



# Draft Genome Sequence of *Amphirosellinia nigrospora* JS-1675, an Endophytic Fungus from *Pteris cretica*

Jongbum Jeon,<sup>a</sup> Sook-Young Park,<sup>b</sup> Jung A. Kim,<sup>c</sup> Nan Hee Yu,<sup>d</sup> Ae Ran Park,<sup>d</sup> Jin-Cheol Kim,<sup>d</sup> Yong-Hwan Lee,<sup>a</sup> Soonok Kim<sup>c</sup>

<sup>a</sup>Department of Agricultural Biotechnology, Interdisciplinary Program in Agricultural Genomics, Center for Fungal Genetic Resources, Seoul National University, Seoul, South Korea

<sup>b</sup>Department of Plant Medicine, Suncheon National University, Suncheon, South Korea

<sup>c</sup>Genetic Resources Assessment Division, National Institute of Biological Resources, Incheon, South Korea

<sup>d</sup>Department of Agricultural Chemistry, Institute of Environmentally Friendly Agriculture, Chonnam National University, Gwangju, South Korea

**ABSTRACT** The fungus *Amphirosellinia nigrospora* strain JS-1675 has been reported to exert antimicrobial effects against various plant-pathogenic bacteria and fungi. Here, we report the draft genome sequence of *A. nigrospora* for the first time. The assembly comprises 48,177,783 bp with 18 scaffolds.

**A** *Amphirosellinia nigrospora* (phylum Ascomycota, class Sordariomycetes, family Xylariaceae) strain JS-1675 is an endophytic fungus isolated from the Cretan brake fern (*Pteris cretica*). It has been reported to produce oxygenated cyclohexanone, a potent antimicrobial compound that is particularly effective against the bacterial wilt pathogen of tomato, *Ralstonia solanacearum*, as well as against various agronomically significant phytopathogenic fungi (1). It is generally accepted that the polyketide synthase (PKS), nonribosomal protein synthetase (NRPS), and terpene synthase (TS) genes are involved in the production of such secondary metabolites. The genome sequence of *A. nigrospora* strain JS-1675 will enable identification of genes involved in the production of oxygenated cyclohexanone.

Genomic DNA was extracted from fresh mycelia grown in potato dextrose broth at 25°C with shaking at 120 rpm for 2 days, using a DNeasy minikit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Library preparation and genome sequencing were performed at Theragen Etx Bio Institute (Suwon, South Korea). A single-molecule real-time (SMRT) sequencing library was prepared according to the PacBio standard library preparation protocol with a fragment size of 20 kb. A total of 443,875 reads with an average size of 10.2 kb, and therefore 4.57 Gb of sequences, were generated on 4 cells of a PacBio RS II sequencer. These reads were assembled into 72 contigs with a total length of 48,704,170 bp and an  $N_{50}$  value of 4,754,034 bp using the overlap-layout-consensus (OLC) algorithm (2). Additionally, two short-read sequencing data sets were generated using 101 cycles of paired-end sequencing on an Illumina HiSeq 2000 platform from one short paired-end (PE) library (average insert size, 350 bp) and one long mate pair (MP) library (average insert size, 10 kb). A total of 45,165,534 reads with a Q30 of 79.38% were generated from the PE library. A total of 10,307,236 high-quality reads out of 75,404,568 reads were obtained from the MP library after filtering with NextClip v1.3 (3). These short reads were assembled using SOAPdenovo v2.04 to generate 1,182 contigs with a total length of 47,310,961 bp and an  $N_{50}$  value of 381,291 bp. The two assemblies were merged with HaploMerger2 (4) to 31 contigs, and scaffolding and gap filling were conducted sequentially using SSPACE-Standard v3.0 (5), SSPACE-LongRead v1.1 (6), and GapFiller v1.10 (7). Finally, from the 16.75 Gb (347.67× genome size) of genome sequences, the genome reads containing the nuclear genome were assembled into 18 scaffolds (31 contigs) with a total length of 48,177,783 bp and an

**Citation** Jeon J, Park S-Y, Kim JA, Yu NH, Park AR, Kim J-C, Lee Y-H, Kim S. 2019. Draft genome sequence of *Amphirosellinia nigrospora* JS-1675, an endophytic fungus from *Pteris cretica*. Microbiol Resour Announc 8:e00069-19. <https://doi.org/10.1128/MRA.00069-19>.

**Editor** Antonis Rokas, Vanderbilt University

**Copyright** © 2019 Jeon et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yong-Hwan Lee, [yonglee@snu.ac.kr](mailto:yonglee@snu.ac.kr), or Soonok Kim, [sokim90@korea.kr](mailto:sokim90@korea.kr).

J.J., S.-Y.P., and J.A.K. contributed equally to this work.

**Received** 5 February 2019

**Accepted** 22 April 2019

**Published** 16 May 2019

$N_{50}$  value of 11.78 Mb. The genome assembly was validated using BUSCO v. 3.0.2b, which showed an 80.3% hit against 290 fungal genes (8), and Core Eukaryotic Genes Mapping Approach (CEGMA) v.2.5, which showed a 93.95% hit against 248 core eukaryotic genes (9). The GC content of the assembled genome was 43.85%.

A total of 10,232 protein-coding genes were identified after gene prediction and annotation based on AUGUSTUS v.3.2.1 (10). After combining BLAST search results against UniProt or NCBI nonredundant (nr) and InterPro scan data, 9,745 genes were functionally annotated. Among these, 22 PKS and 10 NRPS biosynthetic gene clusters (BGCs) were identified through antiSMASH v4.0 (11, 12). Moreover, we found 307 transcription factor genes by the Fungal Transcription Factor Database (FTFD) v1.0 (13), 147 cytochrome P450 genes by the Fungal Cytochrome P450 Database v1.0 (14), 53 genes encoding plant cell wall-degrading enzymes by the Fungal Plant Cell Wall-Degrading Enzyme Database v1.0 (15), and 10 genes encoding laccases and 29 genes encoding peroxidases using the Fungal Peroxide Database (fPoxDB) v1.0 (16).

This is the first genome sequenced in the genus *Amphirosellinia*. This draft genome sequence will provide valuable information to identify genes for biosynthesis of antimicrobial compounds and will also facilitate comparative genomics with other publicly available fungal species, especially those of the family Xylariaceae, for evolutionary studies.

**Data availability.** The whole-genome sequence of *Amphirosellinia nigrospora* JS-1675 obtained in this study has been deposited at GenBank under the accession no. [SCHM00000000](https://doi.org/10.1111/jam.14138). The version described in this article is the first version, SCHM01000000. SRA data of PacBio sequences and Illumina sequences were also deposited at GenBank under accession no. [SRR8463317](https://doi.org/10.1111/jam.14138), [SRR8463318](https://doi.org/10.1111/jam.14138), and [SRR8463316](https://doi.org/10.1111/jam.14138).

## ACKNOWLEDGMENT

This work was supported by the National Institute of Biological Resources, funded by the Ministry of Environment of the Republic of Korea (projects NIBR201830101 and NIBR201921101).

## REFERENCES

- Nguyen HT, Kim S, Yu NH, Park AR, Yoon H, Bae CH, Yeo JH, Kim IS, Kim JC. 2019. Antimicrobial activities of an oxygenated cyclohexanone derivative isolated from *Amphirosellinia nigrospora* JS-1675 against various plant pathogenic bacteria and fungi. *J Appl Microbiol* 126:894–904. <https://doi.org/10.1111/jam.14138>.
- Wang A, Wang Z, Li Z, Li LM. 2018. BAUM: improving genome assembly by adaptive unique mapping and local overlap-layout-consensus approach. *Bioinformatics* 34:2019–2028. <https://doi.org/10.1093/bioinformatics/bty020>.
- Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M. 2014. NextClip: an analysis and read preparation tool for Nextera long mate pair libraries. *Bioinformatics* 30:566–568. <https://doi.org/10.1093/bioinformatics/btt702>.
- Huang S, Chen Z, Huang G, Yu T, Yang P, Li J, Fu Y, Yuan S, Chen S, Xu A. 2012. HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. *Genome Res* 22:1581–1588. <https://doi.org/10.1101/gr.133652.111>.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
- Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* 15:211. <https://doi.org/10.1186/1471-2105-15-211>.
- Nadalín F, Vezzi F, Policriti A. 2012. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 13(Suppl 14):S8. <https://doi.org/10.1186/1471-2105-13-S14-S8>.
- Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Kliuchnikov G, Kriventseva EV, Zdobnov EM. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35:543–548. <https://doi.org/10.1093/molbev/msx319>.
- Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067. <https://doi.org/10.1093/bioinformatics/btm071>.
- Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a Web server for gene finding in eukaryotes. *Nucleic Acids Res* 32:W309–W312. <https://doi.org/10.1093/nar/gkh379>.
- Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, McMenamin C, Mi H, Mutowo-Mueller P, Mulder N, Natale D, Orengo C, Pesseat S, Punta M, Quinn AF, Rivoire C, Sangrador-Vegas A, Selengut JD, Sigrist CJ, Scheremetjew M, Tate J, Thimmajananthan M, Thomas PD, Wu CH, Yeats C, Yong SY. 2012. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res* 40:D306–D312. <https://doi.org/10.1093/nar/gkr948>.
- Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Muller 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>.
- Park J, Park J, Jang S, Kim S, Kong S, Choi J, Ahn K, Kim J, Lee S, Kim S, Park B, Jung K, Kim S, Kang S, Lee YH. 2008. FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics* 24:1024–1025. <https://doi.org/10.1093/bioinformatics/btn058>.
- Park J, Lee S, Choi J, Ahn K, Park B, Park J, Kang S, Lee YH. 2008. Fungal cytochrome P450 database. *BMC Genomics* 9:402. <https://doi.org/10.1186/1471-2164-9-402>.
- Choi J, Kim KT, Jeon J, Lee YH. 2013. Fungal plant cell wall-degrading enzyme database: a platform for comparative and evolutionary genomics in fungi and Oomycetes. *BMC Genomics* 14(Suppl 5):S7. <https://doi.org/10.1186/1471-2164-14-S5-S7>.
- Choi J, Detry N, Kim KT, Asiegbu FO, Valkonen JP, Lee YH. 2014. fPoxDB: fungal peroxidase database for comparative genomics. *BMC Microbiol* 14:117. <https://doi.org/10.1186/1471-2180-14-117>.