

# Genome Sequence of the Thermotolerant Yeast *Kluyveromyces marxianus* var. *marxianus* KCTC 17555

Haeyoung Jeong,<sup>a</sup> Dae-Hee Lee,<sup>a</sup> Sun Hong Kim,<sup>a</sup> Hyun-Jin Kim,<sup>a</sup> Kyusang Lee,<sup>b</sup> Ju Yeon Song,<sup>c</sup> Byung Kwon Kim,<sup>c</sup> Bong Hyun Sung,<sup>a</sup> Jae Chan Park,<sup>b</sup> Jung Hoon Sohn,<sup>a</sup> Hyun Min Koo,<sup>b</sup> and Jihyun F. Kim<sup>a,c</sup>

Systems and Synthetic Biology Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Yuseong-gu, Daejeon, Republic of Korea<sup>a</sup>; Bio Research Center, Samsung Advanced Institute of Technology (SAIT), Nongseo-dong, Giheung-gu, Yongin, Republic of Korea<sup>b</sup>; Department of Systems Biology, Yonsei University, Seodaemun-gu, Seoul, Republic of Korea<sup>c</sup>

***Kluyveromyces marxianus* is a thermotolerant yeast that has been explored for potential use in biotechnological applications, such as production of biofuels, single-cell proteins, enzymes, and other heterologous proteins. Here, we present the high-quality draft of the 10.9-Mb genome of *K. marxianus* var. *marxianus* KCTC 17555 (= CBS 6556 = ATCC 26548).**

The thermotolerant yeast *Kluyveromyces marxianus* var. *marxianus* KCTC 17555 (= CBS 6556 = ATCC 26548), which was isolated from pozol, Mexican fermented corn dough (4), has been used for the production of industrial enzymes, such as inulinase (8) and  $\beta$ -galactosidase (7). Although *Kluyveromyces lactis* has been recognized as a model organism in the genus *Kluyveromyces* (2), *K. marxianus* has a number of advantages over *K. lactis* for development as a new yeast model and a potential host for biotechnological applications (3). *K. marxianus* can grow on a variety of substrates and at higher temperatures, exhibit a higher specific growth rate, and produce less ethanol in the presence of excessive sugar. In an attempt to acquire the key information for its genetic manipulation, aimed to develop engineered *K. marxianus* KCTC 17555 that can convert inulin-rich plant biomass into ethanol and/or platform biochemicals (5), we sequenced and analyzed its 10.9-Mb genome.

Genome sequencing of KCTC 17555 was carried out using Illumina Genome Analyzer IIx. Preprocessing and *de novo* assembly of the reads were conducted using CLC Genomics Workbench version 4.0.2. Initially, 27,789,153 reads ( $\sim$ 2.92 Gb) produced from paired-end sequencing of a 600-bp insert library at a 151-nt cycle were assembled into 346 contigs ( $>$ 200 bp, total length of 10,793,580 bp) with an  $N_{50}$  size of 99,821 bp. Genome annotation was performed using ERGO Genome Analysis Suite (IG Assets, Inc., IL), which utilizes Fgenesh (9) as the gene prediction tool. Scaffold structure was obtained by aligning 5-kb mate-pair library sequencing reads (40,339,119 pairs; 121-nt cycle) with SSPACE (1). Gap closing was done by the Sanger sequencing of PCR products amplified from the gaps spanning contigs in the Consed (<http://www.phrap.org/>) software environment. To generate a long-range scaffold structure and to validate the assembly with regard to paired-read consistency, 2,976 fosmid (pCC1FOS) end reads were aligned with the preexisting scaffolds. We identified 26-bp telomeric repeats located at the 14 physical ends of the scaffolds from fosmid reads, and complete clone sequencing for three selected fosmids was conducted to determine subtelomeric repeats, a process which also revealed that genome sequences produced through next-generation sequencing are perfectly identical to those obtained from Sanger sequencing. The final assembly consisted of 116 contigs (total length of 10,851,738 bp;  $N_{50}$ , 1,189,284 bp; 40.1% G+C content), most of which can be unambiguously allocated into eight chromosomal groups.

When 4,998 putative proteins of *K. marxianus* KCTC 17555, which were predicted by the ERGO system using its prefinished scaffolds, were subject to BLAST analysis against the UniRef90 database, 91% (4,873) of them had homologs in *K. lactis* (E value  $<$  1E-5). Three key enzymes for xylose dissimilation (xylose reductase, xylitol dehydrogenase, and xylulokinase) that are most similar to the corresponding enzymes from *K. marxianus* NBRC 1777 (6, 10, 11) were also identified, suggesting that this yeast can be used for biofuel production from xylose of lignocellulosic hydrolysates.

**Nucleotide sequence accession numbers.** Sequences from this whole-genome shotgun project have been deposited at DDBJ/EMBL/GenBank under the accession number [AKFM000000000](https://www.ncbi.nlm.nih.gov/nuccore/AKFM000000000). The version described in this paper is the first version, [AKFM010000000](https://www.ncbi.nlm.nih.gov/nuccore/AKFM010000000).

## ACKNOWLEDGMENTS

We thank Won-Hyong Chung and Namshin Kim for helpful comments on the *de novo* assembly strategy.

We are grateful for the financial assistance from the Institute of Planning and Evaluation for Technology of the Ministry for Food, Agriculture, Forestry and Fisheries, Republic of Korea. Financial support also came in part from National Research Foundation grants (2012-0001151 and 2012-0005726) and a Global Frontier Intelligent Synthetic Biology Center grant of the Ministry of Education, Science and Technology, Republic of Korea.

## REFERENCES

1. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27: 578–579.
2. Dujon B, et al. 2004. Genome evolution in yeasts. *Nature* 430:35–44.
3. Fonseca GG, Heinze E, Wittmann C, Gombert AK. 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Appl. Microbiol. Biotechnol.* 79:339–354.
4. Herrera T, Ulloa M, Fuentes I. 1973. Descripción de una especie nueva de *Hansenula* y una variedad nueva de *Candida parapsilosis* aisladas del pozol. *Bol. Soc. Mex. Micol.* 7:17–26.

Received 21 September 2012 Accepted 28 September 2012

Address correspondence to Jihyun F. Kim, [jfk1@yonsei.ac.kr](mailto:jfk1@yonsei.ac.kr), or Jae Chan Park, [parkjc@samsung.com](mailto:parkjc@samsung.com).

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/EC.00260-12

5. Lee KS, et al. 22 August 2012, posting date. Characterization of *Saccharomyces cerevisiae* promoters for heterologous gene expression in *Kluyveromyces marxianus*. *Appl. Microbiol. Biotechnol.* doi:10.1007/s00253-012-4306-7.
6. Lulu L, et al. 15 February 2012, posting date. Identification of a xylitol dehydrogenase gene from *Kluyveromyces marxianus* NBRC1777. *Mol. Biotechnol.* doi:10.1007/s12033-012-9508-9.
7. Martins DB, de Souza CG, Jr, Simoes DA, de Morais MA, Jr. 2002. The  $\beta$ -galactosidase activity in *Kluyveromyces marxianus* CBS6556 decreases by high concentrations of galactose. *Curr. Microbiol.* 44:379–382.
8. Rouwenhorst RJ, Visser LE, Van Der Baan AA, Scheffers WA, Van Dijken JP. 1988. Production, distribution, and kinetic properties of inulinase in continuous cultures of *Kluyveromyces marxianus* CBS 6556. *Appl. Environ. Microbiol.* 54:1131–1137.
9. Salamov AA, Solovyev VV. 2000. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* 10:516–522.
10. Wang R, Zhang L, Wang D, Gao X, Hong J. 2011. Identification of a xylulokinase catalyzing xylulose phosphorylation in the xylose metabolic pathway of *Kluyveromyces marxianus* NBRC1777. *J. Ind. Microbiol. Biotechnol.* 38:1739–1746.
11. Zhang B, Zhang L, Wang D, Gao X, Hong J. 2011. Identification of a xylose reductase gene in the xylose metabolic pathway of *Kluyveromyces marxianus* NBRC1777. *J. Ind. Microbiol. Biotechnol.* 38:2001–2010.