



Data in Brief

Genome sequencing and annotation of *Proteus* sp. SAS71Samy Selim^{a,b,*}, Sherif Hassan^{a,c}, Nashwa Hagagy^b^a Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Aljouf University, Sakaka, P.O. 2014, Saudi Arabia^b Microbiology and Botany Department, Faculty of Science, Suez Canal University, Ismailia, P.O. Box 41522, Egypt^c Department of Botany, Faculty of Science, University of Beni-Suef, Beni-Suef 62511, Egypt

ARTICLE INFO

Article history:

Received 4 August 2015

Accepted 7 August 2015

Available online 12 August 2015

Keywords:

Water springs

Proteus

Biosynthesis of biotin

Antimicrobial peptides

Whole genome sequencing

ABSTRACT

We report draft genome sequence of *Proteus* sp. strain SAS71, isolated from water spring in Aljouf region, Saudi Arabia. The draft genome size is 3,037,704 bp with a G + C content of 39.3% and contains 6 rRNA sequence (single copies of 5S, 16S & 23S rRNA). The genome sequence can be accessed at DDBJ/EMBL/GenBank under the accession no. LDIU00000000.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications

Organism/cell line/tissue	<i>Proteus</i> sp.
Strain(s)	SAS71
Sequencer or array type	Sequencer; Roche 454
Data format	Processed
Experimental factors	Microbial strains
Experimental features	Draft genome sequence of <i>Proteus</i> sp. SAS71
Consent	Assembly and annotation
Sample Source location	N/A
	Water spring in Aljouf, Saudi Arabia

1. Direct link to deposited data

<http://www.ncbi.nlm.nih.gov/bioproject/?term=LDIU00000000>.

2. Experimental design, materials and methods

Proteus sp. is a Gram-negative, facultatively anaerobic, rod-shaped bacterium. It shows swarming motility and urease activity. In humans, *Proteus* is found as part of the normal flora of the gut. Its main pathological role is in infections of the urinary tract, but it can also cause wound infections and septicemia. It is widely distributed in soil and water [1].

Genomic DNA was extracted from pure culture of bacterial strain and subsequently sequenced using Roche 454 GS (FLX Titanium) pyrosequencing. All of the reads were assembled using GS De Novo Assembler version 2.8 (454 life science), which generated 1796 contigs with

N50 2082 bp. The G + C content was calculated using the draft genome sequence. The G + C content for the draft genome is 39.3%. The genome contains 40 tRNA genes and 6 rRNA genes (5S–23S–16S) predicted by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP).

A total of 3420 protein coding sequences in 353 subsystems were functionally annotated by Rapid Annotation using the Subsystems Technology (RAST) [2] server (Fig. 1). Genome analysis revealed that the genome of *Proteus* sp. strain SAS71 contains various gene clusters for biosynthesis of secondary metabolites and antimicrobial peptides. Genome analysis revealed that the genome of *Proteus* sp. SAS71 contains various gene clusters for biosynthesis of biotin and antimicrobial peptides. The genome information displays several virulence genes encoding adhesion, flagellin and hemagglutinin.

Functional comparison of genome sequences in the RAST server revealed the closest neighbors of *Proteus mirabilis* ATCC 29906 (score 516) followed by *P. mirabilis* HI 4320 (score 505), *Proteus penneri* ATCC 35198 (score 482), *P. mirabilis* WGLW4 (score 445) and *P. mirabilis* WGLW6 (score 436). On the other hand, the analysis of the complete 16S rRNA sequence in EzTaxonserver (<http://www.ezbiocloud.net/eztaxon>; [3]) under default settings (with matches only against cultured strains) identified *P. mirabilis*. Overall the various in silico results confirmed that the present environmental isolate is a member of the genus *Proteus*, though further characterization work is required to determine its species.

3. Nucleotide sequence accession number

The *Proteus* sp. SAS71 whole genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession no LDIU00000000.

* Corresponding author at: Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Aljouf University, Sakaka, P.O. 2014, Saudi Arabia.

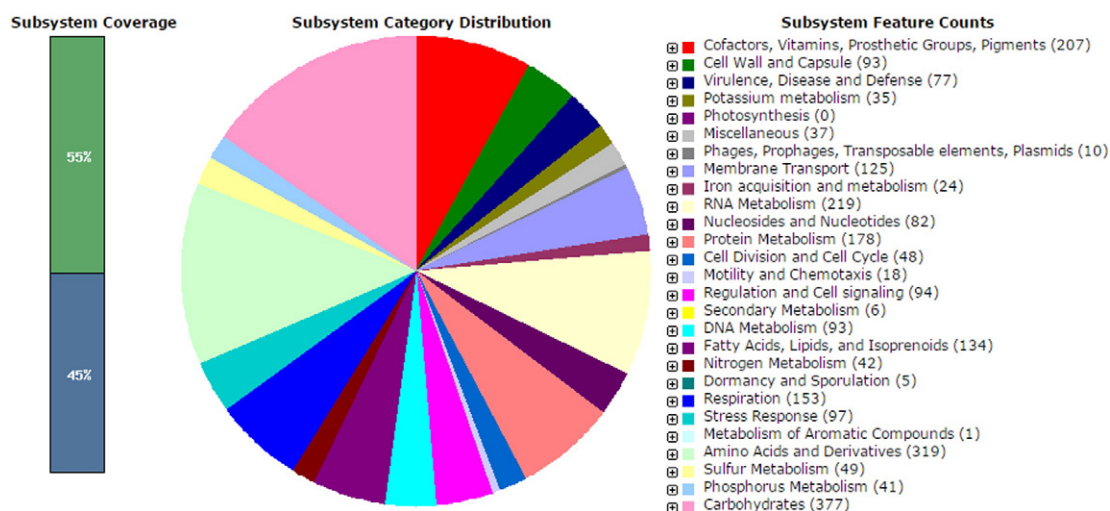


Fig. 1. Subsystem distribution of *Proteus* strain SAS71 (based on RAST annotation server).

Conflict of interest

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

Acknowledgements

This work was funded by the Deanship of Scientific Research (DSR), Aljouf University, Aljouf, KSA, under grant no. (34/155, 2014). The authors, therefore, acknowledge with thanks DSR technical and financial support. We also would like to acknowledge ArrayGen Technologies, Pune, India for contributing to data analysis and bioinformatics support.

References

- [1] M. Gué, V. Dupont, A. Dufour, O. Sire, Bacterial swarming: a biological time-resolved FTIR-ATR study of *Proteus mirabilis* swarm-cell differentiation. *Biochemistry* 40 (2001) 11938–11945, <http://dx.doi.org/10.1021/bi010434m>. PMID 11570895.
- [2] 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.
- [3] 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.