Draft genome sequence of \textit{Halopiger salifodinae} KCY07-B2\textsuperscript{T}, an extremely halophilic archaeon isolated from a salt mine

Wei-Yan Zhang\textsuperscript{1}, Jing Hu\textsuperscript{1}, Jie Pan\textsuperscript{1}, Cong Sun\textsuperscript{1}, Min Wu\textsuperscript{1*} and Xue-Wei Xu\textsuperscript{2*}

\textbf{Abstract}

\textit{Halopiger salifodinae} strain KCY07-B2\textsuperscript{T}, isolated from a salt mine in Kuche county, Xinjiang province, China, belongs to the family \textit{Halobacteriaceae}. It is a strictly aerobic, pleomorphic, rod-shaped, Gram-negative and extremely halophilic archaeon. In this work, we report the features of the type strain KCY07-B2\textsuperscript{T}, together with the draft genome sequence and annotation. The draft genome sequence is composed of 83 contigs for 4,350,718 bp with 65.41 \% G + C content and contains 4204 protein-coding genes and 50 rRNA genes.

\textbf{Keywords:} \textit{Halopiger salifodinae}, Archaea, Extreme halophile, Genome, Salt mine

\textbf{Introduction}

The genus \textit{Halopiger}, which belongs to the family \textit{Halobacteriaceae}, was originally established in 2007 by Gutiérrez et al. \cite{1}. The type species of the genus \textit{Halopiger} is \textit{Halopiger xanaduensis} SH-6\textsuperscript{T}. To date, the genus is comprised of three validly published species and two effectively but not validly published species: \textit{H. xanaduensis} \cite{1}, \textit{Halopiger aswanensis} \cite{2}, \textit{Halopiger salifodinae} \cite{3}, \textit{Halopiger djelfamassiliensis} \cite{4} and \textit{Halopiger goleamassiliensis} \cite{5}. The species of the genus were reported to be isolated from hypersaline environments such as salt lake sediment \cite{1, 4, 5}, hypersaline soil \cite{2} and salt mine \cite{3}. All are Gram-negative, strictly aerobic and extremely halophilic \cite{1–5}. In this genus, three genome sequences, including one finished genome sequence \textit{H. xanaduensis} SH-6\textsuperscript{T}, and two draft genome sequences \textit{H. djelfamassiliensis} IIH2\textsuperscript{T} and \textit{H. goleamassiliensis} IIH3\textsuperscript{T}, are available in Standards in Genomic Sciences \cite{4–6}, except \textit{H. aswanensis} 56\textsuperscript{T} which showed highest 16S rRNA gene similarity to \textit{H. xanaduensis} SH-6\textsuperscript{T} (95.8 \%), followed by \textit{H. aswanensis} 56\textsuperscript{T} (95.5 \%), \textit{H. djelfamassiliensis} IIH2\textsuperscript{T} (94.9 \%) and \textit{H. goleamassiliensis} IIH3\textsuperscript{T} (94.8 \%), and shared low sequence similarities (<94.8 \%) to species of other genera. The phylogenetic tree was reconstructed by the neighbor-joining method using MEGA 5 and Kimura's 2-parameter model for distance calculation \cite{8, 9}. The phylogenetic tree was assessed by bootstrapping for 1000 replications, and the consensus tree was shown in Fig. 1.

\textit{H. salifodinae} KCY07-B2\textsuperscript{T} can tolerate high salinity (5.4 M NaCl) and high temperature (50 °C) \cite{3}. Cells lyse in distilled water. The optimal growth condition of strain KCY07-B2\textsuperscript{T} occurred in medium NOM-3 with 2.9–3.4 M NaCl \cite{3}. The optimum temperature was 37–45 °C. The optimum pH was 7.0, with a growth range of...
pH 6.0–8.0 [3]. Cells of strain KCY07-B2T are strictly aerobic, non-motile and pleomorphic rod-shaped (Fig. 2). Several sugars, organic acids and amino acids can serve as sole carbon and energy sources, and amino acids are not required in the growth medium [3]. The features of *H. salifodinae* KCY07-B2T are listed in Table 1.

**Genome sequence information**

**Genome project history**

This genome was selected for sequencing on the basis of its phylogenetic position and 16S rRNA sequence similarity to other members of the genus *Halopiger*. This whole genome shotgun project of strain *H. salifodinae* KCY07-B2T was deposited at DDBJ/EMBL/GenBank under accession number JROF00000000 and the sequence consisted of 83 contigs (further assembling constructed these contigs into 81 scaffolds). Table 2 shows the project information and its association with MIGS version 2.0 compliance [10].

**Growth conditions and genomic DNA preparation**

*H. salifodinae* KCY07-B2T was cultivated aerobically on 37 °C for 4 days in NOM-3 medium, which contains (per liter distilled water) 5.4 g KCl, 0.3 g K₂HPO₄, 0.25 g CaCl₂, 0.25 g NH₄Cl, 26.8 g MgSO₄·7H₂O, 23.0 g MgCl₂·6H₂O, 184.0 g NaCl, 1.0 g yeast extract, 0.25 g fish peptone, 0.25 g sodium formate, 0.25 g sodium acetate, 0.25 g sodium lactate and 0.25 g sodium pyruvate (adjusted to pH 7.0 with 1 M NaOH) [3]. Genomic DNA was extracted using the method described by Marmur [11]. The purity, quality and the concentration of genomic DNA preparation were analyzed by 0.7 % agarose gel electrophoresis with λ-Hind III digest DNA Marker (TaKaRa, Dalian, China) and measured using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific Inc., USA).

**Genome sequencing and assembly**

The genome of *H. salifodinae* KCY07-B2T was sequenced using Solexa paired-end sequencing technology (HiSeq2000 system, Illumina, Inc., USA) [12]. A shotgun library was constructed with a 500 bp-span paired-end library (~500 Mb available reads, ~130-fold genome coverage) and a 2000 bp-span paired-end library (~250 Mb available reads, ~65-fold genome coverage). The sequence data from an Illumina HiSeq 2000 were assembled with SOAPdenovo v.1.05 [13–15]. The final assembly identified 83 contigs and 81 scaffolds (the minimum length is 523 bp) generating a genome size of 4.35 Mb. The quality of the sequencing reads data was estimated by G + C content and sequencing depth correlation analysis.

**Genome annotation**

The tRNAs and rRNAs were identified using trRNAscan-SE [16], RNAmmer [17] and Rfam database [18]; the open reading frames and the functional annotation of translated ORFs were predicted and achieved by using the RAST server online [19, 20]. Classification of some
predicted genes and pathways were analyzed using COGs [21, 22] and KEGG [23–25] databases. Meanwhile, we used CRISPRs web server [26] to predict CRISPRs and InterPro [27, 28] to obtain the GO annotation with the database of Pfam [29].

To estimate the mean level of nucleotide sequence similarity at the genome level between H. salifodinae KCY07-B2T and the genus Halopiger genomes available to date (H. xanaduensis SH-6T, H. djelfamassiliensis IIH2T and H. goleamassiliensis IIH3T), we compared the ORFs only using comparison sequence based in the server RAST [19] at a query coverage of ≥60 % and a minimum nucleotide length of 100 bp.

**Genome properties**

The draft genome sequence of H. salifodinae KCY07-B2T revealed a genome size of 4,350,718 bp (scaffold length) with a 65.41 % G+C content. Of the 4254

| MG5S ID | Property | Term | Evidence code |
|---------|----------|------|---------------|
| Current classification | Domain Archaea | TAS [31] |
| | Phylum Euryarchaeota | TAS [32] |
| | Class Halobacteria | TAS [33, 34] |
| | Order Halobacteriales | TAS [35–37] |
| | Family Halobacteraeaceae | TAS [38, 39] |
| | Genus Halopiger | TAS [1] |
| | Species Halopiger salifodinae | TAS [3] |
| | Type strain; strain KCY07-B2T = JCM 18547T = CGMCC 1.12284T | TAS [3] |
| Gram stain | negative | TAS [3] |
| Cell shape | pleomorphic rods | TAS [3] |
| Motility | non-motile | TAS [3] |
| Sporulation | non-sporulating | NAS |
| Temperature range | 25–50 °C | TAS [3] |
| Optimum temperature | 37–45 °C | TAS [3] |
| pH range; Optimum | 6.0–8.0; 7.0 | TAS [3] |
| Carbon source | acetate, asparagine, citrate, fumarate, D-glucose, L-glutamate, glycite, isoleucine, L-lysine, L-malate, D-mannose, L-serine, D-sorbitol, starch, succinate and L-threonine | TAS [3] |
| Energy metabolish | heterotrophic | IDA |
| MG5S-6 | Habitat | salt mine | TAS [3] |
| MG5S-6.3 | Salinity | 1.9–5.4 M NaCl (optimum 2.9–3.4 M) | TAS [3] |
| MG5S-22 | Oxygen requirement | aerobic | TAS [3] |
| MG5S-15 | Biotic relationship | free-living | IDA |
| MG5S-14 | Pathogenicity | non-pathogenic | NAS |
| MG5S-4 | Geographic location | Kuche county, Akesu area in Xinjiang province, P.R. China | TAS [3] |
| MG5S-5 | Sample collection time | 2009 | IDA |
| MG5S-4.1 | Latitude | not reported | |
| MG5S-4.2 | Longitude | not reported | |
| MG5S-4.4 | Altitude | not reported | |

Evidence codes, IDA Inferred from Direct Assay, TAS Traceable Author Statement (i.e., a direct report exists in the literature), NAS Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [40, 41]
Predicted genes, 4204 were protein-coding genes, and 50 were rRNA genes. There were one 16S rRNA gene, two 23S rRNA genes and two 5S rRNA genes. A total of 2887 genes (68.67 %) were assigned a putative function (Table 3). Table 4 showed the distribution of genes into COG functional categories.

**Table 2** Project information

| MIGS ID | Property       | Term                        |
|---------|----------------|-----------------------------|
| MIGS-31 | Finishing quality | High-quality draft          |
| MIGS-28 | Libraries used | One pair-end 500 bp library and one pair-end 2 Kb library |
| MIGS-29 | Sequencing platforms | Illumina HiSeq 2000          |
| MIGS-31.2 | Fold coverage | 130 x (based on 500 bp library), 65 x (based on 2 Kb library) |
| MIGS-30 | Assemblers | SOAP denovo                   |
| MIGS-32 | Gene calling method | RAST                        |
|        | Locus Tag | LT39                          |
|        | Genbank ID | JROF00000000                 |
|        | Genbank Date of Release | November 17, 2014          |
|        | GOLD ID | GI0079167                     |
|        | NCBI Project ID | 261874                     |
|        | BIOPROJECT ID | PRJNA261874               |
| MIGS 13 | Source Material Identifier | JCM 18547               |
|        | Project relevance | Phylogenetic diversity, Study of the archaean diversity in a salt mine |

**Table 3** Genome statistics of *Halopiger salifodinae* KCY07-B2T, including nucleotide content and gene count levels

| Attribute                        | Value      | % of total* |
|----------------------------------|------------|-------------|
| Genome size (bp)                 | 4,350,718  | 100.00      |
| DNA coding (bp)                  | 3,567,421  | 82.00       |
| DNA G + C (bp)                   | 2,845,805  | 65.41       |
| DNA scaffolds                     | 81         |             |
| Total genes                      | 4254       | 100.00      |
| Protein coding genes             | 4204       | 98.82       |
| RNA genes                        | 50         | 1.18        |
| Pseudo genes                     | not determined | not determined |
| Genes in internal clusters       | not determined | not determined |
| Genes with function prediction   | 2561       | 60.20       |
| Genes assigned to COGs           | 2887       | 67.87       |
| Genes assigned Pfam domains      | 2694       | 63.33       |
| Genes with signal peptides       | 122        | 2.9         |
| Genes with transmembrane helices | 910        | 21.39       |
| CRISPR repeats                   | 3          | 0.07        |

*The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

**Table 4** Number of genes associated with the 25 general COG functional categories

| Code | value | % age* | Description                                      |
|------|-------|--------|--------------------------------------------------|
| J    | 175   | 4.16   | Translation, ribosomal structure and biogenesis  |
| A    | 1     | 0.02   | RNA processing and modification                  |
| K    | 175   | 4.16   | Transcription                                     |
| L    | 120   | 2.85   | Replication, recombination and repair             |
| B    | 6     | 0.14   | Chromatin structure and dynamics                  |
| D    | 25    | 0.59   | Cell cycle control, Cell division, chromosome partitioning |
| V    | 35    | 0.83   | Defense mechanisms                                |
| T    | 131   | 3.12   | Signal transduction mechanisms                    |
| M    | 118   | 2.81   | Cell wall/membrane biogenesis                     |
| N    | 16    | 0.38   | Cell motility                                     |
| U    | 19    | 0.45   | Intracellular trafficking and secretion           |
| O    | 131   | 3.12   | Posttranslational modification, protein turnover, chaperones |
| C    | 255   | 6.07   | Energy production and conversion                  |
| G    | 200   | 4.76   | Carbohydrate transport and metabolism             |
| E    | 306   | 7.28   | Amino acid transport and metabolism               |
| F    | 77    | 1.83   | Nucleotide transport and metabolism               |
| H    | 159   | 3.78   | Coenzyme transport and metabolism                 |
| I    | 115   | 2.74   | Lipid transport and metabolism                    |
| P    | 215   | 5.11   | Inorganic ion transport and metabolism            |
| Q    | 60    | 1.43   | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 550   | 13.08  | General function prediction only                  |
| S    | 299   | 5.45   | Function unknown                                  |
| -    | 1317  | 31.33  | Not in COGs                                      |

*The total is based on the total number of protein coding genes in the genome.

**Insights from the genome sequence**

Strain *H. salifodinae* KCY07-B2T was isolated from a salt mine sample. The experiments showed this strain could grow at 2.9–3.4 M NaCl for optimal growth, and the cells lysed in distilled water. So the analysis of the genome sequence focused on the adaption mechanism of the halophilic archaea in hypersaline-environments. Strain *H. salifodinae* KCY07-B2T mainly utilized “the salt-in strategy” to maintain osmotic balance. According to the annotation of genome sequence, Trk system potassium uptake protein were found, which were responsible for K+ uptake and transport, including 9 copies TrkH genes and 5 copies TrkA genes. Five copies of Kef-type K+ transport proteins, one copy glutathione-regulated potassium-efflux protein KefB and 8 pH adaptation potassium efflux system proteins were found that were related to K+ efflux. And there also existed 8 copies of potassium channel proteins. In addition, the genome contains 13 copies of Na+/H+ antiporter proteins related...
to Na⁺ efflux. The genome of strain *H. salifodinae* KCY07-B²T contains 12 genes related to the synthesis and transport of the compatible-solute glycine betaine for resistance to osmotic stress including: 7 choline-sulfatases, 2 high-affinity choline uptake protein BetTs, 2 glucose-methanol-choline oxidoreductase and 1 glycine betaine transporter OpuD coding genes. These proteins were also related to the metabolic pathway converting choline sulfate to glycine betaine. All these proteins and systems mentioned played an important role in the adaption of osmotic stress in high salt environment.

Currently, three genomes from *Halopiger* species are available. Here, we compare the genome of strain *H. salifodinae* KCY07-B²T with strains *H. xanaduensis* SH-6ᵀ, *H. djelfamassiliensis* IIH²ᵀ and *H. goleamassiliensis* IIH³ᵀ (Table 5). The size of genome of *H. salifodinae* KCY07-B²T (4.35 Mb) is similar to *H. xanaduensis* SH-6ᵀ (4.35 Mb) but larger than that of *H. djelfamassiliensis* IIH²ᵀ (3.77 Mb) and *H. goleamassiliensis* IIH³ᵀ (3.90 Mb). The G + C content of *H. salifodinae* KCY07-B²T (65.41 %) is similar to *H. xanaduensis* SH-6ᵀ (65.18 %) and higher than that of *H. djelfamassiliensis* IIH²ᵀ (64.30 %) but lower than that of *H. goleamassiliensis* IIH³ᵀ (66.06 %). In addition, *H. salifodinae* KCY07-B²T shares a mean genomic sequence similarity of 79.74 %, 80.16 % and 79.17 % with strains *H. xanaduensis* SH-6ᵀ, *H. djelfamassiliensis* IIH²ᵀ and *H. goleamassiliensis* IIH³ᵀ, respectively.

**Table 5** Genomic comparison of *H. salifodinae* KCY07-B²T with three other *Halopiger* species

| Species                  | Strain          | Genome accession number | Genome size (Mb) | G + C content |
|--------------------------|-----------------|-------------------------|------------------|---------------|
| *H. salifodinae* KCY07-B²T | KROFO00000000   | 4.35                    | 65.41            |
| *H. xanaduensis*, SH-6ᵀ   | NC_015666       | 4.35                    | 65.18            |
| *H. djelfamassiliensis* IIH²ᵀ | PRIEB1777     | 3.77                    | 64.30            |
| *H. goleamassiliensis* IIH³ᵀ | PRIEB1780     | 3.90                    | 66.06            |

Species and strain names, genome accession numbers, sizes and G + C contents.

Conclusions

Strain KCY07-B²ᵀ is the third member of the genus *Halopiger* to be described and the fourth whose genome sequence report is available. These data will provide a new perspective of how microorganisms adapt to halophilic environments, and may also provide a pool of functional enzymes that work at higher salty conditions.

**Abbreviations**

NCBI: National Center for Biotechnology Information; EMBL: European Molecular Biology Laboratory; DDBJ: DNA Data Bank of Japan; BLASTN: Basic Local Alignment Search Tool for Nucleotide; MIGS: Minimum Information about a Genome Sequence; RAST: Rapid Annotations using Subsystems Technology; COG: Cluster of Orthologous Groups of proteins; KEGG: Kyoto Encyclopedia of Genes and Genomes; CRISPR: Clustered Regularly Interspaced Short Palindromic repeat sequences; GO: Gene Ontology; DNA: Deoxyribonucleic Acid; 16S rRNA: ribosomal Ribonucleic Acid; JCM: Japan Collection of Microorganisms; CGMCC: China General Microbiological Culture Collection Center; H. salifodinae *H. salifodinae* KCY07-B²ᵀ; *Halopiger* salifodinae *Halopiger salifodinae* KCY07-B²ᵀ; *Halopiger* xanaduensis *Halopiger xanaduensis* SH-6ᵀ; *Halopiger* goleamassiliensis *Halopiger goleamassiliensis* IIH³ᵀ; *Halopiger* djelfamassiliensis *Halopiger djelfamassiliensis* IIH²ᵀ; *Halopiger* goleamassiliensis *Halopiger goleamassiliensis* IIH³ᵀ.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

YWZ designed the study, isolated strain *Halopiger salifodinae* KCY07-B²ᵀ, performed the laboratory experiments, analyzed the genome and wrote the manuscript. JH worked on genome assembly, annotated the genome and discussed the results. JP and CS participated in the analysis of the genome and checked the manuscript. MU and XWX helped to supervise the study and revised the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

We thank Hong Cheng for her help on offering some websites for data analysis. This work was supported by the China Ocean Mineral Resources R & D Association (COMRA) Special Foundation (grant no. DY125-14-E-02) and the Chinese Natural Science Foundation (grant no. 31170001).

**Received**: 20 April 2015 **Accepted**: 24 November 2015

**Published online**: 10 December 2015

**References**

1. Gutiérrez MC, Castillo AM, Kamekura M, Xue Y, Ma Y, Cowan DA, et al. *Halopiger xanaduensis* gen. nov., sp. nov., an extremely halophilic archaean isolated from saline Lake Shangmatala in Inner Mongolia, China. Int J Syst Evol Microbiol. 2007;57:1402–7.
2. Hezayen FF, Gutiérrez MC, Steinbüchel A, Tindall BJ, Rehm BH. *Halopiger aswanensis* sp. nov., a polymer-producing and extremely halophilic archaean isolated from hypersaline soil. Int J Syst Evol Microbiol. 2010b;60:33–7.
3. Zhang WY, Meng Y, Zhu XF, Wu M. *Halopiger salifodinae* sp. nov., an extremely halophilic archaean isolated from a salt mine. Int J Syst Evol Microbiol. 2013;63:3563–67.
4. Hassani I, Robert C, Michelle C, Raoult D, Hacène H, Desnues C. Non-contiguous finished genome sequence and description of *Halopiger djelfamassiliensis* sp. nov. Stand Genomic Sci. 2013;9:60–74.
5. Hassani I, Robert C, Michelle C, Raoult D, Hacène H, Desnues C. Non-contiguous finished genome sequence and description of *Halopiger goleamassiliensis* sp. nov. Stand Genomic Sci. 2014;9:56–69.
6. Anderson I, Tindall BJ, Rohde M, Lucas S, Han J, Lapidus A, et al. Complete genome sequence of *Halopiger xanaduensis* type strain (SH-6ᵀ). Stand Genomic Sci. 2012;6:31–42.
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
8. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25.
9. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol. 2011;28:2731–9.
10. Fields D, Garrity G, Gray T, Morrison N, Selingeng J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26:541–7.
11. Marmur J, Doty P. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J Mol Biol. 1962;5:159–189.
12. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;457:53–9.
13. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 2010;20:265–72.
Acholeplasma, Halanaerobium, Halobacterium, Methanobacterium, Thermococcales, Thermoproteales are the genera...