Marine Genomics 23 (2015) 19–21

Contents lists available at ScienceDirect

Marine Genomics

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

Genomics/technical resources

Complete genome sequence of *Haloarcula* sp. CBA1115 isolated from non-purified solar salts

Ji-Hyun Yun [a], Hye Seon Song [b], Seong Woon Roh [b], Mi-Ja Jung [a], Pil Soo Kim [a], Jin-Woo Bae [a,*]

[a] Department of Life and Nanopharmaceutical Sciences and Department of Biology, Kyung Hee University, Seoul, Republic of Korea
[b] Division of Life Science, Korea Basic Science Institute, Daejeon 305-806, Republic of Korea

ARTICLE INFO

Article history:
Received 30 January 2015
Received in revised form 25 March 2015
Accepted 25 March 2015
Available online 4 April 2015

Keywords:
Archaea
Genome sequence
Haloarcula sp.
Solar salts
Complete genome

ABSTRACT

Haloarcula sp. CBA1115, isolated from non-purified solar salts from South Korea, is a halophilic archaeon belonging to the family Halobacteriaceae. Here, we present the complete genome sequence of the strain *Haloarcula* sp. CBA1115 (4,225,046 bp, with a G + C content of 61.98%), which is distributed over one chromosome and five plasmids. A comparison of the genome sequence of *Haloarcula* sp. CBA1115 with those of members of its closely related taxa showed that the closest neighbor is *Haloarcula hispanica* Y27, a popular model organism for archaeal studies. The strain was found to possess a number of genes predicted to be involved in osmo-regulatory strategies and metal regulation, suggesting that it might be useful for bioremediation in extreme environments.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The *Haloarcula* genus is part of the Halobacteriaceae family within the order Halobacteriales. Strains belonging to the genus *Haloarcula* have been recovered from various saline ecological niches including marine saltern, saline soils, sea water, salt lakes, and fish sources (Becker et al., 2014; Lynch et al., 2012; Namwong et al., 2011). At present, there are ten validly named *Haloarcula* species (Namwong et al., 2011) and ten sequenced genomes of species in this genus (nine published and one unpublished), namely, those of *Haloarcula hispanica* Y27 (Liu et al., 2011), *Har. hispanica* N601 (Ding et al., 2014), *Har. vallismortis* J.F.54, *Har. sinaiiensis* ATCC33800, *Har. californiae* ATCC33799 (Lynch et al., 2012), *Har. marismortui* ATCC43049 (Baliga et al., 2004), *Har. japonica* TR-1, *Har. argentinensis* arg-1, *Har. amylolytica* BD-3 (Becker et al., 2014), and *Har. salaria* H5-DGR. As model organisms, *Haloarcula* have enormous potential for studying various archaebacterial mechanisms, including archaea–virus interactions (CRISPR–Cas system) (Li et al., 2014), metabolism (bioplastic production, carotenoids, and enzymes) (Camacho et al., 2009; Han et al., 2007; Yatsunami et al., 2014), and genetics (gene characterization and RNA structure) (Mallika and Kundua, 2015; Onodera et al., 2013). Also, as members of the *Haloarcula* genus have the ability to thrive in extreme high salt environments, most studies of *Haloarcula* have focused on their dynamic osmo-adaptive strategies (Becker et al., 2014; Miyashita et al., 2015). Despite its importance, the genomes of only three among ten *Haloarcula* strains, including *Har. hispanica* Y27, *Har. hispanica* N601, and *Har. marismortui* ATCC43049, have been completely sequenced. Thus, genomic analysis of another *Haloarcula* strain would broaden our understanding of the molecular mechanisms used by organisms in hypersaline niches and could facilitate the development of biotechnological applications to improve hypersaline or heavy metal-contaminated field sites.

2. Data description

As part of a study to investigate the archaeal diversity of solar salts, strain CBA1115 was isolated from non-purified solar salt from solar saltern (E126°06′16″, N34°35′22″) in South Korea on August 2013. Serially diluted solar salt was spread onto 20% NaCl-contained DSM medium no. 372 at pH 7.0 and at 37 °C. A. representative genomic 16S rRNA sequence of strain CBA1115 was compared with those of other members of the genus *Haloarcula* using NCBI BLAST. The species with the maximum score was *Har. hispanica* Y27 (CP002921), which shared an identity of 99.4% with strain CBA1115. The strain CBA1115 was deposited in the Japan Collection of Microorganisms (=JCM30477).

The complete genome sequence of *Haloarcula* sp. CBA1115 was determined using the PacBio RSII System on a fragment library generated using Covaris® g-TUBEs. A total of 150,292 polymerase reads, amounting to 93,903,265 bases were generated and 132,163 polymerase reads with low quality were filtered. The remaining 18,129 sequences were assembled using PacBio *de novo* assembler software,

* Corresponding author at: Laboratory of Microbial Ecology, Department of Life and Nanopharmaceutical Sciences and Department of Biology, Kyung Hee University, Seoul, Republic of Korea. Tel.: +82 2 961 2312.
E-mail address: baejw@khu.ac.kr (J.-W. Bae).

http://dx.doi.org/10.1016/j.margen.2015.03.012
1874-7787/© 2015 Elsevier B.V. All rights reserved.

Table 1
Summary of the characteristics of various *Haloarcula* isolates. The information of reference genomes was obtained from NCBI database.

Organism	Isolation source	Genome size, Mb	G + C contents, %	Protein encoding sequence	tRNA	rRNA	ANI value, %	Confirmed CRISPRs no. (questionable CRISPRs no.)
Haloarcula sp. CBA1115	Solar salt, Korea	4.23	61.98	4252	50	10	–	2 (3)
Haloarcula amylolytica BD-3	Aibi salt lake, China	4.23	62.10	4156	67	7	94.60	2 (6)
Haloarcula argentinensis arg-1	Salinas Chica, Valdez Peninsula Argentina	4.15	61.10	4105	67	8	89.47	2 (4)
Haloarcula californiae ATCC33799	Salt brine, Baja, Mexico	4.42	60.82	4373	56	6	89.69	5 (5)
Haloarcula hispanica Y27	Solar saltern, Spain	3.89	62.46	3859	48	11	98.04	1 (3)
Haloarcula hispanica N601	Laboratory-derived mutant of strain Y27	3.90	62.47	3918	47	10	98.02	1 (3)
Haloarcula japonica TR-1	Saltern soil, Japan	4.28	61.20	4235	45	8	89.68	0 (4)
Haloarcula marismortui ATCC43049	Dead Sea, Israel	4.27	61.12	4243	50	10	89.73	5 (5)
Haloarcula sinaiiensis ATCC33800	Salt brine, Red Sea, Israel	4.41	60.77	4344	46	4	89.67	3 (9)
Haloarcula vallismortis J.F.54	Salt pools, Death Valley, California	3.92	61.77	3961	67	4	87.91	0 (4)

HGAP3, which generated six circular contigs with 130,184 reads counting up to 377,158,638 bases (89.27-fold coverage). The total length of strain CBA1115 genome was 4,225,046 bp, with a G + C content of 61.98%. The genome consists of a single circular chromosome (3,423,144 bp, with a G + C content of 63.0%) and five plasmids (405,316 bp, 59.9%; 133,853 bp, 55.4%; 116,496 bp, 54.2%; 104,322 bp, 53.7%; 41,915 bp, 61.8%). The genome sequence of *Haloarcula* sp. CBA1115 (BioProject PRJNA271840) has been deposited at the NCBI GenBank database under accession numbers CP010529–CP010534.

Previous studies showed that *Haloarcula* spp. has two chromosomes (Baliga et al., 2004; Ding et al., 2014; Liu et al., 2011), but strain CBA1115 has only one chromosome, whose size is the largest among the published complete genomes of *Haloarcula*. We searched for genome sequence similarities using the average nucleotide identity (ANI) value, which is used to estimate genomic relatedness genomic relatedness between two genome sequences on an available web program (http://www.ezbiocloud.net/ezgenome/ani). The ANI value of *Har. hispanica* Y27 showed that it is the closest known relative of strain CBA1115 (Table 1). Taxa that showed high similarity (>90% ANI value) were further subjected to whole genome alignments using MAUVE alignment 2.3.1 (progressive Mauve alignment with default settings). Regardless of the ANI analysis, Mauve results indicated higher conservation of gene orientation in the alignment between strain CBA1115 and *Har. hispanica* N601 than between strain CBA1115 and *Har. hispanica* Y27 (Fig. 1).

Putative coding sequence prediction and functional annotation of the assembled genome were performed with Rapid Annotation using Subsystem Technology server databases and gene-caller GLIMMER 3.02. RNAmer 1.2 (Lagesen et al., 2007) and ARAGON (Laslett and Canback, 2004) were used to identify rRNA genes and tRNA genes, respectively. The genome encodes a total of 4309 coding DNA sequences (CDS), including 50 tRNAs and 10 rRNA genes. Among the 4252 protein domains analyzed in the CBA1115, 2829 domains (~66.53%) had matches to entries in public databases, as judged by the SEED (Overbeek et al., 2005). The chromosome is predicted to contain 3507 coding sequences, including 50 tRNA genes, three 5S rRNA genes, three 16S rRNA genes, and three 23S rRNA genes. The largest plasmid contains 368 coding sequences, with one 5S rRNA gene, while the other four plasmids contain 108, 132, 133, and 61 coding sequences, respectively. Additionally, we analyzed clustered regularly interspaced short palindromic repeat (CRISPR) arrays

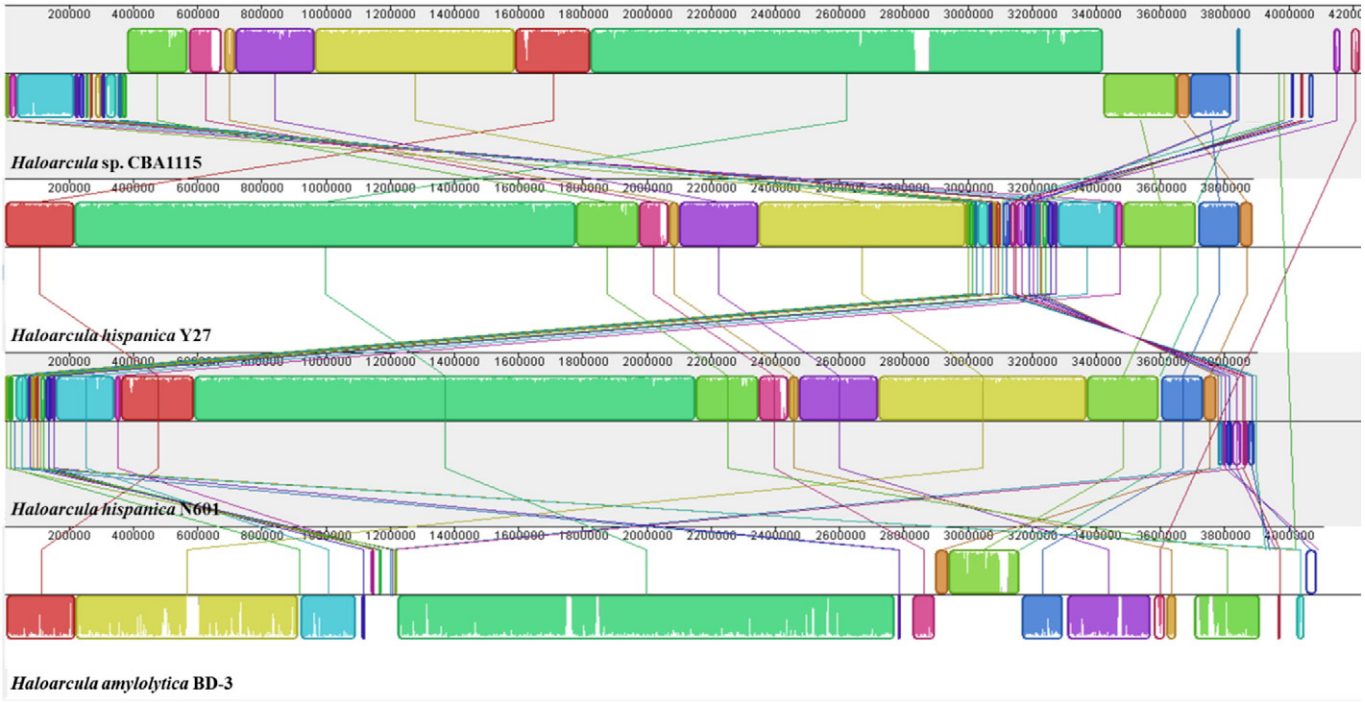

Fig. 1. Whole-genome multiple alignments of the *Haloarcula* sp. CBA1115 genome and its closest relatives (>90% ANI value of genome), *Har. hispanica* and *Har. amylolytica* BD-3. The MAUVE algorithm was used to align the three completely sequenced genomes and one draft genome. The locally collinear block (LCB) indicates a highly homologous region and is connected by the same color bar. The white areas in the LCB indicate similarity gaps in the sequence, and about nine large Indels were identified in *Haloarcula* sp. CBA1115. The genomes were drawn to scale based on the reference *Har. hispanica* Y27 genome.

using the CRISPR finder program (http://crispr.u-psud.fr/Server/). The CBA1115 strain has two CRISPRs with spacers of 26 and 36 direct repeats, respectively, and three putative CRISPRs (Table 1).

Further genomic analysis revealed the presence of numerous metal resistance related genes, including ArsR family transcriptional regulator, ABC transporters and P-type ATPase, in strain CBA1115. As high concentration of metal ions in hypersaline niches such solar salts compared to sea water, microorganisms in hypersaline niches are easily exposed to metal ions. Although metal ions such as calcium, nickel, sodium, potassium and copper are micronutrients, the extreme levels of metal concentration lead to the activation of metal detoxification mechanism in microbes. The numerous transporter systems encoded in the genome of strain CBA1115 may indicate that this strain is able to handle metal accumulation and resist metal stress, and can survive in a heavy metal environment. In addition, strain CBA1115 genome possesses a substantial number of CDS for magnesium transport and solute accumulation (potassium, glutamate, alanine, glutamine, proline, choline, and betaine). In previous reports, haloarchaea species accumulates compatible solutes under osmotic stress (Goh et al., 2011; Kokoeva et al., 2002; Youssef et al., 2014), and magnesium ions were shown to bind to the large ribosomal subunit and contribute to the nucleation and tertiary packing of RNA helices, resulting in enhancement stabilization of RNA structures under stress environments (Klein et al., 2004). Potentially, these genes are a key feature of *Haloarcula* that allows it to adapt to and grow in extreme environments. The genome sequence of *Haloarcula* sp. CBA1115 will enable deeper insights into the molecular mechanisms that underlie survival in extreme environments and may facilitate the development of biotechnological applications to improve the quality of hypersaline or heavy metal-contaminated field sites.

Acknowledgments

This work was carried out with the support of "the Next-Generation Biogreen21 Program (Project Title: Genomic and bioinformatic study on the microbes involving in Korean fermented foods, Project no. PJ008208012014)", Rural Development Administration and "the Strategic Initiative for Microbiomes in Agriculture and Food", Ministry of Agriculture, Food and Rural Affairs, Republic of Korea.

References

Baliga, N.S., Bonneau, R., Facciotti, M.T., Pan, M., Glusman, G., Deutsch, E.W., et al., 2004. Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. Genome Res. 14, 2221–2234.

Becker, E.A., Seitzer, P.M., Tritt, A., Larsen, D., Krusor, M., Yao, A.I., et al., 2014. Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. PLoS Genet. 10, e1004784.

Camacho, R.M., Mateos, J.C., Gonzalez-Reynoso, O., Prado, L.A., Cordova, J., 2009. Production and characterization of esterase and lipase from *Haloarcula marismortui*. J. Ind. Microbiol. Biotechnol. 36, 901–909.

Ding, J.Y., Chiang, P.W., Hong, M.J., Dyall-Smith, M., Tang, S.L., 2014. Complete genome sequence of the extremely halophilic archaeon *Haloarcula hispanica* strain N601. Genome Announc. 2.

Goh, F., Jeon, Y.J., Barrow, K., Neilan, B.A., Burns, B.P., 2011. Osmoadaptive strategies of the archaeon *Halococcus hamelinensis* isolated from a hypersaline stromatolite environment. Astrobiology 11, 529–536.

Han, J., Lu, Q., Zhou, L., Zhou, J., Xiang, H., 2007. Molecular characterization of the phaECHm genes, required for biosynthesis of poly(3-hydroxybutyrate) in the extremely halophilic archaeon *Haloarcula marismortui*. Appl. Environ. Microbiol. 73, 6058–6065.

Klein, D.J., Moore, P.B., Steitz, T.A., 2004. The contribution of metal ions to the structural stability of the large ribosomal subunit. RNA 10, 1366–1379.

Kokoeva, M.V., Storch, K.F., Klein, C., Oesterhelt, D., 2002. A novel mode of sensory transduction in archaea: binding protein-mediated chemotaxis towards osmoprotectants and amino acids. EMBO J. 21, 2312–2322.

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.

Laslett, D., Canback, B., 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32, 11–16.

Li, M., Wang, R., Xiang, H., 2014. *Haloarcula hispanica* CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. Nucleic Acids Res. 42, 7226–7235.

Liu, H., Wu, Z., Li, M., Zhang, F., Zheng, H., Han, J., et al., 2011. Complete genome sequence of *Haloarcula hispanica*, a Model Haloarchaeon for studying genetics, metabolism, and virus–host interaction. J. Bacteriol. 193, 6086–6087.

Lynch, E.A., Langille, M.G., Darling, A., Wilbanks, E.G., Haltiner, C., Shao, K.S., et al., 2012. Sequencing of seven haloarchaeal genomes reveals patterns of genomic flux. PLoS ONE 7, e41389.

Mallika, S., Kundua, S., 2015. Molecular interactions within the halophilic, thermophilic, and mesophilic prokaryotic ribosomal complexes: clues to environmental adaptation. J. Biomol. Struct. Dyn. 33, 639–956.

Miyashita, Y., Ohmae, E., Nakasone, K., Katayanagi, K., 2015. Effects of salt on the structure, stability, and function of a halophilic dihydrofolate reductase from a hyperhalophilic archaeon, *Haloarcula japonica* strain TR-1. Extremophiles 19, 479–493.

Namwong, S., Tanasupawat, S., Kudo, T., Itoh, T., 2011. *Haloarcula salaria* sp. nov. and Haloarcula tradensis sp. nov., isolated from salt in Thai fish sauce. Int. J. Syst. Evol. Microbiol. 61, 231–236.

Onodera, M., Yatsunami, R., Tsukimura, W., Fukui, T., Nakasone, K., Takashina, T., Nakamura, S., 2013. Gene analysis, expression, and characterization of an intracellular alpha-amylase from the extremely halophilic archaeon *Haloarcula japonica*. Biosci. Biotechnol. Biochem. 77, 281–288.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., et al., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 33, 5691–5702.

Yatsunami, R., Ando, A., Yang, Y., Takaichi, S., Kohno, M., Matsumura, Y., et al., 2014. Identification of carotenoids from the extremely halophilic archaeon *Haloarcula japonica*. Front. Microbiol. 5, 100.

Youssef, N.H., Savage-Ashlock, K.N., McCully, A.L., Luedtke, B., Shaw, E.I., Hoff, W.D., Elshahed, M.S., 2014. Trehalose/2-sulfotrehalose biosynthesis and glycine-betaine uptake are widely spread mechanisms for osmoadaptation in the Halobacteriales. ISME J. 8, 636–649.