Complete genome sequence of the molybdenum-resistant bacterium *Bacillus subtilis* strain LM 4–2

Xiao-Yan You¹, Hui Wang¹, Guang-Yue Ren¹, Jing-Jing Li¹, Xu Duan¹, Hua-Jun Zheng² and Zheng-Qiang Jiang¹,³*

**Abstract**

*Bacillus subtilis* LM 4–2, a Gram-positive bacterium was isolated from a molybdenum mine in Luoyang city. Due to its strong resistance to molybdate and potential utilization in bioremediation of molybdate-polluted area, we describe the features of this organism, as well as its complete genome sequence and annotation. The genome was composed of a circular 4,069,266 bp chromosome with average GC content of 43.83 %, which included 4149 predicted ORFs and 116 RNA genes. Additionally, 687 transporter-coding and 116 redox protein-coding genes were identified in the strain LM 4–2 genome.

**Keywords:** Gram-positive, Molybdate, Bioremediation, Molybdenum-resistance, *Bacillus subtilis* LM 4–2

**Introduction**

*Bacillus subtilis* LM 4–2 was a molybdenum-resistant strain isolated from a molybdenum mine. It has been reported that many microbes can resist the toxicity of molybdate ion though reduction of molybdate (Mo⁶⁺) to Mo-blue. Molybdenum-reducing microorganisms came from a variety of genera and included the following species, *Klebsiella* spp. [1, 2], *Acidithiobacillus ferrooxidans* [3], *Enterobacter cloacae* [4], *Serratia marcescens* [5, 6], *Acinetobacter calcoaceticus* [7], *Pseudomonas* spp. [8], and *Escherichia coli* K12 [9]. The capability of molybdate-reduction presents potential possibility of molybdenum bioremediation in many polluted areas [10]. Strain LM 4–2 showed stronger resistance to molybdate (up to 850 mM Na₂MoO₄) than many other reported molybdenum-resistant bacteria [11, 12]. However, no information related to the molecular mechanism of molybdenum-resistance has been identified, also in genus *Bacillus*. Thus, strain LM 4–2 might be a perfect subject for us to unveil the mechanism and evaluate its possibility utilization in bioremediation. Here we present the complete genome sequence and detailed genomic features of *B. subtilis* LM 4–2.

**Organism information**

**Classification and features**

*Bacillus subtilis* LM 4–2 (CGMCC 1.15213) is a Gram-positive, spore-forming, rod-shaped *Bacillus* (0.3-0.5 μm wide and 3.0–4.0 μm long) with an optimum pH 6.0 and optimum temperature of 30 °C (Table 1, Fig. 1). Colonies are milky white and matte with a wrinkled surface when growth on R2A agar medium. Strictly aerobic and catalase formed. Carbon substrates utilized for growth by strain LM 4–2 included D-glucose, maltose, lactose and sucrose. Strain LM 4–2 is closely related to *Bacillus subtilis* species based on the BLAST results of 16S rRNA gene [27]. The identity of 16S rRNA gene sequence between strain LM 4–2 and type strain *B. subtilis* DSM 10² is 100 %. A phylogenetic tree was constructed using the neighbor-joining method under the default settings for complete sequence of 16S rRNA gene derived from genome of strain LM 4–2, along with the sequences of representative members of genus *Bacillus* [28–34]. The phylogenetic tree was assessed by bootstrapped for 1000 times, which is shown in Fig. 2. Average nucleotide
Table 1 Classification and general features of *Bacillus subtilis* LM 4–2 according to the MIGS recommendations [13]

| MIGS ID | Property               | Term                      | Evidence code  |
|---------|------------------------|---------------------------|----------------|
|         | Classification         | Domain *Bacteria*         | TAS [14]       |
|         |                        | Phylum *Firmicutes*       | TAS [15–17]    |
|         |                        | Class *Bacilli*           | TAS [18, 19]   |
|         |                        | Order *Bacillales*        | TAS [20, 21]   |
|         |                        | Family *Bacillaceae*      | TAS [20, 22]   |
|         |                        | Genus *Bacillus*          | TAS [20, 23, 24]|
|         |                        | Species *Bacillus subtilis* | TAS [25]   |
|         | Gram stain             | Positive                  | IDA            |
|         | Cell shape             | Rod-shaped                | IDA            |
|         | Motility               | Motile                    | IDA            |
|         | Sporulation            | Spore-forming             | NAS            |
|         | Temperature range       | 4–45 °C                   | IDA            |
|         | Optimum temperature    | 30 °C                     | IDA            |
|         | pH range; Optimum      | 4–9; 6.0                  | IDA            |
|         | Carbon source          | organic carbon source     | IDA            |
|         | Habitat                | soil                      | IDA            |
|         | Salinity               | salt tolerant             | NAS            |
|         | Oxygen requirement     | aerobic                   | IDA            |
|         | Biotic relationship    | free-living               | NAS            |
|         | Pathogenicity          | non-pathogen              | NAS            |
|         | Geographic location    | Luoyang/Henan/China       | IDA            |
|         | Sample collection      | 2012                      | IDA            |
|         | Latitude               | 33°55'3.21"N              |                |
|         | Longitude              | 111°31'0.42"E             |                |
|         | Altitude               | 1164.78                   |                |

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 [26]

Fig. 1 Transmission electron microscopy of strain LM 4–2. Scale bar corresponds to 1.0 μm
identity (ANI), average amino acid identity (AAI) and in silico Genome-to-Genome Hybridization value (GGDH) were calculated between the genomes of strain LM 4–2 and other 30 *B. subtilis* species that have been completed sequenced [35–40]. Results show that strain LM 4–2 shares high ANI (>95 %, 23 of total 30), AAI (>95 %, 23 of total 30) and GGDH value (>70 %, 24 of total 30) with most of the complete sequenced *B. subtilis* species, and highest ANI (99.00 %), AAI (99.13 %) and GGDH value (92.20 % ± 1.85) with *B. subtilis* strain TO-A JPC (Additional file 1: Table S1).

**Genome sequencing information**

*Genome project history*

*Bacillus subtilis* LM 4–2 was selected for sequencing due to its strong resistance to molybdate and potential utilization in bioremediation of molybdate-polluted areas. The genome sequence was deposited in GenBank under accession number CP011101 and the genome project was deposited in the Genomes on Line Database [42] under Gp0112736. Genome sequencing and annotation were performed by Chinese National Human Genome Center at Shanghai. A summary of the project was given in Table 2.

**Growth conditions and genomic DNA preparation**

*Bacillus subtilis* LM 4–2 was inoculated in 200 mL R2A medium and cultivated for 8 h at 30 °C in a shaker with speed of 200 rpm. 1.2 g of harvested cells was suspended in 5 mL TE (pH8.0) with 10 mg/mL lysozyme at 30 °C.

**Table 2** Genome sequencing project information

| MIGS ID | Property | Term |
|---------|----------|------|
| MIGS 31 | Finishing quality | Complete |
| MIGS 28 | Libraries used | Two libraries, 20 Kb PacBio library, 2 × 150 bplillumina library |
| MIGS 29 | Sequencing platforms | PacBio RS II, Illumina Hi-Seq |
| MIGS 31.2 | Fold coverage | 213-and 409-fold |
| MIGS 30 | Assemblers | HGAP, bowtie2 |
| MIGS 32 | Gene calling method | Glimmer 3.02 and GeneMark |
|         | Locus Tag | BsLM |
|         | Genbank ID | CP011101 |
|         | GenBank Date of Release | April 23, 2015 |
|         | GOLD ID | Gp0112736 |
|         | BIOPROJECT | PRJNA27761 |
| MIGS 13 | Source Material Identifier | CGMCC 1.15213 |
|         | Project relevance | Environmental, Bioremediation |
for 4 h. After centrifugation (12,000 rpm) for 10 min, genomic DNA was extracted by phenol-chloroform methods as described previously [43]. DNA was dissolved in 2 mL sterilized deionized water with a final concentration of 12.67 μg/μL and 2.04 of OD260/OD280 ratio determined by NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). The genomic DNA was stored in −20 °C freezer.

**Genome sequencing and assembly**

The genome of *Bacillus subtilis* LM 4–2 was sequenced by a dual sequencing approach that using a combination of PacBio RS II and Genome Analyzer IIx sequence platforms. Approximately 121,583 PacBio and 1637 million Illumina reads were generated from PacBio platform and the Illumina platform (2 × 150 bp paired-end sequencing) with average sequence coverage of 213-and 409-fold. Sequence reads from the PacBio RS II were assembled by using hierarchical genome-assembly process assembler and finally only one self-cycled super contig was generated. The Illumina reads were quality trimmed with the CLC Genomics Workbench and then utilized for error correction of the PacBio reads by using bowtie2 (version 2.1.0) software [44].

**Table 3** Genome statistics

| Attribute                     | Value         | % of Total |
|-------------------------------|---------------|------------|
| Genome size (bp)              | 4,069,266     | 100.00     |
| DNA coding (bp)               | 3,596,010     | 88.37      |
| DNA G + C (bp)                | 1,811,637     | 44.52      |
| Total genes                   | 4265          | 100.00     |
| Protein coding genes          | 4149          | 97.28      |
| RNA genes                     | 116           | 2.72       |
| rRNA operons                  | 10            | 0.23       |
| Genes with function prediction| 2742          | 64.29      |
| Genes assigned to COGs        | 3111          | 72.94      |
| Genes with Pfam domains       | 3656          | 85.72      |
| Genes with signal peptides    | 541           | 12.68      |
| Genes with transmembrane helices | 778          | 18.24      |
| CRISPR repeats                | 0             | 0          |

**Fig. 3** Graphic representation of circular map of the chromosome of strain LM 4–2. The map was generated with the DNAPlotter [54]. From outside to the center: the first two outer circles represent the positions of genes in the chromosome (Circle 1: plus strand, Circle 2: minus strand). Circle 3 represents tRNA genes (blue), Circle 4 represents G + C content, and Circle 5 represents GC skew.
Genome annotation
The Glimmer 3.02 and GeneMark programs were used to predict the positions of open reading frames [45, 46]. Protein function was predicted by the following methods: 1) homology searches in the GenBank and UniProt protein database [47]; 2) function assignment searches in CDD database [48]; and 3) domain or motif searches in the Pfam databases [49]. The KEGG database was used to reconstruct metabolic pathways [50]. Ribosomal RNAs and Transfer RNAs were predicted by using RNAmmer and tRNAscan-SE programs [51, 52]. Transporters were predicted by searching the TCDB database using BLASTP program [27, 53] with expectation value lower than 1e-05.

Genome properties
The complete strain LM 4–2 genome was composed of a circular 4,069,266 bp chromosome with an overall 43.83 % G + C content. Four thousand one hundred forty-nine ORFs, 10 sets of rRNA operons, and 84 tRNAs were predicted in the LM 4–2 genome (Table 3 and Fig. 3). Two thousand seven hundred forty-two of total 4149 predicted ORFs could be functional assignment, 1415 were annotated as hypothetical proteins. When analyzed for biological roles according to COG categories, amino acid transport and metabolism proteins accounted for the largest percent (7.18 %) of all functionally assigned proteins, followed by carbohydrate transport and metabolism proteins (6.89 %), and Transcription proteins (6.43 %). There are 687 transporter-coding and 116 redox protein-coding genes were identified in the LM 4–2 genome. The distribution of genes into COGs functional categories is presented in Table 4.

Conclusions
Molybdenum pollution has been reported in water and soils all around the world [55]. Some Mo-resistance bacteria can be used to immobilize soluble molybdenum to insoluble forms along with reducing the toxicity. In this study we presented the complete genome sequence of Bacillus subtilis LM 4–2, which was isolated from a molybdenum mine in Luoyang city. Due to its strong resistance to molybdate and potential utilization in bioremediation of molybdate-polluted area, we sequence the genome and try to identify the possible molecular mechanism of molybdenum-resistance. Genomic analysis of strain LM 4–2 revealed 687’ transporter-coding and 116 redox protein-coding genes were separated in the genome. Three genome islands were identified in the strain LM 4–2 genome, covering 2.71 % of the whole genome. Three gene clusters were involved in the non-ribosomal synthesis of lipopeptides, such as surfactin, fengycin, and dipeptide bacilysin. Additionally, one gene clusters for subtilosin A synthesis and one gene clusters for polyketide synthesis. No CRISPRs were identified in the strain LM 4–2 genome.

The complete genome sequence of strain LM 4–2 will facilitate functional genomics to elucidate the molecular mechanisms that underlie molybdenum-resistance and it may facilitate the bioremediation of molybdenum-contaminated areas.

Additional file

Additional file 1: Table S1. The results of ANI, AAI and GGDH value between genomes of strain LM 4-2 and other 30 complete sequenced B. subtilis species. (DOC 58 kb)

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
X-YY and HW participated in the design of the study, carried out the molecular genetic studies and drafted the manuscript. G-YR and XD performed the laboratory experiments. J-JL prepared the genomic DNA. H-JZ performed the bioinformatics analysis. Z-QJ conceived of the study and helped to draft the manuscript. All authors read and approved the final manuscript.

Table 4 Number of genes associated with general COG functional categoriesa

| Code | Value | % age | Description |
|------|-------|-------|-------------|
| J    | 149   | 3.59  | Translation, ribosomal structure and biogenesis |
| A    | 0     | 0.00  | RNA processing and modification |
| K    | 267   | 6.44  | Transcription |
| L    | 114   | 2.75  | Replication, recombination and repair |
| B    | 1     | 0.02  | Chromatin structure and dynamics |
| D    | 36    | 0.87  | Cell cycle control, Cell division, chromosome partitioning |
| V    | 54    | 1.30  | Defense mechanisms |
| T    | 127   | 3.06  | Signal transduction mechanisms |
| M    | 191   | 4.60  | Cell wall/membrane biogenesis |
| N    | 60    | 1.45  | Cell motility |
| U    | 25    | 0.60  | Intracellular trafficking and secretion |
| Q    | 101   | 2.43  | Posttranslational modification, protein turnover, chaperones |
| C    | 166   | 4.00  | Energy production and conversion |
| G    | 286   | 6.89  | Carbohydrate transport and metabolism |
| E    | 298   | 7.18  | Amino acid transport and metabolism |
| F    | 82    | 1.98  | Nucleotide transport and metabolism |
| H    | 114   | 2.75  | Coenzymes transport and metabolism |
| I    | 89    | 2.14  | Lipid transport and metabolism |
| P    | 168   | 4.05  | Inorganic ion transport and metabolism |
| Q    | 72    | 1.74  | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 364   | 8.77  | General function prediction only |
| S    | 347   | 8.36  | Function unknown |
| -    | 1039  | 25.04 | Not in COGs |

*aThe total is based on the total number of protein coding genes in the annotated genome*
Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (31200035, 41201224) and by the Research Fund for the Doctoral Program of Henan University of Science and Technology under Grant (09001608). Transmission electron microscopy was provided by instrument center of Microbiology, Chinese Academy of Sciences.

Author details

1. College of Food and Bioengineering, Henan University of Science and Technology, Luoyang, P. R. China.
2. Chinese National Human Genome Center at Shanghai, Shanghai (CHGC), 201203, P. R. China.
3. College of Food Science and Nutritional Engineering, China Agricultural University, Beijing, P. R. China.

Received: 27 April 2015 Accepted: 3 December 2015

References

1. Lim HK, Syed MA, Shukor MY. Reduction of molybdate to molybdenum blue by Klebsiella sp. strain HKEEM. J Basic Microbiol. 2012;52(3):296–305.
2. Halmi MI, Zuhairiis SW, Yusof MT, Shaharuddin NA, Helmi W, Shukor Y, et al. Hexavalent molybdenum reduction to mo-blue by a sodium-dodecyl-sulfate-degrading Klebsiella oxytoca strain DRY14. Biomed Res Int. 2013;2013:384541.
3. Sugio T, Tsujita Y, Katagiri T, Inagaki K, Tanou T. Reduction of Mo(VI) with elemental sulfur by Thiothrix faciosulfurans. J Bacteriol. 1988;170(1):5956–9.
4. Shukor A, Yunus M, Lee CH, Omar I, Kinari MIA, Syed MA, et al. Isolation and characterization of a molybdenum reducing enzyme in Enterobacter cloacae strain 48. Asia Pac J Mol Biotechnol. 2003;11(2):261–72.
5. Shukor MY, Halmi MI, Rahman MFA, Shamaa NA, Syed MA. Molybdenum reduction to molybdenum blue in Senmis sp. strain DRY5 is catalyzed by a novel molybdenum-reducing enzyme. Biomed Res Int. 2014;2014:853084. doi:10.1155/2014/853084.
6. Shukor MY, Rahman MF, Shamaa NA, Syed MA. Reduction of molybdate to molybdenum blue by Enterobacter sp. strain Dr.Y13. J Basic Microbiol. 2009;49 Suppl 1:543–54.
7. Shukor MY, Rahman MF, Suhaili Z, Shamaa NA, Syed MA. Hexavalent molybdenum reduction to Mo-blue by Acinetobacter calcoaceticus. Folia Microbiol (Praha). 2010;55(2):137–43.
8. Shukor MY, Ahmad SA, Nadirz MWM, Abdullah NP, Shamaa NA, Syed MA. Molybdate reduction by Pseudomonas sp. strain DRY3. J Appl Microbiol. 2010;108(6):2050–8. doi:10.1111/j.1365-2672.2009.04604.x.
9. Campbell AM, del Campillo-Campbell A, Villaret DB. Molybdate reduction by Escherichia coli K-12 and its chi mutants. Proc Natl Acad Sci U S A. 1985;82(12):3177–8.
10. Neunhäuserer C, Berreck M, Insam H. Remediation of soils contaminated with molybdenum using soil amendments and phytoremediation. J Water Air Soil Poll. 2001;128:85–96.
11. Naeemnejad B, Kaghazchi T, Edrisi M, Sohrabi M. Bioleaching of molybdenum from low-grade copper ore. Process Biochem. 1999;35:437–40.
12. Olsson GJ, Clark TR. Biackealing of molybdenite. Hydrometallurgy. 2008;93:10–5.
13. Field D, Garry G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26:541–5.
14. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaebacteria, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87(12):4576–9.
15. Gibbons NE, Murray RG. Proposals for considering the higher taxa of bacteria. Int J Syst Bacteriol. 1978;281–6.
16. Garry GM, Holt JG. The road map to the manual. In: Boone DR, Castenholz RW, Garry GM, editors. Bergey’s manual of systematic bacteriology; Volume 1. 2nd ed. New York: Springer-Verlag, 2001. p. 119–69.
17. Murray RG. The higher taxa, or, a place for everything?...? In: Holt JG, editor. Bergey’s manual of systematic bacteriology. Volume 1. 1st ed. Baltimore: The Williams and Wilkins Co; 1984. p. 31–4.
18. List no.132. List of new names and new combinations previously effectively, but not validly published. Int J Syst Evol Microbiol. 2010;60:469–72.
19. Ludwig W, Schleifer KH, Whitman WB. Bacilli class.c. In: Voig P, Garry G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB, editors. Bergey’s manual of systematic bacteriology. Volume 3. 2nd ed. New York: Springer-Verlag; 2009. p. 19–20.
20. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol. 1980;30:225–420.
46. Borodovsky M, McIninch J. GeneMark: parallel gene recognition for both DNA strands. Computers & Chemistry. 1993;17(19):123–33.
47. UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43(Database issue):D204–D212.
48. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI’s conserved domain database. Nucleic Acids Res. 2015;43(Database issue):D222–D226.
49. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(Database issue):D222–D230.
50. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2015;43:D1070.
51. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100–8.
52. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.
53. Saier MH, Tran OV, Barabote RD. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Res. 2006;34 suppl1:D181–6.
54. Carver T, Thomson N, Bleasby A, Bertram M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics. 2009;25(1):119–20.
55. Ahmad SA, Shukor MY, Shamaan NA, Mac Cormack WP, Syed MA. Molybdate reduction to molybdenum blue by an Antarctic bacterium. BioMed Res Int. 2013;2013:871941. doi:10.1155/2013/871941.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at www.biomedcentral.com/submit