

# Draft Genome Sequence of the Archiascomycetous Yeast *Saitoella complicata*

Kenta Yamauchi,<sup>a</sup> Shinji Kondo,<sup>b</sup> Makiko Hamamoto,<sup>c</sup> Yurika Takahashi,<sup>a</sup> Yoshitoshi Ogura,<sup>d</sup> Tetsuya Hayashi,<sup>d</sup> Hiromi Nishida<sup>a</sup>

Biotechnology Research Center and Department of Biotechnology, Toyama Prefectural University, Imizu, Toyama, Japan<sup>a</sup>; National Institute of Polar Research, Tokyo, Japan<sup>b</sup>; Department of Life Sciences, School of Agriculture, Meiji University, Kanagawa, Japan<sup>c</sup>; Frontier Science Research Center and School of Medicine, University of Miyazaki, Miyazaki, Japan<sup>d</sup>

K.Y. and S.K. contributed equally to this article.

**The draft genome sequence of the archiascomycetous yeast *Saitoella complicata* was determined. The assembly of newly and previously sequenced data sets resulted in 104 contigs (total of 14.1 Mbp;  $N_{50}$ , 239 kbp). On the newly assembled genome, a total of 6,933 protein-coding sequences (7,119 transcripts, including alternative splicing forms) were identified.**

Received 14 February 2015 Accepted 27 April 2015 Published 28 May 2015

**Citation** Yamauchi K, Kondo S, Hamamoto M, Takahashi Y, Ogura Y, Hayashi T, Nishida H. 2015. Draft genome sequence of the archiascomycetous yeast *Saitoella complicata*. *Genome Announc* 3(3):e00220-15. doi:10.1128/genomeA.00220-15.

**Copyright** © 2015 Yamauchi 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 Shinji Kondo, shinji.kondo@gmail.com, or Hiromi Nishida, hnishida@pu-toyama.ac.jp.

The subphylum *Taphrinomycotina* (*Archiascomycetes*) is the earliest ascomycetous lineage that diverged before the separation of the subphyla *Pezizomycotina* (*Euascomycetes*, filamentous ascomycetes) and *Saccharomycotina* (*Hemiascomycetes*, budding ascomycetous yeasts) (1, 2). The anamorphic and saprobic budding yeast *Saitoella complicata* is a member of the *Taphrinomycotina*, which was isolated from Himalayan soil (3). Interestingly, *S. complicata* shares some characteristics with both ascomycetous and basidiomycetous yeasts (3, 4).

We previously attempted to assemble the genome sequence of *S. complicata* using 454 (Roche) sequences (5) and Illumina paired-end read pairs (6). Although these previous assemblies were of a large number of small contigs, at 7,981 contigs (13.0 Mbp) (5) and 1,800 contigs (14.2 Mbp) (6), respectively, we found that the amino acid sequences of protein-coding genes identified on the contigs showed the highest similarity to proteins of *Pezizomycotina* (5, 6).

To elucidate the detailed characteristics of the *S. complicata* genomic DNA sequences, we have refined the genome assembly with additional sequencing of mate-paired DNA libraries of this species. We generated a total of 11.4 million paired-end read pairs (700-bp insert and 100 bp in length) and a total of 23.7 million mate-paired read pairs (6.2 million 3-kb-, 6.2 million 5-kb-, 5.3 million 10-kb-, and 6.0 million 15-kb-long-insert read pairs), respectively, using Illumina HiSeq and MiSeq sequencers. The read pairs were dereplicated by Fulcrum (7) and assembled using the SPAdes assembler (8). The assembly of the dereplicated read pairs by using 21 to 89 bp for the *k*-mer size option yielded a set of 104 contigs of  $\geq 1$  kb, whose total size and  $N_{50}$  are 14.1 Mb and 239 kb, respectively.

Using Augustus (9), a gene prediction software based on the alignment of expressed sequences to the genome, we have determined coding sequences (CDSs) of the genes expressed on the assembled genome of *Saitoella* according to the gene model of *Aspergillus nidulans*, which is thought to have some taxonomic proximity to *Saitoella*. Based on the exon coordinates mapped by a total of 89.3 million RNA sequencing (RNA-seq) paired-end

read pairs (100 bp in length) uniquely mapped to the genome by BLAT (10), Augustus identified 6,933 protein-coding genes (7,119 transcripts, including alternative splicing forms) on the *Saitoella* genome. All this computational work was done on the NIG Supercomputer system (11).

**Nucleotide sequence accession numbers.** The DNA sequences have been deposited in DDBJ under the accession numbers BACD03000001 to BACD03000104.

## ACKNOWLEDGMENTS

We thank Junta Sugiyama for his valuable comments.

This work was supported by JSPS KAKENHI grant no. 25440188 and 221S0002.

## REFERENCES

- Nishida H, Sugiyama J. 1994. Archiascomycetes: detection of a major new lineage within the *Ascomycota*. *Mycoscience* 35:361–366. <http://dx.doi.org/10.1007/BF02268506>.
- Liu Y, Leigh JW, Brinkmann H, Cushion MT, Rodriguez-Ezpeleta N, Philippe H, Lang BF. 2009. Phylogenomic analyses support the monophyly of *Taphrinomycotina*, including *Schizosaccharomyces* fission yeasts. *Mol Biol Evol* 26:27–34. <http://dx.doi.org/10.1093/molbev/msn221>.
- Goto S, Sugiyama J, Hamamoto M, Komagata K. 1987. *Saitoella*, a new anamorph genus in the *Cryptococcaceae* to accommodate two Himalayan yeast isolates formerly identified as *Rhodotorula glutinis*. *J Gen Appl Microbiol* 33:75–85. <http://dx.doi.org/10.2323/jgam.33.75>.
- Sugiyama J, Fukagawa M, Chiu S, Komagata K. 1985. Cellular carbohydrate composition, DNA base composition, ubiquinone systems, and diazotium blue B color test in the genera *Rhodospiridium*, *Leucosporidium*, *Rhodotorula* and related basidiomycetous yeasts. *J Gen Appl Microbiol* 31:519–550. <http://dx.doi.org/10.2323/jgam.31.519>.
- Nishida H, Hamamoto M, Sugiyama J. 2011. Draft genome sequencing of the enigmatic yeast *Saitoella complicata*. *J Gen Appl Microbiol* 57: 243–246. <http://dx.doi.org/10.2323/jgam.57.243>.
- Nishida H, Matsumoto T, Kondo S, Hamamoto M, Yoshikawa H. 2014. The early diverging ascomycetous budding yeast *Saitoella complicata* has three histone deacetylases belonging to the Clr6, Hos2, and Rpd3 lineages. *J Gen Appl Microbiol* 60:7–12. <http://dx.doi.org/10.2323/jgam.60.7>.

7. Burriesci MS, Lehnert EM, Pringle JR. 2012. Fulcrum: condensing redundant reads from high-throughput sequencing studies. *Bioinformatics* 28:1324–1327. <http://dx.doi.org/10.1093/bioinformatics/bts123>.
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
9. 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>.
10. Kent WJ. 2002. BLAT—the BLAST-Like Alignment Tool. *Genome Res* 12:656–664. <http://dx.doi.org/10.1101/gr.229202>.
11. Ogasawara O, Mashima J, Kodama Y, Kaminuma E, Nakamura Y, Okubo K, Takagi T. 2013. DDBJ new system and service refactoring. *Nucleic Acids Res* 41:D25–D29. <http://dx.doi.org/10.1093/nar/gks1152>.