Data in Brief 56 (2024) 110808

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Long-read sequencing and de novo genome assembly data of *Candida parapsilosis* HMC1 and *Rhodotorula mucilaginosa* LBMH1012, two novel isolates with antifungal resistance signatures

Martín R. Ide-Pérez [a,b], Mario León Iza-Arteaga [a], María del Rayo Sánchez-Carbente [a], Edgar Balcázar-López [c,*], Ayixon Sánchez-Reyes [d,*]

[a] *Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, Mexico*
[b] *Facultad de Ciencias Biológicas, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, Mexico*
[c] *Departamento de Farmacobiología, Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara, Mexico*
[d] *Investigador por México-Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico*

ARTICLE INFO

Article history:
Received 23 April 2024
Revised 14 July 2024
Accepted 2 August 2024
Available online 8 August 2024

Dataset link: Nanopore sequencing and assembly of two clinically important yeast (Original data)
Dataset link: Nanopore sequencing and assembly of two clinically important yeast (Original data)
Dataset link: Nanopore sequencing and assembly of two clinically important yeast (Original data)
Dataset link: Nanopore sequencing and assembly of two clinically important yeast (Original data)

ABSTRACT

Candida parapsilosis and *Rhodotorula mucilaginosa* are opportunistic pathogens affecting mostly immunocompromised hosts. Both species have emerged as causes of invasive candidiasis and sepsis respectively. Here we present high-quality long-read genome assemblies for a strain of *C. parapsilosis* isolated from human breast milk, with multiple predicted signatures consistent with Candida Drug Resistance CDR1/CDR2 and Multi Drug Resistance MDR1-type genes, also for an environmental strain of *R. mucilaginosa* with multiresistance to azole antifungals. The genome sequencing was performed using the R9.4.1 flowcell with the MinION Mk1B sequencer (Oxford Nanopore Technologies, Oxford, UK). The draft genome of *C. parapsilosis* HMC1 was assembled from 85,745 long-reads and has 13,114,208 bp in length and comprises 10 contigs making it a highly contiguous assembly. The

* Corresponding authors.
E-mail addresses: edgar.balcazar@academicos.udg.mx (E. Balcázar-López), ayixon.sanchez@ibt.unam.mx (A. Sánchez-Reyes).
Social media: @IDE0717 (M.R. Ide-Pérez), @mdrsc (M.d.R. Sánchez-Carbente), @ayixon1 (A. Sánchez-Reyes)

https://doi.org/10.1016/j.dib.2024.110808

Keywords:
Candidiasis
Azole resistance
Antifungal therapy
Fungemia

R. mucilaginosa LBMH1012 assembly has 23,636,156 bp in length and comprises 54 contigs. The genome completeness was estimated as 94.02 % and 91.40 % respectively using BUSCO. These data may be useful to explore the genetic diversity landscape in both species, infer potential causal genes for antifungal resistance and virulence, and represent an addition to the useful sequence space on emerging fungal pathogens.

Specifications Table

Subject	Microbiology: Fungal Biology
Specific subject area	Bioinformatics, Genomics
Type of data	Raw, Filtered, and Processed.
Data collection	High-molecular-weight (HMW) DNA was isolated using the Quick-DNA HMW MagBead kit (Zymo Research cat number D6060). Long-read nanopore genomic libraries were prepared using the ligation sequencing library preparation kit (SQK-LSK109; ONT, Oxford, UK), followed by sequencing onto a MinION flow cell (R9.4.1; ONT) as described by the manufacturer. Guppy v4.4.1 (ONT) was used to basecall and demultiplexing the raw data, and de novo genome assemblies were carried out using Canu v. 2.2, Flye v2.9, and quickmerge metassembler v0.3. Genome annotation was conducted with the MOSGA tool v2.1.6
Data source location	Data source location Institution: Institute of Biotechnology, National Autonomous University of Mexico; University Center of Exact Sciences and Engineering, Guadalajara University. City/Province/Country: Cuernavaca/Morelos/Mexico and Guadalajara/Guadalajara/Mexico
Data accessibility	1. Raw sequences Repository name: NCBI/SRA Data identification number: SRA Accessions: SRR28047814; SRR28051626 Direct URL to data: https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=SRR28047814&display=metadata https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=SRR28051626&display=metadata 2. Genome sequences Repository name: GenBank Data identification number: JBAJNE000000000.1; JALGWT000000000.1 Direct URL to data: https://www.ncbi.nlm.nih.gov/nuccore/JBAJNE000000000.1 https://www.ncbi.nlm.nih.gov/nuccore/JALGWT000000000.1/

1. Value of the Data

- These data are useful to explore conspecific antifungal resistance profiles, genetic bases related to potential virulence and invasiveness phenotypes, as well as aspects of the biology and evolution of *Candida parapsilosis* and *Rhodotorula mucilaginosa* clades.
- Other researchers could make use of these new genomic assemblies to explore the sequence space and causal elements related to the pathogenicity and emergence of *C. parapsilosis* and *R. mucilaginosa.*
- These genome sequence data are added to the list of assemblies already available, bringing Genome-Wide Association Studies (GWAS) closer to identifying novel variants associated with antimicrobial resistance in yeast.

2. Background

Candida parapsilosis and *Rhodotorula mucilaginosa* are two pathogenic yeasts associated with primary and secondary invasive infections. They mainly affect immunocompromised individuals (neonates, patients under parenteral nutrition, or other types of catheterization). The recent increase in the prevalence of invasive candidiasis and other emerging fungemia have been associated with an increasing antifungal resistance phenotypes [1,2]. We have sequenced the whole genomes of a commensal *C. parapsilosis* strain HMC1 isolated from a clinical environment, with several sequence signatures for antifungal resistance; and an environmental *R. mucilaginosa* strain LBMH1012, resistant to fluconazole, miconazole, and metronidazole. Our goal is to explore conspecific antifungal resistance profiles, genetic bases related to potential virulence and invasiveness phenotypes, as well as aspects of the biology and evolution of both clades. We consider of particular importance the contribution of these genetic contexts to reaching a significant genome number in order to develop GWAS studies in yeast, which allows us to explore elements of causality related to different clinical phenotypes.

3. Data Description

A highly contiguous assembly of 13,114,208 bp was reconstructed for *C. parapsilosis* HMC1, with 9 contigs, 40X coverage, and a mitochondrial genome size of 31589 bp. This assembly level is close to the reported karyotype for this taxon (8 chromosomes). The assembled genome had an N50 of 2085,417 bp and a GC content of 38.5%. MOSGA predicted 5970 protein-coding genes, 108 tRNAs, and 7 rRNAs (6 of eukaryotic origin and 1 mitochondrial); also 16 significant hits similar to Candida Drug Resistance CDR1/CDR2 gene, and 23 Multi-Drug Resistance MDR1 were predicted (Table 1). The genome of strain HMC1 shares a 99.00 % ANI value with *C. parapsilosis* (GCA_000182765.2), while it is more genomically distant from *C. metapsilosis* and *C. orthosilopsis*, indicating that it is a new member of the *parapsilosis* clade. The Whole Genome Shotgun project was deposited at DDBJ/ENA/GenBank under the accession JBAJNE000000000. The raw read sequences were submitted to the Sequence Read Archive (SRA) accession SRR28047814 and are available at https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=SRR28047814&display=metadata. *R. mucilaginosa* LBMH1012 assembly resulted in 54 contigs representing 23,636,156 bp, with 39X coverage, and a mitochondrial genome with 46,983 bp. The assembled genome had an N50 of 600,067 bp and a GC content of 60.5%. The annotation revealed 7495 predicted proteins, 178 tRNAs, and 7 rRNAs (6 of eukaryotic origin and 1 mitochondrial). The genome of strain LBMH1012 shares a 95.82% ANI value with *R. mucilaginosa* C2.5ti (GCA_000931965.1). This strain has a phenotype resistant to fluconazole, miconazole, and metronidazole while it is sensitive to clotrimazole (Fig. 1). The Whole Genome Shotgun project was deposited at DDBJ/ENA/GenBank under the accession JALGWT000000000. The raw read sequences were submitted to the Sequence Read Archive (SRA) accession SRR28051626 and are available at https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=SRR28051626&display=metadata. The genome completeness was estimated as 94.02 % and 91.40 % respectively using BUSCOv5.4.1 [9]. Table 2 shows the SRA-associated information of the high throughput sequencing data for *C. parapsilosis* HMC1 and *C. mucilaginosa* LBMH1012 projects. For both projects, the average read length was greater than 9000 bp and contained >85,000 reads.

4. Experimental Design, Materials and Methods

4.1. Strains isolation

The strain HMC1 was originally isolated in October of 2022 at the Center for Exact Sciences and Engineering at the University of Guadalajara (CUCEI-UdeG), México, from a 24-year-old

Table 1
Whole genome shotgun sequencing project descriptions and genome assembly global statistics.

WGS descriptors			
Organism name	*Candida parapsilosis* (budding yeasts)	*Rhodotorula mucilaginosa* (basidiomycete fungi)	
Infraspecific name	Strain: HMC1	Strain: LBMH1012	
BioSample	SAMN40001617	SAMN27062970	
BioProject	PRJNA1078440	PRJNA821224	
Assembly level	Contig	Contig	
Genome representation	full	full	
Assembly accession	GCA_037952885.1 (latest)	GCA_024500305.1 (latest)	
WGS Project	JBAJNE01	JALGWT01	
Assembly method	Canu v. 2.2; Flye v. 2.9; Quickmerge v. 0.3	Canu v. 2.2	
Genome coverage	40.0x	39.0x	
Sequencing technology	MinION Oxford Nanopore	MinION Oxford Nanopore	
Assembly metrics			
Total sequence length	13,114,208	23,636,156	
Number of contigs	9 nuclear 1 mitochondrial	53 nuclear 1 mitochondrial	
Contig N50	2085,417	600,067	
Contig L50	3	14	
G + C content (%)	38.5	60.5	
Mitochondrion MT	JBAJNE010000003.1	CM044883.1	
Annotation predictions			
Protein-coding genes	5970	7495	
tRNAs	108	178	
rRNAs	7	7	
BUSCO completeness (%)	94.02	91.40	
Antifungal phenotype*	CDR1/CDR2 (16 genes predicted) MDR1 (23 genes predicted)	Fluconazole Miconazole Clotrimazole	(R) (R) (S)

* Antifungal phenotype was assessed in *R. mucilaginosa* LBMH1012 by the disk diffusion plate assay method. **R**: Resistant; **S**: Susceptible. In *C. parapsilosis* HMC1 the phenotype was predicted based on the presence of CDR1, CDR2, and MDR1 genes.

Table 2
Sequence Read Archive data associated with the long-reads for *C. parapsilosis* and *R. mucilaginosa* sequencing projects.

	Candida parapsilosis HMC1	*Rhodotorula mucilaginosa* LBMH1012
Sequencing technology	Nanopore sequencing	Nanopore sequencing
Instrument	MinION Mk1B	MinION Mk1B
Experiment	SRX23699013	SRX23701384
Run	SRR28047814	SRR28051626
Spots	85,745	102,414
Bases	1.0 G	935.9M
Size	864.8MB	802.9MB
GC Content	38.3 %	59 %
Average length (bp)	11,881	9138
Standard deviation	15,495.8	7625.3
Read format	Fastq	Fastq
Access type	Public	Public

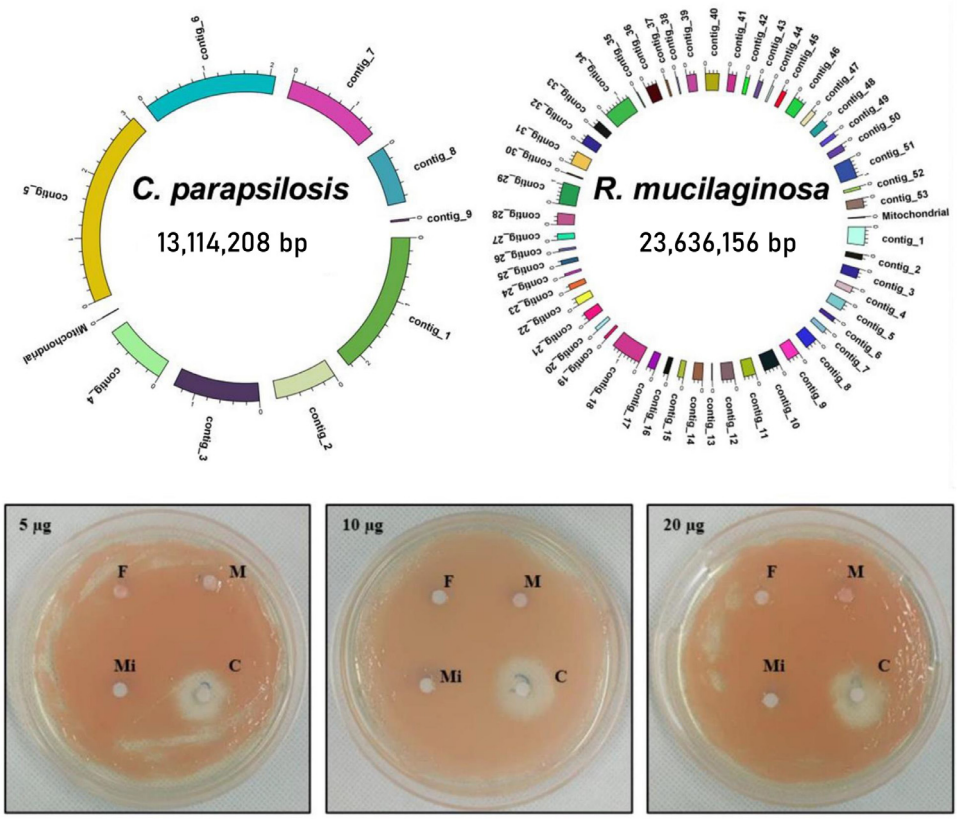

Fig. 1. Circular representation of the *Candida parapsilosis* and *Rhodotorula mucilaginosa* contig-level assemblies obtained with TBtools (upper quadrant). Antifungal resistance phenotypes for *Rhodotorula mucilaginosa* LBMH1012 at different concentrations (5, 10, 15 µg). **F**: fluconazole, **M**: metronidazole, **Mi**: miconazole, **C**: clotrimazole (lower quadrant).

female breast milk sample (healthy female in active breastfeeding). The strain LBMH1012 was isolated in 2020 from soil contaminated with crude oil in the municipality of Santa Isabel, Cunduacán, Tabasco, México as previously described [3].

4.2. Genome sequencing, annotation, and antifungal resistance test and prediction

High-molecular-weight (HMW) DNA was extracted using the Quick-DNA HMW MagBead kit (Zymo Research cat number D6060). Long-read nanopore sequencing was performed on HMW DNA using the ligation sequencing library preparation kit (SQK-LSK109; Oxford Nanopore Technologies, Oxford, UK), followed by sequencing onto a MinION Mk1B flow cell (R9.4.1; Oxford Nanopore Technologies, Oxford, UK) as described by the manufacturer. Guppy v4.4.1 (ONT) was used to basecall and demultiplexing the raw data, and de novo genome assemblies were carried out using Canu v. 2.2, Flye v2.9 and quickmerge v0.3 metassembler [4–6]. Genome annotation was conducted utilizing the MOSGA tool v2.1.6 (https://mosga.mathematik.uni-marburg.de/) [7]. We use the ABC transporter (CDR-1/CDR-2) and the major facilitator (MDR1) sequences reported by [8] as queries for searching homologs using hidden Markov models with the nhmmer tool [9], as has been shown in several studies [10]. The antifungal resistance phenotype was

assessed in *R. mucilaginosa* LBMH1012 by the disk diffusion plate assay method, at the concentrations 5, 10, and 15 µg for fluconazole, metronidazole, miconazole, and clotrimazole according to [11]. We employ the BUSCO (Benchmarking Universal Single-Copy Orthologs) v5.4.1 to assess the completeness and infer the integrity of the genome assemblies [12]. Circular contig maps were recreated using the Toolkit for Biologists integrating various biological data-handling tools (TBtools) [13].

4.3. Ethical considerations

For the isolation of the HMC1 strain, a sample of human breast milk was obtained from a healthy 24-year-old mother who was actively breastfeeding. Written and signed "informed consent" was requested from the donor to comply with the guidelines of the "Ethical Principles for Medical Research Involving Human Subjects" referred to in the Helsinki Declaration. The research involving human participants was reviewed and approved by the Institutional Review Board of Antiguo Hospital Civil de Guadalajara "Fray Antonio Alcalde."

Limitations

Not applicable.

Ethics Statement

All the authors have read and followed the ethical requirements for publication in Data in Brief and confirm that the current work does not involve, animal experiments, or any data collected from social media platforms. The research involving human participants was reviewed and approved by the Institutional Review Board of Antiguo Hospital Civil de Guadalajara "Fray Antonio Alcalde." Informed consent was obtained to collect the human breast milk sample, which was the source for isolating one of the yeasts used in this study.

CRediT Author Statement

A.S.R. and E.B..L. contributed to the study conception and design. Material preparation and data collection were performed by M.R.I.P., M.R.S.C., and E.B.L. The Analyses were performed by A.S.R., M.R.I.P., and M.L.I.A. The first draft of the manuscript was written by A.S.R. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability

Nanopore sequencing and assembly of two clinically important yeast (Original data) (NCBI/SRA/GenBank).
Nanopore sequencing and assembly of two clinically important yeast (Original data) (NCBI/SRA/GenBank).
Nanopore sequencing and assembly of two clinically important yeast (Original data) (NCBI/SRA/GenBank).
Nanopore sequencing and assembly of two clinically important yeast (Original data) (NCBI/SRA/GenBank).

Acknowledgements

This work was supported by the Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT) through the Programa Presupuestario F003, Grant No. CF 2019 265222, awarded to ASR. Additional support was provided by the PROSNI Funds UdeG 2023, granted to EBL.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Duggal, H. Jain, A. Tyagi, A. Sharma, T.D. Chugh, Rhodotorula fungemia: two cases and a brief review, Med. Mycol. 49 (2011), doi:10.3109/13693786.2011.583694.
[2] F. Daneshnia, J.N. de Almeida Júnior, M. Ilkit, L. Lombardi, A.M. Perry, M. Gao, C.J. Nobile, M. Egger, D.S. Perlin, B. Zhai, T.M. Hohl, T. Gabaldón, A.L. Colombo, M. Hoenigl, A. Arastehfar, Worldwide emergence of fluconazole-resistant Candida parapsilosis: current framework and future research roadmap, Lancet Microbe 4 (2023), doi:10.1016/S2666-5247(23)00067-8.
[3] M.R. Ide-Pérez, M.G. Fernández-López, A. Sánchez-Reyes, A. Leija, R.A. Batista-García, J.L. Folch-Mallol, M.D.R. Sánchez-Carbente, Aromatic hydrocarbon removal by novel extremotolerant exophiala and rhodotorula spp. from an oil polluted site in Mexico, J. Fungi 6 (2020) 1–17, doi:10.3390/JOF6030135.
[4] M. Chakraborty, J.G. Baldwin-Brown, A.D. Long, J.J. Emerson, Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage, Nucl. Acids Res. 44 (2016) e147, doi:10.1093/NAR/GKW654.
[5] S. Koren, B.P. Walenz, K. Berlin, J.R. Miller, N.H. Bergman, A.M. Phillippy, Canu: scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation, Genome Res. 27 (2017), doi:10.1101/gr.215087.116.
[6] M. Kolmogorov, J. Yuan, Y. Lin, P.A. Pevzner, Assembly of long, error-prone reads using repeat graphs, Nat. Biotechnol. 37 (2019), doi:10.1038/s41587-019-0072-8.
[7] R. Martin, T. Hackl, G. Hattab, M.G. Fischer, D. Heider, MOSGA: modular open-source genome annotator, Bioinformatics 36 (2020), doi:10.1093/bioinformatics/btaa1003.
[8] K. Khosravi Rad, M. Falahati, M. Roudbary, S. Farahyar, S. Nami, Overexpression of MDR-1 and CDR-2 genes in fluconazole resistance of Candida albicans isolated from patients with vulvovaginal candidiasis, Curr. Med. Mycol. 2 (2016) 24–29, doi:10.18869/acadpub.cmm.2.4.24.
[9] T.J. Wheeler, S.R. Eddy, Nhmmer: DNA homology search with profile HMMs, Bioinformatics 29 (2013), doi:10.1093/bioinformatics/btt403.
[10] A. Kumar, A. Jha, in: Multidrug Resistance and Transporters, Anticandidal Agents, 2017, pp. 49–54, doi:10.1016/B978-0-12-811311-0.00005-3.
[11] S. Khadka, J.B. Sherchand, B.M. Pokhrel, K. Parajuli, S.K. Mishra, S. Sharma, N. Shah, H.P. Kattel, S. Dhital, S. Khatiwada, N. Parajuli, M. Pradhan, B.P. Rijal, Isolation, speciation and antifungal susceptibility testing of Candida isolates from various clinical specimens at a tertiary care hospital, Nepal, BMC Res. Notes 10 (2017) 1–5, doi:10.1186/S13104-017-2547-3/TABLES/2.
[12] M. Manni, M.R. Berkeley, M. Seppey, F.A. Simão, E.M. Zdobnov, BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes, Mol. Biol. Evol. 38 (2021), doi:10.1093/molbev/msab199.
[13] C. Chen, H. Chen, Y. Zhang, H.R. Thomas, M.H. Frank, Y. He, R. Xia, TBtools: an integrative toolkit developed for interactive analyses of big biological data, Mol. Plant 13 (2020) 1194–1202, doi:10.1016/j.molp.2020.06.009.