Genomics Data 5 (2015) 38–39

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Genome sequencing and annotation of *Aeromonas* sp. HZM

Patric Chua *, Zi Mei Har, Christopher M. Austin, Catherine M. Yule, Gary A. Dykes, Sui Mae Lee

School of Science, Monash University, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 28 April 2015
Received in revised form 9 May 2015
Accepted 10 May 2015
Available online 19 May 2015

Keywords:
Tropical peat swamp
Aeromonas
Cellulolytic
Whole genome sequencing

ABSTRACT

We report the draft genome sequence of *Aeromonas* sp. strain HZM, isolated from tropical peat swamp forest soil. The draft genome size is 4,451,364 bp with a G + C content of 61.7% and contains 10 rRNA sequences (eight copies of 5S rRNA genes, single copy of 16S and 23S rRNA each). The genome sequence can be accessed at DDBJ/EMBL/GenBank under the accession no. JEMQ00000000.

Specifications	
Organism/cell line/tissue	*Aeromonas* sp.
Strain(s)	HZM
Sequencer or array type	Sequencer; Illumina MiSeq
Data format	Processed
Experimental factors	Microbial strains
Experimental features	Draft genome sequence of *Aeromonas* sp. HZM, assembly and annotation
Consent	N/A
Sample source location	Tropical peat swamp in Pekan, Pahang, Malaysia

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/bioproject/236450.

2. Experimental design, materials and methods

Aeromonas sp. strain HZM is a Gram negative, rod-shaped bacterium isolated from a soil sample taken at 50 cm depth in Pekan tropical peat swamp forest in Pahang, Malaysia. The isolate was acquired through a series of cellulose-enrichment steps followed by culture-plating on Sizova's minimal salt media [1] with carboxymethyl-cellulose (CMC) as the sole carbon source.

Genomic DNA was extracted from overnight cultures using the GF-1 nucleic acid extraction kit (Vivantis, Malaysia) according to the manufacturer's protocol, and subsequently sequenced using an Illumina MiSeq sequencer (150-bp paired-end reads). Raw reads were trimmed and assembled de novo using CLC Genomics Workbench 6 (CLC Bio, Denmark) to obtain 121 contigs with a total length of 4,451,364 bp (87-fold coverage, N_{50} = 82,273 bp). Predictions using tRNAscan 1.2 [2] and RNAmmer 1.2 [3] revealed 131 tRNAs, 8 copies of 5S rRNA, and a single copy for 16S rRNA and 23S rRNA, respectively. The G + C content for the draft genome is 61.7%.

Functional annotation of the genome sequences was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) available at the Rapid Annotation using Subsystems Technology (RAST) server [4] and resulted in a total of 4072 protein-coding genes for *A.* sp. HZM being identified (Fig. 1). A gene encoding endoglucanase (EC 3.2.1.4) and four β-glucosidase (EC 3.2.1.21) genes identified in the starch and sucrose metabolism subsystem may be responsible for the observed cellulose-degrading capabilities in strain HZM. Neither exoglucanase (EC 3.2.1.91) nor any xylan-degrading enzyme (EC 3.2.1.8 and EC 3.2.1.37) genes were identified, indicating a possible narrow preference and/or dependence on amorphous cellulose as a carbon source. Given its cellulolytic ability and location deep in the peat substrate it is likely that strain HZM is a dedicated degrader of plant detritus. However, eight hemolysin associated genes were also identified, suggesting the potential for an alternative pathway for nutrient acquisition as well as a potentially pathogenic nature that has been associated with other members of this genus [5,6].

Comparison of genome sequences made using the RAST server revealed the closest neighbors as *Aeromonas veronii* AMC34 (score 514) and 3 strains of *Aeromonas hydrophila*: strain SSU (score 415), strain ATCC 7966 (score 405), and strain SNUFPC-A8 (score 404). On the other hand, analysis of the complete 16S rRNA sequence in EzTaxon server (http://www.ezbiocloud.net/eztaxon; [7]) under default settings (with matches only against cultured strains) identified *Aeromonas*

* Corresponding author.
E-mail address: tcchu4@student.monash.edu (P. Chua).

http://dx.doi.org/10.1016/j.gdata.2015.05.008

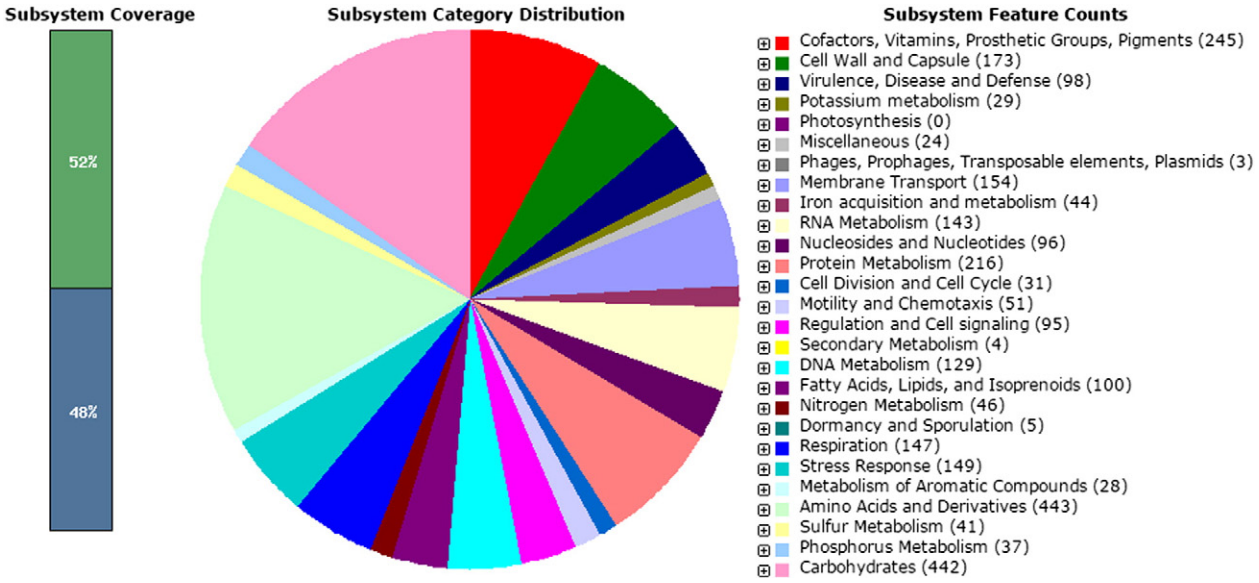

Fig 1. Subsystem distribution of *Aeromonas* sp. strain HZM (based on RAST annotation server).

punctata subsp. caviae (pairwise similarity of 99.52%) as the most closely related species, followed by *A. punctata* subsp. punctata (99.40%), *A. hydrophila* subsp. ranae (99.39%) and *A. hydrophila* subsp. hydrophila (99.33%). This is also evident in the genome-to-genome distance calculator 2.0 (GGDC) analyses provided by DSMZ (http://ggdc.dsmz.de; [8]), which yielded DDH values >80% for multiple *A. caviae* strains. Overall the various *in silico* results confirmed that the present environmental isolate is a member of the genus *Aeromonas*, though further characterization work is required to determine its species.

3. Nucleotide sequence accession number

The *Aeromonas* sp. HZM whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. JEMQ00000000.

Conflict of interest

The authors declare that there is no conflict of interests on the work published in this paper.

Acknowledgments

Funding for this study was provided by the School of Science, Monash University Malaysia. We are grateful to the Pahang State Forestry Department for their assistance in procuring the samples.

References

[1] M.V. Sizova, J.A. Izquierdo, N.S. Panikov, L.R. Lynd, Cellulose- and xylan-degrading thermophilic anaerobic bacteria from biocompost. Appl. Environ. Microbiol. 77 (7) (2011) 2282–2291.
[2] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25 (1997) 955–964.
[3] K. Lagesen, P. Hallin, E.A. Rodland, H.H. Staerfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35 (2007) 3100–3108.
[4] 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, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9 (2008) 75.
[5] G. Wang, C.G. Clark, C. Liu, C. Pucknell, C.K. Munro, T.M.A.C. Kruk, R. Caldeira, D.L. Woodward, F.G. Rodgers, Detection and characterization of the hemolysin genes in *Aeromonas hydrophila* and *Aeromonas sobria* by multiplex PCR. J. Clin. Microbiol. 41 (3) (2003) 1048–1054.
[6] G. Wang, K.D. Tyler, C.K. Munro, W.M. Johnson, Characterization of cytotoxic, hemolytic *Aeromonas caviae* clinical isolates and their identification by determining presence of a unique hemolysin gene. J. Clin. Microbiol. 34 (1996) 3203–3205.
[7] O.S. Kim, Y.J. Cho, K. Lee, S.H. Yoon, M. Kim, H. Na, S.C. Park, Y.S. Jeon, J.H. Lee, H. Yi, S. Won, J. Chun, Introducing EzTaxon: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int. J. Syst. Evol. Microbiol. 62 (2012) 716–721.
[8] J.P. Meier-Kolthoff, A.F. Auch, H.-P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14 (60) (2013).