Complete genome sequence of a plant associated bacterium *Bacillus amyloliquefaciens* subsp. *plantarum* UCMB5033

Adnan Niazi\(^1\)*, Shahid Manzoor\(^1,3\), Sarosh Beja\(^2\), Johan Meijer\(^2\), Erik Bongcam-Rudloff\(^1\)

\(^1\)Department of Animal Breeding and Genetics, SLU Global Bioinformatics Centre, Swedish University of Agricultural Sciences, Uppsala, Sweden.
\(^2\)Department of Plant Biology and Forest Genetics, Uppsala Biocenter, Swedish University of Agricultural Sciences and Linnéan Center for Plant Biology, Uppsala, Sweden
\(^3\)University of the Punjab, Lahore, Pakistan.

*Correspondence: Adnan Niazi (adnan.niazi@slu.se)

Keywords: *Bacillus amyloliquefaciens*, biocontrol, rhizobacteria, priming, stress

*Bacillus amyloliquefaciens* subsp. *plantarum* UCMB5033 is of special interest for its ability to promote host plant growth through production of stimulating compounds and suppression of soil borne pathogens by synthesizing antibacterial and antifungal metabolites or priming plant defense as induced systemic resistance. The genome of *B. amyloliquefaciens* UCMB5033 comprises a 4,071,167 bp long circular chromosome that consists of 3,912 protein-coding genes, 86 tRNA genes and 10 rRNA operons.

Abbreviations: UCM- Ukrainian Collection of Microorganisms, ENA- European Nucleotide Archive, PGPB- Plant growth promoting bacterium

**Introduction**

*Bacillus amyloliquefaciens* is a plant-associated species belonging to the family *Bacillaceae*. The members of the genus *Bacillus* are ubiquitous in nature and include biologically and ecologically diverse species, ranging from those beneficial for economically important plants, to pathogenic species that are harmful to humans. *B. amyloliquefaciens* UCMB5033 is a plant growth promoting bacterium (PGPB) that was isolated from a cotton plant [1]. Studies have shown that *B. amyloliquefaciens* UCMB5033 is an important tool for studies of plant-bacteria associations, has potential to confer protection against soil borne pathogens, and to stimulate growth of oilseed rape (*Brassica napus*) [2]. Such traits make UCMB5033 an important tool for studies of plant-bacteria associations and production of compounds that directly or indirectly promote plant growth or stress tolerance. Here we present a description of the complete genome sequencing of *B. amyloliquefaciens* UCMB5033 and its annotation.

**Classification and features**

Strain UCMB5033 was identified as a member of the *B. amyloliquefaciens* group based on phenotypic analysis [1]. The comparison of 16S rRNA gene sequences with the most recent databases from GenBank using NCBI BLAST [3] under default settings showed that *B. amyloliquefaciens* UCMB5033 shares 99% identity with many *Bacillus* species including *Bacillus atrophaeus* (CP002207.1) and *Bacillus subtilis* subsp. *spizizenii* str. W23 (CP002183.1). Figure 1 shows the phylogenetic relationship of *B. amyloliquefaciens* UCMB5033 with other species within the genus *Bacillus*. The tree highlights the close relationship of UCMB5033 with the *B. amyloliquefaciens* subsp. *plantarum* type strain FZB42. The other *B. amyloliquefaciens* type strain DSM 7\(^\text{r}\) representing subsp. *amyloliquefaciens*, displayed less taxonomic relatedness and strain UCMB5033 can thus be regarded as belonging to the subsp. *plantarum* also in line with its plant associated characteristics [7].
Figure 1. Phylogenetic tree showing the position of B. amyloliquefaciens UCMB5033 in relation to other species within the genus Bacillus. The tree is based on 16S rRNA gene sequences aligned with MUSCLE [4] was inferred under maximum likelihood criterion using MEGA5 [5] and rooted with Geobacillus thermoglucosidasius (a member of the family Bacillaceae). The numbers above the branches are support values from 1,000 bootstrap replicates if larger than 50% [6].

Morphology and physiology
B. amyloliquefaciens UCMB5033 is a Gram-positive, rod shaped, motile, spore forming, aerobic, and mesophilic microorganism (Table 1). Strain UCMB5033 is approximately 0.8 µm wide and 2 µm long that can grow on Luria Broth (LB) and potato dextrose agar (PDA) between 20 °C and 37 °C within the pH range 4–8. B. amyloliquefaciens UCMB5033 has properties as a plant growth promoting rhizobacterium (PGPR) [2]. The ability to catabolize plant derived compounds, resistance to metals and drugs; root colonization and biosynthesis of metabolites presumably give B. amyloliquefaciens UCMB5033 an advantage in developing a symbiotic relationship with plants in competition with other microorganisms in the soil microbiota.

Genome assembly and annotation
Growth conditions and DNA isolation
B. amyloliquefaciens UCMB5033 was grown in LB medium at 28°C for 12 hours (cells were in the early stationary phase). The genomic DNA was isolated using a QIAamp DNA mini kit (Qiagen).

Genome sequencing
B. amyloliquefaciens UCMB5033, originally isolated from cotton plant, was selected for sequencing on the basis of its ability to promote rapeseed growth and inhibit soil borne pathogens. Genome sequencing of B. amyloliquefaciens UCMB5033 using Illumina multiplex technology and Ion Torrent PGM systems was performed by Science for Life Laboratory (SciLifeLab) at Uppsala University. The genome project is deposited in the Genomes On Line Databases [24] and the complete genome sequence is deposited in the ENA database under accession number HG328253. A summary of the project information is shown in Table 2 and its association with MIGS identifiers.

Genome assembly
The genome of B. amyloliquefaciens UCMB5033 was assembled using 21,919,534 Illumina paired-end reads (75bp) and 1,922,725 single-end reads (Ion Torrent). The chromosome of size 4,071,167 bp was assembled by providing paired-end reads to MIRA v.3.4 [25] for reference-guided assembly using the available genome sequence of B. amyloliquefaciens UCMB5036 (accession no. HF563562) [26]. Whereas, single-end reads were assembled with Newbler v.2.8 by a de novo assembly method. Both forms of assemblies were compared after alignment to identify indels and cover gap regions using Mauve genome alignment software [27].
Table 1. Classification and general features of *B. amylo*liquefaciens *subsp. plantarum* UCMB5033 according to the MIGS recommendation [8].

| MIGS ID | Property                | Term                                         | Evidence code<sup>a</sup> |
|---------|-------------------------|----------------------------------------------|---------------------------|
|         | Domain                  | *Bacteria*                                   | TAS [9]                   |
|         | Phylum                  | *Firmicutes*                                 | TAS [10-12]               |
|         | Class                   | *Bacilli*                                    | TAS [13,14]               |
|         | Order                   | *Bacillales*                                 | TAS [15,16]               |
|         | Family                  | *Bacillaceae*                                | TAS [15,17]               |
|         | Genus                   | *Bacillus*                                   | TAS [15,18,19]            |
|         | Species                 | *Bacillus amylo*iquefaciens*                 | TAS [20-22]               |
|         | Strain                  | UCMB5033                                     |                           |
|         | Gram stain              | Positive                                     | IDA                       |
|         | Cell shape              | Rod-shaped                                   | IDA                       |
|         | Motility                | Motile                                       | IDA                       |
|         | Sporulation             | Sporulating                                  | IDA                       |
|         | Temperature range        | Mesophilic                                   | IDA                       |
|         | Optimum temperature     | 28°C                                         | IDA                       |
|         | Carbon source           | Glucose, fructose, trehalose, mannitol, sucrose, arabinose, raffinose | IDA                       |
|         | Energy source           | --                                           |                           |
|         | Terminal electron receptor | --                                         |                           |
| MIGS-6  | Habitat                 | Soil, Host (Plant)                           | IDA                       |
| MIGS-6.3| Salinity                | up to 12% w/v                                | TAS [20,21]               |
| MIGS-22 | Oxygen                  | Aerobic                                      | IDA                       |
| MIGS-15 | Biotic relationship     | Symbiotic (beneficial)                       | TAS [2]                   |
| MIGS-14 | Pathogenicity           | None                                         | NAS                       |
| MIGS-4  | Geographic location     | Tajikistan                                   |                           |
| MIGS-5  | Sample collection time  | --                                           |                           |
| MIGS-4.1| Latitude                | --                                           |                           |
| MIGS-4.2| Longitude               | --                                           |                           |
| MIGS-4.3| Depth                   | --                                           |                           |
| MIGS-4.4| Altitude                | --                                           |                           |

<sup>a</sup> Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [23].
Genome annotation

The genome sequence was annotated using a combination of several annotation tools via the Magnifying Genome (MaGe) Annotation Platform [28]. Genes were identified using Prodigal [29] and AMIGene [30] as part of the MaGe genome annotation pipeline followed by manual curation. Putative functional annotation of the predicted protein coding genes was done automatically by MaGe after BlastP similarity searches against the Uniprot and Trembl, TIGR-Fam, Pfam, PRIAM, COG and InterPro databases. The tRNAscanSE tool [31] was used to find tRNA genes. Ribosomal RNA genes were identified using RNAmmer tool [32].

Genome properties

The B. amyloliquefaciens UCMB5033 genome consists of a circular chromosome of size 4,071,168 bp. The genome having G+C content of 46.19% were predicted to contain 4,095 predicted ORFs including 10 copies each of 16S, 23S, and 5S rRNA; 86 tRNA genes, and 3,912 protein-coding sequences with the coding density of 87.51% (Figure 2). The majority of protein coding genes (81%) was assigned putative functions while those remaining were annotated as hypothetical or conserved hypothetical proteins (Table 3). The distribution into COG functional categories is presented in Table 4.

Table 2. Genome sequencing Project information

| MIGS ID  | Property                  | Term                                                                 |
|----------|---------------------------|----------------------------------------------------------------------|
| MIGS-31  | Finishing quality         | Finished                                                             |
| MIGS-28  | Libraries used            | Illumina PE (75bp reads, insert size of 230bp), IonTorrent single end reads |
| MIGS-29  | Sequencing platforms      | Illumina GAii, IonTorrent PGM Systems                               |
| MIGS-31.2| Fold coverage             | 140× Illumina; 35× IonTorrent                                       |
| MIGS-30  | Assemblers                | MIRA 3.4 and Newbler 2.8                                            |
| MIGS-32  | Gene calling method       | PRODIGAL, AMIGene                                                    |
| ENA Project ID |                | PRJEB3961                                                           |
| Date of Release |          | September 8, 2013                                                  |
| INSDC ID |                          | HG328253                                                            |
| GOLD ID |                           | Gc0053646                                                           |
| Project relevance |                 | Biocontrol, Agriculture                                            |

Table 3. Nucleotide content and gene count levels of the genome

| Attribute                  | Value    | % of total\(^a\) |
|----------------------------|----------|-------------------|
| Genome size (bp)           | 4,071,168| 100               |
| DNA coding region (bp)     | 3,565,936| 87.5              |
| DNA G+C content (bp)       | 1,880,879| 46.1              |
| Total number of genes\(^b\)| 4095     | n/a               |
| RNA genes                  | 116      | n/a               |
| rRNA operons               | 10       | n/a               |
| Protein-coding genes       | 3912     | 100               |
| CDSS with predicted functions | 3170   | 81                |
| Uncharacterized/Hypothetical genes | 742   | 18.1              |
| CDSS assigned to COGs      | 3506     | 89.6              |
| CDSS with signal peptides  | 302      | 7.7               |
| CDSS with transmembrane helices | 1012 | 25.8              |

\(^a\) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

\(^b\) Also includes 36 pseudogenes and 66 non-coding RNA.
Figure 2. Graphical circular map of the *B. amyloliquefaciens* UCMB5033 genome. From outer to inner circle: (1) GC percent deviation (GC window - mean GC) in a 1000-bp window. (2) Predicted CDSs transcribed in the clockwise direction. (3) Predicted CDSs transcribed in the counter-clockwise direction. Red and blue genes displayed in (2) and (3) are MaGe validated annotations and automatic annotations, respectively. (4) GC skew (G+C/G-C) in a 1,000-bp window. (5) rRNA (blue), tRNA (green), non-coding RNA (orange), Transposable elements (pink) and pseudogenes (grey).
Table 4. Number of genes associated with the 25 general COG functional categories

| Code | Value | %agea | Description                                      |
|------|-------|-------|--------------------------------------------------|
| J    | 159   | 4.06  | Translation                                      |
| A    | 1     | 0.02  | RNA processing and modification                 |
| K    | 287   | 7.33  | Transcription                                   |
| L    | 141   | 10.58 | Replication, recombination and repair            |
| B    | 1     | 0.02  | Chromatin structure and dynamics                 |
| D    | 38    | 0.97  | Cell cycle control, mitosis and meiosis          |
| Y    | 0     | 0.00  | Nuclear structure                                |
| V    | 50    | 1.27  | Defense mechanisms                               |
| T    | 167   | 4.26  | Signal transduction mechanisms                   |
| M    | 196   | 5.01  | Cell wall/membrane biogenesis                    |
| N    | 63    | 1.61  | Cell motility                                    |
| Z    | 0     | 0     | Cytoskeleton                                     |
| W    | 0     | 0     | Extracellular structures                         |
| U    | 54    | 1.38  | Intracellular trafficking and secretion          |
| O    | 98    | 2.5   | Posttranslational modification, protein turnover, chaperones |
| C    | 181   | 4.62  | Energy production and conversion                 |
| G    | 270   | 6.9   | Carbohydrate transport and metabolism            |
| E    | 313   | 8     | Amino acid transport and metabolism              |
| F    | 98    | 2.5   | Nucleotide transport and metabolism              |
| H    | 145   | 3.7   | Coenzyme transport and metabolism                |
| I    | 169   | 4.32  | Lipid transport and metabolism                   |
| P    | 167   | 4.26  | Inorganic ion transport and metabolism           |
| Q    | 163   | 4.16  | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 426   | 10.88 | General function prediction only                 |
| S    | 319   | 8.15  | Function unknown                                 |
| -    | 406   | 10.37 | Not in COGs                                      |

a) The total is based on the total number of protein coding genes in the annotated genome.

Conclusion

Comparative genome analysis might reveal mechanisms by which UCMB5033 mediates plant protection and growth promotion, will further enable the investigations of the biochemical and regulatory mechanisms behind the symbiotic relationship, and will shed light on the activity of PGPR in different environments.

Acknowledgement

This work was supported by the grants from Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) and the Higher Education Commission (HEC), Pakistan. The SNP&SEQ Technology Platform and Uppsala Genome Center performed sequencing supported by Science for Life Laboratory (Uppsala), a national infrastructure supported by the Swedish Research Council (VR-RFI) and the Knut and Alice Wallenberg Foundation. The Bioinformatics Infrastructure for the Life Sciences (BILS) supported the SGBC bioinformatics platform at SLU.

http://standardsingenomics.org


**References**

1. Reva ON, Dixielius C, Meijer J, Priest FG. Taxonomic characterization and plant colonizing abilities of some bacteria related to *Bacillus amyloliquefaciens* and *Bacillus subtilis*. FEMS Microbiol Ecol 2004; 48:249-259. PubMed [http://dx.doi.org/10.1016/j.femsec.2004.02.003](http://dx.doi.org/10.1016/j.femsec.2004.02.003)

2. Danielsson J, Reva ON, Meijer J. Protection of oilseed rape (*Brassica napus*) toward fungal pathogens by strains of plant-associated *Bacillus amyloliquefaciens*. Microb Ecol 2007; 54:134-140. PubMed [http://dx.doi.org/10.1007/s00248-006-9181-2](http://dx.doi.org/10.1007/s00248-006-9181-2)

3. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25:3389-3402. PubMed [http://dx.doi.org/10.1093/nar/25.17.3389](http://dx.doi.org/10.1093/nar/25.17.3389)

4. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; 32:1792-1797. PubMed [http://dx.doi.org/10.1093/nar/gkh340](http://dx.doi.org/10.1093/nar/gkh340)

5. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011; 28:2731-2739. PubMed [http://dx.doi.org/10.1093/molbev/msr121](http://dx.doi.org/10.1093/molbev/msr121)

6. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *J Comput Biol* 2010; 17:337-354. PubMed [http://dx.doi.org/10.1089/cmb.2009.0179](http://dx.doi.org/10.1089/cmb.2009.0179)

7. Boriss R, Chen KH, Rueckert C, Blom J, Becker A, Baumgarth B, Fan B, Pukall R, Schumann P, Spröer C, et al. Relationship of *Bacillus amyloliquefaciens* clades associated with strains DSM 7T and FZB42T: a proposal for *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov. and *Bacillus amyloliquefaciens* subsp. *plantarum* subsp. nov. based on complete genome sequence comparisons. *Int J Syst Evol Microbiol* 2011; 61:1786-1801. PubMed [http://dx.doi.org/10.1099/ijss.0.023267-0](http://dx.doi.org/10.1099/ijss.0.023267-0)

8. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; 26:541-547. PubMed [http://dx.doi.org/10.1038/nbt1360](http://dx.doi.org/10.1038/nbt1360)

9. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; 87:4576-4579. PubMed [http://dx.doi.org/10.1073/pnas.87.12.4576](http://dx.doi.org/10.1073/pnas.87.12.4576)

10. Gibbons NE, Murray RGE. Proposals Concerning the Higher Taxa of *Bacteria*. *Int J Syst Bacteriol* 1978; 28:1-6. [http://dx.doi.org/10.1099/00207713-28-1-1](http://dx.doi.org/10.1099/00207713-28-1-1)

11. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), *Berger's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.

12. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), *Berger's Manual of Systematic Bacteriology*, First Edition, Volume 1, The Williams and Wilkins Co., Baltimore, 1984, p. 31-34.

13. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; 60:469-472. [http://dx.doi.org/10.1099/ijs.0.022855-0](http://dx.doi.org/10.1099/ijs.0.022855-0)

14. Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), *Berger's Manual of Systematic Bacteriology*, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 19-20.

15. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; 30:225-420. [http://dx.doi.org/10.1099/00207713-30-1-225](http://dx.doi.org/10.1099/00207713-30-1-225)

16. Prévot AR. *Bacteriol. Soc. Com. 1955, 39. In: Buchanan RE, Gibbons NE (eds), *Bergey's Manual of Determinative Bacteriology*, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 529-550.*
20. Priest FG, Goodfellow M, Shute LA, Berkeley RCW. *Bacillus amyloliquefaciens* sp. nov., nom. rev. *Int J Syst Bacteriol* 1987; 37:69-71. [http://dx.doi.org/10.1099/00207713-37-1-69](http://dx.doi.org/10.1099/00207713-37-1-69)

21. Wang LT, Lee FL, Tai CJ, Kuo HP. *Bacillus velezensis* is a later heterotypic synonym of *Bacillus amyloliquefaciens*. *Int J Syst Evol Microbiol* 2008; 58:671-675. [PubMed](http://dx.doi.org/10.1099/ijs.0.65191-0)

22. Studies on the production of bacterial amylase. I. Isolation of bacteria secreting potent amylase and their distribution. *Nippon Nogeikagaku Kaishi* 1943; 19:487-503. [http://dx.doi.org/10.1271/nogeikagaku1924.19.7_487](http://dx.doi.org/10.1271/nogeikagaku1924.19.7_487)

23. Ashburner M, Ball CA, Blake JA, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25:25-29. [PubMed](http://dx.doi.org/10.1038/75556)

24. Pagani I, Liolios K, Jansson J, Chen IM, Smimova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2012; 40:D571-D579. [PubMed](http://dx.doi.org/10.1038/75556)

25. Chevreux B, Wetter T, Suhai S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *German Conference on Bioinformatics* 1999. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.37465.

26. Manzoor S, Niazi A, Bejai S, Meijer J, Bongcam-Rudloff E. Genome Sequence of a Plant-Associated Bacterium, *Bacillus amyloliquefaciens* Strain UCMB5036. *Genome Announc.* 2013; 1.

27. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 2010; 5:e11147. [PubMed](http://dx.doi.org/10.1371/journal.pone.0011147)

28. Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, Rouy Z, Roche D, Salvignol G, Scarpelli C, Médigue C. MicroScope: a platform for microbial genome annotation and comparative genomics. *Database (Oxford)* 2009; 2009:bp021. [PubMed](http://dx.doi.org/10.1093/database/bap021)

29. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; 25:955-964. [PubMed](http://dx.doi.org/10.1093/nar/gkm160)

30. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; 35:3100-3108. [PubMed](http://dx.doi.org/10.1093/nar/gkm160)