Complete genome sequence and epigenetic profile of *Bacillus velezensis* UCMB5140 used for plant and crop protection in comparison with other plant-associated *Bacillus* strains

Oleg N. Reva[1,*],Safronova A. Larisa[2], Aneth D. Mwakilili[3,4], Donatha Tibuhwa[3], Sylvester Lyantagaye[3], Wai Yin Chan[5,6,7], Stefanie Lutz[8], Christian H. Ahrens[8], Joachim Vater[9] and Rainer Borriss[10]

[1]Centre for Bioinformatics and Computational Biology, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Hillcrest, Lynnwood Rd., Pretoria, South Africa

[2]Innovation and Technology Transfer Laboratory, DK Zabolotny Institute of Microbiology and Virology, 154 Zabolotnogo Str, Kyiv 03143, Ukraine

[3]Department of Molecular Biology and Biotechnology, University of Dar es Salaam, Dar es Salaam, Tanzania

[4]Plant Protection Department, Swedish University of Agricultural Sciences (SLU), Alnarp, Sweden

[5]Biotechnology Platform (BTP), Agricultural Research Council, Onderstepoort Veterinary Research Campus, Old Soutpan Rd, Onderstepoort, South Africa

[6]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

[7]Forestry and Agricultural Biotechnology Institute (FABI), DST-NRF Centre of Excellence in Tree Health Biotechnology (CTHB), University of Pretoria, Pretoria, South Africa

[8]Agroscope, Molecular Diagnostics, Genomics and Bioinformatics &SIB Swiss Institute of Bioinformatics, Müller-Thurgau-Str. 29,8820 Wädenswil, Switzerland

[9]Robert Koch Institut, Berlin, Germany

[10]Institut für Biologie, Humboldt Universität Berlin, Berlin, German

*Correspondence to: oleg.reva@up.ac.za

Abstract

The application of biocontrol biopesticides based on plant growth–promoting rhizobacteria (PGPR), particularly members of the genus *Bacillus,* is considered a promising perspective to make agricultural practices sustainable and ecologically safe. Recent advances in genome sequencing by third-generation sequencing technologies, e.g., Pacific Biosciences' Single Molecule Real-Time (PacBio SMRT) platform, have allowed researchers to gain deeper insights into the molecular and genetic mechanisms of PGPR activities, and to compare whole genome sequences and global patterns of epigenetic modifications. In the current work, this approach was used to sequence and compare four *Bacillus* strains that exhibited various PGPR activities including the strain UCMB5140, which is used in the commercial biopesticide Phytosubtil. Whole genome comparison and phylogenomic inference assigned the strain UCMB5140 to the species *Bacillus velezensis*. Strong biocontrol activities of this strain were confirmed in several bioassays. Several factors that affect the evolution of active PGPR *B. velezensis* strains were identified: (1) horizontal acquisition of novel non-ribosomal peptide synthetases (NRPS) and adhesion genes; (2) rearrangements of functional

modules of NRPS genes leading to strain specific combinations of their encoded products; (3) gain and loss of methyltransferases that can cause global alterations in DNA methylation patterns, which eventually may affect gene expression and regulate transcription. Notably, we identified a horizontally transferred NRPS operon encoding an uncharacterized polypeptide antibiotic in *B. velezensis* UCMB5140. Other horizontally acquired genes comprised a possible adhesin and a methyltransferase, which may explain the strain-specific methylation pattern of the chromosomal DNA of UCMB5140.

Key points

- *Whole genome sequence of the active PGPR Bacillus velezensis UCMB5140.*
- *Identification of genetic determinants responsible for PGPR activities.*
- *Role of methyltransferases and epigenetic mechanisms in evolution of bacteria.*

Keywords

Bacillus velezensis; Biocontrol; Biopesticides; Genomics; Epigenetics; Whole genome sequencing; Plant growth–promoting rhizobacteria

Introduction

Bacillus velezensis and other representatives of the *Bacillus subtilis* taxonomic group (taxid:653685) are among the most popular plant growth–promoting rhizobacteria (PGPR) used in biotechnology (Aloo et al. 2019; Hashem et al. 2019; Saxena et al. 2019). These bacteria are valuable candidates for commercial biopesticides due to their biocontrol activity against various phytopathogens. Their plant growth promotion activity on seedling and plant developments in combination with the stability of spore-based product formulations makes them easy for agricultural application and long-term storage.

Despite being widely applied in agriculture and being the subject of numerous published studies, it is still rather difficult to predict which genetic factors are responsible for the ability of *Bacillus* strains to serve as biopesticides. Recently, the use of complete genome sequences in comparative genomic studies allowed linking numerous genetic markers with varying abilities of potential biocontrol strains to protect agricultural plants from pathogens (De Vrieze et al. 2020). Moreover, it is unclear on the effect of various physical or biological field conditions, i.e., geographical regions, rainfall intensity, physicochemical soil characteristics, and specific plant cultivars to be protected, which may alternate or limit the applicability as biopesticides. Plant growth–promoting rhizobacteria (PGPR) *B. velezensis* strains produce several bioactive peptides which function as antifungal and antibacterial antibiotics (Koumoutsi et al. 2004; Schneider et al. 2007; Chen et al. 2009a, b; Yuan et al. 2012; Liu et al. 2013; Wu et al. 2015; Gu et al. 2017; Lu et al. 2018; Kim et al. 2019). In addition, they are equipped with many enzymes that generate hormone-like compounds and volatiles, which improve plant nourishing and stimulate plant responses to biotic and abiotic stresses (Idriss et al. 2002; Ryu et al. 2004; Li et al. 2015; Wu et al. 2018; Abd El-Daim et al. 2019; Khan et al. 2019). While some of these genes are strain specific, most of them are well conserved in other *B. velezensis* genomes. Importantly, not all *B. velezensis* strains express PGPR activities, and in certain cases, the PGPR activities are induced only in response to root exudate exposure of phytopathogen attacks (Reva et al. 2019). These observations imply that the expression of biocontrol and PGPR activities is likely

controlled by specific gene regulation rather than, or on top of, the presence or absence of certain PGPR-related genes in genomes of selected *Bacillus* strains.

Often the industrial PGPR strains are selected empirically from a collection of promising isolates based on a rather arbitrary set of in vitro and in vivo assays, which do not guarantee the strain to be of biotechnological value (Reva et al. 2019). Eventually, only those strains showing reproducible activity under field conditions may find a biotechnological application. A better understanding of the molecular and genetic determinants and mechanisms of PGPR activities will aid in the development of active biopesticides with predictable activities against certain pathogens affecting agricultural plants in various geographic regions.

The aim of this research was to compare one specific *Bacillus* strain, UCMB5140, selected empirically for plant and crop biocontrol, with several comprehensively studied model organisms showing various bioactivities under laboratory conditions. For this comparison, several bioassays, whole genome sequencing, and profiling of global epigenetic modifications of the chromosomal DNA were used. Recent advances in genome sequencing by third-generation sequencing technologies, e.g., Pacific Biosciences' Single Molecule Real-Time platform (PacBio SMRT), have allowed researchers to gain deeper insights into the molecular and genetic mechanisms of PGPR activities, and to compare whole genome sequences and global patterns of epigenetic modifications. These research tasks have been rather problematic and inaccurate until recently, when Illumina short-read-based genome assemblies were used. Particularly, the genomic assemblies based on Illumina reads frequently missed large fragments of giant genes encoding non-ribosomal polypeptide synthetase (NRPS) and polyketide synthetase (PKS), which play an important key role in PGPR activities (Caulier et al. 2019; Varadarajan et al. 2020).

The strain *Bacillus* sp. UCMB5140 has been used in the biopesticide Phytosubtil patented in Ukraine (Safronova et al. 2012, 2013). This product has been applied to control bacterial and fungal phytopathogens on soybean plantations and to prolong the shelf-life of harvested sugar beet tubers and potatoes. We here sequenced its complete genome and that of the promising PGPR strain *B. subtilis* UCMB5021, which then were compared with two other *B. velezensis* strains UCMB5113 and UCMB5044, whose genomes had been sequenced and characterized previously (Niazi et al. 2014; Reva et al. 2019). For comparative studies, several bioassays, whole genome sequencing, and profiling of global epigenetic modifications of the chromosomal DNA were used to identify an NRPS operon uniquely encoded by strain UCMB5140. Furthermore, an adhesion and a methyltransferase were also identified among the strain-specific genes that may in turn explain some of the observed phenotypic differences.

Materials and methods

Growth conditions

Bacterial strains UCMB5140, UCMB5044, UCMB5113, and UCMB5021 were obtained from the Ukrainian Collection of Microorganisms, section Bacteria (UCMB) at the D.K. Zabolotny Institute of Microbiology and Virology, Kyiv (Ukraine). Lyophilized stock cultures were kept at room temperature or as frozen spore suspensions at − 20 °C. The strains were cultured in Luria–Bertani liquid broth (LB; Sigma-Aldrich, Darmstadt, Germany) or solid LB agar (LBA) medium at 28 or 37 °C. *Bacillus* spores were harvested after 5 days of cultivation on LBA plates at 37 °C and re-suspended in sterile water to achieve 10 OD at 600 nm.

Antagonistic activity assay against the fungal pathogen *Aspergillus* sp.

The antagonistic activity of the selected strains was determined in vitro against cultures of the fungal phytopathogen *Aspergillus* sp. P12 that had been isolated in the Department of Molecular Biology and Biotechnology at the University of Dar es Salaam from rotting sweet potato tubers. The isolate was identified by cell and colony morphology. It was grown on potato dextrose agar (PDA; Sigma-Aldrich, Johannesburg, South Africa) plates at 28 °C until spore production. Spores were harvested by adding sterile water to the plates and scraping the agar surface with a sterilized spatula. Spore suspensions were diluted with sterile water to reach the standard optical density of 10 OD at 600 nm.

A modified dual-culture technique (Gupta et al. 2001) was used, when *Bacillus* strains were co-cultivated with disease-causing microorganisms on solid media. *Bacillus* strains were plated on a crisscrossing/draft way on PDA and cultivated for 24 h at 37 °C. The fungal test culture grown in potato dextrose broth (Sigma-Aldrich) for 24 h at room temperature on a shaker was inoculated as a drop in the middle of the *Bacillus* draft-plates using a sterile wire loop. Positive controls of the disease-causing fungi and bacteria strains were grown on separate PDA plates. The plates were sealed and incubated at 37 °C for 10 days. Each experiment was carried out in triplicate. Fungal growth measurements were done on days 3, 5, and 10, by recording the diameters of the macro-colony (edge-to-edge through the middle of the colony). The percentage of fungal growth inhibition was calculated as follows: (D1 − D2)/D1, using the diameter of a fungal colony without *Bacillus* growth as the control (D1) and the diameter of the fungal colony affected by the *Bacillus* growth as D2.

***Ralstonia solanacearum* biocontrol assays on tomato seedlings**

R. solanacearum MBBC32 causing bacterial wilt of tomato was obtained from the culture collection of phytopathogens at the University of Dar es Salaam (Tanzania). This strain was isolated in 2014 in Tanzania from diseased tomato plants and identified by phenotype, cell morphology, and standard biochemical tests. The isolate was cultured on LB plates at 37 °C for 24 h. A stock suspension of 10 OD (at 600 nm) was obtained by flooding each plate with sterile water prior to scraping off the bacteria using a flame-sterilized spatula.

Tomato seeds (commercial variety The Rio Grande obtained from East-West Seed Tanzania, Kilimanjaro, Tanzania) were surface sterilized by soaking them in 0.1% sodium hypochlorite (Orbit products Africa Limited, Nairobi, Kenya) for 3 min and then rinsed four times with sterile water. Individual seeds were treated separately by soaking each for 6 h in a *Bacillus* spore suspension (10 OD at 600 nm) prepared as described previously. Control seeds were soaked with sterile water for the same period of time. Seeds were air dried after treatments before being planted into sterilized soil in a nursery in a greenhouse with controlled temperature and humidity (relative humidity between 75 and 80% and temperature 25–28 °C), each kept separate according to the treatment. Nurseries were watered every other day for 4 weeks. The seedlings were then transferred on to sterile soil in 18-mm-diameter plastic pots (Jambo Plastics Ltd., Dar es Salaam, Tanzania), three seedlings per pot, in five replicates for every treatment. The pots remained in the greenhouse for the rest of the experimental period. Soil for this experiment was obtained from virgin land near the campus of the University of Dar es Salaam. Soil was prepared by homogenization and sieving. Then it was packed in 1-kg bags and autoclaved at 121 °C for 20 min. General parameters of the prepared soil were analyzed by the soil laboratory of the International Institute for Tropical Agriculture (Dar es Salaam, Tanzania).

After 7 days, 0.2 L *R. solanacearum* MBBC32 working suspension (0.1 × stock solution, 1 OD at 600 nm) was applied as a drench onto each of the *Bacillus*-treated and positive control pots. Negative control pots were uninfected. Seedlings were watered every other day during the experiment. The experiments were conducted in a greenhouse with a relative humidity between 75 and 80% and temperature 25–28 °C. Symptom development and disease progression were recorded every 7 days for a month post-pathogen treatment. Tomato wilting was scored according to the following disease index: 0—no symptoms (healthy plant); 1—one leaf partially wilted; 2—two to three wilted leaves; 3—all but one to three leaves wilted; 4—all leaves wilted; 5—plant dead. Disease progression curves were plotted and the areas under disease progression curve (AUDPC) were calculated using the recorded indices. A smaller AUDCP indicates better disease prevention. AUDPC values were presented as percentages of disease severity index counts calculated as average for the positive control plants (infected with *R. solanacearum* and untreated with *Bacillus*).

Statistical evaluation of plant protection and plant growth promotion experiments

Data were recorded, pooled, and summarized in MS Excel for Mac, version 14.4.5 (Microsoft, Redmond, WA). Average values and SDs were calculated. The graphs were also plotted in MS Excel.

Detection of lipopeptides, siderophores, and polyketides

Secondary metabolites were identified in culture media according to methods published by Reva et al. (2019). *B. velezensis* strains were grown in Landy liquid medium (Landy et al. 1948). Supernatant fractions were collected at 24 and 48 h. Identification of lipopeptides, siderophores, and polyketides was performed using matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) MS. Filtered supernatants were mixed with 50% acetonitrile containing 0.1% trifluoroacetic acid. Then 2-µl portions of extracts and 2 µl of 2.5-dihydroxbenzoic acid matrix solution were mixed, spotted on to the target, and air dried. The MALDI-TOF MS measurements were performed as described by Vater et al. (2002) using a Bruker Autoflex MALDI-TOF instrument (Bruker-Daltonix, Bremen, Germany) containing a 337-nm nitrogen laser. Reagents for this experiment were purchased from Sigma-Aldrich (Darmstadt, Germany).

Separation of the bioactive secondary metabolites produced by the *Bacillus* strains was performed using HPLC on an Agilent 1200 HPLC system (Agilent, Waldbronn, Germany). Acetonitrile (ACN)–water extracts of the freeze-dried supernatants were fractionated by reversed-phase HPLC on a Zorbax Eclipse XDB HPLC C18 3.5-µm (4.6 × 150 mm) column at a flow rate of 1.5 ml/min using an increasing gradient from 0% ACN to 100% ACN with 0.1% formic acid (Sigma-Aldrich, Darmstadt, Germany) in 15 min. The characteristic peaks of eluted compounds were recorded by measuring OD values at two wavelengths, 365 and 230 nm, and compared with the standard reference values of the non-ribosomal peptide and polyketide products, published by Chen et al. (2006).

DNA extraction and sequencing

DNA was extracted from overnight liquid LB cultures grown at 37 °C with the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich, Johannesburg, South Africa). DNA quality was evaluated spectrophotometrically at 260 and 280 nm wavelength absorbance with a NanoDrop 1000 spectrometer (Thermo Fisher Scientific, Johannesburg, South Africa). The complete genome of UCMB5140 was sequenced using the PacBio RSII CCS technology at Inqaba Biotec Ltd. (Pretoria, South Africa). The DNA extracted from UCMB5021 was sequenced on a PacBio RSII machine (1 SMRT cell per strain) at the Functional Genomics Center Zurich, Switzerland. Size selection was performed using the BluePippin system (SAGE Science, Châtel-St-Denis, Switzerland) and resulted in fragments with an average subread length of 8–10 kb. Two 2 × 300 bp Illumina paired-end libraries were

prepared per strain using the Nextera XT DNA kit and sequenced on a MiSeq at Agroscope (Wädenswil, Switzerland).

Genome assembly and annotation

For UCMB5140, a draft assembly was created using the PacBio CCS reads and Ra v.0.2.1 using default parameters (Vaser et al. 2017). The contigs were scaffolded using the PacBio CCS reads and LRScaf v.1.1.6 using default parameters (Qin et al. 2019) before polishing them using two runs of the polish-target module of Flye v.2.4 (Kolmogorov et al. 2019). The quality of the assembly was assessed by mapping the PacBio reads to the contigs using *GraphMap* v.0.5.2 with default parameters (Sović et al. 2016) and inspecting the alignments in the Integrated Genome Viewer (IGV) (Thorvaldsdóttir et al. 2013) as described (Omasits et al. 2017). Genome assembly resulted in three contigs. Gaps between contigs were identified by aligning the contigs against the complete genome sequence of a closely related strain *B. velezensis* UCMB5036 (Manzoor et al. 2013) using the functions Move Contigs and progressiveMauve of the program Mauve 20150226 (Darling et al. 2010). Gaps were patched with the corresponding DNA fragments of the reference genome and then the alignment of the initial PacBio reads was repeated to rectify strain-specific polymorphisms in the sequenced genome.

For UCMB5021, PacBio RSII subreads were extracted using the SMRT Portal and protocol RS_HGAP_Assembly.3 using default parameters with minimum subread length set for 1000, estimated genome size of 4 Mb. The subreads were then de novo assembled using Flye v.2.3.3 (Kolmogorov et al. 2019). Default parameters were applied except for the estimated genome size set to 4 Mb. The assembly resulted in a single contig, which was start-aligned to the gene *dnaA* and polished using two Quiver runs and the RS_Resequencing.1 protocol on the SMRT Portal with default parameters. The assembly was then further polished using 2 × 300 bp Illumina reads and 1–2 Freebayes v.1.2.0 runs (minimum alternate fraction, 0.5; minimum alternate count, 5) (Garrison and Marth 2012), and subsequently corrected using *bcftools* v.0.1.19 (Narasimhan et al. 2016) to correct any potentially remaining small sequencing errors. The filtered PacBio subreads were mapped to the polished and start-aligned contigs using *graphmap* v.0.5.2 to verify the circularity and completeness of the assembly that was then manually inspected in IGV. PlasmidSPAdes (Antipov et al. 2016) was run on the Illumina data as all PacBio approaches involved a BluePippin size selection step, which might have led to the removal of smaller plasmids.

The completeness of the final assembly was further evaluated using the benchmarking universal single-copy orthologous (BUSCO) software (Simão et al. 2015). Assembled genomes were annotated using the online RAST Server (http://rast.nmpdr.org; Aziz et al. 2008) and then manually curated in Artemis 14.0.0 (https://www.sanger.ac.uk/science/tools/artemis) using data from SubtiWiki and AmyloWiki Web resources (Zhu and Stülke 2017; Fan et al. 2019). Orthologous genes were predicted by OrthoFinder (Emms and Kelly 2015). Clusters of genes encoding biosynthesis of secondary metabolites were identified with antiSMASH (Blin et al. 2019). Adenylation and condensation domains in non-ribosomal polypeptide synthetases (NRPS), and their amino acid specificity, were identified with antiSMASH and the PKS/NRPS Analysis Web-tool (http://nrps.igs.umaryland.edu/). SeqWord Genome Island Sniffer (Bezuidt et al. 2009) was used to identify possible insertions of horizontally transferred genomic islands. The replication origin and the terminus of the bacterial chromosomes were determined using GC-skew between the leading and lagging strands (Freeman et al. 1998). Possible origins of genomic islands were predicted by searching through Pre_GI database (Pierneef et al. 2015).

The complete genome sequences of *B. velezensis* UCMB5140 and *B. subtilis* ssp. *subtilis* UCMB5021 were deposited at National Center for Biotechnology Information (NCBI) under accession numbers CP051463.1 and CP051466.1, respectively.

Profiling of epigenetic modifications

Tools available in the SMRT Link 6.0.0 software (https://www.pacb.com/documentation/smrt-link-software-installation-v6-0-0/) were used with an in-house Python script to generate a pipeline to perform base call kinetic analysis on the PacBio reads generated from chromosomal DNA. The Python pipeline consists of the following seven steps. (1) The complete genome consensus sequences in FASTA format, obtained by assembling PacBio reads, were indexed by the program *samtools* (part of the SMRT Link package) to be used as the reference sequence for PacBio read alignment. (2) PacBio reads were converted from the original BAX.H5 format to BAM format by the tool *bax2bam*. (3) Reads stored in BAM files were aligned against the indexed reference sequence by the tool *blasr* (Chaisson and Tesler 2012). (4) Aligned reads in BAM format were sorted by genomic locations and indexed by *samtools sort* and *index* functions. (5) Sorted and indexed BAM files were analyzed by the tool *ipdSummary* to evaluate the base call kinetics for every nucleotide in the reference genome (output file **_kinetics.csv*). Also, for every nucleotide position with a significant base call delay in multiple overlapped reads, the program calculated several statistical parameters such as the interpulse duration (IPD) ratio of the average base call time to the expectation, and the quality values (QV) scores. A QV score of 14 corresponds to *p* value 0.05 and a QV score of 21 to *p* value 0.01. The program stores all the estimated parameters together with context sequences in an output file **_basemods.gff*. (6) Contextual motifs of base modifications were searched by the tool *motifMaker*. (7) Epigenetic profiles of the studied genomes were visualized using an in-house Python script, which uses the **_kinetics.csv* and **_basemods.gff* output files.

Whole genome phylogenomic inference

Phylogenetic relations between the newly sequenced *B. velezensis* and existing reference strains with complete genome sequences available from NCBI (Schmid et al. 2018) were inferred by the supermatrix approach (Delsuc et al. 2005). In total, 24 complete genome sequences were used including the type strains *B. velezensis* FZB42, *B. amyloliquefaciens* DSM7, and *B. subtilis* 168. Orthologous genes in the selected genomes were identified with the OrthoFinder software (Emms and Kelly 2015). Amino acid sequences encoded by these genes were aligned within each cluster of orthologous genes (COG) prior to concatenating the alignments into an artificial super-alignment of 944,912 amino acids in length. The phylogenetic tree was inferred using the neighbor-joining algorithm implemented in MEGA-X (Tamura et al. 2013).

Results

In vitro biocontrol of *Aspergillus* by *Bacillus* strains

The four *Bacillus* strains studied here, i.e., UCMB5140, UCMB5044, UCMB5113, and UCMB5021, showed varying abilities to suppress in vitro the growth of the plant pathogen *Aspergillus* sp. P12 (Fig. 1). *Aspergillus* fungi are common tropical agriculture pests causing plant and crop rotting that leads to a significant production loss (Otusanya and Jeger 1996; Khokhar and Tetarwal 2012). Strain UCMB5140 showed the strongest inhibition activity on days 3 and 5 of the experiment. The percentage of growth inhibition decreased during the experiment, reaching roughly 46% on day 10. In contrast, inhibition of *Aspergillus* by UCMB5044 was delayed. It reached a peak at day 5 showing 52% inhibition, but thereafter it declined again to 35% at day 10 (Fig. 1), which might indicate a different mechanism of action. The strain UCMB5021 showed the weakest inhibitory activity

reaching maximally 40% inhibition at day 3 fast decreasing to 15% at day 10. Strain UCMB5113 showed a lower inhibition compared with strain UCMB5140 at the beginning of the experiment (day 3). However, its inhibitory activity grew stronger and at day 10 of the experiment it showed the highest suppressive activity against the fungal pathogen (roughly 48%, Fig. 1).

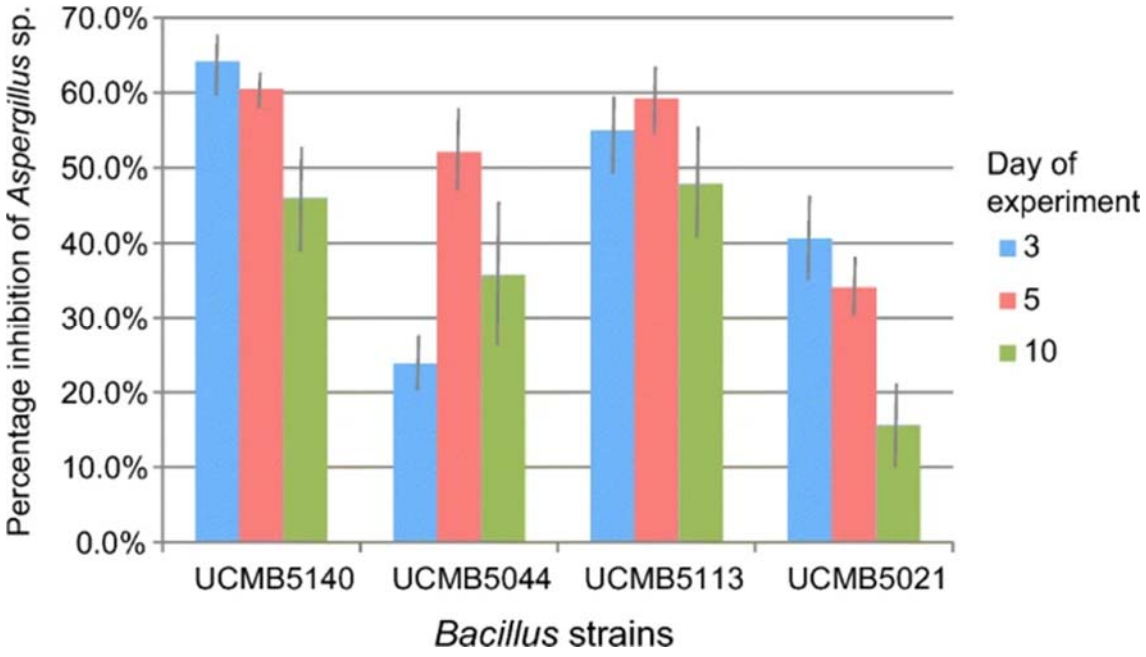

Fig. 1. Summary of the inhibition of growth of *Aspergillus* sp. P12 by *Bacillus* strains after 3, 5, and 10 days of cultivation in in vitro plate assays. Vertical bars depict the SD of inhibition values (based on 3 replicates). Colors represent different days of the experiment (day 3, blue; day 5, red; day 10, green)

In vivo plant protection against *R. solanacearum* infection by treatment with *Bacillus* strains

Next, we tested the potential of the four *Bacillus* strains in terms of the utility as biocontrol agents in in vivo assays against *R. solanacearum* that is a common cause of devastating outbreaks of bacterial wilt disease in tropical agriculture (Genin and Denny 2012). The *Bacillus* sp. mediated protection of tomato seedlings against tomato wilt disease caused by *R. solanacearum* infection was measured in both treated and control plants by calculating the respective AUDPC values (see "Materials and methods" section) (Fig. 2). Higher AUDPC values correspond to a stronger damage of the plants by the pathogen while lower AUDPC values compared with those in the positive control indicated a protective effect of the treatment with the four *Bacillus* strains. While all four *Bacillus* strains exerted some protection, strain UCMB5140 showed the strongest plant protective efficacy in vivo (Fig. 2), reducing the AUDCP to about one fifth of that in the infected control plants that were not treated with *Bacillus* strains. It was followed by the strains UCMB5021, UCMB5113, and UCMB5044 reducing the AUDCP to about half of the infected control plants.

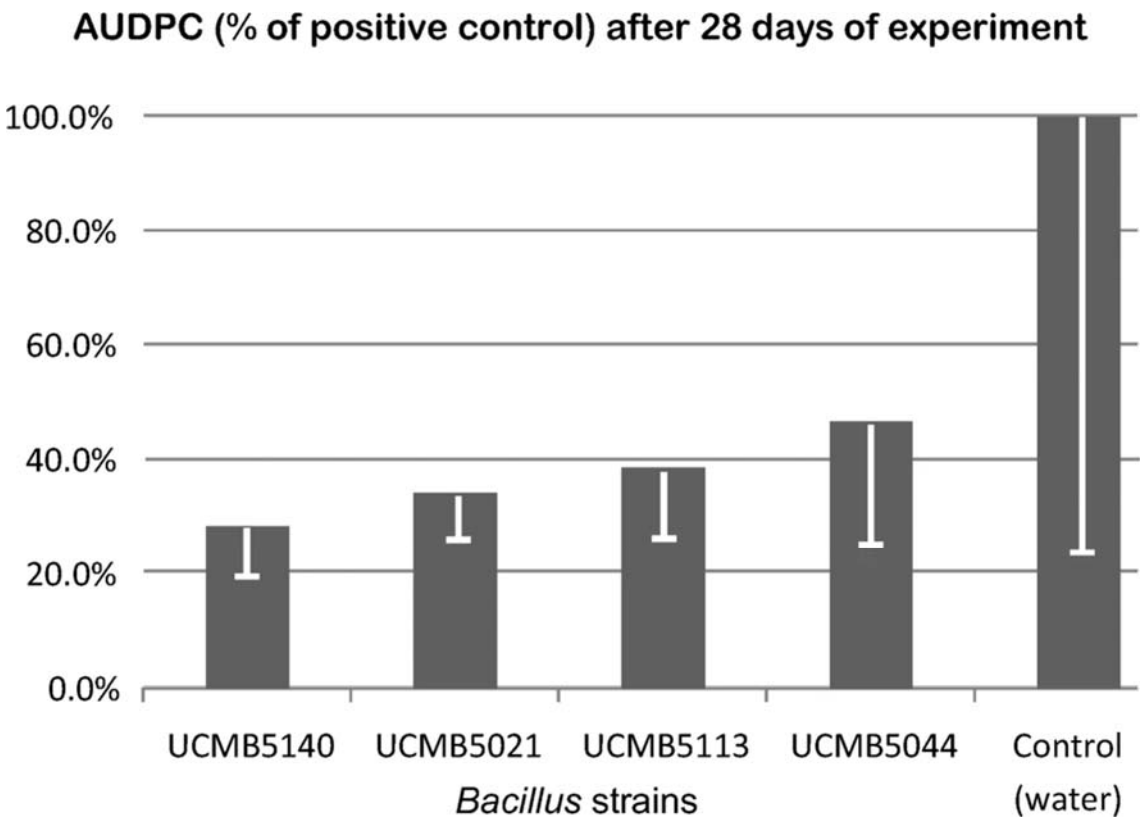

Fig. 2. Extent of wilting symptoms in infected tomato plants treated with *Bacillus* spores and with water treatment as a positive control of the disease development in infected but untreated plants. The columns represent the average AUDCP values and the vertical bars depict the SD (calculated from 3 replicates). Lower AUDCP values indicate a better protection of the plants by treatment with *Bacillus* strains

Biosynthesis of polyketides and lipopeptides

The production of polypeptide antibiotics by selected strains grown on Landy liquid medium after 24 and 48 h of cultivation was analyzed by MALDI-TOF and HPLC. MALDI-TOF analysis revealed the highest diversity of such products for UCMB5113 comprising surfactin, fengycin, bacillibactin, bacillaene, difficidin, and macrolactin D. UCMB5140 produced a detectable amount of surfactin, fengycin, and bacillibactin. Only bacillibactin was detected by this method in the culture medium of UCMB5021 and no significant lipopeptide or polyketide peaks were found in UCMB5044 culture medium (Reva et al. 2019). Results of the HPLC analysis generally confirmed the biosynthetic potential of the selected strains. Table 1 and Supplemental Fig. S1 and S2 show the distribution of characteristic peaks of different lipopeptides and polyketides revealed by HPLC after 24 and 48 h of cultivation. Strain UCMB5044 was not included in Table 1 because no peaks were found at all on this medium at any time of cultivation for this strain. According to the HPLC data, strain UCMB5140 produced bacillaene, bacillomycin, difficidin, and macrolactin D. For some unknown reason, HPLC analysis did not reveal the production of fengycin for this strain, although this product was clearly detected by MALDI-TOF MS. Bacillaene, fengycin, and macrolactin D were produced by UCMB5021 at 48 h of cultivation.

Table 1. Characteristic peak areas retrieved from HPLC chromatograms in culture media after 24 and 48 h of cultivation

Strain	*T*	Products and their column retention time (RT) in minutes											
		mln	bmy/itu	fen	fen	mln	dfn	bae					
		6.29	6.89	7.1	7.31	7.55	8.76	6.52	6.59	6.75	6.84	7.07	7.7
UCMB5140	24 h	×	×				×	×		×		×	
	48 h	×	×				×			×			×
UCMB5113	24 h	×	×	×		×	×						×
	48 h	×		×		×	×	×	×	×	×		×
UCMB5021	24 h											×	
	48 h				×	×					×	×	×

T time of cultivation, *bae* bacillaene, *bmy* bacillomycin D, *dfn* difficidin, *fen* fengycin, *itu* iturin A, *mln* macrolactin D

Complete genome sequencing and assembly

The complete genome sequence of strain UCMB5140 was obtained by assembling long reads generated by PacBio SMRT RSII technology. High quality of base calling was achieved by reading around circular SMRTbell templates. For a de novo genome assembly, 13,350 DNA reads in fastq format were used from the initial PacBio SMRT subreads with an average length of fastq reads of 3480 bp, ranging from 914 to 18,400 bp, with an N50 of 2570 bp. The reads were assembled into a single contig, a 3,980,571-bp circular chromosome with an average GC content of 46.52%. Estimated average depth of coverage was ×11. It should be noted that each fastq read was a consensus of approximately 30 initial PacBio SMRT subreads generated for the same template DNA in the sequencing library, indicating that the actual coverage was much higher. Further alignment of the initial PacBio SMRT subreads against the assembled complete genome sequence, which was performed to profile for epigenetic modifications in the bacterial chromosome, showed an average depth of coverage of 350-fold. The complete genome sequence and raw read data were deposited at the NCBI under accession number CP051463.1. Similarly, the complete genome sequence of the reference strain UCMB5021 (deposited at the NCBI under accession number CP051466.1) was also generated by combining PacBio and Illumina reads. The length of the circular chromosome is 4,060,035 bp, with a GC content of 43.85%.

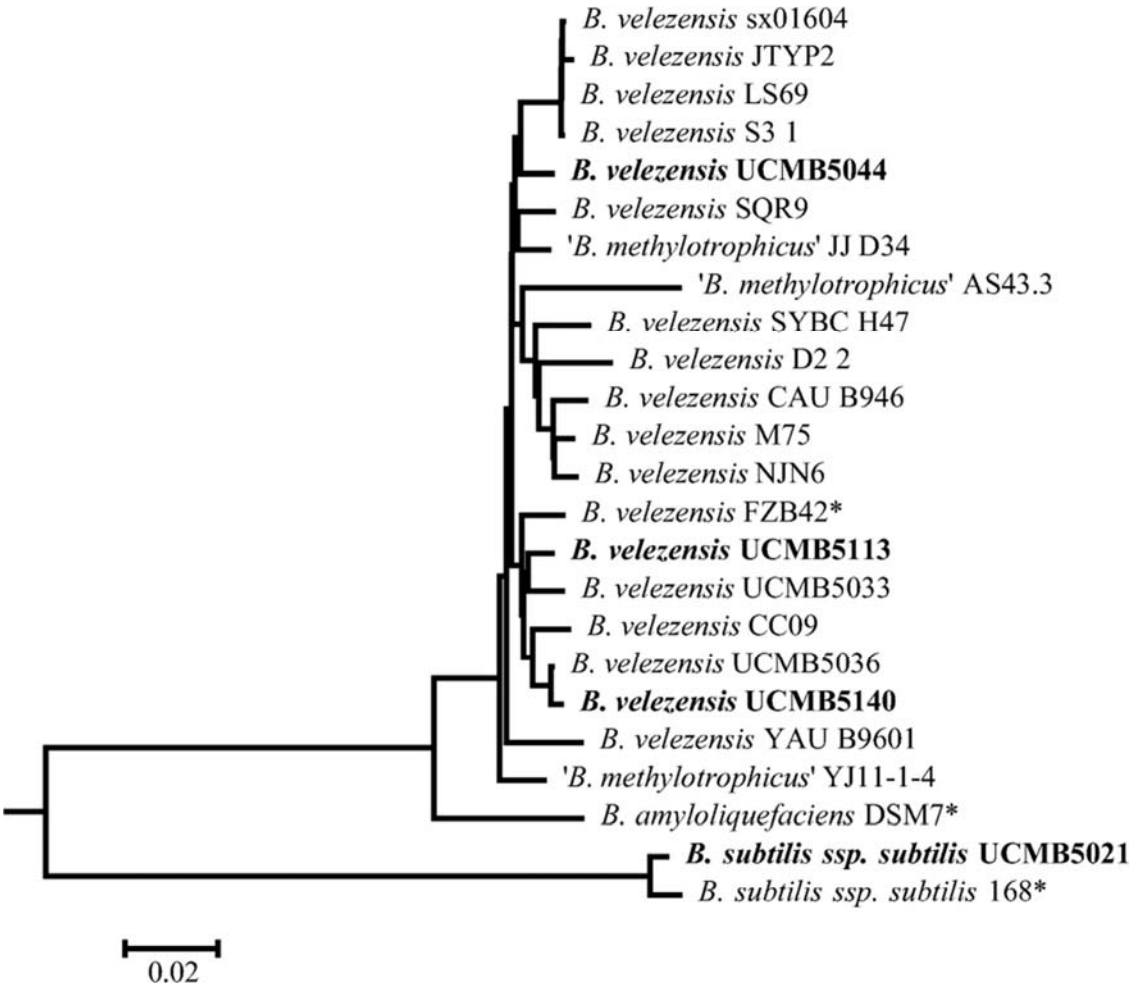

Fig. 3. NJ super-alignment-based phylogenetic tree of the genomes of selected bacterial strains belonging to the species *B. velezensis*, *B. amyloliquefaciens*, and *B. subtilis* ssp. *subtilis*. The strains analyzed in the present study are shown in bold. The species type strains are marked by asterisks

Phylogenomic and pangenomic comparison of the sequenced genomes

In total, 3657 protein coding genes, 9 alleles of ribosomal RNA operons, and 85 tRNAs were identified in the genome of UCMB5140. Comparison of the complete genome sequences of strains UCMB5140, UCMB5044, UCMB5021, and UCMB5113 with the genomes of 20 reference organisms obtained from the NCBI identified a core genome comprising 3216 orthologous protein coding genes shared by all four strains. Concatenated alignment of translated protein sequences was used for a neighbor-joining (NJ) phylogenetic tree inference (Fig. 3). The phylogenetic analysis revealed that UCMB5140 belongs to the species *B. velezensis* and is closely related to the previously sequenced strain *B. velezensis* UCMB5036 (Manzoor et al. 2013). The genome sequence of UCMB5140 was a bit longer (3,980,571 bp vs. 3,910,324 bp) due to the presence of two additional prophages and a specific NRPS operon. Both these strains grouped together with strain UCMB5113 and the reference strain *B. velezensis* FZB42 (Fig. 3). The strain UCMB5044 fell into a separate group of the *B. velezensis* isolates. In contrast, strain UCMB5021 is closely related to the type strain *B. subtilis* ssp. *subtilis* 168.

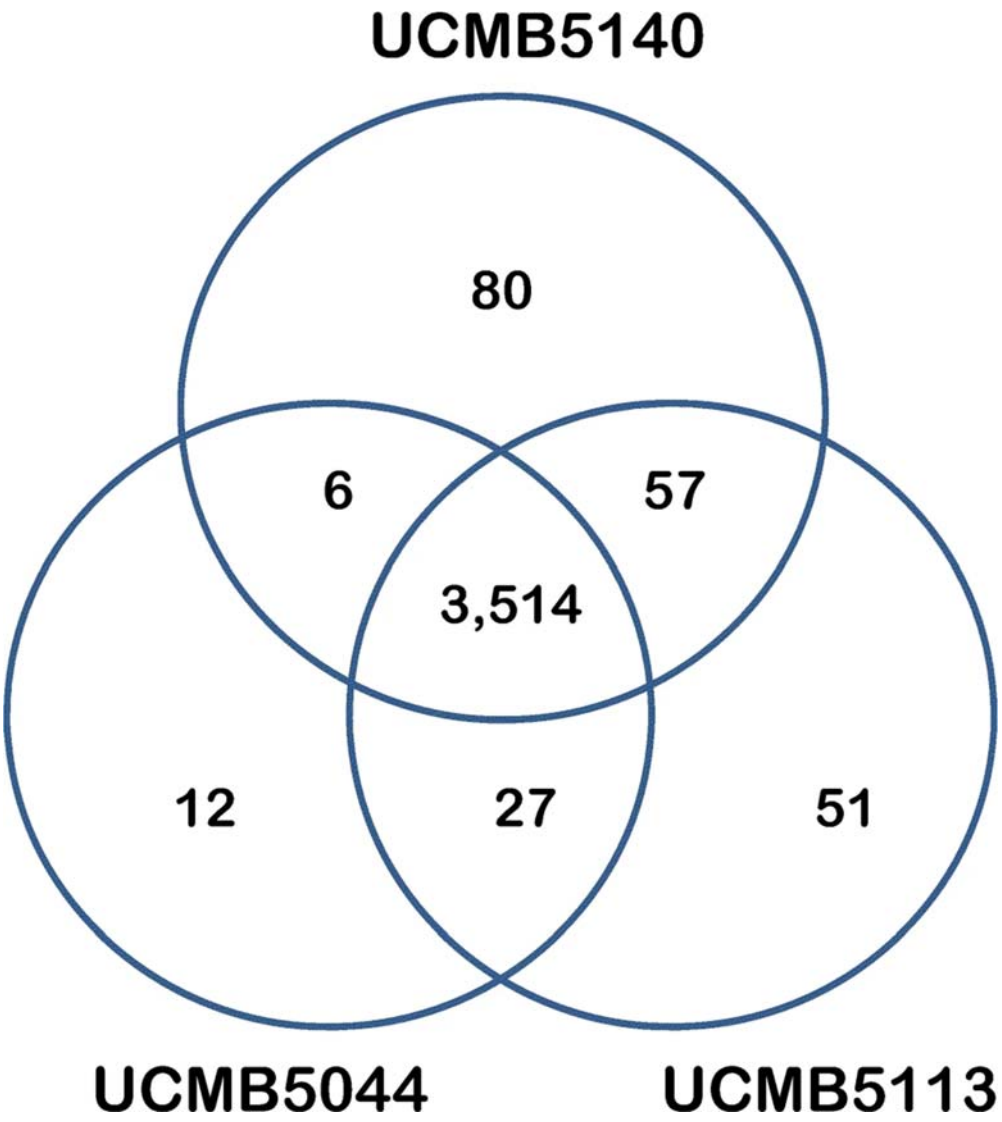

Fig. 4. Venn diagram illustrating the overlap of shared (core) and strain-specific (accessory) protein coding genes in the genomes of three selected *B. velezensis* strain (UCMB5140, UCMB5144, and UCMB5113). Strain UCMB5140 harbors the largest number of strain-specific accessory genes (80)

The *B. velezensis* genomes showed a high level of conservation both in terms of the core genome sequence and gene synteny. The three *B. velezensis* strains used in the present study (excluding *B. subtilis* UCMB5021) share 3514 orthologous genes accounting for 89% of their pangenome (Fig. 4). Their core genome includes nine operons for biosynthesis of the following lipopeptides, polyketides, and aminoglycosidic antibiotics: bacillomycin, fengycin, surfactin, bacillibactin, bacillaene, difficidin, macrolactin, bacilysin, and butirosin-like aminoglycoside.

Genome comparison identified 80 strain-specific genes in UCMB5140, i.e., genes absent in *B. velezensis* strains UCMB5044 and UCMB5113. In terms of genes encoding potential biocontrol activity, the most interesting finding was an operon of six genes, *grmABCDEF,* harbored on one genomic island. The first two genes of this operon were predicted to encode a NRPS complex. The four other genes were annotated as an ATP-dependent ABC transporter, a reductoisomerase/saccharopine dehydrogenase, an unknown conserved protein, and a NRPS-related thioesterase. Homologous genomic islands were found in several other sequenced *B. velezensis* strains, for example, in YAU B9601 and CC09 shown in the phylogenetic tree in Fig. 3, and also in an Antarctic seawater unidentified *Bacillus* sp. Pc3 isolate which was characterized by an outstanding antagonistic activity (Guo et al. 2015), but not in the phylogenetically closest strain UCMB5036. The low GC content of the genomic island harboring this NRPS operon and a specific tetranucleotide pattern identified through the Pre_GI database (Pierneef et al. 2015) suggest its possible recent acquisition from environmental *Lactobacillus*. This NRPS gene cluster encodes a heptapeptide containing two cysteine residues (Fig. 5). The presence of such cysteine residues is characteristic for several antifungal and antibacterial polypeptides containing cysteine-knot structural motifs (Iyer and Acharya 2011; Pushpanathan et al. 2013; Zhao et al. 2018).

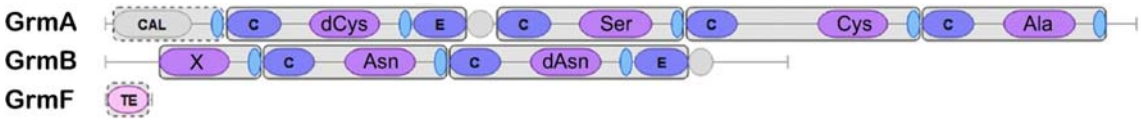

Fig. 5. Schematic distribution of amino acid specific adenylation domains (shown as pink ellipses) in gray blocks representing three NRPS subunits that are unique to strain UCMB5140, GrmA, GrmB, and GrmF. In addition, the names of target amino acids or an X for unidentified amino acids are shown. Other functional protein domains commonly found in NRPS (McErlean et al. 2019) are indicated as CAL for Coenzyme A ligase domain; C, condensation domain; E, epimerization domain; and TE, thioesterase domain

Another protein-coding gene specific for strain UCMB5140, which can also be found in several other *B. velezensis* genomes (UCMB5036, for example), is a large cell wall–associated RHS repeat protein, which may play a role in bacterial cell adhesion to roots, seeds, or other surfaces (Martinez-Garcia et al. 2020). This gene also was associated with a small genomic island of unknown origin. Other strain-specific genes of UCMB5140 encode hypothetical, unknown, and uncharacterized phage-related proteins located in overall 15 horizontally acquired genomic islands shown in Fig. 6a.

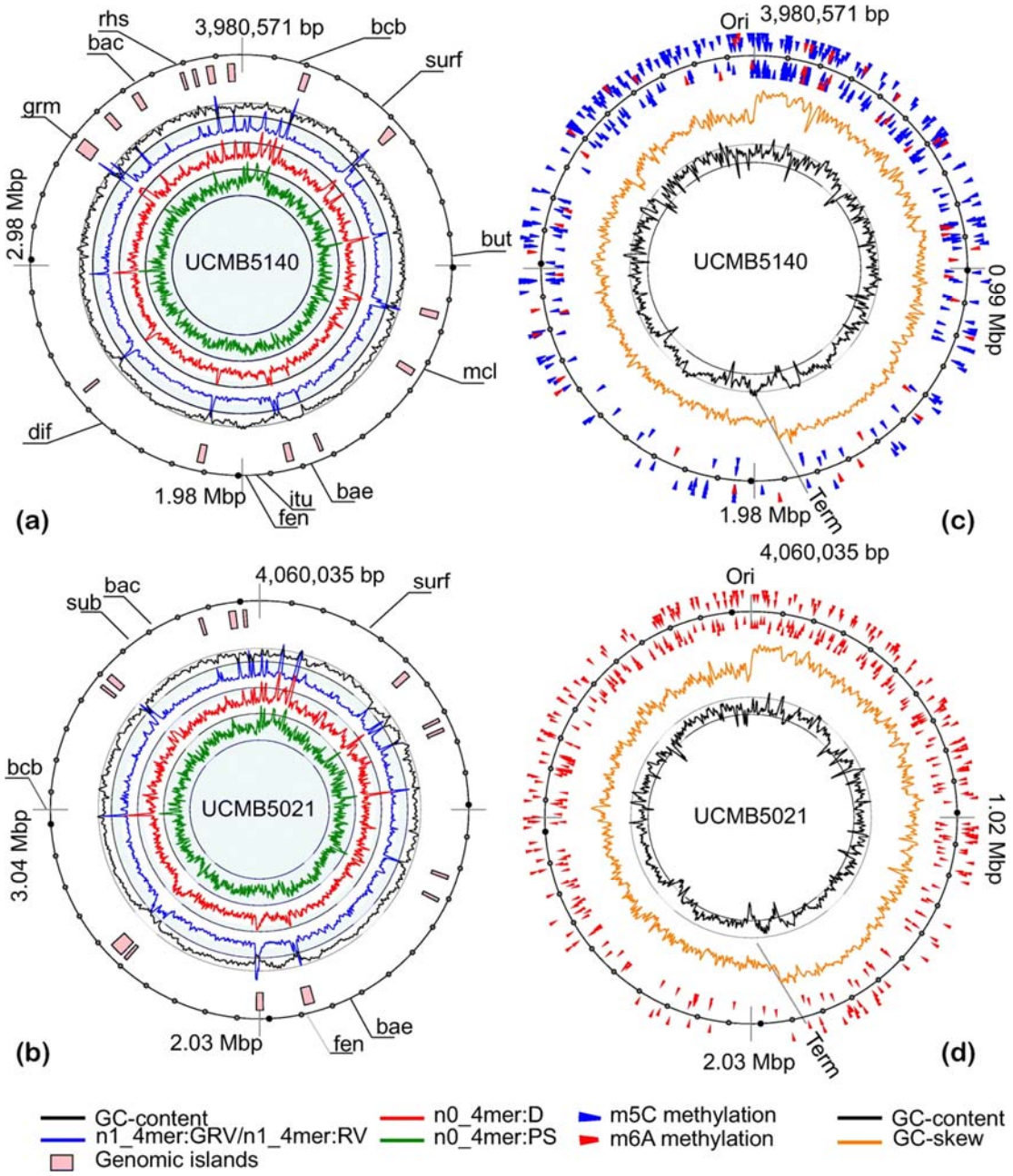

Fig. 6. Maps of the circular chromosomes of strains *B. velezensis* UCMB5140 and *B. subtilis* UCMB5021. **a**, **b** Distribution of horizontally transferred genomic islands as predicted by SeqWord Genomic Island Sniffer (pink blocks) in strains UCMB5140 and UCMB5021, respectively. Histograms of the GC content, local versus global tetranucleotide distribution variance (n1_4mer:RV/n1_4mer:GRV), tetranucleotide pattern deviations (n0_4mer:D), and tetranucleotide pattern skew (n0_4mer:PS) were calculated in 8-kbp sliding windows with steps of 2 kbp. The locations of biosynthetic operons encoding bacillaene (bae), bacillibactin (bcb), bacilysin (bac), difficidin (dif), iturin A (itu), fengycin (fen), macrolactin D (mcl), subtilosin A (sub), surfactin (surf), butirosin-like aminoglycoside (but), gramicidin-related NRPS (grm), and a cell wall–associated RHS repeat protein (rhs) are depicted. **b**, **c** The location of methylated adenine and cytosine residues are indicated respectively by red and blue triangle marks depicting modified nucleotides at the leading and lagging DNA strands as indicated by the location of the marks either externally or internally along the cycle line representing the chromosome. Round notches on the line scale 100 kbp chromosomal segments. Histograms of the GC content and GC skew were calculated in 8-kbp sliding with steps of 2 kbp. The replication origin and terminus are marked as Ori and Term, respectively

The genome of strain *B. subtilis* UCMB5021 shows a smaller number of genes encoding polypeptide antibiotics (Fig. 6b). Particularly, there are no genes encoding the biosynthesis of difficidin, macrolactin, and bacillomycin/iturin, all representing broad-spectrum antifungal and antibacterial antibiotics (Koumoutsi et al. 2004; Schneider et al. 2007; Chen et al. 2006, 2009b). This may explain the reduced antifungal activity of this strain (Fig. 1). Nevertheless, UCMB5021 showed a promising ability to control bacterial wilt caused by *R. solanacearum* (Fig. 2) indicating other mechanisms of improvement of plant resistance by PGPR *B. subtilis*, which compensate the lack of the *B. velezensis*–specific antibiotics. Particularly, UCMB5021 putatively is able to synthesize the antibiotic subtilosin A (Babasaki et al. 1985) as it possesses the corresponding gene cluster on the chromosome (Fig. 6b).

Comparison of patterns of DNA methylation in the sequenced genomes

All three *B. velezensis* strains, UCMB5140, UCMB5044, and UCMB5113, and *B. subtilis* strain UCMB5021 are harboring a unique set of DNA restriction-modification systems. A large methyltransferase operon *hsdMSR* was only present in strain UCMB5113. In contrast, a pair of restriction-modification genes, *bamHI* and *bamHIM*, was only found in UCMB5044. Finally, a class I SAM-dependent methyltransferase was specific for strain UCMB5140. These genes were located outside of the predicted genomic islands; in UCMB5044 and UCMB5113, they are flanked by transposases implying a horizontal gene acquisition of these genes.

Methyltransferases, either belonging to DNA restriction-modification systems or stand-alone proteins, perform methylation of cytosine or adenine residues at specific recognition sites forming strain-specific patterns of epigenetic modifications. For example, cytosine residues of all palindromic sequences G*G*ATCC (the underscored nucleotide is modified and cursive letter marks a nucleotide opposing the modified one on the complement DNA strand) are methylated in UCMB5044 at both DNA strands to protect the chromosomal DNA from cleavage by the *Bam*HI restriction enzyme (Reva et al. 2019). Methylation of adenine residues in this genome was less frequent and distributed sporadically, i.e., there was no association of A-methylated sites with any DNA motifs. In contrast, it was shown previously that the methylation of adenine residues is the most abundant type of epigenetic modifications in strain At1, which is closely related to UCMB5113 (Reva et al. 2019). Often this methylation occurred on both DNA strands with seven non-conserved nucleotides between the modified sites at non-palindromic motifs GYTANNVNNNN*T*GCB and GCADNNNNNN*T*ARC. In the newly sequenced genome of UCMB5140, cytosine methylation was the most abundant type of DNA modification, while adenine methylation was less frequent and sporadic (Fig. 6c), and thus resembled the methylation pattern of UCMB5044. However, in contrast to that, the cytosine methylation in UCMB5140 takes place at an alternative recognition site, GGANNNNAGCTACYCAYN-NC. While there is a palindromic sequence at the central part of the motif, ACGT, which corresponds to the cleavage site of *Alu*I restriction enzymes, the methylation occurs only at one DNA strand due to the non-palindromic character of the upstream and downstream sequences of the recognized motif. In the strain *B. subtilis* UCMB5021, adenine residues were methylated predominantly in the chromosome (Fig. 6d). It was unexpected because two subunits *bsuMA* and *bsuMB* encoding a *Bsu*M DNA-cytosine methyltransferase were found on the chromosome. However, methylation of cytosine residues on the chromosome was rare and sporadic. Another methylase YeeA was found in a genomic island, which potentially may be behind the observed intensive adenine methylation at the motif GACGAG.

The current study showed that the global DNA methylation patterns are strain specific in bacteria of the *B. subtilis* group. The distribution of methylated cytosine residues with QV scores above 50 in the genome of UCMB5140 is shown in Fig. 6c. Figure 6d shows the distribution of methylated sites with QV scores above 320 in the genome UCMB5021. It should be noted that QV scores depend on the coverage that reached much higher values in UCMB5021. The score thresholds were set to select top

25% of methylated nucleotides in both genomes. The frequency of methylated sites decreased around the replication terminus of the chromosome. It may be an artifact of the used methylation detection algorithm as the local coverage values are higher around the replication origin as this chromosomal region is semi-diploidic in bacterial cells due to the continuing DNA replication process starting from the chromosomal origin of replication (Marczynski and Shapiro 1993). Biased coverage of randomly generated reads leads to a decrease in QV values of modified base predictions around the replication terminus, as it is seen in Fig. 6c and d.

Discussion

Complete genome sequencing and phylogenomic inference placed strain UCMB5140 together with other *Bacillus velezensis* strains. Isolates of this species are frequently used as biocontrol agents due to their outstanding antagonistic activities against phytopathogens and the capacity to promote seedling development and plant growth (Aloo et al. 2019; Rabbee et al. 2019). Bioassays conducted in this research have confirmed the high biocontrol potential of strain UCMB5140, which outperformed the other *B. velezensis* and *B. subtilis* strains used in this study, and protected model plants against phytopathogens. Plant protection and growth promotion are often associated with production of nutrient-mobilizing enzymes, hormone-like compounds, polypeptide antibiotics, and lantibiotics (Idriss et al. 2002; Koumoutsi et al. 2004; Schneider et al. 2007; Chen et al. 2009a, b; Liu et al. 2013; Wu et al. 2014, 2015; Gu et al. 2017; Lu et al. 2018; Kim et al. 2019; Rabbee et al. 2019; De Vrieze et al. 2020). The released compounds induce a systemic resistance response in plants against pathogens and improve the abiotic stress management (Ryu et al. 2004; Li et al. 2015; Wu et al. 2018). For example, profound systemic alterations of plant metabolism leading to an enhanced drought tolerance were reported in experiments on chickpea (*Cicer arietinum* L.) treated with PGPR *Bacillus* and exposed to experimental drought conditions (Khan et al. 2019). Multiple polypeptide antibiotics synthesized by PGPR *Bacillus* suppress the growth of bacterial and fungal pathogens (Yuan et al. 2012; Raza et al. 2016). It may be assumed that the various PGPR bioactivities are determined by the presence of specific genes in bacterial genomes and/or that epigenetic modifications may also play a role.

The *B. velezensis* isolates demonstrated a significant level of genome conservation despite variable bioactivities. The genomic architecture only differed in mobile genetic island insertions, mostly prophages. The annotation of genes on the 15 genomic islands specifically found in strain UCMB5140 revealed only a few genes with possible biocontrol relevant functions. Of particular interest were a horizontally acquired NRPS operon encoding an unknown polypeptide antibiotic; a large cell-wall protein, which may contribute to cell adhesion; and a class I SAM-dependent methyltransferase.

B. velezensis strains are equipped with numerous NRPS and PKS genetic complexes. As a substantial fraction of NRPS genes can be missed in incomplete short read–based genome assemblies (Varadarajan et al. 2020), our strategy was to first generate complete genome sequences using long read sequences before subsequently carrying out comparative genomics analyses. Nine large gene clusters were found encoding the biosynthesis of bacillomycin/iturin, fengycin/plipastatin, surfactin, bacillibactin, bacillaene, difficidin, macrolactin, and bacilysin, which are present in the majority of sequenced *B. velezensis* genomes. However, reshuffling and replacement of operational adenylation domains results in the production of individual variants of polypeptides by each *Bacillus* strain. For example, a homologous NRPS operon present in the genomes of *B. velezensis* strains UCMB5113, UCMB5044, and UCMB5140 encodes synthesis of bacillomycin D in the former two strains, and iturin A in the latter one (Supplemental Fig. S3). Both antibiotics are heptapeptides but with a different order of the amino acid residues: Asn-dTyr-dAsn-Ser-Glu-dSer-Thr (bacillomycin D) and Asn-dTyr-dAsn-Gln-Pro-dAsn-Ser (iturin A) (Gong et al. 2015). This operon was not found in *B. subtilis*

UCMB5021 (see Fig. 6a, b). Various antifungal iturin-like lipopeptides produced by different *B. velezensis* strains irrespective of their phylogenetic relations are encoded by a gene operon showing conserved gene organization and chromosomal location. This observation suggests a horizontal exchange of functional modules within NRPS genes or entire operons possibly by the homologous recombination mechanism (Dunlap et al. 2019).

The biosynthetic capacity of PGPR *Bacillus* may also depend on a conditional gene regulation at the transcriptional and post-transcriptional levels. For example, no polypeptide secondary metabolites were found in the culture medium of strain UCMB5044. Our previous study had shown that the NRPS operons are transcribed in this strain, but that the translation from mRNA was probably blocked post-transcriptionally (Reva et al. 2019). UCMB5044 is an oligotrophic microorganism which grows better on minimal media. Biosynthesis of secondary metabolites by this strain is restricted; however, this strain shows a significant ability to control phytopathogens implying that synthesis of polypeptide antibiotics may be enabled only when specific needs or responses were required. In contrast, strain UCMB5113 overproduced all secondary metabolites at the expense of a surcharge of energy and nutrient consumption (Reva et al. 2019). UCMB5140 shows an intermediate state between the oligotrophic and copiotrophic lifestyles. However, a strong inhibition of pathogens identified in vitro on Petri dishes not always guarantees a high in vivo activity. For example, in a recent publication, the *B. velezensis* strain UCMB5007 was introduced, which showed weak PGPR activities despite the intensive inhibition of phytopathogens in vitro (Reva et al. 2019). It was suggested that this strain was not a rhizosphere inhabitant. Therefore, it cannot sense the environmental signals mediated by root exudates and respond to these signals appropriately as PGPR *Bacillus* strains do. Strain *B. velezensis* UCMB5007, which is a quite active intestinal probiotic microorganism, may serve as a negative control for non-PGPR *Bacillus*.

Alternative methylation of nucleotides in gene regulatory elements might play a role in silencing or re-activating various metabolic and biosynthetic activities (Beaulaurier et al. 2019). This factor of bacterial evolution has been generally overlooked until recently due to technical difficulties in identification of methylated nucleotides in bacterial chromosomes. Novel long-read sequencing by PacBio's SMRT technology streamlined the detection of epigenetically modified nucleotides and genome-scale comparison of DNA methylation patterns (Rhoads and Au 2015). Genome comparisons performed in this and in a recent study (Reva et al. 2019) showed that the DNA restriction-modification systems and individual methyltransferases are frequently involved in gene gain-and-loss events in *B. velezensis*. All three sequenced *B. velezensis* strains were found to possess strain-specific methyltransferases, which can render the genomes with strain-specific patterns of epigenetic modifications of the whole chromosomal DNA. Methylated nucleotides may interfere with gene regulation and transcription by either blocking or stimulating these processes (Sánchez-Romero and Casadesús 2020). Replacement of one methyltransferase with another one might immediately affect the global pattern of DNA methylation unavoidably leading to general alternations in the expression of numerous genes. Further studies are needed to investigate whether single nucleotide polymorphisms in coding and non-coding parts of the chromosome may introduce new methylation sites or prevent methylation that may be a way of the micro-evolutionary adaptation of bacteria to new habitats.

In conclusion, this work has identified critical genetic determinants of PGPR activities through whole genome comparative study that involved several pre-selected reference strains of the *B. velezensis* and *B. subtilis* species, which demonstrate promising biocontrol potential. The application of a third-generation sequencing technology, PacBio SMRT, has provided additional information to overcome the limitations of genome sequencing and assembly based on short reads (Ferrarini et al. 2013; Teng et al. 2017; Varadarajan et al. 2020) and allowed profiling of methylation of chromosomal DNA. Several factors that affect the evolution of active PGPR *B. velezensis* strains were identified: (1)

horizontal acquisition of novel non-ribosomal peptide synthetases (NRPS) and adhesion genes; (2) rearrangements of functional modules of NRPS genes leading to strain specific combinations of their encoded products; (3) gain and loss of methyltransferases that can cause global alterations in DNA methylation patterns, which eventually may regulate gene transcription. The positive effects of PGPR *Bacillus* strains on plants are manifold and include a direct suppression of phytopathogens by synthesized antibacterial compounds; induction of plant systemic resistance; solubilization of soil nutrients, e.g., phosphorus and ammonium; water retention by bacterial polysaccharides; and synthesis of small organic compounds acting as plant hormones and plant gene transcription regulators (Khan et al. 2019). Each individual PGPR provides subtle difference in terms of their PCPR activities. Future design of effective biopesticides should therefore be based on combinations of PGPR strains supplementing each other with plant-protective and plant growth–promoting activities.

Acknowledgments

We gratefully acknowledge K. Dietel (ABiTEP GmbH, Berlin, Germany) for preparing samples for mass spectroscopic measurements, and Daniel Frei and Juerg E. Frey (Agroscope, Switzerland) for the Illumina MiSeq sequencing.

Funding

O.N.R. was funded by the South African National Research Foundation (NRF) grants 93134 and 93664; O.N.R., Sy.L., and D.T. received funds from the joint NRF/COSTECH (Tanzanian Commission for Science and Technology) grant 86905 and from joint TIA (Technology Innovation Agency of South Africa)/COSTECH grant TIA 2018-FUN-0050. A.D.M. acquired PhD and MSc student fellowship grants from Southern African Biochemistry and Informatics for Natural Products (SABINA, http://www.sabina-africa.org/). C.H.A. acknowledges support from Agroscope through its research program Microbial Biodiversity.

Contributions

O.N.R.—DNA and RNA sequencing, genome assembly, annotation, epigenetic profiling, funding acquisition, student supervision; S.A.L.—UCMB5140 isolation and characterization; A.D.M.—PGPR assays; D.M.—PGPR assays, student supervision; Sy.L.—PGPR assays, student supervision; W.Y.C.—mass spectrometry studies; St.L.—PacBio and Illumina MiSeq sequencing, genome assembly; C.H.A.—funding acquisition, PacBio and Illumina MiSeq sequencing, genome assembly; J.V.—mass spectroscopic measurements; R.B.—mass spectrometry and study supervision. All co-authors took part in the manuscript writing and edition.

References

Abd El-Daim IA, Bejai S, Meijer J (2019) *Bacillus velezensis* 5113 induced metabolic and molecular reprogramming during abiotic stress tolerance in wheat. Sci Rep 9:16282. https://doi.org/10.1038/s41598-019-52567-x

Aloo BN, Makumba BA, Mbega ER (2019) The potential of bacilli rhizobacteria for sustainable crop production and environmental sustainability. Microbiol Res 219:26–39. https://doi.org/10.1016/j.micres.2018.10.011

Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA (2016) plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics 32:3380–3387. https://doi.org/10.1093/bioinformatics/btw493

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F (2008) The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75

Babasaki K, Takao T, Shimonishi Y, Kurahashi K (1985) Subtilosin A, a new antibiotic peptide produced by *Bacillus subtilis* 168: isolation, structural analysis, and biogenesis. J Biochem 98:585–603. https://doi.org/10.1093/oxfordjournals.jbchem.a135315

Beaulaurier J, Schadt EE, Fang G (2019) Deciphering bacterial epigenomes using modern sequencing technologies. Nat Rev Genet 20:157–172. https://doi.org/10.1038/s41576-018-0081-3

Bezuidt O, Lima-Mendez G, Reva ON (2009) SEQWord Gene Island sniffer: a program to study the lateral genetic exchange among bacteria. World Acad Sci Eng Technol 58:1169–1174

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310

Caulier S, Nannan C, Gillis A, Licciardi F, Bragard C, Mahillon J (2019) Overview of the antimicrobial compounds produced by members of the *Bacillus subtilis* group. Front Microbiol 10:302. https://doi.org/10.3389/fmicb.2019.00302

Chaisson MJ, Tesler G (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics 13:238. https://doi.org/10.1186/1471-2105-13-238

Chen XH, Koumoutsi A, Scholz R, Schneider K, Vater J, Süssmuth R, Piel J, Borriss R (2006) Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. J Bacteriol 188:4024–4036. https://doi.org/10.1128/JB.00052-06

Chen XH, Koumoutsi A, Scholz R, Schneider K, Vater J, Süssmuth R, Piel J, Borriss R (2009a) Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. J Biotechnol 140:27–37. https://doi.org/10.1016/j.jbiotec.2008.10.011

Chen XH, Scholz R, Borriss M, Junge H, Mögel G, Kunz S, Borriss R (2009b) Difficidin and bacilysin produced by plant associated *Bacillus amyloliquefaciens* are efficient in controlling fire blight disease. J Biotechnol 140:38–44. https://doi.org/10.1016/j.jbiotec.2008.10.015

Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147

De Vrieze M, Varadarajan AR, Schneeberger K, Bailly A, Rohr RP, Ahrens CH, Weisskopf L (2020) Linking comparative genomics of nine potato-associated *Pseudomonas* isolates with their differing biocontrol potential against late blight. Front Microbiol 11:857. https://doi.org/10.3389/fmicb.2020.00857

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 6:361–375. https://doi.org/10.1038/nrg1603

Dunlap CA, Bowman MJ, Rooney AP (2019) Iturinic lipopeptide diversity in the *Bacillus subtilis* species group—important antifungals for plant disease biocontrol applications. Front Microbiol 10:1794. https://doi.org/10.3389/fmicb.2019.01794

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157. https://doi.org/10.1186/s13059-015-0721-2

Fan B, Wang C, Ding X, Zhu B, Song X, Borriss R (2019, 2019) AmyloWiki: an integrated database for *Bacillus velezensis* FZB42, the model strain for plant growth-promoting bacilli. Database (Oxford):baz071. https://doi.org/10.1093/database/baz071

Ferrarini M, Moretto M, Ward JA, Šurbanovski N, Stevanović V, Giongo L, Viola R, Cavalieri D, Velasco R, Cestaro A, Sargent DJ (2013) An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. BMC Genomics 14:670. https://doi.org/10.1186/1471-2164-14-670

Freeman JM, Plasterer TN, Smith TF, Mohr SC (1998) Patterns of genome organization in bacteria. Science 279:1827–11827. https://doi.org/10.1126/science.279.5358.1827a

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907

Genin S, Denny TP (2012) Pathogenomics of the *Ralstonia solanacearum* species complex. Annu Rev Phytopathol 50:67–89. https://doi.org/10.1146/annurev-phyto-081211-173000

Gong AD, Li HP, Yuan QS, Song XS, Yao W, He WJ, Zhang JB, Liao YC (2015) Antagonistic mechanism of iturin A and plipastatin A from *Bacillus amyloliquefaciens* S76-3 from wheat spikes against *Fusarium graminearum*. PLoS One 10:e0116871. https://doi.org/10.1371/journal.pone.0116871

Gu Q, Yang Y, Yuan Q, Shi G, Wu L, Lou Z, Huo R, Wu H, Borriss R, Gao X (2017) Bacillomycin D produced by *Bacillus amyloliquefaciens* is involved in the antagonistic interaction with the plant pathogenic fungus *Fusarium graminearum*. Appl Environ Microbiol 83:e01075–e01017. https://doi.org/10.1128/AEM.01075-17

Guo W, Cui P, Chen X (2015) Complete genome of *Bacillus* sp. Pc3 isolated from the Antarctic seawater with antimicrobial activity. Mar Genomics 20:1–2. https://doi.org/10.1016/j.margen.2015.01.004

Gupta CD, Dubey RC, Kang SC, Maheshwari DK (2001) Antibiosis-mediated necrotrophic effect of *Pseudomonas* GRC2 against two fungal plant pathogens. Curr Sci 81:91–94

Hashem A, Tabassum B, Fathi Abd Allah E (2019) *Bacillus subtilis*: a plant-growth promoting rhizobacterium that also impacts biotic stress. Saudi J Biol Sci 26:1291–1297. https://doi.org/10.1016/j.sjbs.2019.05.004

Idriss EE, Makarewicz O, Farouk A, Rosner K, Greiner R, Bochow H, Richter T, Borriss R (2002) Extracellular phytase activity of *Bacillus amyloliquefaciens* FZB45 contributes to its plant-growth-promoting effect. Microbiol 148(Pt 7):2097–2109. https://doi.org/10.1099/00221287-148-7-2097

Iyer S, Acharya KR (2011) Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. FEBS J 278:4304–4322. https://doi.org/10.1111/j.1742-4658.2011.08350.x

Khan N, Bano A, Rahman MA, Guo J, Kang Z, Babar MA (2019) Comparative physiological and metabolic analysis reveals a complex mechanism involved in drought tolerance in chickpea (*Cicer arietinum* L.) induced by PGPR and PGRs. Sci Rep 9:2097. https://doi.org/10.1038/s41598-019-38702-8

Khokhar MK, Tetarwal JP (2012) Management of post-harvest black mould fruit rot of pomegranate (*Punica granatum* L.) caused by *Aspergillus niger* (Tieghem). Agric Res Rev 1:162–165

Kim YH, Choi Y, Oh YY, Ha NC, Song J (2019) Plant growth-promoting activity of beta-propeller protein YxaL secreted from *Bacillus velezensis* strain GH1-13. PLoS One 14:e0207968. https://doi.org/10.1371/journal.pone.0207968

Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8

Koumoutsi A, Chen XH, Henne A, Liesegang H, Hitzeroth G, Franke P, Vater J, Borriss R (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. https://doi.org/10.1128/jb.186.4.1084-1096.2004

Landy M, Warren GH, Roseman SB, Colio LG (1948) Bacillomycin, an antibiotic from *Bacillus subtilis* active against pathogenic fungi. Proc Soc Exp Biol Med 67:539–541. https://doi.org/10.3181/00379727-67-16367

Li Y, Gu Y, Li J, Xu M, Wei Q, Wang Y (2015) Biocontrol agent *Bacillus amyloliquefaciens* LJ02 induces systemic resistance against cucurbits powdery mildew. Front Microbiol 6:1–15. https://doi.org/10.3389/fmicb.2015.00883

Liu Z, Budiharjo A, Wang P, Shi H, Fang J, Borriss R, Zhang K, Huang X (2013) The highly modified microcin peptide plantazolicin is associated with nematicidal activity of *Bacillus amyloliquefaciens* FZB42. Appl Microbiol Biotechnol 97:10081–10090. https://doi.org/10.1007/s00253-013-5247-5

Lu X, Liu SF, Yue L, Zhao X, Zhang YB, Xie ZK, Wang RY (2018) Epsc involved in the encoding of exopolysaccharides produced by *Bacillus amyloliquefaciens* FZB42 act to boost the drought tolerance of *Arabidopsis thaliana*. Int J Mol Sci 19:E3795. https://doi.org/10.3390/ijms19123795

Manzoor S, Niazi A, Bejai S, Meijer J, Bongcam-Rudloff E (2013) Genome sequence of a plant-associated bacterium, *Bacillus amyloliquefaciens* strain UCMB5036. Genome Announc 1:e0011113. https://doi.org/10.1128/genomeA.00111-13

Marczynski GT, Shapiro L (1993) Bacterial chromosome origins of replication. Curr Opin Genet Dev 3:775–782. https://doi.org/10.1016/s0959-437x(05)80098-x

Martinez-Garcia E, Fraile S, Rodriguez-Espeso D, Vecchietti D, Bertoni G, de Lorenzo V (2020) The naked cell: emerging properties of a surfome-streamlined *Pseudomonas putida* strain. bioRxiv. https://doi.org/10.1101/2020.05.17.100628

McErlean M, Overbay J, Van Lanen S (2019) Refining and expanding nonribosomal peptide synthetase function and mechanism. J Ind Microbiol Biotechnol 46:493–513. https://doi.org/10.1007/s10295-018-02130-w

Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R (2016) BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics 32:1749–1751. https://doi.org/10.1093/bioinformatics/btw044

Niazi A, Manzoor S, Asari S, Bejai S, Meijer J, Bongcam-Rudloff E (2014) Genome analysis of *Bacillus amyloliquefaciens* subsp. *plantarum* UCMB5113: a rhizobacterium that improves plant growth and stress management. PLoS One 9:e104651. https://doi.org/10.1371/journal.pone.0104651

Omasits U, Varadarajan AR, Schmid M, Goetze S, Melidis D, Bourqui M, Nikolayeva O, Québatte M, Patrignani A, Dehio C, Frey JE, Robinson MD, Wollscheid B, Ahrens CH (2017) An integrative strategy to identify the entire protein coding potential of prokaryotic genomes by proteogenomics. Genome Res 27:2083–2095. https://doi.org/10.1101/gr.218255.116

Otusanya MO, Jeger MJ (1996) Effect of *Aspergillus niger* on shoot emergence and vine development in field-sown yams (*Dioscorea* spp.) and rot development under long-term storage conditions. Int Biodeterior Biodegrad 38:89–100. https://doi.org/10.1016/S0964-8305(96)00030-3

Pierneef R, Cronje L, Bezuidt O, Reva ON (2015) Pre_GI: a global map of ontological links between horizontally transferred genomic islands in bacterial and archaeal genomes. Database (Oxford) 2015:bav058. https://doi.org/10.1093/database/bav058

Pushpanathan M, Gunasekaran P, Rajendhran J (2013) Antimicrobial peptides: versatile biological properties. Int J Pept 2013:675391–675315. https://doi.org/10.1155/2013/675391

Qin M, Wu S, Li A, Zhao F, Feng H, Ding L, Ruan J (2019) LRScaf: improving draft genomes using long noisy reads. BMC Genomics 20:955. https://doi.org/10.1186/s12864-019-6337-2

Rabbee MF, Ali MS, Choi J, Hwang BS, Jeong SC, Baek KH (2019) *Bacillus velezensis*: a valuable member of bioactive molecules within plant microbiomes. Molecules 24:E1046. https://doi.org/10.3390/molecules24061046

Raza W, Ling N, Yang L, Huang Q, Shen Q (2016) Response of tomato wilt pathogen *Ralstonia solanacearum* to the volatile organic compounds produced by a biocontrol strain *Bacillus amyloliquefaciens* SQR-9. Sci Rep 6:1–13. https://doi.org/10.1038/srep24856

Reva ON, Swanevelder DZH, Mwita LA, Mwakilili AD, Muzondiwa D, Joubert M, Chan WY, Lutz S, Ahrens CH, Avdeeva LV, Kharkhota MA, Tibuhwa D, Lyantagaye S, Vater J, Borriss R, Meijer J (2019) Genetic, epigenetic and phenotypic diversity of four *Bacillus velezensis* strains used for plant protection or as probiotics. Front Microbiol 10:2610. https://doi.org/10.3389/fmicb.2019.02610

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 13:278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Ryu C-M, Farag MA, Hu C-H, Reddy MS, Kloepper JW, Pare PW (2004) Bacterial volatiles induce systemic resistance in *Arabidopsis*. Plant Physiol 134:1017–1026. https://doi.org/10.1104/pp.103.026583

Safronova LA, Zhmurko LG, Karlovsky OA, Evseenko OV (2012) New environmentally safe biopreparation as an effective means of protection against soybean disease. Agroecol J 4:71–76

Safronova LA, Nudga AY, Avdeeva LV (2013) Patent of Ukraine №. 77141. Phytosubtil biopreparation for processing root crops and potatoes during their storage. Bul. № 2, 01/25/2013

Sánchez-Romero MA, Casadesús J (2020) The bacterial epigenome. Nat Rev Microbiol 18:7–20. https://doi.org/10.1038/s41579-019-0286-2

Saxena AK, Kumar M, Chakdar H, Anuroopa N, Bagyaraj DJ (2019) *Bacillus* species in soil as a natural resource for plant health and nutrition. J Appl Microbiol 128:1583–1594. https://doi.org/10.1111/jam.14506

Schmid M, Frei D, Patrignani A, Schlapbach R, Frey JE, Remus-Emsermann MNP, Ahrens CH (2018) Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats. Nucleic Acids Res 46:8953–8965. https://doi.org/10.1093/nar/gky726

Schneider K, Chen XH, Vater J, Franke P, Nicholson G, Borriss R, Süssmuth RD (2007) Macrolactin is the polyketide biosynthesis product of the *pks2* cluster of *Bacillus amyloliquefaciens* FZB42. J Nat Prod 70:417–1423. https://doi.org/10.1021/np070070k

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N (2016) Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nat Commun 7:11307. https://doi.org/10.1038/ncomms11307

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197

Teng JLL, Yeung ML, Chan E, Jia L, Lin CH, Huang Y, Tse H, Wong SSY, Sham PC, Lau SKP, Woo PCY (2017) PacBio but not Illumina technology can achieve fast, accurate and complete closure of the high GC, complex *Burkholderia pseudomallei* two-chromosome genome. Front Microbiol 8:1448. https://doi.org/10.3389/fmicb.2017.01448

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017

Varadarajan AR, Allan RN, Valentin JD, Ocampo OEC, Somerville V, Pietsch F, Buhmann MT, West JJ, Skipp PJ, van der Mei HC, Ren Q, Schreiber F, Webb JS, Ahrens, CH. (2020) An integrated model system to gain mechanistic insights into biofilm formation and antimicrobial resistance development in *Pseudomonas aeruginosa* MPAO1. bioRxiv. https://doi.org/10.1101/2020.02.06.936690

Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746

Vater J, Kablitz B, Wilde C, Franke P, Mehta N, Cameotra SS (2002) Matrix-assisted laser desorption ionization–time of flight mass spectrometry of lipopeptide biosurfactants in whole cells and culture

filtrates of *Bacillus subtilis* C-1 isolated from petroleum sludge. Appl Environ Microbiol 68:6210–6219. https://doi.org/10.1128/aem.68.12.6210-6219.2002

Wu LM, Wu HJ, Chen LN, Xie S, Zang H, Borriss R, Gao X (2014) Bacilysin from *Bacillus amyloliquefaciens* FZB42 has specific bactericidal activity against harmful algal bloom species. Appl Environ Microbiol 80:7512–7520. https://doi.org/10.1128/AEM.02605-14

Wu LM, Wu HJ, Chen L, Yu X, Borriss R, Gao X (2015) Difficidin and bacilysin from *Bacillus amyloliquefaciens* FZB42 have antibacterial activity against *Xanthomonas oryzae* rice pathogens. Sci Rep 5:12975. https://doi.org/10.1038/srep12975

Wu L, Huang Z, Li X, Ma L, Gu Q, Wu H, Liu J, Borriss R, Wu Z, Gao X (2018) Stomatal closure and SA-, JA/ET-signaling pathways are essential for *Bacillus amyloliquefaciens* FZB42 to restrict leaf disease caused by *Phytophthora nicotianae* in *Nicotiana benthamiana*. Front Microbiol 9:847. https://doi.org/10.3389/fmicb.2018.00847

Yuan J, Raza W, Shen Q, Huang Q (2012) Antifungal activity of *Bacillus amyloliquefaciens* NJN-6 volatile compounds against *Fusarium oxysporum* f. sp. *cubense*. Appl Environ Microbiol 78:5942–5944. https://doi.org/10.1128/AEM.01357-12

Zhao J, Yuan S, Gao B, Zhu S (2018) Molecular diversity of fungal inhibitor cystine knot peptides evolved by domain repeat and fusion. FEMS Microbiol Lett 365:365. https://doi.org/10.1093/femsle/fny158

Zhu B, Stülke J (2017) SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. Nucleic Acids Res 46:D743–D748. https://doi.org/10.1093/nar/gkx908