# Draft Genome Sequence of a *Kazachstania humilis* Strain Isolated from Agave Fermentation

Luis F. García-Ortega,[a] Maritrini Colón-González,[b,c] Iván Sedeño,[b] Erick Santiago-Garduño,[b] J. Abraham Avelar-Rivas,[c] Manuel R. Kirchmayr,[d] Alexander DeLuna,[c] Luis Delaye,[a] Lucia Morales,[b] Eugenio Mancera[a]

aDepartamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del IPN, Unidad Irapuato, Irapuato, Mexico
bLaboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Querétaro, Mexico
cUnidad de Genómica Avanzada (Langebio), Centro de Investigación y de Estudios Avanzados del IPN, Irapuato, Mexico
dÁrea de Biotecnología Industrial, Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco A.C. (CIATEJ), Zapopan, Jalisco, Mexico

**ABSTRACT** The ascomycetous yeast *Kazachstania humilis* is an active species in back-slopped sourdough and in the spontaneous fermentation of several traditional foods and beverages. Here, we report the draft genome sequence of a *K. humilis* strain isolated from agave must from a traditional distillery in Mexico.

The ascomycetous yeast *Kazachstania humilis* (*Candida humilis*, *Candida milleri*) is an abundant fungus during the preparation of a variety of fermented foods and beverages (1–5). Most reported isolates are from Asia and Africa, but there are also occurrences in North America (Global Biodiversity Information Facility database), including strains from a natural tequila fermentation in Mexico (6). Despite its wide geographical distribution and its potential for food and beverage production, there is little information about the genome of this species. Here, we report the complete genome assembly of a strain of *K. humilis* that was isolated from fermenting agave must in a traditional distillery in Mexico. The alcoholic beverages produced in these distilleries, such as mezcal, are obtained from the distillation of cooked and fermented juice of *Agave* sp. plants. These artisanal processes are characterized by open and "spontaneous" fermentations in which producers use no commercial inoculum, thus relying on environmental microorganisms.

The *K. humilis* strain (YMX004033) was isolated from a fermentation of cooked agave juice from a distillery located in the Mexican state of Zacatecas. For strain isolation from nitrogen-frozen must containing 25% glycerol, 1 mL was used to inoculate 20 mL of *Saccharomyces sensu stricto* enrichment medium containing 6% ethanol (7). When signs of microbial growth were detected at 25°C, a $10^{-5}$ dilution was plated in Wallerstein Laboratory (WL) medium (Sigma), and single colonies were identified by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS). Strain identity was confirmed from an axenic culture by internal transcribed spacer (ITS)/5.8 rDNA sequencing using primers ITS-1 and ITS-4 (8). The ITS sequence of strain YMX004033 was similar enough to be considered conspecific with the *K. humilis* strains with available sequences at the UNITE database (9); there were 13 different nucleotides (2.45%) compared with the type strain CBS5658 and 11 (2.07%) compared with a tequila fermentation strain also from Mexico [UWO(PS)92-219].

For short-read sequencing, genomic DNA was purified using the MasterPure DNA purification kit, and it was sequenced using the DNBSeq platform (BGI, China), generating 38,025,924 paired-end 150-bp reads. Adapters and low-quality reads (18.54%) were removed with fastp v0.20.0 (10). Additionally, high-molecular-weight DNA obtained with the Qiagen Genomic-tip (20/G) protocol, and fragmented to 20 kb with Covaris G-tubes, was used to prepare a sequencing library with the SQK-LSK109 Ligation protocol from Oxford Nanopore. Sequencing of this library was done in a FLO-MIN106D Spot-ON flow cell vR9. A total of

Address correspondence to Lucia Morales, lmorales@liigh.unam.mx, or Eugenio Mancera, eugenio.mancera@cinvestav.mx.

43,041 raw reads (457.38 Mb) with an estimated $N_{50}$ value of 13.96 kb were obtained after base calling with Guppy v3.2.10. *De novo* genome assembly was performed with SMARTdenovo (11) after read trimming, filtering, and correction with Porechop v0.2, Filtlong v0.2, and Canu v1.8 (12), respectively. The final assembly was polished using two rounds of medaka v0.12.1 with the long reads, followed by three rounds of Pilon v1.23 (13) using the short reads, resulting in 21 scaffolds, with a total length of 14.96 Mb, an $N_{50}$ value of 1.05 Mb, 49.16% GC content, and 30× mean coverage. Using BUSCO v4.0.5 (14), genome completeness was estimated to be 94.5% (Saccharomycetes odb10, $n = 2{,}137$) with 2% of duplicates. All programs were run with default parameters. The genome sequence described here will contribute to our understanding of the poorly described genome diversity of yeasts from Mesoamerica and will enable future research on this species of worldwide importance for food and beverage production.

**Data availability.** The data have been deposited at NCBI under BioProject number PRJNA765343, accession number JAIWYV000000000 (draft genome sequence), SRA numbers SRR16962313 (Nanopore reads) and SRR16016011 (DNBSeq reads), and accession number OK247582 (ITS sequence).

## REFERENCES

1. Gullo M, Romano AD, Pulvirenti A, Giudici P. 2003. Candida humilis: dominant species in sourdoughs for the production of durum wheat bran flour bread. Int J Food Microbiol 80:55–59. https://doi.org/10.1016/S0168-1605(02)00121-6.
2. Pulvirenti A, Solieri L, Gullo M, De Vero L, Giudici P. 2004. Occurrence and dominance of yeast species in sourdough. Lett Appl Microbiol 38:113–117. https://doi.org/10.1111/j.1472-765x.2003.01454.x.
3. Rogalski E, Ehrmann MA, Vogel RF. 2020. Role of Kazachstania humilis and Saccharomyces cerevisiae in the strain-specific assertiveness of Fructilactobacillus sanfranciscensis strains in rye sourdough. Eur Food Res Technol 246:1817–1827. https://doi.org/10.1007/s00217-020-03535-7.
4. Guan Q, Zheng W, Huang T, Xiao Y, Liu Z, Peng Z, Gong D, Xie M, Xiong T. 2020. Comparison of microbial communities and physiochemical characteristics of two traditionally fermented vegetables. Food Res Int 128:108755. https://doi.org/10.1016/j.foodres.2019.108755.
5. You L, Zhou RQ, Tan Y, Wang T, Qiao ZW, Zhao D. 2021. Distribution and function of Kazachstania yeast in the fermentation of strong flavor baijiu. Biotechnol Bull 37:108–116. http://biotech.aiijournal.com/EN/10.13560/j.cnki.biotech.bull.1985.2020-1159.
6. Lachance MA. 1995. Yeast communities in a natural tequila fermentation. Antonie Van Leeuwenhoek 68:151–160. https://doi.org/10.1007/BF00873100.
7. Liti G, Warringer J, Blomberg A. 2017. Isolation and laboratory domestication of natural yeast strains. Cold Spring Harb Protoc 2017:pdb.prot089052. https://doi.org/10.1101/pdb.prot089052.
8. Sylvester K, Wang QM, James B, Mendez R, Hulfachor AB, Hittinger CT. 2015. Temperature and host preferences drive the diversification of Saccharomyces and other yeasts: a survey and the discovery of eight new yeast species. FEMS Yeast Res 15:fov002. https://doi.org/10.1093/femsyr/fov002.
9. UNITE Community. 2017. UNITE general FASTA release. UNITE Community version 01.10.2017 560. https://doi.org/10.15156/BIO/587475.
10. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
11. Liu H, Wu S, Li A, Ruan J. 2021. SMARTdenovo: a de novo assembler using long noisy reads. Gigabyte 1. https://doi.org/10.46471/gigabyte.15.
12. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
13. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
14. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics (Oxford, England) 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.