Draft Genome Sequence of a Red Basidiomycete Yeast, *Symmetrospora coprosmae* Strain UCD350, Isolated from Soil in Ireland

Jameela Almasoud,[a] Laoise McArdle,[a] Ryan Henne,[a] Karen Mathews,[a] Ísla O'Connor,[a] Leah E. McLoughlin,[a] Peadar Ó'Gaora,[a] Kevin P. Byrne,[b] Caoimhe E. O'Brien,[a] Kenneth H. Wolfe,[b] Geraldine Butler[a]

aSchool of Biomedical and Biomolecular Sciences, Conway Institute, University College Dublin, Dublin, Ireland
bSchool of Medicine, Conway Institute, University College Dublin, Dublin, Ireland

ABSTRACT *Symmetrospora coprosmae* is a red yeast from the subphylum Pucciniomycotina in the phylum Basidiomycota. Here, we present the first genome sequence of *S. coprosmae* strain UCD350, from an isolate collected from soil in Ireland. The genome size is 20.2 Mb.

*S*ymmetrospora species are red ballistosporous yeasts in the subphylum Pucciniomycotina of phylum Basidiomycota (1, 2). Many *Symmetrospora* species were previously called *Sporobolomyces* spp. and were initially placed in the *Erythrobasidium* clade of class Cystobasidiomycetes (3). A more detailed analysis proposed placing Symmetrospora in the family Symmetrosporaceae (1), equivalent to the marina clade described by Wang et al. (2). *Symmetrospora* species form nearly symmetrical ballistoconidia (4). The species studied here has been previously isolated from leaves in New Zealand and Germany under the name *Sporobolomyces coprosmae* (5, 6). The yeast grows as pinkish-red colonies and produces the xanthophyll 2-hydroxytorularhodin (6).

We isolated strain UCD350 from soil in a field in Dublin, Ireland (global positioning system coordinates, 53.336938, −6.270591). The yeast was cultured on yeast extract-peptone-dextrose nutrient agar plates containing chloramphenicol (3% [wt/vol]) and ampicillin (10% [wt/vol]) at 30°C, and the species was identified as *S. coprosmae* by amplifying and Sanger sequencing the internal transcribed spacer of ribosomal DNA (rDNA) using primers ITS1 and ITS4 (7) (GenBank accession number MN540641).

Total genomic DNA was extracted and purified using a QiaAMP DNA minikit (Qiagen). Libraries were generated and sequenced by BGI Tech Solutions (Hong Kong). A total of 1 μg genomic DNA was fragmented using Covaris, purified with an AxyPrep Mag PCR clean up kit, and end repaired, and 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 instrument, generating 9.5 million spots.

All parameters used for sequence assembly and analysis are available at https://doi.org/10.6084/m9.figshare.9963617.v1. Reads with low quality were trimmed using Skewer v0.2.2 (8). The genome was assembled using SPAdes v3.11.1 (9). Analysis of Illumina sequence data indicates that cross-contamination may result from multiplexing of several samples (10). We used coverage-versus-length (CVL) plots to identify likely contaminants, as described by Douglass et al. (10). Scaffolds with lengths of <1 kb and a coverage of <40× were discarded; coverage of the major nuclear scaffolds was approximately 58×. A contaminating node containing rDNA of *Kazachstania servazzii* was also removed. Assembly quality was assessed using QUAST v4.6.1 (11). The total
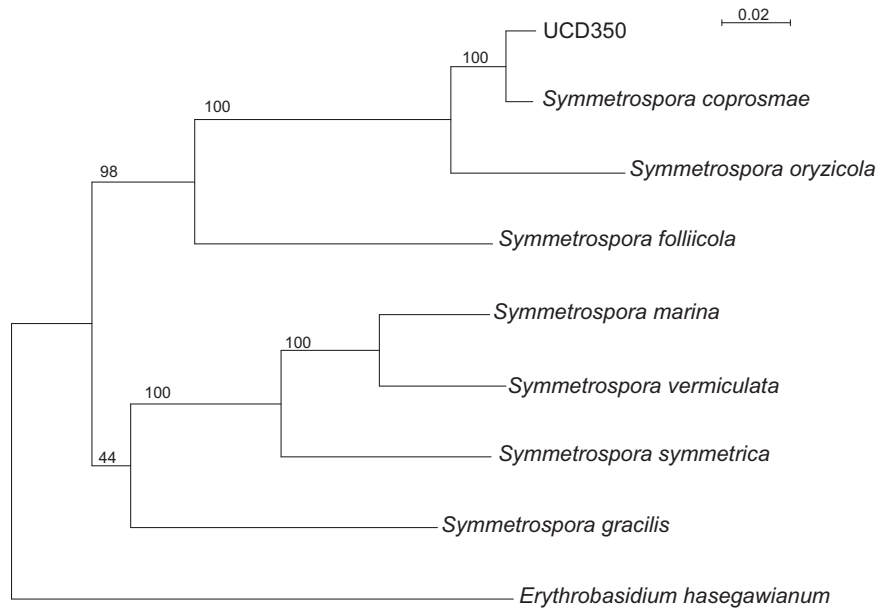
FIG 1 Phylogeny of yeast species in the family Symmetrosporaceae using the combined sequences of SSU rDNA, LSU rDNA D1/D2 regions, ITS regions (including 5.8S rDNA), and *RPB2*, *TEF1*, and *CYTB* genes. *Erythrobasidium hasegawianum* was used as an outgroup. Alignments were constructed using SeaView (15), and the tree was constructed using maximum likelihood (PhyML, with a general-time-reversible model and 100 replicates). Bootstrap supports are shown. Sequences for all other species were obtained from reference 2.

assembly size is 20.2 Mb, the N_{50} value is 484 kb, the L_{50} is 13 scaffolds, and the largest scaffold is 1,583,607 bp. The G+C content in basidiomycetes is normally above 50%, and the G+C content of *S. coprosmae* UCD350 is 59.7%.

The mitochondrial genome is a 26,122-bp circular contig with 5,058× coverage (GenBank accession number VUYT01000067). Phylogenetic analyses using seven loci confirmed that UCD350 lies in the family Symmetrosporaceae and is very closely related to the type strain of *S. coprosmae* (Fig. 1).

Variant analysis was carried out using BWA v0.7.12-r1039 (12), SAMtools v0.1.19 (13), and Genome Analysis Toolkit (GATK) v4.0.1.2 using the commands listed at https://figshare.com/articles/Genome_assembly_parameters_for_Symmetrospora_coprosmae_UCD350/9963617 (14). Variants were filtered by removing clusters (5 variants within 20 bp) that were assumed to result from poor read alignment and by applying the following GATK filters: QualByDepth (QD) of <2.0, Mapping Quality (MQ) of <40.0, FisherStrand (FS) of >60.0, StrandOddsRatio (SOR) of >3.0, MappingQualityRankSumTest (MQRankSum) of less than −12.5, and ReadPosRankSumTest (ReadPosRankSum) of less than −8.0. A very small number of potential variants (~650 heterozygous single nucleotide polymorphisms and ~90 insertions/deletions) were identified, suggesting that the isolate is haploid.

Data availability. This whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank under the accession number VUYT00000000 and the raw reads under SRA accession number SRX6817357. The mitochondrial genome is at GenBank accession number VUYT01000067 and the internal transcribed spacer (ITS) sequence at accession number MN540641. The version described in this paper is the first version. Data are also available under BioProject accession number PRJNA564489.

REFERENCES

1. Wang Q-M, Yurkov AM, Göker M, Lumbsch HT, Leavitt SD, Groenewald M, Theelen B, Liu X-Z, Boekhout T, Bai F-Y. 2015. Phylogenetic classification of yeasts and related taxa within *Pucciniomycotina*. Stud Mycol 81:149–189. https://doi.org/10.1016/j.simyco.2015.12.002.
2. Wang Q-M, Groenewald M, Takashima M, Theelen B, Han P-J, Liu X-Z, Boekhout T, Bai F-Y. 2015. Phylogeny of yeasts and related filamentous fungi within *Pucciniomycotina* determined from multigene sequence analyses. Stud Mycol 81:27–53. https://doi.org/10.1016/j.simyco.2015.08.002.
3. Scorzetti G, Fell JW, Fonseca A, Statzell-Tallman A. 2002. Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. FEMS Yeast Res 2:495–517. https://doi.org/10.1111/j.1567-1364.2002.tb00117.x.
4. Shivas RG, Rodrigues de Miranda L. 1983. Two new species of the genus *Sporobolomyces* and a new *Rhodotorula* species from leaf surfaces. Antonie Van Leeuwenhoek 49:159–165. https://doi.org/10.1007/bf00393674.
5. Hamamoto M, Nakase T. 1995. Ballistosporous yeasts found on the surface of plant materials collected in New Zealand. 1. Six new species in the genus *Sporobolomyces*. Antonie Van Leeuwenhoek 67:151–171. https://doi.org/10.1007/bf00871210.
6. Weber RWS, Madhour A, Anke H, Mucci A, Davoli P. 2005. 2-Hydroxytorularhodin, a new xanthophyll from the red yeast *Sporobolomyces coprosmae*. Helv Chim Acta 88:2960–2966. https://doi.org/10.1002/hlca.200590239.
7. White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p 315–322. *In* Innis MA, Gelfand DH, Sninsky JJ, White TJ (ed), PCR protocols: a guide to methods and applications. Academic Press, Inc., London, United Kingdom.
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. Douglass AP, O'Brien CE, Offei B, Coughlan AY, Ortiz-Merino RA, Butler G, Byrne KP, Wolfe KH. 2019. Coverage-versus-length plots, a simple quality control step for yeast genome sequence assemblies. G3 (Bethesda) 9:879–887. https://doi.org/10.1534/g3.118.200745.
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, 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.
14. 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.
15. Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224. https://doi.org/10.1093/molbev/msp259.