AMERICAN SOCIETY FOR MICROBIOLOGY

genomeAnnouncements™

First Insights into the Genome Sequence of *Clostridium oryzae* DSM 28571, Isolated from the Soil of a Japanese Rice Field

Anja Poehlein,[a] Anna Lena Gippert,[b] Maaike J. Bierenbroodspot,[b] Rolf Daniel[a]

Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University of Göttingen, Göttingen, Germany[a]; Applied Bioinformatics in Microbiology Course of the Microbiology and Biochemistry M.Sc./Ph.D. Program, Georg-August University of Göttingen, Göttingen, Germany[b]

ABSTRACT *Clostridium oryzae* was originally isolated from the soil of a Japanese rice field. *C. oryzae* represents a novel species within the genus *Clostridium* and is associated with anaerobic rice straw degradation. Here, we present the draft genome sequence of *C. oryzae* DSM 28571 (5.076 Mbp), containing 4,590 predicted protein-coding genes.

Clostridium oryzae is an obligate anaerobic, Gram-positive, and spore-forming bacterium which was isolated from the soil of a Japanese rice field, using rice straw as the substrate (1). Phylogenetic analysis based on gyrase A protein sequences revealed that *C. oryzae* is closely related to some members of the genus *Clostridium sensu stricto* and *Clostridium acetobutylicum*. *C. oryzae* represents a novel species within the genus *Clostridium*, with strain KC3 (DSM 28571) as the type strain (1).

Chromosomal DNA was extracted applying the MasterPure complete DNA purification kit, according to the instructions of the manufacturer (Epicentre, Madison, WI, USA). Illumina paired-end seqeuncing libraries were prepared using the Nextera XT DNA library preparation kit, as recommended by the manufacturer (Illumina, San Diego, CA, USA). Sequencing of DNA was performed employing an Illumina MiSeq instrument and Illumina MiSeq reagent kit version 3, as described by the manufacturer. Trimming of reads using Trimmomatic version 0.36 (2) resulted in 2,856,526 paired-end reads. The assembly performed with the SPAdes genome assembler software version 3.10.0 (3) resulted in 176 contigs (>500 bp), with an average coverage of 135-fold. The obtained contigs were validated using QualiMap version 2.1 (4). The draft genome consists of a single chromosome (5.076 Mbp), with an overall G+C content of 33.1%. Automatic gene prediction was performed by using Prokka (5). The draft genome harbored 4,590 predicted protein-coding genes, including 3,514 genes with a predicted function and 1,076 genes coding for hypothetical proteins. In addition, 12 rRNA genes, 1 transfer-messenger RNA (tmRNA), and 69 tRNA genes were predicted. Interestingly, 3 repeat regions were found in the genome of *C. oryzae*. Two of these regions were in close proximity to a clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) gene cluster present in the genome. The third repeat region was associated with another CRISPR-Cas gene cluster identified in the genome.

In addition to *Bacteroidetes* and the *Alphaproteobacteria*, clostridial species are low-abundance components of rice soil bacterial communities. Rice straw mainly contains approximately 37% cellulose, 37% hemicellulose, and 15% lignin (6). *C. oryzae* cannot utilize lignin, cellulose, or hemicellulose, but it uses carbohydrates as growth substrates. One of the main substrates used by *C. oryzae* is soluble starch. We could identify a gene coding for a starch phosphorylase, which converts starch to α-D-

glucose-6-phosphate (7). *C. oryzae* can also grow on cellobiose and melibiose. During the fermentation of cellobiose, *C. oryzae* produces lactate, butyrate, acetate, formate, H_2, and CO_2 (1). Correspondingly, 3 for butyrate production (*puuE*, *buk*, and *ptb*) and 4 genes for acetate synthesis (*paaK, thiH, ackA,* and *pta*) were detected in the genome sequence. *C. oryzae* has been described as positive for different enzymatic activities, including α-galactosidase, β-glucosidase, and α-arabinosidase activities (1). Correspondingly, genes were detected in the genome which were related to these enzyme types.

Accession number(s). The whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number MZGV00000000. The version described here is the first version, MZGV01000000.

REFERENCES

1. Horino H, Ito M, Tonouchi A. 2015. C. *oryzae* sp. nov., from soil of a Japanese rice field. Int J Syst Evol Microbiol 65:943–951. https://doi.org/10.1099/ijs.0.000042.
2. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
3. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-ell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
4. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28:2678–2679. https://doi.org/10.1093/bioinformatics/bts503.
5. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
6. Tsutsuki K, Ponnamperuma FN. 1987. Behavior of anaerobic decomposition products in submerged soils. Soil Sci Plant Nutr 33:13–33. https://doi.org/10.1080/00380768.1987.10557549.
7. Rathore RS, Garg N, Garg S, Kumar A. 2009. Starch phosphorylase: role in starch metabolism and biotechnological applications. Crit Rev Biotechnol 29:214–224. https://doi.org/10.1080/07388550902926063.