

Cracking the code of a coastal oil yeast: the genome of *Rhodotorula pacifica* INDKK

Kiran Kumar Kukkala,¹ Preeti Yadav,¹ Ashish Kumar,² Pannaga Pavan Jutur,² Debarati Paul,³ Naseem A. Gaur¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Rhodotorula pacifica* INDKK, an oleaginous red yeast isolated from mangrove forest soil, efficiently metabolizes biomass hydrolysates and accumulates lipids of more than 60% of its dry weight. Hybrid *de novo* genome assembly (Illumina and Oxford Nanopore) produced 24 scaffolds comprising a 21.05 Mb nuclear and ~26 kb mitochondrial genome.

KEYWORDS genome assembly, Illumina, nanopore, oleaginous yeast, contigs, lipids, *Rhodotorula*, annotation, telomere, *pacifica*

Rhodotorula pacifica INDKK is an oleaginous red yeast isolated from Coringa mangrove forest, Kakinada, Andhra Pradesh (16.83139°N–82.33667°E), India. This strain shows strong potential for low-cost lipid and carotenoid production due to its ability to utilize both pentose and hexose sugars from lignocellulosic hydrolysates and higher tolerance to pretreatment-derived inhibitors (1, 2). Genome sequencing of *R. pacifica* INDKK was initiated for elucidating the metabolic and regulatory pathways related to lipid, carotenoid, and other valuable products with the aim of achieving increased production rates by employing genome engineering.

The strain was isolated by the serial dilution method and cultured in YPD medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L dextrose) at 30°C and 200 rpm. Genomic DNA was extracted using enzymatic lysis and the CTAB (Cetyl Trimethyl Ammonium Bromide) method (3). The quality of the extracted genomic DNA was assessed by NanoDrop and agarose gel electrophoresis. The whole-genome sequencing was performed by combining short-read Illumina and long-read Oxford Nanopore platforms. Illumina libraries were prepared by using 100 ng of genomic DNA (QIASeq FX DNA Library Kit) and sequenced on a HiSeq X Ten (2 × 150 bp) platform. Nanopore libraries (600 ng DNA) were prepared using EXP-NBD114 and sequenced on a GridION X5 platform with R9.4 flow cell.

Oxford Nanopore reads were filtered with NanoFilt v2.6.0, and Illumina reads were trimmed using Fastp and assessed with FastQC v0.11.9. A hybrid *de novo* assembly was generated using MaSuRCA v4.0.3, followed by removal of haplotigs and duplications via Purge_dups v1.2.6. A mitochondrial contig identified via BLASTn was excluded to prepare an exclusively nuclear genome assembly. RepeatModeler v2.0.6 and RepeatMasker v4.1.9 were used for repeat modeling and soft-masking. Tandem Repeats Finder identified canonical telomeric repeats (TTAGGG) at both ends of scaffolds 5, 9, and 10, indicating near chromosome-level assembly. Scaffolds 2, 20, and 22 had canonical repetitions on only one end, while Scaffold-1 contained several internal tandem telomeric-like repeats, indicating subtelomeric or interstitial telomeric repeats. BUSCO v5.8.0 analysis indicated 94.4% complete orthologs, comprising 88.9% single-copy, 5.5% duplicated, 1% fragmented, and 4.6% missing genes. QUAST v5.3.0 was used to assess additional assembly metrics (Table 1). The draft genome sequence of *R. pacifica*

Editor André O. Hudson, Rochester Institute of Technology, Rochester, New York, USA

Address correspondence to Naseem A. Gaur, nasgaur@hotmail.com; naseem@icgeb.res.in, or Debarati Paul, dpaul@amity.edu.

The authors declare no conflict of interest.

Received 4 April 2025

Accepted 12 August 2025

Published 27 August 2025

Copyright © 2025 Kukkala et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Table showing the genome assembly statistics of *R. pacifica* INDKK

Parameter	Number
Total no. of scaffolds	24
N's per 100 kbp	16
No. of scaffolds ($\geq 1,000$ bp)	24
Total length	21,056,800
Largest contig	2,764,732
GC %	59.40
N50 value	1,305,572
N90 value	669,876
Assembly level	Scaffold
Genome coverage	360x
L50	6
L90	15
No. of N's per 100 kbp	16.19

comprises 24 scaffolds with a total genome size of 21.05 Mb and 59.4% GC content, which is consistent with earlier studies involving oleaginous red yeasts (4, 5).

Paired-end RNA-seq reads were aligned to the soft-masked genome using HISAT2 v2.2.1. Splice junctions were identified using GeneMark-ET v4.65 in fungal mode and used by AUGUSTUS v3.4.0 under BRAKER2 for ab initio gene prediction, which predicted 7,809 coding sequences. Functional annotation using EggNOG-mapper v5.0 assigned GO terms to 5,735 predicted proteins.

ACKNOWLEDGMENTS

The authors acknowledge their host institutions and funding from ICGEB, New Delhi, and Department of Biotechnology (DBT), Government of India, for financial support (grant no. BT/PB/Centre/03/ICGEB/2011-PhaseII).

AUTHOR AFFILIATIONS

¹Yeast Biofuel Group, International Centre for Genetic Engineering and Biotechnology, New Delhi, India

²Omics of Algae Group, International Centre for Genetic Engineering and Biotechnology, New Delhi, India

³Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida, Uttar Pradesh, India

AUTHOR ORCID*s*

Debarati Paul  <http://orcid.org/0000-0002-3893-658X>

Naseem A. Gaur  <http://orcid.org/0000-0002-1224-8789>

AUTHOR CONTRIBUTIONS

Kiran Kumar Kukkala, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing, Conceptualization, Data curation | Preeti Yadav, Writing – review and editing | Ashish Kumar, Formal analysis, Investigation, Methodology, Software, Visualization | Pannaga Pavan Jutur, Formal analysis, Investigation, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing | Debarati Paul, Methodology, Validation, Writing – original draft, Writing – review and editing.

DATA AVAILABILITY

The *R. pacifica* INDKK whole-genome sequencing project has been deposited at

DDBJ/EMBL/GenBank under the accession number [CBDEHS030000000.3](#). The revised third version was analyzed and described in this paper.

REFERENCES

1. Kumar KK, Deeba F, Negi YS, Gaur NA. 2020. Harnessing pongamia shell hydrolysate for triacylglycerol agglomeration by novel oleaginous yeast *Rhodotorula pacifica* INDKK. *Biotechnol Biofuels* 13:175. <https://doi.org/10.1186/s13068-020-01814-9>
2. Deeba F, Kiran Kumar K, Ali Wani S, Kumar Singh A, Sharma J, Gaur NA. 2022. Enhanced biodiesel and β -carotene production in *Rhodotorula pacifica* INDKK using sugarcane bagasse and molasses by an integrated biorefinery framework. *Bioresour Technol* 351:127067. <https://doi.org/10.1016/j.biortech.2022.127067>
3. Zhang YJ, Zhang S, Liu XZ, Wen HA, Wang M. 2010. A simple method of genomic DNA extraction suitable for analysis of bulk fungal strains. *Lett Appl Microbiol* 51:114–118. <https://doi.org/10.1111/j.1472-765X.2010.02867.x>
4. Martín-Hernández GC, Müller B, Chmielarz M, Brandt C, Hölzer M, Viehweger A, Passoth V. 2021. Chromosome-level genome assembly and transcriptome-based annotation of the oleaginous yeast *Rhodotorula toruloides* CBS 14. *Genomics* 113:4022–4027. <https://doi.org/10.1016/j.ygeno.2021.10.006>
5. Hu J, Ji L. 2016. Draft genome sequences of *Rhodospiridium toruloides* strains ATCC 10788 and ATCC 10657 with compatible mating types. *Genome Announc* 4:e00098-16. <https://doi.org/10.1128/genomeA.00098-16>