AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of the Yeast *Ogataea degrootiae* Strain UCD465, Isolated from Soil in Ireland

Eoin Ó Cinnéide,[a,b] Max Jones,[a] Elijah Bahate,[a] Elizabeth Boyd,[a] Rebeca Clavero,[a] Harry Doherty,[a] Izabela Drozdz,[a] Maja Dumana,[a] Cristina Gonzales,[a] Jane Kennedy,[a] Maria Khan,[a] Shane Maher,[a] Clare McGurk,[a] Katherine Moreau,[a] Ruby Neville,[a] Jade Norton,[a] Tadhg Ó Cróinín,[a] Elizabeth O'Gorman,[a] Juliana C. Olliff,[a,b] Rhys Orimaco,[a] Ellen Quinn,[a] Sarah Webb Kennedy,[a] Sean A. Bergin,[a] Kevin P. Byrne,[b] 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 *Ogataea degrootiae* is an ascomycete yeast that was first isolated in the Netherlands in 2017. It is a member of the Pichiaceae clade. Here, we present the genome sequence of *O. degrootiae* UCD465, which was isolated from soil in Ireland. This genome is 14.6 Mb and haploid.

Ogataea is an ascomycete yeast genus belonging to the family Pichiaceae in the subphylum Saccharomycotina (1). There are more than 30 species of *Ogataea* (2). The type strain of *Ogataea degrootiae*, CBS 15033, was isolated in 2017 from garden soil in the Netherlands (3). We identified another isolate, *O. degrootiae* UCD465, from soil collected in County Sligo, Ireland (coordinates: 54.227545, −9.031879), by two passages of soil material in 9 ml liquid yeast extract-peptone-dextrose (YPD) containing chloramphenicol (30 μg/ml) and ampicillin (100 μg/ml) and culture on YPD plates at room temperature, similar to the procedure described previously (4). The species was identified from single colonies by PCR amplification and Sanger sequencing of the internal transcribed spacer (ITS) and D1/D2 regions of its ribosomal DNA locus, both of which were identical to those of *O. degrootiae* CBS 15033 (3) (GenBank accession numbers NR_168172.1 and NG_068257.1, respectively).

Total genomic DNA was extracted from a YPD culture using phenol-chloroform-isoamyl alcohol. DNA was precipitated with isopropanol and ammonium acetate, washed twice in 70% ethanol, and dissolved in 100 μl Tris-EDTA (TE) buffer. Libraries were generated and sequenced by BGI Tech Solutions Co. (Hong Kong). A total of 1 μg genomic DNA was fragmented and size selected using a Covaris ultrasonicator, 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 the mixture at 37°C for 30 min. Illumina adapters were ligated by incubation at 16°C for 16 h. Approximately 150 bases were sequenced from each end of ~800-bp inserts with an Illumina HiSeq 4000 instrument, yielding 7.1 million read pairs.

Low-quality reads and adapter sequences were removed using Skewer v.0.2.2 (5) with default parameters. The genome was assembled using SPAdes v.3.11.1 (6) with the careful parameter. Based on coverage-versus-length plot analysis (7), scaffolds with less than 10× coverage or 0.5-kb length were removed, leaving 410 scaffolds. The assembly was analyzed using QUAST v.4.6.1 (8). The total genome size is 14.6 Mb, which is larger than the ~9-Mb genomes of *Ogataea* species *Ogataea polymorpha* and *Ogataea parapolymorpha* but is similar to the assemblies of closer relatives *Ogataea methanolica* and *Ogataea trehalophila* (1, 3). The N_{50} value is 95,261 bp, the L_{50} value is 49 contigs, and the G+C content is 36.2%. The largest contig is 343,079 bp, and the average coverage is 60×. Using BUSCO v.5.2.2 (9), genome completeness was estimated at 94.0% (compared to the Ascomycota lineage data set).

Citation Ó Cinnéide E, Jones M, Bahate E, Boyd E, Clavero R, Doherty H, Drozdz I, Dumana M, Gonzales C, Kennedy J, Khan M, Maher S, McGurk C, Moreau K, Neville R, Norton J, Ó Cróinín T, O'Gorman E, Olliff JC, Orimaco R, Quinn E, Webb Kennedy S, Bergin SA, Byrne KP, Wolfe KH, Butler G. 2021. Draft genome sequence of the yeast *Ogataea degrootiae* strain UCD465, isolated from soil in Ireland. Microbiol Resour Announc 10:e00736-21. https://doi.org/10.1128/MRA.00736-21.

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Geraldine Butler, gbutler@ucd.ie.

Received 19 July 2021
Accepted 3 September 2021
Published 30 September 2021

To examine ploidy and heterozygosity, heterozygous single-nucleotide polymorphisms (SNPs) were identified by aligning the trimmed reads to the assembled genome using BWA v.0.7.12-r1039 (10) (parameters bwa mem -M -Y -t 2 -R, followed by samtools view -S -b, samtools sort, and samtools flagstat with default parameters and then picard-tools MarkDuplicates and picard-tools BuildBamIndex with VALIDATION_STRINGENCY=LENIENT). Variants were called with HaplotypeCaller from GATK v.4.0.1.2 (11) with default parameters. Variants were filtered using GATK VariantFiltration with parameters -cluster-size 5 -cluster-window-size 20, followed by -genotype-filter-expression GQ $<$ 20 -genotype-filter-expression DP $<$ 10. Only 2,547 heterozygous sites were identified, suggesting that *O. degrootiae* UCD465 has a haploid genome.

We found mating-type loci with both *MAT***a** and *MAT*α genotypes, on different contigs, i.e., genes *MAT***a***1* and *MAT***a***2* on JAHDIT010000230.1 and genes *MAT*α1 and *MAT*α2 on JAHDIT010000187.1. It is thus likely that *O. degrootiae* is homothallic and switches mating types by inversion of a section of its genome (12).

Data availability. This whole-genome shotgun project was deposited in DDBJ/ENA/GenBank (accession number JAHDIT000000000). The version described in this paper is version JAHDIT010000000. The raw reads were deposited in the SRA (accession number SRR14551582). These data are also available under BioProject number PRJNA730000. The ITS sequence is available under accession number MZ191776 and the D1/D2 sequence under accession number MZ191775.

ACKNOWLEDGMENTS

This work was supported by undergraduate teaching resources from University College Dublin and by Science Foundation Ireland (grant 19/FFP/6668). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Thanks go to the children of Kilglass National School, Ballyglass, County Sligo, who collected the soil samples from which this strain was isolated.

REFERENCES

1. Shen X-X, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver JH, Wang M, Doering DT, Boudouris JT, Schneider RM, Langdon QK, Ohkuma M, Endoh R, Takashima M, Manabe R-I, Čadež N, Libkind D, Rosa CA, DeVirgilio J, Hulfachor AB, Groenewald M, Kurtzman CP, Hittinger CT, Rokas A. 2018. Tempo and mode of genome evolution in the budding yeast subphylum. Cell 175:1533–1545.e20. https://doi.org/10.1016/j.cell.2018.10.023.
2. Kurtzman CP, Fell JW, Boekhout T. 2011. The yeasts: a taxonomic study, 5th ed, vol 2. Elsevier Science Publishers, Amsterdam, Netherlands.
3. Groenewald M, Lombard L, de Vries M, Lopez AG, Smith M, Crous PW. 2018. Diversity of yeast species from Dutch garden soil and the description of six novel Ascomycetes. FEMS Yeast Res 18:foy076. https://doi.org/10.1093/femsyr/foy076.
4. 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.
5. Jiang H, Lei R, Ding SW, 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.
6. 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.
7. 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 de novo yeast genome sequence assemblies. G3 (Bethesda) 9:879–887. https://doi.org/10.1534/g3.118.200745.
8. 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.
9. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
10. 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.
11. 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.
12. Krassowski T, Kominek J, Shen XX, Opulente DA, Zhou X, Rokas A, Hittinger CT, Wolfe KH. 2019. Multiple reinventions of mating-type switching during budding yeast evolution. Curr Biol 29:2555–2562.e8. https://doi.org/10.1016/j.cub.2019.06.056.