Draft Genome Sequence of *Weissella oryzae* SG25[T], Isolated from Fermented Rice Grains

Yasuhiro Tanizawa,[a,b] Takatomo Fujisawa,[b] Takako Mochizuki,[b] Eli Kaminuma,[b] Yutaka Suzuki,[c] Yasukazu Nakamura,[b] Masanori Tohno[d]

Department of Computational Biology, Graduate School of Frontier Sciences, the University of Tokyo, Chiba, Japan[a]; Genome Informatics Laboratory, National Institute of Genetics, Mishima, Japan[b]; Department of Medical Genome, Graduate School of Frontier Sciences, the University of Tokyo, Chiba, Japan[c]; National Agriculture and Food Research Organization, National Institute of Livestock and Grassland Science, Nasushiobara, Japan[d]

Weissella oryzae was originally isolated from fermented rice grains. Here we report the draft genome sequence of the type strain of *W. oryzae*. This first report on the genomic sequence of this species may help identify the mechanisms underlying bacterial adaptation to the ecological niche of fermented rice grains.

Address correspondence to Masanori Tohno, tohno@affrc.go.jp.

Rice production is the largest single use of agricultural land (over 11% of the total global cultivated area) and provides approximately 21% of global human per capita energy and 15% of per capita protein (1). Recently, there has been a growing interest in the special use of rice not only as food but also as a nutritional feed for livestock. *Weissella oryzae* SG25[T] was originally isolated from fermented rice grains for livestock (2), suggesting that it is associated with adequate fermentation of rice grain silage and has the potential for biotechnological applications such as an effective inoculant for rice grains.

The genomic DNA of the strain SG25[T] was extracted and purified from cells harvested from de Man, Rogosa, and Sharpe broth (Difco) using Qiagen Genomic-tips. Roche 454 pyrosequencing and Illumina HiSeq 2000 produced 169,062 single-end reads and 557,004,694 pair-end reads, respectively. The former reads were assembled into 109 contigs with Newbler assembler (version 2.7). The latter reads were first extended by an overlap of pair-end reads using the Flash program (3), followed by filtering for sequence quality and read length. Subsequently, the reads were aligned by Burrows-Wheeler Aligner (4) to the contigs obtained from Newbler, and the aligned sequences were subjected to the Columbus module of Velvet (version 1.2.08) (5) to perform alignment-assisted assembly. The obtained sequences were further improved by scaffolding using Opera (6) and by gap-closing using GapFiller (7). The closed gaps were manually verified. The final draft genome sequence consisted of 70 scaffolds. The genome was annotated using the Microbial Genome Annotation Pipeline online server (8), by which coding sequences and rRNAs and tRNAs were predicted, and database searches against RefSeq, TrEMBL, and clusters of orthologous groups (COGs) were conducted. The annotated genome was submitted to the GenomeRefine web service (http://genome.annotation.jp/genomerefine/), which refines annotation and assists registration to the International Nucleotide Sequence Database Collaboration through the DNA Data Bank of Japan.

The total length of the draft genome is 2,129,476 bp, with a G+C content of 39%. It contains 2,228 protein-coding genes, among which 1,264 have been assigned to specific COGs. It also contains 72 tRNAs, 5 copies of 5S rRNA, 1 copy of 16S rRNA, and 1 copy of 23S rRNA. This strain possesses gene clusters for the arginine deiminase (ADI) and agmatine deiminase (AgDI) pathways, both of which are believed to serve as alternative pathways to gain ATP and/or serve for acid tolerance by increasing cytoplasmic pH through NH_3 production (9). Comparative analysis with other lactic acid bacteria revealed that the ADI pathway was similar to those found in the genus *Lactococcus* rather than those in the genus *Weissella* and the AgDI pathway was not found in other members of the genus *Weissella*, suggesting that they were acquired through horizontal gene transfer. These pathways may serve for a yet unidentified fermentation process of rice grains.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. BAWR00000000. The version described here is the first version, BAWR01000000.

ACKNOWLEDGMENTS

REFERENCES

1. Maclean JL, Dawe DC, Hardy B, Hettel GP (ed). 2002. Importance of rice, p 1–8. *In* Rice almanac: source book for the most important economic activity on earth, 3rd ed. CABI Publishing, Wallingford, United Kingdom.

2. **Tohno M, Kitahara M, Inoue H, Uegaki R, Irisawa T, Ohkuma M, Tajima K.** 2013. *Weissella oryzae* sp. nov., isolated from fermented rice grain. Int. J. Syst. Evol. Microbiol. **63:**1417–1420. http://dx.doi.org/10.1099/ijs.0.043612-0.

3. **Magoč T, Salzberg SL.** 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics **27:**2957–2963. http://dx.doi.org/10.1093/bioinformatics/btr507.

4. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.

5. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.

6. **Gao S, Sung W-K, Nagarajan N.** 2011. Opera: reconstructing optimal genomic scaffolds with high-throughput paired-end sequences. J. Comput. Biol. **18:**1681–1691. http://dx.doi.org/10.1089/cmb.2011.0170.

7. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with Gap-Filler. Genome Biol. **13:**R56. http://dx.doi.org/10.1186/gb-2012-13-6-r56.

8. **Sugawara H, Ohyama A, Mori H, Kurokawa K.** 2009. Microbial genome annotation pipeline (MiGAP) for diverse users, abstr S-001–1-2. 20th Int. Conf. Genome Informatics, Yokohama, Japan.

9. **Lucas PM, Blancato VS, Claisse O, Magni C, Lolkema JS, Lonvaud-Funel A.** 2007. Agmatine deiminase pathway genes in *Lactobacillus brevis* are linked to the tyrosine decarboxylation operon in a putative acid resistance locus. Microbiology **153:**2221–2230. http://dx.doi.org/10.1099/mic.0.2007/006320-0.