Korean Journal of Microbiology (2022) Vol. 58, No. 2, pp. 99-101
DOI https://doi.org/10.7845/kjm.2022.2023

pISSN 0440-2413
eISSN 2383-9902

Draft genome sequence of *Rhizopus delemar* SSU VMBB-02 isolated from alcohol fermentation starter culture

Jae Yun Lim and Jeong-Ah Seo*

School of Systems Biomedical Science, Soongsil University, Seoul 06978, Republic of Korea

알코올 발효 개시제로부터 분리한 *Rhizopus delemar* SSU VMBB-02 균주의 유전체 염기서열

임재윤 · 서정아*

숭실대학교 의생명시스템학부

(Received April 7, 2022; Revised June 21, 2022; Accepted June 22, 2022)

Filamentous fungi and yeasts were isolated from *banh men*, a fermentation starter culture for the production of traditional liquor in Vietnam. The genus *Rhizopus* was a major group of fungi, including *Rhizopus oryzae*, *Rhizopus delemar*, and *Rhizopus microsporus* isolated from *banh men*. Among twenty-five *Rhizopus* strains isolated, *R. delemar* SSU VMBB-02 (KCTC 46675) strain that highly produces hydrolytic enzymes, was selected for constructing the draft genome. The whole genome was sequenced, assembled and analyzed using Illumina HiSeq platform technology. As a result of genome analysis, the draft genome was assembled with the size of 37.9 Mb containing 34.4% G + C content and 12,076 protein-coding genes. Through additional analyses, the number of genes encoding carbohydrate-active enzymes and proteolytic enzymes, and the number of secondary metabolite biosynthesis gene clusters were revealed. The results of this study will provide useful genomic information that can be compared with other *Rhizopus* species originated from fermentation starter culture.

Keywords: *Rhizopus delemar*, draft genome sequence, fermentation starter culture, next generation sequencing technology

Rhizopus oryzae is a generally known as safe (GRAS) filamentous fungus, commonly associated with production of some oriental traditional foods (Londoño-Hernández *et al*., 2017). *Rhizopus delemar* has been differentiated as a species of *R. oryzae* producing fumaric-malic acid (Abe *et al*., 2007). In this study, we isolated *R. delemar* SSU VMBB-02 from *banh men* made in Daklak province, Vietnam. The *R. delemar* SSU VMBB-02 strain was grown on potato dextrose agar solid medium for 5 days at 25°C and its hyphal mass was collected. The genomic DNA was extracted and purified by using the modified cetyl trimethylammonium bromide method from freeze-dired hyphae (Leslie and Summerell, 2006). The whole genome sequencing was performed by Theragen Bio Institute. The library for sequencing was made by using a TruSeq DNA PCR-free library preparation kit according to the manufacturer's instructions and sequenced on the Illumina HiSeq 2500 platform (Illumina). As a result of a sequencing, total reads of 17,026,137,418 bp were obtained, and the coverage was 447-fold. The sequence reads were assembled by *de novo* by SPAdes assembler (Bankevich *et al*., 2012), which resulted in 5,705 contigs and the whole genome size of 37.9 Mb (*N*50, 16 kb). The G + C content of the assembled draft genome was 34.4%.

Draft genome annotation was performed by using Funannotate pipeline v1.8.9 (Palmer and Stajich, 2017). First, 11.68% of repetitive sequences were masked by RepeatMasker (open-4.0.7) and RepeatModeler (open-1.0.11) (Tarailo-Graovac and Chen,

***For correspondence.** E-mail: sja815@ssu.ac.kr;
Tel.: +82-2-820-0449; Fax: +82-2-824-4383

Table 1. Draft genome features of *Rhizopus delemar* SSU VMBB-02

Features	Value
Draft genome size, bp	37,902,973
GC content, %	34.4
Number of contigs	5,705
Number of contigs ≥ 2 kb	3,067
Contig N_{50}, bp	16,092
Protein coding genes	12,076
Number of genes having InterPro domains	9,149
Coverage of InterPro, %	76
Number of gene ontology assigned	6,709
Number of genes involved in CAZymes	325
Number of protease genes by MEROPS	370
Number of secondary metabolite gene clusters	14

2004). The *ab initio* gene models for the contigs were predicted by using the GeneMark-ES (v4.38) (Ter-Hovhannisyan *et al*., 2008), GlimmerHMM (v3.0.4) (Majoros *et al*., 2004) and AUGUSTUS (v3.3.3) (Keller *et al*., 2011) programs. Evidence-based gene models were made by aligning the contigs with the protein sequence database (UniProtKB) using DIAMOND (v2.0.14) (Buchfink *et al*., 2021) and then polishing using Exonerate (v2.4.0) (Slater and Birney, 2005). EVidenceModeler (v1.1.1) (Haas *et al*., 2008) as implemented in the Funannotate pipeline was used to generate the consensus models from the *ab initio* and evidence-based gene models. The consensus models were functionally annotated after removing short lengths, transposable elements and gaps. A total of 12,076 gene models were used for making 97,106 valid annotations by carrying out sequence similarity searches against the Pfam (v34.0), InterPro (v79.0), BUSCO (v2.0), EggNOG (v4.5), MEROPS (v12.0) and CAZyme (v9.0) databases. We predicted 12,076 protein-coding genes. Among them, 9,149 genes have InterPro domains, of which 6,709 were categorized to Gene Ontology (GO). Further analyses of the genes encoding carbohydrate-active enzymes revealed that 325 genes were involved in CAZymes (21 auxiliary activities, 56 carbohydrate esterases, 7 carbohydrate-binding modules, 118 glycoside hydrolases, 116 glycosyl transferases, and 7 polysaccharide lyases) and 370 genes in proteolytic enzymes (36 aspartic peptidases, 78 cysteine peptidases, 109 metallo peptidases, 7 protease inhibitors, 108 serine peptidases and 32 threonine peptidases) (Table 1). In addition, 549 genes encoding the secreted transcripts were predicted by using the

Table 2. List of 14 secondary metabolite biosynthetic gene clusters

No.	Contig number	Secondary metabolite type	Nucleotide location	
			Start	End
1	ROVContig0012.1	Siderophore	3,838	16,018
2	ROVContig0078.1	Terpene	26,853	41,211
3	ROVContig0123.1	Terpene	1	16,888
4	ROVContig0129.1	Fungal-RiPP	1	34,725
5	ROVContig0291.1	Terpene	11,848	25,530
6	ROVContig0318.1	Terpene	1	16,870
7	ROVContig0413.1	NRPS	1	22,108
8	ROVContig0611.1	Terpene	1	13,711
9	ROVContig0615.1	Terpene	1	14,098
10	ROVContig0813.1	NRPS-like	1	14,269
11	ROVContig0861.1	Terpene	590	13,605
12	ROVContig1115.1	Terpene	1	10,774
13	ROVContig1162.1	NRPS-like	1	10,337
14	ROVContig2258.1	NRPS-like	1	4,049

SignalP secretome prediction program (v4.1) (Armenteros *et al*., 2019). Thirty tRNA genes were predicted by using tRNAscan-SE (v2.0.9) (Lowe and Eddy, 1997). Additionally, fourteen secondary metabolite biosynthesis gene clusters (with 21 biosynthetic enzymes and 15 smCOGs) were found by using antiSMASH 5.0.0 (Blin *et al*., 2019), which may be involved in unknown secondary metabolisms including terpene, sidrophore and non ribosamal peptide synthesis (NRPS) (Table 2).

Nucleotide sequence accession number

The draft genome sequence of *R. delemar* SSU VMBB-02 (KCTC 46675) has been deposited in GenBank under the accession number JAFCNC000000000.

적 요

베트남의 전통주 생산을 위한 발효 개시제인 반멘으로부터 사상성 곰팡이와 효모균을 분리하였다. *Rhizopus* 속은 반멘에서 분리된 *Rhizopus oryzae*, *Rhizopus delemar* 및 *Rhizopus microsporus*를 포함한 주요 곰팡이군이었다. 분리된 *Rhizopus* 25 균주 중 가수분해효소 활성이 높은 *R. delemar* SSU VMBB-02 (KCTC 46675) 균주를 선정하여 전장유전체 염기서열을 Illumina HiSeq 플랫폼 기술을 활용하여 해독, 조립, 분석하였다. 유전체 분석 결과 전체 유전체 크기는 37.9 Mb로 G + C 함

량 34.4%, 생성된 컨티그 5,705개, 단백질 코딩 유전자 12,076개를 포함하고 있었다. 추가적인 분석을 통하여 탄수화물 분해효소와 단백질 분해효소를 코딩한 유전자의 수, 이차 대사물질 생합성 유전자 클러스터 개수 등을 밝혀냈다. 이 연구의 결과는 발효제 유래의 다른 *Rhizopus* 균주들과 비교해 볼 수 있는 유용한 유전체정보를 제공할 것이다.

Acknowledgments

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through Agricultural Microbiome R&D Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA-918010-4).

Conflict of Interest

The authors have no conflict of interest to report.

References

Abe A, Oda Y, Asano K, and Sone T. 2007. *Rhizopus delemar* is the proper name for *Rizopus oryzae* fumaric-malic acid producers. *Mycologia* **99**, 714–722.

Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, and Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol*. **37**, 420–423.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol*. **19**, 455–477.

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, and Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87.

Buchfink B, Reuter K, and Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, and Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol*. **9**, R7.

Keller O, Kollmar M, Stanke M, and Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**, 757–763.

Leslie JF and Summerell BA. 2006. The *Fusarium* Laboratory Manual. Blackwell Publishing, Ames, Iowa, USA.

Londoño-Hernández L, Ramírez-Toro C, Ruiz HA, Ascacio-Valdés JA, Aguilar-Gonzalez MA, Rodríguez-Herrera R, and Aguilar CN. 2017. *Rhizopus oryzae* – ancient microbial resource with importance in modern food industry. *Int. J. Food Microbiol*. **257**, 110–127.

Lowe TM and Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. **25**, 955–964.

Majoros WH, Pertea M, and Salzberg SL. 2004. TigrScan and GlimmerHMM: two open-source *ab initio* eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879.

Palmer JM and Stajich JE. 2017. Funannotate: eukaryotic genome annotation pipeline. Retrieved from https://funannotate.readthedocs.io/en/latest/.

Slater GS and Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31.

Tarailo-Graovac M and Chen N. 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatic* **25**, 4.10.1–4.10.14.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, and Borodovsky M. 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res*. **18**, 1979–1990.