Whole-genome Sequencing Reveals Putative Underlying Mechanisms of Biocontrol Capability of IBFCBF-5

Chen Luo
Central South University of Forestry and Technology
Airong Shen
Central South University of Forestry and Technology
Yesong Ren
Central South University of Forestry and Technology
Yang Zhai
Chinese Academy of Agricultural Sciences
Yi Cheng
Chinese Academy of Agricultural Sciences
Jianping Xu
McMaster University
Lin Wei
Hunan Academy of Agricultural Sciences
Jilie Li
Central South University of Forestry and Technology
Liangbin Zeng (zengliangbin@caas.cn)
Chinese Academy of Agricultural Sciences

Research Article

Keywords: Biocontrol agent, Bacillus, antifungal activity, antibacterial activity, antimicrobial metabolites

Posted Date: July 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-668301/v1

Abstract

Objective: As the world's food safety and environmental safety problems become increasingly severe, the agricultural sectors of various countries are also paying closer attention to the use of biofertilizers and biocontrol agents. Rhizosphere bacteria are a significant source of commonly used biofertilizers and biocontrol agents. This study aims to describe the genome and genomic traits of a biocontrol agent in the genus *Bacillus*.

Results: In this paper, a strain of *Bacillus amyloliquefaciens* IBFCBF-5 was isolated and identified to have an inhibitory effect on several common oomycete and fungal pathogens *Phytophthora capsica, Sclerotinia sclerotiorum, Colletotrichum gloeosporioides,* and *Fusarium oxysporum* f.sp. *cucumerinum*. The genome of strain IBFCB-5 was sequenced, and the assembled genome was 4,338,658bp, with a G+C content of 46.05%. The IBFCBF-5 genome contains abundant GH, GT, CE, PL, AA, and CBM gene families, potentially degrading cellulose and hemicellulose, chitin, starch, xylan, peptidoglycan, etc. In addition, 14 lipopeptide and polypeptide antibiotic gene clusters were found in IBFCBF-5, including those coding for the synthesis of several known antifungal and antibacterial compounds Fengycin, Bacilysin, Bacillibactin, and plantazolicin.

Conclusion: Our results show that *Bacillus amyloliquefaciens* IBFCBF-5 has a broad-spectrum antifungal ability and that its genome contains many genes coding for antimicrobial metabolites.

Background

With international food safety and environmental safety issues becoming increasingly severe, the agricultural sectors in various countries are also paying more attention to the use of bio-fertilizers and biocontrol agents [1]. Consequently, plant rhizosphere growth-promoting bacteria (PGPR) and their biologically active compounds promoting plant growth and/or antagonizing plant pathogens have received increasing recognition. These potential biofertilizers and biocontrol agents are generally considered environmentally friendly, superior to chemical fertilizers and pesticides [2, 3].

Bacillus amyloliquefaciens is closely related to *Bacillus methylotrophicus* and *Bacillus subtilis*. It has a short generation time and is resistant to various stresses. Its spores are highly tolerant to high temperature, drying, ultraviolet and ionizing radiation, and many kinds of toxic chemicals. It can also produce various antibacterial substances such as lipopeptides, bacteriocins, and antibacterial proteins [4]. These antibacterial substances belong to different groups and have shown wide antibacterial activities and no known detrimental effects on the environment. These bacteria and the compounds that they produce have been the focus of research in biological control of plant diseases, animal feed processing, and medical research and development in recent years [5, 6].

There are two primary types of antibacterial substances secreted by *Bacillus amyloliquefaciens*, namely, low molecular weight antibiotics and high molecular weight antibacterial proteins. According to their structural differences, low molecular weight antibiotics can be divided into three categories: Surfactin,

Fengycin, and Iturin. Surfactin mainly inhibits the growth of bacteria, viruses, and mycoplasma. Fengycin has a strong inhibitory effect on filamentous fungi. Iturin mainly inhibits the growth of fungi. The inhibitory effect of high molecular weight antibacterial proteins on fungi is primarily manifested in inhibiting mycelial growth and destroying fungal cells, including chitinases and glucanases. Chitinase and β-1,3-glucanase can degrade the cell wall of pathogenic fungi [7, 8]. Bacteriocins and lipopeptides have been commonly used in medicine, agriculture, and other fields because of their stable physical and chemical properties, broad antibacterial spectrum, and low frequencies of drug resistance, and they have become a focus of bio-pesticide research [9].

Here we screened and identified a strain of *Bacillus amyloliquefaciens* IBFCBF-5 obtained from the rhizosphere soil of a healthy pepper plant. This strain showed vigorous biocontrol activity against a variety of fungal pathogens. To help understand its potential mechanisms of antifungal activities, we sequenced the whole genome of strain IBFCBF-5 and analyzed its potential to produce secondary metabolites related to antimicrobial compounds. Our study provides a basis for the development and utilization of metabolites of *B. amyloliquefaciens* strain IBFCBF-5 in the future.

Results

Screening and identification of strain IBFCBF-5

Fifteen strains of rhizosphere bacteria were isolated from pepper root-soil samples, and one strain IBFCBF-5 showed excellent antimicrobial activity based on plate confrontation culture assay(Fig. 1). Specifically, our assay showed zones of inhibition of 18.60mm, 10.31mm, 24.35mm and 24.06mm respectively against *Sclerotinia sclerotiorum*, *Phytophthora capsica*, *Fusarium oxysporum* f. sp. *cucumerinum*, and *Colletotrichum gloeosporioides* (Table 1).

Table 1
Antagonistic effect of IBFCBF-5 against four phytopathogens

Pathogenic fungi	Diameter of zone of inhibition(mm)
Sclerotinia sclerotiorum	18.60 ± 0.38b
Phytophthora capsica	10.31 ± 0.82 c
Colletotrichumgloeosporioides	24.35 ± 0.40 b
Fusarium oxysporum f. sp. cucumerinum	24.06 ± 0.58 c
Note: the data in the table are average ± standard error. Different lowercase letters indicated that the antagonistic effect of strain IBFCBF-5 against pathogenic fungi was significant.	

Strain IBFCBF-5 was grown on LB (Luria-Bertani) solid medium. The colonies are irregular and sub-radial, off-white, wet, and sticky on the surface without wrinkles. The cells are rod-shaped, and the flagella are visible under scanning electron microscopy (Fig. 2).

Based on BLAST search of the 16S rRNA gene, strain IBFCBF showed a sequence identity of > 90% with *Bacillus amyloliquefaciens* [10, 11]. The 16S rRNA gene sequences of 14 type strains of different species of *Bacillus* were extracted from GenBank to construct a phylogenetic tree. The result is shown in Fig. 3. Strain IBFCBF-5 had the highest similarity with the 16S rRNA gene of *Bacillus amyloliquefaciens* MPA (accession number: 117946.1). The GenBank accession number for strain IBFCBF-5 was SUB9291579. Combined with morphological, physiological, and biochemical identification results, it was preliminarily determined that strain IBFCBF-5 belongs to *Bacillus amyloliquefaciens.*

Genome sequencing and sequence assembly of IBFCBF-5.

As shown in Fig. 4, strain IBFCBF-5 contains a 4,338,658 bp circular chromosome. The relevant genome statistics are shown in Table 2. The genomic G + C content was 46.05%, with a predicted 4546 genes in total, including 4341 protein-coding DNA genes of an average length of 886 bp, 27 tRNA genes, 86 rRNA genes, and 92 other ncRNAs. It also contains a cluster of CRISPR and four gene islands.

Table 2
Basic characteristics of IBFCBF-5 genome

Attribute	Value
Length of sequence (bp)	4338658
G + C content (%)	46.05
Protein-coding gene number	5118
Total repetitive sequence length (bp)	5118
Total genes	4546
Protein coding genes	4341
Non-coding RNA	205
Genomic island number	4
CRISPR	1

Taxonomic status and genetic evolution of IBFCBF-5 strains

Based on the results of ANI online analysis, the ANI values of IBFCBF-5 strain with *Bacillus amyloliquefaciens* and *Bacillus velez* were between 97.65% and 99.00%. In contrast, the ANI values with *Bacillus licheniformis* and *Bacillus subtilis* were less than 78.00% (Table 3). However, the threshold of international identification lower than "species" is 95%. Combined with Fig. 5, it shows that the genetic relationship between strain IBFCBF-5 and *Bacillus amyloliquefaciens* is closer. The DDH values of strain IBFCBF-5 and *Bacillus amyloliquefaciens* and *Bacillus velez* were between 91.80% and 81.70%, while the

DDH values of strain IBFCBF-5, *Bacillus licheniformis*, and *Bacillus subtilis* were between 16.20% and 34.00%, which were lower than 70%. The highest similarity of *Bacillus amyloliquefaciens* ZJU1 (PRJNA544619) ANI and DDH was 99.00% and 91.80%, respectively, which further indicated that strain IBFCBF-5 had the closest genetic relationship with *Bacillus amyloliquefaciens*.

Table 3
The values of ANI and DDH between the IBFCBF-5 strain and its related species

Strain (GenBank accession no.)	IBFCBF-5 strain	
	ANI	DDH
*Bacillus velez*DSYZ_PRJNA474548	98.09%	83.70%
*Bacillus velez*CC09_PRJNA315173	98.02%	86.30%
Bacillus velez S4_PRJNA612570	98.20%	84.30%
Bacillus amyloliquefaciens XJ5_PRJNA715331	98.13%	80.80%
Bacillus amyloliquefaciens FS1092_PRJNA243331	98.15%	86.30%
Bacillus amyloliquefaciens ZJU1_PRJNA544619	99.00%	91.80%
Bacillus amyloliquefaciens WF02_PRJNA631938	97.65%	83.20%
Bacillus amyloliquefaciens YP6_PRJNA488691	97.65%	81.70%
Bacillus licheniformis ATCC14580_PRJNA509976	72.86%	16.20%
Bacillus subtilis CW14_PRJNA377766	77.63%	34.00%
Bacillus subtilis SP1_PRJNA641411	77.20%	32.30%
Bacillus subtilis BS49_PRJEB7327	77.21%	32.10%

Functional annotation of the IBFCBF-5 genome

GO (gene ontology) is a database established by the Gene Ontology Consortium to annotate genes in three ways: Cellular Component, Molecular Function, and Biological Process. As shown in Fig. 6, a total of 3061 genes are annotated, of which 31.7% were in "Cellular Component", followed by membrane and membrane parts, at 23.8% and 17.5% respectively. The most significant proportion of "Molecular Function" is "catalytic activity", accounting for 67.2%. In "Biological Process", metabolic processes and cellular processes accounted for 57% and 46.9%, respectively. There are 25 functions of detoxification genes related to antibiotic antagonism, accounting for 0.8%.

The sequencing data are analyzed based on the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. As shown in Fig. 7, "Metabolism" accounts for the most significant proportion, followed by "Environmental Information Processing". However, biosynthesis of amino acids accounted for a large

proportion of "Metabolism", followed by carbon metabolism. Furthermore, several predicted bacteriostatic substances secreted by *Bacillus amyloliquefaciens* also include amino acids.

The results of COG (Cluster of Orthologous Groups of proteins) database analysis showed that a total of 3512 genes were functionally annotated. Among them, Amino acid transport and metabolism, Transcription, Carbohydrate transport and metabolism, Inorganic ion transport and metabolism, Cell wall/membrane/envelope biogenesis, transcription, ribosomal structure and biogenesis, and other related genes are abundant, and respectively the number of annotated genes is 346, 303, 261, 214, 180, 162 (Fig. 8).

CAZyme Gene Families

As shown in Table 4, there are 174 CAZyme gene families encoded by strain IBFCBF-5, including GH (n = 52), GT (n = 44), CE (n = 29), PL (n = 3), AA (n = 6), and CBM (n = 40) gene families. Among the GH family members, GH5, GH11, GH26, GH43, GH51, and GH53 are mainly related to cellulose and hemicellulose degradation; and GH13 and GH126 are primarily associated with starch hydrolysis. For other CAZyme genes, GH18, GH23, CE9, and CBM50 are related to chitin degradation; GH43, CE1, CE3, CE4, and CE7 are related to xylan degradation; GH43, CE1, PL1, and PL9 are mainly related to pectin-degrading enzymes; GH23 and GH73 are related to peptidoglycan degradation; and GH3, GH16, and GH30 are related to glucanases. Overall, the genome of strain IBFCBF-5 contains great potential to degrade cellulose, hemicellulose, starch, chitin, pectin, peptidoglycan, and glucan.

Table 4
CAZymes families in the genome of *B. amyloliquefaciens* IBFCBF-5

CAZymes families	**gene subfamilies (number)**
Glycoside hydrolases	GH1 (4)、GH3 (1)、GH4 (4)、GH5 (1)、GH11 (1)、GH13 (6)、GH16 (1)、GH18 (2)、GH23 (6)、GH26 (1)、GH28 (1)、GH30 (2)、GH32 (3)、GH43 (4)、GH46 (1)、GH51 (2)、GH53 (1)、GH65 (1)、GH68 (1)、GH73 (3)、GH76 (1)、GH109 (4)、GH126 (1)
glycosyl transferases,	GT1 (3)、GT2 (16)、GT4 (8)、GT8 (1)、GT19 (1)、GT26 (1)、GT28 (3)、GT46 (2)、GT51 (5)、GT83 (4)
carbohydrate esterases	CE1 (8)、CE3 (2)、CE4 (7)、CE7 (2)、CE9 (2)、CE10 (3)、CE12 (2)、CE14 (3)
polysaccharide lyases	PL1 (2)、PL9 (1)
carbohydrate-binding modules	CBM2 (1)、CBM3 (1)、CBM6 (1)、CBM12 (3)、CBM16 (1)、CBM26 (1)、CBM34 (1)、CBM37 (2)、CBM50 (29)
auxiliary activities	AA4(1)、AA6(1)、AA7(3)、AA10(1)

Gene Clusters for Secondary Metabolites

We used the online software anti-SMASH to analyze the secondary metabolites of strain IBFCBF-5. A total of 14 secondary metabolite gene clusters were predicted (Table 5), and 9 cluster types with putative antimicrobial activities were found. The predicted antimicrobial substances included: Plantazolicin[12], Difficidin, Fengycin, Bacillaene[13], Macrolactin H[14], butirosin A/butirosin B[15], Bacillibactin[16], Bacilysin[17] and Surfactin, as shown in Fig. 9.

Table 5
Synthetic gene clusters of secondary metabolites of strains

Cluster ID	Cluster type	Most similar known cluster	Similarity (%)
Cluster 1	phosphonate	-	-
Cluster 2	RRE-containing	plantazolicin	91
Cluster 3	transAT-PKS	difficidin	86
Cluster 4	T3PKS	-	-
Cluster 5	terpene	-	-
Cluster 6	NRPS	fengycin	100
Cluster 7	transAT-PKS-like, transAT-PKS,NRPS,T3PS	bacilllaene	100
Cluster 8	transAT-PKS	macrolactin H	100
Cluster 9	terpene	-	-
Cluster 10	PKS-like	butirosinA /butirosin B	7
Cluster 11	NRPS,RiPP-like	bacillibactin	100
Cluster 12	NRPS	-	-
Cluster 13	other	bacilysin	100
Cluster 14	NRPS	surfactin	82

The amino acid sequence similarity between Cluster 2 and Plantazolicin synthesis gene cluster derived from BGC0000569 is 91%. The similarity between the Difficidin synthesis gene cluster derived from BGC00000176 and the Difficidin synthesis gene cluster derived from Bacilllaene is 86%. The similarity between cluster 6 and Fengycin synthesis gene cluster derived from BGC0001095 is 100%. The similarity between the Bacilllaene synthesis gene cluster derived from BGC000108 and Bacilllaene synthesis gene cluster derived from BGC000108 is 100%. The similarity between the Macrolactin H synthesis gene

cluster derived from BGC0000181 and the Macrolactin H synthesis gene cluster derived from BGC0000181 is 100%. The similarity between Cluster 10 and BGC0000693-derived Butirosin synthesis gene clusters is 7%; the similarity between Cluster11 and BGC0000309-derived Bacillibactin synthesis gene clusters is 100%; that between Cluster13 and BGC0001184-derived Bacilysin synthesis gene clusters is 100%; and that between Cluster 14 and BGC0000433 -derived Surfactin synthesis gene clusters is 100%.

Five functional unknown gene clusters (Clusters 1,4,5,9,12) were found, including one phosphonate (Phosphonate), one T3PKS (Type III PKS), two terpenes (Terpene), and one non-ribosomal peptide synthase (NRPS), indicating that there are gene clusters for the synthesis of new bacteriostatic substances in the strain, and with a great application potential in agriculture and pharmaceutical industry.

Discussion

In 1943, Japanese scientist Fukumoto discovered *Bacillus amyloliquefaciens* for the first time in soil. Because this group of bacteria can produce liquefied amylase and decompose starch, it is named *Bacillus amyloliquefaciens*. Many subsequent studies have shown that *Bacillus amyloliquefaciens* has biological control characteristics, including plant colonization, inhibition of pathogens, promotion of plant growth, and induction of systemic resistance in plants [18]. For example, *B. amyloliquefaciens* strain LX-11 could limit rice bacterial stripe disease by secreting lipopeptides and surfactants [18]; strain LX1 isolated from Hainan saline soil could prevent and control banana wilt by secreting antibacterial proteins [19]; and strain HN06 showed an excellent inhibitory effect on *Aspergillus niger* and *Magnaporthe grisea* [20].

Bacillus can produce various substances with broad-spectrum antimicrobial activities, including lipopeptide antibiotics, bacteriocins, and antibacterial proteins [21]. Here we explored the factors affecting the biocontrol ability of strain IBFCBF-5 from the rhizosphere soil of pepper. Results of the experiment in vitro showed that strain IBFCBF-5 had an apparent inhibitory effect on oomycete and fungal pathogens causing pepper blight, pepper white silk disease, *Camellia oleifera* anthracnose, and cucumber fusarium wilt, indicating that the strain had great potential as a biocontrol agent. To help identify the potential biocontrol mechanism of strain IBFCBF-5, we obtained and analyzed its whole genome sequence.

The genome of strain IBFCBF-5 was compared with the GO database, KEGG database, and COG database. Comparison with the GO database showed that 67.2% of GO terms belonged to catalytic activity, the most significant proportion. Comparison with the KEGG database showed that biosynthesis of amino acids accounted for a large proportion of "Metabolism". Finally, comparison with the COG database showed a high abundance of genes related to transcription, translation, and core metabolisms. Together, these features of the IBFCBF-5 genome are similar to those of most other *Bacillus* strains.

We also compared the IBFCBF-5 genome with the CAZyme database and identified the gene clusters involved in the production of bioactive secondary metabolites. A total of 174 CAZyme family genes were

obtained from strain IBFCBF-5, including glycoside hydrolase (23), glycoside transferase (10), carbohydrate esterase (8), carbohydrate-binding module (9), auxiliary activities(6), and polysaccharide lyase (2), that can degrade cellulose and hemicellulose, chitin, starch, xylan, peptidoglycan, etc. The cell wall of most pathogenic fungi is mainly composed of cellulose, glucan, chitin, and so on. The results of the CAZyme analyses of strain IBFCBF-5 show that it can degrade components of the fungal cell wall, inhibiting fungal spore germination and mycelial growth. Furthermore, we found 14 lipopeptide and polyketone antibiotic gene clusters in the genome of strain IBFCBF-5, including five gene clusters with unknown functions, six antibiotic synthesis gene clusters with similarity > 90% (Fengycin, Bacillaene, macrolactin H, Bacillibactin, Bacilysin, plantazolicin), and one antibiotic gene cluster (Butirosin) with 7% similarity. Among these known antibiotics, Fengycin has solid antifungal activity and weak bacteriostatic activity [22, 23]; Bacilysin is a dipeptide with extensive inhibitory activity against fungi and bacteria [24, 25]; Bacillibactin and plantazolicin have inhibitory activity against bacteria [26]; and Bacillibactin also have antifungal inhibitory activity [27]. Moreover, the strain IBFCBF-5 contains bacteriostatic gene clusters similar to that in Macrolactin and Bacillaene, indicating that this strain has a broad potential of inhibiting pathogenic fungi as well as pathogenic bacteria.

Conclusion

In this paper, a strain of *B. amyloliquefaciens* IBFCBF-5 was isolated and shown to have an inhibitory effect on *Phytophthora capsici* and *Sclerotinia sclerotiorum, Colletotrichum gloeosporioides*, and *Fusarium oxysporum* f. sp. *cucumerinum.* To explore the biocontrol mechanism of the strain, we obtained its whole genome sequence. The whole genome data confirmed its taxonomic identity and showed that the strain has the potential to degrade most components of the fungal cell wall and to secrete antimicrobial metabolites. The genome sequence results identified the putative candidate genes of strain IBFCBF-5 from which further explorations on its biocontrol mechanisms could be targeted.

Method

Strain screening and identification

The rhizosphere bacteria were isolated using protocols described previously [Ref]. All strains were inoculated on LB plate and cultured at 30℃ for 18 hours and the colony morphology was observed. Each of the strains were tested for its ability to inhibit the growth of four plant pathogens. Specifically, using the standard culture method [28], we grew cultures of four plant fungal and oomycete pathogens *Phytophthora capsica*, *Sclerotinia sclerotiorum, Colletotrichum gloeosporioides*, and *Fusarium oxysporum* f. sp. *cucumerinum*. To test for their susceptibility to rhizosphere bacteria, a mycelial block of 5 mm diameter was inoculated in the center of a 9 cm diameter PDA plate, and then strain IBFCBF-5 was spotted at a distance of 2.5 cm from the center of the plate. Each treatment was repeated three times. The control plates had no culture of strain IBFCBF-5. The plates were placed upside down in an incubator at 28℃. The diameters of the zone of inhibition were measured, and the bacteria with the largest inhibition zone was selected as the candidate strain for follow-up research [29, 30, 31].

For species identification, the genomic DNA was extracted from each specimen using the TIANquick FFPE DNA Kit (TIANGEN, China). Forward primers B27F (5'-AGAGTTTGATCCTGGCTCAG-3'), and reverse primer U1492R(5'-GGTTACCTTGTTACGACTT-3') were used to amplify the 16S rRNA gene. Phusion® High-Fidelity PCR Master Mix (New England Biolabs, USA) was used for PCR reaction. Each PCR reaction (25 μL) contains 10×PCR Buffer (2.5 μL, Product name: DBI-2370), template DNA (1 μL), dNTPs(2.5 mmol/L, 4μL), $MgCl_2$ (25 mmol/L, 3 μL), primers (1 mmol/L; 1 μL each, Tsingke Biotechnology Co. Ltd.), TaqDNA polymerase (0.5 μL, KOD DNA Polymerase, Number: BTN101002)and ddH2O (10.5 μL). PCR conditions were set at 94°C for 5 min, followed by 35 cycles of 94℃ for 30s, 56℃ for 30s, and 72℃ for 1 min, and a final extension of 10 min at 72℃. The amplified products were sequenced by Beijing Kinco Biotechnology Co. Ltd. The sequences were compared with those in GenBank for species identity (Available at: http://blast.ncbi.nlm.nih.gov/Blast.cgi) [32].

Extraction of High-Quality DNA from Strain IBFCBF-5 for Genome Sequencing

The strain was inoculated in LB liquid medium and cultured aerobically at 30℃ for 15 hours, until the optical density value reached 2.697. The cells were collected by centrifugation at a speed of 6500 r/min for 15 minutes, and the collected bacteria weighed about 5 g. The genomic DNA was extracted using the bacterial DNA kit (E.Z.N.A. Bacterial DNA Kit, Omega Bio-Tek), following the manufacturer's instructions.

The extracted genomic DNA was sent to Beijing Biomarker Biotechnology Co. Ltd. for sequencing. Sequencing was conducted using both the Illumina Hiseq platform and the PacBio Sequel third-generation single-molecule real-time sequencing system. Low-quality reads were filtered through SMRT 2.3.0 [33].

Genome assembly and sequence analysis of strain IBFCBF-5

The genome of strain IBFCBF-5 was assembled using Canu v1.5 software [34]. The Pilon software [35] was used to correct any mistakes in the assembled genome by using results of Illumina Hiseq sequencing.

For gene prediction, the Repeat Masker software was used to predict and mask the repetitive sequences of the assembled genome [36]. We used the Prodigal software to predict the protein-coding genes [37], tRNAscan-SE to predict the transfer RNA (tRNA) genes, and Infernal 1.1 to predict the ribosomal RNA (rRNA) genes and other ncRNAs except for tRNA and rRNA[38, 39]. Using the predicted protein sequences and the protein sequences in the Swiss-Prot database, we compared the homologous gene sequences in the IBFCBF-5 genome by software GenBlastA [40], and then used the software GeneWise [41]to find premature stop codons and frameshift mutations in gene sequences to obtain pseudogenes. Using the predicted genome information, we obtained estimates of repeat sequences, GC content, etc. We used the program Circos to draw the circular genome map [42].

Analysis of Genome Evolution of strain IBFCBF-5

The genome-wide average nucleotide identity (ANI) refers to the overall similarity of homologous genes between two genomes. It is generally believed that the ANI value between strains of the same species needs to reach more than 95% [43]. To further determine the taxonomic status of *Bacillus* strain IBFCBF-5, the online ANI (https://www.ezbiocloud.net/tools/ani)) was used to analyze the taxonomic status of IBFCBF-5 strains according to its genome sequence. DNA-DNA hybridization (DNA-DAN hybridization, DDH) refers to DNA molecules with complementary base sequences, forming stable double-stranded regions such as hydrogen bonds between base pairs. The DDH value of the IBFCBF-5 strain was analyzed by online DDH (http://ggdc.dsmz.de/ggdc.php) to further determine its taxonomic status.

The relationships among strains were analyzed using their 16S rRNA sequences and by comparing with various reference strains to construct a phylogenetic tree. In addition, the whole-genome sequences of strains CC09, XJ5, 2J01, WF02, YP6, S4, ATCC 14580, CW14, SP1, BS49, and ZJU1 published previously were compared with that of our strain IBFCBF-5 to derive the concatenated SNP profiles. The concatenated SNPs were analyzed using the PhyML (version 3.0) software [44], and the phylogenetic tree was constructed by the ML method (maximum likelihood method).

Functional Annotation of the Genome of strain IBFCBF-5

The predicted gene sequences were compared with available databases including COG [45], KEGG [46], Swiss-Prot [47], TrEMBL [48], Nr [49] and other functional databases. The results of functional gene annotation were obtained. Based on the comparison results of the Nr database, the application software Blast 2 GO [49, 50] annotates the function of the GO database. Furthermore, the software hmmer [51] was used to annotate the Pfam function based on the Pfama [52] database. In addition, we conducted COG and KEGG metabolic pathway enrichment analysis, GO functional enrichment analysis, and other gene function annotation analyses.

Comparative Analysis of CAZyme Database of IBFCBF-5 strains

According to their functions, carbohydrate-active enzymes (carbohydrate active enzymes, CAZyme) mainly include glycoside hydrolase, glycosyltransferase, polysaccharide lyase, carbohydrate esterase, co-oxidoreductase, and carbohydrate-binding modules without catalytic activity. Based on the carbohydrate active enzyme database CAZyme, the functional annotation and analysis of carbohydrate enzyme genes were carried out by using hummer software [52, 53].

Analysis of gene clusters related to secondary metabolites of strain IBFCBF-5

The anti-SMASH software (version5.1.2) was used to analyze and predict the gene clusters related to the synthesis of antagonistic substances in the genome of the IBFCBF-5 strain. According to the preliminary online prediction results, the corresponding genes related to antibacterial substance synthesis were downloaded and compared one by one to determine whether the antagonistic genes in the IBFCBF-5 genome were deleted or mutated.

Statistical analysis

Statistical analysis was performed using Microsoft Excel 2010 (Microsoft Corporation, Redmond, WA, USA) and DPS 7.05 software (Zhejiang University, Hangzhou, China). Mean values were compared using Duncan's multiple range test with $P < 0.05$ as the level of significance.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data set generated and/or analyzed during the current study can be found in the NCBI repository [https://www.ncbi.nlm.nih.gov/], including a total of 1 sequences, with accession numbers: SUB9291413 IBFCBF-5,. The data will be released in October 2021.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Key R&D Program of China (2018YFC0311300), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-IBFC07), and the Biology discipline construction funds.

Authors' contributions

Zeng L-B, Li J-L and Wei L, Chen Y developed the idea for the study and guided it all the way. Luo C and Ren Y-S Shen A-R designed and completed this study. All authors analyzed the data and were involved in writing the manuscript. Xu J-P and Luo C finished revising the English manuscript.

Acknowledgements

Thank you very much for your contribution to this article and thank you for their help and support.

References

1. Ongena M, Jacques P. Bacillus lipopeptides: versatile weapons for plant disease biocontrol – Science Direct[J]. Trends in Microbiology, 2008, 16(3):115-125.

2. Bhardwaj D, Ansari M W, Sahoo R K, Tuteja N. Biofertilizers function as key player in sustainable agriculture by improving soil fertility, plant tolerance and crop productivity[J]. Microbial Cell Factories, 2014, 2014, 13:66.
3. Ali S, Hameed S, Imran A, Iqbal M, Lazarovits G. Genetic, physiological and biochemical characterization of Bacillus sp. strain RMB7 exhibiting plant growth promoting and broad spectrum antifungal activities[J]. Microbial Cell Factories, 2014, 13:144
4. Yang S-Y, Wei J, Zhen X-R. Optimization of fermentation medium for antibacterial substances of *Bacillus amyloliquefaciens*. Food Science , 2015:150-156.
5. Jin Q, Xiao M. Novel antimicrobial peptides: surfactin, iturin and fengycin[J]. Journal of Microbes and Infections, 2018, 13(1): 56-64.
6. Planchot V, Colonna P. Purification and characterization of novel antimicrobial peptides from the skin secretion of *Hylarana guentheri*[J]. Proceedings of the National Academy of Sciences India, 2018, 27(1):3077-3084.
7. Finn R D, Mistry J, Tate J, Coggill P C, Sammut S J, Hotz H R, Ceric G, Forslund K, Eddy S R, Sonnhammer E L. L., Bateman A. The Pfam protein families database[J]. Nuclc Acids Research, 2008, 36:D281-D288.
8. Zhang D-F, Gao Y-X, Ke X-L, Wang Y-J, Ren Y, Shi C-B. Analysis of the genome sequence of *Bacillus Velez* LF01 and the biocontrol effect of its metabolites[J]. Journal of Fisheries, 2021:. doi:10.11964/jfc.20201012442.
9. Huang X-Q, Gao X-P, Zhao G-M, Li M-Y, Liu Y-X. Research Progress on Antibacterial Mechanism of Antimicrobial Peptides[J]. Journal of Biology, 2010, 27(2):62-66.
10. Zhao X, Wang W-L, Li J. Isolation and Biology identification of a strain of high-temperature cellulase-producing bacteria[J]. Gansu Agricultural Science and Technology, 2018, 6:30-33.
11. Song H-M, Zhang B-L, Fan W-W, Xu B-H, Li L-J. Research Progress of Cellular Fatty Acid Analysis Technology in the Field of Microbiology[J]. Chinese Journal of Health Laboratory Science, 2008, 10: 248-249.
12. Molohon K, Blair P, Park S, Doroghazi J R, Maxson T, Hershfield J R, Flatt K M, Schroeder N E, Ha T, Mitchell D A. Plantazolicin is an ultra-narrow spectrum antibiotic that targets the *Bacillus anthracis* membrane[J]. ACS Infectious Diseases, 2016:207.
13. Wang Z-Y, Mao Z-C, Wu X-X, Zhou H-P, Wu Y-X. Preliminary study on the synthesis pathway of the antibacterial active substance difficidin of the biocontrol bacteria B9601-Y2[A]. 2009 Annual Meeting of the Chinese Society of Phytopathology[C], 2009.
14. Yang Q, Han W-J, Zhang W-J, Lu X-L, Liu X-Y, Xu Q-Z, Miao H-N, Jiao B-H. Isolation and identification of a strain of marine *Bacillus amyloliquefaciens* producing a new type of macrolide antibiotic[J]. Chinese Journal of Maritime Medicine and Hyperbaric Medicine, 2009, 16 (2): 92-95.
15. Li X-P, Li Y. Construction and properties of No.44 recombinant plasmid containing the promoter DNA fragment of *Bacillus circulans* NRRL-B3312, which produces Butirosin[J]. Chinese Journal of Antibiotics, 1988, (05): 3-10.

16. Zhang Q-Q, Zhu H-M, Yang X-J, Huang Z-W, Gong M- X, Zhang Z-Q, Niu J-J, Yang H-J. Biosynthesis and Structure Analysis of Catechol Siderophore Produced by *Bacillus subtilis* MB8 [J]. Southwest China Journal of Agricultural Sciences, 2019, 32 (06):1259-1266.
17. Wang T. Preparation of antifungal compound bacilysin, study on antibacterial mechanism and exploration of structural modification[D]. Zhejiang University, 2016.
18. Hunter P. Fight fire with fire. Fight fire with fire. Can biopesticides fill the void left by banning chemical pesticides and herbicides?[J]. Embo Reports, 2009, 10 (5):433-436.
19. Li L, Zhang Z-Y, Sun H-Y, Fu W-Q. Discussion on Pesticide Harm and Green Plant Protection Technology[J]. Hubei Plant Protection, 2018, (04): 63-64.
20. Qiu D-W. Development Status and Trend Analysis of Biological Pesticides[J]. Chinese Journal of Biological Control, 2015, 31(05):679-684.
21. Chu X-M, Lin Y-X, Zhang K, Yan F, Lin H-T. Research Progress on Antibacterial Proteins of *Bacillus amyloliquefaciens*[J]. Packaging and Food Machinery, 2014, (06): 49-54.
22. Ongena M, Jacques P. *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol - ScienceDirect[J]. Trends in Microbiology, 2008, 16(3):115-125.
23. Zhi Y. Metabolic mechanism of highly-efficient biosynthesis of surfactin and its function[D]. Wuxi: Jiangnan University, 2017.
24. Tao W, Liu X-H, Wu M-B, Ge S. Molecular insights into the antifungal mechanism of bacilysin[J]. Journal of Molecular Modeling, 2018, 24(5):118.
25. Chen C-P, Jia P-X. Gene regulation of *Bacillus* protease[J]. Microbiology Bulletin,1992, (02): 40-43.
26. Chen Z-Y. Research and application of bio-fungicide with *Bacillus* spp.[J]. Chinese Journal of Biological Control, 2015, 31(5): 723-732.
27. Zhang D-F, Gao Y-X, Wang Y-J, Liu C, Shi C-B. Advances in taxonomy, antagonistic function and application of *Bacillus velezensis*[J]. Microbiology China, 2020, 47(11): 3634-3649.
28. Ren X, Nan Z, Cao M, Wu K, Shen Q-R, Huang Q-W. Biological control of tobacco black shank and colonization of tobacco roots by a *Paenibacillus polymyxa* strain C5[J]. Biology and Fertility of Soils, 2012, 48(6):613-620.
29. Khedher S B, Kilani-Feki O, Dammak M, Jabnoun-Khiareddine H, Daami-Remadi M, Tounsi S. Efficacy of *Bacillus subtilis* V26 as a biological control agent against *Rhizoctonia solani* on potato[J]. Comptes Rendus Biologies, 2015, 338(12):784-792.
30. Anith K N, Radhakrishnan N V, Manomohandas T P. Screening of antagonistic bacteria for biological control of nursery wilt of black pepper (*Piper nigrum*)[J]. Microbiological Research, 2003, 158(2):91-97.
31. Du N, Shi L, Yuan Y, Sun J, Shu S, Guo S-R. Isolation of a potential biocontrol agent *Paenibacillus polymyxa* NSY50 from vinegar waste compost and its induction of host defense responses against Fusarium wilt of cucumber[J]. Microbiological Research, 2017, 202:1-10.

32. Altschul S F. Basic local alignment search tool (BLAST)[J]. Journal of Molecular Biology, 2012, 215(3):403-410.
33. Berlin K, Koren S, Chin C-S, Drake J P, Landolin J M, Phillippy A M. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing.[J]. Nature biotechnology, 2015, 33:623-630.
34. Koren S, Walenz B P, Berlin K, Miller J R, Bergman N H, Phillippy A M. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation.[J]. Genome research, 2017, 27:722-736.
35. Walker B J, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo C A, Zeng Q-D, Wortman J, Young S K, Earl A M. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement[J]. PLoS ONE, 2014, 9(11):e112963.
36. Tarailo-Graovac M, Chen N-S. Using Repeat Masker to Identify Repetitive Elements in Genomic Sequences[J]. Current Protocols in Bioinformatics, 2009, 25:4.10.1-4.10.14.
37. Hyatt D, Chen G-L, Locascio P F, Land M L, Larimer F W, Hauser L J. Prodigal: prokaryotic gene recognition and translation initiation site identification[J]. BMC Bioinformatics, 2010, 11: 119.
38. Lowe T M, Eddy S R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence[J]. Nucleic Acids Research, 1997, 25(5):955-964.
39. Nawrocki E P, Eddy S R. Infernal 1.1: 100-fold faster RNA homology searches[J]. Bioinformatics, 2013(22):2933-2935.
40. She R, Chu J S.-C., Wang K, Pei J, Chen N-S. genBlastA: Enabling BLAST to identify homologous gene sequences[J]. Genome Research, 2009, 19:143-149.
41. Birney E, Clamp M, Durbin R. GeneWise and Genomewise[J] Genome Research, 2004, 14:988-995.
42. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones S J, Marra M A. Circos: An information aesthetic for comparative genomics[J]. Genome Research, 2009, 19:1639-1645.
43. Hui W-Y, Zhang H-P. Application of genomic analysis in Microbial Taxonomy[J]. Microbiology China, 2016, 43(5):1136-1142.
44. Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0[J]. Systematic Biology, 2010, 59(3):307-321.
45. Tatusov R L, Galperin M Y, Natale D A, Koonin E. The COG database: a tool for genome-scale analysis of protein functions and evolution[J]. Nucleic Acids Research, 2000, 28:133-136.
46. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome[J]. Nucleic Acids Research, 2004, 32(1):D277-D280.
47. Boeckmann, B, Bairoch A, Apweiler R, Blatter M C, Estreicher A, Gasteiger E, Martin M J, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003[J]. Nucleic Acids Research, 2003, 31(1):365-370.
48. Deng Y-Y, Li J-Q, Wu S-F, Zhu Y-P, Chen Y-W, He F-C. Integrated nr database in protein annotation system and its localization[J]. Computer Engineering, 2006, 32(5):71-73, 76.

49. Altschul S F, Madden T L, Schäffer A A，Zhang J-H，Lipman D J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs[J]. Nucleic acids research, 1997, 25(17):3389.
50. Conesa A, Gotz S, Garcia-Gomez J M , Terol J，Talon M，Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research[J]. Bioinformatics, 2005, 21(18):3674-3676.
51. Eddy S R . Profile Hidden Markov Models[J]. Bioinformatics, 1998, 14(9): 755-763.
52. Finn R D, Coggill P, Eberhardt R Y, Eddy S R, Mistry J, Mitchell A L, Potter S C, Punta M, Qureshi M, Sangrador-Vegas A, Salazar G A, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future[J].Nucleic acids research, 2016, 44: D279-D285.
53. Finn R D, Bateman A, Clements J, Coggill P, Eberhardt R Y, Eddy S R, Heger A, Hetherington K, Holm L, Mistry J. Pfam: the protein families database[J]. Nucleic Acids Research, 2014, 42:D222-D230.

Figures

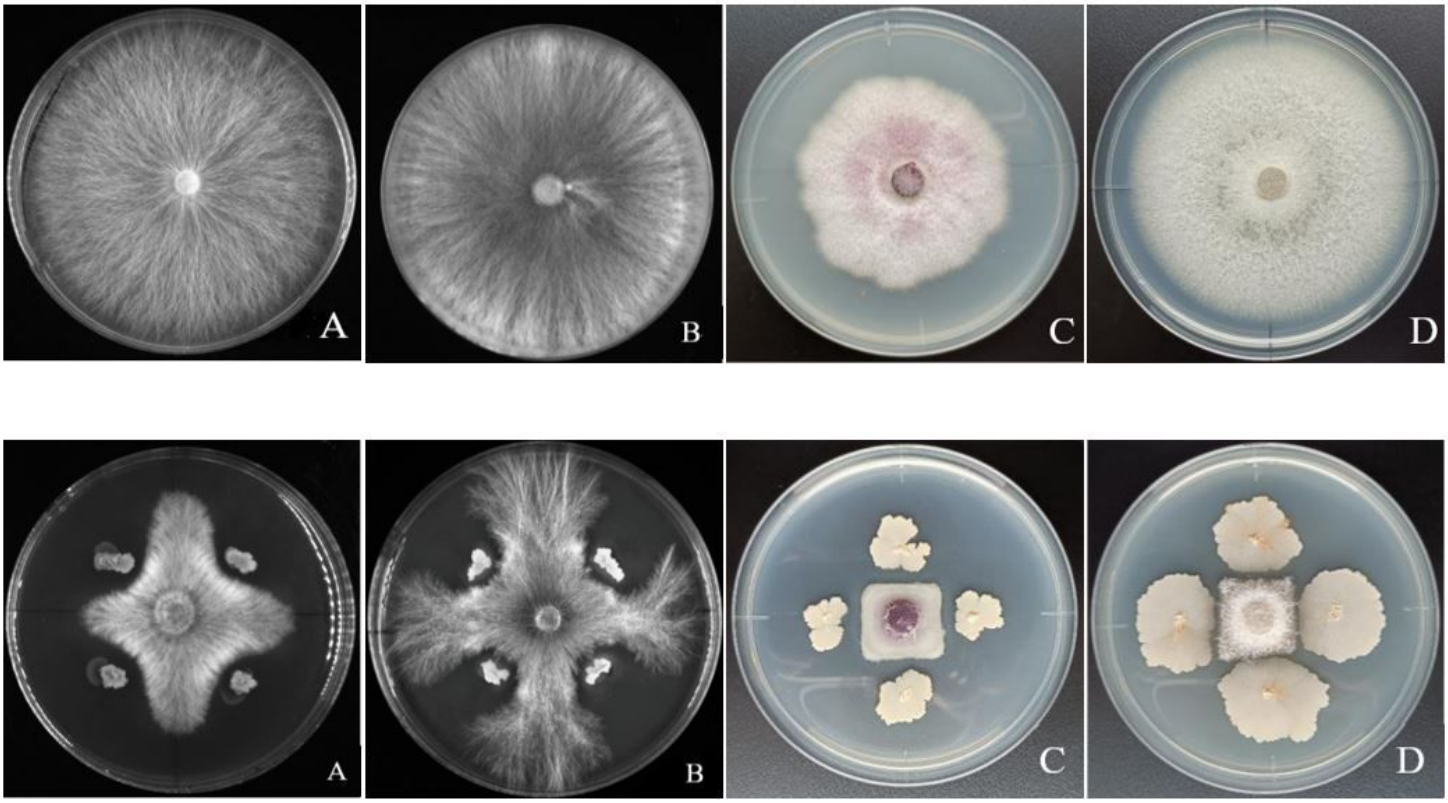

Figure 1

Antagonistic effect of strain IBFCBF-5 on plant fungal pathogens. Note: A: Sclerotinia sclerotiorum; B: Phytophthora capsici C: Fusarium oxysporum f. sp. Cucumerinum; D: Colletotrichum gloeosporioides The upper represents the pathogenic fungus control experiment, the lower represents the antagonistic antibacterial and antibacterial experimental bacteria.

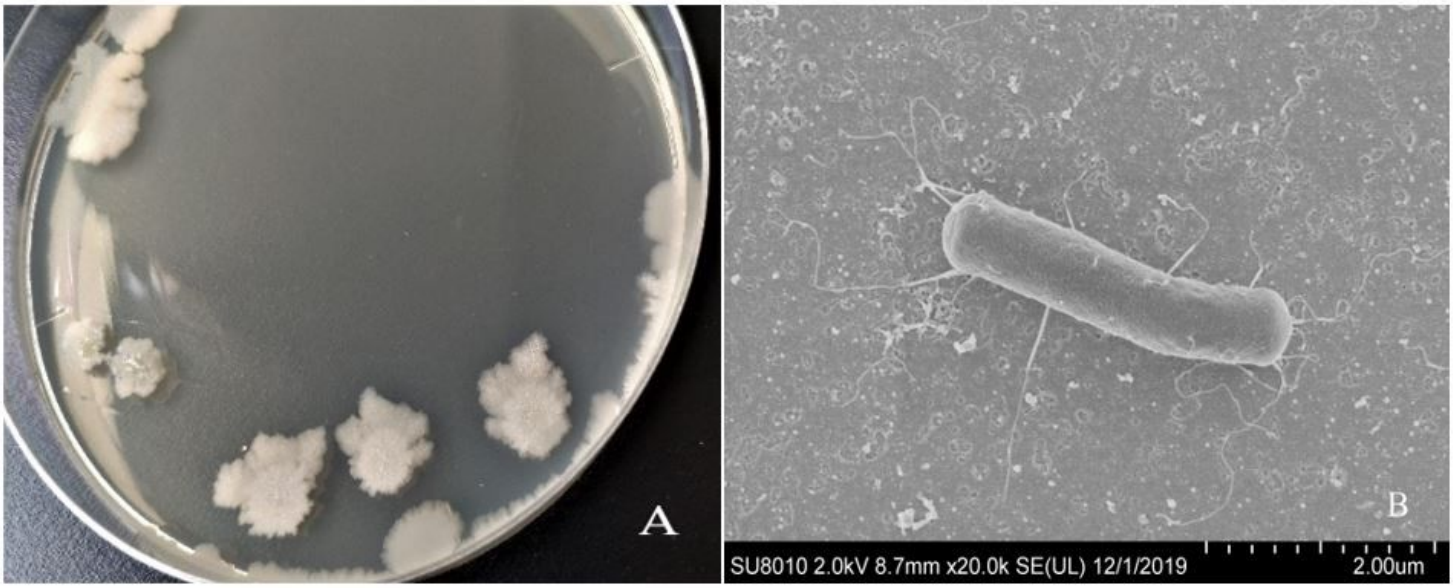

Figure 2

Morphological characteristics of strain IBFCBF-5 Note: A: IBFCBF-5 colony morphology, and B: IBFCBF-5 cell surface morphology based on electron microscopy.

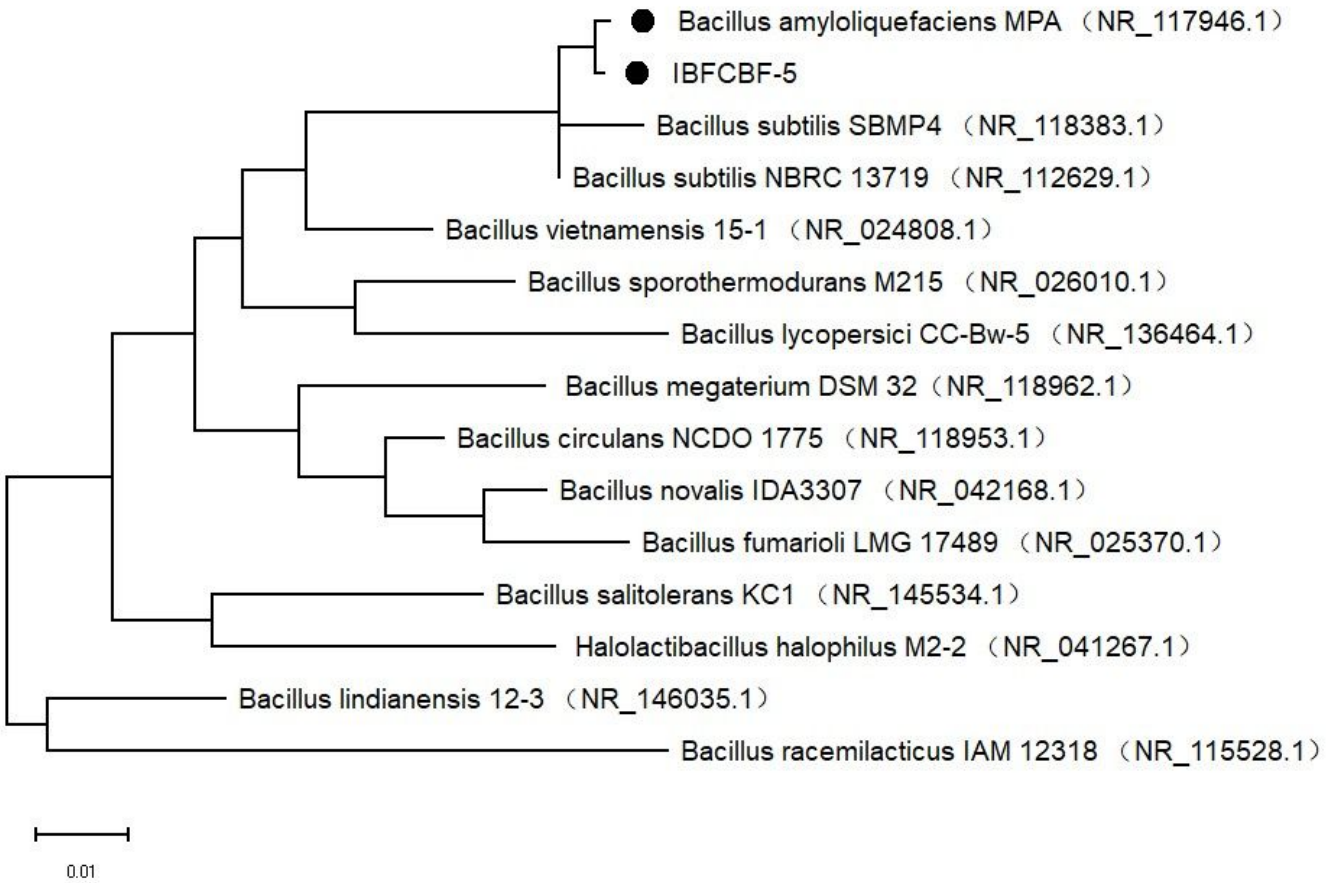

Figure 3

Phylogenetic tree of strain IBFCBF-5 and its related bacteria based on 16S rRNA gene sequences.

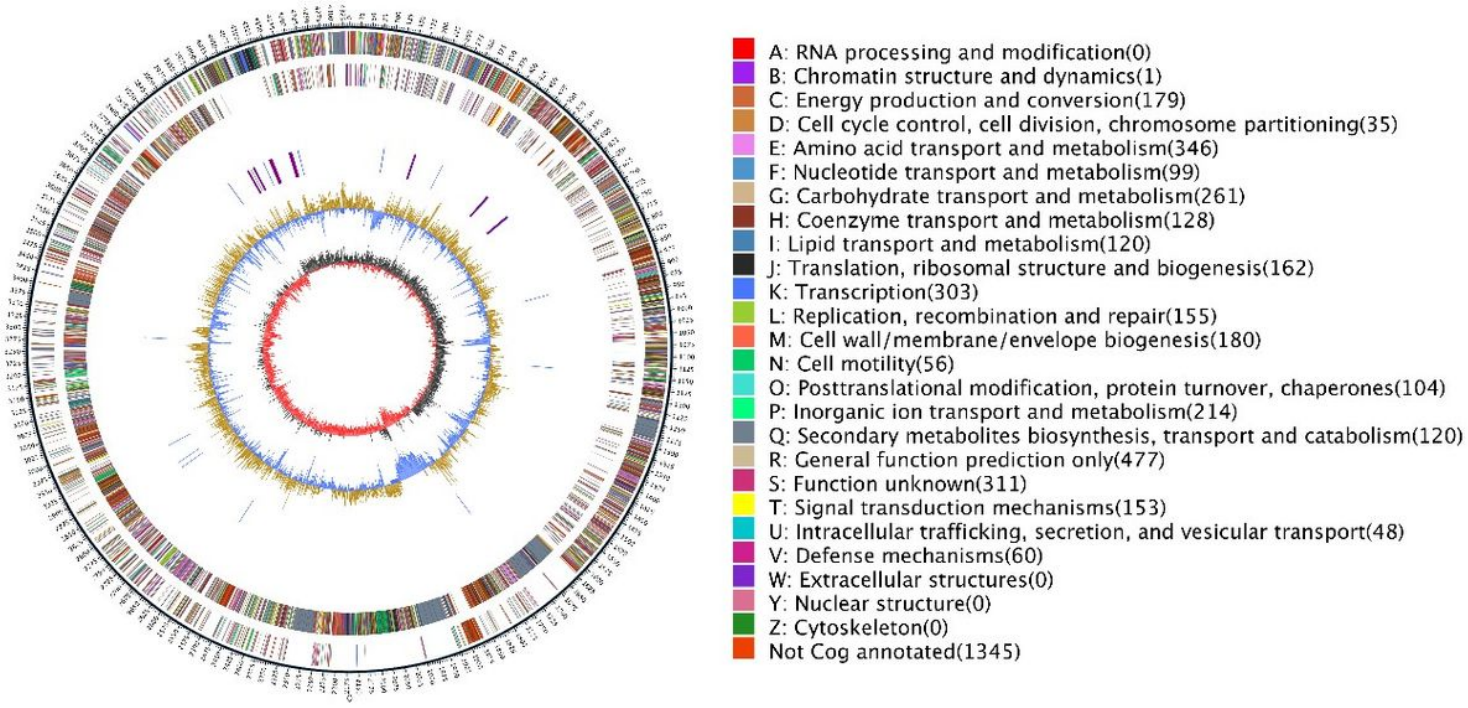

Figure 4

Genome circle diagram of Bacillus amyloliquefaciens IBFCBF-5 strain Note: The outermost circle refers to genome size and gene position, with the distance between two bars representing 5 kb; the second circle and the third circle are genes on the positive and negative strands of the genome, and different colors represent different COG functional classifications; the fourth circle is repetitive sequences; the fifth circle is tRNA and rRNA, with blue representing tRNA, and purple representing rRNA; the sixth circle is the GC content, the light yellow part indicates that the GC content of the region is higher than the average GC content of the genome, (the higher the peak value, the greater the difference), the blue part indicates that the GC content of the region is lower than the average GC content of the genome; the innermost circle is GC-skew, the dark gray represents the area with G content greater than C, and the red represents the area with C content greater than G. Different colors represent different COG function classifications. The colors on the right represent their annotated functions of other groups of genes (the numerals in parenthesis represent the number of genes in each category): A: RNA processing and modification；B: Chromatin structure and dynamics(1)；C: Energy production and conversion (179)；D: Cell cycle control, cell division, chromosome partitioning (35)；E: Amino acid transport and metabolism (346)；F: Nucleotide transport and metabolism (99)；G: Carbohydrate transport and metabolism (261)；H: Coenzyme transport and metabolism (128)；I: Lipid transport and metabolism (120)；J: Translation, ribosomal structure and biogenesis (162)；K: Transcription (303)；L: Replication, recombination and repair (155)；M: Cell wall/membrane/envelope biogenesis (180)；N: Cell motility (56)；O: Posttranslational modification, protein turnover, chaperones (104)； P: Inorganic ion transport and metabolism (214)；Q: Secondary metabolites biosynthesis, transport and catabolism (120)；R: General function prediction only (477)；S: Function unknown (311)；T: Signal transduction mechanisms (153)；U：Intracellular trafficking, secretion, and

vesicular transport (48)□V: Defense mechanisms (60)□W: Extracellular structures (0)□Y: Nuclear structure (0)□Z: Cytoskeleton (0)□□Not Cog annotated(1345).

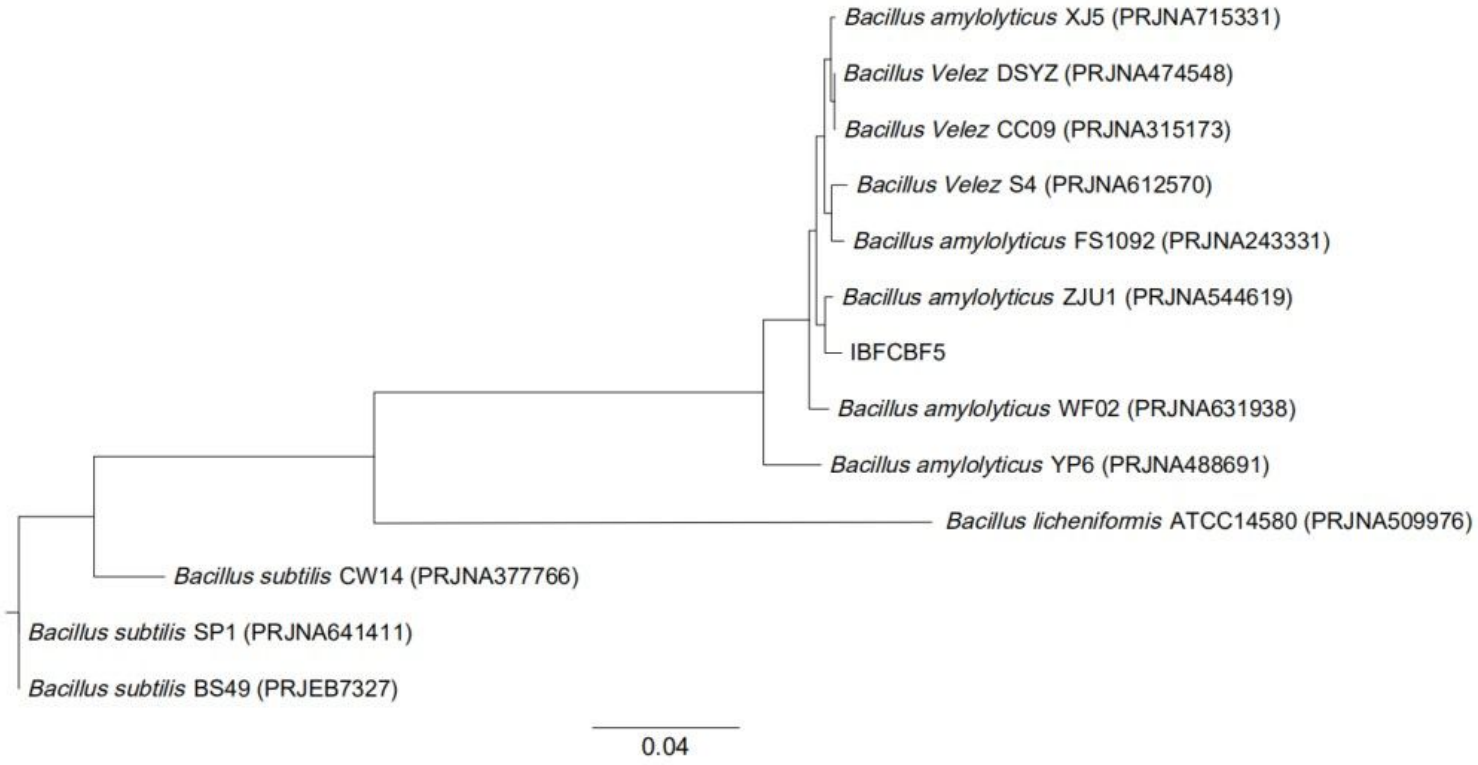

Figure 5

Phylogenetic tree of the IBFCBF-5 and the reference strains base on the whole-genome SNP

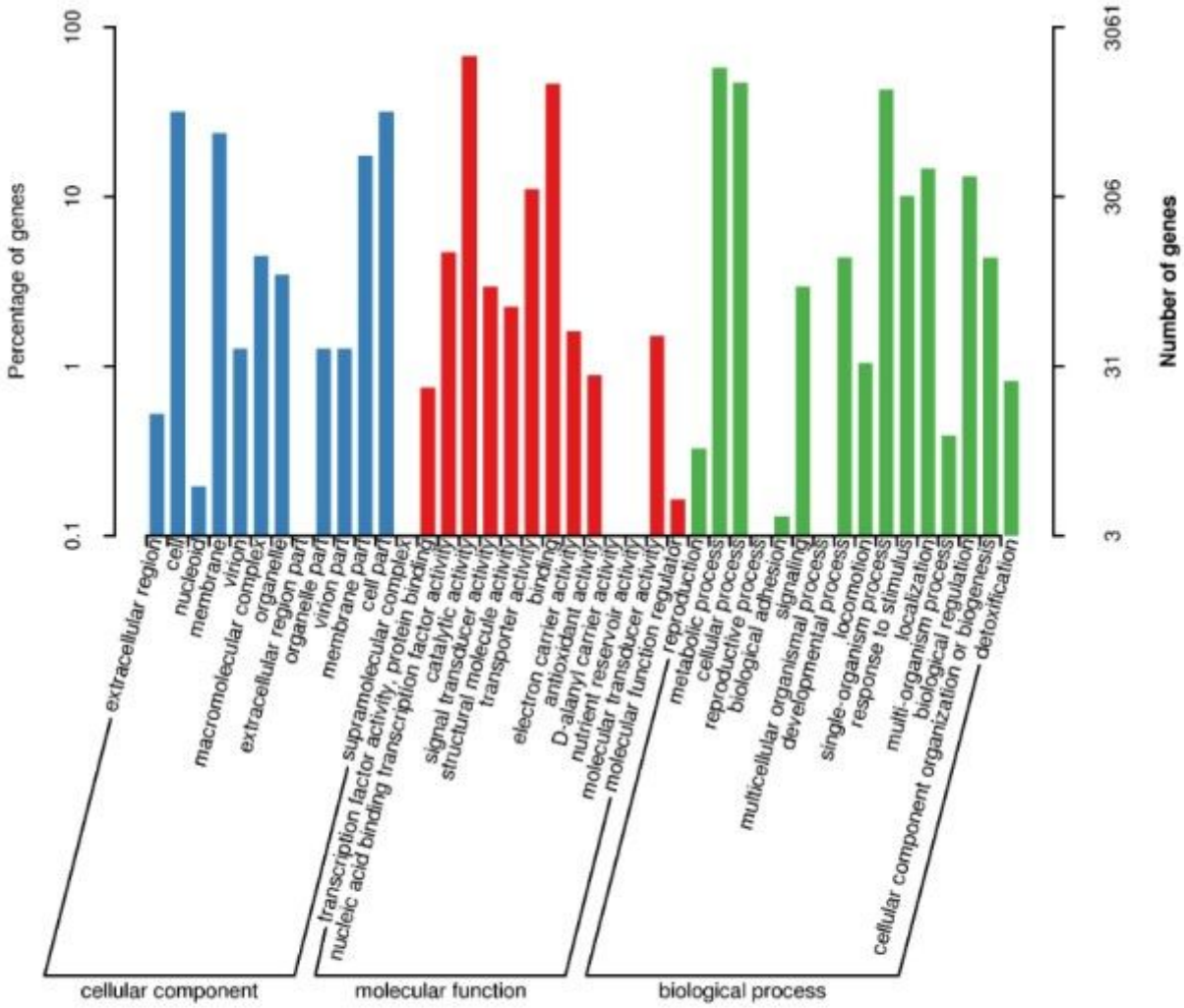

Figure 6

The classification statistics of the genome of strain IBFCBF-5 GO function annotation

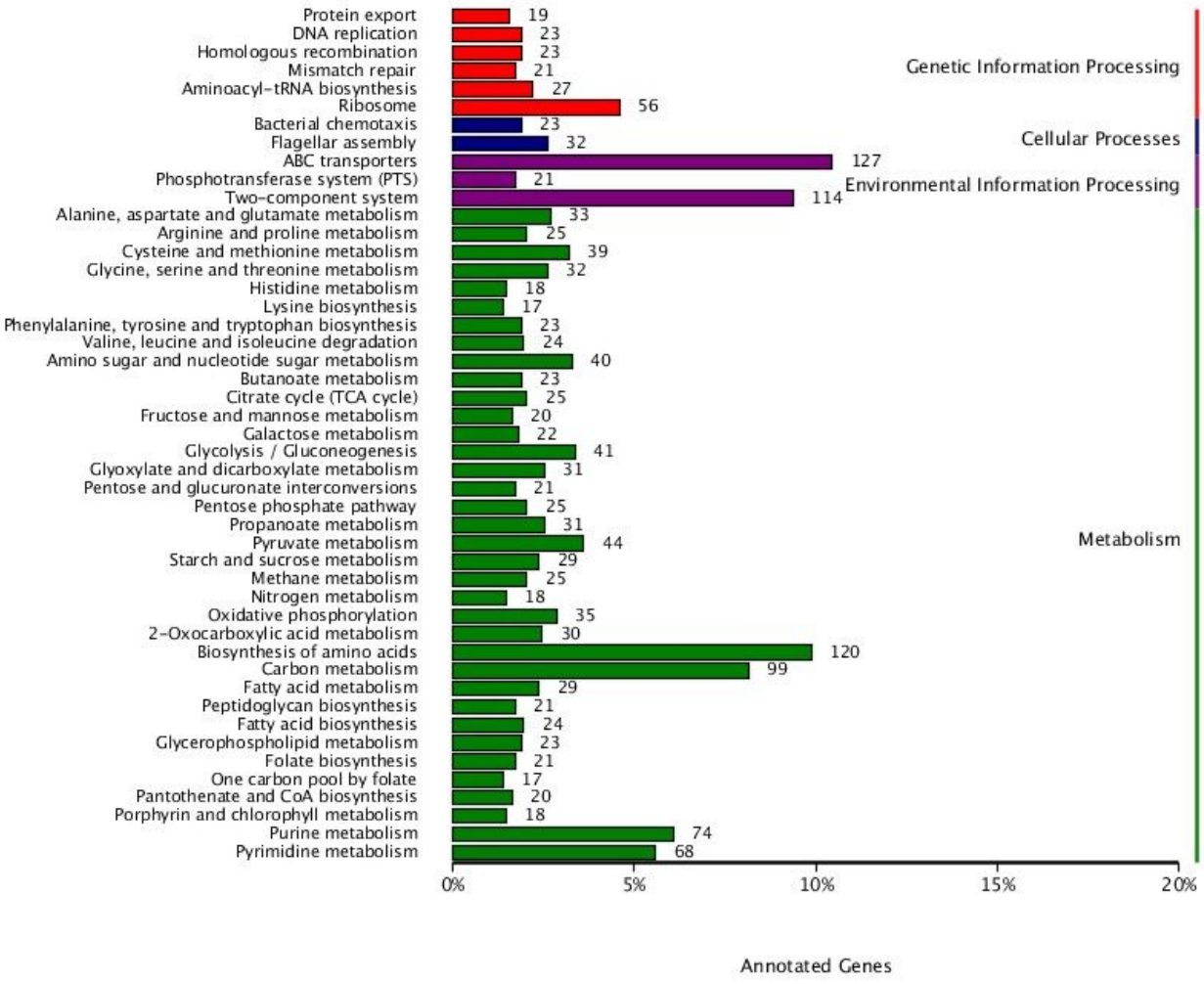

Figure 7

KEGG classification of the genome of strain IBFCBF-5

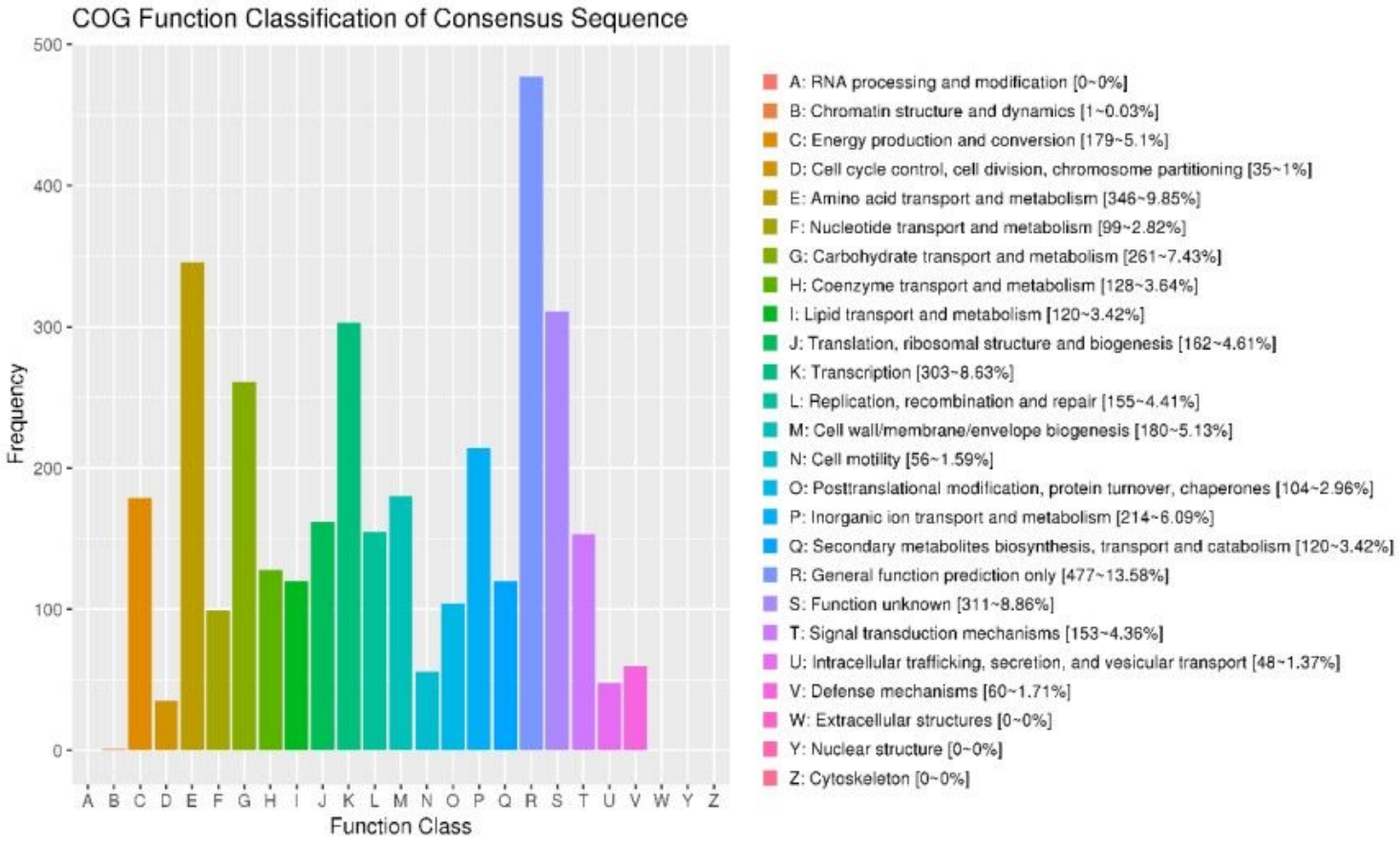

Figure 8

COG distribution of the genome of strain IBFCBF-5 Note: Abscissa is the content of COG classification, the ordinate is the number of genes. The function of A~Z explains the same as the diagram note in figure 4.

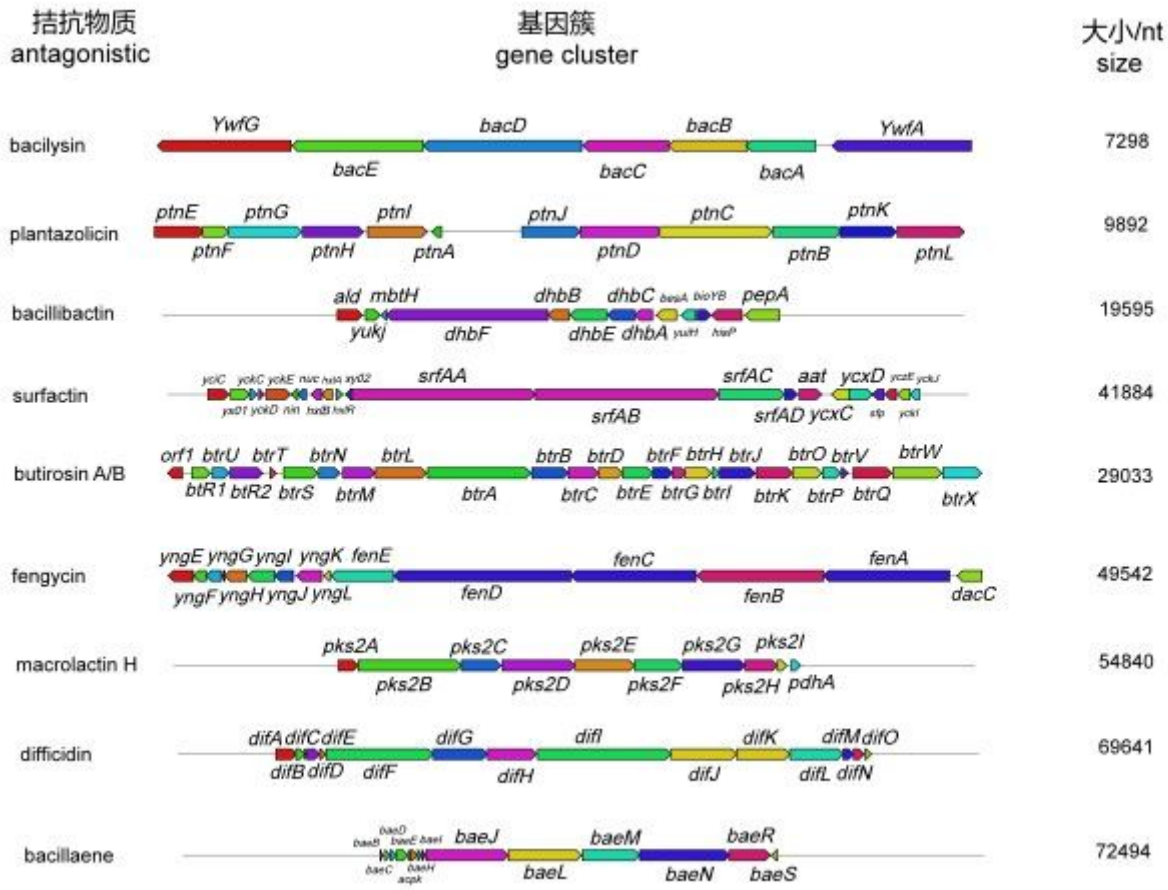

Figure 9

Gene clusters in strain IBFCBF-5 involved in the biosynthesis of nine secondary metabolites related to antimicrobial substances.