antibiotics in large amounts only under stress, such as encountering pathogens [57]. We hypothesized that the essential biosynthetic genes *fenEAB* involved in fengycin synthesis were retained, while another two important biosynthetic genes *fenCD* were lost in the long-term evolution of strain NCD-2. Five fengycins were still produced. Gms1961 might serve the dual roles of FenC and FenE, indicating that NCD-2's fengycin biosynthetic process is unique to the strain and is more energy-efficient than the process used in the other strains.

Conclusions

Genome mining and UHPLC–QTOF–MS/MS analysis revealed 9 gene clusters encoding antimicrobial compounds in the genome of Bacillus subtilis NCD-2. Among them, the fengycin biosynthetic gene cluster containing fenEAB genes is unique to strain NCD-2 compared with the other tested B. subtilis strains. Strain NCD-2 might employ a unique mechanism for synthesizing fengycin, which may shed new light on the synthesis and evolution of antimicrobial lipopeptides through the NRPS pathway.

Methods

Microorganisms and culture conditions

B. subtilis NCD-2 was routinely grown at 37 °C on Luria Bertani medium. For lipopeptide, bacillaene, bacilysin, bacillibactin and subtilosin production, strain NCD-2 was grown in Landy broth [58], PA medium [59], MSA medium [60], and TSB medium [61] at 30 ℃ and 180 rpm. The phytopathogen *Botrytis cinerea* BC-10 was used for antifungal activity test following the method described by Guo et al [28] with some modifications. Briefly, a 6-mm diameter disc of *B. cinerea* was placed in the center of a 9-cm potato dextrose agar (PDA) plate, and the plate was inoculated 2 cm from the center with *B. subtilis* NCD-2 using a sterilized toothpick. Finally, the diameter of the inhibition zone was measured after a 3-d incubation at 25℃.

Genome sequencing of strain NCD-2

The Illumina Solexa platform(BGI, Shenzhen, China) was used for whole-genome sequencing following the method described by Karim [62] with some modifications. The quality of reads was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [63], and paired-end reads were trimmed using Sickle (https://github.com/najoshi/sickle), and were assembled using the software package Velvet [29]. QUAST 5.02 was used to assess the quality of contigs and scaffolds [64]. The assembled scaffolds were annotated using Prokka (version v.1.13) [65]. Annotation of the genome of strain NCD-2 was performed using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (http://www.ncbi.nim.nih.gov/genome/annotation_prok/) utilizing GeneMark, Glimmer, and tRNAscan-SE tools [66], and functional annotation was carried out using the Rapid Annotations by subsystems Technology (RAST) server with the seed database [67]. Finally, the genome of strain NCD-2 was deposited in the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) under the GenBank accession number CP023755.

Evolutionary analysis, signal peptide and CRISPR repeat detection

Whole-genome sequences of *B. subtilis* and closely related species were downloaded from the NCBI database, and the REALPHY website (http://realphy.unibas.ch) [68] was used to perform genome-wide comparisons with the default parameters. A phylogenetic analysis was conducted using MEGA5 [69] with the Maximum Composite Likelihood parameter model [70]. A phylogenetic tree was constructed using the Neighbor-joining algorithm method with bootstrap values based on 1,000 replications. The signal peptide was predicted using the SignalP-5.0 website (www.cbs.dtu.dk/services/SignalP-5.0/) [71]. CRISPR repeats were detected using CRISPRCasFinder (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) [72].

Prediction and specificity analysis of secondary metabolite biosynthetic gene clusters

Secondary metabolite biosynthetic gene clusters for strain NCD-2 were detected using antiSMASH (http://antismash.secondarymetabolites.org) [31, 73] and PRISM (http://grid.adapsyn.com/prism/) [74] with the default parameters. The functional domain predictions for PKS/NRPS in the predicted gene clusters were analyzed using the PKS/NRPS Analysis Website (http://nrps.igs.umaryland.edu/) [75]. Typical PKS and NRPS sequences were selected for genomic and proteomic scanning after using BioEdit software to create a local BLAST based on strain NCD-2's genome and proteome, respectively.

Detection of FenC and FenD loss in the genome of strain NCD-2

FenC and FenD are two important enzymes for synthesizing fengycin. A pair of degenerate primers targeting *fenE* (5 -$\mathbb{C}RT\mathbb{C}ATKAYGATATGATG$ - 3) and *dacC* (5- $TGACAG\forall TGRYGGGMGG\forall C$ - 3) were designed based on the conserved bases of *fenE* and *dacC* in strain NCD-2 and *B. velezensis* strain FZB42. The 16S rDNA (27-F/1492R) primers were used as a positive control [76]. The amplification procedure included a denaturation step at 95 ℃ for 2 min, followed by 32 cycles of 20 sec strand separation at 95 ℃, 20 sec annealing at 55 ℃, and 90 sec elongation at 72 ℃, and an elongation step of 5 min at 72 ℃. The target fragment from NCD-2 was purified by a gel extraction kit (Sangon, Shanghai, China), ligated to blunt-ended vector (Transgen, Beijing, China) and sequenced by BGI company (Shenzhen, China). The fragment sequence was deposited at NCBI with the accession number of MT984302.

Separation of lipopeptides by FPLC

Lipopeptides were extracted using the method described by Guo et al [28]. Briefly, strain NCD-2 or derived strain Δ*gms1961* were cultured in 100 mL Landy broth [58] at 30 ℃ for 72 h with shaking at 180 rpm. The cell-free supernatant was obtained by centrifugation at 8,000×g for 30 min at 4 ℃. The supernatant was adjusted to pH 2.0 with 6 mol/L HCl and stored for 12 h at 4 ℃. After centrifugation at 10,000 ×g, for 20 min, the resulting pellet was extracted with 10 mL

g for 2 h. The obtained extracts were sterilized by passing through 0.45-μm

filters (Millex-GV, Millipore, Billerica, MA, USA) to obtain crude lipopeptides. The crude lipopeptides were separated and purified using an AKTA Purifier (GE Healthcare, Uppsala, Sweden) with a SOURCE 5RPC ST 4.6/150 column as described previously [77]. The lipopeptides were eluted with solvent A [2% acetonitrile containing 0.065% trifluoroacetic acid (TFA) (V/V)] and solvent B [80% acetonitrile containing 0.05% TFA (V/V)] using a linear gradient of 0%–100% acetonitrile over 57 min at a flow rate of 1 mL/min. The detection wavelength was 215 nm. All the main peaks were automatically collected by FPLC. Finally, each peak was concentrated using a rotary evaporator and was analyzed using UHPLC-QTOF–MS/MS.

UHPLC–QTOF–MS/MS

UHPLC–QTOF–MS/MS analysis was conducted on a hybrid quadrupole time-of-flight tandem mass spectrometer (AB SCIEX TripleTOF 5600 Q-TOF/MS, Foster City, CA, USA) with an HPLC (Shimadzu, Kyoto, Japan) equipped with LC-30AD binary pumps, a SIL-30AC autosampler, and a CTO-30AC column oven. A C18 reversed phase LC column (Shim-pack GIST 2-μm particles, 2.1 mm×100 mm) was used for separation. The mobile phases A and B were water and acetonitrile, respectively, with 0.1% formic acid in both phases and with the optimized linear gradient elution procedure as follows: 0.0–0.5 min, 30% B; 0.5–50 min, 60% B; 50–52 min, 95% B; 52–55 min, 95% B; 55–55.1 min, 30% B; and 55.1–60 min, 30% B. The injection volume was 20 μL with a flow rate of 0.30 mL/min. The column oven was set at 40℃. MS analysis was performed using a 5600 TripleTOF system equipped with a DuoSpray™ Ion Source, and the data were processed using Analyst TF 1.7 software (Applied Biosystems Sciex, Toronto, ON, Canada). PeakView™ software 2.0 (Applied Biosystems Sciex, Toronto, ON, Canada) was used to investigate and interpret the mass spectral data, with special tools for processing accurate mass data and structural elucidation. The DuoSpray™ ion source was used in positive ion mode. The instrumental parameters were set as follows: ion spray voltage floating, 5,000 V; nebulizing gas, 50 psi; heater gas, 50 psi; curtain gas, 35 psi; temperature, 350℃; declustering potential, 100 V; collision energy, TOF MS experiments: 10.0 V. TOF-MS/MS experiments: rolling collision energy, with collision energy spread 5 V. The data was acquired using Information Dependent Acquisition for a single run analysis with m/z range of 200–2,000 in TOF MS and 50–1,600 in MS/MS.

Detection of bacillaene, bacilysin, bacillibactin and subtilosin

For bacillaene, strain NCD-2 was cultured in 100 mL Landy broth at 30℃ for 72 h with shaking at 180 rpm, and bacillaene was extracted with methanol using the method described by Reddick et al [78]. For bacilysin, strain NCD-2 was cultured in 100 mL PA medium at 30℃ for 72 h with shaking at 180 rpm, and bacilysin was extracted with ice-cold ethanol as described by Wu et al [59]. For bacillibactin, strain NCD-2 was cultured in 100 mL MSA medium at 30℃ for 72 h, and bacillibactin was extracted with ethanol as described by Li et al [60]. For subtilosin, strain NCD-2 was cultured in 100 mL TSB medium at 30℃ for 72 h, and subtilosin was extracted with precipitation with 65% ammonium sulphate as described by Charles et al [61]. The extracts were detected by UHPLC-QTOF-MS/MS as described above.

Abbreviations

UHPLC-QTOF-MS/MS: ultra-high-performance liquid chromatography coupled to quadrupole-time-of-flight tandem mass spectrometry; A domain: adenylation domain; C domain: condensation domain; T domain: thiolation domain; Te: thioesterase domain; E domain: epimerization domain; N90: the minimum contig length to cover 90 percent of the genome; CRISPR: clustered regularly interspaced short palindromic repeats; PDA: potato dextrose agar; BGC: biosynthetic gene cluster; FPLC: fast protein liquid chromatography; *m/z*: mass-to-charge ratio; TFA: trifluoroacetic acid.β-OHFA: β-hydroxy-fatty acid.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and analyzed in the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the earmarked fund for National Key R & D Projects (2017YFD0200400), the China Agriculture Research System (CARS-18-15), the Natural Science Foundation of Hebei Province (C2019301101), the National Natural Science Foundation of China (31572051 and 31601680), the PhD Fund of Hebei Academy of Agriculture and Forestry (C19R01003), and the Special Fund for Agro-scientific Research in the Public Interest, China (201503109).

Authors' contributions

ZHS, QGG, and PM designed the experiments. ZHS, XYC, and XML performed all the experiments. ZHS and XYC analyzed the data. ZHS, QGG, and PM wrote the manuscript. All the authors reviewed the final manuscript.

Acknowledgements

We would like to thank professor Liqun Zhang from China Agricultural University whose comments and suggestions greatly improved the quality of this article.

References

1. Wang P, Guo Q, Ma Y, Li S, Lu X, Zhang X, Ma P. DegQ regulates the production of fengycins and biofilm formation of the biocontrol agent *Bacillus subtilis* NCD-2. Microbiological Research. 2015;178:42-50.
2. Fan H, Ru J, Zhang Y, Wang Q, Li Y. Fengycin produced by *Bacillus subtilis* 9407 plays a major role in the biocontrol of apple ring rot disease. Microbiological Research. 2017;199:89-97.
3. Wu Y, Wang Y, Zou H, Wang B, Sun Q, Fu A, Wang Y, Wang Y, Xu X, Li W. *Bacillus amyloliquefaciens* probiotic SC06 induces autophagy to protect against pathogens in macrophages. Frontiers in Microbiology. 2017;8:469.
4. Sonenshein AL. Control of sporulation initiation in *Bacillus subtilis*. Current Opinion in Microbiology. 2000;3(6):561-566.
5. Moszer I, Jones L, Moreira S, Fabry C, Danchin A. SubtiList: the reference database for the *Bacillus subtilis* genome. Nucleic Acids Research. 2002;30(1):62-65.
6. Torres MJ, Brandan CP, Sabate DC, Petroselli G, Errabalsells R, Audisio MC. Biological activity of the lipopeptide-producing *Bacillus amyloliquefaciens* PGPBacCA1 on common bean Phaseolus vulgaris L. pathogens. Biological Control. 2017;105:93-99.

7. Agustín L-B, Raunel T-V, Gerardo C, Kohei K, Katsuhiro K, Enrique G, Leobardo S-C. Effects of bacillomycin D homologues produced by *Bacillus amyloliquefaciens* 83 on growth and viability of Colletotrichum gloeosporioides at different physiological stages. Biological Control. 2018;127:145-154.
8. Stein T. Bacillus subtilis antibiotics: structures, syntheses and specific functions. Molecular Microbiology. 2005;56(4):845-857.
9. Kröber M, Wibberg D, Grosch R, Eikmeyer F, Verwaaijen B, Chowdhury SP, Hartmann A, Pühler A, Schlüter A. Effect of the strain *Bacillus amyloliquefaciens* FZB42 on the microbial community in the rhizosphere of lettuce under field conditions analyzed by whole metagenome sequencing. Frontiers in Microbiology. 2014;5:252-252.
10. Beppu T. Secondary metabolites as chemical signals for cellular differentiation. Gene. 1992;115(1-2):159-165.
11. Chaudhary AK, Dhakal D, Sohng JK. An insight into the "-omics" based engineering of streptomycetes for secondary metabolite overproduction. BioMed Research International. 2013;2013:968518-968518.
12. Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S, Fujita N. DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucleic Acids Research. 2012;41(D1):D408-D414.
13. Chen XH, Koumoutsi A, Scholz R, Borriss R. More than anticipated - production of antibiotics and other secondary metabolites by *Bacillus amyloliquefaciens* FZB42. Journal of Molecular Microbiology and Biotechnology. 2009;16(1-2):14-24.
14. Stein T, Vater J, Kruft V, Otto A, Wittmannliebold B, Franke P, Panico M, Mcdowell RA, Morris HR. The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. Journal of Biological Chemistry. 1996;271(26):15428-15435.
15. Du L, Lou L. PKS and NRPS release mechanisms. Natural Product Reports. 2010;27(2):255-278.
16. Arguellesarias A, Ongena M, Halimi B, Lara Y, Brans A, Joris B, Fickers P. *Bacillus amyloliquefaciens* GA1 as a source of potent antibiotics and other secondary metabolites for biocontrol of plant pathogens. Microbial Cell Factories. 2009;8(1):63-63.
17. Tamehiro N, Okamotohosoya Y, Okamoto S, Ubukata M, Hamada M, Naganawa H, Ochi K. Bacilysocin, a novel phospholipid antibiotic produced by *Bacillus subtilis* 168. Antimicrobial Agents and Chemotherapy. 2002;46(2):315-320.
18. Carrillo C, Teruel JA, Aranda FJ, Ortiz A. Molecular mechanism of membrane permeabilization by the peptide antibiotic surfactin. Biochimica et Biophysica Acta. 2003;1611(1-2):91-97.
19. Yu GY, Sinclair JB, Hartman GL, Bertagnolli BL. Production of iturin A by *Bacillus amyloliquefaciens* suppressing Rhizoctonia solani. Soil Biology & Biochemistry. 2002;34(7):955-963.
20. Jacques P, Hbid C, Destain J, Razafindralambo H, Paquot M, De Pauw E, Thonart P. Optimization of biosurfactant lipopeptide production from *Bacillus subtilis* S499 by plackett-burman design. Applied Biochemistry and Biotechnology. 1999;77-79:223-233.
21. Moyne A, Cleveland TE, Tuzun S. Molecular characterization and analysis of the operon encoding the antifungal lipopeptide bacillomycin D. FEMS Microbiology Letters. 2004;234(1):43-49.
22. Tulp M, Bohlin L. Rediscovery of known natural compounds: nuisance or goldmine? Bioorganic & Medicinal Chemistry. 2005;13(17):5274-5282.
23. Oman TJ, van der Donk WA. Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nature Chemical Biology. 2010;6(1):9-18.
24. Lane AL, Moore BS. A sea of biosynthesis: marine natural products meet the molecular age. Natural Product Reports. 2011;28(2):411-428.
25. Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. 3):509-523.

26. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends in genetics. 2014;30(9):418-426.
27. Guo Q, Li S, Lu X, Li B, Ma P. PhoR/PhoP two component regulatory system affects biocontrol capability of *Bacillus subtilis* NCD-2. Genetics and Molecular Biology. 2010;33(2):333-340.
28. Guo Q, Dong W, Li S, Lu X, Wang P, Zhang X, Wang Y, Ma P. Fengycin produced by *Bacillus subtilis* NCD-2 plays a major role in biocontrol of cotton seedling damping-off disease. Microbiological Research. 2014;169(7-8):533-540.
29. Zerbino D, Birney E. Velvet : algorithms for *de novo* short read assembly using de bruijn graphs. Genome Research. 2008;18(5):821-829.
30. Dunlap CA, Bowman MJ, Zeigler DR. Promotion of *Bacillus subtilis* subsp. *inaquosorum*, *Bacillus subtilis* subsp. *spizizenii* and *Bacillus subtilis* subsp. *stercoris* to species status. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology. 2020;113(1):1-12.
31. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Research. 2013;41(W1):W204-W212.
32. Sansinenea E, Ortiz A. Secondary metabolites of soil *Bacillus* spp. Biotechnology Letters. 2011;33(8):1523-1538.
33. Yu D, Fang Y, Tang C, Klosterman SJ, Tian C, Wang Y. Genomewide transcriptome profiles reveal how *Bacillus subtilis* lipopeptides inhibit microsclerotia formation in *Verticillium dahliae*. Molecular Plant-microbe Interactions. 2019;32(5):622-634.
34. Xiao X, Chen H, Chen H, Wang J, Ren C, Wu L. Impact of *Bacillus subtilis* JA, a biocontrol strain of fungal plant pathogens, on arbuscular mycorrhiza formation in *Zea mays*. World Journal of Microbiology & Biotechnology. 2008;24(7):1133-1137.
35. Challis GL, Ravel J. Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. FEMS Microbiology Letters. 2000;187(2):111-114.
36. Basichipalu S, Dischinger J, Josten M, Szekat C, Zweynert A, Sahl H, Bierbaum G. Pseudomycoicidin, a class II lantibiotic from *Bacillus pseudomycoides*. Applied and Environmental Microbiology. 2015;81(10):3419-3429.
37. Seydlová G, Svobodová J. Review of surfactin chemical properties and the potential biomedical applications. Central European Journal of Medicine. 2008;3(2):123-133.
38. Patel PS, Huang S, Fisher S, Pirnik D, Aklonis C, Dean L, Meyers E, Fernandes P, Mayerl F. Bacillaene, a novel inhibitor of procaryotic protein synthesis produced by *Bacillus subtilis*. Journal of Antibiotics. 2006;48(9):997.
39. Ramarathnam R, Bo S, Chen Y, Fernando WG, Xuewen G, de Kievit T. Molecular and biochemical detection of fengycin- and bacillomycin D-producing *Bacillus* spp., antagonistic to fungal pathogens of canola and wheat. Canadian Journal of Microbiology. 2007;53(7):901-911.
40. Miethke M, Klotz O, Linne U, May JJ, Beckering CL, Marahiel MA. Ferri-bacillibactin uptake and hydrolysis in *Bacillus subtilis*. Molecular Microbiology. 2006;61(6):1413-1427.
41. Thennarasu S, Lee DK, Poon A, Kawulka KE, Vederas JC, Ramamoorthy A. Membrane permeabilization, orientation, and antimicrobial mechanism of subtilosin A. Chemistry and Physics of Lipids. 2005;137(1-2):38-51.
42. Kenig M, Abraham EP. Antimicrobial activities and antagonists of bacilysin and anticapsin. Microbiology. 1976;94(1):37-45.
43. Fan B, Wang C, Song X, Ding X, Wu L, Wu H, Gao X, Borriss R. *Bacillus velezensis* FZB42 in 2018: The gram-positive model strain for plant growth promotion and biocontrol. Frontiers in Microbiology. 2018;9:2491.
44. Jin P, Wang H, Liu W, Miao W. Characterization of *lpaH2* gene corresponding to lipopeptide synthesis in *Bacillus amyloliquefaciens* HAB-2. BMC Microbiology. 2017;17(1):227.

45. Chen X, Koumoutsi A, Scholz R, Eisenreich A, Schneider K, Heinemeyer I, Morgenstern B, Voss B, Hess W, Reva O *et al*. Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. Nature Biotechnology. 2007;25(9):1007-1014.
46. Koumoutsi A, Chen X, Henne A, Liesegang H, Hitzeroth G, Franke P, Vater J, Borriss R. Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. Journal of Bacteriology. 2004;186(4):1084-1096.
47. Chen C, Chang L, Chang Y, Liu S, Tschen JS. Transposon mutagenesis and cloning of the genes encoding the enzymes of fengycin biosynthesis in *Bacillus subtilis*. Molecular Genetics and Genomics. 1995;248(2):121-125.
48. Lin G, Chen C, Tschen JS, Tsay S, Chang Y, Liu S. Molecular cloning and characterization of fengycin synthetase gene *fenB* from *Bacillus subtilis*. Journal of Bacteriology. 1998;180(5):1338-1341.
49. Lin TP, Chen CL, Chang LK, Tschen JS, Liu ST. Functional and transcriptional analyses of a fengycin synthetase gene, fenC, from *Bacillus subtilis*. Journal of Bacteriology. 1999;181(16):5060-5067.
50. Lautru S, Deeth RJ, Bailey LM, Challis GL. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. Nature Chemical Biology. 2005;1(5):265-269.
51. Bie XM, Lü FX, Lu ZX, Huang XQ, Shen J. [Isolation and identification of lipopeptides produced by *Bacillus subtilis* fmbJ]. Sheng wu gong cheng xue bao = Chinese journal of biotechnology. 2006;22(4):644-649.
52. Tripathi L, Irorere VU, Marchant R, Banat IM. Marine derived biosurfactants: a vast potential future resource. Biotechnology Letters. 2018;40(11-12):1441-1457.
53. Sivapathasekaran C, Mukherjee S, Samanta R, Sen R. High-performance liquid chromatography purification of biosurfactant isoforms produced by a marine bacterium. Analytical and Bioanalytical Chemistry. 2009;395(3):845-854.
54. Pan H, Tian X, Shao M, Xie Y, Huang H, Hu J, Ju J. Genome mining and metabolic profiling illuminate the chemistry driving diverse biological activities of *Bacillus siamensis* SCSIO 05746. Applied Microbiology and Biotechnology. 2019;103(10):4153-4165.
55. Ongena M, Jacques P. *Bacillus lipopeptides*: versatile weapons for plant disease biocontrol. Trends in Microbiology. 2008;16(3):115-125.
56. Bhat A, Chakraborty R, Adlakha K, Agam G, Chakraborty K, Sengupta S. Ncl1-mediated metabolic rewiring critical during metabolic stress. Life Science Alliance. 2019;2(4).
57. Bae J, Park J, Hahn M, Kim M, Roe J. Redox-dependent changes in RsrA, an anti-sigma factor in *Streptomyces coelicolor*: Zinc release and disulfide bond formation. Journal of Molecular Biology. 2004;335(2):425-435.
58. Landy M, Warren GH. Bacillomycin; an antibiotic from *Bacillus subtilis* active against pathogenic fungi. Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine (New York, NY). 1948;67(4):539-541.
59. Wu L, Wu H, Chen L, Xie S, Zang H, Borriss R, Gao X. Bacilysin from *Bacillus amyloliquefaciens* FZB42 has specific bactericidal activity against harmful algal bloom species. Applied and Environmental Microbiology. 2014;80(24):7512-7520.
60. Li Y, Jiang W, Gao R, Cai Y, Guan Z, Liao X. Fe(III)-based immobilized metal-affinity chromatography (IMAC) method for the separation of the catechol siderophore from CD36. 3 Biotech. 2018;8(9):392.
61. Shelburne C, An F, Dholpe V, Ramamoorthy A, Lopatin D, Lantz M. The spectrum of antimicrobial activity of the bacteriocin subtilosin A. The Journal of Antimicrobial Chemotherapy. 2007;59(2):297-300.
62. Karim A, Poirot O, Khatoon A, Aurongzeb M. Draft genome sequence of a novel *Bacillus glycinifermentans* strain having antifungal and antibacterial properties. Journal of Global Antimicrobial Resistance. 2019;19:308-310.
ol for high throughput sequence data. 2010.

64. Alexey G, Vladislav S, Nikolay V, Glenn T: QUAST: quality assessment tool for genome assemblies. In*: 2013*; 2013: 1072-1075.
65. Torsten S. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-2069.
66. Disz T, Akhter S, Cuevas DA, Olson R, Overbeek R, Vonstein V, Stevens R, Edwards R. Accessing the SEED genome databases via web services API: tools for programmers. BMC Bioinformatics. 2010;11(1):319-319.
67. Aziz RK, Bartels D, Best AA, Dejongh M, Disz T, Edwards R, Formsma K, Gerdes S, Glass EM, Kubal M. The rast server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
68. Haubold B, Klotzl F, Pfaffelhuber P. andi: fast and accurate estimation of evolutionary distances between closely related genomes. Bioinformatics. 2015;31(8):1169-1175.
69. Tamura K, Peterson DS, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution. 2011;28(10):2731-2739.
70. Xu X, Reid N. On the robustness of maximum composite likelihood estimate. Journal of Statistical Planning and Inference. 2011;141(9):3047-3054.
71. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology. 2019;37(4):420-423.
72. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Research. 2018;46(W1):W246-W251.
73. Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Research. 2011;39(2):W339-W346.
74. Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster ALH, Wyatt MA, Magarvey NA. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). Nucleic Acids Research. 2015;43(20):9645-9662.
75. Bachmann BO, Ravel J. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. Methods in Enzymology. 2009;458:181-217.
76. Byers H, Stackebrandt E, Hayward C, Blackall LL. Molecular investigation of a microbial mat associated with the Great Artesian Basin. FEMS Microbiology Ecology. 1998;25(4):391-403.
77. Li B, Lu X, Guo Q, Qian C, Li S, Ma P. Isolation and identification of lipopeptides and volatile compounds produced by *Bacillus subtilis* strain BAB-1, vol. 43; 2010.
78. Reddick J, Antolak S, Raner G. PksS from *Bacillus subtilis* is a cytochrome P450 involved in bacillaene metabolism. Biochemical and Biophysical Research Communications. 2007;358(1):363-367.

Tables

Table 1 Secondary metabolite gene clusters annotated in *B. subtilis* NCD-2 using antiSMASH

Cluster	Type	From	To	Most similar known cluster	Similarity	MIBiG BGC-ID*
cluster 1	NRPS	347853	413245	surfactin	82%	BGC0000433_c1
cluster 2	Terpene	1137768	1158574	-	-	-
cluster 3	NRPS-TransAT PKS-Other KS	1763940	1873766	bacillaene	100%	BGC0001089_c1
cluster 4	NRPS	1936035	2004508	fengycin	93%	BGC0001095_c1
cluster 5	Terpene	2060609	208250	-	-	-
cluster 6	T3PKS	2261562	2302659	-	-	-
cluster 7	NRPS	3225454	3275189	bacillibactin	100%	BGC0000309_c1
cluster 8	Sactipeptide-head to tail	3817363	3838974	subtilosin	100%	BGC0000602_c1
cluster 9	Other	3842273	3883691	bacilysin	100%	BGC0001184_c1

*Identification numbers of the most similar gene clusters from *B. velezensis* FZB42 provided by the MIBiG BGC database. NRPS, non-ribosomal peptide synthetase; PKS, polyketide synthase; T3PKS, type III polyketide; NRPS-TransAT PKS-Other KS, non-ribosomal peptide synthetase-trans-AT polyketide synthase-Other types of polyketide synthase cluster; Sactipeptide-head to tail, head-to-tail cyclised peptide.

Table 2 Fengycin homologs in NCD-2 based on the key product ions of β-OH-FA with different chain lengths.

Fengycin family	$[M+2H]^{2+}$	β-hydroxy fatty acid
fengycin A	718.4, 725.4, 732.4, 739.4, 745.4, 753.4	C14-C19
fengycin B	718.4, 725.4, 732.4, 739.4, 746.4, 753.4, 760.4, 767.4	C12-C19
fengycin A2	718.4, 725.4, 732.4, 739.4	C15-C18
fengycin B2	725.4, 732.4, 739.4, 746.4, 753.4	C14-C18
fengycin C	760.4, 767.4, 774.5	C18-C20

Additional Files

Additional file 1: Supplemental Material.

Files contain supplemental materials, including the supplementary tables and figures referenced in this manuscript. Fig. S1. Fengycin biosynthetic gene clusters of different strains that have a close relation with NCD-2 or the model strains. Fig. S2. Surfactin biosynthetic gene clusters of different strains that have a close relation with NCD-2 or the model

rated from the crude methanolic extract using an AKTA Purifier. Fig. S4.

Primary structures of fengycins and surfactins. Fig. S5. Fengycin A of a β-OH FA with a chain length varying from C14 to C19 identified based on key product ions. Fig. S6. Fengycin B of a β-OH FA with a chain length varying from C12 to C19 identified based on key product ions. Fig. S7. Fengycin A2 of a β-OH FA with a chain length varying from C15-C18 identified based on key product ions. Fig. S8. Fengycin B2 of a β-OH FA with a chain length varying from C14-C18 identified based on key product ions. Fig. S9. Fengycin C of a β-OH FA with a chain length varying from C18-C20 identified based on key product ions. Fig. S10. Surfactin of a fatty acid with a chain length varying from C11-C15 identified based on key product ions. Table S1. All *B. subtilis* strains with the assembly level of chromosome and their RefSeq assembly accessions. Table S2. Homologues of FenC of FZB42 detected by scanning the local NCD-2 proteome in BioEdit. Table S3. Homologues of FenD of FZB42 detected by scanning the local NCD-2 proteome in BioEdit. Table S4. Adenylation domain binding amino acids predicted by PRISM.

Supplementary Fig. S1 Fengycin BGCs of different strains that have a close relation with NCD-2 or model strains. FZB42 belongs to *B. velezensis* and others were *B. subtilis*. Different colour blocks represent genes with conserved functions: taking the FZB42 strain as an example, the fengycin biosynthetic gene cluster includes the genes *fenCDEAB* (also named *ppsABCDE*) in order from right to left. In strain NCD-2, only *fenEAB* genes are found, which was particularly different from the other strains.

Supplementary Fig. S2 Surfactin BGCs of different strains that have a close relation with NCD-2 or model strains. FZB42 belongs to *B. velezensis* and the others were *B. subtilis*. Different colour blocks represent genes with conserved functions: taking the FZB42 strain as an example, the surfactin biosynthetic gene cluster includes the genes *srfAABCD* in order from left to right. In strain NCD-2, *srfAB* is divided into two genes for transcription and translation, which was different from other strains.

Supplementary Fig. S3 Elution of lipopeptides separated from the crude methanolic extract of *B. subtilis* NCD-2. An AKTA Purifier (GE Healthcare, Uppsala, Sweden) with the SOURCE 5RPC ST 4.6/150 column was used, and the fractions (P2-P25) are shown above the peaks. Fractions 12, 13, 14, 15 are fengycins, and fractions 19, 20 are surfactin.

Supplementary Fig. S4 Primary structures of fengycins and surfactins. (a) The overall structure of fengycins; (b) fengycin A, fengycin B, fengycin A2, fengycin B2, and fengycin C. Sites of mass spectrometric cleavage with the *m/z* values for the diagnostic product ions (α and β) are indicated; (c) the overall structure of surfactins.

Supplementary Fig. S5 Fengycin A were identified. β-OH FA with the chain length varied from C14 to C19 based on key product ions ($[M+2H]^{2+}$: 718.4, 725.4, 732.4, 739.4, 745.4, and 753.4).

Supplementary Fig. S6 Fengycin B were identified. β-OH FA with the chain length varied from C12 to C19 were identified based on key product ions ($[M+2H]^{2+}$: 718.4, 725.4, 732.4, 739.4, 746.4, 753.4, 760.4, and 767.4).

Supplementary Fig. S7 Fengycin A2 were identified. β-OH FA with the chain length varied from C15-C18 were identified based on key product ions ($[M+2H]^{2+}$: 718.4, 725.4, 732.4, and 739.4).

Supplementary Fig. S8 Fengycin B2 were identified. β-OH FA with the chain length varied from C14-C18 were identified based on key product ions ($[M+2H]^{2+}$: 725.4, 732.4, 739.4, 746.4, and 753.4).

Supplementary Fig. S9 Fengycin C were identified. β-OH FA with the chain length varied from C18-C20 were identified based on key product ions ($[M+2H]^{2+}$: 760.4, 767.4, and 774.5).

Supplementary Fig. S10 Surfactin were identified. β-OH FA with the chain length varied from C11-C15 were identified based on key product ions ($[M+H]^{+}$: 994.6, 1008.7, 1022.7 and 1036.7).

Supplementary Tables

Table S1 All *B. subtilis* strain with the assembly level of complete genome or chromosome and their RefSeq assembly accession.

strain	RefSeq assembly accession	strain	RefSeq assembly accession	strain	RefSeq assembly accession
168	GCF_000155325.1	SRCM103571	GCF_004103595.1	NBRC 13719	GCF_006741845.1
BEST7003	GCF_000523045.1	SRCM103576	GCF_004119615.1	RO-NN-1	GCF_000227485.1
BSn5	GCF_000186745.1	SRCM103581	GCF_004119655.1	AG1839	GCF_000699525.1
BS49Ch	GCF_000953615.1	SRCM103612	GCF_004119775.1	BAB-1	GCF_000349795.1
HJ5	GCF_000973605.1	SRCM103622	GCF_004119835.1	BSP1	GCF_000321395.1
KCTC 1028	GCF_000971925.1	SRCM103629	GCF_004119815.1	AG174	GCF_000699465.1
PY79	GCF_000497485.1	SRCM103637	GCF_004119875.1	NCIB 3610	GCF_000186085.1
QB928	GCF_000293765.1	SRCM103641	GCF_004119555.1	OH 131.1	GCF_000706705.1
50-1	GCF_003184225.1	SRCM103689	GCF_004119535.1	2KL1	GCF_003665395.1
7702	GCF_002272405.1	SRCM103696	GCF_004119595.1	2RL2-3	GCF_003665275.1
ATCC 11774	GCF_004101945.1	SRCM103697	GCF_004119635.1	3NA	GCF_000827065.1
ATCC 13952	GCF_000772125.1	SRCM103773	GCF_004119675.1	168G	GCF_001703495.1
ATCC 19217	GCF_000772165.1	SRCM103835	GCF_004119715.1	BSD-2	GCF_001465815.1
ATCC 21228	GCF_002982175.1	SRCM103837	GCF_004119695.1	CU1050	GCF_001541905.1
B-1	GCF_000769515.1	SRCM103862	GCF_004101345.1	D12-5	GCF_001596535.1
BJ3-2	GCF_002893805.1	SRCM103881	GCF_004101445.1	delta6	GCF_001660525.1
Bs-916	GCF_000772205.1	SRCM103886	GCF_004101365.1	G7	GCF_004328925.1
BS16045	GCF_001720505.1	SRCM103923	GCF_004101405.1	GFR-12	GCF_003665195.1
CW14	GCF_002163815.1	SRCM103971	GCF_004101465.1	IITK SM	GCF_003426125.1
DKU_NT_02	GCF_002269175.1	SRCM104005	GCF_004101425.1	KCTC 3135	GCF_001697265.1
DKU_NT_03	GCF_002269195.1	SRCM104008	GCF_004101485.1	MH-1	GCF_003665235.1
FDAARGOS_606	GCF_006364495.1	SRCM104011	GCF_004101565.1	N1-1	GCF_003665335.1
ge28	GCF_002202055.1	SX01705	GCF_002216085.1	N2-2	GCF_003665315.1
GS 188	GCF_002220075.1	SZMC 6179J	GCF_001604995.1	N3-1	GCF_003665355.1
H19	GCF_005234095.1	TLO3	GCF_002290305.1	N4-2	GCF_003665295.1
HJ0-6	GCF_001704095.1	TO-A JPC	GCF_001037985.1	PJ-7	GCF_003665215.1
MBI 600	GCF_005160425.1	UD1022	GCF_001015095.1	SRCM100333	GCF_002201995.1
MZK05	GCF_003612735.1	WB800N	GCF_003610955.1	SRCM100757	GCF_002173715.1
NRS 231	GCF_005153965.1	DE111	GCF_001534785.1	SRCM100761	GCF_002201955.1
PR10	GCF_005849145.1	KCTC 13429	GCF_003148415.1	SRCM101392	GCF_002202035.1
PS832	GCF_000789295.1	BEST195	GCF_000209795.2	SRCM101441	GCF_002173615.1
QB61	GCF_002148255.1	CGMCC 2108	GCF_001565875.1	SRCM101444	GCF_002173695.1

SEM-9	GCF_006165085.1	ATCC 6633	GCF_006094475.1	SSJ-1	GCF_003665255.1
SG6	GCF_000782835.1	W23	GCF_000146565.1	XF-1	GCF_000338735.1
SRCM103517	GCF_004103535.1	TU-B-10	GCF_000227465.1	NCD-2	GCF_002556525.1
SRCM103551	GCF_004103555.1	6051-HGW	GCF_000344745.1		

Table S2 The homologues of FenC of FZB42 detected by scanning the local NCD-2 proteome in BioEdit.

protein number	score	similarity	E-value	function description
Gms1961	2701	55	0.0	FenE
Gms1960	2036	43	0.0	FenA
Gms0365	1639	37	0.0	SrfAA
Gms0366	1296	34	0.0	SrfAB
Gms3368	1127	34	0.0	DhbF
Gms1826	572	27	e-164	PKSJ
Gms1829	516	30	e-147	PKSN
Gms0367	489	39	e-138	Surfactin synthase subunit 2
Gms1959	478	30	e-135	FenB
Gms0368	462	29	e-130	SrfAC
Gms4064	234	32	6e-062	DltA

Table S3 The homologues of FenD of FZB42 detected by scanning the local NCD-2 proteome in BioEdit.

protein number	score	similarity	E-value	function description
Gms1960	1719	38	0.0	FenA
Gms0365	1715	37	0.0	SrfAA
Gms1961	1706	39	0.0	FenE
Gms0366	1481	35	0.0	Surfactin synthase subunit 1
Gms3368	1179	35	0.0	DhbF
Gms1959	814	41	0.0	FenB
Gms0368	739	38	0.0	SrfAC
Gms1826	620	27	e-178	PKSJ
Gms1829	560	32	e-160	PKSN
Gms0367	519	40	e-148	Surfactin synthase subunit 2

Table S4 Adenylation domain binding amino acids predicted by PRISM.

Gms1961	A domain A9	Val	Ile	Leu	Val	Phe	Asp	Tyr	Glu	Ala	N5-hydroxy-Orn
	score	943.0	595.0	566.0	556.5	515.2	483.3	482.8	482.7	475.2	109.9
Gms1959	A domain A13	Ile	Val	Val	Leu	Phe	Ala	Tyr	Leu	Glut	β-Phe
	score	819.1	645.3	535.3	508.4	469.7	430.9	424.1	421.7	402.9	109.1

Predicted by PRISM (http://grid.adapsyn.com/prism/). The scores represent the ability of the adenylation domain binding amino acids.

Figures

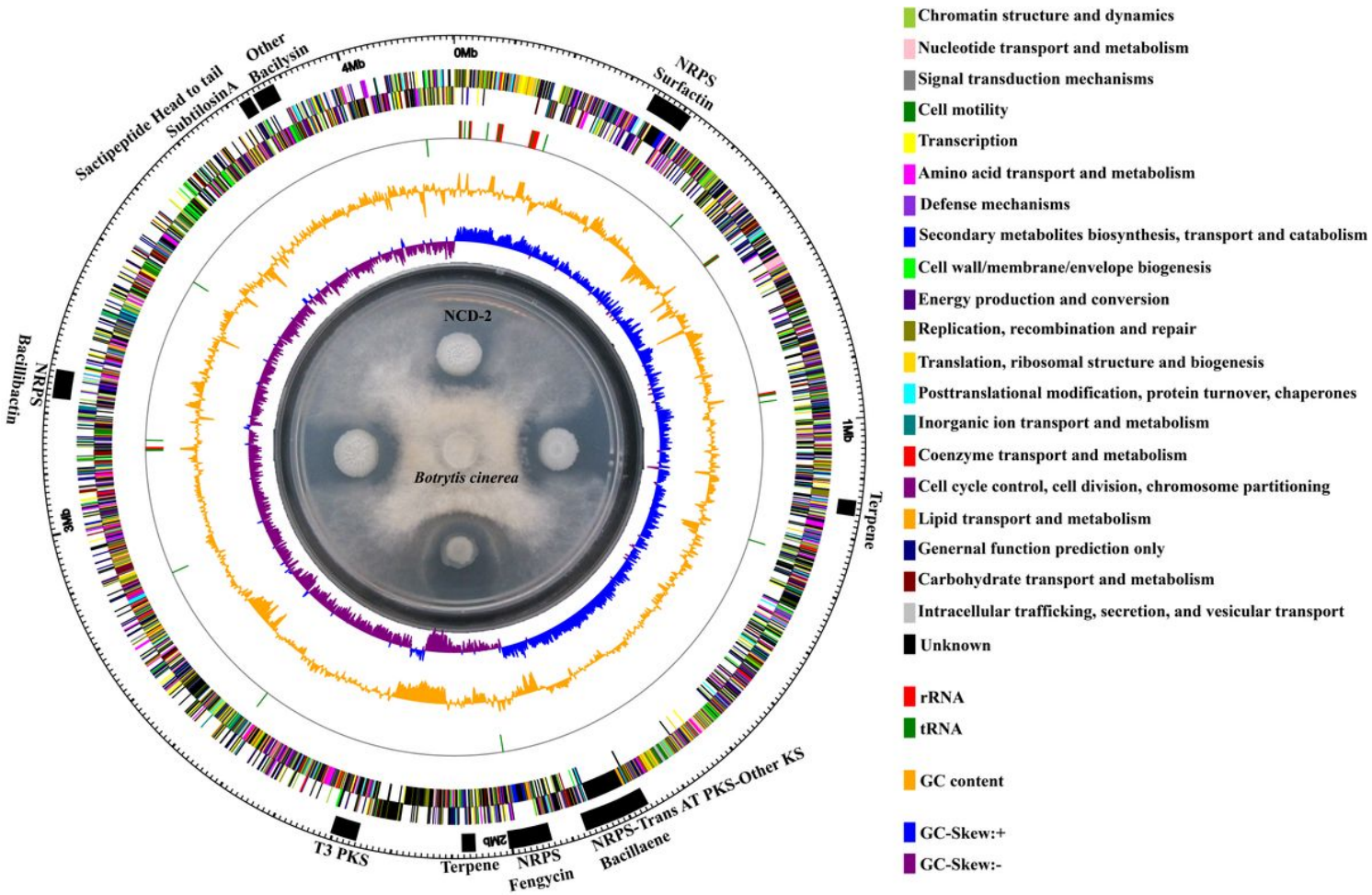

Figure 1

Circular genome of strain NCD-2 with specific features. The circular genome map was created using Circos v0.66 with COG (Cluster of Orthologous Groups of proteins) function annotation. From outside to inside: circle 1, the size of the complete genome; circles 2 and 3, the predicted protein-coding genes on the + and - strands, respectively, where different

colours represent different COG function classifications; circle 4, tRNA (green) and rRNA (red); circle 5, G + C content, where peaks outside/inside the circle indicate above or below average GC content, respectively; the inner circle, G + C skew, with G% < C% in purple and with G% > C% in blue. The potato dextrose agar plate inside the representation of the circular genome shows the antifungal activity of strain NCD-2 and its derived strain constructed by atmospheric and room temperature plasma (ARTP) against Botrytis cinerea. The black bars outside the circular genome indicate the secondary metabolite biosynthetic gene clusters.

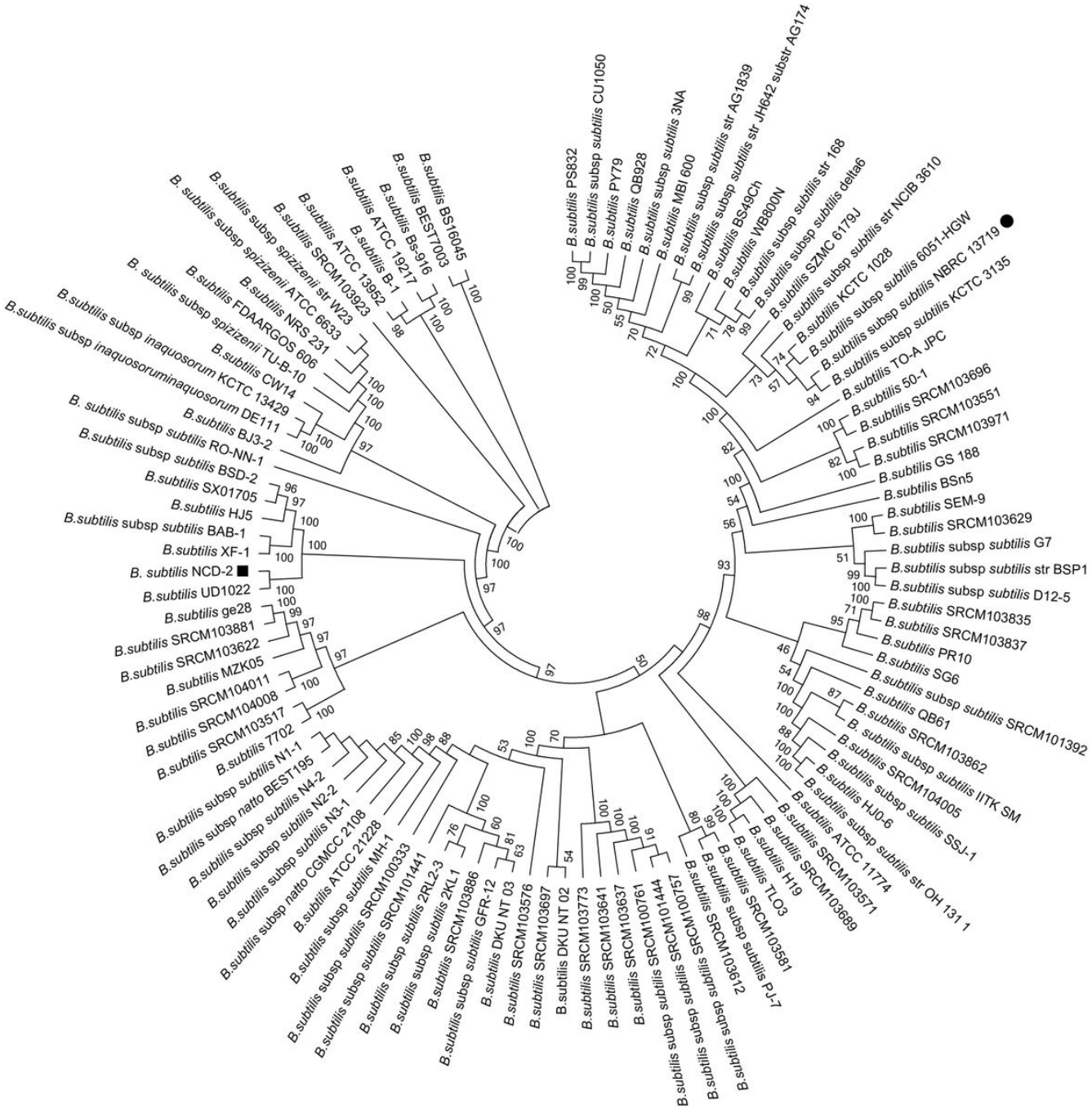

Figure 2

Phylogenetic tree of 113 B. subtilis strains based on whole genome alignments. The position of strain NCD-2 in the phylogenetic tree is indicated by a black square mark, and the position of the reference strain B. subtilis NBRC 13719 is indicated by a black circle mark. Single Nucleotide Polymorphisms (SNPs) and short insertions or deletions (indels) within the multiple sequence alignments constructed by the REALPHY pipeline were extracted for subsequent phylogeny

reconstruction. The phylogenetic tree was constructed using MEGA 5.0 by the Neighbor-joining method, with a bootstrap of 1,000 replications. Bootstrap confidence levels > 50% are indicated at the internodes.

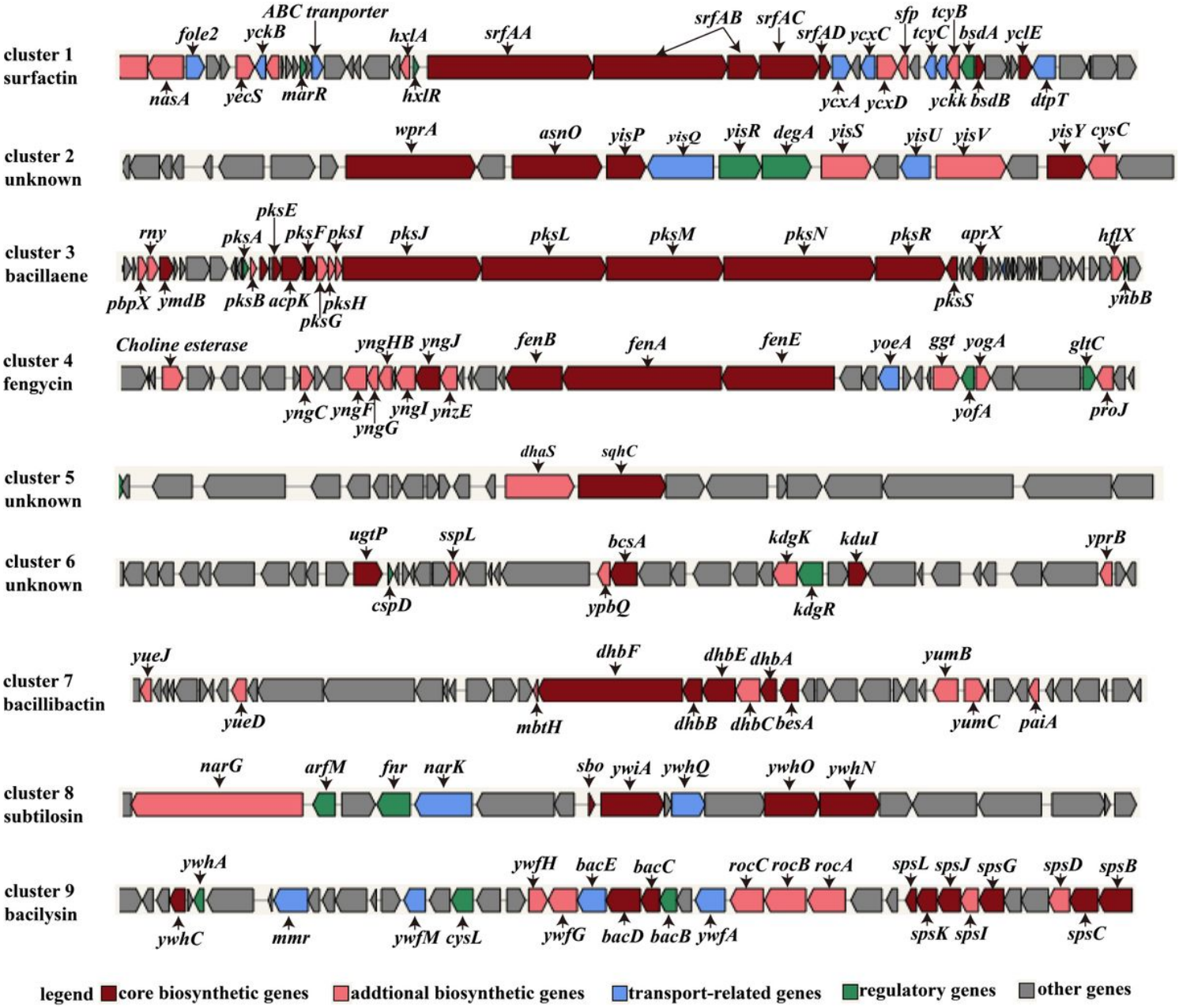

Figure 3

Schematic diagram of nine secondary metabolite biosynthetic gene clusters in B. subtilis strain NCD-2. antiSMASH was used to predict potential secondary metabolite biosynthetic gene clusters. Different colour blocks represent genes with different functions; the genes marked with dark red, light red, blue, green, and gray are core biosynthetic, additional biosynthetic, transport-related, regulatory, and other genes, respectively.

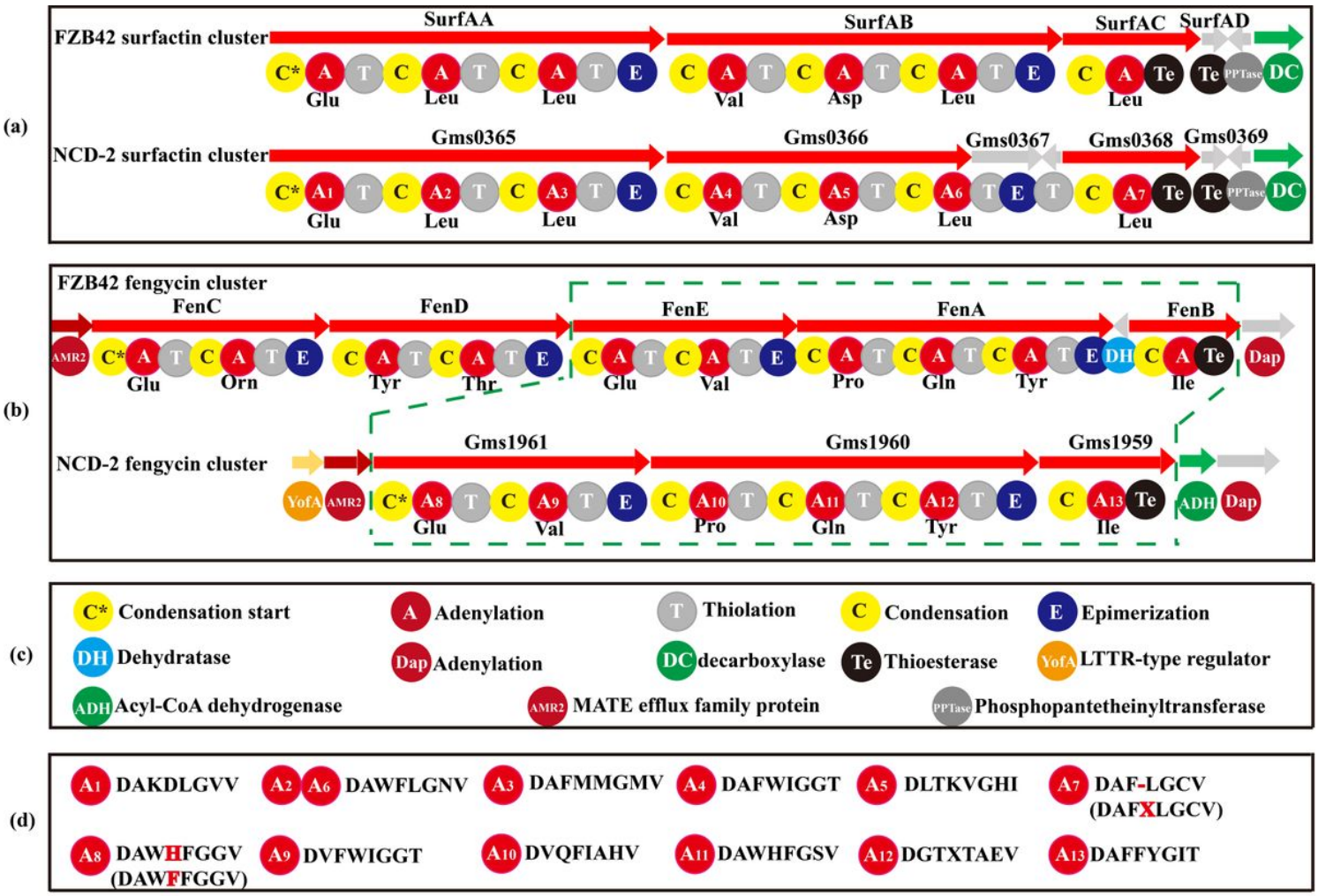

Figure 4

Comparisons of functional domains of core genes involved in synthesizing surfactin and fengycin in NCD-2. The functional domains of the core genes of clusters 1 (a) and 3 (b) in B. subtilis NCD-2. (c) The abbreviations indicate the functions of the corresponding structural domains. (d) The conserved binding pockets for substrates formed by amino acids in different adenylation domains.

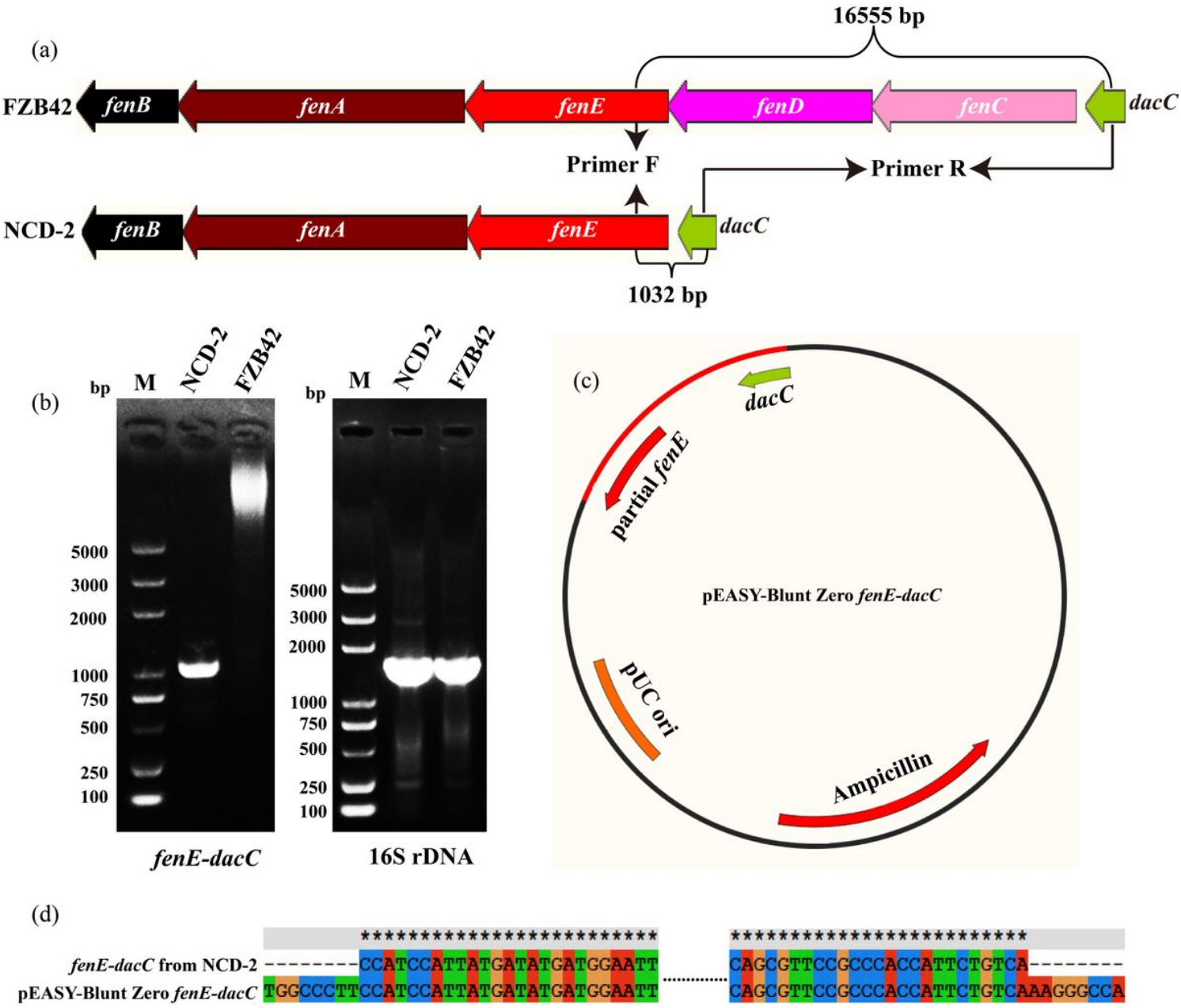

Figure 5

PCR and sequence of the fragment between fenE and dacC. (a) Schematic diagram used to design primers according to conserved bases from NCD-2 and B. velezensis FZB42; (b) PCR of the fenE-dacC fragment using the genomic templates NCD-2 and FZB42, with 16S rDNA as an internal reference control; (c) Schematic diagram of the constructed sequencing vector by ligating the fenE-dacC fragment to the pEASY-Blunt Zero vector, (d) BLAST of the fenE-dacC fragments from NCD-2 and pEASY-Blunt Zero fenE-dacC, in which the two sequences of fenE-dacC were complete same.

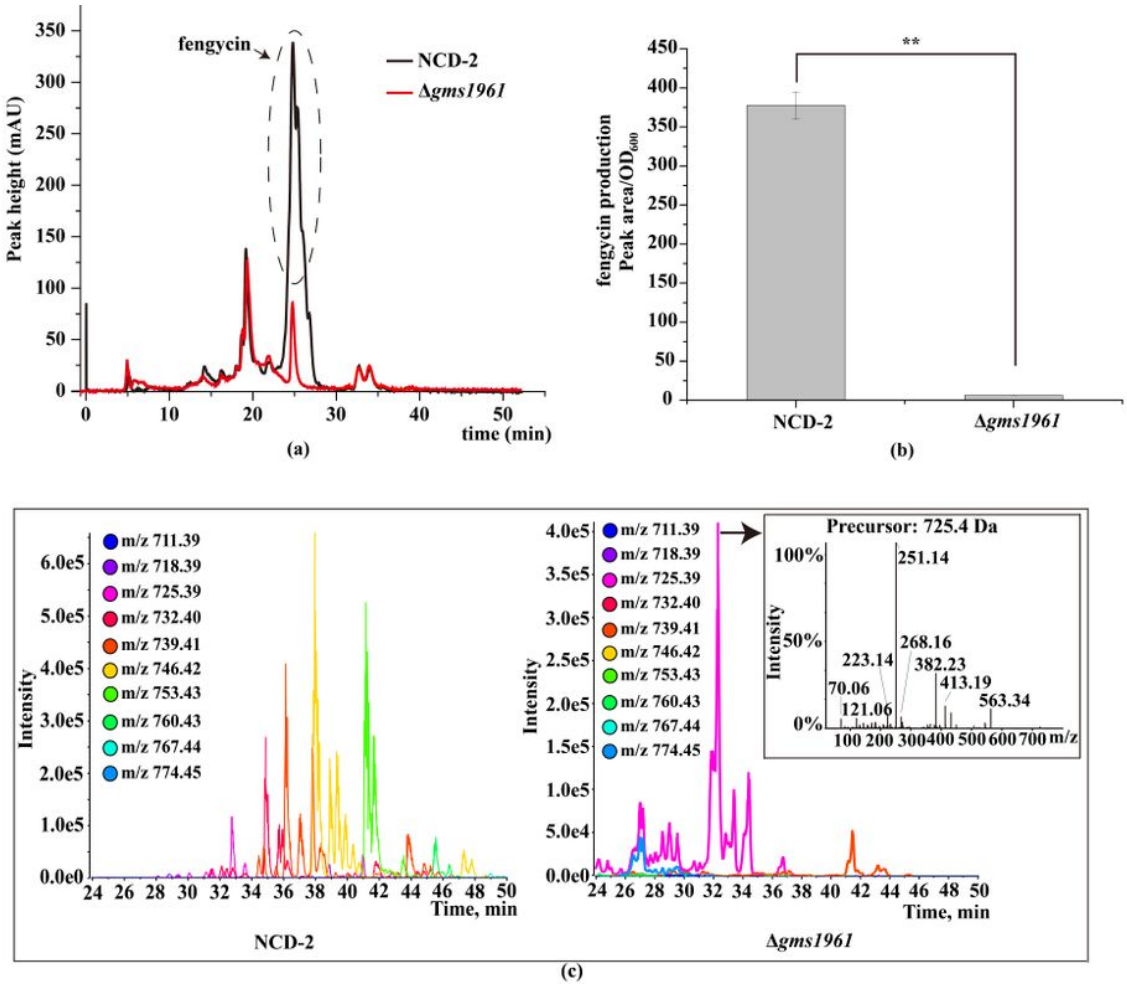

Figure 6

The role of gms1961 in synthesizing fengycin. (a) FPLC of the lipopeptides of strain NCD-2 and Δgms1961, (b) quantitative production of fengycin in strain NCD-2 and Δgms1961, where the error bars represent the standard deviation and asterisks depict significant differences as measured by the t-test (**p< 0.01), (c) Extract Ions Using Dialog (XIC) and UHPLC-QTOF-MS of fengycin from NCD-2 and Δgms1961. The lipopeptide fengycin exhibited a difference at 25-50 min between strains NCD-2 and Δgms1961, and only the precursor related to m/z 725.4 was same (the light purple line), but the fragments were absolutely different from those of fengycin from strain NCD-2.

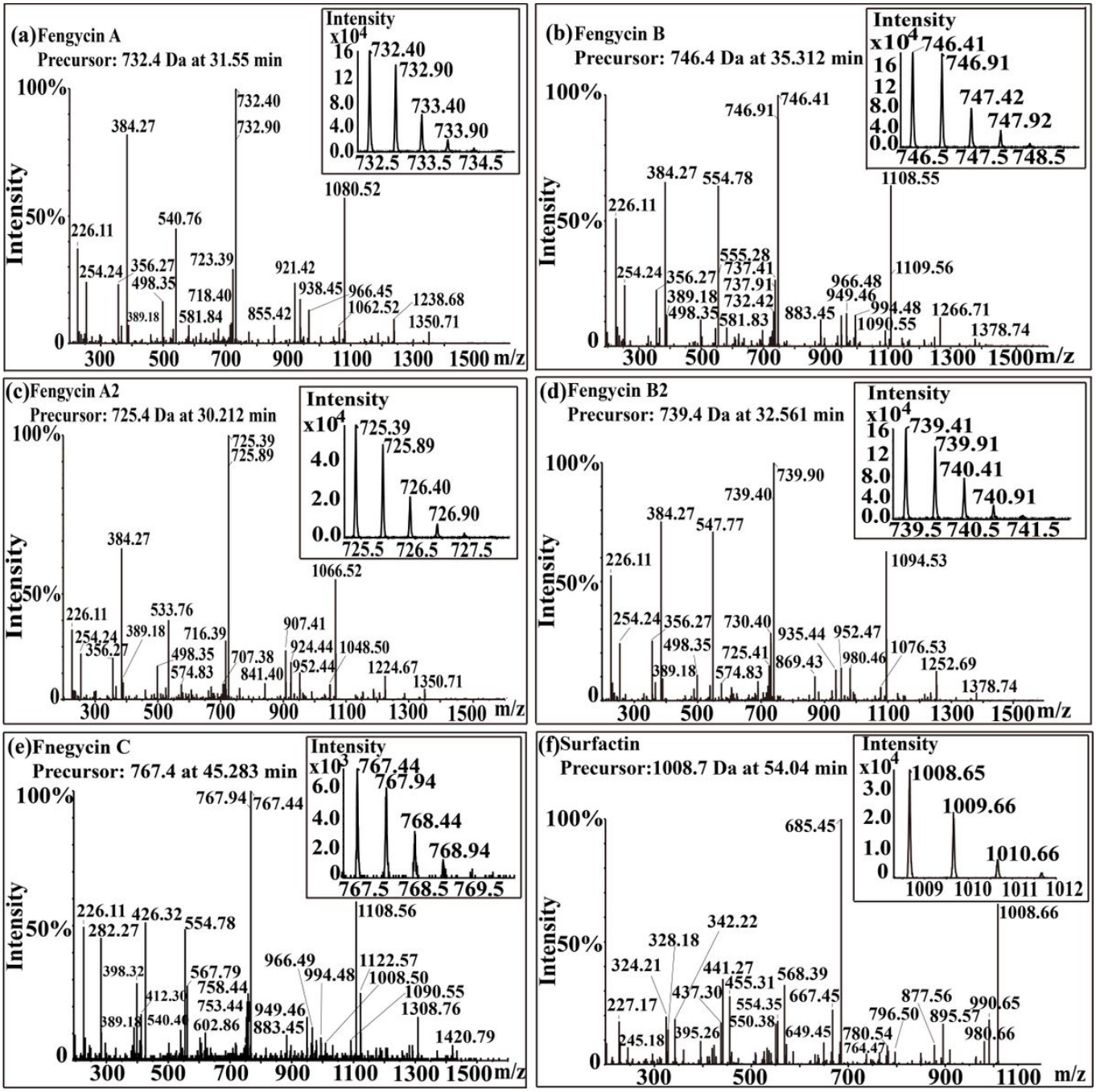

Figure 7

MS/MS spectra of protonated cyclic fengycin and surfactin ions. (a) m/z 732.4, (b) m/z 746.4, (c) m/z 725.4, (d) m/z 739.4, (e) m/z 767.4, and (f) m/z 1,008.7.

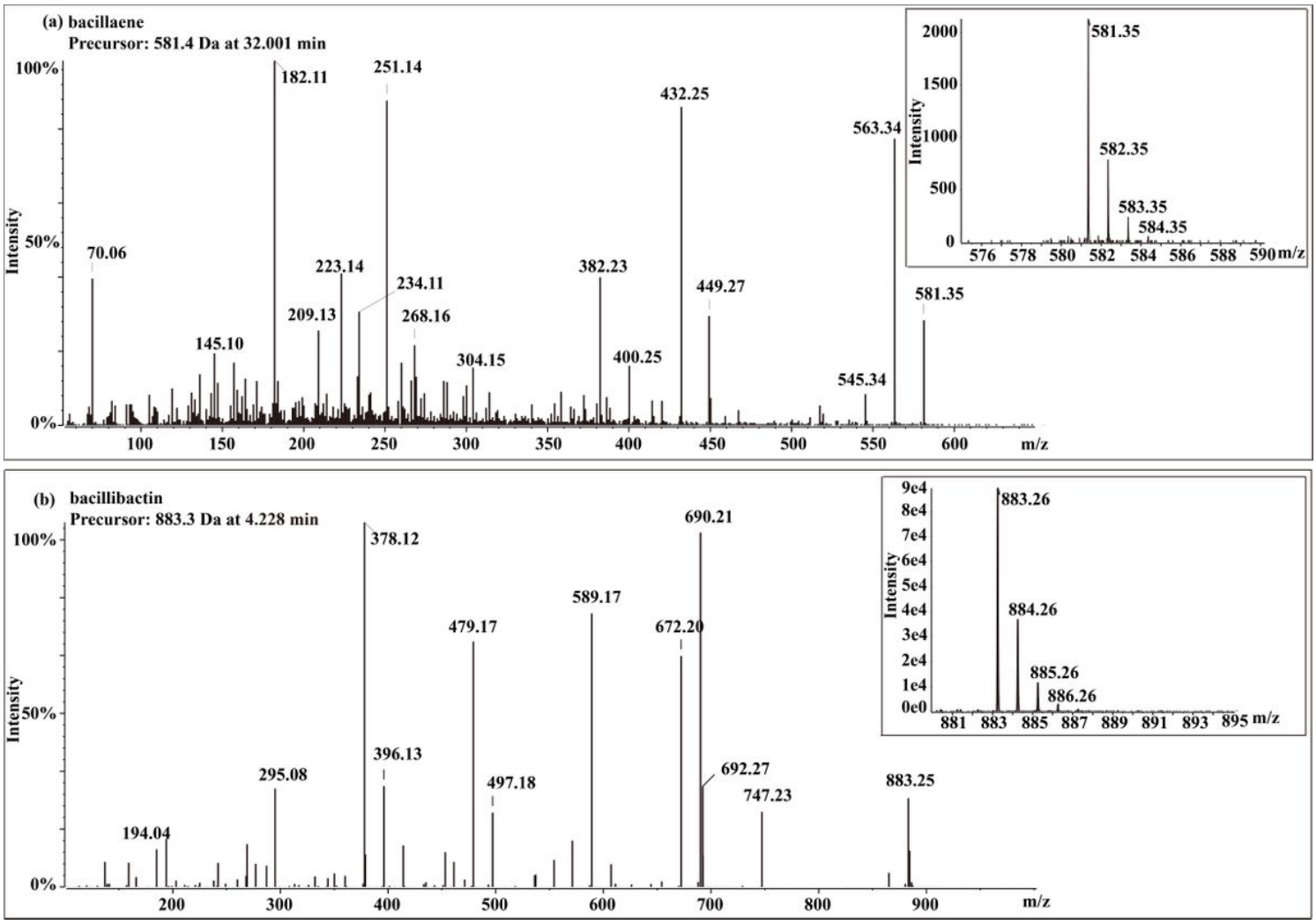

Figure 8

MS/MS spectra of protonated cyclic bacillaene and bacillibactin ions. (a) m/z 581.4, (b) m/z 883.3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Fig.S1.jpg
- Fig.S2.jpg
- Fig.S3.jpg
- Fig.S4.jpg
- Fig.S5.jpg
- Fig.S6.jpg
- Fig.S7.jpg
- Fig.S8.jpg
- Fig.S9.jpg
- Fig.S10.jpg