Genome mining and UHPLC-MS/MS illuminate the specificity of secondary metabolite synthetic 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

DOI: https://doi.org/10.21203/rs.3.rs-33091/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
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 all the secondary metabolite synthetic gene clusters and related bioactive compounds in NCD-2. An integrative approach, which coupled genome mining with structural identification technologies using ultra-high-performance liquid chromatography coupled to quadrupole time-of-flight tandem mass spectrometry (UHPLC-MS/MS), was conducted to interpret the chemical origins of the significant biological activities in NCD-2.

Results

Genome mining revealed that NCD-2 contained nine gene clusters having predicted functions involving secondary metabolites with bioactive abilities. They encoded six known products—fengycin, surfactin, bacillaene, subtilosin, bacillibactin, and bacilysin—as well as three unknown products. Interestingly, the synthetic gene clusters for surfactin and fengycin showed 83% and 92% amino acid sequence similarity levels with the corresponding products in *Bacillus velezensis* strain FZB42. A further comparison of gene clusters encoding fengycin and surfactin revealed that strain NCD-2 had lost the *fenC* and *fenD* genes in the fengycin biosynthetic operon, and that the surfactin synthetic enzyme-related gene *srfAB* was divided into two parts. A bioinformatics analysis showed that *fenEA* may function as *fenCD* in synthesizing fengycin and that the structure of this fengycin synthetic gene cluster is likely unique to NCD-2. Five kinds of fengycin, with 26 homologs, and surfactin, with 4 homologs, were detected from strain NCD-2, which indicated the non-typical and unique nature of this fengycin biosynthetic gene cluster. To the best of our knowledge, this is the first report of a non-typical gene cluster related to fengycin synthesis.

Conclusions

The data provide the genetic characteristics of secondary metabolite synthetic gene clusters for fengycin and surfactin in *B. subtilis* NCD-2, including the unique synthetic mechanism for fengycin, and suggest that bioactive secondary metabolites explain the biological activities of NCD-2.

Background

*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[1]. The *Bacillus* genus has received considerable attention as a biological resource used in the development of microbial pesticides, in part because its members form stress-resistant spores that do not harm the environment and are useful in pesticide production[2–4]. The mechanisms used by *B. subtilis* to suppress plant soil-borne diseases include competing with phytopathogens for nutrients and spatial sites, inducing the systematic
resistance of plants, and inhibiting pathogen growth by producing antimicrobial compounds [5]. The latter is a general characteristic of B. subtilis' biocontrol capability and plays an important role in the biological control of plant diseases [6, 7]. B. subtilis produces more than two dozen antimicrobial compounds having amazing structural variety. On the basis of the biosynthetic pathway, 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 polyketones [8]. Most antimicrobial compounds are secondary metabolites, with very complex chemical structures, that are not necessary for the growth and reproduction of microorganisms. Secondary metabolites function as essential chemical signals for the induction of cellular differentiation in the producing organism and for controlling its metabolism [9, 10]. They also function as antibiotics, and their antimicrobial properties may lead to shifts within rhizospheric microbial functional subsystems, such as affecting the availability of nutrients for the plant [11].

The genes encoding the secondary metabolites commonly exist in clusters and encode complex enzymes with multiple functions [12]. The polyketide synthase/non-ribosomal peptide synthase (PKS/NRPS) gene clusters have been well studied. The PKS pathway peptides require at least three domains, an acyl transferase, a ketosynthase, and an acyl carrier protein [13]. The NRPS pathway shares a common mode of synthesis, the multicarrier thiotemplate mechanism, requiring the cooperation of three basic domains [14]. The adenylation domain selects its 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 in the metabolite is synthesized by the core PKS and NRPS enzymes, and then, the final product is formed with the assistance of various modifying enzymes [15]. The bioactive secondary metabolites produced by the PKS/NRPS pathway in species of B. subtilis include bacilysin [16], bacilysocin [17], surfactin [18], iturin A [19], fengycin [20], mycosubtilin [21], bacillomycins [8], and difficidin [16].

The traditional method of screening for new active products is based on testing for biological activity. However, this method is time-consuming and the same products have been repeatedly discovered [22]. Thus, the discovery of natural products had encountered a bottleneck [23], and the development of a more rapid and effective screening strategy to detect new secondary metabolites was necessary [24, 25]. Genome mining is a technology that uses modern bioinformatics to recognize specific functional genes or gene clusters from genome sequences [26]. With the rapid development of gene sequencing technology and the decreasing cost of genome sequencing, increasing numbers of microbial genome sequences have been determined [27]. Therefore, genome mining has become a more accurate and efficient screening strategy for discovering new metabolites [26].

B. subtilis strain NCD-2 is a promising biological control agent against plant soil-borne diseases that produces lipopeptides, fengycin, and surfactin [28]. Fengycin has an antifungal activity, and surfactin facilitates the root colonization ability of strain NCD-2. Both fengycin and surfactin play important roles in strain NCD-2’s ability to suppress plant soil-borne diseases [29]. The purpose of this study was to
identify potential secondary metabolites with antifungal activities in strain NCD-2 using genome mining. Then, a bioinformatics analysis was conducted to reveal the differences between gene clusters for these secondary metabolites in strain NCD-2 and reference strain *Bacillus velezensis* FZB42. Finally, ultra-high-performance liquid chromatography coupled to quadrupole time-of-flight tandem mass spectrometry (UHPLC-QTOF-MS/MS) was used to identify the different homologs of fengycin and surfactin found in 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 Velvet[30]. The genome of *B. subtilis* NCD-2 contained 189 contigs (> 133 bp; N90, 16,187) of 4,644,322 bp, with an average G + C content of 63.74%. The final assembled genome comprised 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 CRISPR repeat 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 are 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 42.9–46.6%. These genome sequences were downloaded from the GenBank database, and their accession numbers are listed (Additional file 1, Table S1). To analyze the evolution of different *B. subtilis* strains, a phylogenetic tree was constructed based on the genome sequences. The 272 strains of *B. subtilis* were divided into four subspecies, *subtilis*, *inaquosorum*, *spizizenii*, and *stercoris*[31]. As shown in Fig. 2, strain NCD-2 (represented by the black bar) clustered together 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 the online website antiSMASH [32]. In total, nine secondary metabolic gene clusters were identified in the NCD-2 genome sequences (Table 1), including three NRPS, two terpenes, one heterozygous Nrps-Transatpks-Otherks, one type III polyketide gene cluster, one Sactipeptide-head to tail, and a gene cluster with an 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 these nine gene clusters, clusters 3, 7, 8, and 9 had 100% amino acid sequence
homology 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 synthetic gene cluster in \textit{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 those of the model strain 168 and \textit{B. subtilis} strains closely related phylogenetically to strain NCD-2. The fengycin synthetic gene cluster of strain NCD-2 contained three genes, \textit{fenEAB}, while the other strains contained five genes, \textit{fenCDEAB} (Additional file 1, Fig. S1). SrfAB of surfactin was synthesized by the normal transcription and translation of \textit{srfAB} in the 11 strains. However, the same SrfAB was assembled with Gms0365 and Gms0366 and then transcribed and translated by \textit{gms0365} and \textit{gms0366} separately 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 may be different from those of the other \textit{B. subtilis} strains.

| 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-Transatpks-Otherks | 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 \textit{Bacillus velezensis} FZB42 provided by the MIBiG BGC database. NRPS, non-ribosomal peptide synthase; PKS, polyketide synthase; T3pks, type III polyketide; Nrps-Transatpks-Otherks, non-ribosomal peptide synthetase-trans-AT polyketide synthase-Other types of polyketide synthase cluster; Sactipeptide-head to tail, head-to-tail cyclised peptide.
Specifcity of surfactin and fengycin synthetase gene clusters in B. subtilis NCD-2

The surfactin synthetic gene cluster in strain NCD-2 was analyzed using PRISM, and the core genes were selected for a PKS/NRPS analysis. This gene cluster contained four genes: gms0365, gms0366, gms0367, and gms0368. Gms0365 showed an identical conserved structural and functional 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 had lost the T and E domains, but the amino acid residues for the binding pockets of 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 had only T and E domains, with no specific substrate-binding domain. The superposition of Gms0367 and Gms0366 domains formed a complete SrfAB. The T domain was reversed between Gms0367 and Gms0368. The domains of Gms0368 were CATe, in which the thioesterase domain releases linear peptide chains. The domains of Gms0368 were exactly the same as those of SrfAC, but the amino acid residues forming the binding pockets were not completely conserved. The residue sequence was DAF-LGCV, compared with DAFXLGCV of strain FZB42, revealing a difference of one residue.

The fengycin synthetic gene cluster was analyzed by PRISM, and the core genes were selected for a PKS/NRPS analysis. This cluster contained five genes in strain FZB42's genome, they were ordered as fenCDEAB (Fig. 4b). However, the fengycin synthetic gene cluster in strain NCD-2 contained only three genes: gms1961, gms1960, and gms1958. Gms1961 of strain NCD-2 corresponded to FenE in strain FZB42 and they had conserved residues of A8, which bound two amino acids. Gms1960 and Gms1959 in strain NCD-2 had amino acids sequences identical 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 from strain FZB42 were compared with the strain NCD-2 proteome using BioEdit. Gms1961 was most similar to FenC, and Gms1960 was most similar to FenD (Additional file 1, Tables. S2, S3). Therefore, it was hypothesized that Gms1961 and Gms1960 performed the functions of FenC and FenD in strain NCD-2, respectively. Thus, Gms1961 and Gms1960 might have dual functions in the synthesis of fengycin, including those of FenEA as well as FenCD. However, the FenD domain varied greatly between Gms1960 and FZB42, and other enzymes might have similar functions.

To further investigate whether the structure of the fengycin synthetase gene cluster in NCD-2 is strain specific, the fengycin synthetic gene clusters from 11 different B. subtilis strains that are closely related to strain NCD-2 or are model strains were compared (Additional file 1, Fig. S1). The gene cluster sequences of all 11 strains were ppsABCDE (also fenCDEAB), and only that of strain NCD-2 was fenEAB. Therefore, the fengycin synthetic gene cluster of strain NCD-2 is unique.

**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), and the QTOF–MS/MS analysis revealed five fractions in the fengycin cluster (Fig. 5a–e). The five fractions 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 show the distributions of key fragmentation ions ($\alpha$ and $\beta$), representing the linear N-terminal and the cyclic C-terminal segments, respectively, of diverse fengycin species (Additional file 1, Fig. S4AB) and (Fig. 5a–e). 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. 5b) yielded key product ions at $m/z$ 994.5 and 1,108.6, representing fengycin B (Fig. 5b). 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. 5c), while the MS/MS spectrum of the fengycin ion at $m/z$ 739.4 (Fig. 5d) yielded key product ions at $m/z$ 980.5 and 1,094.5 representing fengycin B2 (Fig. 5d). The MS/MS spectrum of the fengycin ion at $m/z$ 767.4 yielding two intense product ions at $m/z$ 994.5/1,008.5 and 1,108.6/1,122.6 representing fengycin C (Fig. 5e). Five classes of fengycins were identified based on the key product ions of $\beta$-hydroxy fatty acid ($\beta$-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, representing surfactin (Additional file 1, Fig. S4C) and (Fig. 5f). Based on these key product ion, one class of surfactin was identified, which were the surfactins ($m/z$ values of 994.6, 1,008.7, 1,022.7, and 1,036.7) of fatty acids with chain lengths varying from C11 to C15 (Fig. S10).

Table 2

| Fengycin homologs in NCD-2 based on key product ions of $\beta$-OH-FA with different chain lengths |
|-------------------------------------------------|---------------------------------|---------------------------------|
| Fengycin family | $[M + 2H]^2+$ | $\beta$-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 |

**Discussion**

Species of *B. subtilis* have the potential to produce two dozen antimicrobial substances, and 5–8% of the *B. subtilis* genome contributes to the production of antimicrobial substances [33]. Some inhibit the growth of pathogens and the germination of spores. The lipopeptide mixture of *B. subtilis* C232 inhibits
the formation of *Verticillium dahliae* microsclerotia [34], and the volatile compounds secreted by *B. subtilis* JA inhibit the conidial formation and mycelial growth of *Glomus etunicatum* [35].

However, certain bioactive compounds are synthesized only under special conditions or as the result of external stimulation; therefore, it is difficult to obtain all the antimicrobial compounds produced by *Bacillus* using traditional cultivation and extraction methods, and this limited the comprehensive understanding of the mechanisms of biological control and biocontrol bacteria [22]. Genome mining allows the prediction of metabolites based on genome sequences and is widely used in obtaining new antibiotics [26]. It was used to identify a new NRPS pathway product, coelichelin, in *Streptomyces coelicolor* [36]. Pseudomycoicidin in *Bacillus pseudomycooides* DSM 12442 was discovered through the heterologous expression of its BGC in *Escherichia coli* [37]. Traditional cultivation and extraction methods were used to identify lipopeptide, fengycin, and surfactin from *B. subtilis* NCD-2, and fengycin showed strong antifungal abilities against *V. dahliae* and *B. cinerea*. However, the fengycin-deficient mutant of strain NCD-2 still has a certain antifungal ability, but it is less than that of wild-type strain NCD-2. Therefore, other antifungal active compounds, besides fengycin, may be produced by strain NCD-2 [2, 28, 29]. In this study, genome mining was conducted to analyze the potential antimicrobial compounds of the NCD-2 strain, and some of them were identified using MS. In total, nine kinds of secondary metabolite gene clusters related to surfactin, bacillaene, fengycin, bacillibactin, subtilosin, bacilysin, two terpenes, and one unknown product were identified from the genome of strain NCD-2. The surfactin [38], bacillaene [39], fengycin [40], bacilliactin [41], subtilosin [42], and bacilysin [43] showed antimicrobial abilities and played different roles in suppressing plant diseases. Only fengycin and surfactin were identified from the lipopeptide extract of NCD-2 despite the presence of other gene clusters in the genome. These other antimicrobial compounds may not have been detected because the acid precipitation extraction method was not suitable. Some bioactive compounds, such as bacillaene, bacillibactin, subtilosin, and bacilysin, are not lipopeptides. Therefore, these substances were not extracted using hydrochloric acid precipitation [44–47].

Fengycin comprises a peptide ring circled by 10 amino acids with a fatty acid chain tail. The mechanism of fengycin synthesis has been well studied in *B. velezensis* strain FZB42 [48], and the fengycin synthetic gene cluster in the strain consists of five genes (38 kb) that encode the synthetases FenCDEAB, of 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 [49–51]. However, NCD-2 only had fenEAB, lacking fenC and fenD, compared with the typical cluster structure of fenCDEAB in the FZB42 strain and 10 other *Bacillus* strains (Fig. 4b) and (Additional File 1, Fig. S1). 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 Gms1961 protein in the NCD-2 strain 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, and the molecular weight was 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 [52]. The UHPLC-QTOF MS/MS of the fengycins revealed that all the structures possessed the amino acid ornithine at position 2 (Fig. 5a–e), indicating that there was a protein that transports ornithine in the NCD-2 strain. Thus, it was hypothesized that Gms1961 functions as FenC and FenE. The analysis was performed using the Gms1960 protein and it 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 may have functions similar to those of FenD.

Although the fengycin synthetic gene cluster in the NCD-2 strain lacked two important genes-phenC and phenDhat synthesize enzymes compared with the reported fengycin synthetic gene cluster, the NCD-2 strain was capable of producing 26 homologs of 5 kinds of fengycins. The amino acids at position 6 and 10 of the fengycin cyclic peptide ring determine the type of fengycin. There are currently five types of reported fengycins, A, B, A2, B2, and C (Additional File 1, Fig. S4). When the amino acid at position 6 was valine and at position 10 was isoleucine or valine, then fengycin B or fengycin B2, respectively, was produced (Fig. 5a, b) and (Additional File 1, Fig. S4); however, if the amino acid at position 6 was alanine, then fengycin A or fengycin A2, respectively, was produced (Fig. 5c, d) and (Additional File 1, Fig. S4). When the amino acid at position 6 was isoleucine or leucine and at position 10 was valine, then fengycin C was produced (Fig. 5e) and (Additional File 1, Fig. S4). The MS analysis of the fengycins in the NCD-2 strain revealed that the strain was capable of producing these five kinds of fengycins. Based on differences in the number of carbon atoms in the β-OH FA, fengycin had different homologs, and the molecular weight of each homologs differed by 14 (-CH2) [53]. The molecular structure of the lipopeptide determines its biological activity, and long-chain fatty acids increase the hydrophobic activities of lipopeptides, making them more likely to have membrane-bound antimicrobial effects [54]. A Bacillus circulans strain produces four fengycin homologs, but only fengycins with C16 and C17 carbon atoms in their β-OH FA chains had antibacterial activities [55]. The NCD-2 strain produced 14 fengycin homologs having more than 16 carbon atoms, and they accounted for a large proportion of all the homologs. It was speculated that these long-chain fengycins play important roles in the antimicrobial functions of NCD-2. The Bacillus siamensis SCSIO 05746 strain produces a great number of fengycin homologs, including 19 homologs of fengycin B [56]. Using an MS analysis, the five fengycins produced by the NCD-2 strain were divided into 26 homologs (Fig. 5a–e) and (Additional File 1, Fig. S5-S9). Therefore, NCD-2 is currently the strain with the largest number of known fengycin homologs [57].

During the microbial synthesis of secondary metabolites, such as lipopeptide, the relatively high energy-consuming process of protein synthesis takes priority [58]. Excessive energy consumption is not conducive to the normal growth of microbes, and, generally, microbes produce antibiotics in large amounts only when encountering pathogens or other stresses [59]. In the long-term evolution of NCD-2, the key synthetic genes fenEAB involved in synthesizing fengycin were conserved, while two important synthetic genes fenCD were lost. However, five fengycins are still produced. Gms 1961 played the dual roles of FenC and FenE, indicating that NCD-2’s fengycin synthetic process, which is unique to the strain, was more energy-efficient than the process used in the other strains.
Conclusions

In this study, genome mining and UHPLC–QTOF–MS/MS were performed. They determined that there were more gene clusters encoding antimicrobial compounds in the genome of the NCD-2 strain and that the fengycin synthetic gene cluster was unique. The results indicated that the NCD-2 strain has a unique mechanism for synthesizing fengycin. Using molecular genetics and biochemistry to analyze the new mechanism of fengycin synthesis may provide a new theory for the synthesis of antimicrobial compounds through the NRPS pathway.

Methods

Microorganisms and culture conditions

*Bacillus subtilis* NCD-2 was routinely grown at 37 °C on Luria Bertani medium. For secondary metabolite production, strain NCD-2 was grown in Landy broth at 30°C and 180 rpm [60]. Phytopathogen *Botrytis cinerea* BC-10 was used for antifungal activity tests following the method described by Guo[29] 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 plates were inoculated with *B. subtilis* NCD-2 using a sterilized toothpick 2 cm from the center. Finally, the diameter of the inhibition zone was measured after a 3-d incubation at 25°C.

Genome sequencing of NCD-2

The Illumina Solexa platform was used for the whole-genome sequencing following the method described by Karim[61] with some modifications. The quality of reads was checked using FastQC([http://www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/))[62], paired-end reads were trimmed using Sickle([https://github.com/najoshi/sickle](https://github.com/najoshi/sickle)), and were assembled using the software Velvet[30]. QUAST 5.02 was used to assess the quality of contigs and scaffolds[63]. The assembled scaffolds were annotated using Prokka (version v.1.13)[64]. The annotation of strain NCD-2 genome was performed using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline ([http://www.ncbi.nim.nih.gov/genome/annotation_prok/](http://www.ncbi.nim.nih.gov/genome/annotation_prok/)) utilizing GeneMark, Glimmer, and tRNAscan-SE tools [65], and the functional annotation was carried out using the Rapid Annotations by subsystems Technology (RAST) server with the seed database [66]. Finally, the genome of strain NCD-2 was deposited in the National Center for Biotechnology Information (NCBI;[https://www.ncbi.nlm.nih.gov/](https://www.ncbi.nlm.nih.gov/)), and the GenBank accession number is CP023755.

Evolutionary analysis, signal peptide and CRISPR repeat detection

The whole-genome sequences of *B. subtilis* and closely related species were downloaded from the NCBI database, and the REALPHY website ([http://realphy.unibas.ch](http://realphy.unibas.ch)) [67] was used for genome-wide comparisons with default parameters. A phylogenetic analysis was conducted using MEGA5 [68] with the
Maximum Composite Likelihood parameter model [69]. Aphylogenetic 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/) [70]. CRISPR repeats were detected using CRISPRCasFinder (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) [71].

**Predictions and a specificity analysis of secondary metabolite synthetic gene clusters**

Secondary metabolite synthetic gene clusters for strain NCD-2 were detected using antiSMASH (http://antismash.secondarymetabolites.org) [32, 72] and PRISM (http://grid.adapsyn.com/prism/) [73] with the parameters selected by default. Functional domain predictions for PKS/NRPS in the predicted gene clusters were analyzed using the PKS/NRPS Analysis Website (http://nrps.igs.umaryland.edu/) [74]. 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.

**Separation of lipopeptides by FPLC**

Lipopeptides were extracted using the method described by Guo [29]. Briefly, strain NCD-2 was cultured in Landy broth [60] at 30°C for 72 h with shaking at 180 rpm. The cell-free supernatant was obtained by centrifugation at 8,000 x g for 30 min at 4°C. The supernatant was adjusted to pH 2.0 with 6 mol/L HCl and stored for 12 h at 4°C. After centrifugation at 10,000 x g, for 20 min, the resulting pellet was extracted with methanol under continuous magnetic stirring 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 the SOURCE 5RPC ST 4.6/150 column as described previously [75]. The lipopeptides were eluted by 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 collected by FPLC automatically. Finally, each peak was concentrated using a rotary evaporator and was analyzed using UHPLC-QTOF–MS/MS.

**UHPLC–QTOF–MS/MS**

The 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) that was 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 5-µm particles, 2.1 mm × 100 mm) was used for separation. The mobile phases A and B were water and acetonitrile with 0.1% formic acid, respectively, in both phases with an optimized linear gradient eluting 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; 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°C. The 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 for investigating and interpreting 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°C; declustering potential (in TOF MS experiments, 100 V; and collision energy, 10.0 V. During the TOF-MS/MS declustering potential, the collision energy spread was between 100 V and 5 V, with rolling collision energy. The MS was operated in full-scan TOF-MS (m/z 200–2,000) and MS/MS (m/z 200–1,600) modes using Information Dependent Acquisition for a single run analysis.

**Additional Files**

**Additional file 1: Supplemental Material.**

Files contain supplemental materials, including supplementary tables and figures referenced in this manuscript. Fig.S1. Fengycin synthetic gene clusters of different strains which have a close relative with NCD-2 or model strains. Fig.S2. Surfactin synthetic gene clusters of different strains which have a close relative with NCD-2 or model strains. Fig.S3. Elution of lipopeptides separated from the crude methanolic extract using an AKTA Purifier. Fig.S4. Primary structures of fengycins and surfactins. Fig.S5. FengycinA of β-OH FA with the chain length varying from C14 to C19 are identified based on key product ions. Fig.S6. FengycinB of β-OH FA with the chain length varying from C12 to C19 are identified based on key product ions. Fig.S7. FengycinA2 of β-OH FA with the chain length varying from C15-C18 are identified based on key product ions. Fig.S8. FengycinB2 of β-OH FA with the chain length varying from C14-C18 are identified based on key product ions. Fig.S9. FengycinC of β-OH FA with the chain length varying from C18-C20 are identified based on key product ions. Fig.S10. Surfactin of fatty acid with the chain length varying from C11-C15 are identified based on key product ions. Table S1. All the *B. subtilis* strain with the assembly level of chromosome and their RefSeq assembly accession. Table S2. The homologues of FenC of FZB42 by scanning the local NCD-2 proteome in BioEdit. Table S3. The homologues of FenD of FZB42 by scanning the local NCD-2 proteome in BioEdit. Table S4. Adenylation domain binding amino acids predicted by PRISM.

**Abbreviations**

- UHPLC-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
- 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.

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). We thank Lesley Benyon, PhD, from Liwen Bianji, Edanz Group China (www.liwenbianji.cn/ac), for editing the English text of a draft of this 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.

Availability of data and materials

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

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.

Author details

Institute of Plant Protection, Hebei Academy of Agricultural and Forestry Sciences, Integrated Pest Management Center of Hebei Province, Key Laboratory of IPM on Crops in Northern Region of North China, Ministry of Agriculture, Baoding 071000, China
References

1. Sonenshein AL. Control of sporulation initiation in *Bacillus subtilis*. Curr Opin Microbiol. 2000;3(6):561–6.

2. 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. Microbiol Res. 2015;178:42–50.

3. 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. Microbiol Res. 2017;199:89–97.

4. 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. Front Microbiol. 2017;8:469.

5. Moszer I, Jones L, Moreira S, Fabry C, Danchin A. SubtList: the reference database for the *Bacillus subtilis* genome. Nucleic Acids Res. 2002;30(1):62–5.

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. Biol Control. 2017;105:93–9.

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. Biol Control. 2018;127:145–54.

8. Stein T. *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. Mol Microbiol. 2005;56(4):845–57.

9. Beppu T. Secondary metabolites. as chemical signals for cellular differentiation. Gene. 1992;115(1–2):159–65.

10. Chaudhary AK, Dhakal D, Sohng JK. An insight into the "-omics" based engineering of streptomycetes for secondary metabolite overproduction. Biomed Res Int. 2013;2013:968518–8.

11. 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. Front Microbiol. 2014;5:252–2.

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 Res. 2012;41(D1):D408–14.

13. Chen XH, Koumoutsi A, Scholz R, Borriss R. More than anticipated - production of antibiotics and other secondary metabolites by *Bacillus amyloliquefaciens* FZB42. J Mol Microbiol Biotechnol. 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. J Biol Chem. 1996;271(26):15428–35.
15. Du L, Lou L. PKS and NRPS release mechanisms. Nat Prod Rep. 2010;27(2):255–78.

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. Microb Cell Fact. 2009;8(1):63–3.

17. Tamehiro N, Okamotosooya Y, Okamoto S, Ubakata 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. Biochem Biophys Acta. 2003;1611(1–2):91–7.

19. Yu GY, Sinclair JB, Hartman GL, Bertagnoll BL. Production of iturin A by *Bacillus amyloliquefaciens* suppressing Rhizoctonia solani. Soil Biol Biochem. 2002;34(7):955–63.

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. Appl Biochem Biotechnol. 1999;77–79:223–33.

21. Moyne A, Cleveland TE, Tuzun S. Molecular characterization and analysis of the operon encoding the antifungal lipopeptide bacillomycin D. FEMS Microbiol Lett. 2004;234(1):43–9.

22. Tulp M, Bohlin L. Rediscovery of known natural compounds: nuisance or goldmine? Bioorg Med Chem. 2005;13(17):5274–82.

23. von Bubnoff A. Seeking new antibiotics in nature's backyard. Cell. 2006;127(5):867–9.

24. Oman TJ, van der Donk WA. Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat Chem Biol. 2010;6(1):9–18.

25. Lane AL, Moore BS. A sea of biosynthesis: marine natural products meet the molecular age. Nat Prod Rep. 2011;28(2):411–28.

26. Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol. 2015;13(8):509–23.

27. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends in genetics. 2014;30(9):418–26.

28. 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 Molecular Biology. 2010;33(2):333–40.

29. 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. Microbiol Res. 2014;169(7–8):533–40.

30. Zerbino D, Birney E. Velvet: algorithms for *de novo* short read assembly using de bruijn graphs. Genome Res. 2008;18(5):821–9.

31. 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.
32. 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 Res. 2013;41(W1):W204–12.

33. Sansinenea E, Ortiz A. Secondary metabolites of soil Bacillus spp. Biotech Lett. 2011;33(8):1523–38.

34. 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):622–34.

35. 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 J Microbiol Biotechnol. 2008;24(7):1133–7.

36. 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 Microbiol Lett. 2000;187(2):111–4.

37. Basichipalu S, Dischinger J, Josten M, Szekat C, Zweynert A, Sahl H, Bierbaum G. Pseudomycoicidin, a class II lantibiotic from Bacillus pseudomycoides. Appl Environ Microbiol. 2015;81(10):3419–29.

38. Seydlová G, Svobodová J. Review of surfactin chemical properties and the potential biomedical applications. Central European Journal of Medicine. 2008;3(2):123–33.

39. Patel PS, Huang S, Fisher S, Pirnik D, Aklonis C, Dean L, Meyers E, Fernandes P, Mayerl F. Bacillaene, a novel inhibitor of proaryotic protein synthesis produced by Bacillus subtilis. J Antibiot. 2006;48(9):997.

40. 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. Can J Microbiol. 2007;53(7):901–11.

41. Miethke M, Klotz O, Linne U, May JJ, Beckering CL, Marahiel MA. Ferri-bacillibactin uptake and hydrolysis in Bacillus subtilis. Mol Microbiol. 2006;61(6):1413–27.

42. Thennarasu S, Lee DK, Poon A, Kawulka KE, Vederas JC, Ramamoorthy A. Membrane permeabilization, orientation, and antimicrobial mechanism of subtilosin A. Chem Phys Lipid. 2005;137(1–2):38–51.

43. Kenig M, Abraham EP. Antimicrobial activities and antagonists of bacilysin and anticapsin. Microbiology. 1976;94(1):37–45.

44. Li B, Li Q, Xu Z, Zhang N, Shen Q, Zhang R. Responses of beneficial Bacillus amyloliquefaciens SQR9 to different soilborne fungal pathogens through the alteration of antifungal compounds production. Front Microbiol. 2014;5:636–6.

45. Wu L, Wu H, Chen L, Yu X, Borris R, Gao X. Difficidin and bacilysin from Bacillus amyloliquefaciens FZB42 have antibacterial activity against Xanthomonas oryzae rice pathogens. Sci Rep. 2015;5:12975.

46. Nonejuie P, Trial RM, Newton GL, Lamsa A, Perera VR, Aguilar J, Liu W, Dorrestein PC, Pogliano J, Pogliano K. Application of bacterial cytological profiling to crude natural product extracts reveals the
antibacterial arsenal of *Bacillus subtilis*. The Journal of Antibiotics. 2016;69:353–61.

47. Goodson JR, Klupt S, Zhang C, Straight PD, Winkler WC. LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in *Bacillus amyloliquefaciens*. Nature Microbiology. 2017;2(5):1–10.

48. 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. Front Microbiol. 2018;9:2491.

49. 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*. Mol Genet Genomics. 1995;248(2):121–5.

50. Lin G, Chen C, Tschen JS, Tsay S, Chang Y, Liu S. Molecular cloning and characterization of fengycin synthetase gene fenB from *Bacillus subtilis*. J Bacteriol. 1998;180(5):1338–41.

51. Lin TP, Chen CL, Chang JK, Tschen JS, Liu ST. Functional and transcriptional analyses of a fengycin synthetase gene, fenC, from *Bacillus subtilis*. J Bacteriol. 1999;181(16):5060–7.

52. Lautru S, Deeth RJ, Bailey LM, Challis GL. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. Nat Chem Biol. 2005;1(5):265–9.

53. 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):644–9.

54. Tripathi L, Irorere VU, Marchant R, Banat IM. Marine derived biosurfactants: a vast potential future resource. Biotech Lett. 2018;40(11–12):1441–57.

55. Sivapathasekaran C, Mukherjee S, Samanta R, Sen R. High-performance liquid chromatography purification of biosurfactant isoforms produced by a marine bacterium. Anal Bioanal Chem. 2009;395(3):845–54.

56. 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. Appl Microbiol Biotechnol. 2019;103(10):4153–65.

57. Ongena M, Jacques P. *Bacillus lipopeptides*: versatile weapons for plant disease biocontrol. Trends Microbiol. 2008;16(3):115–25.

58. 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).

59. 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. J Mol Biol. 2004;335(2):425–35.

60. 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.

61. 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
62. Andrews S. FastQC A quality control tool for high throughput sequence data. 2010.
63. Alexey G, Vladislav S, Nikolay V, Glenn T. QUAST: quality assessment tool for genome assemblies. In: 2013; 2013: 1072–1075.
64. Torsten S. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
65. 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):319–9.
66. 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 Genom. 2008;9:75.
67. Haubold B, Klotzl F, Pfaffelhuber P. andi: fast and accurate estimation of evolutionary distances between closely related genomes. Bioinformatics. 2015;31(8):1169–75.
68. 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. Mol Biol Evol. 2011;28(10):2731–9.
69. Xu X, Reid N. On the robustness of maximum composite likelihood estimate. Journal of Statistical Planning Inference. 2011;141(9):3047–54.
70. 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. Nat Biotechnol. 2019;37(4):420–3.
71. 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 Res. 2018;46(W1):W246–51.
72. 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 Res. 2011;39(2):W339–46.
73. 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 Res. 2015;43(20):9645–62.
74. Bachmann BO, Ravel J. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. Methods Enzymol. 2009;458:181–217.
75. LiB LuX. GuoQ, QianC, LiS, MaP. Isolation and identification of lipopeptides and volatile compounds produced by Bacillussubtilis strain BAB-1, vol. 43; 2010.

Figures
Figure 1

Circular genome of strain NCD-2 with specific features. The potato dextrose agar plate inside the representation of the circular genome shows the antifungal activity of strain NCD-2 against Botrytis cinerea. The black bars outside the circular genome indicate the secondary metabolite synthetic gene clusters.
Figure 2

Phylogenetic tree of 113 Bacillus subtilis strains having whole-genome sequences deposited in the GenBank database. The position of strain NCD-2 in the phylogenetic tree is indicated by a black bar. 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 synthetic gene clusters in Bacillus subtilis strain NCD-2. Different color 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, other genes, respectively.
Figure 4

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 5

Comparisons of functional domains of core genes involved in synthesizing surfactin and fengycin in NCD-2. The functional domains of 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.

Supplementary Files

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

- Fig.S10.jpg
- Fig.S9.jpg
- Fig.S8.jpg
- Fig.S7.jpg
- Fig.S6.jpg
- Fig.S5.jpg
