

# Draft Genome Sequence of the Yeast *Starmerella bombicola* NBRC10243, a Producer of Sophorolipids, Glycolipid Biosurfactants

Tomohiko Matsuzawa,<sup>b</sup> Hideaki Koike,<sup>b</sup> Azusa Saika,<sup>a</sup> Tokuma Fukuoka,<sup>a</sup> Shun Sato,<sup>a</sup> Hiroshi Habe,<sup>a</sup> Dai Kitamoto,<sup>a</sup> Tomotake Morita<sup>a</sup>

Research Institute for Innovation in Sustainable Chemistry, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan<sup>a</sup>;  
Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan<sup>b</sup>

**The yeast *Starmerella bombicola* NBRC10243 is an excellent producer of sophorolipids (SLs) from various feedstocks. Here, we report the draft genome sequence of *S. bombicola* NBRC10243. Analysis of the sequence may provide insight into the properties of this yeast that make it superior for use in the production of functional glycolipids and biomolecules, leading to the further development of *S. bombicola* NBRC10243 for industrial applications.**

Received 9 February 2015 Accepted 13 February 2015 Published 26 March 2015

**Citation** Matsuzawa T, Koike H, Saika A, Fukuoka T, Sato S, Habe H, Kitamoto D, Morita T. 2015. Draft genome sequence of the yeast *Starmerella bombicola* NBRC10243, a producer of sophorolipids, glycolipid biosurfactants. *Genome Announc* 3(2):e00176-15. doi:10.1128/genomeA.00176-15.

**Copyright** © 2015 Matsuzawa 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/4.0/).

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

The ascomycetous yeast *Starmerella bombicola* produces large amounts of sophorolipids (SLs) that consist of a sophorose head group (2-O-β-D-glucopyranosyl-D-glucopyranose), the anomeric carbon atom of which is attached to an ω- or (ω1)-hydroxylated C18 or C16 fatty acid. Massive amounts of SLs are produced by *S. bombicola* from feedstocks as a mixture of lactone-form and acid-form SLs (1, 2). SLs and their derivatives show surface-active and emulsification properties, as well as other beneficial properties such as antimicrobial, anticancer, and antiviral activities, and are useful in the food and detergent industries (3).

Here, we describe the draft genome sequence of *S. bombicola* NBRC10243 (=ATCC 22214) as a typical SL producer. A paired-end DNA library of *S. bombicola* NBRC10243 genomic DNA was sequenced using the MiSeq system (Illumina) to approximately 300-fold coverage, comprising a total of 11,269,740 reads with a length of 2 × 250 nucleotides and an insert size of approximately 300 nucleotides. Running the SOAPdenovo assembler with a k-mer size of 63 (4) using the genomic fragments generated an assembly of 296 contigs (≥1,000 bp), resulting in 9.37 Mb for the whole *S. bombicola* genome (longest contig, 432 kb; shortest contig, 1.0 kb; N<sub>50</sub>, 87.6 kb) with a G+C content of 48.1%.

Protein coding genes were automatically predicted by Augustus (5) using *Saccharomyces cerevisiae*, and the models created resulted in a 4,599 protein-coding gene set. The gene cluster responsible for SL biosynthesis in *S. bombicola* (6) was identified on the two scaffolds. Two oligonucleotide primers, 5'-CTCACACACAACGATTGCAGT ATATTTACC-3' and 5'-CAGGATCGGGCCCTCGCTCCAATGA ATAAC-3', were prepared to amplify the gap region between the two contigs. The obtained gene fragment was sequenced using the conventional Sanger method, after which the two contigs were connected to conserve the complete length of the gene cluster for SL biosynthesis. The first step for SL biosynthesis is terminal hydroxylation of a fatty acid by the reaction of a cytochrome P450

monooxygenase (CYP52M1). SL is then formed via the reactions of two glucosyltransferases (UGTA1 and UgTB1) and an acetyltransferase (AT) and is secreted by an ABC transporter (MDR) (7). The five CDSs of the cluster were identical to *ugt1* (HM440974), *mdr* (HQ660581), *at* (HQ670751), *ugt1* (HM440973), and *cyp52m1* (EU552419), respectively. This genome sequence will provide a novel aspect for the use of *S. bombicola* as a platform organism for the production of various biomolecules.

**Nucleotide sequence accession numbers.** The nucleotide sequence of the *S. bombicola* genome has been deposited in DDBJ/EMBL/GenBank under the accession numbers [BBSW01000001](https://www.ncbi.nlm.nih.gov/nuccore/BBSW01000001) to [BBSW01000295](https://www.ncbi.nlm.nih.gov/nuccore/BBSW01000295) (295 entries).

## ACKNOWLEDGMENTS

We thank all members of the Biochemical group at AIST.

This research received no grants from any funding agencies.

## REFERENCES

- Gao R, Falkeborg M, Xu X, Guo Z. 2013. Production of sophorolipids with enhanced volumetric productivity by means of high cell density fermentation. *Appl Microb Biotechnol* 97:1103–1111. <http://dx.doi.org/10.1007/s00253-012-4399-z>.
- Imura T, Kawamura D, Morita T, Sato S, Fukuoka T, Yamagata Y, Takahashi M, Wada K, Kitamoto D, Fell JW. 2013. Production of sophorolipids from non-edible Jatropha oil by *Starmerella bombicola* NBRC 10243 and evaluation of their interfacial properties. *J Oleo Sci* 62:857–864.
- Van Bogaert IN, Saerens K, de Muynck C, Develter D, Soetaert W, Vandamme EJ. 2007. Microbial production and application of sophorolipids. *Appl Microb Biotechnol* 76:23–34. <http://dx.doi.org/10.1007/s00253-007-0988-7>.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1: <http://dx.doi.org/10.1186/2047-217X-1-18>.

5. 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. 7:62. BMC Bioinformatics. <http://dx.doi.org/10.1186/1471-2105-7-62>.
6. Van Bogaert IN, Holvoet K, Roelants SL, Li B, Lin YC, Van de Peer Y, Soetaert W. 2013. The biosynthetic gene cluster for sophorolipids: a biotechnological interesting biosurfactant produced by *Starmerella bombicola*. Mol Microbiol 88:501–509. <http://dx.doi.org/10.1111/mmi.12200>.
7. Roelants SL, De Maeseneire SL, Ciesielska K, Van Bogaert IN, Soetaert W. 2014. Biosurfactant gene clusters in eukaryotes: regulation and biotechnological potential. Appl Microbiol Biotechnol 98:3449–3461. <http://dx.doi.org/10.1007/s00253-014-5547-4>.