

# Draft Genome Sequence of *Saccharomyces cerevisiae* IR-2, a Useful Industrial Strain for Highly Efficient Production of Bioethanol

Takehiko Sahara,<sup>a</sup> Kazuhiro E. Fujimori,<sup>a</sup> Maiko Nezu,<sup>b</sup> Masatoshi Tsukahara,<sup>b</sup> Yuki Tochigi,<sup>c,d</sup> Satoru Ohgiya,<sup>a,c</sup> Yoichi Kamagata<sup>a,c</sup>

Bioproduction Research Institute (BPRI), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan<sup>a</sup>; BioJet Co., Ltd., Uruma, Japan<sup>b</sup>; Bioproduction Research Institute (BPRI), National Institute of Advanced Industrial Science and Technology (AIST), Toyohira, Sapporo, Japan<sup>c</sup>; Laboratory of Veterinary Physiology, Nippon Veterinary and Life Science University, Musashino, Tokyo, Japan<sup>d</sup>

**We sequenced the genome of *Saccharomyces cerevisiae* IR-2, which is a diploid industrial strain with flocculation activity and the ability to efficiently produce bioethanol. The approximately 11.4-Mb draft genome information provides useful insights into metabolic engineering for the production of bioethanol from biomass.**

Received 12 December 2013 Accepted 13 December 2013 Published 16 January 2014

Citation Sahara T, Fujimori KE, Nezu M, Tsukahara M, Tochigi Y, Ohgiya S, Kamagata Y. 2014. Draft genome sequence of *Saccharomyces cerevisiae* IR-2, a useful industrial strain for highly efficient production of bioethanol. *Genome Announc.* 2(1):e01160-13. doi:10.1128/genomeA.01160-13.

Copyright © 2014 Sahara et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Takehiko Sahara, t-sahara@aist.go.jp.

*Saccharomyces cerevisiae*, the most popular ethanol-producing microorganism, has been used in the production of various alcoholic beverages. In the last few decades, a detailed understanding of the metabolic and genetic regulations of ethanol production from *S. cerevisiae* has become important for improved production of ethanol as a biofuel. Recently, several *S. cerevisiae* industrial strains have been sequenced (1–4). By comparing the publicly available genome sequences, a recent study revealed that *S. cerevisiae* species have a large number of genetic divergences (5). However, the genetic basis of their efficient bioethanol production remains unclear.

The flocculating *S. cerevisiae* strain IR-2 was originally isolated from a fermented food of Indonesia (6). Previous studies demonstrated that this strain has a high fermentation activity (7, 8). Further, the IR-2 strain, engineered with three enzyme genes for xylose metabolism (xylose reductase, xylitol dehydrogenase, and xylulose kinase), has been shown to efficiently produce ethanol from xylose (9, 10).

These data reveal that IR-2 potentially has a genetic background that is beneficial for effective bioethanol production from C<sub>6</sub> and C<sub>5</sub> carbon sources. A genomic analysis of IR-2 is therefore essential for further refinement of the xylose metabolic pathways and also provides a gateway for improving and developing effective systems for ethanol fermentation from different types of biomass.

Here, we report the draft genome sequence of the industrial diploid strain IR-2. The IR-2 genome was *de novo* sequenced with the GS FLX Titanium system (Roche Diagnostics, Switzerland) to highly oversample the genome (26.1-fold coverage), with a total of 956,160 reads and the generation of a paired-end library, enabling the assembly of 916 contigs into 90 supercontigs (scaffolds) using the GS *de novo* assembler software (Roche). A genome of 11.4 Mb was covered by 90 scaffolds (N<sub>50</sub> scaffold length, 517,926 bases). Whole-genome resequencing analysis of the IR-2 genome was performed using the SOLiD 3 system (Life Technologies, Inc., Carlsbad, CA) to improve the sequence quality of the draft ge-

nome, and 251,770,705 50-base reads were obtained. The SOLiD 3 reads were aligned to the scaffolds by BWA (11), Bowtie (12), and SAMtools (13) to detect sequencing errors in the scaffolds. As a result, 3,606 nucleotide differences between the scaffolds generated by the GS FLX Titanium system and the reads generated by the SOLiD 3 system were revised. Five hundred ninety of the intercontig gaps were confirmed by PCR and closed by Sanger sequencing of the amplicons with the 3730xl DNA analyzer (Life Technologies).

Gene prediction and annotation were performed using Augustus software with the training set that is available for *S. cerevisiae* (14) and Exonerate software (15). The predicted proteins were searched against the curated open reading frames of the *Saccharomyces* Genome Database (SGD) (<http://www.yeastgenome.org>) (16) by BLASTp software (17), and matches were found for 5,628 protein-encoding genes at an *E* value cutoff of 10<sup>-6</sup>.

**Nucleotide sequence accession numbers.** The nucleotide sequences of the *S. cerevisiae* IR-2 draft genome have been deposited in DDBJ/EMBL/GenBank under the accession no. [BAUI01000001](https://www.ncbi.nlm.nih.gov/nuccore/BAUI01000001) to [BAUI01000322](https://www.ncbi.nlm.nih.gov/nuccore/BAUI01000322). The version described in this paper is the first version.

## ACKNOWLEDGMENT

This work was supported through project P07015 by the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

## REFERENCES

1. 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](https://doi.org/10.1093/dnares/dsr029).
2. Argueso JL, Carazzolle MF, Mieczkowski PA, Duarte FM, Netto OV, Missawa SK, Galzerani F, Costa GG, Vidal RO, Noronha MF, Dominiska 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>.
3. 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>.
  4. 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>.
  5. 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>.
  6. Hiroshi K, Yoshio S, Toshio M, Harumi K, Yorikazu S. 1985. Continuous ethanol fermentation with cell recycling using flocculating yeast. *J. Ferment. Technol.* 63:159–165.
  7. Kida K, Morimura S, Kume K, Suruga K, Sonoda Y. 1991. Repeated-batch ethanol fermentation by a flocculating yeast, *Saccharomyces cerevisiae* IR-2. *J. Ferment. Bioeng.* 71:340–344. [http://dx.doi.org/10.1016/0922-338X\(91\)90347-J](http://dx.doi.org/10.1016/0922-338X(91)90347-J).
  8. Kida K, Kume K, Morimura S, Sonoda Y. 1992. Repeated-batch fermentation process using a thermotolerant flocculating yeast constructed by protoplast fusion. *J. Ferment. Bioeng.* 74:169–173. [http://dx.doi.org/10.1016/0922-338X\(92\)90078-9](http://dx.doi.org/10.1016/0922-338X(92)90078-9).
  9. Matsushika A, Inoue H, Watanabe S, Kodaki T, Makino K, Sawayama S. 2009. Efficient bioethanol production by a recombinant flocculent *Saccharomyces cerevisiae* strain with a genome-integrated NADP<sup>+</sup>-dependent xylitol dehydrogenase gene. *Appl. Environ. Microbiol.* 75: 3818–3822. <http://dx.doi.org/10.1128/AEM.02636-08>.
  10. Matsushika A, Inoue H, Murakami K, Takimura O, Sawayama S. 2009. Bioethanol production performance of five recombinant strains of laboratory and industrial xylose-fermenting *Saccharomyces cerevisiae*. *Bioresour. Technol.* 100:2392–2398. <http://dx.doi.org/10.1016/j.biortech.2008.11.047>.
  11. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>.
  12. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. <http://dx.doi.org/10.1186/gb-2009-10-3-r25>.
  13. 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>.
  14. Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24:637–644. <http://dx.doi.org/10.1093/bioinformatics/btn013>.
  15. Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. <http://dx.doi.org/10.1186/1471-2105-6-31>.
  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>.
  17. 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>.