Microbiol. Biotechnol. Lett. (2023), 51(4), 551–554
http://dx.doi.org/10.48022/mbl.2311.11004
pISSN 1598-642X eISSN 2234-7305

MBL
Microbiology and Biotechnology Letters

Genome Reports

Genome Sequence of the Yeast Strain *Sporobolomyces phaffii* RJAF-17, Which Produces the Lipoamino Acid Surfactants

Parthiban Subramanian[1], Jeong-Seon Kim[2], Jun Heo[2], and Yiseul Kim[2]*

[1]National Agrobiodiversity Center, National Institute of Agricultural Sciences, Rural Development Administration, Wanju 55365, Republic of Korea
[2]Agricultural Microbiology Division, National Institute of Agricultural Sciences, Rural Development Administration, Wanju 55365, Republic of Korea

Received: November 7, 2023 / Accepted: December 6, 2023

We report the draft genome sequence of *Sporobolomyces phaffii* RJAF-17, a basidiomycetous yeast strain producing lipoamino acid surfactants, N-palmitoyl leucine and N-parmitoleyl glutamine. The annotation and classification of protein-coding genes provided the basic information for the genome of strain RJAF-17, including prediction of abundant genes as well as detection of genes involved in the biosynthesis of lipoamino acids. With the molecular importance of lipoamino acids as promising alternatives to chemical surfactants, the genomic information of strain RJAF-17 can help us understand the role of biomolecules in yeasts and explore possibilities of large-scale synthesis for industrial applications.

Keywords: **Yeast, *Sporobolomyces phaffii*, biosurfactant, genome**

In response to high demand for environmental and industrial applications of microorganisms, efforts have been made to explore microorganisms with biotechnological potential. Among several useful products derived from microorganisms, biosurfactants have a significant importance as they are widely used around the world for a variety of purposes. According to a recent survey, the global market size of biosurfactants is projected to reach a compound annual growth rate of 5.5% from 2020 to 2026, while the market size of chemical surfactants is estimated to grow at around 5.3% from 2020 to 2027 (https://www.alliedmarketresearch.com). Adverse effects of climate change and surge of overall pollution urge the need to look for environmentally friendly biosurfactants for everyday applications. With this in mind, wild flowers were collected in Gwangyang-si, Jeollanam-do, Republic of Korea for isolation of yeast strains, of which *Sporobolomyces phaffii* RJAF-17 from *Prunus mume* for. *alphandii* was selected for further analyses. Briefly, approximately 3 g of the sample was suspended in 10 ml of sterile saline. The suspension was serially diluted and 0.1 ml of each dilution was spread onto Yeast Malt Agar (Difco). After incubation at 25℃ for 3–4 days, colonies with different colors and morphologies were isolated. Analysis of the internal transcribed spacer region of strain RJAF-17 exhibited the highest similarity value to *Sporobolomyces phaffii* in the phylum *Basidiomycota*. Characterization of the compounds produced by strain RJAF-17, including measurement of surface tension and determination of chemical structure revealed its capability to synthesize lipoamino acid surfactants (patent application number 10-2023-0149087). A genomic investigation of strain RJAF-17 would assist in identifying the genetic sources which can be developed later for industrial applications. Hence, we present here the genomic information of biosurfactant-producing

***Corresponding author**
Phone: +82-63-238-3028
E-mail: dew@korea.kr

yeast strain RJAF-17.

The genome of strain RJAF-17 was sequenced using a combination of HiFi sequencing (Sequel II System) and Illumina HiSeq X-ten (Illumina, USA) platforms provided by Macrogen, Republic of Korea. High quality *de novo* genome assemblies were generated based on HiFi reads by using genome assembly application from PacBio. First, PanCake 1.1.2 [1] was used to overlap the reads and Nighthawk to phase the overlapped reads. Following the removal of chimeras and duplicates from the overlapped reads, a string graph was constructed and primary contigs as well as haplotigs were generated. The primary contigs and haplotigs were polished using Racon 1.5.0 [2]. To remove haplotype duplications in the primary contig set, purge_dups was used for retrieving potential haplotype duplications and move them to the haplotig set. Following this, the genome assembly generated was subjected to further analyses. The raw sequencing data was submitted to the National Center for Biotechnology Information (NCBI) under the BioProject PRJNA1005338 with the accession number SRX21362516 and SRX21362517.

All further analyses were carried out on the Galaxy Web platform (https://usegalaxy.org). Assessments of assembly completeness and repetitive elements were conducted with BUSCO 4.1.4 [3] using the dothideomycetes_odb10 lineage dataset and with RepeatMasker 4.1.5., respectively. For quality assessment of gene prediction, annotations of strain RJAF-17 were performed using AUGUSTUS and MAKER software with *Neurospora crassa* as a model for training [4]. A comparison of the annotations using the two different tools are provided in Table 1. The functional annotation was carried out based on publicly available databases, including Clusters of Orthologous Groups (COG), Gene Ontology (GO), Carbohydrate-Active Enzymes (CAZy), and Pfam using eggNOG-mapper 2 [5]. Additionally, the genome was screened for Kyoto Encyclopedia of Genes and Genomes and Protein Analysis THrough Evolutionary Relationships using KOBAS 2.0 to study functional metabolism of genes [6]. Prediction of secondary metabolite gene clusters was accomplished with the fungal version of antiSMASH 7.0 (https://fungismash.secondarymetabolites.org/#!/start). Average nucleotide identity (ANI) was calculated by ANI calculator (http://enve-omics.ce.gatech.edu/ani/index).

Table 1. Statistics of gene prediction using the different programs.

Attribute	AUGUSTUS	MAKER
Contigs	13	13
Number of genes predicted	5,861	7,864
Number of transcripts predicted	5,861	7,864
Complete BUSCOs	1,468	1,658
Missing BUSCOs	212	88
Number of selected queries by EggNOG-mapper	4,362 (74.4%)	6,015 (76.4%)
Pfam hits*	4,086	5,618
GO hits*	2,656	3,582
EC hits*	1,237	1,680
CAZy hits*	80	117

*Number of predicted genes that contain at least one Pfam domain, one GO term, one enzyme, and one CAZy hit.

The genome assembly was 19.7 Mb (19,710,391 bp) in size and consisted of 13 scaffolds. The largest scaffold was 3,769,252 bp long and the shortest was 463,633 bp long with N50 value of 1.9 Mb (1,875,694 bp). The GC content was estimated to be 57.5%. Completeness of the genome assembly was 92.2%. Analysis of repetitive elements exhibited very few repeat elements in the genome (2.86%). The predicted genes using AUGUSTUS (5,861 genes) and MAKER (7,864 genes) were submitted to eggNOG-mapper for scanning (Table 1). Although there is currently no ANI threshold range for yeast species demarcation, calculation of ANI was performed between strain RJAF-17 and three available genomes of the genus *Sporobolomyces* at NCBI. The ANI results indicated that strain RJAF-17 showed the highest values with *Sporobolomyces salmonicolor* (GCA 001373355.1, 76.7%), followed by *Sporobolomyces pararoseus* (GCA 010758995.1, 76.4%) and *Sporobolomyces roseus* (GCA 016617785.1, 76.3%).

Functional annotation using outputs from both AUGUSTUS and MAKER elucidated functional traits of the coding sequences (Fig. 1). In terms of COG categories, strain RJAF-17 contained many genes related with “post-translational modification, protein turnover, chaperones” followed by “translation, ribosomal structure and biogenesis”. Analysis using fungal antiSMASH resulted in identification of non-ribosomal peptide synthetase (NRPS domain, contigs 1 and 3), terpene synthesis domain (contigs 2 and 4), and betalactone synthesis domain (contig 2) (Table 2). As mentioned earlier, strain

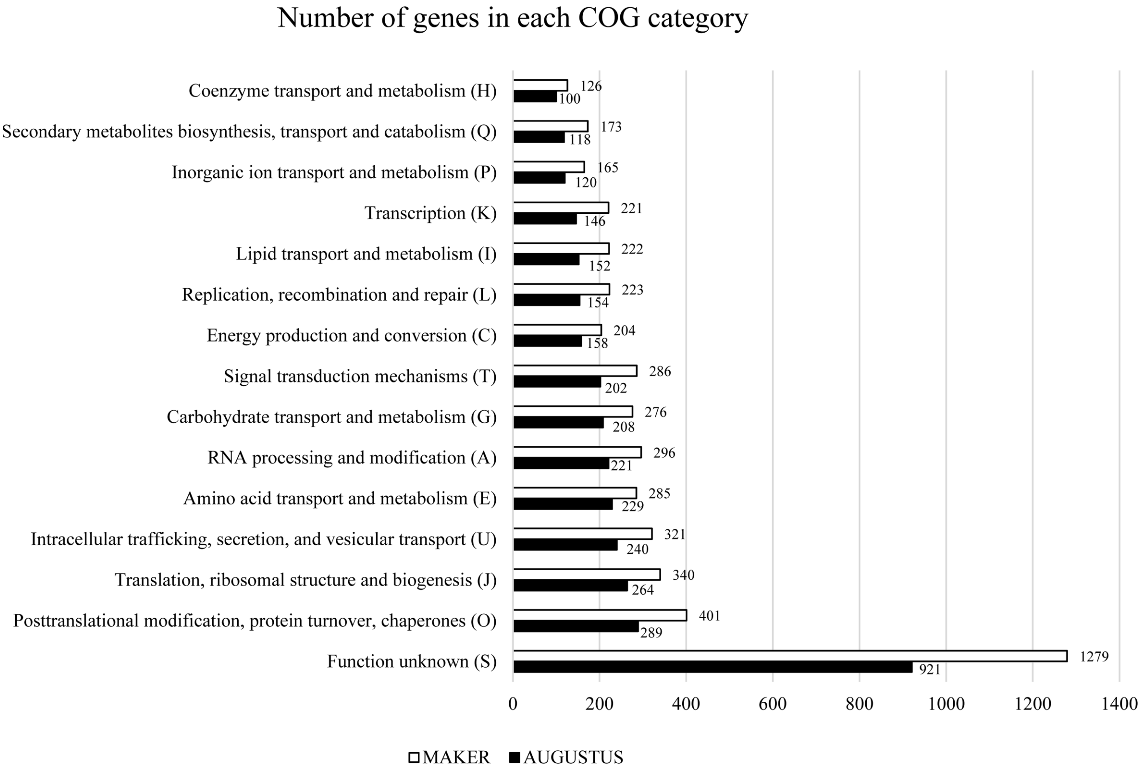

Fig. 1. Functional annotation of the genome of strain RJAF-17 indicating top 15 COG categories.

Table 2. Secondary metabolite gene clusters determined by the fungal version of antiSMASH.

Contig	Region	Type	From	To	Most similar known clusters	Similarity
Contig 1	1.1	NRPS	1,464,762	1,510,798	Nonribosomal peptide synthase of yeast *Rhodotorula toruloides*	54.9%
Contig 2	2.1	Terpene	56,421	78,302	Hypothetical protein from *Rhodotorula paludigena* involved in ergosterol biosynthetis	70.3%
	2.2	Betalactone	1,585,039	1,617,996	6-Coumarate-CoA ligase of *Rhodotorula toruloides* NP11	68.3%
Contig 3	3.1	NRPS-like	438,470	483,419	L-Aminoadipate-semialdehyde dehydrogenase of *Sporobolomyces salmonicolor*	81.3%
Contig 4	4.1	Terpene	75,068	102,257	Squalene cyclase (SQCY) domain gene of *Sporobolomyces salmonicolor*	66.0%
	4.2	Terpene	572,343	593,838	Lycopene cyclase/phytoene synthase of *Sporobolomyces pararoseus*	71.3%

RJAF-17 was observed to produce surfactant molecules namely N-palmitoyl leucine and N-parmitoleyl glutamine, which are categorized as lipoamino acids (patent application number 10-2023-0149087). These lipoamino acids have been well established for their surfactant property [7, 8]. Formed by the association of a polar amino acid and a non-polar long-chain compound, these molecules have high surface activity resulting in surfactant characteristics. In bacteria, this acylation reaction is reported to be carried out by two enzymes, an N-acetyltransferase and an O-acetyltransferase of which the former catalyzes the initial conjugation of the amino acid to a beta hydroxy fatty acid followed by conjugation of a second fatty acid to the lysolipid by the later [9]. In this study, we found five genes for N-acetyltransferase in the annotated genome of strain RJAF-17 but not for O-acetyltransferases. As *Sporobolomyces* spp. have been reported for their ability to reduce animal wastes and biosurfactant activity [10], further experimental and genomic analyses can help identify biosynthetic pathway for the lipoamino acids.

Acknowledgment

This study (Project No. 015675) was carried out with the support of National Institute of Agricultural Sciences, Rural Development Administration, Republic of Korea.

Conflict of Interest

The authors have no financial conflicts of interest to declare.

References

1. Ernst C, Rahmann S. 2013. PanCake: A data structure for pangenomes. *German Conference on Bioinformatics* **34**. DOI: 10.4230/OASIcs.GCB.2013.35.
2. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**: 737-746.
3. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: novel and streamlined workflows along with broader and deeeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**: 4647-4654.
4. 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.
5. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**: 5825-5829.
6. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, *et al.* 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **39**: W316-W322.
7. Gerova M, Rodrigues F, Lamère JF, Dobrev A, Fery-Forgues S. 2008. Self-assembly properties of some chiral *N*-palmitoyl amino acid surfactants in aqueous solution. *J. Colloid Interface Sci.* **319**: 526-533.
8. Ribeiro MHL, Carvalho P, Martins TS, Faustino CMC. 2019. Lipoaminoacids enzyme-based production and application as gene delivery vectors. *Catalysts* **9**: 977.
9. Stirrup R, Mausz MA, Silvano E, Murphy A, Guillonneau R, Quareshy M, *et al.* 2023. Aminolipids elicit functional trade-offs between competitiveness and bacteriophage attachment in *Ruegeria pomeroyi*. *ISME J.* **17**: 315-325.
10. Szotkowski M, Byrtusova D, Haronikova A, Vysoka M, Rapta M, Shapaval V, *et al.* 2019. Study of metabolic adaptation of red yeasts to waste animal fat substrate. *Microorganisms* **7**: 578.