

Draft Genome Sequence of the Yeast *Pseudozyma antarctica* Type Strain JCM10317, a Producer of the Glycolipid Biosurfactants, Mannosylerythritol Lipids

Azusa Saika,^a Hideaki Koike,^b Tomoyuki Hori,^c Tokuma Fukuoka,^a Shun Sato,^a Hiroshi Habe,^a Dai Kitamoto,^a Tomotake Morita^a

Research Institute for Innovation in Sustainable Chemistry, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan^a; Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan^b; Research Institute for Environmental Management Technology, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan^c

The basidiomycetous yeast *Pseudozyma antarctica* is known as a producer of industrial enzymes and the extracellular glycolipids, mannosylerythritol lipids. Here, we report the draft genome sequence of the type strain JCM10317. The draft genome assembly has a size of 18.1 Mb and a G+C content of 60.9%, and it consists of 197 scaffolds.

Received 5 August 2014 Accepted 26 August 2014 Published 25 September 2014

Citation Saika A, Koike H, Hori T, Fukuoka T, Sato S, Habe H, Kitamoto D, Morita T. 2014. Draft genome sequence of the yeast *Pseudozyma antarctica* type strain JCM10317, a producer of the glycolipid biosurfactants, mannosylerythritol lipids. *Genome Announc.* 2(5):e00878-14. doi:10.1128/genomeA.00878-14.

Copyright © 2014 Saika et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Tomotake Morita, morita-tomotake@aist.go.jp.

Pseudozyma antarctica is a basidiomycetous yeast belonging to the class *Ustilaginomycetes*, which includes the plant pathogen, *Ustilago maydis* (1, 2). The species is known to produce attractive enzymes, which are used in a wide range of industrial fields (3, 4). In recent years, the type strain has also received considerable attention as a producer of the glycolipid biosurfactants known as mannosylerythritol lipids (MELs) (5). The type strain of *P. antarctica* is thus a promising organism for the industrial production of several attractive materials; however, little about this strain is known. Here, we report the draft genome sequence of *P. antarctica* JCM10317 (CBS 214.83), the type strain isolated from lake sediment at Antarctica, formerly *Sporobolomyces antarctica* (2).

The paired-end DNA library (insert size: ~500 bp) was prepared using a NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioLabs, Ipswich, MA, USA), and sequencing was performed using the Illumina platform (MiSeq). Sequence data totaling 11,086,474 paired-end reads, each 250 bp in length, were generated. Additionally, a mate-paired library (insert size: ~3200 bp) was prepared using a Nextera Mate Pair Sample Prep Kit (Illumina, San Diego, CA, USA), and sequencing resulted in 10,198,878 mate-paired reads. The Allpaths-LG assembler (6) was used to generate 276 contigs with 65.5× and 69.2× genome coverage from the paired-end and mate paired libraries, respectively. The contigs derived from the two libraries were assembled into 197 scaffolds. Accordingly, the draft genome size of strain JCM10317 was 18.1 Mb (G+C content: 60.9%). The length of the longest scaffold was 2,397 kb, and the N_{50} length was 701,213 kb with eight scaffolds.

Using the database of *U. maydis* UM521 (MIPS *U. maydis* Database, <http://mips.helmholtz-muenchen.de/genre/proj/ustilago>) as references, 6,845 protein-coding genes were automatically predicted using the AUGUSTUS program (7), and 5,901 (86.2%) and 5,877 (85.9%) of the genes were found to be homologous to proteins in the *P. antarctica* T-34 (accession no. DF196767 to DF196793) (8) and RefSeq protein database (9) of the National

Center for Biotechnology Information (BLASTp e-value: $1e-10$; aligned region: $\geq 70\%$ of the reference sequence; sequence identity: $\geq 25\%$), respectively. In addition, 57.8% of the genes were also tentatively assigned protein function(s) based on EuKaryotic Orthologous Groups (KOG) classification by searching against the eggNOG database (10).

The gene responsible for lipase B (CALB), which is a major industrial enzyme, was located on scaffold 7. The five genes encoding erythritol/mannose transferase (*emt1*), acyl transferase (*mac1* and *mac2*), putative transporter (*mmf1*), and acetyl transferase (*mat1*), which are responsible for MEL biosynthesis, were conserved on scaffold 31 as a gene cluster identical to those of the other MEL producers, *P. antarctica* T-34 (11), and shared 94.8%, 91.1%, 86.8%, 93.4%, and 92.4% identity. The genome data of the *P. antarctica* JCM10317 will facilitate understanding of its genomic characteristics and lead to further utilization of the species in a range of industries.

Nucleotide sequence accession numbers. The nucleotide sequence of the *P. antarctica* JCM10317 genome has been deposited in DDBJ/EMBL/GenBank under the accession numbers [BBIZ01000001](https://www.ncbi.nlm.nih.gov/nuccore/BBIZ01000001) to [BBIZ01000276](https://www.ncbi.nlm.nih.gov/nuccore/BBIZ01000276) (as WGS entries) and [DF830068](https://www.ncbi.nlm.nih.gov/nuccore/DF830068) to [DF830264](https://www.ncbi.nlm.nih.gov/nuccore/DF830264) (as 197 entries).

ACKNOWLEDGMENTS

We thank all the members of the Biochemical group at AIST.

This research was financially supported by the Science and Technology Research Promotion Program for Agriculture, Forestry, Fisheries and Food Industry, Japan.

REFERENCES

- Boekhout T. 1995. *Pseudozyma* Bandoni emend. Boekhout, a genus for yeast-like anamorphs of Ustilaginales. *J. Gen. Appl. Microbiol.* 41: 359–366. <http://dx.doi.org/10.2323/jgam.41.359>.
- Boekhout T, Fell JW. 1998. *Pseudozyma* Bandoni emend. Boekhout and a comparison with the yeast state of *Ustilago maydis* (de Candolle) Corda, p 790–797. In Kurtzman CP, Fell JW (ed), *The yeasts: a taxonomic study*, 4th ed. Elsevier Science, Amsterdam, Netherlands.

3. Gotor-Fernández V, Busto E, Gotor V. 2006. *Candida antarctica* lipase B: an ideal biocatalyst for the preparation of nirtogenated organic compounds. *Adv. Synth. Catal.* 348:797–812. <http://dx.doi.org/10.1002/adsc.200606057>.
4. De Mot R, Verachtert H. 1987. Purification and characterization of extracellular α -amylase and glucoamylase from yeast *Candida antarctica* CBS 6678. *Eur. J. Biochem.* 164:643–654. <http://dx.doi.org/10.1111/j.1432-1033.1987.tb11175.x>.
5. Morita T, Fukuoka T, Imura T, Kitamoto D. 2013. Production of mannosylerythritol lipids and their application in cosmetics. *Appl. Microbiol. Biotechnol.* 97:4691–4700. <http://dx.doi.org/10.1007/s00253-013-4858-1>.
6. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 108: 1513–1518. <http://dx.doi.org/10.1073/pnas.1017351108>.
7. Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:62. <http://dx.doi.org/10.1186/1471-2105-7-62>.
8. Morita T, Koike H, Koyama Y, Hagiwara H, Ito E, Fukuoka T, Imura T, Machida M, Kitamoto D. 2013. Genome sequence of the basidiomycetous yeast *Pseudozyma antarctica* T-34, a producer of the glycolipid biosurfactants mannosylerythritol lipids. *Genome Announc.* 1(2):e00064-13. <http://dx.doi.org/10.1128/genomeA.00064-13>.
9. Tatusova T, Ciufu S, Fedorov B, O'Neill K, Tolstoy I. 2014. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* 42:D553–D559. <http://dx.doi.org/10.1093/nar/gkt1274>.
10. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei T, Creevey C, Kuhn M, Jensen LJ, von Mering C, Bork P. 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* 42:D231–D239. <http://dx.doi.org/10.1093/nar/gkt1253>.
11. Morita T, Koike H, Hagiwara H, Ito E, Machida M, Sato S, Habe H, Kitamoto D. 2014. Genome and transcriptome analysis of the basidiomycetous yeast *Pseudozyma antarctica* producing extracellular glycolipids, mannosylerythritol lipids. *PLoS One* 9:e86490. <http://dx.doi.org/10.1371/journal.pone.0086490>.