AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequences of Two Isolates of the Yeast *Kazachstania servazzii* Recovered from Soil in Ireland

Lynne Faherty,[a] Clifton Lewis,[a] Matt McElheron,[a] Niall Garvey,[a] Róisín Duggan,[a] Ben Shovlin,[a] Tadhg Ó Cróinín,[a] Kevin P. Byrne,[b] Caoimhe E. O'Brien,[a] Kenneth H. Wolfe,[b] Geraldine Butler[a]

[a]School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Dublin, Ireland
[b]School of Medicine, Conway Institute, University College Dublin, Dublin, Ireland

ABSTRACT We sequenced two isolates of *Kazachstania servazzii*, UCD13 and UCD335, from soil in Ireland. Heterozygosity in these diploid genomes differs 19-fold between the two strains. Most currently available *K. servazzii* genome sequences come from Korean kimchi isolates, so our data will facilitate analysis of diversity in this species.

Kazachstania servazzii is an ascomycete yeast in the family Saccharomycetaceae. It was formerly called *Saccharomyces servazzii*. The type strain was isolated from soil in Finland in 1967 (1). *K. servazzii* is not pathogenic. It contributes to enhanced flavor through isoamyl alcohol production in fermented foods, including Camembert cheese, kimchi, and sourdough (2–4). It has also been implicated in changing flavor profiles through ethyl decanoate production during Chinese liquor fermentation (5). Conversely, *K. servazzii* is associated with food spoilage, such as swelling of pizza packaging due to gas production (3) and formation of undesirable white yeast colonies on kimchi (6).

Genome survey sequencing of the *K. servazzii* type strain was reported in 2000 (7), and the whole-genome sequence of an isolate from kimchi (strain CBA6004) was recently reported (6). The genome sequence of a second kimchi isolate is also available (NCBI BioProject accession number PRJNA390859).

Strain UCD13 was isolated in 2017 from a soil sample from woodland in County Tipperary, Ireland (global positioning system [GPS] coordinates, 52.52355, −8.13510). UCD335 was isolated in 2018 from soil in a Dublin, Ireland, garden (GPS coordinates, 53.28961, −6.28822). The internal transcribed spacer (ITS) region of rRNA genes was amplified using the primers ITS1 and ITS4 (8) and sequenced using Sanger technology (GenBank accession numbers MN540706 and MN540707).

The yeasts were grown in yeast extract-peptone-dextrose broth. Total genomic DNA was isolated by phenol-chloroform extraction and purified using a Zymo Research DNA Clean & Concentrator-25 kit (UCD13) or a Qiagen QiaAMP DNA minikit (UCD335). Libraries were generated and sequenced by BGI Tech Solutions (Hong Kong). Genomic DNA (1 μg) was fragmented using Covaris, purified with an AxyPrep Mag PCR cleanup kit, and end repaired; A tails were added by using an A-tailing mix and incubating at 37°C for 30 min. Illumina adapters were ligated by incubating at 16°C for 16 h. Insert sizes of ~800 bp were selected, and 150 bases were sequenced from each end with an Illumina HiSeq 4000 platform, yielding 6.4 million spots for UCD13 and 9.6 million spots for UCD335. Scaffolds were assembled using Redundans v0.14a (9). All parameters used for sequence assembly and analysis are available at https://doi.org/10.6084/m9.figshare.9963923.

QUAST v4.6.1 (10) showed that the UCD13 assembly totaled 12.0 Mb, with a G+C content of 34.2%. The N_{50} value was 169,222 bp, and the L_{50} value was 22. The UCD335 assembly was 11.76 Mb, the N_{50} value was 281,905 bp, and the L_{50} value was 14. The

Citation Faherty L, Lewis C, McElheron M, Garvey N, Duggan R, Shovlin B, Ó Cróinín T, Byrne KP, O'Brien CE, Wolfe KH, Butler G. 2019. Draft genome sequences of two isolates of the yeast *Kazachstania servazzii* recovered from soil in Ireland. Microbiol Resour Announc 8:e01257-19. https://doi.org/10.1128/MRA.01257-19.

Editor Antonis Rokas, Vanderbilt University

Copyright © 2019 Faherty et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Geraldine Butler, geraldine.butler@ucd.ie.

Received 4 October 2019
Accepted 10 October 2019
Published 31 October 2019

mitochondrial genomes were identified by tBLASTn searches using *S. cerevisiae* mitochondrial proteins as queries and have GenBank accession numbers VWSC01000085 and VWSB01000061.

Variant analysis used the Burrows-Wheeler Aligner MEM (BWA-MEM) algorithm v0.7.12-r1039 (11), SAMtools v1.1.19 (12), and the Genome Analysis Toolkit (GATK) v4.0.1.2 (13). Variants were filtered by removing clusters (5 variants within 20 bp) that were assumed to result from poor read alignment and by applying GATK filters (QualByDepth [QD], <2.0; Mapping Quality [MQ], <40.0; FisherStrand [FS], >60.0; StrandOddsRatio [SOR], >3.0; MappingQualityRankSumTest [MQRankSum], less than −12.5; and ReadPosRankSumTest [ReadPosRankSum], less than −8.0). There were approximately 73,500 heterozygous single nucleotide polymorphisms (SNPs) and 9,400 insertions/deletions (indels) in UCD13. In contrast, UCD335 had only approximately 3,750 heterozygous SNPs and 960 indels. Analysis of biallelic SNP frequency distribution indicated that UCD13 has a heterozygous diploid genome (i.e., a 50/50 distribution of the reference and nonreference alleles was observed). The level of heterozygosity in UCD13 (6.1 SNPs per kb) is slightly less than the highest level seen in a survey of 1,011 *S. cerevisiae* isolates (14).

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers VWSB00000000 and VWSC00000000 (UCD335 and UCD13, respectively), and the raw reads are at the Sequence Read Archive under accession numbers SRX6818742 and SRX6818741. The ITS sequences have accession numbers MN540706 and MN540707, and the mitochondrial genomes are at VWSB01000061 and VWSC01000085. The data are also available under BioProject accession number PRJNA564535.

ACKNOWLEDGMENTS

This work was supported by an undergraduate teaching award from University College Dublin and by Science Foundation Ireland (13/IA/1910). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Thanks go to Sean Bergin and to the boys from 4th Class, St Mary's Boys National School, Rathfarnham, Dublin, for help collecting soil samples.

REFERENCES

1. Capriotti A. 1967. *Saccharomyces servazzii* n. sp. A new yeast from Finland soil. Ann Microbiol Enzimol 17:79–84.
2. Kato S, Kitamura E, Ohshima S. 1991. Factors influencing the growth of *Debaryomyces hansenii* and *Saccharomyces servazzii* isolated from salted "daikon" (Japanese radish). J Food Sci Technol 38:357–359. https://doi.org/10.3136/nskkk1962.38.357.
3. Spanoghe M, Godoy Jara M, Rivière J, Lanterbecq D, Gadenne M, Marique T. 2017. Development and application of a quantitative real-time PCR assay for rapid detection of the multifaceted yeast *Kazachstania servazzii* in food. Food Microbiol 62:133–140. https://doi.org/10.1016/j.fm.2016.10.015.
4. Mei J, Guo Q, Wu Y, Li Y. 2014. Microbial diversity of a Camembert-type cheese using freeze-dried Tibetan Kefir coculture as starter culture by culture-dependent and culture-independent methods. PLoS One 9:e111648. https://doi.org/10.1371/journal.pone.0111648.
5. Kong Y, Wu Q, Zhang Y, Xu Y. 2014. In situ analysis of metabolic characteristics reveals the key yeast in the spontaneous and solid-state fermentation process of Chinese light-style liquor. Appl Environ Microbiol 80:3667–3676. https://doi.org/10.1128/AEM.04219-13.
6. Kim JY, Kim J, Cha I-T, Jung MY, Song HS, Kim YB, Lee C, Kang S-Y, Bae J-W, Choi Y-E, Kim T-W, Roh SW. 2019. Community structures and genomic features of undesirable white colony-forming yeasts on fermented vegetables. J Microbiol 57:30–37. https://doi.org/10.1007/s12275-019-8487-y.
7. Casaregola S, Lépingle A, Bon E, Neuvéglise C, Nguyen H-V, Artiguenave F, Wincker P, Gaillardin C. 2000. Genomic exploration of the hemiascomycetous yeasts: 7. *Saccharomyces servazzii*. FEBS Lett 487:47–51. https://doi.org/10.1016/S0014-5793(00)02278-X.
8. White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p 315–322. *In* Innes MA, Gelfand DH, Sninsky JJ, White TJ (ed), PCR protocols: a guide to methods and applications. Academic Press, London, United, Kingdom.
9. Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 44:e113. https://doi.org/10.1093/nar/gkw294.
10. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
11. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
13. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110.
14. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored A, Cruaud C, Labadie K, Aury J-M, Istace B, Lebrigand K, Barbry P, Engelen S, Lemainque A, Wincker P, Liti G, Schacherer J. 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556:339–344. https://doi.org/10.1038/s41586-018-0030-5.