Microbiol. Biotechnol. Lett. (2023), 51(4), 531–534
http://dx.doi.org/10.48022/mbl.2310.10004
pISSN 1598-642X eISSN 2234-7305

MBL
Microbiology and Biotechnology Letters

Genome Reports

Draft Genome Sequence of the Yeast Strain *Hormonema macrosporum* POB-4, which Produces the Biosurfactant Glycocholic Acid

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: October 5, 2023 / Accepted: November 7, 2023

We report the draft genome sequence of the yeast strain *Hormonema macrosporum* POB-4, capable of producing the biosurfactant glycocholic acid, one of the bile acids. A majority of genes with known function were associated with metabolism and transport of amino acid and carbohydrate as well as secondary metabolites biosynthesis, transport, and catabolism. We observed genes of eleven C-N hydrolases and two CoA transferases which have been reported to be involved in the biosynthesis of glycocholic acid. Further experimental studies can help to elucidate the specific genes responsible for biosurfactant production in strain POB-4.

Keywords: **Yeast, *Hormonema macrosporum*, biosurfactant, genome**

With an aim of supporting the sustainable development, there has been a growing interest in microorganisms with industrial potential to reduce overuse of environment-polluting synthetic materials. We report here, the genomic information of biosurfactant, glycocholic acid producing yeast *Hormonema macrosporum* POB-4, isolated from flower pollen in the Republic of Korea. In order to study the metabolic potential of strain POB-4, it was subject to genome sequencing. The genome of strain POB-4 was sequenced using a combination of HiFi sequencing (Sequel II System) and Illumina HiSeq X-ten (Illumina, USA) platforms by Macrogen Inc., (Republic of Korea). Genome assembly application from PacBio was used to generate high quality *de novo* assemblies using HiFi reads. Briefly, PanCake 1.1.2 [1] was used to overlap the reads followed by Nighthawk to phase the overlapped reads. After the removal of chimeras and duplicates from the overlapped reads, a string graph was constructed generating primary contigs as well as haplotigs. The primary contigs and haplotigs were polished using Racon 1.5.0 [2] with phased reads. To remove haplotype duplications from the primary contig set, purge_dups was used for retrieving potential haplotype duplications and move them to the haplotig set. The generated genome assembly was subject to further analyses.

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 [3] with the lineage dataset dothideomycetes_odb10. Repetitive elements were studied using RepeatMasker 4.1.5 followed by gene prediction using AUGUSTUS software with *Neurospora*

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

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

ATTRIBUTE	AUGUSTUS	MAKER
Contigs	16	16
Number of genes predicted	8,640	9,854
Number of transcripts predicted	8,640	9,854
Complete BUSCOs	3,352	3,261
Missing BUSCOs	291	130
Number of selected queries by EggNOG-mapper	7,133 (82.6%)	8,455 (85.8%)
Single EggNOG	6,758	6,791
Multi EggNOG	375	929
Pfam hits*	6,461	7,613
GO hits*	3,540	4,021
EC hits*	1,753	2,012
CAZy hits*	137	162

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

crassa as a model for training [4]. For quality assessment of gene prediction, annotations of strain POB-4 using the different tools, AUGUSTUS and Maker, were compared via genome annotation statistics tool available at the Galaxy Web server (Table 1). Subsequently, the functional annotation of the data from AUGUSTUS and MAKER was carried out using eggNOG-mapper 2 [5]. 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 of genes [6]. Secondary metabolite production by strain POB-4 was analyzed using fungal version of antiSMASH 7.0 [7] (https://fungismash.secondarymetabolites.org/#!/start). The raw sequencing data of strain POB-4 was deposited in GenBank with the accession number of SRX21170146 and SRX21170147.

The genome assembly was 28.4 Mb (28,419,067 bp) in size and consisted of 16 scaffolds with N50 value of 2.2 Mb (2,213,373 bp). The largest scaffold was 2,872,214 bp long and the shortest was 53,043 bp long. The GC content was estimated to be 49.8%. Completeness of the genome assembly was 97.3%, showing the following profile C: 97.3% [S: 96.9%, D: 0.4%], F: 0.2%, M: 2.5%, n: 3786 when dothideomycetes_odb10 dataset was used as reference. Analysis of repetitive elements exhibited very few repeat elements in the genome (0.92%). Functional annotation using outputs from AUGUSTUS (8,640 predicted genes) and MAKER (9,854 predicted genes) resulted in identification of functional traits of the coding sequences (Table 2). Strain POB-4 contained a majority of genes for metabolism and transport of amino acid as well as carbohydrate followed by secondary metabolite biosynthesis, transport, and catabolism. Analysis using fungal antiSMASH resulted in three hits, namely melanin (100% match, contig 6), neosatorin (52% match, contig 1), and polyketide synthesis (33% match, contig 10).

The compound glycocholic acid, identified as the major component of the biosurfactant by strain POB-4 during HPLC and NMR analyses, is one of the bile acids (patent application number 10-2023-0147900). Microbial production of bile acid conjugates such as glycocholic acid has been documented previously. Two studies have shown the production of glycocholic acid by fungus *Penicillum* sp., which belongs to subdivision *Pezizomycotina* of *Ascomycota* same as strain POB-4 [8, 9]. Moreover, bacterial strains with marine origin as well as gut bacteria have been reported to produce bile acid conjugates [10–12]. In particular, Garcia *et al.* proposed amino acid N-acyltransferases as a mechanism for the production of microbially conjugated bile acids such as glycocholic acid [12]. In the KEGG database, glycocholic acid (compound C01921) leads to 1,027 reported genes, which are orthologues of two genes, namely bile acid-CoA: amino acid N-acyltransferase (K00659) and choloylglycine hydrolase (EC 3.5.1.24) or linear amide C-N hydrolase (K01442). Although genes associated with neither bile acid-CoA: amino acid N-acyltransferases nor choloylglycine hydrolase were detected, we identified eleven genes for C-N hydrolase as well as two genes for acyl-CoA thioesterase (EC 3.1.2.2). Future addition of genomic information on various yeast species with similar metabolism might help identification of bile acid pathway in strain POB-4.

http://dx.doi.org/10.48022/mbl.2310.10004

Table 2. Metabolism related genes of the yeast strain *Hormonema macrosporum* POB-4.

COG CATAGORIES	AUGUSTUS	MAKER
INFORMATION STORAGE AND PROCESSING	**1,030**	**1,208**
[J] Translation, ribosomal structure and biogenesis	319	355
[A] RNA processing and modification	272	302
[K] Transcription	206	275
[L] Replication, recombination and repair	155	184
[B] Chromatin structure and dynamics	78	92
CELLULAR PROCESSES AND SIGNALING	**1,282**	**988**
[D] Cell cycle control, cell division, chromosome partitioning	82	90
[Y] Nuclear structure	5	5
[V] Defense mechanisms	47	59
[T] Signal transduction mechanisms	234	277
[M] Cell wall/membrane/envelope biogenesis	47	54
[N] Cell motility	3	3
[Z] Cytoskeleton	75	85
[W] Extracellular structures	4	5
[U] Intracellular trafficking, secretion, and vesicular transport	353	410
[O] Posttranslational modification, protein turnover, chaperones	432	0
METABOLISM	**2,168**	**2,491**
[C] Energy production and conversion	283	314
[G] Carbohydrate transport and metabolism	432	511
[E] Amino acid transport and metabolism	398	458
[F] Nucleotide transport and metabolism	78	88
[H] Coenzyme transport and metabolism	141	165
[I] Lipid transport and metabolism	242	280
[P] Inorganic ion transport and metabolism	205	229
[Q] Secondary metabolites biosynthesis, transport and catabolism	389	446
POORLY CHARACTERIZED	**1,701**	**2,104**
[R] General function prediction only	0	0
[S] Function unknown	1,701	2,104
No hits	577	735

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. Ernst C, Rahmann S. 2013. PanCake: A Data Structure for Pangenomes. German Conference on Bioinformatics.
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 deeper 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. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, *et al*. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res*. **51**: W46-W50.
8. Pil GB, Won HS, Shin HJ. 2016. Bile acids from a marine sponge-associated fungus *Penicillium* sp. *J. Korean Magnetic Reson. Soc.* **20**: 41-45.
9. Ohashi K, Miyagawa Y, Nakamura Y, Shibuya H. 2008. Bioproduction of bile acids and the glycine conjugates by *Penicillium fungus. J. Nat. Med.* **62**: 83-86.
10. Maneerat S, Nitoda T, Kanzaki H, Kawai F. 2005. Bile acids are new products of a marine bacterium, *Myroides* sp. strain SM1. *Appl. Microbiol. Biotechnol.* **67**: 679-683.
11. Kim D, Lee JS, Kim J, Kang SJ, Yoon JH, Kim WG, *et al.* 2007. Biosynthesis of bile acids in a variety of marine bacterial taxa. *J. Microbiol. Biotechnol.* **17**: 403-407.
12. Garcia CJ, Kosek V, Beltrán D, Tomás-Barberán FA, Hajslova J. 2022. Production of new microbially conjugated bile acids by human gut microbiota. *Biomolecules* **12**: 687.

http://dx.doi.org/10.48022/mbl.2310.10004