



Comparative Genomics Analysis Provides New Strategies for Bacteriostatic Ability of *Bacillus velezensis* HAB-2

Peidong Xu^{1,2†}, Shangqian Xie^{1,3†}, Wenbo Liu^{1†}, Pengfei Jin¹, Dandan Wei¹, Dahar Ghulam Yaseen¹, Yu Wang¹ and Weiguo Miao^{1*}

¹ Key Laboratory of Green Prevention and Control of Tropical Plant Diseases and Pests, Ministry of Education, College of Plant Protection, Hainan University, Haikou, China, ² School of Life and Pharmaceutical Sciences, Hainan University, Haikou, China, ³ College of Forestry, Hainan University, Haikou, China

OPEN ACCESS

Edited by:

Yoshio Yamaoka,
Oita University, Japan

Reviewed by:

Manoj Kumar Solanki,
Agricultural Research Organization
(ARO), Israel
Ivica Z. Dimkjae,
University of Belgrade, Serbia

*Correspondence:

Weiguo Miao
miao@hainu.edu.cn

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 12 August 2020

Accepted: 29 October 2020

Published: 17 November 2020

Citation:

Xu P, Xie S, Liu W, Jin P, Wei D,
Yaseen DG, Wang Y and Miao W
(2020) Comparative Genomics
Analysis Provides New Strategies
for Bacteriostatic Ability of *Bacillus*
velezensis HAB-2.
Front. Microbiol. 11:594079.
doi: 10.3389/fmicb.2020.594079

Biocontrol formulations prepared from biocontrol bacteria are increasingly applied in sustainable agriculture. Notably, inoculants prepared from *Bacillus* strains have been proven efficient and environmentally friendly alternatives to chemical bactericides. The bacterium *Bacillus velezensis* HAB-2 (formerly classified as *B. amyloliquefaciens* HAB-2) is used as a biological control agent in agricultural fields. In this study, we reported a high-quality genome sequence of HAB-2 using third-generation sequencing technology (PacBio RS II). The 3.89 Mb genome encoded 3,820 predicted genes. Comparative analysis among the genome sequences of reference strains *B. velezensis* FZB42, *B. amyloliquefaciens* DSM7 and *B. subtilis* 168 with the HAB-2 genome revealed obvious differences in the variable part of the genomes, while the core genome shared by FZB42 and HAB-2 was similar (96.14%). However, there were differences in the prophage region among the four strains. The numbers of prophage regions and coding genes in HAB-2 and FZB42 were smaller than the other two strains. The HAB-2 genome showed superior ability to produce secondary metabolites and harbored 13 gene clusters involved in synthesis of antifungal and antibacterial acting secondary metabolites. Furthermore, there were two unique clusters: one cluster which encoded lanthipeptide was involved in mersacidin synthesis and another cluster which encoded ladderane was shown to direct an unknown compound. Multidomain enzymes, such as non-ribosomal peptide synthetase and polyketide synthase, control the biosynthesis of secondary metabolites and rely on 4'-phosphopantetheinyl transferases (PPTases). Key genes *lpaH2* and *a* encoded PPTases in HAB-2 encoded 224 and 120 amino acids, respectively. The genomic features revealed that HAB-2 possesses immense potential to synthesize antimicrobial acting secondary metabolites by regulating and controlling gene clusters. The prophage regions and genes encoding PPTases may provide novel insight for the bacteriostatic mechanism of *Bacillus* in the biological control of plant diseases.

Keywords: *Bacillus velezensis*, genome, secondary metabolites, prophage, 4'-phosphopantetheinyl transferases

INTRODUCTION

Pathogenic microorganisms are a major and chronic threat to food production and ecosystem stability all over the world by infecting plant tissue and affecting plant health (Compant et al., 2005). As the human population has increased and agricultural production has intensified over the past few decades, producers have been increasingly dependent on agrochemicals to control plant diseases (Schäfer and Adams, 2015). However, excessive use of chemical pesticides has had a huge negative impact on the ecological environment. Increasingly, researchers and producers realize the need for alternatives to chemical pesticides to avoid damaging the environment. Use of beneficial bacteria as alternatives to chemical pesticides in plant protection is steadily increasing and is beginning to replace some chemical pesticides (Qiao et al., 2014). In China, according to the China Pesticide Information Network¹, a total of 41,614 pesticide products were registered in China at the end of May 2020. Among them, 10,971 antiseptics accounted for 26.3%. There were 434 actual microbial pesticides, and registered biological pesticides only accounted for 2.9% of the total pesticide products. Thus, prevention and control of plant diseases are still dominated by chemical pesticides in China, and the road to replace them with biological pesticides is long, but the development potential is huge. As a consequence, researchers are also focusing on the antibacterial mechanisms of biocontrol bacteria.

Bacillus velezensis is a biocontrol bacterium that is efficient, environmentally friendly and suitable for wide use for plant protection in agriculture. It produces natural products that are antimicrobial, antiviral and nematocidal (Mariappan et al., 2012). Recently, whole-genome assemblies of *Bacillus* spp. have been produced using high-throughput sequencing, and genome analysis of *B. velezensis* reveals their potential for biocontrol of plant pathogens (Chen et al., 2008). The *B. velezensis* genome harbors some important gene clusters involved in the synthesis of secondary metabolites. The cluster of protective molecules is generally made by non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) (Schneider et al., 2007; Chen et al., 2008; Beld et al., 2014). Five gene clusters (*srf*, *bmy*, *fen*, *nrs*, and *dhb*) were shown to direct synthesis of surfactin, fengycin and iturin (including bacillomycinD and mycosubtilin). Three gene clusters, *mln(pks1)*, *bae(pks2)*, and *dfn(pks3)*, were shown to direct synthesis of the polyketides macrolactin, bacillaene and diffidin, respectively (Koumoutsis et al., 2004). The size of these gene clusters generally accounts for 7.5–8.5% of the genome of the strain (Koumoutsis et al., 2004; Chen et al., 2008).

In *B. velezensis*, NRPS and PKS control the biosynthesis of secondary metabolites (Metz et al., 2001; Cox, 2007). However, these enzymes are inactive until they are post-translationally modified. The activation of apo-peptidyl carrier proteins of NRPS and apo-acyl carrier proteins of PKS to their active holo form is accomplished by dedicated 4'-phosphopantetheinyl transferases (PPTases) (Lambalot et al., 1996). The synthesis of non-ribosomal lipopeptides and polyketides are dependent on the PPTases, which are encoded by a functional *sfp* gene (Chen et al., 2008).

A PPTase *sfp* was first identified by Nakano et al. (1990, 1992) in attempts to transform the non-surfactin-producing strain *B. subtilis* JH642. Borchert et al. (1994) identified a gene, *gsp*, with 34% identity and 54% homology to *sfp*. Huang et al. (1993) discovered *lpa-14*, showing homology to *sfp*, which is related to iturin A. Because of their cornerstone role in metabolism, PPTases are considered antibiotic targets in agriculture. Being indispensable for the biosynthesis of natural products with antibiotic action, the PPTase super family genes play a leading role in antibiotic engineering efforts for agriculture.

The bacterium *B. velezensis* HAB-2 has the potential as a biological pesticide due to its excellent antibacterial ability. Our previous study provided evidence that HAB-2 was evaluated against 17 plant pathogens such as *Colletotrichum gloeosporioides*, *Alternaria solani*, and *Fusarium oxysporum*. The inhibition rates were ranged from 42.99–69.76%. Furthermore, the strain could effectively inhibit the infection of *C. gloeosporioides* on the mango leaf and fruits, *Alternaria solani* on the tomato fruits, and prevent and control *Oidium heveae* (Jin et al., 2018). Analysis of the whole genome of *B. velezensis* will allow uncovering the antibacterial mechanism of biocontrol bacteria. In this study, we assembled the whole genome of *B. velezensis* HAB-2, which not only showed a similar ability to produce antimicrobial secondary metabolites as for most *Bacillus* spp., but also demonstrated some unique abilities in our previous study (Jin et al., 2017). Furthermore, the assembled genome of HAB-2 was compared with the genomes of *B. velezensis* FZB42, *B. amyloliquefaciens* DSM7, and *B. subtilis* 168; genome annotation and comparative genomics analysis showed some important features including prophage regions, unique clusters, and key genes of these strains and also revealed that HAB-2 had great potential for biocontrol of plant pathogens.

MATERIALS AND METHODS

Strain Culture and Type Genomes

Bacillus velezensis HAB-2 was isolated from cotton (*Gossypium hirsutum*) root-soil in Xinjiang Province, China, and preserved in the Key Laboratory of Green Prevention and Control of Tropical Plant Diseases and Pests (Hainan University), Ministry of Education. The HAB-2 strain was cultivated in Luria-Bertani agar medium and broth at 28°C with shaking at 180 r/min, and was used to extract genomic DNA. Type genome sequences of strains *B. velezensis* FZB42 (Chen et al., 2007, NC_009725.1), *B. amyloliquefaciens* DSM7 (Rückert et al., 2011, NC_014551.1) and *B. subtilis* 168 (Barbe et al., 2009, NC_000964.3) were collected from the National Center for Biotechnology Information (NCBI).

Genomic DNA Extraction and Sequencing

Genomic DNA from HAB-2 was extracted using commercial kits according to the instructions (Sigma Aldrich, St. Louis, MO, United States). The DNA quality was determined using Qubit and Nanodrop (both Thermo Fisher Scientific, Waltham, MA, United States). Then the qualified genomic DNA from *B. velezensis* HAB-2 was fragmented with G-tubes (Covaris,

¹<http://www.chinapesticide.org.cn/>

Woburn, MA, United States) and end-repaired to prepare SMRTbell DNA template libraries (with fragment size > 10 kb selected by bluepippin system) according to the specification (PacBio, Menlo Park, CA, United States). Library quality was detected by Qubit, and average fragment size was estimated on a Bioanalyzer 2100 (Agilent, Santa Clara, CA, United States). The SMRT sequencing was performed on a Pacific Biosciences RSII sequencer (PacBio, Menlo Park, CA, United States) according to standard protocols (MagBead Standard Seq v2 loading, 1 × 180 min movie) using P4-C2 chemistry.

Genome Assembly

Continuous long reads were generated from three SMRT sequencing runs. Reads > 500 bp with a quality value > 0.75 were merged into a single dataset. Next, the Hierarchical Genome Assembly Process (HGAP) pipeline (Chin et al., 2013) was used to correct for random errors in the long seed reads (seed length threshold 6 kb). The resulting corrected and preassembled reads were used for *de novo* assembly using Celera Assembler with an overlap-layout-consensus strategy (Myers et al., 2000). Since SMRT sequencing features very little variation in quality throughout the reads (Koren et al., 2012), no quality values were used during the assembly. To validate the assembly quality and determine the final genome sequence, the Quiver consensus algorithm (Chin et al., 2013) was used to polish the assembly. Finally, the ends of the assembled sequence were trimmed to circularize the genome.

Genomic Features Prediction and Annotations

The open reading frame was predicted using GeneMarkS (Besemer et al., 2001). Repetitive elements were identified by RepeatMasker (Tarailo-Graovac and Chen, 2009). Non-coding RNA, such as rRNA, prediction was carried out using rRNAmmer (Lagesen et al., 2007), tRNAs were identified by tRNAscan (Lowe and Eddy, 1997) and compared with the Rfam database, and sRNAs were obtained using Infernal (Gardner et al., 2009). Several complementary approaches were utilized to annotate the assembled sequences. The genes were annotated by aligning with those deposited in diverse protein databases (*E*-value ≤ 1e-5, minimal alignment length percentage ≥ 40%) including the NCBI non-redundant protein (Nr) database, UniProt/Swiss-Prot (UniProt Consortium, 2015), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2016), Gene Ontology (GO) (Ashburner et al., 2000; Jones et al., 2014), Cluster of Orthologous Groups of proteins (COG) (Galperin et al., 2015; Makarova et al., 2015) and protein families (Pfam) (Finn et al., 2014). Additional annotation was carried out using the following databases: Pathogen Host Interactions (PHI) (Winnenburg et al., 2006), Virulence Factors of Pathogenic Bacteria (VFDB) (Chen et al., 2016), Antibiotic Resistance Genes Database (ARDB) (Liu and Pop, 2009), and Carbohydrate-Active enZymes (CAZy) (Levasseur et al., 2013). Prophage was predicted using the PHAge SearchTool (PHAST²) as described by Zhou et al. (2011). Based on Nr annotation,

²<http://phast.wishartlab.com/>

GO annotation was carried out using Blast2GO (Conesa et al., 2005), and Pfam annotation was applied with PfamScan³ (Finn et al., 2014).

Comparative Genomic Analysis

In order to ensure the same annotation conditions, the genome data of the reference strains are analyzed based on the updated data in 2020 from NCBI, and the genome features are based on NCBI annotation. Genomics comparison of HAB-2, FZB42, DSM7, and 168 was performed using MUMmer (Kurtz et al., 2004) and Mauve (version 2.3.1) (Darling et al., 2010). Gene families were constructed using genes of the reference and the target bacteria, and then the gene families were analyzed using the EDGAR software framework (Blom et al., 2009) to produce a Venn diagram. Phylogenetic analysis was performed to validate the phylogenetic relationship of HAB-2. The phylogenetic tree was constructed by the array of SNPs obtained from alignment of complete sample genomes against the reference. The phylogenetic tree drawn from the genomes of the four strains was constructed by TreeBeST (Nandi et al., 2010) using the method of PHYML (Wang D. et al., 2010) with bootstrap setting as 1,000. Based on 16S rRNA sequences, the phylogenetic tree was constructed using the maximum likelihood method in MEGA7 software (Tamura et al., 2007) with bootstrap setting as 1,000. The secondary metabolite antibacterial gene clusters of the four strains were compared and analyzed using anti-SMASH 4.1.0 (Weber et al., 2015).

RESULTS

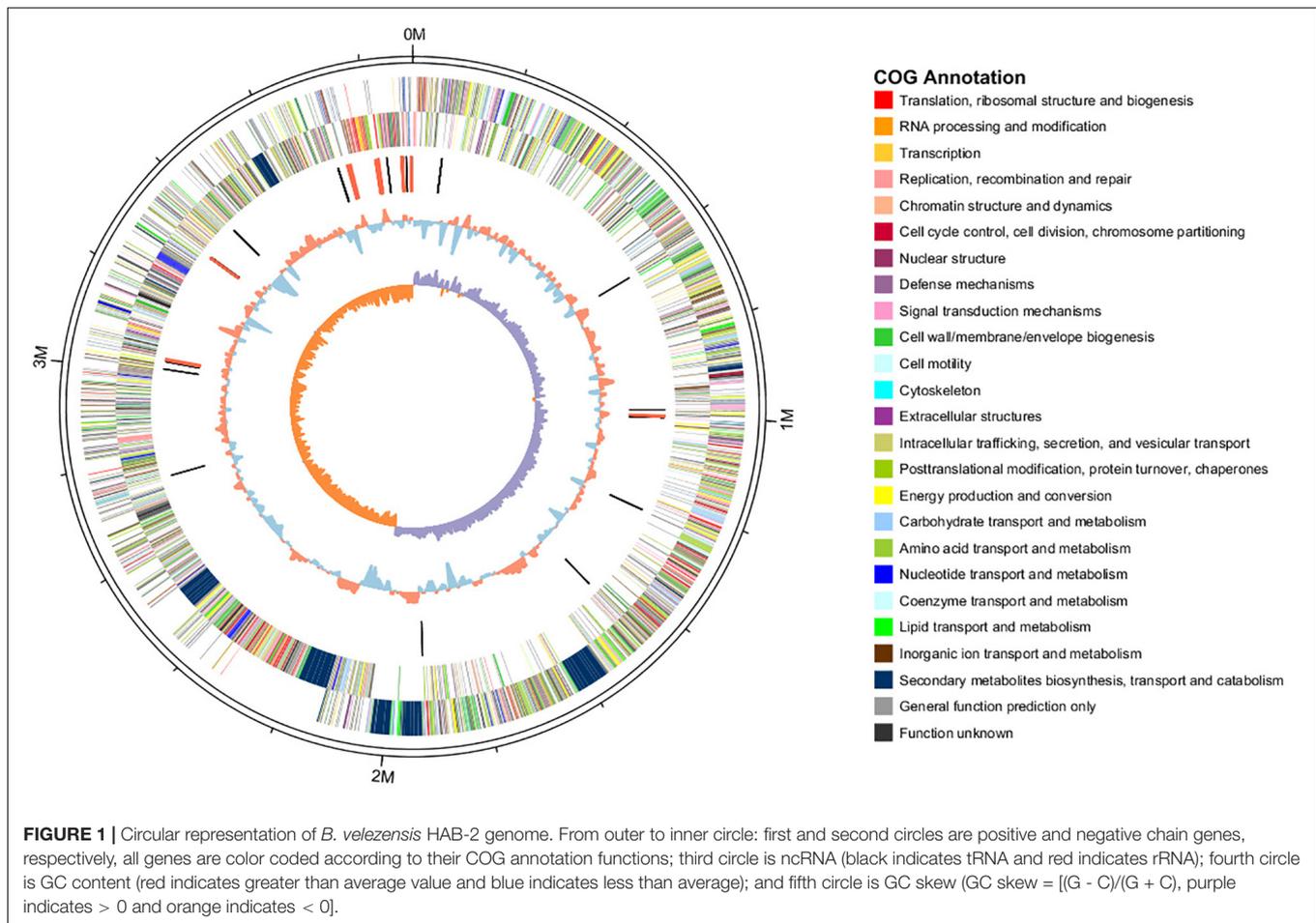
Distinctive Genomic Features in the Prophage Region of *Bacillus* Strains

The features of the genomes of *Bacillus* reference strains may differ due to the classification of the strains. Thus, we downloaded the genomes of the reference strains from the NCBI database and compared their genomic features with HAB-2. The summary features of the complete genome sequences of HAB-2, FZB42, DSM7 and 168 are summarized in **Table 1**. There were 143,574 reads with a total of 1,350,725,830 base pairs (bp) in HAB-2, and the sequencing data of HAB-2 were *de novo* assembled

³<http://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/PfamScan.tar.gz>

TABLE 1 | Genomic feature comparisons among *B. velezensis* HAB-2, *B. velezensis* FZB42, *B. amyloliquefaciens* DSM7, and *B. subtilis* 168.

	HAB-2	FZB42	DSM7	168
Genome size (bp)	3,894,648	3,918,596	3,980,199	4,215,606
G + C content (mol%)	46.6	46.4	46.1	43.5
Genes	3,820	3,855	4,102	4,536
Percent of coding region	88.8	88.0	86.7	87.2
Ribosomal RNA operons	9	10	10	10
Number of tRNAs	86	89	94	86
Phage-associated genes	58	25	273	324



using HGAP. Genome circular representation can fully display the characteristics of the genome. We obtained a genome size of approximately 3,894,648 bp, consisting of a circular chromosome and leading to the one scaffold with N50 length of 3.89 Mb and 46.64% GC content (**Figure 1**). The genome of HAB-2 was smaller than of 168 (4,215,606 bp) and comparable to those of FZB42 (3,918,596 bp) and DSM7 (3,980,199 bp). A total of 3,820 genes were predicted, in which the maximum gene size was 16,302 bp and the minimum was 114 bp. Furthermore, 86 tRNAs and nine sets of 16s-23s-5s ribosomal RNA operons were predicted in the genome (**Table 1**). The numbers of genes representing the core genome of strains HAB-2, FZB42, DSM7 and 168 were 3820, 3855, 4102, and 4536, respectively (**Table 1**). The majority of the genes unique to DSM7 and 168 but not present in HAB-2 and FZB42 were phage related (**Figure 2**). Compared with FZB42, there was one more intact prophage region [Pr 2, with 46 coding sequences (CDS)] in HAB-2 (**Figure 2**).

Phylogenetic Analysis of *B. velezensis* HAB-2

The classification of *Bacillus* strains has been constantly revised based on the development of molecular technology

and the popularity of whole-genome sequencing, especially for *B. velezensis*. A phylogram based on *Bacillus* 16S rRNA, the most useful and is commonly used marker in the systematic classification of bacterium (Coenye and Vandamme, 2010) suggested a close relationship between the genomes of *B. velezensis* HAB-2 and FZB42 (**Figure 3A**). To further clarify the relationships of HAB-2 with the other relative species, the whole genomic SNPs was used to determine the evolutionary positions of the strains. The phylogenetic tree for HAB-2 and the reference strains was based on genomic scale SNPs (**Figure 3B**) and showed that HAB-2 was most closely related to *B. velezensis* FZB42. Overall, evolutionary analysis showed that HAB-2 was in the same branch with FZB42.

Analysis of the Core Genome Among *Bacillus* Species

Research on core genomes is important for determining the functional differences and similarities between strains, and provides molecular evidence for phenotypic differences and similarities. Core-genome plot analyses were performed for the four *Bacillus* genomes (**Figure 4**). There were 2921 gene families in the final core genome. Among them, 357 (364 CDS), 135

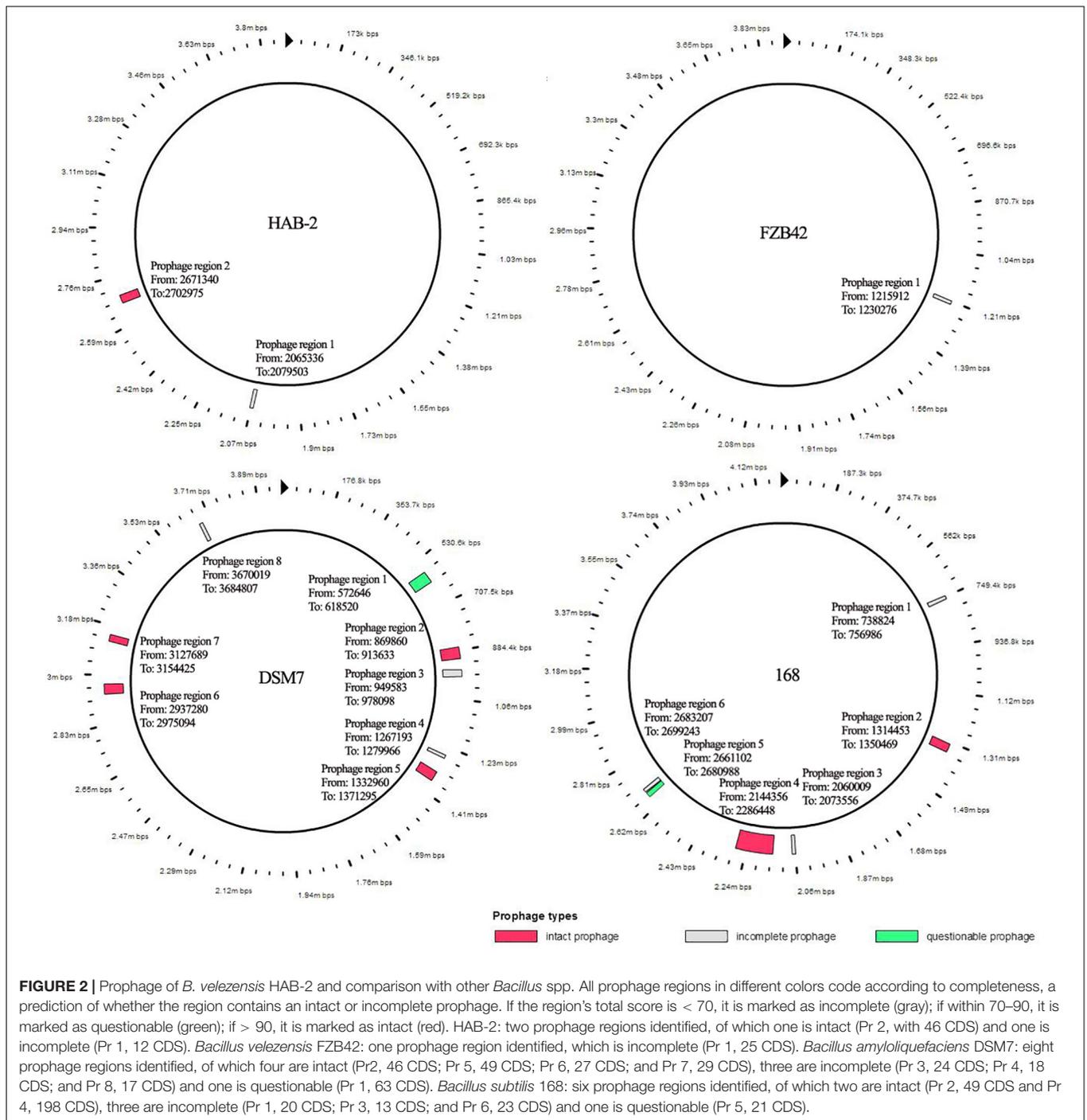


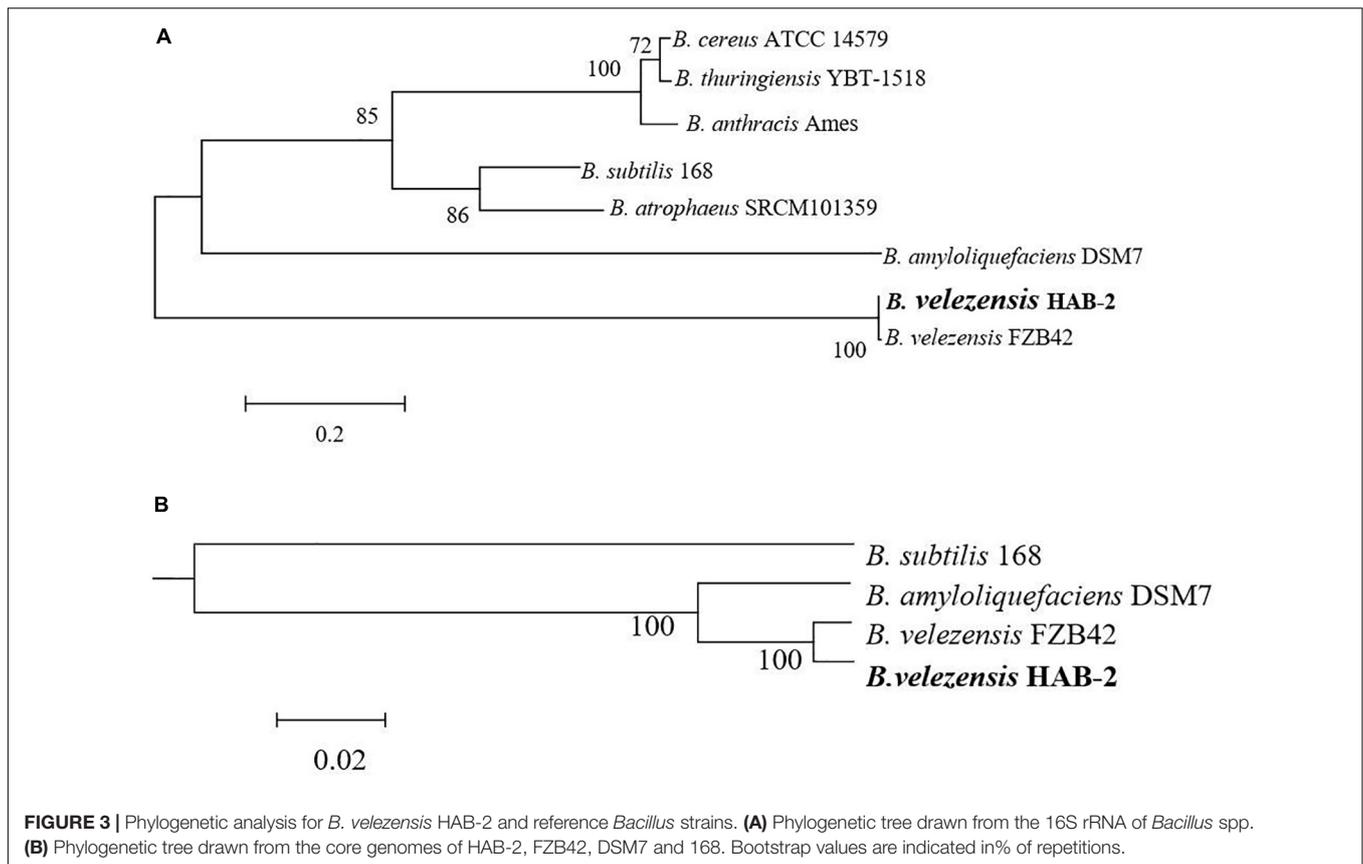
FIGURE 2 | Prophage of *B. velezensis* HAB-2 and comparison with other *Bacillus* spp. All prophage regions in different colors code according to completeness, a prediction of whether the region contains an intact or incomplete prophage. If the region's total score is < 70, it is marked as incomplete (gray); if within 70–90, it is marked as questionable (green); if > 90, it is marked as intact (red). HAB-2: two prophage regions identified, of which one is intact (Pr 2, with 46 CDS) and one is incomplete (Pr 1, 12 CDS). *Bacillus velezensis* FZB42: one prophage region identified, which is incomplete (Pr 1, 25 CDS). *Bacillus amyloliquefaciens* DSM7: eight prophage regions identified, of which four are intact (Pr 2, 46 CDS; Pr 5, 49 CDS; Pr 6, 27 CDS; and Pr 7, 29 CDS), three are incomplete (Pr 3, 24 CDS; Pr 4, 18 CDS; and Pr 8, 17 CDS) and one is questionable (Pr 1, 63 CDS). *Bacillus subtilis* 168: six prophage regions identified, of which two are intact (Pr 2, 49 CDS and Pr 4, 198 CDS), three are incomplete (Pr 1, 20 CDS; Pr 3, 13 CDS; and Pr 6, 23 CDS) and one is questionable (Pr 5, 21 CDS).

(135 CDS), 396 (427 CDS) and 739 gene families (782 CDS) in HAB-2, FZB42, DSM7, and 168 strains were singletons, respectively. Comparison of homology of the gene families showed that HAB-2 and FZB42 had 3,360 gene families in common with an average 87.22% identity; HAB-2 and DSM7 had 3,230 gene families in common with average 86.84% identity; and HAB-2 and 168 had 3,153 gene families in common with average 81.85% identity. A Venn diagram based on computing of the *Bacillus* gene families suggested high genetic homogeneity among

the four *Bacillus* strains and close homology among *B. velezensis* and *B. amyloliquefaciens* genomes.

Collinearity Analysis of *B. velezensis* HAB-2

We performed a collinearity analysis to further compare the genomics similarities and differences among the genome of HAB-2 and the genomes of FZB42, DSM7 and 168. The results



showed that the HAB-2 genome displayed different synteny to those of the other strains (Figure 5). The collinearity of the genome of HAB-2 with those of FZB42, DSM7 and 168 represented approximately 96.14, 89.15, and 22.46%, respectively, of the total length of HAB-2. Of all the sequenced genomes, HAB-2 showed the highest synteny with *B. velezensis* FZB42, indicating their evolutionary stages were the closest, and their genomes were more related.

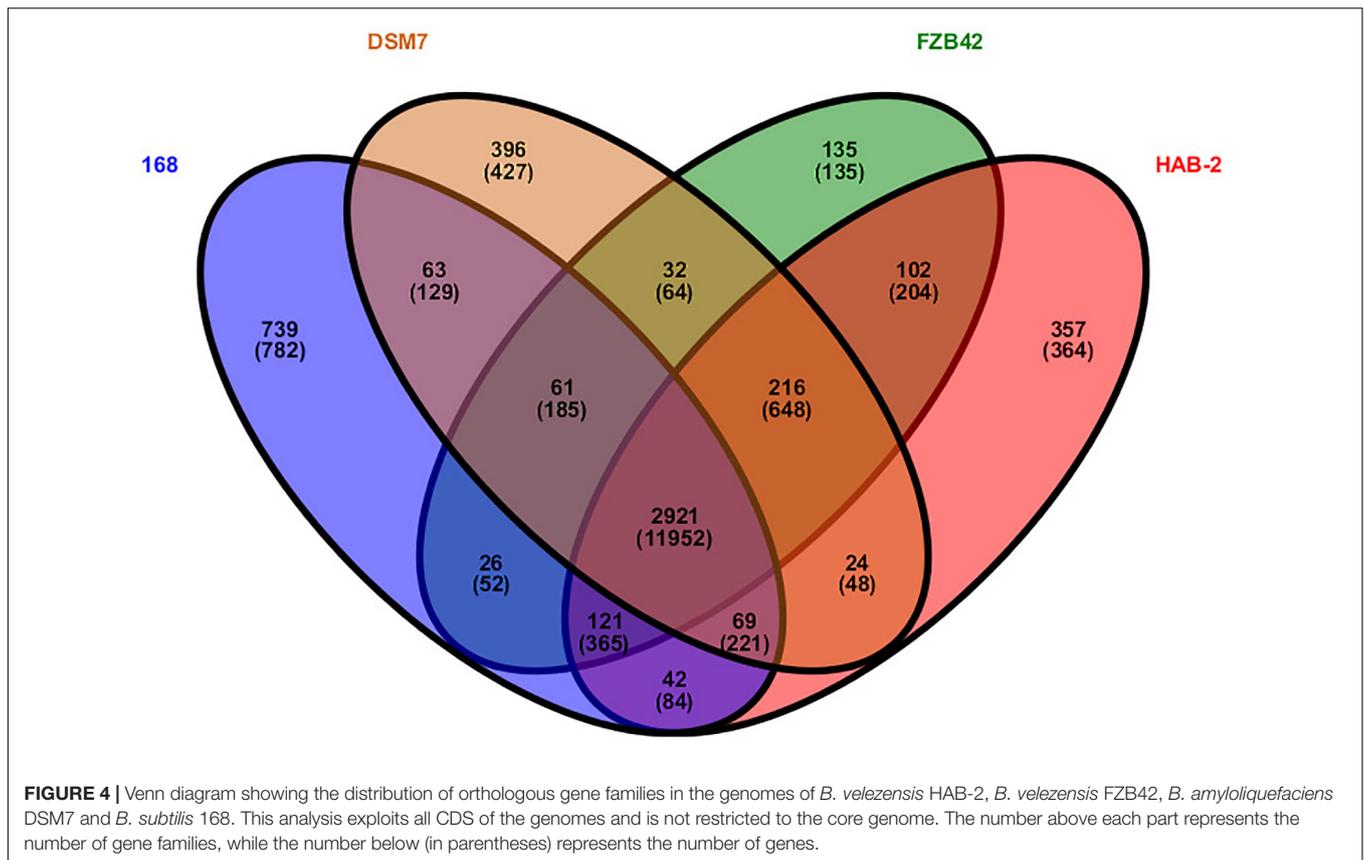
Gene Clusters Analysis of *B. velezensis* HAB-2

The genomes of *Bacillus* harbor some important gene clusters involved in the synthesis of secondary metabolites, with antimicrobial, antiviral and nematocidal action. Through use of anti-SMASH genome analysis, 13 clusters of secondary metabolites were identified in the HAB-2 genome (Table 2 and Figure 6), covering 20.08% (782.23 kb) of the whole genome and involved in the synthesis of antifungal and antibacterial acting secondary metabolites. This result revealed the potential for biocontrol of plant pathogens in HAB-2. Five gene clusters with a total length of 164.04 kb were shown to direct unknown compound. There were eight gene clusters related to encoding NRPS and PKS, including Type III PKS cluster and PKS-like cluster directing two unknown compounds. Eight clusters were identified as involved in mersacidin synthesis (GenBank AJ250862.2), bacilysin (GenBank CP000560.1), bacillibactin (GenBank AL009126.3), diffidin

(GenBank AJ634062.2), fengycin (GenBank CP000560.1), bacillaene (GenBank AJ634060.2), macrolactin (GenBank AJ634061.2) and surfactin (GenBank AJ575642.1) (Table 2 and Figure 6). Secondary metabolites clusters were compared to strains FZB42, DSM7 and 168 (Figure 7). Among the 13 gene clusters, two (Clusters 1 and 12) were specific and only existed in HAB-2, two (Clusters 4 and 9) existed in both HAB-2 and FZB42, one (Cluster 11) existed in three strains excluding 168, and the other eight (Clusters 2, 3, 5–8, 10, and 13) existed in all four strains. Strain HAB-2 had two clusters that were not present in any of the other three strains (Table 2): one (Cluster 1) encoding lanthipeptide was involved in mersacidin synthesis and one (Cluster 12) encoding ladderane was shown to direct an unknown compound. In the same way, we found five clusters (Clusters 5, 6, and 10–12) encoding potential novel metabolites with no previously reported description, and four of these were similar to gene clusters in FZB42.

Biosynthesis and Regulation Genes of Secondary Metabolites

Although several gene clusters were shared by HAB-2, FZB42, DSM7, and 168, the gene structures differed (Figure 8). The fengycin cluster lacked three core genes in DSM7: *fenB*, *fenC*, and *fenD*. In the surfactin cluster, HAB-2 lacked *ycxE*, DSM7 lacked *ycxC* and *ycxD*, and 168 lacked *ycx01* (Table 3). Furthermore, the core biosynthetic genes were similar among the four strains.



There was a key gene *sfp* involved in non-ribosomal synthesis, encoding PPTases. A previous study of our research group showed that a *sfp* homolog *lpaH2*, which encoded PPTase, was present in HAB-2 (Jin et al., 2017), was 675 bp in size and encoded 224 amino acids. The Nr annotation information allowed us to obtain the GO function annotation. The GO annotation resulted in three classifications: describing the molecular function (MF), cellular component and biological process (BP) of the gene. Further analysis of the GO annotated genes revealed that *lpaH2* appeared only in MF and BP; and *lpaH2* appeared five times in MF and nine times in BP. Bioinformatics analysis of the GO node was performed on these five and nine branches respectively and showed that the HAB-2_2473 gene was consistent with *lpaH2* function, was 366 bp in size and encoded 120 amino acids. The NCBI alignment showed that HAB-2_2473 had 100% homology with *acpS* and encoded PPTase, a member of the PPTase superfamily. The multiple gene functional bioinformatic methods and resources (including Nr, Swiss-Prot, KOG, KEGG, and Pfam) used to analyze the functions of *lpaH2* and HAB-2_3473 (*acpS*) showed that their functions were highly consistent (Table 4).

DISCUSSION

This study is surefire to identify and classify *B. velezensis* strain HAB-2 based on genomic information using third-generation

sequencing technology. The mechanism of biocontrol of HAB-2 was elucidated by analyzing the gene clusters of secondary metabolites. *Bacillus* species, including *B. velezensis*, *B. amyloliquefaciens*, and *B. subtilis*, have been widely researched because of their potential to produce bacteriostatic secondary metabolites (Fan et al., 2018). *Bacillus velezensis*, isolated from the Vélez River in Málaga in southern Spain, was described in 2005 (Ruiz-García et al., 2005), and *B. velezensis* was later classified as a heterotypic synonym of *B. amyloliquefaciens* (Wang et al., 2008). The model *B. amyloliquefaciens* strain FZB42 was later recognized as synonymous with *B. velezensis* using phylogenomic analysis (Dunlap et al., 2016). The classification of *Bacillus* spp. has become clearer due to the availability of an increasing number of genome sequences and the former *B. amyloliquefaciens* was reclassified and divided into *B. amyloliquefaciens*, *B. siamensis*, and *B. velezensis* (Fan et al., 2017). In this study, we compared genome sequences of reference strains *B. velezensis* FZB42, *B. amyloliquefaciens* DSM7, and *B. subtilis* 168, and conducted phylogenomic analysis; the result showed that HAB-2 was *B. velezensis*, which was previously classified as *B. amyloliquefaciens*.

In this study, we reported a high-quality genome sequence of HAB-2 using third-generation sequencing technology, and made a comparative genome sequence analysis with strains FZB42, DSM7 and 168. The 3.89 Mb genome of HAB-2 encoded 3,820 predicted genes. The four strains were basically identical in regard to whole genome size and the number of genes encoded. The

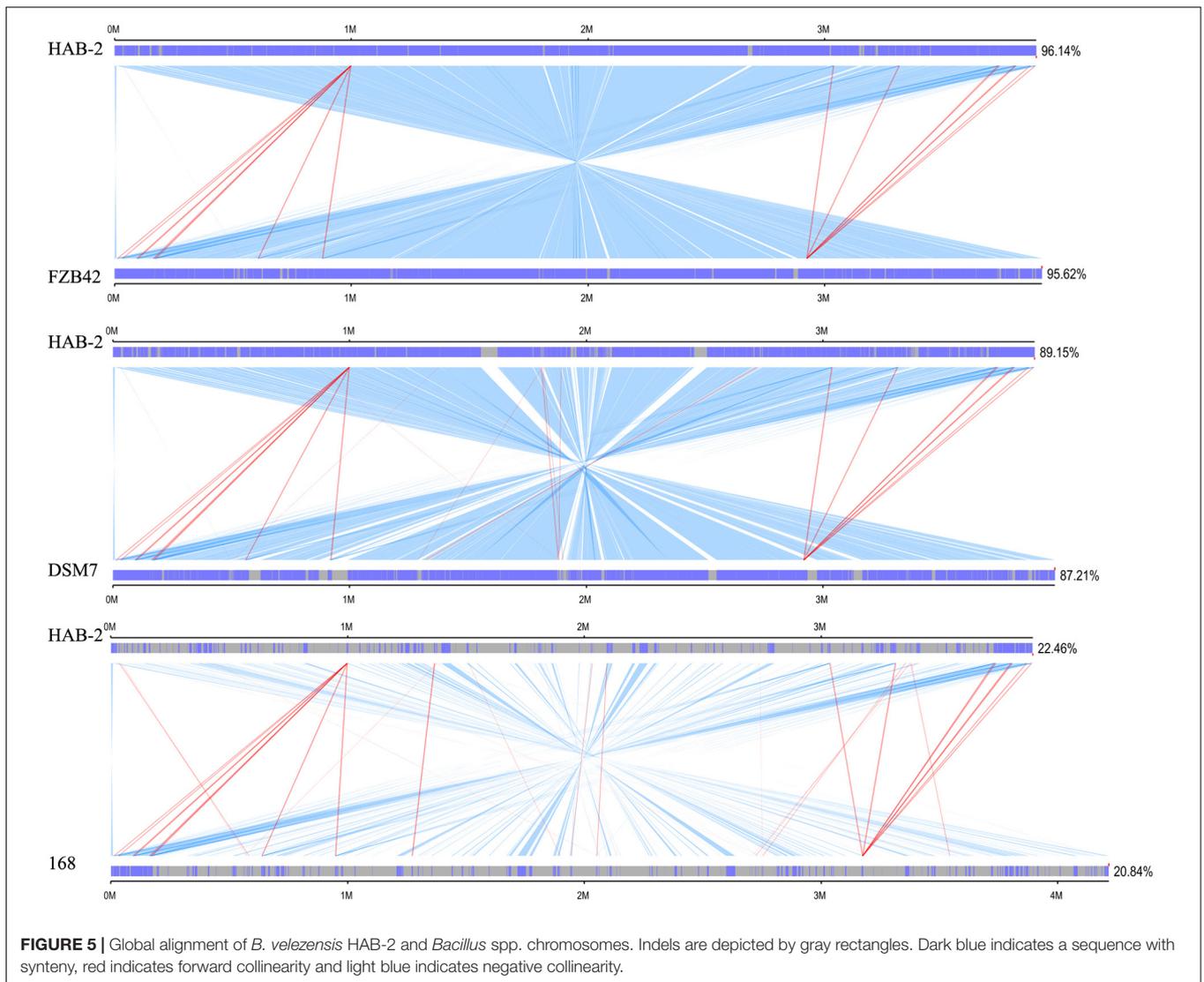


TABLE 2 | Comparative analysis of secondary metabolite clusters of *B. velezensis* HAB-2 with *B. velezensis* FZB42, *B. amyloliquefaciens* DSM7, and *B. subtilis* 168,

Cluster number	Compound	HAB-2				Size(kb)	Presence (+) or absence (-)		
		From	To	Type	FZB42		DSM7	168	
1	Mersacidin	140,027	163,215	Lanthipeptide	23.18	-	-	-	
2	Bacilysin	314,409	355,827	Other	41.42	+	+	+	
3	Bacillibactin	870,296	920,797	Bacteriocin-nrps	50.50	+	+	+	
4	Difficidin	1,536,021	1,641,780	Transatpks	105.76	+	-	-	
5	Unknown	1,757,688	1,798,401	t3pks	40.71	+	+	+	
6	Unknown	1,864,153	1,884,379	Terpene	20.23	+	+	+	
7	Fengycin	1,913,038	2,047,159	Transatpks-nrps	134.12	+	+	+	
8	Bacillaene	2,109,669	2,219,248	Transatpks-nrps	109.58	+	+	+	
9	Macrolactin	2,440,918	2,529,136	Transatpks	88.22	+	-	-	
10	Unknown	2,825,278	2,846,018	TERPENE	20.74	+	+	+	
11	Unknown	2,928,806	2,970,050	pks-like	41.24	+	+	-	
12	Unknown	3,201,340	3,242,461	Ladderane	41.12	-	-	-	
13	Surfactin	3,519,488	3,584,895	nrps	65.41	+	+	+	

Secondary metabolite antibacterial gene clusters of the four strains were compared and analyzed using anti-SMASH 4.1.0 (Weber et al., 2015).

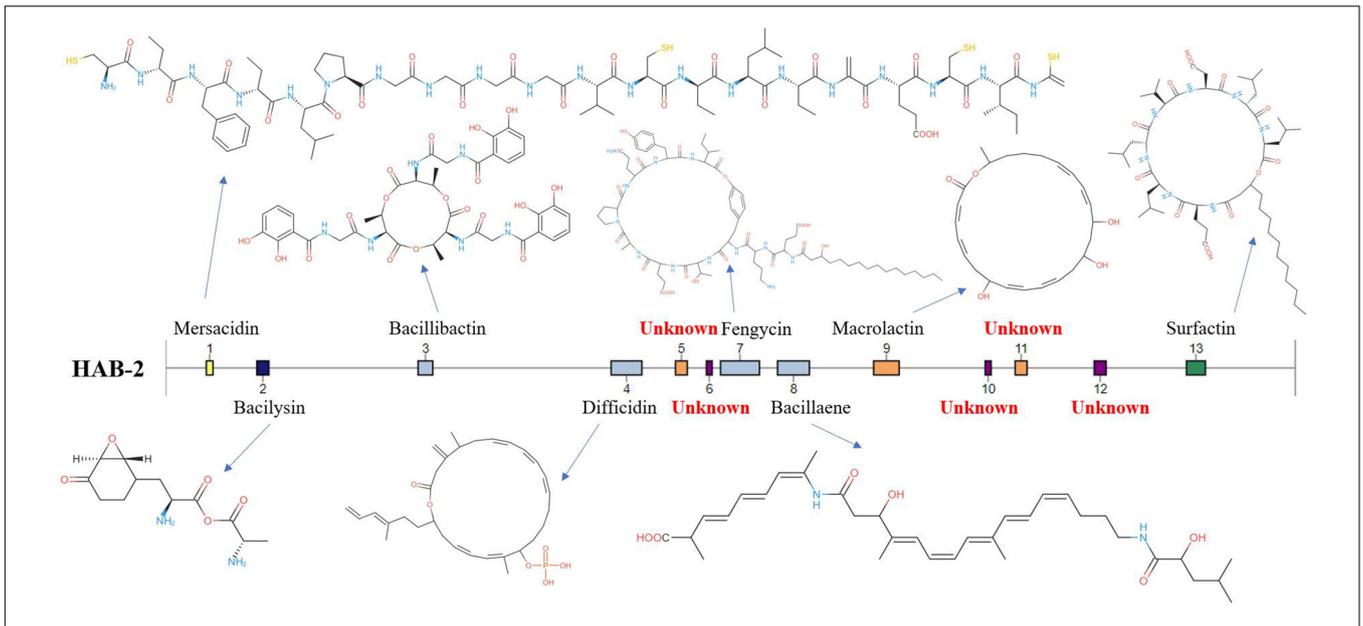


FIGURE 6 | Gene clusters and encoded secondary metabolites identified using the *B. velezensis* HAB-2 genome.

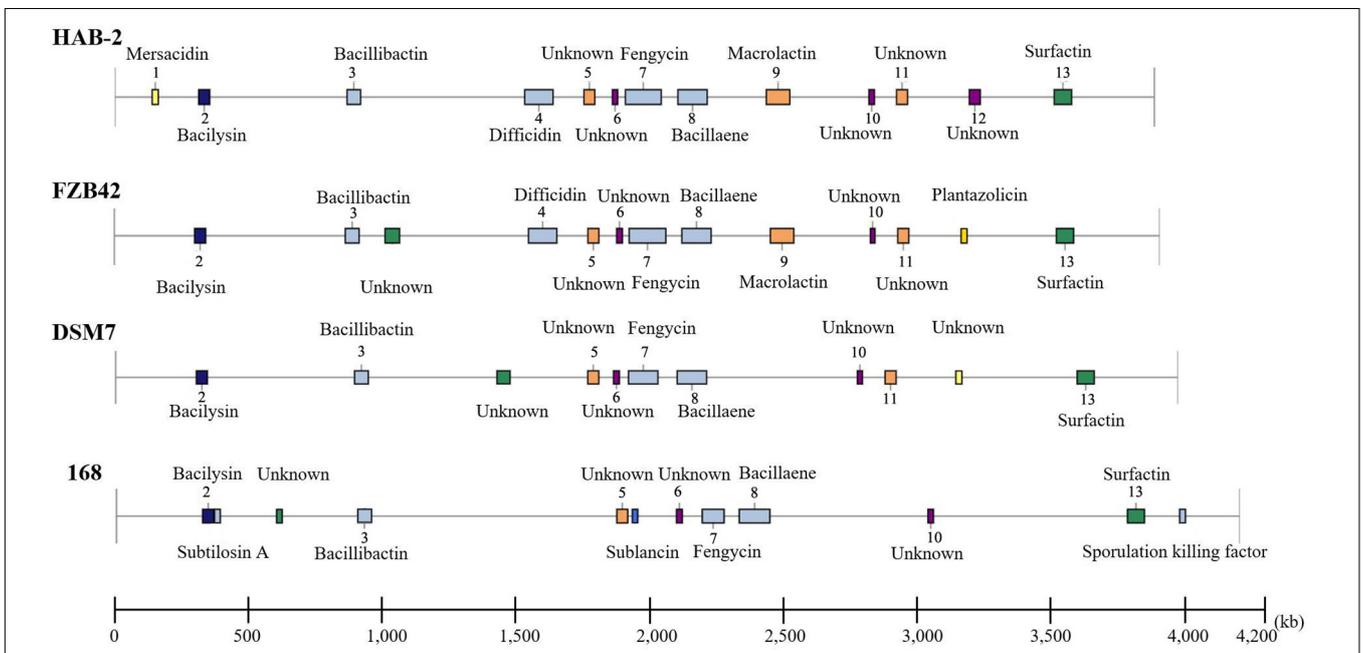


FIGURE 7 | Comparative analysis of the location and products of secondary metabolite gene clusters in *B. velezensis* HAB-2, *B. velezensis* FZB42, *B. amyloliquefaciens* DSM7 and *B. subtilis* 168.

16S rRNA and genome phylogenetic trees showed high similarity with each other and suggested that HAB-2 was most closely related to FZB42 and DSM7. The differences in the genomes of the four strains were the number and size of prophage regions, and the gene clusters encoding secondary metabolites.

Prophages can constitute 3.1–20% of a bacterium’s genome and are major sources of diversification of bacterial genomes; many prophages appear to be defective and are in a state

of mutational decay (Casjens, 2005; Touchon et al., 2016). In our study, the prophage regions of the four strains differed and each strain had a defective prophage. This result also confirmed previous studies. Prophages, including defective ones, can contribute important biological properties to their bacterial hosts, such as expressing virulence genes, to sporulate or to differentiate, and serve as active regulatory switches in bacteria by regulating bacterial genes via genome excision (Casjens, 2005;

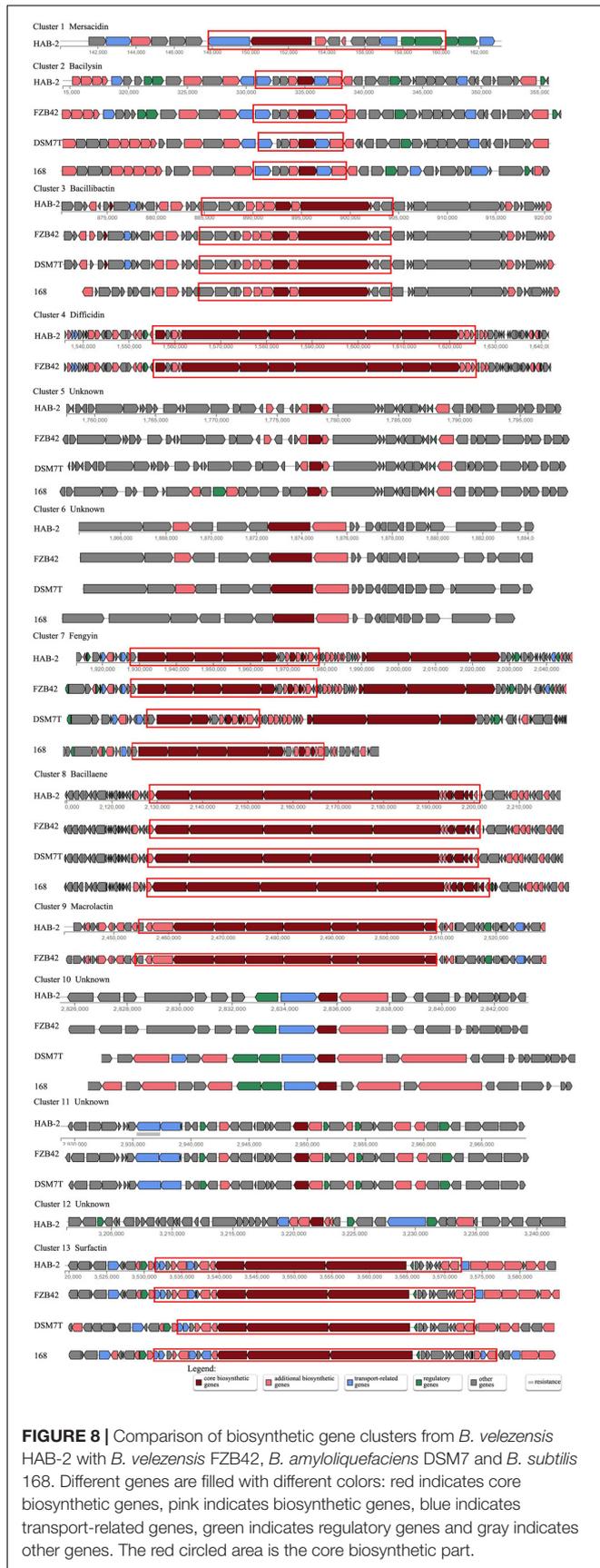


FIGURE 8 | Comparison of biosynthetic gene clusters from *B. velezensis* HAB-2 with *B. velezensis* FZB42, *B. amyloliquefaciens* DSM7 and *B. subtilis* 168. Different genes are filled with different colors: red indicates core biosynthetic genes, pink indicates biosynthetic genes, blue indicates transport-related genes, green indicates regulatory genes and gray indicates other genes. The red circled area is the core biosynthetic part.

Feiner et al., 2015). Prophages may indirectly affect the evolvability of bacteria (Touchon et al., 2016). Wang X. et al. (2010) showed that prophages helped the cell respond to various stresses, and inhibiting key proteins encoded by prophages might be a novel means to combat antibiotic resistance. *Pseudomonas aeruginosa* produces three types of bacteriocins, and genetically related phages exist for each type – the gene organization of two types of gene clusters suggested that phage tails have been evolutionarily specialized as bacteriocins (Nakayama et al., 2000). Furthermore, prophages have non-homologous modules that can assemble gene clusters (Casjens, 2005; Gilcrease et al., 2005). Thus, the prophage in *Bacillus* strains may affect the evolutionary, classification and the production of metabolites of the strains.

Based on the *B. velezensis* genomes analysis, more and more secondary metabolite gene clusters have been discovered. *B. velezensis* K26 harbors seven gene clusters and it contains a 1-Deoxynojirumycin biosynthetic gene cluster which is a representative iminosugar with alpha-glucosidase inhibition activity (Lee et al., 2020). *B. velezensis* B-4 harbors 12 gene clusters relating to synthesis of antimicrobial metabolites (Zhu et al., 2020). The HAB-2 genome harbors 13 gene clusters involved in synthesis of antifungal and antibacterial acting secondary metabolites. The number of gene clusters is larger than in other type strains. Notably, HAB-2 has a complete gene cluster, and FZB42 has an incomplete gene cluster, directing mersacidin. Mersacidin is a lantionine-containing antimicrobial peptide belonging to the family of lantibiotics, and has been identified as a type-B lantibiotic (Bierbaum et al., 1995). It was found in *Bacillus* sp. Strain HIL Y-85,54728 (Chatterjee et al., 1992). People found that mersacidin could act against methicillin-resistant *Staphylococcus aureus* (Brötz et al., 1997, 1998) and was regarded as a lead member of the type-B lantibiotics. Altena et al. (2000) explored the mersacidin biosynthetic gene cluster and reported the complete sequence of a gene cluster of a type-B lantibiotic. Herzner et al. (2011) transferred genomic DNA of a mersacidin-producing strain to FZB42, which resulted in successful production of mersacidin by FZB42.

The GO function annotation showed that gene function was mainly concentrated in MF and BP. Analysis of gene function revealed genes relevant to cellular process, metabolic process, single-organism process, binding and catalytic activity. As previously mentioned, multidomain enzymes (e.g., PKs and NRPSs) control the biosynthesis of secondary metabolites (Marahiel et al., 1997; Von Dohren et al., 1997; Metz et al., 2001; Cox, 2007). A same common characteristic of these enzymes is that they require post-translational modification to become active. The apoprotein form (inactive) becomes active (holo form) after the covalent attachment of the 4'-phosphopantetheine moiety, which derives from coenzyme A (Lambalot et al., 1996; García-Estrada et al., 2008). This means that this reaction is catalyzed by PPTase. There is a key gene *spf* encoding the PPTase. Borchert et al. (1994) identified that the *gsp* gene complements a *spf* deficiency in a *B. subtilis* mutant. Huang et al. (1993) and Tsuge et al. (1996) separately discovered *lpa-14* and *lpa-8*, which regulate production of iturin A, surfactin and plipastatin. Grover et al. (2010) found that *B. subtilis* strain RP24, which possesses antifungal activity, can produce iturin A, surfactin and

TABLE 3 | Core genes involved with fengycin and surfactin in *B. velezensis* HAB-2, *B. velezensis* FZB42, *B. amyloliquefaciens* DSM7, and *B. subtilis* 168.

Compound	Gene	Presence (+) or absence (-)				Functions or product
		HAB-2	FZB42	DSM7	168	
Fengycin	<i>dacC</i>	+	+	+	+	DacC protein
	<i>fenA</i>	+	+	+	+	Fengycin synthetase A, scaffold biosynthesis
	<i>fenB</i>	+	+	-	+	Fengycin synthetase B, scaffold biosynthesis
	<i>fenC</i>	+	+	-	+	Fengycin synthetase C, scaffold biosynthesis
	<i>fenD</i>	+	+	-	+	Fengycin synthetase D, scaffold biosynthesis
	<i>fenE</i>	+	+	+	+	Fengycin synthetase E, scaffold biosynthesis
Surfactin	<i>yxO1</i>	+	+	+	-	YxO1 protein, alcohol dehydrogenase
	<i>yckE</i>	-	+	+	+	YckE protein
	<i>srfAA</i>	+	+	+	+	Surfactin synthetase A, scaffold biosynthesis
	<i>srfAB</i>	+	+	+	+	Surfactin synthetase B, scaffold biosynthesis
	<i>srfAC</i>	+	+	+	+	Surfactin synthetase C, scaffold biosynthesis
	<i>srfAD</i>	+	+	+	+	Surfactin synthetase D, other enzymatic
	<i>aat</i>	+	+	+	+	Amino transferase
	<i>ycxC</i>	+	+	-	+	Transporter
	<i>ycxD</i>	+	+	-	+	Transcriptional regulator containing an aminotransferase domain
	<i>sfp</i>	+	+	+	+	Phosphopantetheinyl transferase involved in non-ribosomal synthesis
	<i>yczE</i>	+	+	+	+	Integral membrane protein involved in non-ribosomal synthesis
	<i>yckI</i>	+	+	+	+	YckI protein
	<i>yckJ</i>	+	+	+	+	YckJ protein

Secondary metabolite antibacterial gene clusters of the four strains were compared using anti-SMASH 4.1.0 (Weber et al., 2015), reference gene clusters and core genes analyzed with MIBIG (Minimum Information about a Biosynthetic Gene cluster) (Kautsar et al., 2020).

TABLE 4 | Comparison of gene function bioinformatics analysis between *lpaH2* and *acpS*.

	<i>lpaH2</i>	<i>acpS</i>
Size(bp)	675	366
Nr annotation	4'-Phosphopantetheinyl transferase	4'-Phosphopantetheinyl transferase
Swiss-Prot annotation	4'-Phosphopantetheinyl transferase	Holo-[acyl-carrier-protein] synthase
KOG function description	Phosphopantetheinyl transferase	Phosphopantetheinyl transferase (holo-ACP synthase)
KEGG pathway	Pantothenate and CoA biosynthesis/etabolism of cofactors and vitamins/Metabolism	Pantothenate and CoA Biosynthesis/Metabolism of cofactors and vitamins/Metabolism
PfamA definition	4'-phosphopantetheinyl transferase superfamily	4'-Phosphopantetheinyl Transferase superfamily

fengycin and also contains a homolog to *lpa-14* found to encode a PPTase. Morikawa et al. (1992) identified the *B. pumilus* A-1 gene *psf-1*, which regulates surfactin production, and showed that it encoded an active PPTase. Previous study by our research group detected a *sfp* homolog *lpaH2*, which encoded PPTase, in HAB-2 (Jin et al., 2017). The relevant genes encoding PPTases can be screened using a cluster analysis of gene function. In this study, bioinformatics analysis revealed that key gene *acpS*, which encoded 120 amino acids, encoded the PPTases in HAB-2. This provides significant guidance for subsequent analysis of the antibacterial mechanism of *Bacillus* spp.

CONCLUSION

We showed the genetic relationship of *B. velezensis* HAB-2 with other *Bacillus* spp. strains through comparative genomic analysis, which is instructive in guiding us to reveal more characteristics of this strain. The differences in prophage regions may provide a new strategy for taxonomic classification and secondary metabolite production of *Bacillus*. The fact that abundant secondary metabolite gene clusters and unique genes exist in the *B. velezensis* HAB-2 genome illustrates that this strain has potential as a biocontrol bacterium for plant protection in agriculture.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/genbank/>, MT375545; <https://www.ncbi.nlm.nih.gov/genbank/>, MT386600; <https://www.ncbi.nlm.nih.gov/>, CP060085.

AUTHOR CONTRIBUTIONS

WM and PX designed the experiments and wrote the manuscript. PX, SX, WL, and PJ performed collection and bioinformatics analysis. SX, DW, DY, and YW revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Key Research and Development Program of Hainan Province (No.ZDYF2018240), the National Key R&D Program of China (No. 2018YFD0201105), the National Natural Science Foundation of China (Grant Nos. 31560495, 31760499 and 31660033), the National Natural Rubber Industry Technical System (No. CARS-34-GW8) and

Hainan Major Research Fund of Science and Technology (No. ZDKJ201817).

ACKNOWLEDGMENTS

We gratefully acknowledge Prof. Xiangfeng Wang and Dr. Jianan Wang for useful discussions about this article.

REFERENCES

- Altena, K., Guder, A., Cramer, C., and Bierbaum, G. (2000). Biosynthesis of the lantibiotic mersacidin: organization of a type b lantibiotic gene cluster. *Appl. Environ. Microbiol.* 66, 2565–2571. doi: 10.1128/aem.66.6.2565-2571.2000
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., et al. (2009). From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 155, 1758–1775. doi: 10.1099/mic.0.027839-0
- Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P., and Burkart, M. D. (2014). The phosphopantetheinyl transferases: catalysis of a posttranslational modification crucial for life. *Nat. Prod. Rep.* 31, 61–108. doi: 10.1039/c3np70054b
- Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618. doi: 10.1093/nar/29.12.2607
- Bierbaum, G., Brotz, H., Koller, K. P., and Sahl, H. G. (1995). Cloning, sequencing and production of the lantibiotic mersacidin. *FEMS Microbiol. Lett.* 127, 121–126. doi: 10.1016/0378-1097(95)00048-A
- Blom, J., Albaum, S., Doppmeier, D., Pühler, A., Vorhölder, F., Zakrzewski, M., et al. (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* 10:154. doi: 10.1186/1471-2105-10-154
- Borchert, S., Stachelhaus, T., and Marahiel, M. A. (1994). Induction of surfactin production in *Bacillus subtilis* by *gsp*, a gene located upstream of the gramicidin S operon in *Bacillus brevis*. *J. Bacteriol.* 176, 2458–2462. doi: 10.1128/jb.176.8.2458-2462.1994
- Brötz, H., Bierbaum, G., Leopold, K., Reynolds, P. E., and Sahl, H. G. (1998). The lantibiotic mersacidin inhibits peptidoglycan synthesis by targeting lipid II. *Antimicrob. Agents Chemother.* 42, 154–160. doi: 10.1128/aac.42.1.154
- Brötz, H., Bierbaum, G., Reynolds, P. E., and Sahl, H. G. (1997). The lantibiotic mersacidin inhibits peptidoglycan biosynthesis at the level of trans glycosylation. *Eur. J. Biochem.* 246, 193–199. doi: 10.1111/j.1432-1033.1997.t01-1-00193.X
- Casjens, S. R. (2005). Comparative genomics, and evolution of the tailed-bacteriophages. *Curr. Opin. Microbiol.* 8, 451–458. doi: 10.1046/j.1365-2958.2003.03580.X
- Chatterjee, S., Chatterjee, S., Lad, S. J., Phansalkar, M. S., Rupp, R. H., Ganguli, B. N., et al. (1992). Mersacidin, a new antibiotic from *Bacillus*: fermentation, isolation, purification and chemical characterization. *J. Antibiot.* 45, 832–838. doi: 10.7164/antibiotics.45.832
- Chen, L. H., Zheng, D. D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239
- Chen, X. H., Koumoutsis, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., et al. (2007). Comparative analysis of the complete genome sequence of the plantgrowth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nat. Biotechnol.* 25, 1007–1014. doi: 10.1002/9781118297674.ch83
- Chen, X. H., Koumoutsis, A., Scholz, R., Schneider, K., Vater, J., Süßmuth, R., et al. (2008). Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. *J. Biotechnol.* 140, 27–37. doi: 10.1016/j.jbiotec.2008.10.011
- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Coenye, T., and Vandamme, P. (2010). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol. Lett.* 228, 45–49. doi: 10.1016/S0378-1097(03)00717-1
- Compant, S., Duffy, B., Nowak, J., Clement, C., and Barka, E. A. (2005). Use of plant growth-promoting bacteria for biocontrol of plant diseases: principles, mechanism of action, and future prospects. *Appl. Environ. Microbiol.* 71, 4951–4959. doi: 10.1128/AEM.71.9.4951-4959.2005
- Conesa, A., Gotz, S., Garcia-Gomez, J. M., Talón, J., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Cox, R. J. (2007). Polyketides. Proteins and genes in fungi: programmed nanomachines begin to reveal their secrets. *Cheminform* 5, 2010–2026. doi: 10.1039/b704420h
- Darling, A. E., Mau, B., and Perna, N. T. (2010). progressive Mauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147
- Dunlap, C., Kim, S. J., Kwon, S. W., and Rooney, A. P. (2016). *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*, *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and '*Bacillus oryzicola*' are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. *Int. J. Syst. Evol. Microbiol.* 66, 1212–1217. doi: 10.1099/ijsem.0.000858
- Fan, B., Blom, J., Klenk, H. P., and Rainer, B. (2017). *Bacillus amyloliquefaciens*, *Bacillus velezensis*, and *Bacillus siamensis* form an “operational group *B. amyloliquefaciens*” within the *B. subtilis* species complex. *Front. Microbiol.* 8:22. doi: 10.3389/fmicb.2017.00022
- Fan, B., Wang, C., Song, X., Ding, X. L., Wu, L. M., Wu, H. J., et al. (2018). *Bacillus velezensis* FZB42 in 2018: the Gram-positive model strain for plant growth promotion and biocontrol. *Front. Microbiol.* 9:2491. doi: 10.3389/fmicb.2018.02491
- Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A. A. (2015). A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat. Rev. Microbiol.* 13, 641–650. doi: 10.1038/nrmicro3527
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Eugene, V. K. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. doi: 10.1093/nar/gku1223
- García-Estrada, C., Ricardo, V. U., Velasco-Conde, T., Ramiro, P. G., Fernando, T., Inmaculada, V., et al. (2008). Post-translational enzyme modification by the phosphopantetheinyl transferase is required for lysine and penicillin biosynthesis but not for roquefortine or fatty acid formation in *Penicillium chrysogenum*. *Biochem. J.* 415, 317–324. doi: 10.1042/BJ20080369
- Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., et al. (2009). Rfam: updates to the RNA families database. *Nucleic Acids Res.* 37, D136–D140. doi: 10.1093/nar/gkn766
- Gilcrease, E. B., Winn-Stapley, D. A., Hewitt, F. C., Joss, L., and Casjens, S. R. (2005). Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. *J. Bacteriol.* 187, 2050–2057. doi: 10.1128/JB.187.6.2050-2057.2005

- Grover, M., Nain, L., Singh, S. B., and Saxena, A. K. (2010). Molecular and biochemical approaches for characterization of antifungal trait of a potent biocontrol agent *Bacillus subtilis* RP24. *Curr. Microbiol.* 60, 99–106. doi: 10.1007/s00284-009-9508-6
- Herzner, A. M., Dischinger, J., Szekat, C., Josten, M., Schmitz, S., Yakéléba, A., et al. (2011). Expression of the lantibiotic mersacidin in *Bacillus amyloliquefaciens* FZB42. *PLoS One* 6:e22389. doi: 10.1371/journal.pone.0022389
- Huang, C. C., Ano, T., and Shoda, M. J. (1993). Nucleotide sequence and characteristics of the gene, *lpa-14*, responsible for biosynthesis of the lipopeptide antibiotics iturin A and surfactin from *Bacillus subtilis* RB14. *J. Ferment. Bioeng.* 76, 445–450. doi: 10.1016/0922-338x(93)90238-4
- Jin, P. F., Wang, H. N., Liu, W. B., Fan, Y. M., and Miao, W. G. (2018). A new cyclic lipopeptide isolated from *Bacillus amyloliquefaciens* HAB-2 and safety evaluation. *Pestic. Biochem. Physiol.* 147, 40–45. doi: 10.1016/j.pestbp.2017.08.015
- Jin, P. F., Wang, H. N., Liu, W. B., and Miao, W. G. (2017). Characterization of *lpaH2* gene corresponding to lipopeptide synthesis in *Bacillus amyloliquefaciens* HAB-2. *BMC Microbiol.* 17:227. doi: 10.1186/s12866-017-1134-z
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W. Z., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070
- Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., Hooft, J. J. J., et al. (2020). Mibig 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 48, D454–D458. doi: 10.1093/nar/gkz882
- Koren, S., Schatz, M. C., Walenz, P. B., Jeffrey, M., Jason, T. H., Ganeshkumar, G., et al. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30, 693–700. doi: 10.1038/nbt.2280
- Koumoutsis, A., Chen, X. H., Henne, A., Liesegang, H., Hitzeroth, G., Franke, P., et al. (2004). Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *J. Bacteriol.* 186, 1084–1096. doi: 10.1128/JB.186.4.1084-1096.2004
- Kurtz, S., Phillippy, A., Delcher, A. L., Michael, S., Martin, S., Corina, A., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12
- Lagesen, K., Hallin, P., Rodland, E., Staerfeldt, A., Rognes, T. H. H., and Ussery, D. W. (2007). RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160
- Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., and Walsh, C. T. (1996). A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem. Biol.* 3, 923–936. doi: 10.1016/S1074-5521(96)90181-7
- Lee, H., Jung, D. H., Seo, D. H., Chung, W. H., and Seo, M. J. (2020). Genome analysis of 1-deoxynojirimycin (1-DNJ)-producing *Bacillus velezensis* K26 and distribution of *Bacillus* sp. harboring a 1-DNJ biosynthetic gene cluster. *Genomics* doi: 10.1016/j.ygeno.2020.09.061 [Epub ahead of print].
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZY database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6:41. doi: 10.1186/1754-6834-6-41
- Liu, B., and Pop, M. (2009). ARDB-antibiotic resistance genes database. *Nucleic Acids Res.* 37, D443–D447. doi: 10.1093/nar/gkn656
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955
- Makarova, K., Wolf, Y., and Koonin, E. (2015). Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between Thermococcales, Methanococcales, and Methanobacteriales. *Life* 5, 818–840. doi: 10.3390/life5010818
- Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* 97, 2651–2674. doi: 10.1021/cr960029e
- Mariappan, A., Makarewicz, O., Chen, X. H., and Borriss, R. (2012). Two-component response regulator DegU controls the expression of bacilysin in plant-growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *J. Biotechnol.* 22, 114–125. doi: 10.1159/000338804
- Metz, J. G., Roessler, P., Facciotti, D., Levering, C., Dittrich, F., Lassner, M., et al. (2001). Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science* 293, 290–293. doi: 10.1126/science.1059593
- Morikawa, M., Ito, M., and Imanaka, T. (1992). Isolation of a new surfactin producer *Bacillus pumilus*, A-1, and cloning and nucleotide sequence of the regulator gene, *psf-1*. *J. Ferment. Bioeng.* 74, 255–261. doi: 10.1016/0922-338X(92)90055-Y
- Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., et al. (2000). A whole-genome assembly of *Drosophila*. *Science* 287, 2196–2204. doi: 10.1126/science.287.5461.2196
- Nakano, M. M., Corbell, N., Besson, J., and Zuber, P. (1992). Isolation and characterization of *sfp*: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. *Mol. Genet. Genomics* 232, 313–321. doi: 10.1007/bf00280011
- Nakano, M. M., Marahiel, M. A., and Zuber, P. (1990). Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *Bacillus subtilis*. *Genet. Biotechnol. Bacilli* 170, 397–405. doi: 10.1016/B978-0-12-274162-3.50047-5
- Nakayama, K., Takashima, K., Ishihara, H. T., Shinomiya, M., Kageyama, S. K., et al. (2000). The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. *Mol. Microbiol.* 38, 213–231. doi: 10.1046/j.1365-2958.2000.02135.x
- Nandi, T., Ong, C., Singh, A. P., Justin, B., Timothy, A., Mitali, S. T., et al. (2010). A genomic survey of positive selection in *Burkholderia pseudomallei* provide insights into the evolution of accidental virulence. *PLoS Pathog.* 6:e1000845. doi: 10.1371/journal.ppat.1000845
- Qiao, J. Q., Wu, H. J., Huo, R., Gao, X. W., and Borriss, R. (2014). Stimulation and biocontrol by *Bacillus amyloliquefaciens* subsp. *plantarum* FZB42 engineered for improved action. *Chem. Biol. Technol. Agric.* 1:12. doi: 10.1186/s40538-014-0012-2
- Rückert, C., Blom, J., Chen, X. H., Oleg, R., and Rainer, B. (2011). Genome sequence of *B. amyloliquefaciens* type strain DSM7(T) reveals differences to plant-associated *B. amyloliquefaciens* FZB42. *J. Biotechnol.* 155, 78–85. doi: 10.1016/j.jbiotec.2011.01.006
- Ruiz-García, C., Béjar, V., Martínez-Checa, F., Llamas, I., and Quesada, E. (2005). *Bacillus velezensis* sp. nov., a surfactant-producing bacterium isolated from the river Vélez in Málaga, Southern Spain. *Int. J. Syst. Evol. Microbiol.* 55, 191–195. doi: 10.1099/ijs.0.63310-0
- Schäfer, T., and Adams, T. (2015). “The importance of microbiology in sustainable agriculture,” in *Principles of Plant-Microbe Interactions. Microbes for Sustainable Agriculture*, ed. B. Lugtenberg (Berlin: Springer), 5–6. doi: 10.1007/978-3-319-08575-3
- Schneider, K., Chen, X. H., Vater, J., Franke, P., Nicholson, G., Borriss, R., et al. (2007). Macrolactin is the polyketide biosynthesis product of the *pk2* cluster of *Bacillus amyloliquefaciens* FZB42. *J. Nat. Prod.* 70, 1417–1423. doi: 10.1021/np070070k
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092
- Tarailo-Graovac, M., and Chen, N. S. (2009). Using repeat masker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* 5, 4.10.1–4.10.14. doi: 10.1002/0471250953.bi0410s05
- Touchon, M., Bernheim, A., and Rocha, E. P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J.* 10, 2744–2754. doi: 10.1038/ismej.2016.47
- Tsuge, K., Ano, T., and Shoda, M. (1996). Isolation of a gene essential for biosynthesis of the lipopeptide antibiotics plipastatin B1 and surfactin in *Bacillus subtilis* YB8. *Arch. Microbiol.* 165, 243–251. doi: 10.1007/s002030050322
- UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989
- Von Dohren, H., Keller, U., Vater, J., and Rainer, Z. (1997). Multifunctional peptide synthetases. *Chem. Rev.* 97, 2675–2706. doi: 10.1021/cr9600262
- Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

- Wang, L. T., Lee, F. L., Tai, C. J., and Kuo, H. P. (2008). *Bacillus velezensis* is a later heterotypic synonym of *Bacillus amyloliquefaciens*. *Int. J. Syst. Evol. Microbiol.* 58, 671–675. doi: 10.1016/j.diamond.2007.11.010
- Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J., et al. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.* 1:147. doi: 10.1038/ncomms1146
- Weber, T., Blin, K., Duddela, S., Daniel, K., Hyun, U. K., Robert, B., et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43, 237–243. doi: 10.1093/nar/gkv437
- Winnenburg, R., Baldwin, T. K., Urban, M., Rawlings, C., Köhler, J., Hammond-Kosack, K. E., et al. (2006). PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res.* 34, D459–D464. doi: 10.1093/nar/gkj047
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39(Suppl. 2), W347–W352. doi: 10.1093/nar/gkr485
- Zhu, Z. Y., Peng, Q., Man, Y. L., Li, Z. R., Zhou, X. M., Bai, L. Y., et al. (2020). Analysis of the antifungal properties of *Bacillus velezensis* B-4 through a bioassay and complete-genome sequencing. *Front. Genet.* 7:703. doi: 10.3389/fgene.2020.00703

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Xu, Xie, Liu, Jin, Wei, Yaseen, Wang and Miao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.