Relationship of *Bacillus amyloliquefaciens* clades associated with strains DSM 7^T and FZB42^T: a proposal for *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov. and *Bacillus amyloliquefaciens* subsp. *plantarum* subsp. nov. based on complete genome sequence comparisons

Rainer Borriss,¹ Xiao-Hua Chen,¹ Christian Rueckert,² Jochen Blom,² Anke Becker,^{2†} Birgit Baumgarth,² Ben Fan,¹ Rüdiger Pukall,³ Peter Schumann,³ Cathrin Spröer,³ Helmut Junge,⁴ Joachim Vater,⁵ Alfred Pühler² and Hans-Peter Klenk³

Correspondence

Rainer Borriss

Rainer.Borriss@rz.hu-berlin.de

¹Bakteriengenetik, Institut für Biologie, Humboldt Universität Berlin, Chausseestrasse 117, 10115 Berlin, Germany

²Computational Genomics, Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

³DSMZ – German Collection of Microorganisms and Cell Cultures, Inhoffenstrasse 7B, 38124 Braunschweig, Germany

⁴ABiTEP, Glienicke Weg 185, 12489 Berlin, Germany

⁵Institut für Chemie, Technische Universität Berlin, Franklinstrasse 29, 10587 Berlin, Germany

The whole-genome-sequenced rhizobacterium *Bacillus amyloliquefaciens* FZB42^T (Chen *et al.*, 2007) and other plant-associated strains of the genus *Bacillus* described as belonging to the species *Bacillus amyloliquefaciens* or *Bacillus subtilis* are used commercially to promote the growth and improve the health of crop plants. Previous investigations revealed that a group of strains represented a distinct ecotype related to *B. amyloliquefaciens*; however, the exact taxonomic position of this group remains elusive (Reva *et al.*, 2004). In the present study, we demonstrated the ability of a group of *Bacillus* strains closely related to strain FZB42^T to colonize *Arabidopsis* roots. On the basis of their phenotypic traits, the strains were similar to *Bacillus amyloliquefaciens* DSM 7^T but differed considerably from this type strain in the DNA sequences of genes encoding 16S rRNA, gyrase subunit A (*gyrA*) and histidine kinase (*cheA*). Phylogenetic analysis performed with partial 16S rRNA, *gyrA* and *cheA* gene sequences revealed that the plant-associated strains of the genus *Bacillus*, including strain FZB42^T, formed a lineage, which could be distinguished from the cluster of strains closely related to *B. amyloliquefaciens* DSM 7^T. DNA–DNA hybridizations (DDH) performed with genomic DNA from strains DSM 7^T and FZB42^T yielded relatedness values of 63.7–71.2%. Several methods of genomic analysis, such as direct whole-genome comparison, digital DDH and microarray-based comparative genomic

[†]Present address: Molekulare Genetik, Institut für Biologie III, Albert-Ludwigs-Universität Freiburg, Schänzlestrasse 1, D-79104 Freiburg, Germany.

Abbreviations: AZCL, azurine cross-linked; CDS, coding sequence; CLSM, confocal laser scanning microscopy; DDH, DNA–DNA hybridization; ESI, electrospray ionization; FAME, fatty acid methyl ester; GGDC, genome-to-genome distance comparison; HSP, high-scoring segment pair; NJ, neighbour-joining; M-CGH, microarray-based comparative genomic hybridization; ML, maximum-likelihood; MLST, multi-locus sequence typing; MP, maximum-parsimony; MSP, mass spectral profiles; SNP, single nucleotide polymorphism.

The GenBank/EMBL/DDBJ accession number for the whole-genome sequence of strain DSM 7^T is FN597644. Accession numbers for partial 16S rRNA gene sequences and *gyrA* and *cheA* gene sequences determined in this study are provided in Figs 4 and 5, respectively.

hybridization (M-CGH) were used as complementary tests. The group of plant-associated strains could be distinguished from strain DSM 7^T and the type strain of *B. subtilis* by differences in the potential to synthesize non-ribosomal lipopeptides and polyketides. Based on the differences found in the marker gene sequences and the whole genomes of these strains, we propose two novel subspecies, designated *B. amyloliquefaciens* subsp. *plantarum* subsp. nov., with the type strain FZB42^T (=DSM 23117^T=BGSC 10A6^T), and *B. amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov., with the type strain DSM 7^T (=ATCC 23350^T=Fukumoto Strain F^T), for plant-associated and non-plant-associated representatives, respectively. This is in agreement with results of DDH and M-CGH tests and the MALDI-TOF MS of cellular components, all of which suggested that the ecovars represent two different subspecies.

INTRODUCTION

A group of Gram-positive, aerobic, endospore-forming bacteria called, in the vernacular, the '*Bacillus subtilis*' group, is traditionally of great importance in basic and applied microbiology (Fritze, 2004). All members of this group, which originally consisted of *B. subtilis*, *B. licheniformis*, and *B. pumilus*, are placed in 16S rRNA/DNA group 1. The first novel species added to this group was *B. amyloliquefaciens*, originally described as a potent producer of liquefying amylase and other extracellular enzymes of industrial importance (Fukumoto, 1943). In 1987, *B. amyloliquefaciens* was recognised as a distinct species (Priest *et al.*, 1987) after a long lasting debate about its taxonomical position (Welker & Campbell, 1967). Its separation from *B. subtilis* was based mainly on significantly low DNA relatedness values, which were found to be less than 25, 13, and 5 % in DNA–DNA hybridization (DDH) experiments performed with DNA from *B. subtilis*, *B. licheniformis* and *B. pumilus*, respectively (Priest *et al.*, 1987). Other names of species in the *B. subtilis* group that have been validly published within the last 20 years, include the black-pigmented *Bacillus atrophaeus* (Nakamura, 1989), *Bacillus mojavensis* (Roberts *et al.*, 1994), *Bacillus vallismortis* (Roberts *et al.*, 1996), *B. sonorensis* (Palmisano *et al.*, 2001) and *B. tequilensis* (Gatson *et al.*, 2006). Except for *B. licheniformis*, which is propionate-positive, grows at temperatures up to 55 °C and is a facultative anaerobe, and *B. pumilus*, which is starch-negative and hippurate-positive, all other species belonging to the *B. subtilis* group have been distinguished from *B. subtilis* exclusively by their fatty acid composition and by genetic differences in DDH tests. In the case of *B. amyloliquefaciens*, faster acid production from lactose and slower production of gluconate distinguishes it from *B. subtilis* (Fritze, 2004). Until now, *B. subtilis* Cohn, 1872 is the only member of the group that has been subdivided into several subspecies, namely, *B. subtilis* subsp. *subtilis*, *B. subtilis* subsp. *spizizenii* (Nakamura *et al.*, 1999) and *B. subtilis* subsp. *inaquosorum* (Rooney *et al.*, 2009).

Three strains belonging to the genus *Bacillus*, strains FZB13, FZB24 and FZB42^T, were previously isolated from plant-pathogen-infested soil of a sugar beet field in Brandenburg, Germany (Krebs *et al.*, 1998) and have since been shown to promote plant growth (Bochow *et al.*, 2001; Grosch *et al.*,

1999; Idriss *et al.*, 2002; Idris *et al.*, 2007; Schmiedeknecht *et al.*, 1998), a property typical of many plant-associated rhizobacteria (Kloepper *et al.*, 1980) but not the type strains *B. subtilis* DSM 10^T and *B. amyloliquefaciens* DSM 7^T. The three strains were commercialized for increasing crop yield (Yao *et al.*, 2006). According to phage-susceptibility studies, the strains were found to be most closely related to *B. amyloliquefaciens* (Krebs *et al.*, 1998). A more extensive analysis corroborated the close taxonomic relationship to *B. amyloliquefaciens*, although, deviations within 16S rRNA gene sequences and differences in ribotype patterns and *Sfi*I macrorestriction profiles between the three strains and *B. amyloliquefaciens* DSM 7^T were noticed. In addition, the strains could be distinguished from *B. subtilis* by their ability to produce acid from lactose, a feature attributed to *B. amyloliquefaciens* (Idriss *et al.*, 2002). Plant-associated strains of the genus *Bacillus* that are closely related to *B. amyloliquefaciens* are widely distributed. Reva *et al.* (2004) reported 7 out of 17 *Bacillus* strains, isolated from plants and soil, which formed a cluster distinct from *B. amyloliquefaciens* type strain DSM 7^T. These strains were generally better adapted to colonization of the rhizosphere than other members of the *Bacillus subtilis* group and were considered as a distinct ecotype of *B. amyloliquefaciens*.

Analysis of the whole genome of strain FZB42^T revealed an unexpected potential to produce secondary metabolites (Chen *et al.*, 2007). More than 8.5 % of the genome, corresponding to 340 kb, was dedicated to synthesizing antibiotics and siderophores by pathways not involving ribosomes (Chen *et al.*, 2009b). This exceeds the genomic capacity of *B. subtilis* strain 168^T (Barbe *et al.*, 2009) and, more remarkably, that of the type strain *B. amyloliquefaciens* DSM 7^T (accession no. FN597644) by more than two times. Here we present novel whole-genome sequences obtained from *B. amyloliquefaciens* DSM 7^T and from three plant-associated strains of *B. amyloliquefaciens*, known for their potential to promote plant growth. The differences between the whole genome sequences of strains FZB42^T and *B. amyloliquefaciens* DSM 7^T inspired us to reinvestigate the taxonomy of strain FZB42^T as a related subgroup to the type strains of *B. amyloliquefaciens* and *B. subtilis*.

Due to deviations in core genomes, changes in the variable portion of the genomes and differences in specific marker

gene sequences, we propose that members of the strain FZB42^T subgroup be classified as *B. amyloliquefaciens* subsp. *plantarum* subsp. nov. To the best of our knowledge, this is the first time that methods based on whole genome comparisons have been used extensively together with 'classical' methods of modern bacterial taxonomy to describe a novel subspecies within the genus *Bacillus*.

METHODS

Bacterial strains, plant hosts and growth conditions used. The type strain *B. amyloliquefaciens* DSM 7^T (strain F^T, first described by Fukumoto 1943) was obtained from the DSMZ. *B. amyloliquefaciens* FZB42^T, originally isolated from plant-pathogen-infested soil in Germany (Krebs *et al.*, 1998; Chen *et al.*, 2007) was deposited as strain 10A6 at the culture collection of the Bacillus Genetic Stock Center (BGSC) and as DSM 23117 at the DSMZ. Other strains of *B. amyloliquefaciens* used throughout in this study are listed in Table 1. Strains S23 (F. Priest, Edinburgh, Scotland), ATCC 15841 (American Type Culture Collection), ZF178 (Institute of Technical Microbiology; Riedel *et al.*, 1987) and BE20/78 (lab stock; Idriss *et al.*, 2002) represented strains related to the type strain DSM 7^T (see Results section). Commercial strains FZB13 (*T*=99; used for bioformulation Terranal; ECOstyle, 8426 ZM Appelscha, The Netherlands), FZB24 (used for bioformulations TAE-022 or 'Taegro', RhizoPlus; ABiTEP) and FZB45 (Idriss *et al.*, 2002) were isolated from plant-pathogen-infested soil in Germany (Krebs *et al.*, 1998). Strains CAU-B946 (Chinese Agricultural University, Beijing, China), NAU-B3 (Nanjing Agricultural University, Nanjing, China) and YAU Y2 (Yunnan Agricultural University, Kunming, China) were isolated as plant-root-associated bacteria from different locations in China and were previously described as representatives of *B. amyloliquefaciens* (Chen *et al.*, 2007). Other plant-associated strains of *B. amyloliquefaciens* used in this study were NAU B2, NAU B55 and NAU B85 (Nanjing Agricultural University, Nanjing, China), FZB45 (ABiTEP, Berlin; Krebs *et al.*, 1998; Idriss *et al.*, 2002), A1/3 (=FZB109; Hofemeister *et al.*, 2004), GB03 (used for biofungicide Kodiak, Bayer Crop Science LP, NC 27709, USA; Brannen & Kenney, 1997), QST713 (isolated from a California peach orchard in 1995, active ingredient of biofungicide Serenade, AgraQuest, CA, USA), FDK21 (S. Kunz, University of Konstanz, Germany; Chen *et al.*, 2009a), UCM B5113 (isolated from soil; red pigmented) and UCM B5044 (isolated from a cotton plant; Ukrainian Collection of Microorganisms, Kyiv, Ukraine; Reva *et al.*, 2004), and *Bacillus*

pumilus INR7 (J. Kloepper, Auburn University, AL, USA; Raupach & Kloepper, 1998). Bacteria were cultivated routinely in Luria broth (LB) solidified with 1.5% agar. For biosurfactant production and MALDI-TOF MS characterization, the bacteria were grown in Landy medium as described previously (Koumoutsis *et al.*, 2004).

The ability to colonize plant roots was studied to distinguish members of the ecovars from one another. In all plant colonization experiments six plants were used in each test group along with a negative control group (seeds treated with sterile distilled water). *Arabidopsis thaliana* Columbia seeds were sterilized in 10% sodium hypochlorite solution for 3 mins and rinsed with sterilized distilled water three times. The seeds were allowed to germinate at 24 °C with 18 h day-light in Petri dishes on Murashige and Skoog (MS) medium containing 0.6% agar and 1% sucrose. The seedlings were then transferred to a vertically positioned square dish (12 × 12 cm) containing, in its bottom part, MS medium without sucrose and solidified with 0.8% agar. After 7 days of growth, the roots of the seedlings were soaked for 2 min in LB containing 10⁶ bacteria ml⁻¹ and were plated onto 1% MS agar diluted 1:1. After seven days, the infected plants were gathered, washed three times with 1 ml saline, put into a test tube containing 1 ml saline and vortexed for 10 s, then transferred into a new tube containing 1 ml saline and vortexed for a further 30 s. 100 µl cell suspension was plated onto LB agar in appropriate dilution steps and incubated at 37 °C. The experiments were repeated five times, the cell numbers were normalized and standard deviations were calculated.

GFP-labelled *B. amyloliquefaciens* FZB42^T (P_{spacgfp}) was grown in LB until OD₆₀₀ 1.0 was reached. The culture was diluted 1000 times with fresh LB and then incubated at 37 °C with shaking for 15 min before being used as an inoculum. The roots of the plant seedlings were inoculated by dipping in gently swirling bacterial culture for two minutes.

Construction of GFP-labelled cells of strain FZB42^T. The two border sequences of the *amyE* gene of strain FZB42^T were amplified from genomic DNA using primers amyBack-1 (5'-AGCGAAATT-ACCTGACGGCAG-3'), amyBack-2 (5'-AGCTCAAGTTCGGTCAC-AC CTG-3'), amyFront(aatII)-1 (5'-AGTTTGACGTC TCTCCGAT-TTCGCCGACAACAC-3') and amyFront-2 (5'-TCGATTTGTTTG-CAGTTTCAGCG-3'). Primers were designed according to sequence information obtained from the whole genome sequence of strain FZB42^T (Chen *et al.*, 2007). The amplified sequences were inserted into vector plasmid pUC18^{Emr}. The *gfp* gene, together with an upstream located P_{spac} promoter element, was derived from plasmid ECE149 (BGSC) and introduced into recombinant plasmid pUC18^{Emr} bearing the two *amyE* border sequences. The resulting

Table 1. Characteristics of plant-associated and non-plant-associated strains of *B. amyloliquefaciens*

Strains: 1, DSM 7^T (=F^T); 2, S23; 3, ATCC 15841; 4, BE20/78; 5, ZF178; 6, FZB13; 7, FZB24; 8, FZB42^T; 9, NAU B3; 10, CAU B946; 11, YAU B9601-Y2. Additional plant-associated strains of *B. amyloliquefaciens* were included in the single investigations listed here and in MLST tests (see text). Bacterial strains, with newly determined sequences in this study include DSM7^T, CAU B946, NAU B3 and YAU Y2; those with known whole genome sequences include strain FZB42^T. +, Positive; -, negative; (+), weakly positive. FAME analysis, MSP determination, M-CGH, MLST (16S rRNA, *gyrA*, *cheA*) and MALDI-TOF MS for the non-ribosomal synthesis of lipopeptides and polyketides were performed on all strains except for strains 4 and 5, for which results of MLST, M-CGH and MALDI-TOF MS were not determined.

Characteristic	Strains related to <i>B. amyloliquefaciens</i> DSM7 ^T					Plant-associated strains of <i>B. amyloliquefaciens</i>					
	1	2	3	4	5	6	7	8	9	10	11
Plant colonizing	-	-	-	-	-	+	+	+	+	+	+
AZCL-HE cellulose liquefaction	-	-	-	(+)	-	+	+	+	(+)	(+)	+
Growth in lactose minimal medium	(+)	(+)	(+)	(+)	(+)	+	+	+	(+)	+	+

integrative plasmid, pFB01, was transformed into competent cells of strain FZB42^T as described previously (Idriss *et al.*, 2007). The *amy*-transformants were selected on LB plates supplemented with 1% starch, 1 µg erythromycin ml⁻¹ and 25 µg lincomycin ml⁻¹. Homologous recombination was confirmed by PCR and fluorescence microscopy.

Confocal Laser Scanning Microscopy (CLSM). The *Arabidopsis* seedlings were sampled 7 days after inoculation. Bacteria growing on *Arabidopsis* roots could be observed directly with CLSM. All samples were thoroughly washed with 5 ml distilled water prior to observation. CLSM was carried out with a Leica DM IRE2 & DM IRB system (Leica) using an excitation wavelength of 488 nm (argon laser) and an emission wavelength band of 500–550 nm for GFP fluorescence. Transmission light was collected to visualize root structure and was designated red in colour in later image reconstruction in order to contrast the green colour. Images were acquired and reconstructed by using Leica Confocal Software (LCS 2.6).

Assay of metabolic features. A microplate-based Biolog phenotyping system (GEN III microplates) was used for assaying substrates utilized by plant-associated and non-plant-associated strains of *B. amyloliquefaciens*. Substrates that the phenotyping indicated as being used differently by the strains tested were reanalysed according to Gordon *et al.* (1973). The Biolog system was used to test the metabolic phenotype of the organism by using a 96-well plate format, each well containing a different carbon source. The test organism was inoculated into the plate and incubated, usually for 24 h. The pattern of wells that were positive or negative for substrate utilization was then analysed on a plate reader and matched to a database which provided a presumptive identification for the organism. Endo-cellulase activity was determined by using insoluble azurine cross-linked (AZCL)-HE-cellulose (Megazyme) according to the manufacturer's instructions. Culture plates with nutrient agar containing 0.1% AZCL-HE-cellulose were used for the assay.

Fatty acid methyl ester (FAME) analysis. Together with DDH, FAME analysis was used to distinguish plant-associated from non-plant-associated representatives of *B. amyloliquefaciens*. FAMES were prepared and analysed as previously described (Klatte *et al.*, 1994) using the standard Microbial Identification System (MIDI) for automated GC analyses (Sasser, 1990). The resulting chromatogram provided a quantitative profile of fatty acids that were used for identification (Suzuki *et al.*, 1993).

MALDI-TOF MS. In recent years, MALDI-TOF MS profiling of cellular components, especially surface proteins, has proven to be a valuable tool for modern phenotyping. Spectra have been obtained using a Microflex L20 mass spectrometer (Bruker Daltronics) equipped with an N₂ laser. Sample preparation for MALDI-TOF analyses was carried out according to the OS-extraction protocol of Bruker Daltronics.

MALDI-TOF MS was also used for detection of secondary metabolites synthesized non-ribosomally. For the detection of the lipopeptide products from whole cells, strains were grown on agar plates with Landy medium. To record mass spectra, cell material was picked from the agar plate, spotted onto the target and covered with matrix medium for analysis by MALDI-TOF MS as previously described (Leenders *et al.*, 1999). Alternatively, a small sample of the freeze-dried culture filtrate was extracted with 70% acetonitrile/0.1% trifluoroacetic acid and treated further as previously described (Vater *et al.*, 2002). Post-source decay mass spectra were obtained with the same samples. Monoisotopic mass numbers were recorded. Polyketides were identified in extracts of lyophilized culture filtrates and XAD7 extracts by MALDI-TOF MS. Mass spectra were recorded with a Bruker Daltronic Autoflex MALDI-TOF instrument containing a 337 nm N₂ laser for desorption and ionization. One hundred single scans were

accumulated for every spectrum and 2,5-dihydroxybenzoic acid was used as the matrix. For MS analysis, 1–2 µl portions of extracts were mixed with an equal volume of matrix solution, spotted onto the target, air dried and measured as previously described (Chen *et al.*, 2006).

HPLC-electrospray ionization (ESI) MS. Alternatively, polyketides were identified by HPLC-ESI MS on a QTRAP 2000 system (Applied Biosystems) coupled with an Agilent 1100 HPLC system as described previously (Chen *et al.*, 2006).

DNA fragment amplification and reconstruction of phylogenetic trees. Isolation of chromosomal DNA from growing bacilli, DNA amplification and sequencing of the amplified fragments were performed as described previously (Idriss *et al.*, 2002). Primers, previously described by Idriss *et al.* (2002) and Reva *et al.* (2004), were used for amplification of partial 16S rRNA, *gyrA* and *cheA* genes. The GenBank/EMBL/DBJ accession numbers for sequences obtained in this study are given in Figs 4 and 5. DNA sequences were aligned by using the CLUSTAL W program (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>; Thompson *et al.*, 1994). A distance matrix was calculated from this alignment by performing DNA distance matrix calculations in the DNADIST PROGRAM. The matrix was then transformed into a tree by using the NEIGHBOR program. In order to assess reliability of the trees, multiple datasets were generated with the SEQBOOT program using 1000 bootstrap replicates. A tree was built from each replicate using DNADIST and bootstrap values were computed with the CONSENSE program. The phylogenetic tree was visualized with the TreeView32 program (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>). Programs used to reconstruct the phylogenetic tree were obtained from the PHYLIP version 3.69 software package (<http://evolution.genetics.washington.edu/phylip.html>; Felsenstein, 1989).

Spectroscopic DDH. Cells were disrupted by using a French pressure cell (Thermo Spectronic) and the DNA in the crude lysate was purified by chromatography on hydroxyapatite (Cashion *et al.*, 1977). DDH was carried out as described by De Ley *et al.* (1970) with modifications as described by Huß *et al.* (1983) using a Cary 100 Bio UV/VIS-spectrophotometer equipped with a Peltier-thermostatted 6 × 6 multicell changer and a temperature controller with *in situ* temperature probe (Varian) in 2 × SSC at 69 °C with two repetitions.

Microarray-based comparative genomic hybridization (M-CGH). A *Bacillus amyloliquefaciens* FZB42^T strain-specific oligonucleotide microarray was performed as a complementation to whole-genome-based analysis to identify genes that were absent or divergent in the test strains listed in Fig. 6 and Table 4. Each array was spotted with 3931 50- to 70-mer oligonucleotides; 3816 spots represented the predicted gene set of strain FZB42^T and 238 spots represented the hypothetical smRNAs (Chen *et al.*, 2007). All oligonucleotides were spotted four times in a microarray plate. The comparative hybridizations were repeated two or three times for each test strain and included at least one hybridization where the labelling regimen was switched to rule out potential bias introduced by inherent differences in Cy dye incorporation. Five micrograms of purified *MspI*- and *TaqI*-digested genomic DNA was labelled with either Cy3- or Cy5-NHS ester as described previously (Giuntini *et al.*, 2005). Following hybridization and scanning, data analysis was performed using ImaGene 6.0 software (Biodiscovery) for acquisition of the mean signal and mean local background intensity for each spot of the microarray and EMMA 2.2 software for normalization and *t*-statistics (Dondrup *et al.*, 2003; Becker *et al.*, 2004). A gene was considered to have a statistically significant difference in hybridization if the log₂-ratio of the intensities (*M* value) was ≥ 1 or ≤ -1, if the mean intensity (*A* value; $A_i = \log_2(R_i G_i)^{0.5}$) was ≥ 7 and if, in two of the three repeats, the *P*_{adjusted} value was ≤ 0.1. In this study a positive log₂-ratio of the intensities (*M* value) indicated that the respective gene was missing in the genome of the tested strain.

Genome sequencing, assembly and annotation. Genomic DNA prepared from *Bacillus amyloliquefaciens* DSM 7^T was used for construction of a 3 kb-long paired-end library with a GS FLX library preparation kit in combination with GS FLX paired end adaptors (Roche) according to the manufacturer's protocol. Sequencing was performed using an FLX genome sequencer (Roche) delivering 851 700 reads with a total of 108 926 617 base pairs. The reads were assembled using the GS *de novo* Assembler resulting in 13 scaffolds containing 49 contigs (with 55 contigs larger than 500 bp in total), indicating a genome size of approximately 3 913 433 bp. The scaffolds were oriented based on comparison to the genome of *B. amyloliquefaciens* FZB42^T and the remaining gaps were closed by long range PCR using Phusion polymerase (New England Biolabs) and subsequent Sanger sequencing, resulting in 177 reads in total (IIT Biotech). Prediction of protein-encoding sequences was initially accomplished by using REGANOR (Linke *et al.*, 2006). Automatic annotation was performed using the annotation software GenDB 2.4* (<https://www.cebitec.uni-bielefeld.de/groups/brf/software/gendb-2.2/cgi-bin/login.cgi>; Meyer *et al.*, 2003). 1278 ORFs out of a total of 3921 were manually curated using comparative annotation with the genomes of strain FZB42^T (Chen *et al.*, 2007) and *B. subtilis* 168^T (Barbe *et al.*, 2009; http://subtiwiki.uni-goettingen.de/wiki/index.php/Main_Page).

Comparative genome analysis was performed using EDGAR software framework (Blom *et al.*, 2009; http://edgar.cebitec.uni-bielefeld.de/cgi-bin/edgar_login.cgi). A private project was constructed comprising *B. amyloliquefaciens* strains DSM 7^T, CAU B946, NAU B3, YAU Y2, FZB42^T and four other selected strains of the genus *Bacillus* with known genome sequences. To construct a phylogenetic tree for this project, the core genes of the genomes were computed. In a following step, multiple alignments of the core genes were generated using MUSCLE, non-matching parts of the alignment were masked by GBLOCKS and subsequently removed. The remaining parts of all alignments were concatenated into one large alignment. The PHYLIP package (Felsenstein, 1989) was used to create a phylogenetic tree of this alignment, represented in newick format.

RESULTS

Ability to colonize plant roots is restricted to a specific group of *B. amyloliquefaciens* strains

Colonization of *Arabidopsis* roots by strains of the genus *Bacillus* was determined as described in Methods. Colonization rates of more than 5×10^5 c.f.u. per root were

registered in all plant-associated strains except for strain UCM B5044. Strain UCM B5044, although having been isolated from cotton plants, was reported as being unable to colonize plant roots (Reva *et al.*, 2004). In contrast to this, cell numbers detected in all strains related to *Bacillus amyloliquefaciens* DSM 7^T were not distinct from that of the control, lying below 0.21×10^5 c.f.u. per root (Fig. 1).

In addition, the plant-colonizing ability of strain FZB42^T was demonstrated with GFP-labelled cells of strain FZB42^T, detected by using CLSM. Seedlings of *Arabidopsis thaliana* were incubated with the bacteria. The roots appeared to be preferentially colonized at their tips and at laterally emerging root hairs (Fig. 2). By contrast, similar experiments performed with GFP-labelled *B. subtilis* DSM 10^T and *B. amyloliquefaciens* DSM 7^T failed, suggesting that the ability to colonize plant roots is not commonly distributed within the members of the *B. subtilis* group and is restricted to the plant-associated members of this taxonomic unit.

Metabolic features of plant-associated and type-strain-related strains of *B. amyloliquefaciens*

All 11 strains listed in Table 1 were found to be strictly aerobic and were able to grow in a medium containing 0.001 % lysozyme. No apparent differences in substrate utilization were detected between strains FZB42^T and DSM 7^T after 72 h of incubation in the ID module of the Biolog system, suggesting a close relationship to *B. subtilis* and *B. amyloliquefaciens*. In contrast to *B. subtilis*, the plant-associated strains of *B. amyloliquefaciens* grew well on lactose minimal medium, as did the type strain *B. amyloliquefaciens* DSM 7^T. A more remarkable feature, useful in distinguishing between groups of *B. amyloliquefaciens*, was the ability of the plant-associated strains of *B. amyloliquefaciens* to liquefy AZCL-HE cellulose, indicating endo- β -1,4-glucan-glucano-hydrolase (endo-cellulase) activity. This property is also common in *B. subtilis* but the majority of the strains related to the type strain *B. amyloliquefaciens* DSM 7^T were unable to hydrolyse cellulose

Fig. 1. Colonization of *Arabidopsis* roots by 19 *B. amyloliquefaciens* strains. The number of colony forming units isolated from *Arabidopsis* roots is the mean value of at least three independent experiments.

Fig. 2. CLSM images showing the colonization by GFP labelled cells of strain FZB42^T at a root hair (a) and root tip (b) of *Arabidopsis thaliana*.

(Table 1). The plant-associated strains FZB13, FZB24, FZB42^T, NAU B3, CAU B946 and YAU Y2 could not be clearly distinguished from the type strain *B. amyloliquefaciens* DSM 7^T by the spectrum of fatty acids obtained by FAME analysis (Table 2). In summary, the phenotypic traits established for the plant-associated strains did not allow clear classification of these strains as members of one of the two species, although they did appear to be more related to *B. amyloliquefaciens* than to *B. subtilis*, as proposed previously (Krebs *et al.*, 1998; Idriss *et al.*, 2002).

MALDI-TOF MS

The mass spectral profiles (MSP), mainly reflecting molecular masses (m/z) of surface macromolecules, which were obtained after irradiating of whole cells, were used to distinguish the plant-associated and non-plant-associated strains of the genus *Bacillus*. The MSP dendrogram deduced from our data revealed that these bacterial strains clearly fell into two different categories (Fig. 3), allowing clear discrimination between both groups of bacteria.

16S rRNA phylogenetic analysis

The nearly complete 16S rRNA gene sequence of the plant-associated strain FZB42^T was compared with that of *B. amyloliquefaciens* DSM 7^T and other members of the *B. subtilis* group. This comparison indicated variable residues located at positions ranging from 89 to 1462. Most of them were located within the hypervariable 5' region (Supplementary Table S1, available in IJSEM Online). A phylogenetic tree was reconstructed with nucleotide sequences extracted from the hypervariable 5' region (nucleotides 115–485) of the 16S rRNA genes of plant-associated and non-plant-associated strains of *B. amyloliquefaciens* and other members of the *B. subtilis* subgroup using the corresponding region of *Bacillus cereus* as an outgroup. This region has been useful in previous studies for the classification and identification of endospore-forming bacteria (Goto *et al.*, 2000; Idriss *et al.*, 2002; Reva *et al.*, 2004). Neighbour-joining (NJ) analysis revealed that strains of *B. amyloliquefaciens* together with the type strains of *B. atrophaeus* and *B. vallismortis* formed one cluster which could be distinguished from another cluster formed by *B. mojavensis*, *B. tequilensis* and the subspecies *B. subtilis* subsp. *subtilis*, *B. subtilis* subsp. *spizizenii* and *B. subtilis* subsp. *inaquosorum*. Within the *B. amyloliquefaciens* cluster, strains FZB13, FZB24 and FZB42^T, together with other strains known for their plant-growth-promoting and biocontrol activity (Reva *et al.*, 2004; Li *et al.*, 2008; Sharma *et al.*; X. Gao; Q. Wang; Y. He; unpublished results), comprised a monophyletic lineage closely related to a second cluster comprising *B. amyloliquefaciens* strains DSM 7^T, S23 and ATCC15841. Bootstrap analysis corroborated the branch points for the *B. amyloliquefaciens* and *B. subtilis* lineages but did not sufficiently support the separation of the *B. amyloliquefaciens* type-strain-related strains and their plant-associated counterparts, including strains FZB13, FZB24 and FZB42^T (Fig. 4).

gyrA and *cheA* gene sequences

In order to further clarify the degree of relatedness of the plant-associated strains of *B. amyloliquefaciens* to strains related to the type strain DSM 7^T, we used protein-coding genes, which exhibit much greater genetic variation and can be used for classification and identification of closely related taxonomic groups and ecotypes (Cole *et al.*, 2010; Connor *et al.*, 2010). The highly conserved *gyrA* gene encodes the DNA gyrase subunit A; the more variable *cheA* gene encodes the two-component sensor histidine kinase CheA, which is crucial for regulating bacterial chemotaxis. The *gyrA* and *cheA* genes have been used previously for resolving closely related taxa of the *B. subtilis* group (Chun & Bae, 2000; Reva *et al.*, 2004). Alignments performed with full-length (not published) and partial *gyrA* and *cheA* nucleotide sequences derived from closely related representatives of the *B. subtilis* group revealed identical results, which suggested that strains FZB13, FZB24 and FZB42^T represent a separate cluster, together with other

Table 2. Cellular fatty acid compositions (%) of type strain-related and plant-associated strains of *B. amyloliquefaciens*Strains: 1, DSM 7^T; 2, S23; 3, ATCC 15841; 4, BE20/78; 5, ZF178; 6, FZB42^T; 7, FZB13; 8, FZB24; 9, NAU B3; 10, CAU B946; 11, YAU B9601-Y2.

Fatty acid	Strains related to <i>B. amyloliquefaciens</i> DSM7 ^T					Plant-associated strains of <i>B. amyloliquefaciens</i>					
	1	2	3	4	5	6	7	8	9	10	11
iso C _{13:0}	0.38	0.19	0.46	0.27	0.22	0.31	0.34	0.35	0.60	0.38	0.70
iso C _{14:0}	0.99	0.69	0.69	0.86	0.87	0.43	0.45	0.48	0.41	0.98	0.53
C _{14:0}	0.36	0.20	0.76	0.48	0.40	1.21	1.58	1.03	1.95	1.98	2.80
iso C _{15:0}	40.29	38.65	62.78	33.69	25.78	31.00	28.89	26.10	38.65	29.82	36.64
anteiso C _{15:0}	28.32	25.25	17.74	34.81	33.98	31.73	32.98	35.90	25.25	32.03	31.43
C _{16:1} ω7c	0.42	0	0.29	0.59	0.51	0.19	0	0	0	0.28	0.16
iso C _{16:0}	2.13	0.65	1.80	1.47	2.70	1.01	0.94	1.21	0.65	1.68	0.71
C _{16:1} ω11c	1.23	1.24	0.64	1.70	1.41	2.59	2.82	2.26	3.21	2.86	3.25
C _{16:0}	3.02	2.20	2.60	2.79	3.94	7.60	9.21	7.48	9.14	10.70	9.33
iso C _{17:1} ω10c	2.59	5.02	1.26	3.69	2.83	2.70	2.29	2.05	2.66	1.67	1.81
iso C _{17:0}	13.14	19.27	7.42	9.95	14.72	12.11	11.53	11.92	11.69	10.22	7.64
C _{17:0}	6.46	9.93	3.29	8.05	11.54	7.70	7.83	10.04	4.50	6.34	3.92

plant-associated strains of *B. amyloliquefaciens*, which is distinct from the type strain *B. amyloliquefaciens* DSM 7^T. Reva *et al.* (2004) previously reported a similar finding. Notably, other strains with known plant-growth-promoting and biocontrol activity formerly mislabelled as *Bacillus subtilis* also grouped with this cluster.

The NJ phylogenetic tree constructed from partial *gyrA* sequences (Fig. 5a), shows a clear distinction between the three *B. subtilis* subspecies, *B. vallismortis*, and the two *B. amyloliquefaciens* groups, containing DSM 7^T- and FZB42^T-related strains. The tree topology was corroborated by 100 % bootstrap values. Within the *B. amyloliquefaciens* group, discrimination between the newly proposed subspecies 'amyloliquefaciens' and 'plantarum' was supported by bootstrap values of 100 % and 52 %, respectively. In addition to strains FZB13, FZB24 and FZB42^T, the proposed 'plantarum' subspecies group also contained many representatives of plant-associated bacilli with known plant-growth-promoting properties, collected from

distant sites in Europe, North America, Australia, and Asia, some of which (e.g. strains GB03 and QST713) have been used commercially as biocontrol and biofertilizer agents in the agriculture industry (Joshi & McSpadden Gardener, 2006).

The NJ tree obtained with the more variable partial *cheA* gene sequences (Fig. 5b) also distinguished between plant-associated and non-plant-associated strains of *B. amyloliquefaciens*. Branching of the two clades within *B. amyloliquefaciens* was supported by bootstrap values of 76 % and 100 %, respectively. Again, strains FZB13, FZB24 and FZB42^T formed a separate taxonomical unit, together with other plant-associated bacilli.

DNA–DNA-hybridization (DDH)

The development of nucleic-hybridization methods, introduced into prokaryotic systematics from the 1960s onwards, has allowed the indirect comparison of gene

Fig. 3. Dendrogram deduced from MALDI-TOF mass spectral profiles of 11 plant-associated and non-plant-associated strains of *B. amyloliquefaciens*.

Fig. 4. NJ phylogenetic tree based on partial 16S rRNA nucleotide sequences using *B. cereus* ATCC 14579^T as an outgroup. The consensus tree was reconstructed from 1000 trees according to the extended majority rule (SEQBOOT program). Bootstrap values >50 %, based on 1000 repetitions, are indicated at branch points. Strain and accession numbers are indicated. Type strains for the two subspecies of *B. amyloliquefaciens* are in bold. Bar, 0.01 substitutions per nucleotide position. Strain numbers with asterisks indicate sequences obtained in this study.

sequences. DDH is typically applied when strains share more than 97 % 16S rRNA gene sequence similarity. Based on a threshold of 70 % recommended for the delineation of bacterial species, DDH values <70 % generally indicate that the tested organism belongs to a different species than the strains used as a reference (Tindall *et al.* 2010). An interesting ‘*in silico*’ alternative to the cumbersome ‘wet lab’ experimental DDH estimate, named genome-to-genome distance comparison (GGDC), has recently been developed (Auch *et al.*, 2010a, b). The method is based on whole genome data and allows the inclusion of unfinished draft genome sequences. We took advantage of this method to determine the genomic distance between strains FZB42^T and DSM 7^T. The complete *B. subtilis* 168^T genome and draft genomes of three other plant-associated strains related to *B. amyloliquefaciens*, namely, strains YAU Y2, CAU B946 and NAU B3, were also included in the analysis. The results demonstrated that *B. subtilis* and *B. amyloliquefaciens* were distinct from one another at the species level, their digital DDH values being much lower than 70 %, whilst the DDH

values between strains FZB42^T and DSM 7^T were calculated as being around 77 %. Values between 70–80 % are considered as sufficient for the discrimination of subspecies. GGDC analysis of draft genomes of the three strains YAU Y2, CAU B946 and NAU B3 yielded DDH values of 86–88 % with strain FZB42^T but only 74–77 % when compared with *B. amyloliquefaciens* DSM 7^T (Table 3), supporting a closer taxonomic relatedness to the plant-associated strains of *B. amyloliquefaciens*. Spectroscopic DDH performed with hydroxyapatite-purified chromosomal DNA isolated from strains DSM 7^T and FZB42^T corroborated the results obtained with the ‘*in silico*’ calculation. The DNA–DNA relatedness values ranged from 63.7 to 71.2 %. Taken together, these results did not sufficiently support the separation of both groups at the species level but were clearly sufficient to distinguish the plant- and non-plant-associated strains of *B. amyloliquefaciens* as separate subspecies.

Microarray-based comparative genomic hybridization (M-CGH)

M-CGH is a powerful method for rapidly identifying regions of genomic diversity among closely related organisms in the absence of complete genome datasets. This technique allows one to predict gene absence (or divergence) versus gene presence by measuring the relative hybridization efficiencies of two differentially Cy-labelled pools of genomic DNA taken from two strains. The method has previously been applied to discriminate between different members of the *B. subtilis* clade (Earl *et al.* 2007). We used *Bacillus amyloliquefaciens* FZB42^T strain-specific oligonucleotide microarrays to identify genes that were absent or divergent in the strains tested (Table 4). As controls, FZB42^T–FZB42^T and FZB42^T–CH40 (FZB42^TΔ*nrs::cm* and Δ*dhb::em*) hybridizations were performed. As expected, the self-self control experiments yielded no genes with a log₂ fluorescence ratio greater than 1. The results of the FZB42^T–CH40 hybridization did, however, reveal a potential limitation of the array; the values for only 8 of the 11 genes known to be deleted in this strain were above the cut-off ratio for gene absence or divergence. This may have been a consequence of cross-hybridization between gene spots as the missing genes, *nrsA*, *nrsC* and *nrsF*, contain peptide synthetase modules, which exist quite frequently in the genome of strain FZB42^T. Fig. 6 shows a composite view of the results of M-CGH experiments in heat-map format organized according to the results of hierarchical clustering using Spearman rank correlation to calculate similarity among all datasets (Eisen *et al.* 1998). Strain FZB42^T, its derivative strain CH40 and, surprisingly, strain FZB24 were nearly identical, underlining the close taxonomic relationship between strain FZB42^T and FZB24. The other plant-associated strains of *B. amyloliquefaciens* formed a cluster closely related to strain FZB42^T. Strain DSM 7^T, together with the two other non-plant-associated strains of *B. amyloliquefaciens*, ATCC 15841 and S23, were more diverse and formed a separate cluster. *Bacillus pumilus* INR7 was used as outgroup in this comparative analysis. For the strains examined, a range in the amount of gene absence or

Fig. 5. NJ phylogenetic tree based on partial *gyrA* (a) and *cheA* (b) nucleotide sequences. The consensus tree was reconstructed from 1000 trees according to the extended majority rule (SEQBOOT program). Bootstrap values >50%, based on 1000 repetitions, are indicated at branch points. Strain numbers with asterisks indicate sequences obtained in this study. The type strains for the two subspecies of *B. amyloliquefaciens* are in bold. *B. cereus* ATCC 14579^T *gyrA* and *cheA* gene sequences have been used as outgroups. Bar, 0.1 substitutions per nucleotide position.

divergence was detected by the array (Table 4). Mirroring the results obtained in *gyrA* and *cheA* gene sequence analyses, the plant-associated strains of *B. amyloliquefaciens* exhibited less diversity relative to strain FZB42^T (0.2–6.9% divergence among the CDS and sRNAs tested) than the non-plant-associated strains (11.4–14.3%). Interestingly, the phylogeny obtained when the degree and pattern of gene variation measured by the arrays was used as a marker of relatedness was in almost perfect concordance with the phylogeny obtained when *gyrA* and *cheA* genes were used as markers of relatedness. In both cases, the cluster of *B. amyloliquefaciens* FZB42^T-related strains was distinguishable from the cluster of *B. amyloliquefaciens* DSM 7^T-related strains. As shown previously for *B. subtilis* (Earl *et al.* 2007), M-CGH was proven to be a reliable phylogenetic tool for subtyping strains of *B. amyloliquefaciens*.

Genome comparison

The complete genome of *B. amyloliquefaciens* DSM 7^T (3 980 199 bp in size, DNA GC content 46.08%, 3 922

ORFs) was determined and, using EDGAR software (Blom *et al.*, 2009), significant differences were found to the genome sequence of strain FZB42^T. Strains FZB42^T and DSM 7^T shared 3345 genes (CDS) in their core genomes; 547 and 344 CDS were unique to strains DSM 7^T and FZB42^T, respectively. The core genome sequence shared by both strains exhibited 97.89% similarity on an amino acid level. For comparison, *B. subtilis* 168^T (DSM 10^T) had a similar number of CDS in common with *B. amyloliquefaciens* strains DSM 7^T and FZB42^T (3222 and 3182 CDS, respectively). In addition to strains FZB42^T and DSM 7^T, the recently resolved draft genomes of three further plant-associated strains of *B. amyloliquefaciens*, CAU B946, YAU Y2 and NAU B3, (Table 5) were included in the analysis. The number of genes representing the core genome from these strains and *B. subtilis* DSM 10^T was calculated as being 2974 and their deduced amino acid sequence identity was calculated as being 92.25%. The phylogenetic tree deduced from the core genomes of members of the *B. subtilis* group, including five strains of *B. amyloliquefaciens* and using *B. cereus* as outgroup, suggested that strains

Fig. 6. M-CGH composite view of genome diversity exhibited by 17 *B. amyloliquefaciens* strains and *B. pumilus* INR7. Each row shows the results of one strain, and each column represents genes as they are found along the length of the *B. amyloliquefaciens* FZB42^T genome. Black indicates a gene that is likely to be present in the test strain (mean Cy control/Cy test log₂ ratio <1) and white indicates a gene that is either absent or too highly divergent to hybridize with equal efficiency to the gene of strain FZB42^T (mean Cy control/Cy test log₂ ratio >1). The image was generated using the CLUSTER and TreeView programs.

FZB42^T and DSM 7^T represent taxonomically related but distinct units; bootstrap values of 95 % and 74 %, respectively, indicated that, on the genomic level, *B. amyloliquefaciens* and *B. subtilis* represented distinct species but also supported our proposal to separate the plant-associated strains of *B. amyloliquefaciens* from the type strain at the subspecies level (Fig. 7).

Occurrence of polysaccharide-degrading enzymes

It is well known that members of the *B. amyloliquefaciens* DSM 7^T clade secrete a starch-liquefying α -amylase (*amyA*) with high industrial potential, whilst *B. subtilis* secretes a saccharifying enzyme (*amyE*; Keay, 1970). Unlike the type

Table 3. Comparison of spectroscopic DNA–DNA hybridization and *in silico* GGDC analysis of genomic DNA of strains of the genus *Bacillus* using the program BLAT

Regression-based DDH estimates (%) are indicated. Data in bold are results of comparisons between the same strain.

Query/reference	Formula*	DSM 7 ^T	FZB42 ^T	<i>B. subtilis</i> 168 ^T
<i>B. amyloliquefaciens</i> DSM 7 ^T	1	97.7	80.3	37.1
length=3980199 bp	2	87.1	64.4	14.9
(acc. no. FN597644)	3	100	77.6	30.3
DNA–DNA relatedness†			63.7 (69.0)	
<i>B. amyloliquefaciens</i> FZB42 ^T	1	80.3	97.7	38.2
length=3918589 bp	2	64.4	87.1	15.0
(acc. no. CP000560.1)	3	77.6	100	31.2
DNA–DNA relatedness†		66.0 (71.2)		
<i>B. amyloliquefaciens</i> YAU Y2‡	1	76.0	84.6	35.6
length=4198660 bp	2	64.1	79.1	14.2
	3	73.7	85.7	29.0
<i>B. amyloliquefaciens</i> NAU B3‡	1	77.2	86.3	36.5
length=4154898 bp	2	64.1	79.1	14.3
	3	74.8	87.4	29.7
<i>B. amyloliquefaciens</i> CAU B946‡	1	80.8	88.1	35.2
length=3978635 bp	2	63.8	77.2	14.0
	3	77.89	88.4	28.7
<i>B. subtilis</i> 168	1	37.1	38.2	97.7
length=4215606 bp	2	14.9	15.0	87.1
(acc. no. AL009126.3)	3	30.3	31.2	100

*Formulae: 1, High-scoring segment pair (HSP) length/total length; 2, identities/HSP length; 3, identities/total length.

†Spectroscopic DNA–DNA hybridization was performed with *B. amyloliquefaciens* DSM7^T and FZB42^T only in two independent experiments. Values in parentheses are the results of the repeat measurement.

‡Draft genomes without rRNA gene clusters (available on personal request from the corresponding author).

Table 4. Number of 'variable' loci exhibited by each strain and presence of selected genes and gene clusters involved in non-ribosomal synthesis of secondary metabolites as determined by M-CGH with strain FZB42^T.

Strains: FZB42^T; 2, FZB24; 3, NAU B2; 4, NAU B3; 5, NAU B55; 6, CAU B946; 7, YAU Y2; 8, FDK21; 9, A1/3; 10, GB03; 11, QST713; 12, UCM B-5044; 13, UCM B-5113; 14, DSM 7^T; 15, S23; 16, ATCC 15841.

Locus/gene	<i>B. amyloliquefaciens</i> subsp. <i>plantarum</i>													<i>B. amyloliquefaciens</i> subsp. <i>amyloliquefaciens</i>		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Number																
–	0	7	213	214	254	271	228	166	173	214	214	182	206	458	529	563
+	3946	3939	3733	3732	3692	3675	3718	3780	3773	3732	3732	3764	3740	3488	3417	3383
Polyketide																
dfn	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–
mln	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–
bae	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Lipopeptide																
itu	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
fen	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–
srf	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
dhb	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nrs	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–

strain DSM 7^T, the genomes of strains FZB42^T, CAU B946, YAU Y2, NAU B3 and FZB24 have not been found to contain the *amyA* sequence (M. Rey, personal communication), corroborating an earlier finding of Reva *et al.* (2004) who were unable to amplify *amyA* gene-like sequences in plant-associated strains of *B. amyloliquefaciens*. These strains possessed, instead, an *amyE*-like gene, which also occurs in *B. subtilis* (Table 5).

Environmental and industrial *B. subtilis*-group strains secrete different hydrolases enabling these bacteria to use external cellulosic and hemicellulosic substrates present in plant cell walls. Two of these enzymes, endo-1,4- β -glucanase or cellulose (BglC, EC 3.2.1.4) and endo-1,4- β -xylanase (xylanase, 1,4- β -xylan xylanohydrolase, EC 3.2.1.8) are encoded in *B. subtilis* 168^T by the genes *bglC* (*eglS*) and *xynA*, respectively (Wolf *et al.*, 1995). The same genes were also present in the genomes of plant-associated strains of *B. amyloliquefaciens* but were not detected in *B. amyloliquefaciens* DSM 7^T (Table 5).

Non-ribosomal synthesis of secondary metabolites

Besides the five gene clusters known in *B. subtilis* to mediate non-ribosomal synthesis of secondary metabolites, four additional giant clusters, *bmyD*, *mln*, *dfn* and *nrs*, were previously identified by Chen *et al.*, 2007, which were absent in *B. subtilis* 168^T. Whilst the function of the *nrs* gene product remains undetermined, the functions of the *bmyD* (bacillomycin D), *mln* (macrolactin) and *dfn* (difficidin) genes have been identified (Koumoutsis *et al.*, 2004; Chen *et al.*, 2006; Schneider *et al.*, 2007). Like *B. subtilis* 168^T, the genome of *B. amyloliquefaciens* DSM 7^T

harboured a significantly lower number of gene clusters involved in non-ribosomal synthesis of secondary metabolites than strain FZB42^T. No gene clusters involved in the synthesis of the antibacterial polyketides macrolactin and difficidin could be identified. Surprisingly, a gene cluster involved in fengycin production was only partially present in strain DSM 7^T (Table 5). A comparative survey of presence of non-ribosomal polyketides, lipopeptides and siderophores (bacillibactin) in the plant-associated (related to strain FZB42^T) and non-plant-associated (related to DSM 7^T) strains of *B. amyloliquefaciens* supported our genomic findings. *B. amyloliquefaciens* DSM 7^T (=F^T) and related strains S23 and ATCC 15841 did not produce difficidin, macrolactin or fengycin, whilst the plant-associated strains (except for UCM B-5113), including strain FZB42^T did produce these compounds. Interestingly, the iturin-like compounds produced by the plant-associated strains of *B. amyloliquefaciens* were different. Whilst the majority of the strains produced bacillomycin D, the production of bacillomycin L (by strain A1/3) and iturin A (by strains QST713, FDK21, UCM B-5044 and CAU B946) was also detected (Table 6).

Restriction and modification

Plant-associated strains of *B. amyloliquefaciens* and *B. subtilis* DSM 10^T evolved unique restriction modification systems without sequence similarity (results not shown). On the contrary, the type II restriction modification genes present in strain DSM 7^T shared 98 % (methyltransferase) and 100 % (restrictase) similarity with the *Bam*HI system known to be present in *B. amyloliquefaciens* H (Roberts *et al.*, 1977).

Table 5. Comparative analysis of genomes of plant-associated strains of *B. amyloliquefaciens*, *B. amyloliquefaciens* DSM 7^T and *B. subtilis* DSM 10^T.

Features	Plant-associated <i>B. amyloliquefaciens</i> strains				Type strain DSM 7 ^T emb FN597644	<i>B. subtilis</i> DSM 10 ^T Gb AL009126.3
	FZB42 ^T gb CP000560.1	CAU B946 draft	YAU Y2 draft	NAU B3 draft		
General features						
DNA G+C content	46.5 %	46.5 %	45.9 %	46.0 %	46.1 %	43.5 %
Genome size (bp)	3 918 589	4 019 861	4 247 774	4 193 003	3 980 199	4 215 606
Number of CDs	3660	4139	4363	4307	3921	4105
Unique genes	93	230	223	162	263	681
Extracellular carbohydrate degrading enzymes						
Amylase AmyA	—	—	—	—	514/514 (100 %)	—
Amylase AmyE	659/659 (100 %)	640/659 (97 %)	614/644 (95 %)	629/659 (95 %)	—	575/659 (100 %)
Cellulase BglC	499/499 (100 %)	473/499 (94 %)	483/501 (96 %)	480/499 (96 %)	—	464/499 (92 %)
Xylanase XynA	213/213 (100 %)	211/213 (99 %)	211/213 (99 %)	212/213	—	201/213 (94 %)
Non-ribosomal synthesis of lipopeptides and polyketides						
Bacillomycin D/ iturin A	100 % (<i>bmyA–D</i>)	96–97 % (<i>ituA–D</i>)	97–98 % (<i>bmyA–D</i>)	98 % (<i>bmyA–D</i>)	68–90 % (<i>ituA–D</i>)	—
Fengycin	100 % (<i>fenA–E</i>)	96–97 % (<i>fenA–E</i>)	97–98 % (<i>fenA–E</i>)	97 % (<i>fenA–E</i>)	71–93 % (<i>fenD–E</i>)	73–93 % (<i>ppsA–E</i>)
Macrolactin	100 % (<i>mlnA–I</i>)	97 % (<i>mlnA–I</i>)	97–98 % (<i>mlnA–I</i>)	97 % (<i>mlnA–I</i>)	—	—
Difficidin	100 % (<i>dfnA–M</i>)	96–100 % (<i>dfnA–M</i>)	97–100 % (<i>dfnA–M</i>)	97–99 % (<i>dfnA–M</i>)	—	—

DISCUSSION

Despite the enormous progress in microbial whole-genome sequencing in recent years, the minimal standards for the description of new taxa of aerobic, endospore-forming bacteria proposed by Logan *et al.* (2009) are founded mainly on a set of microscopic and macroscopic features including cell, spore and colony morphology, physiological and biochemical characteristics and profiles of fatty acid and cell-wall constituents. In addition, whole 16S rRNA

gene sequence analysis and, in the case of closely related species, DNA–DNA hybridization are also recommended tools. Unfortunately, these standards are not always completely sufficient to discriminate between closely related taxa, as was found with members of the *B. subtilis* species complex. For many years, it has been recognized that species belonging to this group cannot be differentiated on the basis of their phenotypic characteristics and 16S rRNA gene nucleotide sequences alone. Also, fatty acid profiles, including those of the bacterial groups investigated in this study, did not yield satisfying results for discriminating closely related bacterial groups. Phylogenetic analysis of multiple protein-coding loci has been used as a complementary approach to detect and differentiate novel taxa of the genus *Bacillus* (Gatson *et al.*, 2006; Rooney *et al.*, 2009). In the present study, we used this approach to describe a group of plant-associated strains of the genus *Bacillus* related to the type strain *B. amyloliquefaciens* F^T (=DSM 7^T).

Taxonomic trees reconstructed from *gyrA* and *cheA* gene sequences suggested that strains FZB13, FZB24 and FZB42^T, which were isolated from soil in the vicinity of sugar beet (*Beta vulgaris*) roots (Krebs *et al.* 1998), are closely related to a group of plant-colonizing and/or endophytic strains of the genus *Bacillus* formerly described by Reva *et al.* (2004) as forming an ecomorph distinct from the type strain *B. amyloliquefaciens* F^T (=DSM 7^T).

Fig. 7. Phylogenetic tree reconstructed using the core genomes of *B. amyloliquefaciens* strains B9601-Y2 (=YAU Y2), GaoB3 (=NAU B3), FZB42^T, B946 (=CAU B946), DSM 7^T, *B. subtilis* 168^T (DSM 10^T), *B. licheniformis* DSM 13, *B. pumilus* SAFR-032 and *B. cereus* ATCC 14579^T. Bootstrap values (%) are indicated at branch points. Bar, 0.05 substitutions per amino acid within the coding regions of the core genomes.

Table 6. Occurrence of non-ribosomal polyketides (bacillaene, difficidin, macrolactin), lipopeptides (surfactin, fengycin, iturins) and siderophores (bacillibactin) in plant-associated and non-plant-associated strains of *B. amyloliquefaciens*

Strains: 1, DSM 7^T; 2, S23; 3, ATCC 15841; 4, FZB42^T; 5, FZB24; 6, A1/3; 7, GB03; 8, QST713; 9, FDK21; 10, UCM B-5113; 11, UCM B-5044; 12, NAU B2; 13, NAU B3; 14, YAU Y2; 15, CAU B946. Data for polyketides were compiled from Schneider *et al.* (2007). Data for lipopeptides and bacillibactin were obtained by MALDI-TOF MS as described by Koumoutsis *et al.* (2004). +, Detected; –, not detected; ituA, IturinA; bmyD, bacillomycin D; bmyL, bacillomycin L.

Secondary metabolite	Plant-associated strains of <i>B. amyloliquefaciens</i>														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bacillaene	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Difficidin	–	–	–	+	+	+	+	+	+	+	+	+	+	+	+
Macrolactin	–	–	–	+	+	+	+	+	+	+	+	+	+	+	+
Surfactin	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fengycin	–	–	–	+	+	+	+	+	+	+	+	+	+	+	+
Iturins	ituA	ituA	ituA	bmyD	bmyD	bmyL	bmyD	ituA	ituA	–	ituA	bmyD	bmyD	bmyD	ituA
Bacillibactin	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Moreover, a great number of strains of the genus *Bacillus* with known plant-growth-promoting and biocontrol activities were shown to display close similarity in their partial 16S rRNA, *gyrA* and *cheA* gene sequences to the three FZB strains described in this study (Figs 4 and 5). Bootstrap analysis performed with *gyrA* and *cheA* gene sequences supported the NJ phylograms and the assertion that the plant-associated strains of *B. amyloliquefaciens* form a cluster distinct from those strains more closely related to *B. amyloliquefaciens* DSM 7^T.

While sequencing of rRNA genes and analysis of a few selected CDS marker genes (multilocus sequence typing, MLST) is used successfully in bacterial phylogenetics and taxonomy, there are several limitations in that the use of only a few single genes does not necessarily reflect the whole-cell/whole-genome relatedness. It has been shown that ribosomal phylogeny is sufficient to discriminate organisms at the genus level and higher but does not afford strain distinction at the species and, in particular, subspecies levels (Cole *et al.*, 2010). Due to novel, highly improved sequencing technologies, the availability of an increasing number of genome sequences opens up new possibilities in bacterial taxonomy. Perhaps the most important contribution of genome sequencing is towards deepening our understanding of the microbial world by exposing the differences between closely related micro-organisms (Cole *et al.*, 2010). It is now possible to distinguish related taxonomic units at the species and subspecies levels on the basis of sequence comparisons despite similar or identical phenotypic features. A comparative analysis between the genome sequences of the plant-associated strain FZB42^T and the type strain of the species, *B. amyloliquefaciens* DSM 7^T, revealed obvious differences in the variable part of the genomes, whilst the core genome, conserved in all representatives of the *B. subtilis* group, was very similar and did not go beyond 97.8 % similarity at the amino acid sequence level. Whole-genome sequence comparison between strains FZB24 and

FZB42^T revealed surprisingly few differences; except for a limited number of single-nucleotide polymorphisms (SNPs), no major genome variations have been found (M. Rey, personal communication), indicating a close taxonomic relationship between these strains. Moreover, analysis of draft genomes of three other plant-associated strains of *B. amyloliquefaciens* and direct measurements of lipopeptides and polyketides by MALDI-TOF MS provided corroborating evidence that the plant-associated strains of *B. amyloliquefaciens* are a distinct group and are discernable by their potential to synthesize a huge spectrum of different secondary metabolites non-ribosomally, many of which have antibacterial and/or antifungal properties (Schneider *et al.*, 2007). A survey of the genomes of four other *B. subtilis* strains (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi) and the genome of *B. amyloliquefaciens* DSM 7^T supported the evidence that strains related to *B. subtilis sensu stricto* and the type strain of *B. amyloliquefaciens* did not produce the polyketides difficidin and macrolactin and are often impaired in their ability to produce lipopeptides other than surfactin. In summary, the genomic data separating plant-associated and non-plant-associated strains of *B. amyloliquefaciens* on subspecies level, were perfectly supported by the results of genomic DDH and MALDI-TOF MS analysis of cellular components as well as the spectra of secondary metabolites present in culture fluids.

Based on DNA genome sequence comparisons, diversity in 16S rRNA genes and the molecular marker genes *gyrA* and *cheA*, the strain-specific occurrence of either *amyE* or *amyA*, *bglC* and *xynA*, differing capabilities in the non-ribosomal synthesis of lipopeptides and polyketides and the observation of different life-styles, strains of the *B. amyloliquefaciens* FZB42^T subgroup represent members of a novel subspecies for which the name *B. amyloliquefaciens* subsp. *plantarum* subsp. nov. is proposed. Strain FZB42^T is the neotype culture for this subspecies and may serve as a

guide for comparison of newly isolated plant-associated strains of *B. amyloliquefaciens*. This proposal automatically creates *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov.

Description of *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov.

Bacillus amyloliquefaciens subsp. *amyloliquefaciens* (am.yl.o.li.que.fac'i.ens. L. n. *amylum* starch; M.L. part. adj. *amyloliquefaciens* starch liquefying).

Cells are Gram-reaction-positive, motile rods, 0.6–0.9 × 1.8–4 µm and often in chains. Cylindrical spores are formed centrally or paracentrally in non-swollen sporangia. Strictly aerobic. Catalase positive. Grows optimally at 30–40 °C but does not grow below 15 °C or above 50 °C. Starch is hydrolysed by a liquefying amylase. Casein, elastin, gelatin, lichenan, tributyrin and Tweens 20, 40 and 60 are degraded but cellulose, guanine, hypoxanthine, pectin, testosterone, tyrosin, xanthin and Tween 80 are not. Acetyl-methylcarbinol (Voges–Proskauer test), acetoin and phosphatase are produced. Nitrate is reduced to nitrite. Aesculin and arbutin are hydrolysed. Citrate is utilized but propionate is not. Neither fengycin nor representatives of the iturin group are synthesized. Bacillaene is produced but no other polyketides are. Growth occurs in the presence of 7 % (w/v) NaCl and, for most strains, growth also occurs in 10 % (w/v) NaCl. Acid is produced from cellobiose, fructose, glucose, glycerol, lactose, maltose, mannose, mannitol, raffinose, salicin, sorbitol, sucrose and trehalose when the medium of Gordon *et al.* (1973) is used (Priest *et al.*, 1987).

The type strain, DSM 7^T (=ATCC 23350^T =Fukumoto strain F^T), is deposited at the DSMZ Culture Collection and the American type Culture Collection (ATCC).

Description of *Bacillus amyloliquefaciens* subsp. *plantarum* subsp. nov.

Bacillus amyloliquefaciens subsp. *plantarum* (plan.ta'rum. L. fem.n. *planta* a sprout; M.L. *planta* a plant; M.L. gen. pl. noun *plantarum* of plants).

The morphological, physiological and biochemical characteristics are the same as those of *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* except that soluble cellulose is decomposed and starch is hydrolysed by a saccharifying amylase. Strains are able to colonize plant roots and to produce the plant growth hormone indole-3-acetic acid. Relatedness to the core genome of *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T is 97.89 % (deduced amino acid sequence level). Due to presence of additional giant gene clusters involved in non-ribosomal production of secondary metabolites, representatives of the subspecies are able to produce macrolactin, difficidin, a representative of the iturin group (either iturin A or bacillomycin) and fengycin. Phylogeny based on inferred nucleotide

sequences of the hyper variable regions of 16S rRNA and partial *gyrA* and *cheA* genes distinguishes strains from *B. amyloliquefaciens* subsp. *amyloliquefaciens*. In total, 344 genes, including *amyE*, *eglS*, *xynA*, *mlnA–I* and *dfn A–M*, which are not present in the type strain of the species, *B. amyloliquefaciens* DSM 7^T, are different in both subspecies.

The type strain, plant-associated strain FZB42^T, is deposited as strain *Bacillus amyloliquefaciens* 10A6^T in the culture collection of BGSC and as strain *Bacillus amyloliquefaciens* DSM 23117^T (=BGSC 10A6^T) in the DSMZ culture collection.

ACKNOWLEDGEMENTS

We gratefully acknowledge Christiane Müller and Birgit Grün for their technical assistance. The financial support in-frame of the competence network Genome Research on Bacteria (GenoMikPlus) and the Chinese-German collaboration program by the German Ministry for Education and Research is gratefully acknowledged. We are very grateful to Michael Rey (Novozymes; Davis, CA, USA), Fergus Priest (Edinburgh, Scotland), the unknown referees and IJSEM editor Dr Fred Rainey, for their many valuable suggestions for improving the manuscript. We thank our colleagues Joe Kloepper (Auburn University, AL, USA), Brian McSpadden-Gardener (Wooster, OH, USA), Oleg Reva (Kyiv, Ukrainian Collection of Microorganisms; UCM), Professor Qi Wang (CAU, Beijing, China), Professor Xuewen Gao (NAU, Nanjing, China) and Professor Yueqiu He (YAU, Kunming, China) for delivering plant-growth-promoting strains and their trustful collaboration. *B. amyloliquefaciens* strain S23 was obtained from Fergus Priest. We are indebted to Professor D. Naumann and Dr P. Lasch enabling us to perform the MALDI-TOF MS with the Autoflex instrument at the Robert Koch Institut, Berlin. We thank Dr Oleg Reva, University of Pretoria, for support in preparing Fig. 7.

REFERENCES

- Auch, A. F., von Jan, M., Klenk, H.-P. & Göker, M. (2010a). Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2, 117–134.
- Auch, A. F., Klenk, H.-P. & Göker, M. (2010b). Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci* 2, 142–148.
- Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., Vallenet, D., Wang, T., Moszer, I. & other authors (2009). From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 155, 1758–1775.
- Becker, A., Bergès, H., Krol, E., Bruand, C., Rüberg, S., Capela, D., Lauber, E., Meilhoc, E., Ampe, F. & other authors (2004). Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Mol Plant Microbe Interact* 17, 292–303.
- Blom, J., Albaum, S. P., Doppmeier, D., Pühler, A., Vorhölter, F.-J., Zakrzewski, M. & Goesmann, A. (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* 10, 154.
- Bochow, H., El-Sayed, S. F., Junge, H., Stavropoulos, A. & Schmiedeknecht, G. (2001). Use of *Bacillus subtilis* as biocontrol agent. IV. Salt-stress tolerance induction by *Bacillus subtilis* FZB24 seed treatment in tropical vegetable field crops, and its mode of action. *J Plant Dis Prot* 108, 21–30.

- Brannen, P. M. & Kenney, D. S. (1997). Kodiak—a successful biological-control product for suppression of soil-borne plant pathogens of cotton. *J Ind Microbiol Biotechnol* **19**, 169–171.
- Cashion, P., Holder-Franklin, M. A., McCully, J. & Franklin, M. (1977). A rapid method for the base ratio determination of bacterial DNA. *Anal Biochem* **81**, 461–466.
- Chen, X. H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K. A., Koumoutsis, A., Hitzeroth, G., Grammel, N. & other authors (2006). Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J Bacteriol* **188**, 4024–4036.
- Chen, X. H., Koumoutsis, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., Morgenstern, B., Voss, B., Hess, W. R. & other authors (2007). Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nat Biotechnol* **25**, 1007–1014.
- Chen, X. H., Scholz, R., Borriss, M., Junge, H., Mögel, G., Kunz, S. & Borriss, R. (2009a). Difficidin and bacilysin produced by plant-associated *Bacillus amyloliquefaciens* are efficient in controlling fire blight disease. *J Biotechnol* **140**, 38–44.
- Chen, X. H., Koumoutsis, A., Scholz, R. & Borriss, R. (2009b). More than anticipated-production of antibiotics and other secondary metabolites by *Bacillus amyloliquefaciens* FZB42. *J Mol Microbiol Biotechnol* **16**, 14–24.
- Chun, J. & Bae, K. S. (2000). Phylogenetic analysis of *Bacillus subtilis* and related taxa based on partial *gyrA* gene sequences. *Antonie van Leeuwenhoek* **78**, 123–127.
- Cole, J. R., Konstantinidis, K., Farris, R. J. & Tiedje, J. M. (2010). Microbial diversity and phylogeny: extending from rRNAs to genomes. In *Environmental Molecular Microbiology*. pp.1–20. Edited by W.-T. Liu & J. K. Jansson. UK: Calster Academic Press Norfolk.
- Connor, N., Sikorski, J., Rooney, A. P., Kopac, S., Koepfel, A. F., Burger, A., Cole, S. G., Perry, E. B., Krizanc, D. & other authors (2010). Ecology of speciation in the genus *Bacillus*. *Appl Environ Microbiol* **76**, 1349–1358.
- De Ley, J., Cattoir, H. & Reynaerts, A. (1970). The quantitative measurement of DNA hybridization from renaturation rates. *Eur J Biochem* **12**, 133–142.
- Dondrup, M., Goesmann, A., Bartels, D., Kalinowski, J., Krause, L., Linke, B., Rupp, O., Sczyrba, A., Pühler, A. & Meyer, F. (2003). EMMA: a platform for consistent storage and efficient analysis of microarray data. *J Biotechnol* **106**, 135–146.
- Earl, A. M., Losick, R. & Kolter, R. (2007). *Bacillus subtilis* genome diversity. *J Bacteriol* **189**, 1163–1170.
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863–14868.
- Felsenstein, J. (1989). PHYLIP: phylogeny interference package. *Cladistics* **5**, 154–166.
- Fritze, D. (2004). Taxonomy of the genus *Bacillus* and related genera: the aerobic endospore-forming bacteria. *Phytopathology* **94**, 1245–1248.
- Fukumoto, J. (1943). [Studies on the production of bacterial amylase. I. Isolation of bacteria secreting potent amylases and their distribution.] *J Agric Chem Soc Jpn* **19**, 487–503 (in Japanese).
- Gatson, J. W., Benz, B. F., Chandrasekaran, C., Satomi, M., Venkateswaran, K. & Hart, M. E. (2006). *Bacillus tequilensis* sp. nov., isolated from a 2000-year-old Mexican shaft-tomb, is closely related to *Bacillus subtilis*. *Int J Syst Evol Microbiol* **56**, 1475–1484.
- Giuntini, E., Mengoni, A., De Filippo, C., Cavalieri, D., Aubin-Horth, N., Landry, C. R., Becker, A. & Bazzicalupo, M. (2005). Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of *Sinorhizobium meliloti* natural strains. *BMC Genomics* **6**, 158.
- Gordon, R. E., Haynes, W. C. & Hor-Nai Pang, C. (1973). *The genus Bacillus*. Washington, DC: US Department of Agriculture.
- Goto, K., Omura, T., Hara, Y. & Sadaie, Y. (2000). Application of the partial 16S rDNA sequence as an index for rapid identification of species in the genus *Bacillus*. *J Gen Appl Microbiol* **46**, 1–8.
- Grosch, R., Junge, H., Krebs, B. & Bochow, H. (1999). Use of *Bacillus subtilis* as biocontrol agent. III. Influence of *Bacillus subtilis* on fungal root diseases and on yield in soilless culture. *J Plant Dis Prot* **106**, 568–580.
- Hofemeister, J., Conrad, B., Adler, B., Hofemeister, B., Feesche, J., Kucheryava, N., Steinborn, G., Franke, P., Grammel, N. & other authors (2004). Genetic analysis of the biosynthesis of non-ribosomal peptide- and polyketide-like antibiotics, iron uptake and biofilm formation by *Bacillus subtilis* A1/3. *Mol Genet Genomics* **272**, 363–378.
- Huß, V. A. R., Festl, H. & Schleifer, K. H. (1983). Studies on the spectrophotometric determination of DNA hybridization from renaturation rates. *Syst Appl Microbiol* **4**, 184–192.
- Idris, E. E., Iglesias, D. J., Talon, M. & Borriss, R. (2007). Tryptophan-dependent production of indole-3-acetic acid (IAA) affects level of plant growth promotion by *Bacillus amyloliquefaciens* FZB42. *Mol Plant Microbe Interact* **20**, 619–626.
- Idriss, E. E., 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. *Microbiology* **148**, 2097–2109.
- Joshi, R. & McSpadden Gardener, B. B. (2006). Identification and characterization of novel genetic markers associated with biological control activity in *Bacillus subtilis*. *Phytopathology* **96**, 145–154.
- Keay, L. (1970). The action of *Bacillus subtilis* saccharifying amylase on starch and β -cyclodextrin. *Die Stärke (Starch)* **22**, 153–157.
- Klatte, S., Kroppenstedt, R. M. & Rainey, F. A. (1994). *Rhodococcus opacus* sp. nov., an unusual nutritionally versatile *Rhodococcus*-species. *Syst Appl Microbiol* **17**, 355–360.
- Kloepper, J. W., Leong, J., Teintze, M. & Schroth, M. (1980). Enhancing plant growth by siderophores produced by plant-growth-promoting rhizobacteria. *Nature* **286**, 885–886.
- Koumoutsis, A., Chen, X. H., 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.
- Krebs, B., Höding, B., Kübart, S. M., Workie, A., Junge, H., Schmiedeknecht, G., Grosch, R., Bochow, H. & Hevesi, M. (1998). Use of *Bacillus subtilis* as biocontrol agent. 1. Activities and characterization of *Bacillus subtilis* strains. *J Plant Dis Prot* **105**, 181–197.
- Leenders, F., Stein, T. H., Kablitz, B., Franke, P. & Vater, J. (1999). Rapid typing of *Bacillus subtilis* strains by their secondary metabolites using matrix assisted laser desorption/ionisation mass spectrometry of intact cells. *Rapid Commun Mass Spectrom* **13**, 943–949.
- Li, B., Xu, L. H., Lou, M. M., Li, F., Zhang, Y. D. & Xie, G. L. (2008). Isolation and characterization of antagonistic bacteria against bacterial leaf spot of *Euphorbia pulcherrima*. *Lett Appl Microbiol* **46**, 450–455.
- Linke, B., McHardy, A. C., Neuweget, H., Krause, L. & Meyer, F. (2006). REGANOR: a gene prediction server for prokaryotic genomes and a database of high quality gene predictions for prokaryotes. *Appl Bioinformatics* **5**, 193–198.

- Logan, N. A., Berge, O., Bishop, A. H., Busse, H.-J., De Vos, P., Fritze, D., Heyndrickx, M., Kämpfer, P., Rabinovitch, L. & other authors (2009). Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria. *Int J Syst Evol Microbiol* **59**, 2114–2121.
- Meyer, F., Goesmann, A., McHardy, A. C., Bartels, D., Bekel, T., Clausen, J., Kalinowski, J., Linke, B., Rupp, O. & other authors (2003). GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31**, 2187–2195.
- Nakamura, L. K. (1989). Taxonomic relationship of black-pigmented *Bacillus subtilis* strains and a proposal for *Bacillus atrophaeus* sp. nov. *Int J Syst Bacteriol* **39**, 295–300.
- Nakamura, L. K., Roberts, M. S. & Cohan, F. M. (1999). Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: a proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov. and *Bacillus subtilis* subsp. *spizizenii* subsp. nov. *Int J Syst Bacteriol* **49**, 1211–1215.
- Palmisano, M. M., Nakamura, L. K., Duncan, K. E., Istock, C. A. & Cohan, F. M. (2001). *Bacillus sonorensis* sp. nov., a close relative of *Bacillus licheniformis*, isolated from soil in the Sonoran Desert, Arizona. *Int J Syst Evol Microbiol* **51**, 1671–1679.
- Priest, F. G., Goodfellow, M., Shute, L. A. & Berkeley, W. (1987). *Bacillus amyloliquefaciens* sp. nov., nom. rev. *Int J Syst Bacteriol* **37**, 69–71.
- Raupach, G. S. & Kloepper, J. W. (1998). Mixtures of plant growth-promoting rhizobacteria enhance biological control of multiple cucumber pathogens. *Phytopathology* **88**, 1158–1164.
- Reva, O. N., Dixelius, C., Meijer, J. & Priest, F. G. (2004). Taxonomic characterization and plant colonizing abilities of some bacteria related to *Bacillus amyloliquefaciens* and *Bacillus subtilis*. *FEMS Microbiol Ecol* **48**, 249–259.
- Riedel, K., Schroeter, A., Liebs, P., Graba, J.-P., Hecker, M. & Schrapel, D. (1987). Formation of extracellular neutral proteinase and the stringent response in *Bacillus subtilis*. *Folia Microbiol (Praha)* **32**, 96–100.
- Roberts, R. J., Wilson, G. A. & Young, F. E. (1977). Recognition sequence of specific endonuclease BamH.I from *Bacillus amyloliquefaciens* H. *Nature* **265**, 82–84.
- Roberts, M. S., Nakamura, L. K. & Cohan, F. M. (1994). *Bacillus mojavensis* sp. nov., distinguishable from *Bacillus subtilis* by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. *Int J Syst Bacteriol* **44**, 256–264.
- Roberts, M. S., Nakamura, L. K. & Cohan, F. M. (1996). *Bacillus vallismortis* sp. nov., a close relative of *Bacillus subtilis*, isolated from soil in Death Valley, California. *Int J Syst Bacteriol* **46**, 470–475.
- Rooney, A. P., Price, N. P., Ehrhardt, C., Swezey, J. L. & Bannan, J. D. (2009). Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. *inaquosorum* subsp. nov. *Int J Syst Evol Microbiol* **59**, 2429–2436.
- Sasser, M. (1990). Identification of bacteria by gas chromatography of cellular fatty acids. *USFCC Newsl* **20**, 16.
- Schmiedeknecht, G., Bochow, H. & Junge, H. (1998). Use of *Bacillus subtilis* as biocontrol agent. II. Biological control of potato diseases. *J Plant Dis Prot* **105**, 376–386.
- Schneider, K., Chen, X.-H., Vater, J., Franke, P., Nicholson, G., Borriß, R. & Süssmuth, R. D. (2007). Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. *J Nat Prod* **70**, 1417–1423.
- Suzuki, K., Goodfellow, M. & O'Donnell, A. G. (1993). Cell envelopes and classification. In *Handbook of New Bacterial Systematics*. pp. 195–250. Edited by M. Goodfellow & A. G. O'Donnell. London, San Diego, New York: Academic Press.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Tindall, B. J., Rosselló-Móra, R., Busse, H.-J., Ludwig, W. & Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* **60**, 249–266.
- Vater, J., Kablitz, B., Wilde, C., Franke, P., Mehta, N. & Cameotra, S. S. (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.
- Welker, N. E. & Campbell, L. L. (1967). Unrelatedness of *Bacillus amyloliquefaciens* and *Bacillus subtilis*. *J Bacteriol* **94**, 1124–1130.
- Wolf, M., Geczi, A., Simon, O. & Borriß, R. (1995). Genes encoding xylan and β -glucan hydrolysing enzymes in *Bacillus subtilis*: characterization, mapping and construction of strains deficient in lichenase, cellulase and xylanase. *Microbiology* **141**, 281–290.
- Yao, A. V., Bochow, H., Karimov, S., Boturov, U., Sanginboy, S. & Sharipov, K. (2006). Effect of FZB42 *Bacillus subtilis* as a biofertilizer on cotton yields in field tests. *Arch Phytopathol Plant Prot* **39**, 323–328.