Draft genome sequence of *Arthrobacter* sp. strain B6 isolated from the high-arsenic sediments in Datong Basin, China

Linghua Xu 1,2, Wanxia Shi 1, Xian-Chun Zeng 1*, Ye Yang 1, Lingli Zhou 1, Yao Mu 1 and Yichen Liu 1

**Abstract**

*Arthrobacter* sp. B6 is a Gram-positive, non-motile, facultative aerobic bacterium, isolated from the arsenic-contaminated aquifer sediment in the Datong basin, China. This strain displays high resistance to arsenic, and can dynamically transform arsenic under aerobic condition. Here, we described the high-quality draft genome sequence, annotations and the features of *Arthrobacter* sp. B6. The G + C content of the genome is 64.67%. This strain has a genome size of 4,663,437 bp; the genome is arranged in 8 scaffolds that contain 25 contigs. From the sequences, 3956 protein-coding genes, 264 pseudo genes and 89 tRNA/rRNA-encoding genes were identified. The genome analysis of this strain helps to better understand the mechanism by which the microbe efficiently tolerates arsenic in the arsenic-contaminated environment.

**Keywords:** *Arthrobacter* sp. B6, Genome, Arsenate reduction, High-arsenic sediment, Datong basin

**Organism information**

**Classification and features**

*Arthrobacter* sp. B6 is a Gram-positive, non-motile, facultative aerobic bacterium. Cells are straight or slightly curved rods during log phase of bacterial growth (Fig. 1) and become coccoid in stationary phase. The bacteria cells formed white colonies on 0.1× Trypticase Soy Broth agar plate. Colonies are convex and circular with entire margin. The strain can grow at a wide range of temperatures from 4 to 37 °C; the optimum is 30 °C. It can proliferate in a pH range of 6.0–8.5; the optimum is 7.0. The strain tolerates high concentrations of NaCl up to approximately 7% (Table 1). It is catalase- and oxidase-positive. It hydrolyzes starch and tyrosine, but not o-nitrophenyl-β-d-galactoside, gelatin, aesculin, chitin, casein or cellulose. It is negative for nitrate reduction, H2S production, citrate utilization, indole production, arginine dihydrolase and urease activity.

*Correspondence:* xianchun_zeng@hotmail.com; xianchun.zeng@gmail.com

1 State Key Laboratory of Biogeology and Environmental Geology & Department of Biological Science and Technology, School of Environmental Studies, China University of Geosciences (Wuhan), Wuhan, China

© The Author(s). 2017 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Fig. 1 Images of Arthrobacter sp. B6 using scanning electron microscopy (Left) and the appearance of colony morphology on 0.1× Trypticase Soy Broth solid media (Right)

Table 1 Classification and general features of Arthrobacter sp. B6 [19]

| MIGS ID | Property            | Term                                                                 | Evidence code |
|---------|---------------------|----------------------------------------------------------------------|---------------|
|         | Classification      | Domain Bacteria                                                     | TAS [24]      |
|         | Phylum              | Actinobacteria                                                      | TAS [25]      |
|         | Class               | Actinobacteria                                                      | TAS [26]      |
|         | Order               | Actinomycetales                                                     | TAS [27, 28]  |
|         | Family              | Micrococcaceae                                                      | TAS [27, 29]  |
|         | Genus               | Arthrobacter                                                        | TAS [1, 2]    |
|         | Species undetermined| Strain: B6                                                           |               |
|         | Gram stain          | Positive                                                            | IDA           |
|         | Cell shape          | Polymorphic: rod to coccus shaped                                    | IDA           |
|         | Motility            | Non-motile                                                          | IDA           |
|         | Sporulation         | Non-sporulating                                                     | IDA           |
|         | Temperature range   | 4–37 °C                                                             | IDA           |
|         | Optimum temperature | 30 °C                                                               | IDA           |
|         | pH range; Optimum   | 6.0–8.5; 7                                                          | IDA           |
|         | Carbon source       | Dextrin, Tween 40, D-fructose, Gentiobiose, α-D-glucose, Lactulose,  | IDA           |
|         |                     |  D-mannose, D-mannitol, D-melezitose, Palatinose, D-psicose, D-raffinose,  |               |
|         |                     | L-rhamnose, D-ribose, D-sorbitol, Sucrose, Turanose, α- hydroxybutyric acid,  |               |
|         |                     | α-ketoglutaric acid, L-malic acid, Pyruvic acid, D-alanine, L-alanine, L-serine, Glycerol,  |               |
|         |                     | Adenosine, 2-deoxy adenosine, Inosine.                               |               |

*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 [30]
The strain utilizes dextrin, tween 40, D-fructose, gentiobiose, α-D-glucose, lactulose, maltotriose, D-mannose, D-mannitol, D-melezitose, palatinose, D-psicose, D-raffinose, L-rhamnose, D-ribose, D-sorbitol, sucrose, turannose, α-hydroxybutyric acid, α-ketoglutaric acid, L-malic acid, pyruvic acid, D-alanine, L-alanine, L-serine, glycerol, adenosine, 2-deoxy adenosine and inosine as tested using the Biolog GP2 microplate system. The major fatty acids of strain B6 are anteiso-C15:0 (56.58%), anteiso-C17:1ω9c (8.89%), anteiso-C17:0 (8.22%), iso-C15:0 (7.63%), iso-C16:0 (5.26%), sum in feature 3 (4.31%), summed feature 3 (containing C16:1ω6c and/or C16:1ω7c) (4.31%) and iso-C16:1 H (2.32%). These data suggested that the morphological and biochemical traits and fatty acid profile of B6 are consistent with those of other described species of the genus Arthrobacter.

The 16S rRNA gene sequence of strain B6 shares 94.67–99.59% identities with those of other known species of the genus Arthrobacter. In order to evaluate the evolutionary relationships between B6 and other known

![Phylogenetic tree](image-url)
strains of the genus *Arthrobacter*, the 16S rRNA gene sequence of all of these bacteria were aligned using ClustalW [17], and a phylogenetic tree was conducted using the maximum-likelihood and neighbor-joining algorithms implemented in MEGA 6.0, respectively [18]. The phylogeny illustrated that the strain B6 is closely associated with *Arthrobacter oryzae*, *A. globiformis*, *A. pascens* and *A. humicola*; suggesting that B6 is affiliated with the genus *Arthrobacter* (Fig. 2). We also found that *Arthrobacter* sp. B6 showed high resistance to arsenic, with maximal inhibitory concentrations of 150.0 mM for arsenate and 5.0 mM for arsenite. A dynamic transformation of arsenic catalyzed by strain B6 was observed when it was cultured aerobically with arsenate.

**Genome sequencing information**

**Genome project history**

*Arthrobacter* sp. strain B6 was selected for sequencing on the basis of its high resistance to arsenic and dynamic arsenic transformation capability. The Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank database under the accession number LQAP00000000. A summary of the main project information on compliance with MIGS version 2.0 is shown in Table 2 [19].

**Growth conditions and genomic DNA preparation**

Strain B6 was grown at 30 °C in 0.1× Trypticase Soy Broth liquid medium to mid-exponential phase. Genomic DNA was extracted from 0.5 to 1.0 g of cells using the modified method of Marmur [20]. The purity of DNA, expressed as the value of A260/A280, was assessed on a NanoDrop™ ND-1000 Spectrophotometer (Biolab).

**Genome sequencing and assembly**

The draft genome of *Arthrobacter* sp. B6 was sequenced at the Beijing Genomics Institute (BGI, Shenzhen) using

| Table 2 | Project information |
|---------|---------------------|
| MIGS 31 | Finishing quality    |
| MIGS-28 | Libraries used       |
| MIGS 29 | Sequencing platforms |
| MIGS 31:2| Fold coverage       |
| MIGS 30 | Assemblers           |
| MIGS 32 | Gene calling method  |
| Locus Tag | Illumina HiSeq 2000 |
| Genbank ID | Illumina HiSeq 2000 |
| GenBank Date of Release | Jun 15, 2016 |
| GOLD ID | Gs0118476            |
| BIOPROJECT | PRJNA306410 |
| MIGS 13 | Source Material Identifier |

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

| Table 3 | Genome statistics |
|---------|-------------------|
| Attribute | Value | % of Total |
| Genome size (bp) | 4,663,437 | 100.00 |
| DNA coding (bp) | 4,100,739 | 87.93 |
| DNA G+C (bp) | 3,015,845 | 64.67 |
| DNA scaffolds | 8 | 100.00 |
| Total genes | 4309 | 100.00 |
| Protein coding genes | 3956 | 91.81 |
| RNA genes | 89 | 2.07 |
| Pseudo genes | 264 | 6.12 |
| Genes in internal clusters | 4250 | 98.63 |
| Genes with function prediction | 3527 | 81.85 |
| Genes assigned to COGs | 2210 | 51.29 |
| Genes with Pfam domains | 3464 | 80.39 |
| Genes with signal peptides | 220 | 5.11 |
| Genes with transmembrane helices | 249 | 5.78 |
| CRISPR repeats | 125 | 2.90 |

| Table 4 | Number of genes associated with general COG functional categories |
|---------|----------------------------------------------------------|
| Code   | Value | %age | Description                                  |
| J       | 145   | 6.56 | Translation, ribosomal structure and biogenesis |
| A       | 1     | 0.05 | RNA processing and modification              |
| K       | 162   | 7.33 | Transcription                               |
| L       | 110   | 4.98 | Replication, recombination and repair         |
| B       | 1     | 0.05 | Chromatin structure and dynamics             |
| D       | 12    | 0.54 | Cell cycle control, Cell division, chromosome partitioning |
| V       | 26    | 1.18 | Defense mechanisms                          |
| T       | 58    | 2.62 | Signal transduction mechanisms               |
| M       | 72    | 3.26 | Cell wall/membrane biogenesis               |
| N       | 0     | 0    | Cell motility                               |
| U       | 18    | 0.81 | Intracellular trafficking and secretion      |
| O       | 65    | 2.94 | Posttranslational modification, protein turnover, chaperones |
| C       | 168   | 7.60 | Energy production and conversion            |
| G       | 225   | 10.18| Carbohydrate transport and metabolism       |
| E       | 272   | 12.31| Amino acid transport and metabolism         |
| F       | 71    | 3.21 | Nucleotide transport and metabolism         |
| H       | 111   | 5.02 | Coenzyme transport and metabolism           |
| I       | 103   | 4.66 | Lipid transport and metabolism              |
| P       | 127   | 5.75 | Inorganic ion transport and metabolism      |
| Q       | 66    | 2.99 | Secondary metabolites biosynthesis, transport and catabolism |
| R       | 266   | 12.04| General function prediction only             |
| S       | 131   | 5.93 | Function unknown                            |
| -       | 2099  | 48.71| Not in COGs                                |
the high throughout sequencing technique. A standard Illumina shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform; this generated 8,355,450 clean reads totaling 752 Mbp. These reads were assembled using the Short Oligonucleotide Analysis Package (SOAPdenovo v2.04) with all parameters set to default [21]. The final draft assembly contains 25 contigs in 8 scaffolds. Final assembly was based on all clean reads that provide an average of 161-fold coverage of the genome. The total size of the genome is 4.66 Mbp.

**Genome annotation**
Genes were identified using Glimmer v3.02 [22]. The predicted CDSs were translated into amino acid sequences that were used as queries to BLAST the GenBank, Swissprot, InterPro, KEGG, COG and GO databases, respectively. These data were combined to assert a product description for each predicted protein. Additional gene prediction analysis and functional annotation was performed using the Integrated Microbial Genomes-Expert Review (IMG-ER) platform [23].

**Genome properties**
The assembly of the draft genome sequence consists of 8 scaffolds amounting to 4,663,437 bp. The G + C content is 64.67% (Table 3). From the genome, 4309 genes were predicted, of which 3956 are protein-coding genes. Among these protein-coding genes, 154 were assigned to putative functions, and 275 were annotated as hypothetical proteins. The assignment of genes into COGs functional categories is presented in Table 4 and Fig. 3.

![Fig. 3 A graphical circular map of the genome performed with CGview comparison tool](image-url)
**Insights from the genome sequence**

Genome comparison using the RAST Prokaryotic Genome Annotation Server revealed that the genome sequence of *Arthrobacter* sp. B6 is most similar to that of *Arthrobacter* sp. FB24 (comparison score: 536), but less similar to those of other *Arthrobacter* strains. *Arthrobacter* sp. B6 shares 2035, 2011, 1958, 1930, 1850 and 1829 genes with the strains *A. globiformis* NBRC 12137, *Arthrobacter* sp. FB24, *A. enclensis* NIO-1008, *A. nitrophilicus* SJCon, *A. castelli* DSM 16402 and *A. crystalllopiaetes* BAB-32, respectively.

A three-gene (*arsR-acr3-arsC*) operon involved in the regulation of arsenate tolerance and reduction was identified from the genome of *Arthrobacter* sp. B6. The putative arsenate reductase (*ArsC*) of strain B6 shows 96% and 95% sequence identities to those of *Arthrobacter* sp. Leaf137 and *Pseudarthrobacter phanenthrenivorans* Sphe3, respectively. It also shows 89% identities to those of *A. globiformis* NBRC 12137, *A. nitrophilicus* SJCon, *A. enclensis* NIO-1008 and *Arthrobacter* sp. FB24, respectively. The amino acid sequence of ACR3 displays 85% identity to that of the arsenic transporter from *Arthrobacter* sp. FB24. Numerous genes responsible for tolerance or detoxification of metals were identified from the genome of *Arthrobacter* sp. B6, including copper resistance protein CopC and CopD, copper chaperone, copper-translocating P-type ATPase, cobalt-zinc-cadmium resistance protein CzcD, mercuric reductase, DNA gyrase subunit A and B involved in fluoroquinolones resistance, various polyols ABC transporter and DedA protein involved in the uptake of selenate and selenite. In addition, there are some genes in the genome responsible for osmotic stress. The high tolerance of salt (7% NaCl) of strain B6 may be explained by the presence of glycine betaine ABC transport system permease protein in the genome.

**Conclusions**

In the present study, we characterized the genome of *Arthrobacter* sp. B6 that was isolated from the arsenic-contaminated aquifer sediment in the Datong Basin, China. It contains numerous genes involved in heavy metal tolerance and detoxification. The knowledge of the genome sequence of *Arthrobacter* sp. B6 lays foundation for better understanding of the special metabolic abilities of the strain and for elucidation of the metabolic diversity of bacteria inhabiting in the high-arsenic environment. Further functional analyses of the identified genes may gain insights into the detailed molecular mechanisms by which the microbes tolerate and transform arsenic in the arsenic-contaminated environments.

**Abbreviations**

ABC: ATP-binding cassette; ACR3: Arsenite transporter; ArsC: Arsenate reductase; ArsR: Arsenite responsive repressor; BLAST: Basic local alignment search tool; CDS: Coding DNA sequence; CRISPR: Clustered regularly interspaced short; DedA: Integral membrane protein; IMG-ER: Integrated Microbial Genomes-Expert Review; MIGS: Minimum information on the genome sequence

**Acknowledgements**

This work was financially supported by the National Natural Science Foundation of China (grants nos. 41272257, 41472219, 41072181 and 41521001), and the Research Projects of the Educational Commission of Hubei Province of China (grant no. Q20154401).

**Authors’ contributions**

LHX performed laboratory experiments, analyzed the data and wrote the draft manuscript. YY and YM cultured the bacterial cells. WXS, LLZ and YCL analyzed the data and revised the manuscript. XCZ revised the manuscript and provided financial supports. All authors read and approved the final manuscript.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1State Key Laboratory of Biogeology and Environmental Geology & Department of Biological Science and Technology, School of Environmental Studies, China University of Geosciences (Wuhan), Wuhan, China. 2School of Chemistry and Chemical Engineering, Hubei Polytechnic University, Huangshi, China.

**Received:** 4 July 2016 **Accepted:** 12 January 2017 **Published online:** 23 January 2017

**References**

1. Conn HJ, Dimmick I. Soil Bacteria Similar in Morphology to *Mycobacterium* and *Corynebacterium*. J Bacteriol. 1947;54(3):291–303.
2. Busse HJ. Review of the taxonomy of the genus *Arthrobacter*, emendation of the genus *Arthrobacter sensu lato*, proposal to reclassify selected species of the genus *Arthrobacter* in the novel genera *Glutamicibacter* gen. nov., *Paeniglutamicibacter* gen. nov., *Pseudoglutamicibacter* gen. nov., *Paenarthrobacter* gen. nov. and *Pseudarthrobacter* gen. nov. Int J Syst Evol Microbiol. 2016;66(1):119–37.
3. Stackebrandt E, Schumann P. Introduction to the taxonomy of Actinobacteria. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors. The Prokaryotes. New York: Springer; 2006. p. 297–321.
4. Nakatsu CH, Barabote R, Thompson S, Bruce D, Detter C, Brettin T, et al. Complete genome sequence of *Arthrobacter* sp. strain FB24. Stand Genomic Sci. 2013;9(1):106–16.
5. Arora PK, Sharma A. New metabolic pathway for degradation of 2-nitrobenzoate by *Arthrobacter* sp. SPG. Front Microbiol. 2015;6:551.
6. Ren L, Shi Y, Jia Y, Yan Y. Genome Sequence of *Arthrobacter* sp. YC-RL1, an Aromatic Compound-Degrading Bacterium. Genome Announc. 2015; 3(4):e00749–15.
7. Overhage J, Siefker S, Homburg S, Parschat K, Fetzner S. Identification of large linear plasmids in *Arthrobacter* spp. encoding the degradation of quinaldine to anthranilate. Microbiology. 2005;151:491–500.
8. Isbella C, Daniela L. *Arthrobacters*: successful arid soil bacteria: a review. Arid Land Res Manag. 1987;1:1–30.
9. Crocker FH, Fredrickson JK, White DC, Ringelberg DB, Balkwill DL. Phylogenetic and physiological diversity of *Arthrobacter* strains isolated from unconsolidated subsurface sediments. Microbiology. 2000;146:1295–310.
10. Hanbo Z, Changqun D, Qiyong S, Weimin R, Tao S, Lizhong C, et al. Genetic and physiological diversity of phylogenetically and geographically distinct groups of *Arthrobacter* isolated from lead-zinc mine tailings. FEMS Microbiol Ecol. 2004;49:333–41.
rhizosphere soils of metal-tolerant Chenopodium ambrosioides grown on lead-zinc mine tailings. Chemosphere. 2012;87:1171–8.
14. Doouza M, Taylor MW, Turner SJ, Aislabie J. Genomic and phenotypic insights into the ecology of Arthrobacter from Antarctic soils. BMC Genomics. 2015;16:36.
15. Fredrickson JK, Zachara JM, Balkwell DL, Kennedy D, Li SM, Kostandarithes HM, et al. Geomicrobiology of high-level nuclear waste-contaminated vadose sediments at the Hanford site, Washington state. Appl Environ Microbiol. 2004;70:4230–41.
16. He J, Charlet L. A review of arsenic presence in China drinking water. J Hydrol (Amst). 2013;492(10):79–88.
17. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(2):2947–8.
18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.
19. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26(5):541–7.
20. Marmur J, Doty P. Thermal Renaturation of Deoxyribonucleic Acids. J Mol Biol. 1961;3(5):585–94.
21. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 2012;1:18.
22. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics. 2007;23(6):673–9.
23. Markowitz VM, Chen IW, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 2012;40:D115–22.
24. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87(12):4576–9.
25. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey’s Manual of Systematic Bacteriology, vol. 1. Second ed. New York: Springer; 2001. p. 119–69.
26. Stackebrandt E, Rainey F, Ward-Rainey N. Proposal for a new hierarchic classification system, Actinobacteria classis nov. Int J Syst Bacteriol. 1997;47:479–91.
27. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. Am Soc Microbiol. 1980;30(1):225–420.
28. Buchanan RE. Studies in the nomenclature and classification of bacteria: II. The primary subdivisions of the Schizomycetes. J Bacteriol. 1917;2(2):155–64.
29. Pribram E. A contribution to the classification of microorganisms. J Bacteriol. 1923;18(6):361–94.
30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology tool for the unification of biology. Nat Genet. 2000;25:25–9.
31. Grant JR, Arantes AS, Stothard P. Comparing thousands of circular genomes using the CGView Comparison Tool. BMC Genomics. 2012;13:202.