Microbiol. Biotechnol. Lett. (2023), 51(3), 328–331
http://dx.doi.org/10.48022/mbl.2307.07002
pISSN 1598-642X eISSN 2234-7305

MBL
Microbiology and Biotechnology Letters

Genome Reports

Draft Genome Sequence of the *Neodothiora populina*-Like Yeast Strain JAF-11, Which Produces the Biosurfactant *myo*-Inositol Lipids

Jeong-Seon Kim[1†], Parthiban Subramanian[2†], Seunghwan Kim[1], Jun Heo[1], Bong-Sik Yun[3], and Yiseul Kim[1*]

[1]*Agricultural Microbiology Division, National Institute of Agricultural Sciences, Rural Development Administration, Wanju 55365, Republic of Korea*
[2]*National Agrobiodiversity Center, National Institute of Agricultural Sciences, Rural Development Administration, Wanju 55365, Republic of Korea*
[3]*Department of Biotechnology, College of Environmental and Bioresource Sciences, Jeonbuk National University, Wanju 54596, Republic of Korea*

Received: July 12, 2023 / Revised: September 20, 2023 / Accepted: September 21, 2023

Genomic information of biotechnologically and industrially important microorganisms provides the basis for understanding their metabolic potential. Here, we report the draft genome sequence of the *Neodothiora populina*-like yeast strain JAF-11 capable of producing biosurfactant *myo*-inositol lipids. The draft genome contained genes associated with secondary metabolite biosynthesis, including transport and metabolism of lipids, which are a major component of fungal surfactants. Identification of *myo*-inositol and acyl chain synthesis genes in the draft genome corresponded to the specific biosurfactant produced by strain JAF-11. Further experimental studies could help to elucidate the genes responsible for the production of biosurfactant by strain JAF-11.

Keywords: **Yeast, *Neodothiora*, biosurfactant, genome**

With the growing interest in the development of microbial surfactants, strain JAF-11 was isolated and screened for its biosurfactant producing ability [1]. Initial analysis of the D1/D2 domain of the large subunit (LSU) rRNA gene and the internal transcribed spacer (ITS) region of strain JAF-11 exhibited similarities less than the thresholds, representing a novel yeast species most closely related to *Neodothiora populina* CPC 39399^T in the family *Dothioraceae*. This strain JAF-11 was shown to produce a novel biosurfactant of *myo*-inositol lipid type, which is known as a common and essential polyol found in yeast cells. This compound has a similar chemical structure to a previously reported yeast-derived surfactant, pullusurfactan E [2]. In order to gain a better understanding of the biosurfactant-producing yeast, strain JAF-11 was subject to genome sequencing. Here, we report the genomic information of the *Neodothiora populina*-like yeast strain JAF-11, isolated from flower of *Prunus mume* Sieb. *et* Zucc. in the Republic of Korea.

The genome of strain JAF-11 was sequenced using a combination of Oxford Nanopore MinION (Oxford Nanopore Technologies, UK) and Illumina HiSeq X-ten (Illumina, USA) platforms by Seeders Inc. (Republic of Korea). *De novo* genome assembly based on high-quality reads of Nanopore and Illumina data was constructed

***Corresponding author**
Phone: +82-63-238-3028, Fax: +82-63-238-3845
E-mail: dew@korea.kr
†These authors contributed equally to this work and are listed alphabetically.

using HASLR 0.8a1 [3]. Unless otherwise specified, all further analyses were carried out on the Galaxy Web server (https://usegalaxy.org). Completeness assessment of genome assembly was examined using BUSCO 4.1.4 [4]. Repetitive elements were investigated using RepeatMasker 4.1.5 followed by gene prediction using AUGUSTUS 3.4.0 with *Neurospora crassa* as a model for training [5]. Subsequently, the functional annotation of the data from AUGUSTUS was carried out using eggNOG-mapper 2 [6]. During mapping, the query genome was screened for Clusters of Orthologous Genes (COGs), Gene Ontology (GO terms), Carbohydrate-Active enZYmes (CAZy), and Pfam. The genome was also queried at the main pathway databases, including KEGG and PANTHER using KOBAS 2.0 to study functional metabolism genes [7]. Secondary metabolite production by strain JAF-11 was analyzed using fungal version of antiSMASH 6.1.1 (https://fungismash.secondarymetabolites.org/#!/start). The genome of strain JAF-11 has been deposited in GenBank with the accession number of JASVWA000000000.

The genome assembly was 27.2 Mbp in size with 14 scaffolds, N50 value of 2.2 Mbp, and a GC content of 50.8%. The largest scaffold was 2.61 Mbp and the shortest was 0.53 Mbp long. Completeness of the genome assembly was 94.3%, showing the following profile C:94.3% [S: 94.0%, D: 0.3%], F: 1.9%, M: 3.8%, n: 3786 when dothideomycetes_odb10 dataset was used as reference. Analysis of repetitive elements exhibited very few repeat elements in the genome (0.72%). Gene prediction using AUGUSTUS indicated the presence of 10,883 genes, of which 7,398 were selected by eggNOG-mapper for scanning. For quality assessment of gene prediction, annotations of strain JAF-11 based on AUGUSTUS and Maker were compared using the genome annotation statistics tool available at the Galaxy server [8]. A comparison of the annotations using the different tools are provided in Table 1.

Functional annotation revealed about 24% of genes (1,762 genes) having unknown function. Genes with known functions were mainly associated with secondary metabolites biosynthesis, including transport and catabolism (369 genes), translation, ribosomal structure, and biogenesis (336 genes), signal transduction mechanisms (250 genes), and lipid transport and metabolism (236 genes) (Fig. 1). Lipids are known as a major component of fungal surfactants [9] and observation of a large number of genes belonging to the COG category of lipid transport and metabolism is in line with the biosurfactant producing ability of strain JAF-11. KEGG analysis showed presence of metabolic pathways of inositol phosphate as well as sphingolipid. We found five contigs with high similarity to genes coding for inositol monophosphatase family proteins, which convert 1D-*myo*-inositol

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

Attribute	AUGUSTUS 3.4.0	Maker 2.31.11
Number of genes predicted	10,883	8,997
Number of transcripts predicted	10,883	8,997
Mean gene locus size (first to last exon)	1,752.7	1708.6
Mean exon size	304.6	507.6
Mean number of distinct exons per gene	3.02	2.8
Number of distinct exons	32,906	25,263
Number of single-exon genes	2,252 (20.7%)	2,196 (24.4%)
Number of multi-exon genes	8,330 (76.5%)	6,801 (75.6%)
Number of selected queries by EggNOG-mapper	7,398 (68%)	7,708 (85%)
Pfam hits*	6,674	6,946
Pfam domains	10,484	10,830
GO hits*	3,618	3,774
EC hits*	1,813	1,855
CAZy hits*	141	149

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

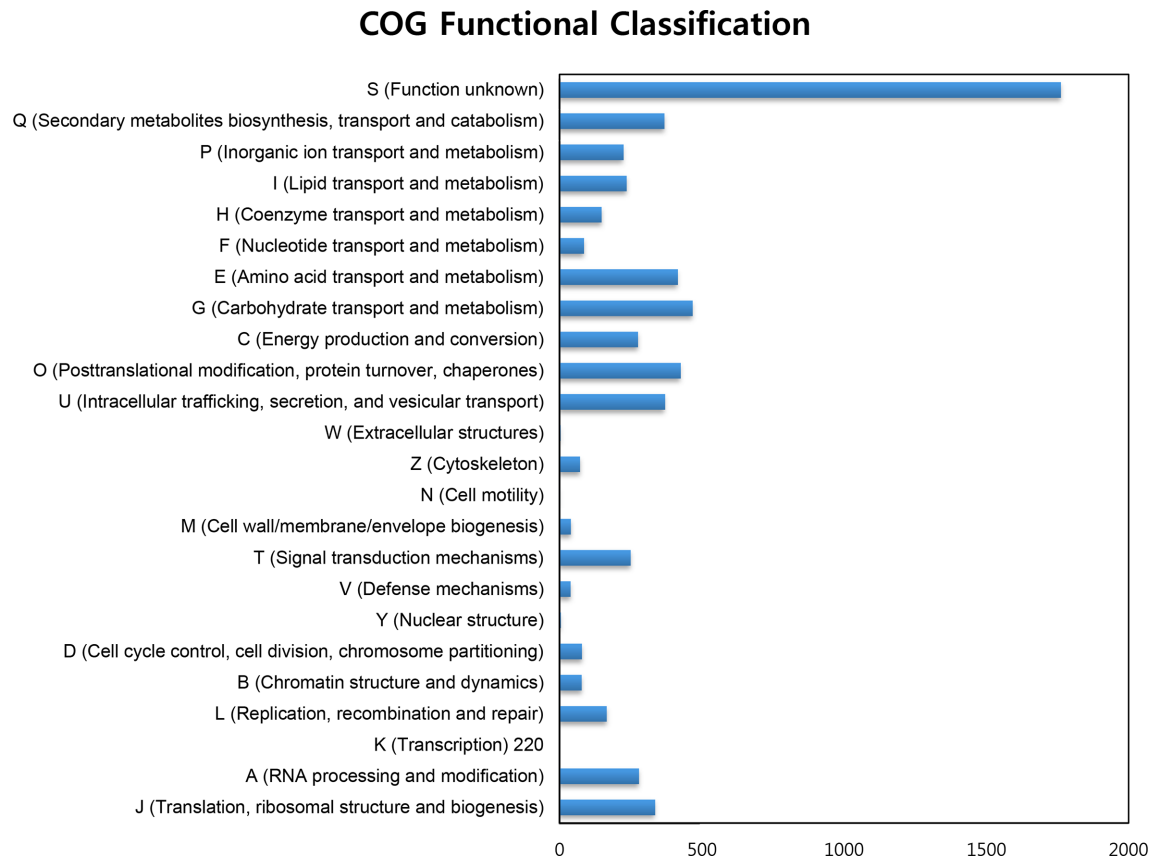

Fig. 1. Functional COG classification of the genome of the *Neodothiora populina*-like yeast strain JAF-11.

3-phosphate to *myo*-inositol. In addition, a contig was found exhibiting high similarity to genes coding for *Saccharomyces cerevisiae* inositol phospholipid synthesis and fat storage-inducing transmembrane protein Scs3p [10].

Furthermore, Pfam search led to identification of 22 Glycosyl transferase family genes, which help in elongation of complex oligosaccharides and putative acyltransferase family genes. Three of these acyltransferases also showed similarity to ketoacyl-synt domain of β-ketoacyl synthases, previously reported to be involved in fatty acid synthesis [11]. Identification of three acyl chains in the structure of the biosurfactant molecule [1] and our observation of these acyltransferases could be hypothesized to have played a role in formation of the acyl moieties. Investigation of the effect of acyltransferases activity on the incorporation of the acyl moieties may yield further insights into their metabolic potential.

Lastly, although there is currently no threshold of average nucleotide identity (ANI) used for yeast species demarcation, calculation of ANI was performed between strain JAF-11 and *Delphinella strobiligena* CBS 735.71^{T} (78.0%), which is the only genome available from the phylogenetic tree [1] and from the family *Dothioraceae*. Beside the phylogenetic assessment, additional analysis including phenotypic characterization could help provide genomic insights for accurate species identification.

Acknowledgment

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

References

1. Kim JS, Lee M, Ki DW, Kwon SW, Ko YJ, Kim JS, *et al.* 2023. Production of a new biosurfactant by a new yeast species isolated from *Prunus mume* Sieb. *et* Zucc. *J. Microbiol. Biotechnol.* **33**: 1023-1029.
2. Kim JS, Lee IK, Yun BS. 2018. Pullusurfactans A-E, new biosurfac-

tants produced by *Aureobasidium pullulans* A11211-4-57 from a fleabane, *Erigeron annus* (L.) pers. *J. Antibiot (Tokyo)* **71**: 920-926.
3. Haghshenas E, Asghari H, Stoye J, Chauve C, Hach F. 2020. HASLR: Fast hybrid assembly long reads. *iScience* **23**: 101389.
4. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**: 4647-4654.
5. 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.
6. 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.
7. 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.
8. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, *et al.* 2008. Maker: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**:188-96.
9. da Silva AF, Banat IM, Giachini AJ, Robl D. 2021. Fungal biosurfactants, from nature to biotechnological product: bioprospection, production and potential applications. *Bioprocess Biosyst. Eng.* **44**: 2003-2034.
10. Hosaka K, Nikawa J, Kodaki T, Ishizu H, Yamashita S. 1994. Cloning and sequence of the SCS3 gene which is required for inositol prototrophy in *Saccharomyces cerevisiae. J. Biochem.* **116**: 1317-1321.
11. Witkowski A, Joshi AK, Smith S. 2002. Mechanism of the β-Ketoacyl synthase reaction catalyzed by the animal fatty acid synthase. *Biochemistry* **41**: 10877-10887.