AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Draft Genome Sequences of *Rhodosporidium toruloides* Strains ATCC 10788 and ATCC 10657 with Compatible Mating Types

Jie Hu, Lianghui Ji

Temasek Life Sciences Laboratory, National University of Singapore, Singapore

Rhodosporidium toruloides **ATCC 10788 (haploid, A1 mating type) and ATCC 10657 (haploid, A2 mating type) were derived from the same diploid parent strain *Rhodotorula glutinis* ATCC 90781 and are important strains for metabolic engineering. Draft genome sequences of both strains are reported here. The current assembly of strain ATCC 10788 comprises 61 scaffolds with a total size of 20.75 Mbp and a GC content of 62.01%, while that of strain ATCC 10657 comprises 137 scaffolds with a total size of 21.49 Mbp and a GC content of 61.81%. Genome annotation predicts 7,730 and 7,800 protein encoding genes for strain ATCC 10788 and strain ATCC 10657, respectively.**

Received 25 January 2016 **Accepted** 29 January 2016 **Published** 10 March 2016

Citation Hu J, Ji L. 2016. Draft genome sequences of *Rhodosporidium toruloides* strains ATCC 10788 and ATCC 10657 with compatible mating types. Genome Announc 4(2): e00098-16. doi:10.1128/genomeA.00098-16.

Address correspondence to Lianghui Ji, jilh@tll.org.sg.

Rhodosporidium toruloides has attracted increasing interest since the 1980s because of its capability for high-cell-density fermentation and high-level lipid production (1–3). It is a promising production host for renewable fuels and chemicals. Four genome sequences of *R. toruloides* strains have been published since 2012, i.e., MTCC 457 (4), NP 11 (5), CECT 1137 (6), and ATCC 204091 (previously *Rhodotorula glutinis*) (7). Haploid strains ATCC 10788 and ATCC 10657, obtained from the American Type Culture Collection, were derived from the same parent strain *Rhodotorula glutinis* ATCC 90781 with A1 and A2 mating types, respectively, and have been targets for metabolic engineering for the production of high-value bioproducts (8–10).

Whole-genome sequencing was carried out by Macrogen, Inc. (Republic of Korea) with the Illumina HiSeq 2000 platform using paired-end (insert length of 200 bp) and mate-pair (10-kb insert) libraries. Approximately 5 Gb of raw data (101-bp reads with about 100× sequencing depth) were generated from each strain. Several *de novo* assemblies, like SOAPdenovo (11), ALLPATHS-LG (12), CLC genomics workbench (Qiagen), Velvet (13), ABySS (14), IDBA-UD (15), and MaSuRCA (16), were used to perform the assembly. The best assemblies (by ALLPATHS-LG) were evaluated and chosen by the quality assessment tool for genome assemblies (QUAST) (17). The genes were predicted by GeneMark-ES (18) and MAKER2 (19). Gene functions and evolutionary relationship were identified by BLAST (20) against the NCBI nonredundant databases (nt and nr).

The draft genome sequence of ATCC 10788 comprises 61 scaffolds with a total size of 20.75 Mbp and a GC content of 62.01%, while that of strain ATCC 10657 comprises 137 scaffolds, with a total size of 21.49 Mbp and a GC content of 61.81%. A total of 7,730 genes for strain ATCC 10788 and 7,800 genes for strain ATCC 10657 were predicted by GeneMark-ES without a reference annotated genome, whereas 7,181 and 7,085 for ATCC 10788 and ATCC 10657, respectively, were predicted by MAKER2 based on the *Rhodotorula glutinis* ATCC 204091 protein database. A comparison of the genome assemblies to published ones by QUAST (genome fraction %) reveals that 98.78 to 99.63% of contig bases of ATCC 10788 could be aligned to the genome of MTCC 457 (4), NP 11 (5), or CECT 1137 (6), while 0.14% could be aligned to the genome of ATCC 204091 (7). On the other hand, 99.51% of the contig bases of ATCC 10657 could be aligned to the genome of ATCC 204091, but 0.11 to 0.16% could be aligned to the genome of MTCC 457 (4), NP 11 (5), or CECT 1137 (6). These data suggest that *R. toruloides* of different mating types have diversified extensively in nucleotide sequences and gene organizations.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession numbers LNQQ00000000 and LNKU00000000.

ACKNOWLEDGMENTS

This work was supported by the Singapore National Research Foundation under CRP award no. NRF-CRP8-2011-02 and by the Temasek Trust.

FUNDING INFORMATION

This work, including the efforts of Jie Hu, was funded by Singapore National Research Foundation CRP Award (NRF-CRP8-2011-02).

REFERENCES

1. **Zhao X, Hu C, Wu S, Shen H, Zhao ZK.** 2011. Lipid production by *Rhodosporidium toruloides* Y4 using different substrate feeding strategies. J Ind Microbiol Biotechnol **38:**627–632. http://dx.doi.org/10.1007/s10295-010-0808-4.
2. **Li Y, Zhao Z, Bai F.** 2007. High-density cultivation of oleaginous yeast *Rhodosporidium toruloides* Y4 in fed-batch culture. Enzyme Microb Technol **41:**312–317. http://dx.doi.org/10.1016/j.enzmictec.2007.02.008.
3. **Pan JG, Kwak MY, Rhee JS.** 1986. High density cell culture of rhodotorula glutinis using oxygen-enriched air. Biotechnol Lett **8:**715–718. http://dx.doi.org/10.1007/BF01032568.
4. **Kumar S, Kushwaha H, Bachhawat AK, Raghava GP, Ganesan K.** 2012. Genome sequence of the oleaginous red yeast *Rhodosporidium toruloides* MTCC 457. Eukaryot Cell **11:**1083–1084. http://dx.doi.org/10.1128/EC.00156-12.
5. **Zhu Z, Zhang S, Liu H, Shen H, Lin X, Yang F, Zhou YJ, Jin G, Ye M,**

Zou H, Zhao ZK. 2012. A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. Nat Commun **3:**1112. http://dx.doi.org/10.1038/ncomms2112.

6. **Morin N, Calcas X, Devillers H, Durrens P, Sherman DJ, Nicaud JM, Neuvéglise C**. 2014. Draft genome sequence of *Rhodosporidium toruloides* CECT1137, an oleaginous yeast of biotechnological interest. Genome Announc **2**(4):e00641-14. http://dx.doi.org/10.1128/genomeA.00641-14.
7. **Paul D, Magbanua Z, Arick M, II, French T, Bridges SM, Burgess SC, Lawrence ML**. 2014. Genome sequence of the oleaginous yeast *Rhodotorula glutinis* ATCC 204091. Genome Announc **2**(1):e00046-14. http://dx.doi.org/10.1128/genomeA.00046-14.
8. **Liu Y, Koh CM, Ngoh ST, Ji L**. 2015. Engineering an efficient and tight D-amino acid-inducible gene expression system in *Rhodosporidium/Rhodotorula* species. Microb Cell Fact **14:**170. http://dx.doi.org/10.1186/s12934-015-0357-7.
9. **Koh CM, Liu Y, Moehninsi, Du M, Ji L**. 2014. Molecular characterization of *KU70* and *KU80* homologues and exploitation of a *KU70*-deficient mutant for improving gene deletion frequency in *Rhodosporidium toruloides*. BMC Microbiol **14:**50. http://dx.doi.org/10.1186/1471-2180-14-50.
10. **Liu Y, Koh CM, Sun L, Hlaing MM, Du M, Peng N, Ji L**. 2013. Characterization of glyceraldehyde-3-phosphate dehydrogenase gene Rt-*GPD1* and development of genetic transformation method by dominant selection in oleaginous yeast *Rhodosporidium toruloides*. Appl Microbiol Biotechnol **97:**719–729. http://dx.doi.org/10.1007/s00253-012-4223-9.
11. **Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam TW, Li Y, Xu X, Wong GK, Wang J**. 2014. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. BioInformatics **30:**1660–1666. http://dx.doi.org/10.1093/bioinformatics/btu077.
12. **Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB**. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci USA **108:**1513–1518. http://dx.doi.org/10.1073/pnas.1017351108.
13. **Zerbino DR, Birney E**. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
14. **Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I**. 2009. ABySS: a parallel assembler for short read sequence data. Genome Res **19:**1117–1123. http://dx.doi.org/10.1101/gr.089532.108.
15. **Peng Y, Leung HC, Yiu SM, Chin FY**. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. BioInformatics **28:**1420–1428. http://dx.doi.org/10.1093/bioinformatics/bts174.
16. **Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA**. 2013. The MaSuRCA genome assembler. Bioinformatics **29:**2669–2677. http://dx.doi.org/10.1093/bioinformatics/btt476.
17. **Gurevich A, Saveliev V, Vyahhi N, Tesler G**. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics **29:**1072–1075. http://dx.doi.org/10.1093/bioinformatics/btt086.
18. **Borodovsky M, Lomsadze A**. 2011. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Curr Protoc Bioinformatics **35:**. http://dx.doi.org/10.1002/0471250953.bi0406s35.
19. **Holt C, Yandell M**. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics **12:**491. http://dx.doi.org/10.1186/1471-2105-12-491.
20. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL**. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10:**421. http://dx.doi.org/10.1186/1471-2105-10-421.