AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

EUKARYOTES

Draft Genome Sequence of a Highly Heterozygous Yeast Strain from the *Metschnikowia pulcherrima* Subclade, UCD127

Anjan Venkatesh,[a] Anthony L. Murray,[a] Adrian B. Boyle,[a] Lisa Quinn Farrington,[a] Timothy J. Maher,[a] Peadar Ó'Gaora,[a] Kenneth H. Wolfe,[b] Caoimhe E. O'Brien,[a] Geraldine Butler[a]

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

ABSTRACT *Metschnikowia* strain UCD127 was isolated from soil in Ireland and sequenced. It is a highly heterozygous diploid strain with 385,000 single nucleotide polymorphisms (SNPs). Its ribosomal DNA has the highest similarity to that of *M. chrysoperlae*, but its *ACT1* and *TEF1* loci and mitochondrial genome show affinity to those of *M. fructicola*, whose genome is significantly larger.

Metschnikowia spp. are yeasts that make characteristic needle-shaped spores. One of its subclades contains nine very closely related species, of which the best known is *M. pulcherrima* (1). The subclade also includes *M. fructicola*, *M. chrysoperlae*, and *M. zizyphicola*, among others (2–4). Species in this subclade are used commercially as biocontrol agents to prevent fruit spoilage, because they can kill molds (5, 6). However, because they are autogamous, defining species in this subclade has been based solely on molecular data and not reproductive isolation (7). A genome sequence is available for *M. fructicola* (5) but not for any other species in the subclade.

We isolated strain UCD127 from soil in a forest in Ireland (global positioning system [GPS] coordinates, N53.405750, W7.041967). It was cultured on yeast-peptone-dextrose (YPD) nutrient agar plates containing chloramphenicol (3% [wt/vol]) and carbenicillin (10% [wt/vol]) at 30°C. Sequencing the internal transcribed spacer (ITS) region suggested affinity to the *M. pulcherrima* subclade. Total genomic DNA was extracted with phenol-chloroform, and purified using a DNA Clean & Concentrator-25 kit (Zymo Research). Genomic DNA was sequenced by BGI Tech Solutions (Illumina HiSeq 4000) from libraries containing fragments of 170 to 800 bp. A total of 6.7 million paired-end reads (2 × 150 bp) were generated. Low-quality reads were trimmed using Skewer v0.2.1 (8).

The genome was assembled separately using SPAdes (9) and dipSPAdes v3.11.1 (10). QUAST v4.6 was used to assess assembly quality (11). Since the dipSPAdes assembly had substantially fewer contigs (33 contigs versus 7,594 contigs > 1 kb) and a higher N_{50} value (151 kb versus 2.5 kb), the genome was hypothesized to be highly heterozygous. The total assembly sizes were 16.1 Mb from dipSPAdes and 17.1 Mb from SPAdes, ignoring contigs of <1 kb. Analysis of variants was carried out using BWA (12) and SAMtools v1.4 (13) to map reads to the dipSPAdes assembly. A total of 385,486 SNPs and 45,673 indels were found. Histogram analysis of biallelic SNP frequencies confirmed that the genome is diploid.

Annotation of the dipSPAdes assembly using AUGUSTUS (14) predicted 5,807 protein-coding genes, which is in line with those of other ascomycete yeasts but much fewer than the 9,674 predicted in the 26-Mb *M. fructicola* genome (5). tRNAscan-SE identified 173 tRNA genes, including two genes for $tRNA^{Ser}$(CAG) with characteristic G_{33} and G_{73} positions, indicating that UCD127 translates CUG codons as serine, as expected for this genus (15).

Received 14 May 2018 **Accepted** 15 May 2018 **Published** 21 June 2018

Citation Venkatesh A, Murray AL, Boyle AB, Quinn Farrington L, Maher TJ, Ó'Gaora P, Wolfe KH, O'Brien CE, Butler G. 2018. Draft genome sequence of a highly heterozygous yeast strain from the *Metschnikowia pulcherrima* subclade, UCD127. Genome Announc 6:e00550-18. https://doi.org/10.1128/genomeA.00550-18.

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

UCD127 is in the *M. pulcherrima* subclade, but its exact species designation is uncertain. Phylogenetic analysis of the D1/D2 rDNA region clustered it with *M. chrysoperlae* (1 difference in 516 bp) and the unnamed strain NRRL Y-6148 (2 differences) rather than *M. pulcherrima* (5 differences) or *M. fructicola* (7 differences). However, the *ACT1* and *TEF1* sequences of UCD127 show a closer relationship to *M. fructicola* and *M. pulcherrima* than to *M. chrysoperlae*, and the mitochondrial genome has 99% sequence identity with *M. fructicola* (mitochondrial DNAs [mtDNAs] of the other species have not sequenced).

Accession number(s). This whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank under accession no. QBLL00000000. The version described in this paper is the first version, QBLL01000000.

ACKNOWLEDGMENTS

This work was supported by an undergraduate teaching award from University College Dublin.

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

We thank Lisa Lombardi, Elise Iracane, João Pacheco, and Siobhán Turner for help running the GENE30060 module.

REFERENCES

1. Lachance M-A. 2016. *Metschnikowia*: half tetrads, a regicide, and the fountain of youth. Yeast 33:563–574. https://doi.org/10.1002/yea.3208.
2. Suh S-O, Gibson CM, Blackwell M. 2004. *Metschnikowia chrysoperlae* sp. nov., *Candida picachoensis* sp. nov. and *Candida pimensis* sp. nov., isolated from the green lacewings *Chrysoperla comanche* and *Chrysoperla carnea* (Neuroptera: Chrysopidae). Int J Syst Evol Microbiol 54:1883–1890. https://doi.org/10.1099/ijs.0.63152-0.
3. Guzmán B, Lachance M-A, Herrera CM. 2013. Phylogenetic analysis of the angiosperm-floricolous insect-yeast association: have yeast and angiosperm lineages co-diversified? Mol Phyl Evol 68:161–175. https://doi.org/10.1016/j.ympev.2013.04.003.
4. Xue M-L, Zhang L-Q, Wang Q-M, Zhang J-S, Bai F-Y. 2006. *Metschnikowia sinensis* sp. nov., *Metschnikowia zizyphicola* sp. nov. and *Metschnikowia shanxiensis* sp. nov., novel yeast species from jujube fruit. Int J Syst Evol Microbiol 56:2245–2250. https://doi.org/10.1099/ijs.0.64391-0.
5. Hershkovitz V, Sela N, Taha-Salaime L, Liu J, Rafael G, Kessler C, Aly R, Levy M, Wisniewski M, Droby S. 2013. *De-novo* assembly and characterization of the transcriptome of *Metschnikowia fructicola* reveals differences in gene expression following interaction with *Penicillium digitatum* and grapefruit peel. BMC Genomics 14:168. https://doi.org/10.1186/1471-2164-14-168.
6. Hilber-Bodmer M, Schmid M, Ahrens CH, Freimoser FM. 2017. Competition assays and physiological experiments of soil and phyllosphere yeasts identify *Candida subhashii* as a novel antagonist of filamentous fungi. BMC Microbiol 17:4. https://doi.org/10.1186/s12866-016-0908-z.
7. Lachance MA. 2011. *Metschnikowia* Kamienski (1899), p 575–620. *In* Kurtzman CP, Fell JW, Boekhout T (ed), The yeasts, a taxonomic study. Elsevier, Amsterdam, the Netherlands.
8. Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15:182. https://doi.org/10.1186/1471-2105-15-182.
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
10. Safonova Y, Bankevich A, Pevzner PA. 2015. dipSPAdes: assembler for highly polymorphic diploid genomes. J Comput Biol 22:528–545. https://doi.org/10.1089/cmb.2014.0153.
11. 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.
12. 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.
13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroups. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
14. Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics 24:637–644. https://doi.org/10.1093/bioinformatics/btn013.
15. Santos MAS, Gomes AC, Santos MC, Carreto LC, Moura GR. 2011. The genetic code of the fungal CTG clade. C R Biol 334:607–611. https://doi.org/10.1016/j.crvi.2011.05.008.