Draft Genome Sequence of *Saccharomyces cerevisiae* Strain NCIM3186 Used in the Production of Bioethanol from Sweet Sorghum

Burragoni Sravanthi Goud, Kandasamy Ulaganathan

Centre for Plant Molecular Biology, Osmania University, Hyderabad, Telangana, India

Here, we report the draft genome sequence of *Saccharomyces cerevisiae* strain NCIM3186 used in bioethanol production from sweet sorghum. The size of the genome is approximately 11.9 Mb and contains 5,347 protein-coding genes.

Address correspondence to Kandasamy Ulaganathan, kulaganathan123@gmail.com.

*S*accharomyces cerevisiae is the most preferred host organism for lignocellulosic bioethanol production due to its ability in fermentative production of high ethanol, inhibitor tolerance, and suitability for heterologous expression of genes (1–3). Understanding the genomic variations that facilitate high ethanol production by *S. cerevisiae* is necessary for engineering strains for lignocellulosic bioethanol production. Many strains used in bioethanol production have been sequenced, and a number of variations have been identified (4–8). In our effort to select a suitable strain for lignocellulosic bioethanol production, we have sequenced strains differing in their ability to produce bioethanol from plant biomass and reported the genome sequence of a moderate bioethanol-producing strain, NCIM3107 (9). Here, we report the draft genome sequence of *S. cerevisiae* strain NCIM3186 used in bioethanol production from sweet sorghum (10).

S. cerevisiae strain NCIM3186 was obtained from the Microbial Type Culture Collection, Chandigarh, India. Genome sequencing was done using the Illumina MiSeq sequencing platform (Illumina, San Diego, CA). The sequencing generated paired-end read files R1 and R2 consisting of a total of 15,216,196 reads. Quality assessment was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adapter trimming by the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The genome assembly was performed by using both *de novo* and reference-based genome assemblers. The *S. cerevisiae* S288C R64-1.1 reference genome was used to perform a reference-based assembly of raw reads by using Bowtie 2 (11), which generated a SAM alignment file as output. Using this SAM file BAM file, sorted BAM and indexed BAM files were made using SAMtools (12). The overall coverage was estimated to be 87.7×. *De novo* assembly of reads was performed using the Velvet assembler (13). Using SAMtools, a consensus sequence of size 11.9 Mb was created. Variant calling was performed using *S. cerevisiae* S288C R64-1.1 reference genome and ≥20-read depth and ≥30 mapping quality score as cutoffs. Further annotation of variants was done using SnpEFF 3.4 (14). Gene prediction was done by using Augustus (15). A total of 5,347 protein-coding genes were found, which showed homology to *S. cerevisiae* genes in the *Saccharomyces* Genome Database (SGD) (16).

Nucleotide sequence accession numbers. The nucleotide sequences of the NCIM3186 strain of the *S. cerevisiae* genome have been deposited in GenBank under the accession numbers CP011810 to CP011825 and the mitochondrial genome under the accession no. CP011826.

REFERENCES

1. **Hahn-Hägerdal B, Galbe M, Gorwa-Grauslund MF, Lidén G, Zacchi G.** 2006. Bio-ethanol—the fuel of tomorrow from the residues of today. Trends Biotechnol **24**:549–556. http://dx.doi.org/10.1016/j.tibtech.2006.10.004.
2. **Hahn-Hägerdal B, Karhumaa K, Fonseca C, Spencer-Martins I, Gorwa-Grauslund MF.** 2007. Towards industrial pentose fermenting yeast strains. Appl Microbiol Biotechnol **74**:937–953. http://dx.doi.org/10.1007/s00253-006-0827-2.
3. **Ulaganathan K, Goud BS, Reddy MM, Kumar VP, Balsingh J, Radhakrishna S.** 2015. Proteins for breaking barriers in lignocellulosic bioethanol production. Curr Protein Pept Sci **16**:100–134. http://dx.doi.org/10.2174/1389203716602150215165718.
4. **Akao T, Yashiro I, Hosoyama A, Kitagaki H, Horikawa H, Watanabe D, Akada R, Ando Y, Harashima S, Inoue T, Inoue Y, Kajiwara S, Kitamoto K, Kitamoto N, Kobayashi O, Kuhara S, Masubuchi T, Mizoguchi H, Nakao Y, Nakazato A, Namise M, Oba T, Ogata T, Ohta A, Sato M, Shibasaki S, Takatsume Y, Tanimoto S, Tsuboi H, Nishimura A, Yoda K, Ishikawa T, Iwashita K, Fujita N, Shimoi H.** 2011. Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no. 7. DNA Res **18**:423–434. http://dx.doi.org/10.1093/dnares/dsr029.
5. **Argueso JL, Carazzolle MF, Mieczkowski PA, Duarte FM, Netto OV, Missawa SK, Galzerani F, Costa GG, Vidal RO, Noronha MF, Dominska M, Andrietta MG, Andrietta SR, Cunha AF, Gomes LH, Tavares FC, Alcarde AR, Dietrich FS, McCusker JH, Petes TD, Pereira GA.** 2009. Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. Genome Res **19**:2258–2270. http://dx.doi.org/10.1101/gr.091777.109.
6. **Babrzadeh F, Jalili R, Wang C, Shokralla S, Pierce S, Robinson-Mosher A, Nyren P, Shafer RW, Basso LC, de Amorim HV, de Oliveira AJ, Davis RW, Ronaghi M, Gharizadeh B, Stambuk BU.** 2012. Whole-genome sequencing of the efficient industrial fuel-ethanol fermentative

saccharomyces cerevisiae strain CAT-1. Mol Genet Genomics **287:** 485–494. http://dx.doi.org/10.1007/s00438-012-0695-7.

7. **Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ**. 2011. Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet **7:**e1001287. http://dx.doi.org/10.1371/journal.pgen.1001287.

8. **Zheng DQ, Wang PM, Chen J, Zhang K, Liu TZ, Wu XC, Li YD, Zhao YH**. 2012. Genome sequencing and genetic breeding of a bioethanol *Saccharomyces cerevisiae* strain YJS329. BMC Genomics **13:**479. http://dx.doi.org/10.1186/1471-2164-13-479.

9. **Ulaganathan K, Sravanthi Goud B, Reddy MM, Praveen Kumar V, Radhakrishna S, Balsingh J**. 2015. Genome sequence of *Saccharomyces cerevisiae* NCIM3107, used in bioethanol production. Genome Announc **3**(1):e01557-14. http://dx.doi.org/10.1128/genomeA.01557-14.

10. **Prasad S, Lata JHC, Pathak H**. 2009. Selection of efficient *Saccharomyces cerevisiae* strain for ethanol production from sorghum stalk juice. Curr Adv J Agric Sci **1:**70–72.

11. **Langmead B, Salzberg SL**. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357–359. http://dx.doi.org/10.1038/nmeth.1923.

12. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup**. 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25:**2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.

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. **Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM**. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) **6:**80–92. http://dx.doi.org/10.4161/fly.19695.

15. **Stanke M, Diekhans M, Baertsch R, Haussler D**. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics **24:**637–644. http://dx.doi.org/10.1093/bioinformatics/btn013.

16. **Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED**. 2012. *Saccharomyces* Genome Database: the genomics resource of budding yeast. Nucleic Acids Res **40:**D700–D705. http://dx.doi.org/10.1093/nar/gkr1029.