Draft Genome Sequence of *Streptomyces cavourensis* YBQ59, an Endophytic Producer of Antibiotics Bafilomycin D, Nonactic Acid, Prelactone B, and 5,11-Epoxy-10-Cadinanol

Huy Quang Nguyen,[a,b,e] Nguyen Thi-Hanh Vu,[b] Ha Hoang Chu,[b,e] Son Ky Chu,[c] Ha Hoang,[b] Thanh Trung Tran,[b] Cuong Nguyen,[d] Linh Thi-My Dinh,[b] Hien Thi-Thu Trinh,[b] Tien Quyet Phi[b,e]

[a]University of Science and Technology of Hanoi (USTH), Vietnam Academy of Science and Technology (VAST), Hanoi, Vietnam
[b]Institute of Biotechnology (IBT), Vietnam Academy of Science and Technology (VAST), Hanoi, Vietnam
[c]Hanoi University of Science and Technology (HUST), Hanoi, Vietnam
[d]Gene Technology Department, Vinmec Research Institute of Stem Cell and Gene Technology, Hanoi, Vietnam
[e]Graduate University of Science and Technology (GUST), Vietnam Academy of Science and Technology (VAST), Hanoi, Vietnam

ABSTRACT This study reports the draft genome sequence of the endophytic *Streptomyces cavourensis* strain YBQ59, produces the antibiotics bafilomycin D, nonactic acid, prelactone B, and 5,11-epoxy-10-cadinanol. The draft genome sequence comprises ~10.2 Mb, with a GC content of 64% and 8,958 predicted protein-coding genes, of which 14 gene clusters were found to associate with antibiotic biosynthetic pathways.

The endophyte *Streptomyces cavourensis* strain YBQ59 was isolated from the root of the medicinal plant *Cinnamomum cassia* Presl in Vietnam (1). Eight pure secondary metabolites were purified from culture broth of the strain YBQ59, among which the five compounds 1-monolinolein, bafilomycin D, nonactic acid, daidzein, and 3'-hydroxydaidzein showed antimicrobial and cytotoxic activities against multidrug-resistant bacteria and human cancer cell lines, respectively (1).

The strain *S. cavourensis* YBQ59 was grown on YIM38 medium (1) for 48 h, and cells were harvested by centrifugation at 10,000 rpm for 5 min. Genomic DNA was extracted by using the PureLink genomic DNA kit (Invitrogen, CA, USA), and the purity was determined with a NanoDrop spectrophotometer (ThermoFisher Scientific, Inc., Waltham, MA, USA). The genomic DNA of the strain YBQ59 was sequenced by using Ion Torrent PGM technology with the Ion PGM template OT2 200 and sequencing 200 kits (ThermoFisher Scientific, Inc.). The obtained raw data consisted of 512,914,080 bp, comprising 2,856,000 reads (25 to 373 bp), in which the quality was controlled by using Trimmomatic (with parameters Slidingwindow, 10:30; Crop, 215; Minlen, 50) (2). The *de novo* genome assembly was performed by using VelvetOptimiser with default parameters (3), which revealed the genome size to be 10,232,294 bp, comprising 4,428 contigs (N_{50}, 12,307 bp; longest contig length, 161,472 bp) with coverage of 26× and an average GC content of 64%. Gene prediction and annotation were processed by combining three programs, Prodigal (4), GeneMarkS (5), and NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 4.5 (6) with default parameters, which allowed for the identification of 8,958 overlapped predicted protein-coding sequences, 76 tRNA genes, 4 complete rRNA operons, and 3,145 pseudogenes. Blast2GO PRO (7) was used for the functional annotation of the whole genome based on Gene Ontology (GO) databases, which showed 6,292, 3,083, and 5,961 genes assigned to known biological processes, synthesis of cellular components, and molecular functions (GO categories),

respectively. Analysis of the genome annotation based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (8) revealed at least 140 distinct metabolic pathways. Notably, 717 enzymes linked to secondary metabolic biosynthesis were predicted, among which 158 enzymes are directly involved in antibiotic biosynthetic pathways, such as macrolides; ketolides; 12-, 14-, and 16-membered macrolides; non-ribosomal peptide structures; and polyketide type I and type II structures. The biosynthetic pathways of the specific antibiotics cephamycin C, nocardicin A, clavaminate, erythromycin A/B, oleandomycin, picromycin, methymycin, neomethymycin, avermectin, tetracycline, oxytetracycline, chlortetracycline, mithramycin, tetracenomycin C, alamycin, rebeccamycin, vancomycin, and validamycin A were predicted based on the KEGG database (8), supporting the broad-spectrum antimicrobial activity of the strain YBQ59.

Furthermore, analysis using antiSMASH v3.0 (9) with default parameters revealed that the *S. cavourensis* YBQ59 genome contains 37 putative biosynthetic gene clusters (BGCs) involved in 14 different biosynthetic pathways, including the following: nonribosomal peptide synthetase (NRPS), type II polyketide synthases (PKS-II), hybrid PKS-NRPS, type III polyketide synthase (PKS-III), beta-ketoacyl synthase, and other BGCs for producing bacteriocin, bleomycin, calyculin, coelimycin, colonic acid, chlorizidine A, desferrioxamine B, ectoine, fatty acids, macrotetrolides, naringenin, landepoxcin, terpenes, and svaricin. Of those antibiotics, macroterrolides exhibiting the broad spectrum of antifungal, antibacterial, and anticancer properties have been found in *Streptomyces griseus* DSM 40695 (10). The analyzed genomic data provide valuable insights into the genetic determinants associated with antimicrobial and cytotoxic properties of *S. cavourensis* YBQ59 (1).

Data availability. The draft genome sequence of the strain *Streptomyces cavourensis* YBQ59 has been deposited at DDBJ/ENA/GenBank under the accession number QLNH00000000. The version described in this paper is the first version, QLNH01000000.

REFERENCES

1. Vu H-NT, Nguyen DT, Nguyen HQ, Chu HH, Chu SK, Van Chau M, Phi Q-T. 2018. Antimicrobial and cytotoxic properties of bioactive metabolites produced by *Streptomyces cavourensis* YBQ59 isolated from *Cinnamomum cassia* Prels in Yen Bai Province of Vietnam. Curr Microbiol https://doi.org/10.1007/s00284-018-1517-x.
2. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
3. Gladman S, Seemann T. 2012. VelvetOptimiser 2.2.5. https://github.com/tseemann/VelvetOptimiser.
4. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
5. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29:2607–2618. https://doi.org/10.1093/nar/29.12.2607.
6. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
7. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. https://doi.org/10.1093/bioinformatics/bti610.
8. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27.
9. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:W237–W243. https://doi.org/10.1093/nar/gkv437.
10. Shen B, Kwon H-J. 2002. Macrotetrolide biosynthesis: a novel type II polyketide synthase. Chem Rec 2:389–396. https://doi.org/10.1002/tcr.10042.