Whole genome sequencing of *Microbacterium* sp. AIS03 from polluted San Jacinto River sediment reveals high bacterial mobility, metabolic versatility and heavy metal resistance

Rupa Iyer\textsuperscript{a}\textsuperscript{⁎}, Ashish Damania\textsuperscript{b}, Brian Iken\textsuperscript{a}

\textsuperscript{a} Center for Life Sciences Technology, Department of Engineering Technology, University of Houston, Houston, TX, USA
\textsuperscript{b} Department of Tropical Medicine-Pediatrics, Baylor College of Medicine, Houston, TX, USA

**A B S T R A C T**

The genus *Microbacterium* is composed of high GC content, Gram-positive bacteria of the phylum *Acintobacteria* known for their antibiotic production. *Microbacterium* species commonly colonize agricultural rhizospheres and more infrequently have been found to colonize and infect human tissues as well. Here we report the 3,696,310 bp draft genome (chromosome and plasmids) sequence assembled at the scaffold level from 232 contigs of *Microbacterium* sp. strain AIS03, isolated from polluted San Jacinto River sediment in Channelview, Texas. The nucleotide sequence of this genome was deposited into NCBI GenBank under the accession NHRF00000000.

**Specifications**

| Organism/cell line/tissue | *Microbacterium* sp. |
|---------------------------|---------------------|
| Strain                    | AIS03               |
| Sequencer or array type   | Illumina Miseq      |
| Data format               | Draft genome (Scaffolds) |
| Experimental factors      | Bacterial strain    |
| Experimental features     | Whole genome analysis and gene annotation of AIS03 |
| Sample source location    | San Jacinto River sediment near the Battleship Texas Historic Site in Channelview, Texas |
| GPS coordinates           | 29° 45′ 20.916″ N, 95° 5′ 25.2564″ W |

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/nuccore/NHRF00000000.

2. Experimental design, materials and methods

Polluted sediment containing a heterogenous mixture of polychlorinated dioxins, furans, biphenyls, petroleum hydrocarbons, and agricultural waste was collected from the Battleship Texas Historic Site in Channelview, Texas along the banks of the San Jacinto River [1]. Selective media was prepared to screen for microorganisms within this sample with the capacity to tolerate, degrade and/or metabolize organophosphate compounds. For this purpose utilized Carbon Selective Media (CSM) which has a composition of 2 mM NTA, 0.8 mM MgSO\(_4\)·7H\(_2\)O, 0.17 mM Ca(NO\(_3\))\(_2\), 0.018 mM FeSO\(_4\)·7H\(_2\)O, 20% v/v Phosphate Buffer. A total of 5 mL of CSM media was aliquoted into culture tubes with 100 μg/mL ethyl paraoxon as a screening agent. These tubes were prepared fresh each week for each new subculture set for a period of five weeks. The culture was then diluted into minimal media with glycerol added as a supplementary carbon source and plated onto an agar plate with 100 μg/mL ethyl paraoxon. A single orange colored bacterium was isolated from the agar plate and shipped to Genewiz (South Plainfield, NJ), where library construction and whole genome sequencing of the bacterium was performed as described below.

Samples were visually inspected upon receipt and genomic DNA was extracted from bacterial colonies using the PureLink Genomic DNA extraction kit as per manufacturer’s protocols. The resulting genomic DNA was quantified using both the Nanodrop and the Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). A total of 50–60 ng of each sample was run on a 0.6% agarose gel to check for quality. The Illumina Nextera XT library preparation, clustering, and sequencing reagents were used throughout the process following the...
manufacturer’s recommendations (Illumina, San Diego, CA, USA). DNA libraries were analyzed on the Agilent TapeStation (Agilent Technologies, Palo Alto, CA, USA) and quantified using the Qubit 2.0 Fluorometer. The DNA libraries were quantified by real time PCR (Applied Biosystems, Carlsbad, CA, USA), and multiplexed in equal molar mass.

DNA libraries were multiplexed and loaded on an Illumina MiSeq instrument according to manufacturer’s instructions (Illumina, San Diego, CA, USA). Sequencing was performed using a 2 × 250 paired-end (PE) configuration (Table 1). Sequence reads were checked for quality using Fastqc [2] and filtered using BBTools [3] with minimum Phred score of 20. Paired-end reads were assembled into contigs with the Spades 3.10.1 program [4]. The Mash program [5] was used for species identification using k = 21 and sketch size of 1000 against the Mash Refseq (release 70) database. Fasta files for the five top bacterial hits sorted by distance were downloaded from RefSeq database and used to calculate the Mash distance. The resulting distance file from the previous step was imported into R using the reader package [6]. Finally, the Ggdendrogram [7] package was used to create a dendrogram plot through hclust function output using UPGMA method (Fig. 1). The Quast program [8] was used to calculate assembly statistics using scaffold mode. Preliminary reference based annotation using PATRIC [9] web resources was carried out to identify conserved pathways. Final de novo annotation was performed through the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (http://www.ncbi.nlm.nih.gov/ genomes/static/Pipeline.html) and the Rapid Annotation System Technology (RAST) server [10,11].

3. Data description

Microbacterium species are versatile, ubiquitous microorganisms capable of thriving in both endaphic and aquatic environments. More commonly known for their role in plant promotion, Microbacterium has also recently emerged as a rare opportunistic human pathogen [12–16]. The draft genome of Microbacterium sp. AISO3 includes 232 contigs with a GC content of 69.82%, consists of 3,478,976 bp and 2 plasmids totaling 217,334 bp. Combined, the draft genome and plasmid of AISO3 contains approximately 3623 gene sequences, 3563 coding sequences including 62 pseudogenes, 8 rRNA genes, 48 tRNAs, and 4 noncoding RNA (ncRNA) genes. The closest neighboring genome to strain AISO3 is Microbacterium sp. TS-1 (See Fig. 1). An overview of genome subsystem features (Fig. 2) shows a microorganism that has not only retained high metal resistance similar to other Acintobacteria and plant promoting rhizosphere bacteria, but also the capacity for opportunistic pathogenicity through a combination of virulence and antibiotic resistance factors including multidrug efflux, β-lactamase activity, Type VII secretion systems, and a Mycobacterium-like virulence operon. In addition, Microbacterium sp. AISO3 notably encodes for a greater number of motility, chemotaxis, and carbohydrate metabolism genes than other

| Table 1 Microbacterium sp. AISO3 genome statistics. |
|---------------------------------|---------------------------------|
| Assembly statistics             |                                  |
| platform                        | Illumina MiSeq (2*250) paired end |
| total raw reads                 | 3,125,578                        |
| total filtered reads            | 2,847,047                        |
| genome size (bp)                | 3,696,310                        |
| Contigs (≥ 200 bp)              | 230 (chromosome) + 2 (plasmids)  |
| Contigs ≥ 500 bp                | 55                               |
| Largest contig                  | 598,210                          |
| N50                             | 294,906                          |
| L50                             | 4                                |
| Average coverage                | 282.44                           |
| Annotation statistics           |                                  |
| GC content (%)                  | 69.82 (chromosome) + 68.33(plasmid) |
| total genes                     | 3623                             |
| coding genes                    | 3563                             |
| rRNAs                           | 8                                |
| tRNAs                           | 48                               |

Fig. 1. Dendrogram of Microbacterium sp. AISO3 and the five closest neighboring genomes.
Microbacterium strains (Table 2) [17,18].

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

Whole genome sequencing of Microbacterium sp. Strain AISO3 was funded through NIST (G110008 58106).

Table 2
Number of genes identified in representative Microbacterium species. Categories in which the greatest number of genes was identified in strain AISO3 are indicated in bolded text.

| Category                | Strain AISO3 | Strain BH-3-3-3 | Strain TS-1 |
|-------------------------|--------------|-----------------|-------------|
| Cell wall/capsule       | 72           | 70              | 52          |
| Virulence               | 51           | 35              | 47          |
| Amino acids             | 321          | 313             | 232         |
| Carbohydrates           | 175          | 188             | 143         |
| Nucleosides/nucleotides | 95           | 99              | 59          |
| Carbohydrates           | 392          | 355             | 325         |
| Fatty acids/lipids      | 135          | 136             | 115         |
| Sulfur                  | 22           | 8               | 18          |
| Phosphorus              | 36           | 31              | 35          |
| Protein metabolism      | 190          | 225             | 143         |
| DNA metabolism          | 88           | 69              | 56          |
| RNA metabolism          | 86           | 78              | 55          |
| Aromatic metabolism     | 17           | 51              | 16          |
| Membrane transport      | 132          | 100             | 121         |
| Cell signaling          | 22           | 31              | 16          |
| Cell mobility/chemotaxis| 49           | 8               | 42          |
| Stress response         | 74           | 69              | 72          |
| Respiration             | 40           | 48              | 40          |
| Cell division/cell cycle| 23           | 22              | 20          |
| Miscellaneous           | 30           | 32              | 27          |

Representative Microbacterium strains (Table 2) [17,18].

References

[1] R. Iyer, J. Aggarwal, B. Iken, A review of the Texas, USA San Jacinto Superfund site and the deposition of polychlorinated dibenzop-dioxins and dibenzofurans in the San Jacinto River and Houston Ship Channel, Environ. Sci. Poll. Res. 23 (2016) 23321–23338.

[2] B. Bioinformatics, FastQC a quality control tool for high throughput sequence data, Babraham Institute, Cambridge, United Kingdom, 2011http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[3] B. Bushnell, BBMap Short Read Aligner.http://sourceforge.net/projects/bbmap, (2016).

[4] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lenin, S.I. Nikolaenko, S. Pham, A.D. Pijlskei, A.V. Pshikh, A.V. Sirotnik, N. Vyahhi, G. Tesler, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477.

[5] B.D. Ondov, T.J. Treangen, P. Melsted, A.B. Mallonee, N.H. Bergman, A.M. Phillippy, Mash: fast genome and metagenome distance estimation using minhash, Genome Biol. 17 (1) (2016) 132.

[6] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics 29 (8) (2013) 1072–1075.

[7] H. Wickham, J. Hester, R. Francois, Readr: read tabular data.https://CRAN.R-project.org/package=readr, (2016).

[8] A. de Vries, B.D. Ripley, ggdendro: create dendrograms and tree diagrams using ‘ggplot2’. R package version 0.1–20. https://CRAN.R-project.org/package=ggdendro, (2016).

[9] A.R. Wattam, D. Abraham, O. Dalay, T.L. Dize, T. Driscoll, J.L. Disz, T. Driscoll, J.J. Gillespie, R. Gough, D. Hix, R. Kenyon, D. Machi, C. Mao, E.K. Nordberg, R. Olson, R. Overbeek, G.D. Pusch, M. Shukla, J. Schulman, R.L. Steven, D.E. Sullivan, V. Vonstein, A. Warren, R. Will, M.J.C. Wilson, Y.H. Seung, C. Zhang, Y. Zhang, B.W. Sobral, PATRIC, the bacterial bioinformatics database and analysis resource, Nucleic Acids Res. 42 (2014) D581–D591.

[10] T. Tatusova, M.J. DiCuccio, A. Badretlin, V. Chetvernin, S. Ciufo, W. Li, Prokaryotic genome annotation pipeline, The NCBI Handbook, second ed., National Center for Biotechnology Information, Bethesda, MD, 2013.

[11] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, The RAST server: rapid annotations using subsystems technology, BMC Genomics 9 (2008) 75.

[12] Y. Ma, M.N.V. Prasad, M. Rajkumar, H. Freitas, Plant growth promoting rhizobacteria and endophytes accelerate phytoremediation of metalliferous soils, Biotechnol. Adv. 29 (2) (2011) 248–258.

[13] M. Madhaiyan, S. Poonguzhali, J.S. Lee, K.C. Lee, V.S. Saravanan, P. Santhanakrishnan, Microbacterium azadirachtae sp. nov., a plant-growth-promoting actinobacterium isolated from the rhizoplane of neem seedlings, Int. J. Syst. Evol. Microbiol. 60 (7) (2010) 1687–1692.
Gaubert, J.M. Rolain, Phenotypic and genotypic properties of *Microbacterium yannicii*, a recently described multidrug resistant bacterium isolated from a lung transplanted patient with cystic fibrosis in France, BMC Microbiol. 13 (1) (2013) 97.

[15] E. Gallois, T. Lamy, M. Fines-Guyon, T. Lobbedez, V. Cattoir, Recurrent peritoneal dialysis-related peritonitis caused by *Microbacterium resistens*, Diagn. Microbiol. Infect. Dis. 79 (1) (2014) 111–113.

[16] S.N. Buss, R. Starlin, P.C. Iwen, Bacteremia caused by *Microbacterium binotii* in a patient with sickle cell anemia, J. Clin. Microbiol. 52 (1) (2014) 379–381.

[17] S. Fujinami, K. Takeda, T. Onodera, K. Sato, M. Sano, I. Narumi, M. Ito, Draft genome sequence of sodium-independent alkaliphilic *Microbacterium* sp. strain TS-1, Genome Announc. 1 (6) (2013) e01043-13.

[18] M.W. Dees, M.B. Brurberg, E. Lysøe, Complete genome sequence of the biofilm-forming *Microbacterium* sp. strain BH-3-3-3, isolated from conventional field-grown lettuce (Lactuca sativa) in Norway, Genom. Data 11 (2017) 7–8.