



# Draft Genome Sequence of an Acid-Tolerant Yeast, *Candida zemplinina* NP2, a Potential Producer of Organic Acids

Hyeok-Jin Ko,<sup>a</sup> Hyun Joo Park,<sup>a</sup> Sun Hee Lee,<sup>a</sup> Haeyoung Jeong,<sup>b</sup> Jung-Hoon Bae,<sup>a</sup> Bong Hyun Sung,<sup>a</sup> In-Geol Choi,<sup>c</sup> Jung-Hoon Sohn<sup>a</sup>

Cell Factory Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Republic of Korea<sup>a</sup>; Infectious Disease Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Republic of Korea<sup>b</sup>; Department of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul, Republic of Korea<sup>c</sup>

**ABSTRACT** Here, we report the draft genome sequence of the acid-tolerant yeast *Candida zemplinina* NP2, which was isolated from peach peels. This genome sequence will aid in the understanding of the organism's physiological properties as a potential producer of organic acids in acidic environments.

*Candida zemplinina* is synonymously known as *Starmerella bacillaris* and is an acidophilic yeast species frequently isolated from fruit sources and wine environments (1). Owing to its osmotolerant, psychrotolerant, and acid-tolerant properties (2, 3), *C. zemplinina* has been applied in mixed fermentation with *Saccharomyces cerevisiae* to reduce acetic acid byproducts generated by *S. cerevisiae* in botrytized wine fermentation (4). Recently, organic acids, such as succinic acid, 3-hydroxypropionic acid, and lactic acid, have become increasingly attractive as important platform chemicals in biorefining applications (5). With an ability to grow at a low pH (pH < 2), *C. zemplinina* is explored for its potential as the producer of a variety of organic acids. Here, we report the draft genome sequence of *C. zemplinina* NP2, an acid-tolerant yeast isolated from peach peels, in order to obtain further insight into the acid-tolerant properties of NP2 in fermentation of a variety of organic acids under acidic conditions.

Genome sequencing of *C. zemplinina* NP2 was performed by the Illumina HiSeq 2500 platform using paired-end libraries at the Core Facility Management Center in the Korea Research Institute of Bioscience and Biotechnology (KRIBB). We obtained 15.6 million paired-end reads, with 78.9-fold coverage. Low-quality ( $Q < 30$ ) reads were identified and eliminated by the HTQC program (6). *De novo* assembly of the filtered reads was performed using Velvet, version 1.2.10, and VelvetOptimiser, version 2.2.5 (7). The final assembly was 9,311,634 bp, with 117 scaffolds and 145 gaps. The G+C content was 39.46%, and 131 tRNA-coding sequences, including 2 tRNA-like pseudogenes, were identified by tRNAscan-SE (8). The  $N_{50}$  value was 481,187 bp, and the length of the longest contig was 1,363,694 bp. Gene prediction of the genome sequence was performed using AUGUSTUS with a training set of *Candida albicans* (9). In total, 3,741 protein-coding genes were predicted in the draft genome, and 3,285 of these candidates were functionally characterized using an InterProScan search against the InterPro protein signature databases (10).

**Accession number(s).** The nucleotide sequence has been deposited in GenBank under the accession number [NQLE00000000](https://www.ncbi.nlm.nih.gov/nuclink/NQLE00000000).

## ACKNOWLEDGMENTS

This work was supported by the Advanced Biomass R&D Center of the Global Frontier Program (grant 2010-0029737) through the National Research Foundation of

Received 23 August 2017 Accepted 31 August 2017 Published 28 September 2017

**Citation** Ko H-J, Park HJ, Lee SH, Jeong H, Bae J-H, Sung BH, Choi I-G, Sohn J-H. 2017. Draft genome sequence of an acid-tolerant yeast, *Candida zemplinina* NP2, a potential producer of organic acids. Genome Announc 5:e01052-17. <https://doi.org/10.1128/genomeA.01052-17>.

**Copyright** © 2017 Ko 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 Jung-Hoon Sohn, [sohn4090@kribb.re.kr](mailto:sohn4090@kribb.re.kr).

H.-J.K. and H.J.P. contributed equally to this work.

Korea, funded by the Ministry of Science and ICT, the Renewable Energy Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), and the Research Initiative Program of KRIBB.

## REFERENCES

1. Masneuf-Pomarede I, Juquin E, Miot-Sertier C, Renault P, Laizet Y, Salin F, Alexandre H, Capozzi V, Cocolin L, Colonna-Ceccaldi B, Englezos V, Girard P, Gonzalez B, Lucas P, Mas A, Nisiotou A, Sipiczki M, Spano G, Tassou C, Bely M, Albertin W. 2015. The yeast *Starmerella bacillaris* (synonym *Candida zemplinina*) shows high genetic diversity in winemaking environments. *FEMS Yeast Res* 15:fov045. <https://doi.org/10.1093/femsyr/fov045>.
2. Sipiczki M. 2003. *Candida zemplinina* sp. nov., an osmotolerant and psychrotolerant yeast that ferments sweet botrytized wines. *Int J Syst Evol Microbiol* 53:2079–2083. <https://doi.org/10.1099/ijs.0.02649-0>.
3. Sipiczki M. 2004. Species identification and comparative molecular and physiological analysis of *Candida zemplinina* and *Candida stellata*. *J Basic Microbiol* 44:471–479. <https://doi.org/10.1002/jobm.200410449>.
4. Rantsiou K, Dolci P, Giacosa S, Torchio F, Tofalo R, Torriani S, Suzzi G, Rolle L, Cocolin L. 2012. *Candida zemplinina* can reduce acetic acid produced by *Saccharomyces cerevisiae* in sweet wine fermentations. *Appl Environ Microbiol* 78:1987–1994. <https://doi.org/10.1128/AEM.06768-11>.
5. Xiao H, Shao Z, Jiang Y, Dole S, Zhao H. 2014. Exploiting *Issatchenkia orientalis* SD108 for succinic acid production. *Microb Cell Fact* 13:121. <https://doi.org/10.1186/s12934-014-0121-4>.
6. Yang X, Liu D, Liu F, Wu J, Zou J, Xiao X, Zhao F, Zhu B. 2013. HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics* 14:33. <https://doi.org/10.1186/1471-2105-14-33>.
7. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
8. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–W57. <https://doi.org/10.1093/nar/gkw413>.
9. Hoff KJ, Stanke M. 2013. WebAUGUSTUS—a Web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res* 41:W123–W128. <https://doi.org/10.1093/nar/gkt418>.
10. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.