

# Complete Genomic Sequence of the Equol-Producing Bacterium *Eggerthella* sp. Strain YY7918, Isolated from Adult Human Intestine

Shin-ichiro Yokoyama,<sup>1</sup> Kenshiro Oshima,<sup>2</sup> Izumi Nomura,<sup>3</sup>  
Masahira Hattori,<sup>2</sup> and Tohru Suzuki<sup>3\*</sup>

Department of Environmental and Chemical Research, Industrial Technology Center, Gifu Prefectural Government, 47 Kitaoyobi, Kasamatsu-cho, Hashima-gun, Gifu 501-6064, Japan<sup>1</sup>; Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan<sup>2</sup>; and The United Graduate School of Agricultural Science, Gifu University, 1-1 Yanagido, Gifu 501-1193, Japan<sup>3</sup>

Received 24 June 2011/Accepted 18 July 2011

***Eggerthella* sp. strain YY7918 was isolated from the intestinal flora of a healthy human. It metabolizes daidzein (a soybean isoflavonoid) and produces S-equol, which has stronger estrogenic activities than daidzein. Here, we report the finished and annotated genomic sequence of this organism.**

The complete genomic sequence of *Eggerthella* sp. strain YY7918 (8) was determined by a whole-genome shotgun strategy with the Sanger method. Genomic libraries containing 2-kb inserts were constructed with pTS1 plasmids, and 36,480 sequences were generated, providing 8.1-fold coverage from both ends of the genomic clones. Sequence reads were assembled with the Phred-Phrap-Consed program (1). Remaining gaps between contigs were closed by direct sequencing of the fosmid clones. Prediction and annotation of protein-coding genes were performed by MiGAP (6).

The genome of strain YY7918 consists of a circular 3,123,671-bp chromosome with 56.2% GC content and contains 2,680 predicted protein-coding sequences (CDSs) but no plasmid. It also has 2 sets of rRNA operons and 44 tRNA genes. The strain possesses two sets of clustered, regularly interspaced short palindromic repeats. Similarity at the sequence level (1,772 CDSs) was observed between strain YY7918 and the closely related type strain *Eggerthella lenta* VPI0255 (GenBank accession no. CP001726) (3), with overall genome synteny. However, a reciprocal BLASTP search revealed 913 (34.0%) protein-coding genes that are present in strain YY7918 but absent in *E. lenta* VPI0255<sup>T</sup>. The genome of strain YY7918 is 509 kb shorter than that of *E. lenta* VPI0255<sup>T</sup> (circular; 3,632,260 bp) (3).

The predicted CDSs were submitted to the Kyoto Encyclopedia of Genes and Genomes Automatic Annotation Server (<http://www.genome.jp/tools/kaas/>) (2) and the virulence factor database (VFDB) (<http://www.mgc.ac.cn/VFs/>) (7). We could assign 975 CDSs (36%) to known functions, 108 (4%) as conserved hypothetical genes, and 1,597 (60%) as novel hypothetical genes. As expected, strain YY7918 possesses an incomplete carbohydrate metabolic pathway in KEGG, indicating that the strain is nonsaccharolytic. By contrast, this organism

possesses genes for arginine deiminase (*arcA*), ornithine carbamoyltransferase (*argF* or *agl*), and carbamate kinase (*arcC*), suggesting that the strain utilizes arginine as an important energy source, similarly to *E. lenta* (5). The VFDB result indicates that the strain expresses genes for streptomycin adenylyltransferase (*aadE*) and erythromycin resistance methylase (*ermB*). This organism also carries antibiotic ABC transport systems, such as permease (*yadH*) and ATP-binding protein (*yadG*) genes. These results correspond with our previous report that the strain is resistant to erythromycin. However, the strain is sensitive to aminoglycoside antibiotics despite possessing *aadE* (7). Excluding these elements, other known virulence factors were not present.

This organism also has a notable gene cluster (EGYY\_15730 to EGYY\_15760). Recently, Shimada et al. reported a gene that might encode NADP(H)-dependent daidzein reductase (L-DZNR; DDBJ accession no. AB558141) from *Lactococcus* strain 20-92 (4). The EGYY\_15730 gene has 99.6% homology with both the nucleoside and protein sequences of L-DZNR. This suggests that EGYY\_15730 retains L-DZNR activity. Because EGYY\_15750 and EGYY\_15760 construct a cluster with EGYY\_15730, these genes are expected to be related to equol production from dihydrodaidzein. We will report the characterization of these gene products in the near future.

**Nucleotide sequence accession number.** The sequence data of the *Eggerthella* sp. strain YY7918 genome have been deposited in GenBank/DDBJ/EMBL under accession number AP012211.

We thank K. Furuya, C. Shindo, H. Inaba, E. Iioka, and Y. Hattori for technical support.

This research was supported by a Grant-in-Aid for Scientific Research on Priority Areas 'Comprehensive Genomics' (M.H.), the global COE project 'Genome Information Big Bang' (M.H. and K.O.), and a Grant-in Aid for Scientific Research (C), 20510189 (T.S.), from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Iijima Memorial Foundation for the Promotion of Food Science and Technology, 2011 (T.S.).

\* Corresponding author. Mailing address: The United Graduate School of Agricultural Science, Gifu University, 1-1 Yanagido, Gifu 501-1193, Japan. Phone and fax: 81-58-293-2991. E-mail: [suzuki@gifu-u.ac.jp](mailto:suzuki@gifu-u.ac.jp).

## REFERENCES

1. **Gordon, D., C. Desmarais, and P. Green.** 2001. Automated finishing with Autofinish. *Genome Res.* **11**:614–625.
2. **Moriya, Y., M. Itoh, S. Okuda, A. C. Yoshizawa, and M. Kanehisa.** 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**:W182–185.
3. **Saunders, E., et al.** 2009. Complete genome sequence of *Eggerthella lenta* type strain (VPI 0255<sup>T</sup>). *Stand. Gnom. Sci.* **1**:174–182.
4. **Shimada, Y., et al.** 2010. Cloning and expression of a novel NADP(H)-dependent daidzein reductase, an enzyme involved in the metabolism of daidzein, from equol-producing *Lactococcus* strain 20-92. *Appl. Environ. Microbiol.* **76**:5892–5901.
5. **Sperry, J. F., and T. D. Wilkins.** 1976. Arginine, a growth-limiting factor for *Eubacterium lentum*. *J. Bacteriol.* **127**:780–784.
6. **Sugawara, H., A. Ohyama, H. Mori, and K. Kurokawa.** 2009. Microbial Genome Annotation Pipeline (MiGAP) for diverse users, software demonstration S001-1–2. *In* The 20th International Conference on Genome Informatics (Yokohama). Japanese Society for Bioinformatics, Tokyo, Japan.
7. **Yang, J., L. H. Chen, L. L. Sun, J. Yu, and Q. Jin.** 2008. VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. *Nucleic Acids Res.* **36**:D539–542.
8. **Yokoyama, S., and T. Suzuki.** 2008. Isolation and characterization of a novel equol-producing bacterium from human feces. *Biosci. Biotechnol. Biochem.* **72**:2660–2666.