IM Release Number: LLNL-JRNL-489129

Complete Genome Sequence of the Anaerobic, Halophilic Alkalithermophile *Natranaerobius thermophilus* JW/NM-WN-LF

Baisuo Zhao[1], Noha M. Mesbah[1]†, Eileen Dalin[2], Lynne Goodwin[3], Matt Nolan[2], Sam Pitluck[2], Olga Chertkov[3], Thomas S. Brettin[4], James Han[2], Frank W. Larimer[2], Miriam L. Land[4], Loren Hauser[4], Nikolaos Kyrpides[2], and Juergen Wiegel[1]*

[1] *Department of Microbiology, University of Georgia, Athens, Georgia 30602*1

[2] *DOE Joint Genome Institute, Walnut Creek/Lawrence Berkeley National Lab, California 94598*

[3] *Los Alamos National Laboratory, Los Alamos, New Mexico, 87545*

[4] *Oak Ridge National Laboratory, Oak Ridge, Tennessee 378314*

**To whom correspondence may be addressed: Juergen Wiegel* - Mailing address: Department of Microbiology, University of Georgia, Athens, GA 30602. Phone: (706) 542-2651. Fax: (706) 542-2674. E-mail: juergenwiegel@gmail.com.

15 June 2011

ACKNOWLEDGMENTS:

This work was supported by grants MCB-060224 from the National Science Foundation and AFOSR 033835-01 from the Air Force Office of Scientific Research to J.W. and DE-AC02-05CH11231 from the Office of Science of the U.S. Department of Energy to the DOE Joint Genome Institute.

DISCLAIMER:

[**LLNL**] This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes

IM Release Number: LLNL-JRNL-489129

[**LBNL**] This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Complete Genome Sequence of the Anaerobic, Halophilic Alkalithermophile *Natranaerobius thermophilus* JW/NM-WN-LF△

Baisuo Zhao[1], Noha M. Mesbah[1]†, Eileen Dalin[2], Lynne Goodwin[3], Matt Nolan[2], Sam Pitluck[2], Olga Chertkov[3], Thomas S. Brettin[4], James Han[2], Frank W. Larimer[2], Miriam L. Land[4], Loren Hauser[4], Nikolaos Kyrpides[2], and Juergen Wiegel[1]*

[1] *Department of Microbiology, University of Georgia, Athens, Georgia 30602*1

[2] *DOE Joint Genome Institute, Walnut Creek/Lawrence Berkeley National Lab, California 94598*

[3] *Los Alamos National Laboratory, Los Alamos, New Mexico, 87545*

[4] *Oak Ridge National Laboratory, Oak Ridge, Tennessee 378314*

The genome of the anaerobic halophilic alkalithermophile *Natranaerobius thermophilus* consists of one 3,165,557-bp chromosome and two plasmids (17,207 bp and 8,689 bp). The present study is the first to report the completely sequenced genome of an anaerobic polyextremophile and genes associated with roles in regulation of intracellular osmotic pressure, pH homeostasis, and growth at elevated temperatures.

The polyextremophile *Natranaerobius thermophilus* strain JW/NM-WN-LFT represents the novel order *Natranaerobiales* (*Firmicutes*), harboring obligately anaerobic alkalithermophiles. It grows optimally at 3.5MNa_ with a pH of 9.5 at 55°C and 53°C and was isolated from the sediment of the sun-heated salt lake Fazda (Wadi An Natrun, Egypt) (5). *N. thermophilus* is an excellent model organism for understanding the unique combinations of adaptation mechanisms in anaerobic halophilic alkalithermophiles necessary for thriving in the presence of multiple environmental extremes (4).

The genome of *N. thermophilus* was sequenced at the Joint Genome Institute (JGI) using a combination of 3-, 8-, and 40-kb insert libraries and random shotgun sequencing. The Phred/Phrap/Consed software package was used for sequence assembly and quality. Possible misassemblies were corrected, and gaps between contigs were closed by editing in Consed, custom primer walks, or PCR amplification (1). The error rate of the completed genome sequence of *N. thermophilus* was less than 1 in 50,000. Genes were identified using the Prodigal software program (2), followed by a round of manual curation using the JGI GenePRIMP pipeline (6).

The *N. thermophilus* genome contains a single, circular chromosome (3,165,557 bp) and two circular plasmids, pNTHE01 and pNTHE02 (17,207 bp and 8,689 bp, respectively). The average GC content of the three elements is 36.4% for the chromosome, 34.1% for pNTHE01, and 35.7% for pNTHE02. Genome sequence analysis revealed the presence of a large group of genes encoding putative proteins potentially associated with the adaptation of *N. thermophilus* to life under multiple conditions of high salinity, alkaline pH, and elevated temperature: the *gsmA* gene and the *sdmA* gene, encoding *de novo* synthesis of the solute glycine betaine, three genes for L-glutamine synthetase, 15 genes for glycine betaine ABC transporters, four genes for glycine betaine/L-proline ABC transporters, and three for betaine/carnitine/choline transporters. Furthermore, five genes encoding Na_/proline symporters, two for Na_/glutamate symporters, and seven for K_ transport systems, which together with a specific K_/H_ antiporter regulate the intracellular K_ concentration, were found. *N. thermophilus* is an extremely halophilic alkaliphile and therefore must have mechanisms for cytoplasm acidification. The chromosome of *N. thermophilus* contains genes for 11 monovalent cation/proton antiporters of the NhaC type, a gene cluster encoding a multisubunit cation/proton antiporter of the CPA-3 family, four monovalent cation/proton antiporters of the CPA1 and CPA2 family, and one gene encoding a cation/proton antiporter of the NdhF-a family (3, 4). The annotations of seven sodium and potassium proton antiporters, one specific K_ antiporter, and 3 K_ transporters were experimentally verified (4). Four genes were identified encoding orthologous rRNA MTases and three orthologous tRNA MTases, all involved in the structural stabilization of DNA and RNA at high temperature. Two genes were identified for L-isoaspartate *O*methyltransferase, one gene for spermine synthase, and 12 genes for heat shock proteins. A full genome analysis combined with data from verified experiments will be the subject of a future publication.

Nucleotide sequence accession numbers. The genome sequence of *N. thermophilus* was deposited in GenBank with the accession numbers CP001034 for the chromosome, CP001035 for the plasmid pNTHE01, and CP001036 for the plasmid pNTHE02.

This work was supported by grants MCB-060224 from the National Science Foundation and AFOSR 033835-01 from the Air Force Office of Scientific Research to J.W. and DE-AC02-05CH11231 from the Office of Science of the U.S. Department of Energy to the DOE Joint Genome Institute.

REFERENCES

1. **Allen, M. A., et al.** 2009. The genome sequence of the psychrophilic archaeon, Methanococcoides burtonii: the role of genome evolution in cold adaptation. ISME J. **3:**1012–1035.

2. **Hyatt, D., et al.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119.

3. **Krulwich, T. A., D. B. Hicks, and M. Ito.** 2009. Cation/proton antiporter complements of bacteria: why so large and diverse? Mol. Microbiol. **74:**257–260.

4. **Mesbah, N. M., G. M. Cook, and J. Wiegel.** 2009. The halophilic alkalithermophile *Natranaerobius thermophilus* adapts to multiple environmental extremes using a large repertoire of Na_(K_)/H_ antiporters. Mol. Microbiol. **74:**270–281.

5. **Mesbah, N. M., D. B. Hedrick, A. D. Peacock, M. Rohde, and J. Wiegel.** 2007. *Natranaerobius thermophilus* gen. nov., sp. nov., a halophilic, alkalithermophilic bacterium from soda lakes of the Wadi An Natrun, Egypt, and proposal of *Natranaerobiaceae* fam. nov. and *Natranaerobiales* ord. nov. Int. J. Syst. Evol. Microbiol. **57:**2507–2512.

6. **Pati, A., et al.** 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat. Methods **7:**455–457.

---

* Corresponding author. Mailing address: Department of Microbiology, University of Georgia, Athens, GA 30602. Phone: (706) 542-2651. Fax: (706) 542-2674. E-mail: juergenwiegel@gmail.com.

† Present address: Department of Biochemistry, Suez Canal University, Ismailia 41522, Egypt.

△ Published ahead of print on 3 June 2011.