Complete genome sequence of a marine-sediment-derived bacterial strain *Bacillus velezensis* SH-B74, a cyclic lipopeptides producer and a biopesticide

Zongwang Ma^{1,2} · Jiangchun Hu²

Received: 28 June 2018 / Accepted: 26 March 2019 / Published online: 1 April 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

A marine-sediment sample-derived strain *Bacillus velezensis* SH-B74 has the capacity to produce cyclic lipopeptides (CLPs), and these CLPs secreted by the strain show biological activities against various pests under both in vitro and in planta conditions, such evidence has supported that the strain SH-B74 is a biopesticide. To get a better insight into the mechanisms on the control of the pesticides by the strain, a genome sequencing project has been applied to the genomic DNA of the strain SH-B74. The results show that the strain SH-B74 has a chromosome size of 4,042,190 bp, with a GC content of 46.5%, in addition, the strain contains a 61,634 bp plasmid *p*SH-B74, with a GC content of 40.8%. Data from bioinformatic analysis reveal that the strain SH-B74 has genes with the capacity to increase environmental adaptation, promote the rhizosphere fitnesses and secrete a spectrum of antibiotics, including nonribosomal peptide synthetases (NRPSs)-derived CLPs bacilloleptin, plipastatin, and surfactin. The presence of CLPs in the bacterial cultures of the strain SH-B74 was confirmed further by LC–MS analysis. Thus, genome sequencing and analyses together with chemical analysis reveal the promising perspectives of the strain SH-B74 that are of spectacular importance to its trait as a plant beneficial microbe to be used in agriculture practices.

Keywords *Bacillus velezensis* · Marine-sediment-derived bacterium · Full genome sequence · Cyclic lipopeptides · Environmental adaptation and fitness

Developing microbe-based biological control agent to manage pests is a promising strategy to the sustainable agriculture, moreover, such a strategy is an alternative way to alleviate the side effects (including environmental pollution, increasing pesticides resistance, etc.) caused by the traditional chemical therapy (Ongena and Jacques 2008; Stehle and Schulz 2015). Certain *Bacillus* species are popular microbes possessing the capacity to be developed as biopesticides, either in laboratory research or in commercial market, due to their excellent properties on dealing with

the environmental stress and colonizing rhizosphere niche, on secreting diverse secondary metabolites that not only can directly suppress pests but also can mount the innate immunity in plants, on further promoting the plant growth and enhancing the productivity of plants (Avis et al. 2008; Ongena and Jacques 2008; Velivelli et al. 2014).

A *Bacillus* isolate SH-B74 has been previously isolated from a marine sediment sample collected in the depth of 3601 m from the South China Sea (17°53'59.545 N, 114°34'58.614 E), and the strain has been formally identified as *B. amyloliquefaciens*. However, a report has recently confirmed that *B. amyloliquefaciens* belongs to *B. velezensis* (Dunlap et al. 2016), such evidence allows us to feasibly dub the strain SH-B74 as *B. velezensis*. Further data from dual confrontation assay show that the strain *B. velezensis* SH-B74 has good antagonistic activity against *Rhizoctonia solani* (Data not shown), indicating that the strain *B. velezensis* SH-B74 has the potential to be developed as a biopesticide. In addition, previous results display that the strain *B. velezensis* SH-B74 has the ability to produce bacilloleptins

✉ Zongwang Ma
zongwang.ma@nwnu.edu.cn

✉ Jiangchun Hu
hujc@iae.ac.cn

¹ College of Life Science, Northwest Normal University, Lanzhou 730070, China

² Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China

(Niu et al. 2011; Ma et al. 2014), plipastatins (Ma and Hu 2018) and a group of uncharacterized CLPs (Ma et al., unpublished data). Bacillopeptin A, B, a recently identified new CLP bacillopeptin B₁, and plipastatin A1 produced by the strain *B. velezensis* SH-B74 indicate direct antagonistic activity against a spectrum of plant pathogens, moreover, CLP plipastatin A1 has the capacity of controlling the infection of *Botrytis cinerea* (a causal agent of gray mold disease) in tomato plants. Collectively, the strain *B. velezensis* SH-B74 is a potent CLPs producer and a good candidate for biopesticide.

To gain more detailed insight into the genetic information of the strain *B. velezensis* SH-B74, a project has been launched to sequence the genomic DNA of the strain *B. velezensis* SH-B74. Multiple bioinformatic tools have been used to analyze the genetic features of the strain *B. velezensis* SH-B74, and the detailed genomic data of the strain *B. velezensis* SH-B74 will be described in this report.

Genomic DNA of the strain *B. velezensis* SH-B74 were extracted using a kit (EZ-10 Spin Column Bacterial Genomic DNA Isolation Kit) from Sangon Biotech (Shanghai) Co., Ltd., China, and the extraction approach was followed according to the manufacturer's instructions. The genomic DNA of the strain *B. velezensis* SH-B74 was sequenced using a 20-kb Pacific Biosciences (PacBio) RSII single-molecule real-time (SMRT) cell over PacBio sequencing technology. Subreads data were assembled by Canu software (Koren et al. 2017). The encoding genes from the genome sequence of the strain *B. velezensis* SH-B74 were predicted by Prodigal (Hyatt et al. 2010). Repetitive sequences from the genome sequence of the strain *B. velezensis* SH-B74 were predicted by RepeatMasker (Tarailo-Graovac and Chen 2009). rRNA, tRNA and other noncoding RNA from the genome sequence of the strain *B. velezensis* SH-B74 were analyzed by Infernal 1.1 based on the Rfam database (Nawrocki and Eddy 2013). CRT software was used for prediction of the Clustered Regularly Interspaced Palindromic Repeats (CRISPRs) sequence in the chromosome of the strain *B. velezensis* SH-B74 (Bland et al. 2007). Genome islands were predicted by IslandPath-DIMOB from the genome sequence of the strain *B. velezensis* SH-B74 (Langille et al. 2008). Prophage was detected by the PhiSpy software from the genome sequence of the strain *B. velezensis* SH-B74 (Akhter et al. 2012).

The results from sequencing and assembling of the genomic DNA of the strain *B. velezensis* SH-B74 indicate that the strain contains a chromosome and a plasmid pSH-B74. The chromosome size of the strain *B. velezensis* SH-B74 is 4,042,190 bp, with a GC content of 46.5%, alongside with a circular plasmid pSH-B74 of 61,634 bp, GC content of which is 40.8%. Data from bioinformatic analysis revealed that the chromosome of the strain *B. velezensis*

SH-B74 contains 3888 coding DNA sequences (CDSs), 27 rRNA operons (including 9 of 23S rRNA, 9 of 16S rRNA and 9 of 5S rRNA, respectively), 86 transfer RNA (tRNA) genes and 87 other noncoding RNA genes.

The total length of the repetitive gene sequence from the chromosome of the strain *B. velezensis* SH-B74 was predicted as 5300 bp. CRISPRs sequences are kinds of important repetitive sequences, and the results from bioinformatic analysis display that the chromosome of the strain *B. velezensis* SH-B74 contains 16 CRISPRs sequences. Nine genome islands were detected in the chromosome of the strain *B. velezensis* SH-B74. Two prophage genes, with a length of 35,579 bp and 52,886 bp, respectively, were detected in the chromosome of the strain *B. velezensis* SH-B74.

Circular map of the chromosome of the strain *B. velezensis* SH-B74 was shown in Fig. 1a. The plasmid pSH-B74 contains 24 CDSs, the function of most genes in the plasmid pSH-B74 remains unclear. The circular map of the plasmid pSH-B74 was shown in Fig. 1b. Different color (Fig. 1c) illustrates different functions of the COG-annotated CDSs of the chromosome (Fig. 1a) and the plasmid (Fig. 1b). The CDSs in the chromosome and plasmid of the strain *B. velezensis* SH-B74 were annotated by COG database, and the results were shown in Table 1.

The annotation of the CDSs in the chromosome of the strain *B. velezensis* SH-B74 by COG database indicates that the chromosome contains CDSs for cell motility, inorganic ion transport and metabolism, secondary metabolites biosynthesis, as well as defense. These biological functions are important for the strain *B. velezensis* SH-B74 in dealing with stress from natural environment and competition with other organisms for nutrition, space, etc.

Genome mining results by antiSMASH 4.0 (Blin et al. 2017) showed that the strain *B. velezensis* SH-B74 has the gene clusters for synthesizing a set of secondary metabolites, including those metabolites predicted with high scores ($\geq 91\%$) to know gene clusters, such as macrolactin, bacillaene, difficidin, bacillibactin, bacilysin, CLPs bacillopeptins, plipastatins and surfactins, and two metabolites, butirosin and rhizoctin, with low scores ($\leq 22\%$) to corresponding known gene clusters (Table 2). CLPs produced by *Bacillus* sp. have been identified as the main antibiotics displaying fungicidal activities (Cawoy et al. 2015), and as elicitors to trigger immunity in plants (Farace et al. 2015; Ongena et al. 2007). Additionally, the presence of a gene cluster that codes for plant hormone auxin has been confirmed by bioinformatic analysis in the chromosome of the strain *B. velezensis* SH-B74, displaying additional evidence of the strain on the regulation of plant growth, development, and immunity (De Vleeschauwer et al. 2013).

Fig. 1 The genome map of the chromosome (**a**) and the plasmid (**b**) of the strain *B. velezensis* SH-B74 generated by Circos (Krzyszewski et al. 2009). From inner circle to outer circle: the first circle means GC skew, in which gray color region means that the content of G>C content, while red color region means that the content of C>G content. The yellow part means that the GC content is higher than 46.5%, while the blue part means that the GC content is lower than 46.5% in the second circle. The third circle indicates tRNA (blue bars) and rRNA (purple bars), respectively. The fourth circle means

the repetitive sequences. The fifth (the forward strand) and the sixth (the reverse strand) circles show the predicted CDSs of chromosome of the strain *B. velezensis* SH-B74. The different types of Clusters of Orthologous Groups of proteins (COGs) function showing the different color in **c** were marked as the same color in the fifth and sixth circles of the **a** and **b**. The genes from the strain *B. velezensis* SH-B74 not annotated by COG database were marked as 'Not COG annotated' in **c**

The bacterial cultures and a crude extract of CLPs from the strain *B. velezensis* SH-B74 were prepared following a protocol published previously (Ma and Hu 2018). Subsequently, the fractions of different types of CLPs from the crude extract of the strain *B. velezensis* SH-B74 were separated using solid-phase extraction process as published previously (Ma and Hu 2018). Then, the fractions were submitted to a TSQ Quantum Access Max Triple Quadrupole mass spectrometer (Thermo Scientific, San Jose, CA) equipped with a RP-HPLC column (ϕ 2.1 $\times$ 150 mm, Waters Atlantis T3 Column, Waters Corporation, Dublin, Ireland)

for analysis. A series of ions (mass-to-charge ratio, m/z , see Fig. 2 and Table 3) were detected by LC-MS and these ions confirmed that the strain *B. velezensis* SH-B74 has the capacity to produce CLPs bacillopeptins, plipastatins, and surfactins. The assignment of different CLPs family was conducted by comparing the ion data reported previously for these CLPs (Niu et al. 2011; Ma et al. 2014; Ma and Hu 2015, 2018).

To sum up, this study sheds additional sights on the genomic perspectives of the biocontrol property, environmental fitness and adaptation of the strain *B. velezensis*

Table 1 Clusters of Orthologous Groups of proteins (COGs) from the chromosome and the plasmid *pSH-B74* of the strain *B. velezensis* SH-B74 were annotated by COG database (Tatusov et al. 2000)

Annotated functions for COGs	Numbers of COGs for	
	Chromosome	Plasmid
RNA processing and modification	0	0
Chromatin structure and dynamics	1	0
Energy production and conversion	184	0
Cell cycle control, cell division, chromosome partitioning	35	2
Amino acid transport and metabolism	358	0
Nucleotide transport and metabolism	78	0
Carbohydrate transport and metabolism	262	0
Coenzyme transport and metabolism	126	0
Lipid transport and metabolism	121	0
Translation, ribosomal structure and biogenesis	164	0
Transcription	300	4
Replication, recombination and repair	130	4
Cell wall/membrane/envelope biogenesis	179	2
Cell motility	56	0
Posttranslational modification, protein turnover, chaperones	102	0
Inorganic ion transport and metabolism	224	0
Secondary metabolites biosynthesis, transport and catabolism	131	0
General function prediction only	466	6
Function unknown	301	6
Signal transduction mechanisms	146	0
Intracellular trafficking, secretion, and vesicular transport	48	0
Defense mechanisms	52	0
Extracellular structures	0	0
Nuclear structure	0	0
Cytoskeleton	0	0
Not COG annotated	595	96

Annotated functions, numbers of COGs for chromosome and plasmid are shown in this table

Table 2 The gene clusters responsible for secondary metabolite biosynthesis from the chromosome of the strain *B. velezensis* SH-B74 predicted by antiSMASH 4.0 are shown (Blin et al. 2017)

Number	Type	From ^a	To ^a	Size (kb)	Metabolite
1	NRPS	61,995	127,363	65.368	Surfactin
2	Others	666,099	707,343	41.244	Butirosin
3	Terpene	790,144	810,884	20.740	Unknown
4	TransatPKS	1,160,513	1,246,409	85.896	Macrolactin
5	TransatPKS-NRPS	1,475,938	1,578,630	102.692	Bacillaene
6	TransatPKS-NRPS	1,641,931	1,779,779	137.848	Plipastatin/Iturin
7	Terpene	1,845,794	1,867,677	21.883	Unknown
8	T3PKS	1,918,106	1,959,206	41.100	Unknown
9	TransatPKS	2,075,121	2,175,571	100.450	Difficidin
10	Bacteriocin-NRPS	2,799,286	2,866,080	66.794	Bacillibactin
11	Other	3,377,145	3,418,563	41.418	Bacilysin
12	TransatPKS-NRPS	3,913,840	3,991,570	77.730	Rhizocticin

^a‘From’ and ‘To’ indicate the location of gene clusters responsible for secondary metabolite biosynthesis in the chromosome of the strain *B. velezensis* SH-B74

NRPS nonribosomal peptide synthetase, PKS polyketide synthase

Fig. 2 LC–MS analysis of CLPs (**a** for bacillopeptins, **b** for plipastatins and **c** for surfactins, respectively) produced by the strain *B. velezensis* SH-B74

SH-B74 as a biopesticide. We have experimentally confirmed that the strain *B. velezensis* SH-B74 has the capacity to produce CLPs and these CLPs are important molecules for the biocontrol of certain plant pathogens. However, the detailed mechanisms of the strain *B. velezensis* SH-B74 on biocontrol of the other pests need to be studied and elucidated in more detail.

Nucleotide sequence accession numbers: the complete sequence of the chromosome of the strain *B. velezensis* SH-B74 and the sequence of the plasmid *p*SH-B74 of the

strain *B. velezensis* SH-B74 have been deposited in GenBank database under the accession number CP030097 and CP030098, respectively. The strain *B. velezensis* SH-B74 is publicly available from China Center for Type Culture Collection (CCTCC) under the accession number CCTCC AB 2018012.

Table 3 Ions were detected by LC–MS analysis from the samples of CLPs produced by the strain *B. velezensis* SH-B74

No.	Ions			CLPs family
	[M+2H] ²⁺	[M+H] ⁺	[M+Na] ⁺	
1	511.10	1021.28	1043.26	Bacillopeptin
2	518.22	1035.31	1057.31	
3	n.d.	1049.18	1071.32	
4	711.14	1421.47	n.d.	Plipastatin
5	718.34	1435.75	n.d.	
6	725.13	1449.62	n.d.	
7	732.13	1463.75	n.d.	
8	739.20	1477.68	n.d.	
9	746.27	1491.75	n.d.	
10	753.90	1505.81	n.d.	
11	755.23	1509.80	n.d.	
12	760.20	1519.95	n.d.	
13	762.29	1523.87	n.d.	
14	769.30	1537.91	n.d.	
15	n.d. ^a	994.40	1016.39	Surfactin
16	n.d.	1008.42	1030.47	
17	n.d.	1022.44	1044.46	
18	n.d.	1036.45	1058.45	
19	n.d.	1050.49	1072.52	

^an.d. means that the corresponding ion signals were not detected by LC–MS analysis

Acknowledgements Z. Ma conceived and designed the study, conducted the experiments and wrote the paper. This work was supported by a grant from Science and Technology Service Network Initiative of Chinese Academy of Sciences (CAS) (Grant No. KJFJ-EW-STS-143) to J. Hu.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Akhter S, Aziz RK, Edwards RA (2012) PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res* 40:e126–e126. <https://doi.org/10.1093/nar/gks406>
- Avis TJ, Gravel V, Antoun H, Tweddell RJ (2008) Multifaceted beneficial effects of rhizosphere microorganisms on plant health and productivity. *Soil Biol Biochem* 40:1733–1740. <https://doi.org/10.1016/j.soilbio.2008.02.013>
- Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P (2007) CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209
- Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T,

- Medema MH (2017) antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>
- Cawoy H, Debois D, Franzil L, Pauw ED, Thonart P, Ongena M (2015) Lipopeptides as main ingredients for inhibition of fungal phytopathogens by *Bacillus subtilis/amyoliquefaciens*. *Microb Biotechnol* 8:281–295. <https://doi.org/10.1111/1751-7915.12238>
- De Vleeschauwer D, Gheysen G, Hofte M (2013) Hormone defense networking in rice: tales from a different world. *Trends Plant Sci* 18:555–565. <https://doi.org/10.1016/j.tplants.2013.07.002>
- Dunlap CA, Kim S-J, Kwon S-W, Rooney AP (2016) *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyoliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyoliquefaciens* subsp. *plantarum* and ‘*Bacillus oryzicola*’ are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. *Int J Syst Evol Micr* 66:1212–1217. <https://doi.org/10.1099/ijsem.0.000858>
- Farace G, Fernandez O, Jacquens L, Coutte F, Krier F, Jacques P, Clément C, Barka EA, Jacquard C, Dorey S (2015) Cyclic lipopeptides from *Bacillus subtilis* activate distinct patterns of defence responses in grapevine. *Mol Plant Pathol* 16:177–187. <https://doi.org/10.1111/mpp.12170>
- Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645
- Langille MG, Hsiao WW, Brinkman FS (2008) Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics* 9:329. <https://doi.org/10.1186/1471-2105-9-329>
- Ma Z, Hu J, Wang X, Wang S (2014) NMR spectroscopic and MS/MS spectrometric characterization of a new lipopeptide antibiotic bacillopeptin B₁ produced by a marine sediment-derived *Bacillus amyoliquefaciens* SH-B74. *J Antibiot* 67:175–178. <https://doi.org/10.1038/ja.2013.89>
- Ma Z, Hu J (2018) Plipastatin A1 produced by a marine sediment-derived *Bacillus amyoliquefaciens* SH-B74 contributes to the control of gray mold disease in tomato. *3 Biotech* 8:125. <https://doi.org/10.1007/s13205-018-1144-z>
- Ma Z, Hu J (2015) Production and characterization of surfactin-type lipopeptides as bioemulsifiers produced by a *Pinctada martensii*-derived *Bacillus mojavensis* B0621A. *Appl Biochem Biotechnol* 177:1520–1529. <https://doi.org/10.1007/s12010-015-1832-7>
- Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29:2933–2935. <https://doi.org/10.1093/bioinformatics/btt509>
- Niu LX, Wang N, Wang XM, Hu JC, Wang SJ (2011) Screening and identification of a deep-sea derived *Bacillus* sp. producing lipopeptide bacillopeptin A. *Chin J Antibio* 10:738–744
- Ongena M, Jourdan E, Adam A, Paquot M, Brans A, Joris B, Arpigny JL, Thonart P (2007) Surfactin and fengycin lipopeptides of *Bacillus subtilis* as elicitors of induced systemic resistance in plants. *Environ Microbiol* 9:1084–1090
- Ongena M, Jacques P (2008) *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. *Trends in Microbiol* 16:115–125. <https://doi.org/10.1016/j.tim.2007.12.009>
- Stehle S, Schulz R (2015) Agricultural insecticides threaten surface waters at the global scale. *Proc Natl Acad Sci USA* 112:5750–5755. <https://doi.org/10.1073/pnas.1500232112>

- Tarailo-Graovac M, Chen N (2009) Using repeatmasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinform* 25:4.10.11–14.10.14. <https://doi.org/10.1002/0471250953.bi0410s25>
- Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33–36
- Velivelli SL, De Vos P, Kromann P, Declerck S, Prestwich BD (2014) Biological control agents: from field to market, problems, and challenges. *Trends in Biotechnol* 32:493–496. <https://doi.org/10.1016/j.tibtech.2014.07.002>