Title

Draft genome sequence of Halorubrum tropicale strain V5, a novel halophilic archaeon isolated from the solar salterns of Cabo Rojo, Puerto Rico

Authors

Sánchez-Nieves, Rubén
Facciotti, Marc T
Saavedra-Collado, Sofía
et al.

Publication Date

2016-03-01

DOI

10.1016/j.gdata.2016.02.004

Peer reviewed

Genomics Data 7 (2016) 284–286

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Draft genome sequence of *Halorubrum tropicale* strain V5, a novel halophilic archaeon isolated from the solar salterns of Cabo Rojo, Puerto Rico

Rubén Sánchez-Nieves [a], Marc T. Facciotti [b], Sofía Saavedra-Collado [a], Lizbeth Dávila-Santiago [a], Roy Rodríguez-Carrero [a], Rafael Montalvo-Rodríguez [a,*]

[a] Biology Department, Box 9000, University of Puerto Rico, Mayaguez PR 00681-9000, Puerto Rico
[b] Biomedical Engineering and Genome Center, 451 Health Sciences Drive, Davis, CA 95618, United States

ARTICLE INFO

Article history:
Received 2 February 2016
Accepted 5 February 2016
Available online 6 February 2016

Keywords:
Halophilic archaea
Genome sequence
Marine solar saltern
Puerto Rico

ABSTRACT

The genus *Halorubrum* is a member of the family *Halobacteriaceae* which currently has the highest number of described species (31) of all the haloarchaea. Here we report the draft genome sequence of strain V5, a new species within this genus that was isolated from the solar salterns of Cabo Rojo, Puerto Rico. Assembly was performed and rendered the genome into 17 contigs (N50 = 515,834 bp), the largest of which contains 1,031,026 bp. The genome consists of 3.57 MB in length with G + C content of 67.6%. In general, the genome includes 4 rRNAs, 52 tRNAs, and 3246 protein-coding sequences. The NCBI accession number for this genome is LIST00000000 and the strain deposit number is CECT9000.

Specifications

Organism	*Halorubrum tropicale*
Strain	V5
Sequencer or array type	MiSeq systems (Illumina)
Data format	Analyzed
Experimental factors	Microbial strain
Experimental features	Assembled and annotated whole genome
Consent	*N/A*
Sample source location	Solar salterns of Cabo Rojo, Puerto Rico 17°57′12″ N, 67°11′45″W

Direct link to deposited data:
http://www.ncbi.nlm.nih.gov/bioproject/PRJNA293638

1. Summary

Summary of the genome information for strain V5.

Organism	*H. tropicale*, strain V5
Source	Cabo Rojo, Puerto Rico
Genome Size, Mb	3.57
GC content, %	67.6
tRNA	52
rRNA	4
Protein coding sequences	3246

The genus *Halorubrum* belongs to the halophilic archaea and it currently comprises 31 species [1]. Members of this genus have been isolated from different hypersaline environments around the world including salterns, soda lakes, saline soil, and fermented foods [2–5]. The solar salterns of Cabo Rojo in Puerto Rico has been subjected to microbial studies over the years where several new species had been isolated and described [6,7]. A recent microbial survey of this environment isolated a novel strain (V5) using pyruvate as the sole carbon and energy source. Analysis using the 16S rRNA gene revealed that this strain belongs to the genus *Halorubrum*. Phylogenetic analyses using the *rpoB*, *ppsA* and *atpB* genes by multilocus sequence analysis and other physiological properties revealed that strain V5 represents a new species within this genus. The name *H. tropicale* was selected for this strain and a formal taxonomic description is in progress.

The draft genome was 3,572,834 bp in length with G + C content of 67.6%. The genome was predicted to include 3405 open reading frames (ORFs), 3246 of which are protein-coding sequences, 4 rRNA (5S (2), 16S (1), 23S (1)), and 52 tRNA genes. Based on the RAST functional classification ontology (Fig. 1) (http://rast.nmpdr.org/), 1666 predicted genes were classified into subsystems, with 229 genes classified into

* Corresponding author.
E-mail address: rafael.montalvo@upr.edu (R. Montalvo-Rodríguez).

http://dx.doi.org/10.1016/j.gdata.2016.02.004

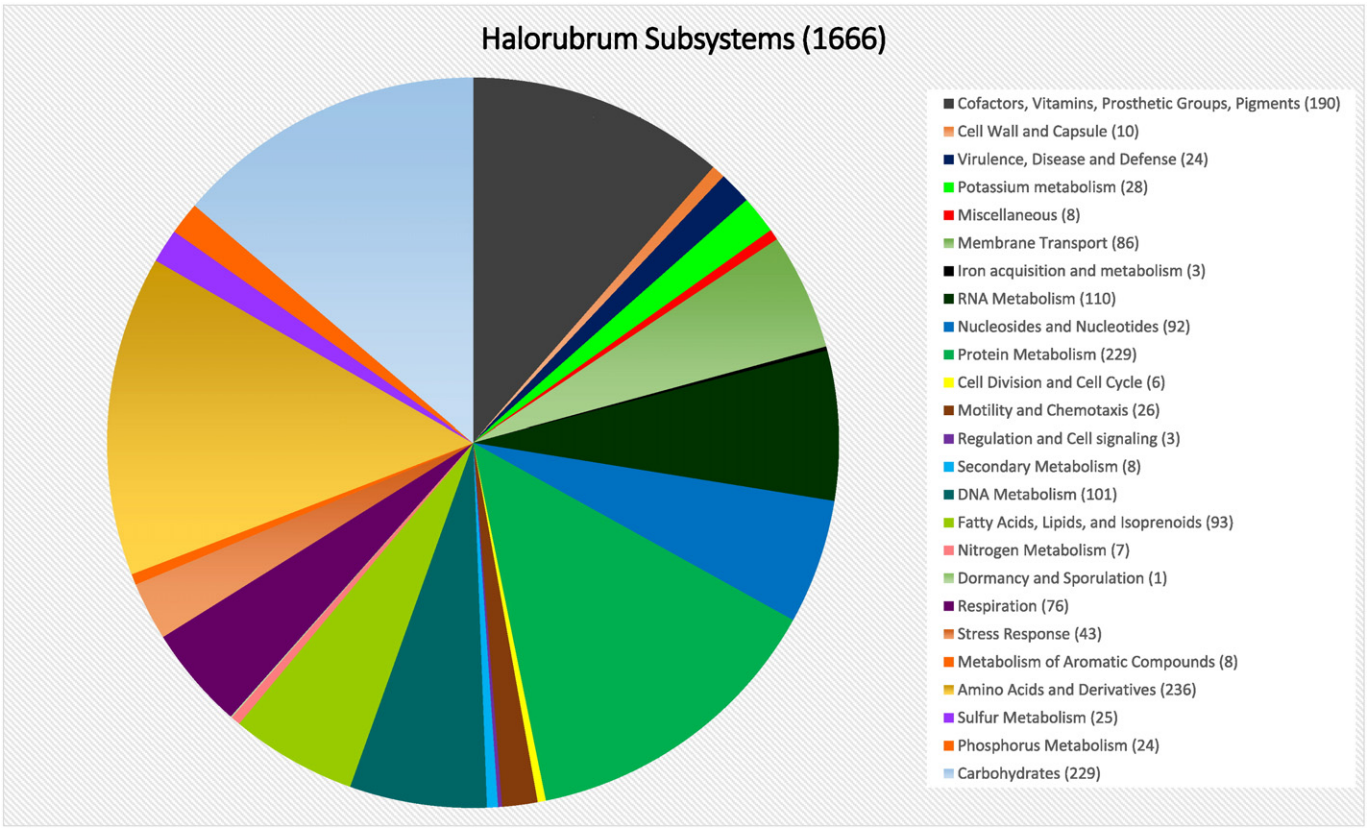

Fig. 1. The subsystem category distribution of strain V5 (*H. tropicale*). The chart represents the coverage of proteins which were grouped into subsystems. Each section represents a subsystem and the number of proteins within that subsystem. A total of 1666 proteins were categorized within these subsystems. This chart was generated by RAST (Rapid Annotation System Technology).

the carbohydrate transport and metabolism subsystem. Carbohydrate-active enzyme analysis performed by dbCAN (http://csbl.bmb.uga.edu/dbCAN/) showed that most of these enzymes belonged to the glycoside hydrolase group, with 19 total enzymes in 10 different families.

2. Experimental design

Strain V5 was grown in solid media containing pond water from the sampling site which was diluted to 20% NaCl (w/v). The media also contained (in g/L), yeast extract 5, sodium pyruvate 5, and agar 20. Genomic DNA extraction was performed using a Promega Wizard® Genomic DNA Purification Kit. The DNA sample was sequenced at the Molecular Research Lab (MR DNA) facility in Shallowater, TX, USA. The Nextera DNA Sample preparation kit (Illumina) was used following the manufacturer's instructions to prepare the genomic library. The initial DNA concentration was determined using the Qubit® dsDNA HS Assay Kit (Life Technologies). To achieve the recommended DNA input of 50 ng, samples were diluted accordingly to a concentration of 2.5 ng/μL. Then samples underwent fragmentation and the addition of adapter sequences. These adapters are utilized during a limited-cycle (5 cycles) PCR in which unique index was added to the sample. After the library was prepared, the final concentration was measured with the Qubit® dsDNA HS Assay Kit (Life Technologies), and the Agilent 2100 Bioanalyzer (Agilent Technologies) was utilized to determine the average library size. The library was then pooled in equimolar ratios of 2 nM, and 12 pM of the library pool was sequenced paired end for 600 cycles using the MiSeq system (Illumina). Assembly was performed using the NGEN Assembler and resulted in 17 contigs (N50 = 515,834 bp), the largest of which included 1,031,026 bp. Protein coding sequences were predicted using RAST (Rapid Annotation Subsystem Technology) [8], ribosomal RNA genes were detected using RNAmmer 1.2 server [9] and tRNA genes were detected using ARAGORN [10]. Carbohydrate-active enzymes were annotated by dbCAN [11].

Conflict of interest

The authors declare that no conflict of interest exists about the work published in this paper.

Acknowledgments

This study was supported by Howard Hughes Medical Institute (HHMI) "Enhancing Advanced Educational Opportunities in STEM Fields for Minority Students at UPRM" Grant Number 52007566. The authors would like to thank members of the Extremophile Laboratory at UPRM for their help.

References

[1] A.C. Parte, LPSN–list of prokaryotic names with standing in nomenclature. Nucleic Acids Res. 42 (Database issue) (2014) D613–D616.
[2] P. Corral, R.R. de la Haba, C. Sánchez-Porro, M.A. Amoozegar, R.T. Papke, A. Ventosa, *Halorubrum persicum* sp. nov., an extremely halophilic archaeon isolated from sediment of a hypersaline lake. Int. J. Syst. Evol. Microbiol. 65 (6) (2015) 1770–1778.
[3] J. Feng, P. Zhou, Y.-G. Zhou, S.-J. Liu, K. Warren-Rhodes, *Halorubrum alkaliphilum* sp. nov., a novel haloalkaliphile isolated from a soda lake in Xinjiang, China. Int. J. Syst. Evol. Microbiol. 55 (1) (2005) 149–152.
[4] P.T. Pesenti, M. Sikaroodi, P.M. Gillevet, C. Sánchez-Porro, A. Ventosa, C.D. Litchfield, *Halorubrum californiense* sp. nov., an extreme archaeal halophile isolated from a crystallizer pond at a solar salt plant in California, USA. Int. J. Syst. Evol. Microbiol. *58* (12) (2008) 2710–2715.
[5] K.J. Yim, I.-T. Cha, H.-W. Lee, H.S. Song, K.-N. Kim, S.-J. Lee, ... D. Kim, *Halorubrum halophilum* sp. nov., an extremely halophilic archaeon isolated from a salt-fermented seafood. Antonie Van Leeuwenhoek 105 (3) (2014) 603–612.
[6] R. Montalvo-Rodríguez, R.H. Vreeland, A. Oren, M. Kessel, C. Betancourt, J. López-Garriga, *Halogeometricum borinquense* gen. nov., sp. nov., a novel halophilic archaeon from Puerto Rico. Int. J. Syst. Bacteriol. 48 (1998) 1305–1312.

[7] R. Montalvo-Rodríguez, J. López-Garriga, R.H. Vreeland, A. Oren, A. Ventosa, M. Kamekura, *Haloterrigena thermotolerans* sp. nov., a halophilic archaeon from Puerto Rico. Int. J. Syst. Evol. Microbiol. 50 (2000) 1065–1071.
[8] R. Overbeek, R. Olson, G.D. Pusch, G.J. Olsen, J.J. Davis, T. Disz, R. Stevens, The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42 (Database issue) (2014) D206–D214.
[9] K. Lagesen, P.F. Hallin, E. Rødland, H.H. Stærfeldt, T. Rognes, D.W. Ussery, RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. 35 (9) (2007) 3100–3108.
[10] D. Laslett, B. Canback, ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32 (1) (2004) 11–16.
[11] Y. Yin, X. Mao, J.C. Yang, X. Chen, F. Mao, Y. Xu, dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40 (Web Server issue) (2012) W445-51.