JB Journals.ASM.org

GENOME ANNOUNCEMENT

Genome Sequence of an Oligohaline Hyperthermophilic Archaeon, *Thermococcus zilligii* AN1, Isolated from a Terrestrial Geothermal Freshwater Spring

Byung Kwon Kim,[a] Seong Hyuk Lee,[b,c] Seon-Young Kim,[d,e] Haeyoung Jeong,[a,f] Soon-Kyeong Kwon,[a,f] Choong Hoon Lee,[a] Ju Yeon Song,[a] Dong Su Yu,[a] Sung Gyun Kang,[b,c] and Jihyun F. Kim[a,f,g]

Systems and Synthetic Biology Research Center[a] and BioMedical Genomics Research Center,[d] Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Republic of Korea; Department of Marine Biotechnology,[b] Department of Functional Genomics,[e] and Biosystems and Bioengineering Program,[f] University of Science and Technology, Daejeon, Republic of Korea; Korea Ocean Research and Development Institute, Ansan, Republic of Korea[c]; and Department of Systems Biology, Yonsei University, Seoul, Republic of Korea[g]

***Thermococcus zilligii*, a thermophilic anaerobe in freshwater, is useful for physiological research and biotechnological applications. Here we report the high-quality draft genome sequence of *T. zilligii* AN1^T. The genome contains a number of genes for an immune system and adaptation to a microbial biomass-rich environment as well as hydrogenase genes.**

As extreme thermophiles belonging to the family *Thermococcaceae*, species in the genus *Thermococcus* generally grow at temperatures between 75°C and 88°C (13). They require elemental sulfur or protons as an electron acceptor and release hydrogen, hydrogen sulfide, or carbon dioxide as an end product. Most of the species have been isolated from marine hydrothermal systems, whereas only two species have been isolated from freshwater hot springs (6, 12). *Thermococcus zilligii*, reported in 1982 (16), is the first of the two.

AN1 (DSM 2770), the type strain of *T. zilligii*, was isolated from a terrestrial geothermal spring in Kuirau Park, Rotorua, New Zealand, an environment rich in microbial biomass (9, 19). It is an obligately anaerobic and chemoorganotrophic sulfur-reducing microorganism and absolutely requires a low concentration of sodium or lithium ion for growth. A number of enzymes from AN1 have been investigated to understand the physiology and the evolution of thermophilic archaea (2, 4, 17, 18, 22). Also, the strain has been regarded as an attractive reservoir of enzymes, such as thermostable proteases and hemicellulases, for biotechnological applications (15, 21). In spite of scientific and industrial importance, the genome of this strain has not been analyzed.

The genome sequence of AN1 was determined using a massive parallel sequencing technology. Approximately 6.45 Gb of reads (~3,398-fold genome coverage) were produced from a 400-bp paired-end library using Ilumina/Solexa HiSeq 2000 (NICEM, Republic of Korea). The paired reads were *de novo* assembled with the CLC Genomics Workbench (CLC bio, Inc.). Twenty-four contigs were orderly rearranged in five scaffolds by SSPACE (1). *In silico* gap closing was performed by Image (20) and Perl scripts developed in-house. All tRNA and rRNA genes were predicted by tRNAscan-SE and RNAmmer, respectively. A total of 1,958 protein-coding sequences (91.84% of the coding density) were predicted by Glimmer and Gene Marks. All genes were annotated using the information from GenBank, Pfam, TIGRfam, KEGG, and RAST. Clustered, regularly interspaced short palindromic repeats (CRISPRs) were predicted by a web-based program, CRISPRFinder (7).

The final assembly consists of five contigs of 1,764,559 bp (54.48% G+C content). The genome has a single 16S-23S rRNA operon, two distantly located 5S rRNA genes, and 46 tRNA genes. Like other *Thermococcus* spp., the genome contains genes encoding two sulfhydrogenase complexes, three [NiFe]-hydrogenase complexes, and sodium-dependent transporters (5, 11, 14). However, 10 CRISPR loci composed of 5 to 58 short direct repeats were detected in the AN1 genome, whereas other completely sequenced genomes of *Thermococcus* contain fewer than six CRISPR loci. The average length of the direct repeats is 30 bases, and spacer regions between repeats range from 32 to 70 bases. Thirty-five genes encoding CRISPR-associated protein families that belong to the core and two subtypes, Mtube and RAMP module, are found near six CRISPR loci (3, 8). These CRISPR-related genes may contribute to controlling viral infection or plasmid conjugation in microbial biomass-rich aquatic habitats.

Nucleotide sequence accession number. The draft genome sequence of AN1 has been deposited in GenBank under the accession number AJLF00000000. The sequence is also available from the Genome Encyclopedia of Microbes (GEM; http://www.gem.re.kr) (10).

ACKNOWLEDGMENTS

This project was conducted in part as the Microbial Genome Sequencing and Analysis course at the University of Science and Technology in fall 2011. We are grateful to students in the class for critical comments on the manuscript and to members of GEM and the KRIBB sequencing team for technical help.

This work was supported by the National Research Foundation of Korea (2011-0017670 to J.F.K.), the Global Frontier Biosystems Design and Engineering Center (2011-0031952) of the Ministry of Education, Science and Technology, and the Development of Biohydrogen Production Technology Using Hyperthermophilic Archaea program (to S.G.K.) of the Ministry of Land, Transport, and Maritime Affairs, Republic of Korea.

Received 18 April 2012 **Accepted** 30 April 2012

Address correspondence to Jihyun F. Kim, jfk1@yonsei.ac.kr, or Sung Gyun Kang, sgkang@kordi.re.kr.

B.K.K. and S.H.L. contributed equally to this study.

doi:10.1128/JB.00655-12

REFERENCES

1. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**: 578–579.
2. **Collin RG, Morgan HW, Musgrave DR, Daniel RM.** 1988. Distribution of reverse gyrase in representative species of eubacteria and archaebacteria. FEMS Microbiol. Lett. **55**:235–240.
3. **Deveau H, Garneau JE, Moineau S.** 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu. Rev. Microbiol. **64**:475–493.
4. **Dinger ME, Baillie GJ, Musgrave DR.** 2000. Growth phase-dependent expression and degradation of histones in the thermophilic archaeon *Thermococcus zilligii*. Mol. Microbiol. **36**:876–885.
5. **Fukui T, et al.** 2005. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. Genome Res. **15**:352–363.
6. **Gonzalez JM, et al.** 1999. *Thermococcus waiotapuensis* sp. nov., an extremely thermophilic archaeon isolated from a freshwater hot spring. Arch. Microbiol. **172**:95–101.
7. **Grissa I, Vergnaud G, Pourcel C.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **35**:W52–W57.
8. **Haft DH, Selengut J, Mongodin EF, Nelson KE.** 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol. **1**:e60. doi: 10.1371/journal.pcbi.0010060.
9. **Hetzer A, Morgan HW, McDonald IR, Daughney CJ.** 2007. Microbial life in Champagne Pool, a geothermal spring in Waiotapu, New Zealand. Extremophiles **11**:605–614.
10. **Jeong H, Yoon SH, Yu DS, Oh TK, Kim JF.** 2008. Recent progress of microbial genome projects in Korea. Biotechnol. J. **3**:601–611.
11. **Kanai T, et al.** 2011. Distinct physiological roles of the three [NiFe]-hydrogenase orthologs in the hyperthermophilic archaeon *Thermococcus kodakarensis*. J. Bacteriol. **193**:3109–3116.
12. **Klages KU, Morgan HW.** 1994. Characterization of an extremely thermophilic sulphur-metabolizing archaebacterium belonging to the *Thermococcales*. Arch. Microbiol. **162**:261–266.
13. **Kobayashi T.** 2001. Genus I. Thermococcus Zillig 1983, 673^{VP}, p 342–346. *In* Boone DR, Castenholz RW, Garrity GM (ed), Bergey's manual of systematic bacteriology, 2nd ed. Springer, New York, NY.
14. **Lee HS, et al.** 2008. The complete genome sequence of *Thermococcus onnurineus* NA1 reveals a mixed heterotrophic and carboxydotrophic metabolism. J. Bacteriol. **190**:7491–7499.
15. **Michael K, Fuad H, Garabed A.** 1991. Properties of extremely thermostable proteases from anaerobic hyperthermophilic bacteria. Appl. Microbiol. Biotechnol. **34**:715–719.
16. **Morgan HW, Daniel RM.** 1982. Abstr. 13th Int. Congress Microbiol., abstr P-20.
17. **Ronimus RS, Koning J, Morgan HW.** 1999. Purification and characterization of an ADP-dependent phosphofructokinase from *Thermococcus zilligii*. Extremophiles **3**:121–129.
18. **Ronimus RS, Musgrave DR.** 1996. Purification and characterization of a histone-like protein from the archaeal isolate AN1, a member of the *Thermococcales*. Mol. Microbiol. **20**:77–86.
19. **Ronimus RS, Reysenbach A, Musgrave DR, Morgan HW.** 1997. The phylogenetic position of the *Thermococcus* isolate AN1 based on 16S rRNA gene sequence analysis: a proposal that AN1 represents a new species, *Thermococcus zilligii* sp. nov. Arch. Microbiol. **168**:245–248.
20. **Tsai IJ, Otto TD, Berriman M.** 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol. **11**:R41.
21. **Uhl AM, Daniel RM.** 1999. The first description of an archaeal hemicellulase: the xylanase from *Thermococcus zilligii* strain AN1. Extremophiles **3**:263–267.
22. **Xavier KB, da Costa MS, Santos H.** 2000. Demonstration of a novel glycolytic pathway in the hyperthermophilic archaeon *Thermococcus zilligii* by ^{13}C-labeling experiments and nuclear magnetic resonance analysis. J. Bacteriol. **182**:4632–4636.