



Genome Sequence of the Lager-Brewing Yeast *Saccharomyces* sp. Strain M14, Used in the High-Gravity Brewing Industry in China

Chunfeng Liu,^{a,b} Qi Li,^{a,b} Chengtuo Niu,^{a,b} Feiyun Zheng,^{a,b} Yongxian Li,^{a,b} Yun Zhao,^{a,b} Xiangsheng Yin^c

Key Laboratory of Industrial Biotechnology, Ministry of Education, Jiangnan University, Wuxi, Jiangsu Province, People's Republic of China^a; Lab of Brewing Science and Engineering, Jiangnan University, Wuxi, Jiangsu Province, People's Republic of China^b; Cargill Malt, Wayzata, Minnesota, USA^c

ABSTRACT Lager-brewing yeasts are mainly used for the production of lager beers. Illumina and PacBio-based sequence analyses revealed an approximate genome size of 22.8 Mb, with a GC content of 38.98%, for the Chinese lager-brewing yeast *Saccharomyces* sp. strain M14. Based on *ab initio* prediction, 9,970 coding genes were annotated.

Beer is one of the most consumed alcoholic beverages, and China's beer production has been the largest in the world for more than 10 years (1). In China, the majority of the beer products are light beer using high-gravity brewing (HGB) with dilution technology (2). Brewing yeasts play important roles during HGB beer fermentation. To satisfy the need of the HGB process, the brewing yeast M14, with suitable characteristics for fermentation in high-gravity wort, was obtained, and it is now widely used in the Chinese brewing industry (3). In this study, we aimed to sequence the genome of the brewing yeast *Saccharomyces* sp. strain M14 and to gain detailed insights into its genomic features. To our knowledge, this is the first report of the genome sequence of an industrial lager-brewing yeast from China.

The genome DNA of M14 was extracted using the Yeast genomic DNA kit (CW0569S; CWBio, Beijing, China). Genomic DNA library construction and draft genome sequencing were performed in Shanghai Personal Biotechnology Co., Ltd. (Shanghai, China) using the Illumina MiSeq system. Quality control procedures removed DNA spike-in, artifacts, and ambiguous or low-quality reads (4, 5). Paired ends having at least 90% of bases with a quality score greater than or equal to Q₂₀ were filtered before assembly. These sequences were assembled *de novo* using the SPAdes (version 3.7.1) software package (6), and the results were calibrated using the Pilon software. To predict genes in the M14 genome, we used a Web server for gene prediction in eukaryotes, AUGUSTUS (7), which is specifically trained for this genome, and one homology-based gene predictor, Exonerate (8). Using a heuristic approach implemented in a homemade pipeline, we combined all predicted gene models to produce a nonredundant set of genes, in which a single best-gene model per locus was selected on the basis of sequence similarity to known proteins. We annotated and classified genes according to four databases: Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Evolutionary Genealogy of Genes: Nonsupervised Orthologous Groups (eggNOG), and Swiss-Prot.

The resulting genome sequence of M14 has an estimated size of 22.84 Mb and a GC content of 38.98%. Three sequencing libraries (PE 400, PE 5K, and S-10K) were sequenced, resulting in 28,064,706 paired reads (8,167,798,336 bp), 18,906,632 paired reads (4,472,881,298 bp), and 2,922,174 paired reads (6,712,756,353 bp), respectively.

Received 25 September 2017 Accepted 2 October 2017 Published 26 October 2017

Citation Liu C, Li Q, Niu C, Zheng F, Li Y, Zhao Y, Yin X. 2017. Genome sequence of the lager-brewing yeast *Saccharomyces* sp. strain M14, used in the high-gravity brewing industry in China. *Genome Announc* 5:e01194-17. <https://doi.org/10.1128/genomeA.01194-17>.

Copyright © 2017 Liu et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Qi Li, liqi@jiangnan.edu.cn.

These reads were assembled in 133 scaffolds, with a scaffold N_{50} value of 575,172 bp. This representative set included 9,970 protein-coding genes. The majority (66%) of the predicted genes contained multiple exons, with an average of 1.00 exons per gene. The average gene density, similar to that of larger scaffolds, was 1.0 kb per gene. Proteins including molecular function proteins, cellular component proteins, and biological process proteins were annotated. We assigned the Swiss-Prot database to 9,936 (99.66%) of the predicted M14 proteins. We also assigned 6,560 (65.80%) proteins to the nonredundant (NR) database, 5,124 (51.39%) proteins to the eggNOG database, 3,724 (37.35%) proteins to the KEGG database, and 2,159 (21.65%) proteins to the GO database.

Accession number(s). This whole-genome shotgun project has been deposited at GenBank under the accession number [MVPU00000000](https://www.ncbi.nlm.nih.gov/nuccore/MVPU00000000). The version described in this paper is the first version, MVPU01000000.

ACKNOWLEDGMENTS

We appreciate Shanghai Personal Biotechnology Co., Ltd. (Shanghai, People's Republic of China) for providing laboratory space and equipment.

This study was financially supported by the National Science Foundation (grants 31571942 and 31601558) and the Program of Introducing Talents of Discipline to Universities (grant 111-2-06).

REFERENCES

- Jiang Y. 2017. China beer yield in 2016. *Liquor Mak Sci Technol* 271:89.
- Sigler K, Matoulková D, Dienstbier M, Gabriel P. 2009. Net effect of wort osmotic pressure on fermentation course, yeast vitality, beer flavor, and haze. *Appl Microbiol Biotechnol* 82:1027–1035. <https://doi.org/10.1007/s00253-008-1830-6>.
- Li Q. 1998. The research of beer flavor and flavor stability. Thesis. Jiang Nan University, Wuxi, Jiangsu, China.
- Lindgreen S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes* 5:337. <https://doi.org/10.1186/1756-0500-5-337>.
- Kelley DR, Schatz MC, Salzberg SL. 2010. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol* 11:R116. <https://doi.org/10.1186/gb-2010-11-11-r116>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Stanke M, Morgenstern B. 2005. AUGUSTUS: a Web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 33:W465–W467. <https://doi.org/10.1093/nar/gki458>.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. <https://doi.org/10.1186/1471-2105-6-31>.