Genome Sequence of *Saccharomyces cerevisiae* NCIM3107, Used in Bioethanol Production

Kandasamy Ulaganathan, Burragoni Sravanthi Goud, Mettu Madhavi Reddy, Vanaparthi Praveen Kumar, Surabhi Radhakrishna, Jatoth Balsingh

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

Here, we report the genome of *Saccharomyces cerevisiae* strain NCIM3107, used in bioethanol production. The genome size is approximately 11.8 Mb and contains 5,435 protein-coding genes.

An ideal organism for lignocellulosic bioethanol production should be able to utilize lignocellulose, ferment hexose and pentose sugars, have a high ethanol yield, and tolerate high ethanol concentration (1, 2). Although a number of organisms, like *Escherichia coli*, *Zymomonas mobilis*, *Clostridium* species, and *Trichoderma reesei*, have been tried for use in lignocellulosic ethanol production, *Saccharomyces cerevisiae* is the most widely used organism (3, 4). *S. cerevisiae* meets most needs for a bioethanol producer, and its inability to utilize pentose sugars has been addressed by genetic engineering (4–6). Understanding the genetic background of the strain is important for its use in industrial production and metabolic engineering. As part of our effort to select a suitable host strain for the heterologous expression of genes associated with lignocellulosic bioethanol production, we sequenced the NCIM3107 strain of *S. cerevisiae* obtained from the Microbial Type Culture Collection, Chandigarh, India. This strain has been tested for bioethanol production and found to be a moderate producer compared to other tested strains (7). Recent comparative genomics work has revealed that many yeast strains possess unique variations and genes, so sequencing the genome would be the first step in understanding the genetic background of a strain to be used in industrial production and for use in metabolic engineering (8–12). Here, we report the genome sequence of *S. cerevisiae* strain NCIM3107.

The genome was sequenced with the Illumina MiSeq system. A total of 1,459,146 reads were generated, with a coverage of 27.35-fold. The genome was assembled using the Genotypic tool. A genome of 11.8 Mb was assembled, covering 16 chromosomes. The raw reads obtained were aligned to the reference *S. cerevisiae* R64 genome with the Bowtie 2 tool (13). Using SAMtools, the variants were detected in comparison to the *S. cerevisiae* R64 reference genome, with cutoffs of $\geq$20-read depth and $\geq$30 mapping quality score (14). The variants detected with SAMtools were further annotated using SnpEFF 3.4 to give the locations (intronic/exonic/untranslated region [UTR]), gene names, protein changes, and functions of the variants (15).

Gene prediction and annotation were performed using the Augustus software with the training sets available for *S. cerevisiae* (16). There were 5,435 protein-coding genes, which showed homology to *S. cerevisiae* genes found in the *Saccharomyces* Genome Database (SGD) (17).

Nucleotide sequence accession numbers. The nucleotide sequences of the *S. cerevisiae* NCIM3107 genome have been deposited in GenBank under the accession numbers CP009944 to CP009960.

REFERENCES

1. **Fischer CR, Klein-Marcuschamer D, Stephanopoulos G.** 2008. Selection and optimization of microbial hosts for biofuels production. Metab Eng **10:**295–304. http://dx.doi.org/10.1016/j.ymben.2008.06.009.
2. **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, in press.
3. **Roehr M, Kosaric N, Vardar-Sukan F, Pieper HJ, Senn T.** 2001. The biotechnology of ethanol: classical and future applications. Wiley-VCH Press, Weinheim, Germany.
4. **Demain AL, Newcomb M, Wu JH.** 2005. Cellulase, clostridia, and ethanol. Microbiol Mol Biol Rev **69:**124–154. http://dx.doi.org/10.1128/MMBR.69.1.124-154.2005.
5. **Barnett JA, Barnett L.** 2011. Yeast research: a historical overview. ASM Press, Herndon, VA.
6. **Gray KA, Zhao L, Emptage M.** 2006. Bioethanol. Curr Opin Chem Biol **10:**141–146. http://dx.doi.org/10.1016/j.cbpa.2006.02.035.
7. **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.
8. **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.

9. **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.

10. **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.

11. **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.

12. **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.

13. **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.

14. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin, 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.

15. **Cingolani P, Platts A, Wang le L, 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.

16. **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.

17. **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.