

1 **GENOME ANNOUNCEMENT**

2

3 **Complete Genome Sequence of the Haloalkaliphilic, Hydrogen Producing**  
4 ***Halanaerobium hydrogeniformans***

5

6 Steven D. Brown<sup>1</sup>, Matthew B. Begemann<sup>2</sup>, Melanie R. Mormile<sup>3</sup>, Judy D. Wall<sup>4</sup>, Cliff S.  
7 Han<sup>5,6</sup>, Lynne A. Goodwin<sup>5,6</sup>, Samuel Pitluck<sup>6,7</sup>, Miriam L. Land<sup>1,6</sup>, Loren J. Hauser<sup>1,6</sup>,  
8 and Dwayne A. Elias<sup>1\*</sup>

9

10

11 <sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37831

12 <sup>2</sup>Microbiology Doctoral Training Program, University of Wisconsin, Madison,  
13 Wisconsin, 53706

14 <sup>3</sup>Department of Biological Sciences, Missouri University of Science & Technology,  
15 Rolla, Missouri, 65409.

16 <sup>4</sup>Biochemistry Department, University of Missouri, Columbia, Missouri, 65211

17 <sup>5</sup>Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

18 <sup>6</sup>DOE Joint Genome Institute, Walnut Creek, California 94598

19 <sup>7</sup>Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

20

21 \*Corresponding author

22 Biosciences Division

23 Oak Ridge National Laboratory

24 Oak Ridge, TN 37831

25 Tel: 865- 574-0956

26 Fax: 865-576-3989

27 E-mail: [eliasda@ornl.gov](mailto:eliasda@ornl.gov)

28

29 **Abstract**

30 *Halanaerobium hydrogeniformans* is an alkaliphilic bacterium capable of  
31 biohydrogen production at pH 11 and 7% (w/v) salt. We present the 2.6 Mb genome  
32 sequence to provide insights into its physiology and potential for bioenergy  
33 applications.

34

35 Hydrogen has numerous applications as a clean energy carrier with water and heat  
36 as byproducts. For industrial purposes, it is currently produced abiotically from natural  
37 gas and coal (10). Microorganisms are being explored for chemical and biofuel  
38 production, including future biohydrogen generation at industrial scales. Growth in  
39 biofuel production will most likely be derived from lignocellulosic biomass. However,  
40 pretreatment of biomass, required so the sugars can be accessible, is carried out with  
41 regimes ranging from alkaline to acidic conditions, often including elevated temperatures  
42 with each step adding to the final product cost (7).

43 *Halanaerobium hydrogeniformans* (formerly *H. sapolanicus*) was isolated from  
44 haloalkaline, (pH ~10, 15-140 g/L NaCl) anaerobic sediments of Soap Lake, Washington  
45 with extraordinarily high sulfide concentrations of up to 10g/L (1, 3, 4, 9, 13). It is an  
46 obligately anaerobic, Gram negative, nonmotile, non-sporulating, elongated rod. It can  
47 utilize a range of C-5 and C-6 sugars with optimal growth at pH 11, 7% (w/v) NaCl and  
48 33°C, producing acetate, formate and hydrogen as major metabolic end-products. The  
49 genome sequence for *H. hydrogeniformans* was determined to improve assessment of its  
50 metabolic and bioenergy potential, particularly towards improving alkaline or  
51 haloalkaline pretreatment regimes for robust hydrogen production by this bacterium.

52 The *H. hydrogeniformans* genome sequence was determined through a  
53 combination of Illumina (8) and 454 technologies (8). We constructed and sequenced an  
54 Illumina GAii shotgun library which generated 27,639,916 reads totaling 2100 Mb, a 454  
55 Titanium standard library generated from 77,351 reads and a paired end 454 library with  
56 an average insert size of 10.607 +/- 2.651 kb that generated 160,293 reads totaling 82.3  
57 Mb of 454 data. A total of 486 additional reactions and 6 shatter libraries were necessary  
58 to close gaps and to raise the finished sequence quality. Methods for determining the  
59 genome sequence are previously described (6) and it is a 'finished' genome (2).

60           The total genome size is 2,613,116 bp, with final assembly based on 52.2 Mb of  
61 454 draft data providing an average 21.5x genome coverage and 463 Mb of Illumina draft  
62 data providing an average 178x genome coverage. The genome is 33.1% G+C and  
63 encodes 2,295 candidate protein-encoding gene models. The genome contains four  
64 separate rRNA operons, each containing a 5S, 16S, and 23S rRNA gene with 99.9-100%  
65 identity between 16S rRNA genes. The closest significant 16S rRNA gene matches  
66 (GenBank Accession Number GQ215697) were to *Halanaerobium* sp. *AN-B15B* (97.2%  
67 similarity, AM157647.1), *Halanaerobium praevalens* DSM 2228 (97%, AB022035.1),  
68 *Halanaerobium* sp. *KT-2/3-3* (96.9%, AJ309519.1), and *Halanaerobium acetethylicum*  
69 (96.4% NR\_036958.1) (5, 11, 12). All comparative species are physiologically different  
70 as they are neutrophilic. The *Halanaerobium hydrogeniformans* genome sequence will  
71 enable a range of studies into halotolerance and alkaliphily, will facilitate a better  
72 understanding of the potential of this organism for bioenergy applications, and will  
73 expand our knowledge of the physiology of this genus.

74

#### 75 **Nucleotide sequence accession number**

76 This whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under  
77 the accession CP002304. The version described in this paper is the first version,  
78 CP002304.

79

#### 80 **Acknowledgements**

81 Supported by the Office of Biological and Environmental Research in the DOE Office of  
82 Science through the BioEnergy Science Center, a US DOE Bioenergy Research Center.  
83 Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S.  
84 Department of Energy under contract DE-AC05-00OR22725. The work conducted by the  
85 U.S. Department of Energy Joint Genome Institute is supported by the Office of Science  
86 of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

87

88

89 **References**

- 90 1. **Anderson, G. C.** 1958. Seasonal characteristics of two saline lakes in  
91 Washington. *Limnol. Ocean.* **3**:51-68.
- 92 2. **Chain, P. S. G., D. V. Grafham, R. S. Fulton, M. G. FitzGerald, J. Hostetler,**  
93 **D. Muzny, J. Ali, B. Birren, D. C. Bruce, C. Buhay, J. R. Cole, Y. Ding, S.**  
94 **Dugan, D. Field, G. M. Garrity, R. Gibbs, T. Graves, C. S. Han, S. H.**  
95 **Harrison, S. Highlander, P. Hugenholtz, H. M. Khouri, C. D. Kodira, E.**  
96 **Kolker, N. C. Kyrpides, D. Lang, A. Lapidus, S. A. Malfatti, V. Markowitz,**  
97 **T. Metha, K. E. Nelson, J. Parkhill, S. Pitluck, X. Qin, T. D. Read, J.**  
98 **Schmutz, S. Sozhamannan, P. Sterk, R. L. Strausberg, G. Sutton, N. R.**  
99 **Thomson, J. M. Tiedje, G. Weinstock, A. Wollam, Genomic Standards**  
100 **Consortium Human Microbiome Project Jumpstart Consortium, and J. C.**  
101 **Detter.** 2009. Genome project standards in a new era of sequencing. *Science*  
102 **326**:236-237.
- 103 3. **Dimitriu, P. A., H. C. Pinkart, B. M. Peyton, and M. R. Mormile.** 2008.  
104 Spatial and temporal patterns in the microbial diversity of a meromictic soda lake  
105 in Washington State. *Appl. Environ. Microbiol.* **74**:4877-4888.
- 106 4. **Dimitriu, P. A., H. C. Pinkart, B. M. Peyton, and M. R. Mormile.** 2008.  
107 Spatial and temporal patterns in the microbial diversity of a meromictic soda lake  
108 in Washington State. *Appl. Environ. Microbiol.* **74**:4877-4888.
- 109 5. **Eder, W., L. L. Jahnke, M. Schmidt, and R. Huber.** 2001. Microbial diversity  
110 of the brine-seawater interface of the Kebrit Deep, Red Sea, studied via 16S  
111 rRNA gene sequences and cultivation methods. *Appl. Environ. Microbiol.*  
112 **67**:3077-3085.
- 113 6. **Elkins, J. G., A. Lochner, S. D. Hamilton-Brehm, K. W. Davenport, M.**  
114 **Podar, S. D. Brown, M. L. Land, L. J. Hauser, D. M. Klingeman, B. Raman,**  
115 **L. A. Goodwin, R. Tapia, L. J. Meincke, J. C. Detter, D. C. Bruce, C. S. Han,**  
116 **A. V. Palumbo, R. W. Cottingham, M. Keller, and D. E. Graham.** 2010.  
117 Complete Genome Sequence of the Cellulolytic Thermophile *Caldicellulosiruptor*  
118 *obsidiansis* OB47T. *J. Bacteriol.*:JB.00950-00910.

- 119 7. **Johnson, D. K., and R. T. Elander.** 2008. Pretreatments for enhanced  
120 digestibility of feedstocks, p. 436-453. *In* M. E. Himmel (ed.), Biomass  
121 recalcitrance: deconstructing the plant cell wall for bioenergy. Blackwell  
122 Publishing Ltd, Oxford, United Kingdom.
- 123 8. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A.**  
124 **Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L.**  
125 **Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho,**  
126 **G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B.**  
127 **Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li,**  
128 **K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E.**  
129 **W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T.**  
130 **Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R.**  
131 **Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P.**  
132 **Weiner, P. G. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing  
133 in microfabricated high-density picolitre reactors. *Nature* **437**:376-380.
- 134 9. **Mormile, M. R., M. F. Romine, M. T. Garcia, A. Ventosa, T. J. Bailey, and B.**  
135 **M. Peyton.** 1999. *Halomonas campisalis* sp. nov., a denitrifying, moderately  
136 haloalkaliphilic bacterium. *Syst. Appl. Microbiol.* **22**:551-558.
- 137 10. **Mueller-Langer, F., E. Tzimas, M. Kaltschmitt, and S. Peteves.** 2007. Techno-  
138 economic assessment of hydrogen production processes for the hydrogen  
139 economy for the short and medium term. *Int. J. Hydrogen Energy* **32**:3797-3810.
- 140 11. **Perreault, N. N., D. T. Andersen, W. H. Pollard, C. W. Greer, and L. G.**  
141 **Whyte.** 2007. Characterization of the prokaryotic diversity in cold saline  
142 perennial springs of the Canadian high Arctic. *Applied and Environmental*  
143 *Microbiology* **73**:1532-1543.
- 144 12. **Rengpipat, S., T. A. Langworthy, and J. G. Zeikus.** 1988. *Halobacteroides-*  
145 *acetioethylicus* sp-nov, a new obligatory anaerobic halophile isolated from deep  
146 subsurface hypersaline environments. *Syst. Appl. Microbiol.* **11**:28-35.
- 147 13. **Sorokin, D. Y., M. Foti, H. C. Pinkart, and G. Muyzer.** 2007. Sulfur-oxidizing  
148 bacteria in Soap Lake (Washington State), a meromictic, haloalkaline lake with an  
149 unprecedented high sulfide content. *Appl. Environ. Microbiol.* **73**:451-455.