Comparative genomics of *Tetragenococcus halophilus*

(Received December 9, 2016; Accepted February 3, 2017; J-STAGE Advance publication date: October 17, 2017)

Ikuko Nishimura,1 Yuh Shiwa,2 Atsushi Sato,1 Tetsuya Oguma,1 Hirofumi Yoshikawa,2,3 and Yasuji Koyama1,*

1 Noda Institute for Scientific Research, Chiba 278-0037, Japan
2 Genome Research Center, Tokyo University of Agriculture, Tokyo 156-8502, Japan
3 Department of Bioscience, Tokyo University of Agriculture, Tokyo 156-8502, Japan

Key Words: genome analysis; halotolerant; lactic acid bacteria; *Tetragenococcus halophilus*

*Tetragenococcus halophilus*, a highly halotolerant gram-positive bacterium, is a member of the lactic acid bacteria (LAB) and is routinely isolated from a wide variety of fermented foods, especially soy sauce mash. *T. halophilus* has many variants and plays an important role in the quality of soy sauce. Although the molecular mechanisms underlying its salt tolerance remain to be elucidated, Robert et al. (2000) reported some characteristics of *T. halophilus*, such as the accumulation of compatible solutes in the cell and conversion of choline to glycine betaine, which other LAB don’t synthesize. The genome of *T. halophilus* NBRC 12172 has been completely sequenced (http://www.bio.nite.go.jp/dogan/project/view/TH1). In this study, we have sequenced the genomes of 11 different strains of *T. halophilus* and performed comparative genomics analysis of these along with NBRC 12172, and confirmed previously reported characteristics of *T. halophilus* from genome sequences. The sequenced strains are shown in Table 1. Recently, the species *T. halophilus* was divided in two subspecies: *T. halophilus* subsp. halophilus isolated from high-salt environments and *T. halophilus* subsp. flandriensis isolated from high-sugar environments (Justé et al., 2012). NIRC 00987 and DSM 237667 are the type strains of each subspecies. The other 10 strains (including NBRC 12172) were isolated from soy sauce mash.

The genomic DNAs of the *T. halophilus* strains were extracted using lysozyme, SDS, and Proteinase K, and were purified using NucleoSpin Mini spin columns (Macherey-Nagel). Whole-genome sequencing was performed using an Illumina Genome Analyzer II system. De novo assembly was performed using Velvet (v. 1.1.02) (Zerbino and Birney, 2008) with parameters optimized by the VelvetOptimizer. ORF prediction and annotation were performed by MiGAP (http://www.migap.org/). The draft-sequencing results for the 11 strains are summarized in Table 2. The 11 genomes ranged from 2.3 to 2.6 Mb with a mean G+C content of 35.37 to 35.92%. These results are similar to those of NBRC 12172 (2.6 Mb and 36.04%, respectively). The draft genome sequences have been deposited in the DDBJ/EMBL/GenBank databases under the accession numbers shown in Table 2.

Protein alignment was performed by Scipio (Keller et al., 2008) to determine homolog genes and strain-specific
Table 1. *Tetragenococcus halophilus* strains used in this study.

| Strains        | NRIC No. | ATCC No. | Taxon and sample origin                                      |
|----------------|----------|----------|-------------------------------------------------------------|
| NBRC 12172     |          |          | *T. halophilus* subsp. *halophilus*, isolated from soy sauce mash |
| NISL 7116\(^a\) | 1633     |          | *T. halophilus* subsp. *halophilus*, isolated from soy sauce mash |
| NISL 7118\(^a\) | 1989     |          | *T. halophilus* subsp. *halophilus*, isolated from soy sauce mash |
| NISL 7125\(^a\) | 1990     |          | *T. halophilus* subsp. *halophilus*, isolated from soy sauce mash |
| NISL 7126\(^a\) | 1991     |          | *T. halophilus* subsp. *halophilus*, isolated from soy sauce mash |
| NISL 7128\(^a\) | 1519     |          | *T. halophilus* subsp. *halophilus*, isolated from soy sauce mash |
| D10\(^b\)      |          |          | *T. halophilus* subsp. *halophilus*, isolated from soy sauce mash |
| D-86\(^c\)     |          |          | *T. halophilus* subsp. *halophilus*, isolated from soy sauce mash |
| 11\(^d\)       |          |          | *T. halophilus* subsp. *halophilus*, isolated from soy sauce mash |
| NISL 7121\(^a\) |          | 33315    | *T. halophilus* subsp. *halophilus* type strain, isolated from anchovy |
| NISL 7125\(^a\) |          |          | *T. halophilus* subsp. *halophilus* type strain, isolated from soy sauce mash |
| NISL 7126\(^a\) |          |          | *T. halophilus* subsp. *halophilus* type strain, isolated from thick juice |

a) Sakaguchi (1958), b) Higuchi et al. (1998), c) Uchida and Kanbe (1993), d) Strains in Kikkoman Corporation culture collection, e) Justé et al. (2012).

Fig. 2. Comparison of the module completion ratio (MCR) between the 12 strains of *T. halophilus* sequenced in this study and 13 other strains of LAB.

The MCR in each metabolic functional module was analyzed by MAPLE (Metabolic And Physiological potential. Evaluator) and the results were subjected to a Principal Component Analysis (PCA). Abbreviations of the lactic acid bacteria strains used in this study are as follows: Entfa, *Enterococcus faecalis* V583; Lacbr, *Lactobacillus brevis* ATCC 367; Lacca, *Lactobacillus paracasei* ATCC 334; Laccr, *Lactococcus lactis* subsp. *cremoris* SK11; Lacde, *Lactobacillus delbrueckii* subsp. *bulgaricus* ATCC BAA-365; Lacla, *Lactococcus lactis* subsp. *lactis* IL1403; Lacpl, *Lactobacillus plantarum* WCFS1; Leume, *Leuconostoc mesenteroides* subsp. *mesenteroides* ATCC 8293\(^3\); Oenoe, *Oenococcus oeni* PSU-1; Pedpe, *Pediococcus pentosaceus* ATCC 25745; Strth, *Streptococcus thermophilus* LMD-9.

genes in the 12 strains. The number of homolog genes was 1697 of which identity and coverage were over 60% (Fig. 1). DCM 23766\(^T\) had the most strain-specific genes (334 genes) compared to those in other strains (0–117 genes). Although the functions of these genes are almost unknown, they might play a role in an environment with a high sugar concentration.

Given that the most remarkable characteristic of *T. halophilus* is halotolerance, the related predicted genes were analyzed. In a previous study, it was assumed that the halotolerance mechanism of *T. halophilus* mainly involved the accumulation of compatible solutes (Robert et al., 2000); the genes encoding GbsA and B, which convert choline to glycine betaine, were found only in *T. halophilus* NBRC 12172 among LAB (Ichige, A., pers. comm.). Using BLAST search, we found that all the eleven strains sequenced in this study possessed the gbs\(AB\) gene cluster (Fig. S1A).

In addition, the gene cluster encoding the ABC-type transporters related to the import of compatible solutes and the BCCT (betaine-carnitine-choline transporter)-type transporters related to the symport of protons and compatible solutes were investigated. In the genome of NBRC 12172, 9 genes encoding glycine betaine-binding protein (OpuAC) and one gene encoding a BCCT-type transporter were found (Ichige, A., pers. comm.). The genomes of the other 11 strains were scanned for the presence or absence of the genes encoding OpuAC or BCCT transporter using hmmsearch. The ABC-type transporter was searched against all proteins in the sequenced genomes for PF04069,
hidden Markov model (HMM) of OpuAC in Pfam database, as the query (e-value cutoff $e^{-10}$). Nine homologs of OpuAC that were present in the reference strain NBRC 12172 were found in all the strains, and NISL 7118 and DSM 23766 showed one extra homolog (Fig. S1B). For the BCCT family transporter PF02028, HMM in Pfam, was searched as the query (e-value cutoff $e^{-10}$) and all the strains had homologs that were present in strain NBRC 12172 (Fig. S1C). Generally, LAB have zero to three OpuAC homologs and no BCCT family transporter homolog (Ichige, A., pers. comm.).

$T. halophilus$ strains including NBRC 12172 have more transporters for compatible solutes than those in other LAB. This is supposed to be one of the reasons for the halotolerance of $T. halophilus$.

We performed the metabolic and physiological characterization of the 12 sequenced strains of $T. halophilus$ and the other 13 strains of LAB for which the module completion ratio in each functional module was analyzed using MAPLE (Metabolic And Physiological potentialI Evaluator; http://www.genome.jp/tools/maple/; Takami, 2014; Takami et al., 2012), and the results were subjected to a Principal Component Analysis (PCA) (Fig. 2). PCA was performed using the prcomp function in R (Ihaka and Gentleman, 1996). PCA showed that the $T. halophilus$ strains formed one group in the third quadrant and were clearly separated from the other LAB (Fig. 2). It is supposed that this is mainly because only $T. halophilus$ strains possess the metabolic pathway that converts choline to glycine betaine, considering the eigenvectors and the principal component scores of the 1st and 2nd principal components (Table S1).

In conclusion, we determined the draft genome sequences of 9 strains of $T. halophilus$ isolated from soy sauce mash and 2 standard strains. Their genome sizes were almost similar to that of $T. halophilus$ NBRC 12172, which was used as the reference sequence. Homolog analysis showed that the core genome of the 12 strains consisted of 1697 genes. The number of strain-specific genes in $T. halophilus$ DSM23766 is 334 and is the largest of the 12 strains. This result is in agreement with the fact that only DSM 23766 belongs to a different subspecies. The most characteristic feature of the $T. halophilus$ genome is the presence of more genes related to the biosynthesis of compatible solutes and their transporters compared to those in other LAB.

**Supplementary Materials**

Supplementary figure and table are available in our J-STAGE site (http://www.jstage.jst.go.jp/browse/jgam).

**References**

Higuchi, T., Uchida, K., and Abe, K. (1998) Aspartate decarboxylation encoded on the plasmid in the soy sauce lactic acid bacterium, *Tetragenococcus halophila* D10. *Biosci. Biotechnol. Biochem.*, 62, 1601–1603.

Ihaka, R. and Gentleman, R. (1996) R: A language for data analysis and graphics. *J. Comput. Graph. Stat.*, 299–314.

Justé, A., Van Trappen, S., Verreth, C., Cleenwerck, I., De Vos, P. et al. (2012) Characterization of *Tetragenococcus* strains from sugar thick juice reveals a novel species, *Tetragenococcus osmophilus* sp. nov., and divides *Tetragenococcus halophilus* into two subspecies, *T.
*halophilus* subsp. *halophilus* subsp. nov. and *T. halophilus* subsp. *flandriensis* subsp. nov. *Int. J. Syst. Evol. Microbiol.*, 62, 129–137.

Keller, O., Odonitz, F., Stanke, M., Kollmar, M., and Waack, S. (2008) Scipio: Using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC Bioinformatics*, 9, 278.

Robert, H., Le Marrec, C., Blanco, C., and Jebbar, M. (2000) Glycine betaine, carnitine, and choline enhance salinity tolerance and prevent the accumulation of sodium to a level inhibiting growth of *Tetragenococcus halophilus*. *Appl. Environ. Microbiol.*, 66, 509–517.

Sakaguchi, K. (1958) Studies on the activities of bacteria in soy sauce brewing. Part III. Taxonomic studies on *Pediococcus soyae* nov. sp., the soy sauce lactic acid bacteria. *Bull. Agric. Chem. Soc. Jpn.*, 22, 353–362.

Takami, H. (2014) New method for comparative functional genomics and metagenomics using KEGG MODULE. *In Encyclopedia of Metagenomics*, ed. by Nelson, K. E., Springer-Verlag, New York, pp. 525–539.

Takami, H., Taniguchi, T., Moriya, Y., Kuwahara, T., Kanehisa M. et al. (2012) Evaluation method for the potential functionome harbored in the genome and metagenome. *BMC Genomics*, 13, 699.

Uchida, K. and Kanbe, C. (1993) Occurrence of bacteriophages lytic for *Pediococcus halophilus*, a halophilic lactic acid bacterium. *J. Gen. Appl. Microbiol.*, 39, 429–437.

Zerbino, D. R. and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.*, 18, 821–829.