Eukaryotic Cell

Draft Genome Sequence of Salt-Tolerant Yeast Debaryomyces hansenii var. hansenii MTCC 234

Shailesh Kumar, Anmoldeep Randhawa, Kaliannan Ganesan, Gajendra Pal Singh Raghava and Alok K. Mondal
Eukaryotic Cell **2012, 11(7):961. DOI: 10.1128/EC.00137-12.**

EC Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequence of Salt-Tolerant Yeast *Debaryomyces hansenii* var. *hansenii* MTCC 234

Shailesh Kumar, Anmoldeep Randhawa, Kaliannan Ganesan, Gajendra Pal Singh Raghava, and Alok K. Mondal

CSIR—Institute of Microbial Technology, Sector 39-A, Chandigarh, India

***Debaryomyces hansenii* is one of the most halotolerant species of yeast, and the genome sequence of *D. hansenii* strain CBS767 is already available. Here we report the 11.46-Mb draft genome of *D. hansenii* strain MTCC 234, which is even more halotolerant than strain CBS767. Comparative analysis of these sequences would definitely provide further insight into the halotolerance of this yeast.**

Debaryomyces hansenii is a biotechnologically important yeast with interesting genetic and biochemical properties (3). It was originally isolated from saline environments such as seawater and concentrated brines. *D. hansenii* strains are also associated with cheese and meat processing. It is one of the most halotolerant species of yeast, and the presence of sodium in the medium appears to stimulate its growth. The molecular basis for the halophilic nature of this yeast has drawn considerable attention in the recent past (1, 11). Unlike other yeasts, *D. hansenii* is considered a sodium includer, and the accumulation of a large amount of NaCl does not have any adverse effect on its physiology (13). *D. hansenii* is one of the important extremophilic yeasts that can utilize xylose. Attempts have been made to produce xylitol from wood hydrolysate using *D. hansenii* strains. Besides xylitol, strains of *D. hansenii* are also known to produce arabitol and riboflavin (3). Because of its halotolerance and unique phylogenetic position, it was selected as one of the hemiascomycetous yeast species used for comparative genomic and evolutionary studies in the Génolevures project (5).

Here we report the genome sequence of *D. hansenii* strain MTCC 234, originally isolated from New Zealand soil. Compared to *D. hansenii* strain CBS767, whose genome was sequenced previously, MTCC 234 is more halotolerant and it also produces riboflavin and arabitol. The genome of *D. hansenii* MTCC 234 was sequenced by using Illumina GA IIX at Genotypic Technology, Bangalore, India. A total of 29,714,918 single-end reads with a length of 72 nucleotides were generated with a genome coverage of 175-fold. We have used SeqQC (http://genotypic.co.in/SeqQC.html?mnu=1) and Fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) to get high-quality (cutoff read length for HQ, 70%; cutoff quality score, 20), vector/adaptor-free reads for genome assembly. A total of 25,957,404 high-quality, vector-filtered reads were assembled into 542 contigs (size of 11,462,699 nucleotides) with an N50 contig length of 68,507 bp by the Velvet 1.1.06 (16) software (at a hash length of 57).

The draft genome (542 contigs) has 35.42% G+C content and encodes 69 tRNAs and 3 rRNAs (5S, 18S, and 28S rRNAs), as predicted by tRNAscan-SE v 1.23 (10) and the RNAmmer 1.2 Server (8), respectively. Gene prediction and annotation were done by the MAKER (4) pipeline by using several executables—RepeatMasker (12), BLAST (2), SNAP (7), Exonerate (14), Genemark (9), and Augustus (15). The predicted proteins (5,313; minimum length of 22 amino acids, maximum length of 4,972 amino acids) were searched against the nonredundant NCBI database, and matches were found for 5,294 proteins at an E value cutoff of 10^{-6}. Of these, 5,069 proteins could be mapped to the UniProt database. We found the following gene ontology terms after mapping: biological process, 1,904; cellular component, 1,817; molecular function, 2,863. We also mapped the proteins to KEGG (6) and found four genes for riboflavin metabolism and five genes for the pentose and glucuronate interconversion pathway.

Nucleotide sequence accession number. This whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. AHBE00000000. The version described in this paper is the first version (accession no. AHBE01000000).

ACKNOWLEDGMENTS

Financial support for this work was from CSIR, India. S.K. and A.R. are recipients of a senior research fellowship from CSIR, India.

REFERENCES

1. **Aggarwal M, Mondal AK.** 2006. Role of N-terminal hydrophobic region in modulating the subcellular localization and enzyme activity of the bisphosphate nucleotidase from *Debaryomyces hansenii*. Eukaryot. Cell **5**:262–271.
2. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
3. **Breuer U, Harms H.** 2006. *Debaryomyces hansenii*—an extremophilic yeast with biotechnological potential. Yeast **23**:415–437.
4. **Cantarel BL, et al.** 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. **18**:188–196.
5. **Dujon B, et al.** 2004. Genome evolution in yeasts. Nature **430**:35–44.
6. **Kanehisa M, Goto S.** 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. **28**:27–30.
7. **Korf I.** 2004. Gene finding in novel genomes. BMC Bioinformatics **5**:59.
8. **Lagesen K, et al.** 2007. RNAmmer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. **35**:3100–3108.
9. **Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M.** 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. **33**:6494–6506.
10. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.

Received 28 April 2012 Accepted 30 April 2012

Address correspondence to Alok K. Mondal, alok@imtech.res.in, or Gajendra Pal Singh Raghava, raghava@imtech.res.in.

doi:10.1128/EC.00137-12

11. **Minhas A, et al.** 2012. Conserved Ser/Arg-rich motif in PPZ orthologs from fungi is important for its role in cation tolerance. J. Biol. Chem. **287**:7301–7312.
12. **Price AL, Jones NC, Pevzner PA.** 2005. De novo identification of repeat families in large genomes. Bioinformatics **21**(Suppl 1):i351–i358.
13. **Prista C, Loureiro-Dias MC, Montiel V, Garcia R, Ramos J.** 2005. Mechanisms underlying the halotolerant way of Debaryomyces hansenii. FEMS Yeast Res. **5**:693–701.
14. **Slater GS, Birney E.** 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics **6**:31. doi:10.1186/1471-2105-6-31.
15. **Stanke M, Steinkamp R, Waack S, Morgenstern B.** 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. **32**:W309–312.
16. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.