



Contents lists available at ScienceDirect

Marine Genomics

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

Genome sequence of facultatively anaerobic marine bacterium *Maribacter thermophilus* strain HT7-2^T

Jing Hu^a, Fang Wang^a, Shuai-Bo Han^a, Sang-Ling Wu^b, Min Wu^{a,*}, Xue-Wei Xu^{c,*}

^a College of Life Sciences, Zhejiang University, Hangzhou, PR China

^b Analysis Center of Agrobiological and Environmental Sciences, Faculty of Agriculture, Life and Environment Sciences, Zhejiang University, PR China

^c Laboratory of Marine Ecosystem and Biogeochemistry, State Oceanic Administration, Hangzhou, PR China

ARTICLE INFO

Article history:

Received 2 August 2015

Received in revised form 8 August 2015

Accepted 12 August 2015

Available online xxxxx

Keywords:

Sea area

Higher optimum temperature

Facultatively anaerobic

Robustness

Genome sequence

ABSTRACT

Maribacter thermophilus strain HT7-2^T, isolated from *Ulva prolifera* collected from the intertidal zone of Qingdao sea area, China. To date, *M. thermophilus* strain HT7-2^T is the only one which has been found that has a relatively higher optimum temperature compared to other *Maribacter* species and it is also the first strain to be described that is facultatively anaerobic while other identified strains within genus *Maribacter* are strictly aerobic. Meanwhile, *M. thermophilus* strain HT7-2^T harbors several coding genes related to heavy metal resistance and to antibiotics, which increase the robustness of the strain. Here, we report the genome sequence and annotation of *M. thermophilus* strain HT7-2^T, which comprises 4,050,606 bp with G + C content of 38.93%. A total of 3585 protein coding genes, 41 tRNAs and 6rRNAs were obtained. The genome annotation may provide basic information on some special traits and pathways in this strain.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The genus *Maribacter*, belonging to the family *Flavobacteriaceae* (Nedashkovskaya et al., 2004), consists of thirteen validly described species and one invalid species, including *Maribacter aquivivus*, *Maribacter orientalis*, *Maribacter sedimenticola* and *Maribacter ulvicola* (Nedashkovskaya et al., 2004), *Maribacter dokdonensis* (Yoon et al., 2005), *Maribacter polysiphoniae* (Nedashkovskaya et al., 2007), *Maribacter arcticus* (Cho et al., 2008), *Maribacter forsetii* (Barbeyron et al., 2008), *Maribacter antarcticus* (Zhang et al., 2009), *Maribacter stanieri* (Nedashkovskaya et al., 2010), *Maribacter chungangensis* (Weerawongwiwat et al., 2013), *Maribacter aestuarii* (Lo et al., 2013), *Maribacter thermophilus* (Hu et al., 2015) and *Maribacter caenipelagi* (Jung et al., 2014). Until now, species of the genus are exclusively known from marine environments, such as seawater, sediment, alga and sponges.

M. thermophilus strain HT7-2^T, obtained from the intertidal zone of Qingdao sea area, China, demonstrates tolerance to high temperature (optimum 40–42 °C). Meanwhile, this strain is the only strain to be described that is facultatively anaerobic. We here report its genome

sequence supporting its potential in adapting to diverse environments and in industrial applications.

2. Data description

In this study, the sea water sample was collected from the intertidal zone of Qingdao sea area, China. Strain HT7-2^T was isolated and has been deposited with accession number CGMCC 1.12207^T (=JCM 18466^T). *M. thermophilus* strain HT7-2^T is Gram-negative, rod-shaped, motile by gliding and non-sporulating. Within the genus *Maribacter*, *M. thermophilus* strain HT7-2^T has been found to present a higher optimal temperature (40–42 °C) that makes it adapt to some thermophilic conditions (the general features of strain HT7-2^T was summarized in Table 1). Thus, its proteins may be applied to some industrial applications. In addition, *M. thermophilus* is facultatively anaerobic while other species in the genus are strictly aerobic. Moreover, we noticed that the genome of *M. thermophilus* strain HT7-2^T contains several genes grouped into three main mechanisms involved in metal resistance. Thus, *M. thermophilus* strain HT7-2^T has the potential to resist in conditions contaminated by heavy metals and applied to bioremediation. Here, we report the draft genome sequence of *M. thermophilus* strain HT7-2^T.

Genome sequencing of *M. thermophilus* strain HT7-2^T was conducted at the BGI (Wuhan, China) employing Illumina HiSeq2000 sequencing platform. The Illumina data included a 500 bp paired-end library and a

* Corresponding authors.

E-mail addresses: extremelab@126.com (M. Wu), xuxw@sio.org.cn (X.-W. Xu).

Table 1
General features of *M. thermophilus* strain HT7-2^T and MIGS mandatory information.

Property	Term
Classification	Domain <i>Bacteria</i> Phylum <i>Bacteroidetes</i> Class <i>Flavobacteria</i> Order <i>Flavobacteriales</i> Family <i>Flavobacteriaceae</i> Genus <i>Maribacter</i> Species <i>Maribacter thermophilus</i> (Type) strain: HT7-2 ^T (= CGMCC 1.12207 ^T = JCM 18466 ^T)
Gram stain	Negative
Cell shape	Rod
Motility	Motile
Sporulation	Non-sporulating
Temperature range	4–50 °C
Optimum temperature	40–42 °C
pH range; Optimum	5.5–8.5; 5
Carbon source	L-arabinose, D-fructose, D-galactose, D-glucose, lactose, D-mannitol, D-mannose, raffinose and D-xylose
Salinity	0–8% (w/v) NaCl (optimum 2–3%)
Oxygen requirement	Facultatively anaerobic
Biotic relationship	Free-living
Pathogenicity	Non-pathogen
Altitude	Not reported
Submitted to insdc	LDAS00000000.1
Investigation type	Bacteria_archaea
Project name	<i>Maribacter thermophilus</i> strain HT7-2 Genome sequencing
Geographic location (latitude and longitude)	36.05 °N, 120.35 °E
Geographic location (country and/or sea region)	Intertidal zone of Qingdao sea area, P.R. China
Collection date	2010–07
Environment (biome)	Sea water
Environment (feature)	Intertidal zone of sea area
Environment (material)	Sea water
Environmental package	Water
Number of replicons	2
Reference for biomaterial	PMID:25269849
Isolation and growth condition	Marine agar 2216 (MA, BD) at 40–42 °C
Sequencing method	Illumina Hiseq 2000 platform
Assembly	
Assemble method	The assembly result is local assembled and optimized according to paired-end and overlap relationship via mapping reads to Contig
Estimated error rate	Not reported
Method of calculation	De Bruijn Graph
Assembly name	SOAPdenovo 2.04
Finishing strategy	Draft, 150×, 7 contigs in 2 scaffolds

6 k bp mate-pair library. Both two libraries were sequenced with a 2 × 125 bp pair-end strategy. The result reads were filtered with following criteria: 1) Remove reads with a certain proportion of low quality (20) bases; 2) remove reads with a certain proportion of Ns; 3) remove adapter contamination; and 4) remove duplication contamination. Thus, a total of 3406 Mb clean data, which provided 150× coverage of the genome, was gathered. The clean data was used for assembling by performing the SOAPdenovo 2.04 (Luo et al., 2012) and produced 7 contigs in 2 scaffolds.

Gene prediction was performed by using Glimmer v3.02 (Salzberg et al., 1998; Delcher et al., 2007) according to the manufacturers' instruction. Ribosomal RNAs were found by performing RNAmmer v1.2 (Lagesen et al., 2007), and tRNAs were identified by tRNAscan-SE v1.21 (Lowe and Eddy, 1997). To annotate predicted genes, the result from Glimmer was translated and analyzed by RAST (Rapid Annotation using Subsystem Technology) (Aziz et al., 2008). Meanwhile, the predicted genes were searched against the NCBI nr, Swiss-Prot, Pfam, COG database by blastp with e-value < 1e-5. We used KAAS (KEGG

Automatic Annotation Server) (Moriya et al., 2007) to assign translated amino acids into KEGG pathways with BBH method. Genes with signal peptides were defined by using SignalIP v4.1 (Petersen et al., 2011) with default parameters. THMMER 2.0 (Krogh et al., 2001) was performed to identify genes with transmembrane helices. CRISPR repeats were predicted by CRISPRfinder (Grissa et al., 2007a,b). The circular map of genome was obtained by using CGview (Stothard and Wishart, 2005) with an uploaded GFF file.

The draft genome sequence of *M. thermophilus* strain HT7-2^T was assembled into 2 scaffolds with a genome size of 4,050,606 bp and a G + C content of 38.93%. These two scaffolds consist 3633 predicted genes, and 3585 of which are protein coding genes. 41 tRNAs and 6 rRNAs (2 5S rRNAs, 2 16S rRNAs and 2 23S rRNAs) and 1 sRNA 1102 genes were categorized into COG functional groups (Fig. 1).

The genomic features of strain HT7-2^T was summarized in Table 2. Strain HT7-2^T has a relatively higher GC content which may contribute to its higher optimal temperature (*M. forsetii* with GC content 35.20% and *M. antarcticus* with GC content 37.40%).

The analyzed result from BLAST showed that the strain HT7-2^T contained 25 coding genes in lipid metabolism which produces some long-chain saturated lipids that help the strain adapt to the thermophilic environment. In addition, the genome analysis has shown the presence of genes coding anaerobic regulatory protein named Crp/Fnr family transcriptional regulator, FNR is found in bacteria and is involved in oxygen-regulated gene expression (Matsui et al., 2013). Furthermore, the Crp/Fnr super family and its target motifs may reveal the evolutionary distance among the bacteria. The genome also contains heat shock proteins (such as Hsp20 and Hsp90) and cold shock proteins (like cspA) which may help the strain adapt to temperature changes.

The genome of *M. thermophilus* strain HT7-2^T harbors genes related to heavy metal resistance such as RND (resistance-nodulation-cell division) proteins which can cooperate with another two proteins named MFPs and OMFs forming efflux protein complexes that export cations to the outside (Nies, 2003). Meanwhile, a gene coding for cobalt transporter named ZntA from CDF (Cation Diffusion Factor) family was found. Additionally, *M. thermophilus* strain HT7-2^T also contains three P-type ATPases which is driven by ATP hydrolysis. This system can export inorganic cations such as Cu⁺, Ag⁺, Zn²⁺, Cd²⁺ and Pb²⁺. The three systems presented above may offer the strain strong ability to resist heavy metals because of the interaction of different systems. Moreover, dozens of genes coding for multiple antibiotic resistance related protein were detected. These genes are the direct evidence that *M. thermophilus* strain HT7-2^T is a robust strain and may tolerate diverse environments.

The genome sequence has shown that strain HT7-2^T is obligate chemo-organotrophic, heterotrophic containing required coding genes which cover all major metabolic pathways. Additionally, the genome includes several copies of genes for glucide utilization and a set of genes coding for the required enzymes catalyze in degrading macromolecular compounds. Hence, the presence of the above genes which undergo important roles in different metabolism pathways coincides with the physiological features of strain HT7-2^T.

3. Nucleotide sequence accession number

The Whole Genome Shotgun project is deposited in DDBJ/EMBL/GenBank under the accession number LDAS00000000. The version presented in this paper is the first version.

Acknowledgments

We gratefully acknowledge the support of the IBM high performance computing cluster of Bio-macromolecules Analysis Lab, Analysis Center of Agrobiological and Environmental Sciences, Zhejiang University. This work was supported by the China Ocean Mineral Resources R & D

Length: 4,050,606 bp

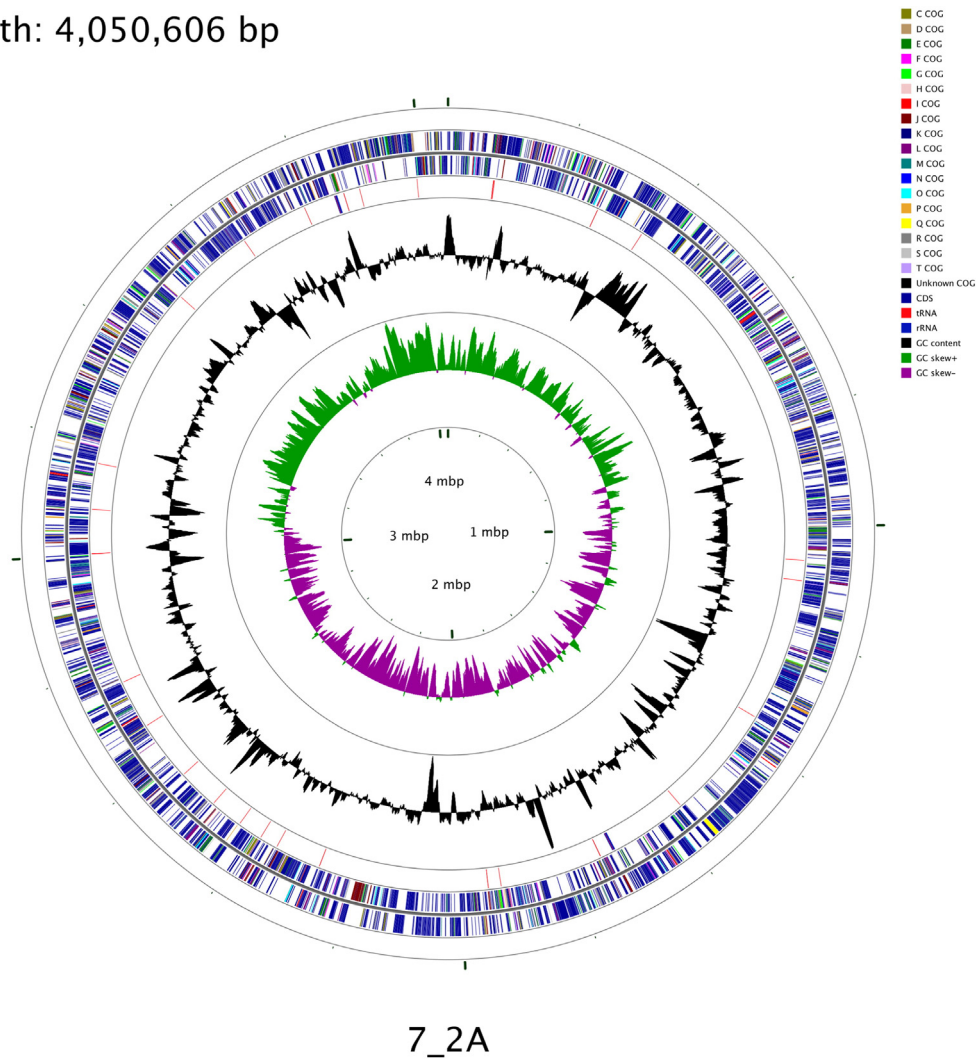


Fig. 1. Graphical map of *M. thermophilus* strain HT7-2^T. From the outside to center: genes on forward strand (different colors represent different COG categories), genes on reverse strand (different colors represent different COG categories), RNA genes (tRNAs red, rRNAs blue), GC content, GC skew. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Association (COMRA) Special Foundation (No. DY125-14-E-02) and National Natural Science Foundation of China (31470005).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.margen.2015.08.003>.

Table 2

Genome features of *M. thermophilus* strain HT7-2^T.

Attribute	Value
Genome size (bp)	4,050,606
DNA coding (bp)	3,681,153
DNA G + C (bp)	1,576,901
DNA scaffolds	2
Total genes	3633
Protein coding genes	3585
RNA genes	48
Genes with function prediction	2641
Genes assigned to COGs	997
Genes with Pfam domains	2186
Genes with signal peptides	311
Genes with transmembrane helices	883
CRISPR repeats	0

References

- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., other authors, 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9.
- Barbeyron, T., Carpentier, F., L'Haridon, S., Schuler, M., Michel, G., Amann, R., 2008. Description of *Maribacter forsetii* sp. nov., a marine *Flavobacteriaceae* isolated from north sea water, and emended description of the genus *Maribacter*. *Int. J. Syst. Evol. Microbiol.* 58, 790–797.
- Cho, K.H., Hong, S.G., Cho, H.H., Lee, Y.K., Chun, J., Lee, H.K., 2008. *Maribacter arcticus* sp. nov., isolated from arctic marine sediment. *Int. J. Syst. Evol. Microbiol.* 58, 1300–1303.
- Delcher, A.L., Bratke, K.A., Powers, E.C., Salzberg, S.L., 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* 23, 673–679.
- Grissa, I., Vergnaud, G., Pourcel, C., 2007a. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57.
- Grissa, I., Vergnaud, G., Pourcel, C., 2007b. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8, 172.
- Hu, J., Yang, Q.Q., Ren, Y., Zhang, W.W., Zheng, G., Sun, C., Pan, J., Zhu, X.F., Zhang, X.Q., other authors, 2015. *Maribacter thermophilus* sp. nov., isolated from an algal bloom in an intertidal zone, and emended description of the genus *Maribacter*. *Int. J. Syst. Evol. Microbiol.* 65, 36–41.
- Jung, Y.T., Lee, J.S., Yoon, J.H., 2014. *Maribacter caenipelagi* sp. nov., a member of the *Flavobacteriaceae* isolated from a tidal flat sediment of the yellow sea in Korea. *Antonie Van Leeuwenhoek* 106, 733–742.
- Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.
- Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., Ussery, D.W., 2007. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108.

- Lo, N., Jin, H.M., Jeon, C.O., 2013. *Maribacter aestuarii* sp. nov., isolated from tidal flat sediment, and an emended description of the genus *Maribacter*. *Int. J. Syst. Evol. Microbiol.* 63, 3409–3414.
- Lowe, T.M., Eddy, S.R., 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., other authors, 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1, 18.
- Matsui, M., Tomita, M., Kanai, A., 2013. Comprehensive computational analysis of bacterial CRP/FNR superfamily and its target motifs reveals stepwise evolution of transcriptional networks. *Genome Biol. Evol.* 5, 267–282.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185.
- Nedashkovskaya, O.I., Kim, S.B., Han, S.K., Lysenko, A.M., Rohde, M., Rhee, M.S., Frolova, G.M., Falsen, E., Mikhailov, V.V., other authors, 2004. *Maribacter* gen. nov., a new member of the family *Flavobacteriaceae*, isolated from marine habitats, containing the species *Maribacter sedimenticola* sp. nov., *Maribacter aquivivus* sp. nov., *Maribacter orientalis* sp. nov. and *Maribacter ulvicola* sp. nov. *Int. J. Syst. Evol. Microbiol.* 54, 1017–1023.
- Nedashkovskaya, O.I., Kim, S.B., Mikhailov, V.V., 2010. *Maribacter stanieri* sp. nov., a marine bacterium of the family *Flavobacteriaceae*. *Int. J. Syst. Evol. Microbiol.* 60, 214–218.
- Nedashkovskaya, O.I., Vancanneyt, M., De Vos, P., Kim, S.B., Lee, M.S., Mikhailov, V.V., 2007. *Maribacter polysiphoniae* sp. nov., isolated from a red alga. *Int. J. Syst. Evol. Microbiol.* 57, 2840–2843.
- Nies, D.H., 2003. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol. Rev.* 27, 313–339.
- Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786.
- Salzberg, S.L., Delcher, A.L., Kasif, S., White, O., 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 26, 544–548.
- Stothard, P., Wishart, D.S., 2005. Circular genome visualization and exploration using CGView. *Bioinformatics* 21, 537–539.
- Weerawongwiwat, V., Kang, H., Jung, M.Y., Kim, W., 2013. *Maribacter chungangensis* sp. nov., isolated from a green seaweed, and emended descriptions of the genus *Maribacter* and *Maribacter arcticus*. *Int. J. Syst. Evol. Microbiol.* 63, 2553–2558.
- Yoon, J.H., Kang, S.J., Lee, S.Y., Lee, C.H., Oh, T.K., 2005. *Maribacter dokdonensis* sp. nov., isolated from sea water off a Korean island, dokdo. *Int. J. Syst. Evol. Microbiol.* 55, 2051–2055.
- Zhang, G.I., Hwang, C.Y., Kang, S.H., Cho, B.C., 2009. *Maribacter antarcticus* sp. nov., a psychrophilic bacterium isolated from a culture of the Antarctic green alga *Pyramimonas gelidicola*. *Int. J. Syst. Evol. Microbiol.* 59, 1455–1459.