Zymomonas mobilis ZM481 (ATCC 31823) is an ethanol-tolerant strain that can produce the highest level of ethanol in Z. mobilis from glucose in the shortest time. Here, we report a draft genome sequence of ZM481, which can help us understand the genes related to the ethanol tolerance of this strain.

Received 10 February 2016 Accepted 24 February 2016 Published 7 April 2016
Citation Zhao N, Pan Y, Liu H, Cheng Z. 2016. Draft genome sequence of Zymomonas mobilis ZM481 (ATCC 31823). Genome Announc 4(2):e00193-16. doi:10.1128/genomeA.00193-16.
Copyright © 2016 Zhao et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.
Address correspondence to Hongguang Liu, simonliu@mail.neu.edu.cn, or Zhen Cheng, zcheng@stanford.edu.

Zymomonas mobilis quickly metabolizes glucose to ethanol via the unique Entner-Doudoroff pathway, which improves some of the most important technoeconomic factors of production, ethanol yield, and productivity, because of low biomass accumulation and high specific surface area for sugar uptake during fermentation (1). However, as one of the most important bottlenecks, poor tolerance to the high level of accumulated ethanol during fermentation makes it hard to use Z. mobilis in industrial production. Z. mobilis ATCC 31823 (ZM481) is a high-yielding mutant strain derived from Z. mobilis ATCC 31821 (ZM4), which cannot only produce the highest level of ethanol in the shortest time but also get better ethanol tolerance than the original strain (2). These advantages warrant ZM481 as a potential strain for large-scale production of ethanol under high-gravity fermentation. Here, we aim to study the genome information of ZM481 and the genomic differences between ZM481 and ZM4 to understand the genes related to ethanol tolerance.

We sequenced the genome of ZM481 using the Illumina HiSeq 2500 PE126. A library with a fragment length of 500 bp was constructed, and a total of 1 Gbp of clean data from paired-end reads were generated. The genome was assembled using SOAPdenovo 2.04 (3), and 500× coverage resulted in a final assembly of 2,200,722 bp, with a G+C content of 46.06%. The draft genome consists of 30 scaffolds of 2,200,722 bp, with a G+C content of 46.06%. The draft genome consists of 30 scaffolds of >500 bp, and the N50 is 385,689 bp. Tandem repetitive sequences were analyzed using Tandem Repeats Finder 4.04 (4), and 56 tandem repeats were found in the genome. Protein-coding sequences were predicted by Glimmer 3.02 (5), and 86.3% of the genome was composed of coding genes. Next, we used BLAST (http://blast.ncbi.nlm.nih.gov/) to accomplish functional annotation by comparing coding gene with nonredundant protein sequences from KEGG (6), Swiss-Prot (7), COG (8), IPS (9), and CAZy (10), tRNA, tRNA, and small RNA (sRNA) genes were detected by RNAmmer 1.2 (11), tRNAscan-SE 1.23 (12), and Rfam 12.0 (13), respectively. The entire genome of ZM481 contains 5 rRNA, 44 tRNA, and 2 sRNA genes.

Because ZM481 was derived from chemical mutagenesis of ZM4, whose genome was completely sequenced and well annotated in 2005 (14), the genomic differences between ZM481 and ZM4 were investigated. Compared with ZM4, single-nucleotide polymorphisms (SNPs) and insertion-deletions (indels) of ZM481 were analyzed via MUMmer 3.22 (15) and LASTZ 1.02.00 (16). The results showed that 125 SNPs (85.62%) localized in the coding regions, and 21 SNPs (14.38%) fell into noncoding regions. No indels were found in the whole genome of ZM481. Synthony analysis of the ZM481 and ZM4 genomes indicated that they were almost the same. These results highlight that SNPs take major responsibility for the ethanol tolerance of ZM481.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. LSFP00000000. The version described in this paper is the first version, LSFP01000000.

ACKNOWLEDGMENTS
We thank Dawei Zhao and Zhenyu Yang for valuable discussions and assistance for this project.

FUNDING INFORMATION
This work, including the efforts of Ning Zhao and Hongguang Liu, was funded by Fundamental Research Funds for the Central University of China (N130520001 and N142003001). This work, including the efforts of Hongguang Liu, was funded by Liaoning Provincial Educational Commission (L2010995). This work, including the efforts of Hongguang Liu, was funded by National Natural Science Foundation of China (NSFC) (11271141). This work, including the efforts of Zhen Cheng, was funded by U.S. Department of Energy (DOE) (DE-SC0008397). This work, including the efforts of Zhen Cheng, was funded by DOE | Office of Science (SC).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES
1. He MX, Wu B, Qin H, Ruan ZY, Tan FR, Wang JL, Shui ZX, Dai LC, Zhu QL, Pan K, Tang YT, Wang WG, Hu QC. 2014. Zymomonas mobilis: a novel platform for future biorefineries. Biotechnol Biofuels 7:101. http://dx.doi.org/10.1186/1754-6834-7-101.
2. Tribe DE, Rogers PL. 1984. Ethanol production in a continuous process with cell recycle. U.S. patent 4445344.
3. 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 S-M, Peng S, Xiaojian Z, Liu G, Liao X, Li Y, Yang
Zhao et al.

1. H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18. http://dx.doi.org/10.1186/2047-217X-1-18.

2. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580. http://dx.doi.org/10.1093/nar/27.2.573.

3. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679. http://dx.doi.org/10.1093/bioinformatics/btm009.

4. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res 36:D480–D484. http://dx.doi.org/10.1093/nar/gkm882.

5. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kyrutin B, Koonin EV, Krylov DM, Makarova KS, Natale DA, Fedorova ND, Izgi Y, Kyrutin B, Makarova KS, Natale DA. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:1. http://dx.doi.org/10.1186/1471-2105-4-1.

6. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Naka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. http://dx.doi.org/10.1093/bioinformatics/btu031.

7. Kurtz S, Phillippy A, Delcher AL, Mergen M, Smith HO, Bergman NH, Weidman JF, Halpern AL, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. http://dx.doi.org/10.1186/gb-2004-5-2-r12.

8. Harris RS. 2007. Improved pairwise alignment of genomic DNA. Ph.D. thesis. Pennsylvania State University, University Park, PA.