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 inulinas (8) and β-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 (~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 N50 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; N50, 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 AKFM00000000. The version described in this paper is the first version, AKFM01000000.

**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**


