Plant Disease • 2022 • 106:720-722 • https://doi.org/10.1094/PDIS-04-21-0787-A

Genome Resource: Draft Genome of *Fusarium avenaceum*, Strain F156N33, Isolated from the Atmosphere Above Virginia and Annotated Based on RNA Sequencing Data

Shu Yang,[1] Jeffrey J. Coleman,[2] and Boris A. Vinatzer[1,†]

[1] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA 24061
[2] Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849

Abstract

Fusarium avenaceum is a filamentous fungus commonly associated with plants and soil. It is a causal agent of Fusarium head blight (FHB) on maize and small-grain cereals and blights on other plant species, and is one of the very few fungal species known to have ice nucleation activity (i.e., it catalyzes ice formation). Here, we report the draft genome of the ice-nucleation-active *F. avenaceum* strain F156N33 isolated from the atmosphere above Virginia. The genome assembly is 41,175,306 bp long, consists of 214 contigs, and is predicted to encode 11,233 proteins, which were annotated using RNA-sequencing data obtained from the same strain.

Genome Announcement

Fusarium avenaceum (previously also referred to by the teleomorphic name *Gibberella avenacea*) is a common plant pathogen widely distributed in soils with a diverse host range that has also been isolated from the atmosphere, suggesting long-distance aerial dispersal (Lin 2013). This species is one of the causal agents of Fusarium head blight (FHB) on maize (*Zea mays* L.) and small-grain cereals. It also causes blights of other plant species such as pine (*Pinus* spp.) and lisianthus (*Eustoma grandiflorum* L.) (Desjardins 2003; Nalim et al. 2009; Parry et al. 1995). *F. avenaceum* is one of the most important species causing FHB in warmer and temperate regions of the world (Doohan et al. 2003; Parry et al. 1995) and is the dominant species isolated from diseased plants in colder climates such as those in Canada and Northern Europe (Uhlig et al. 2007). Moreover, *F. avenaceum* is known for its ice nucleation activity, catalyzing ice formation at temperatures as high as –5°C (Hasegawa et al. 1994; Kunert et al. 2019; Pouleur et al. 1992; Richard et al. 1996).

Five genome assemblies of *F. avenaceum* strains are publicly available (Fa05001: GCA_000769215.1; FaLH03: GCA_000769305.1; FaLH27: GCA_000769295.1; KA13: GCA_012959155.1; and NRRL 13321: GCA_013753855.1). Four of these genomes do not have accompanying annotation and the one that is annotated, Fa05001, has an annotation that is not based on RNA-sequencing (RNA-seq) data of *F. avenaceum*. Therefore, here, we report the genome assembly and RNA-seq-based annotation of *F. avenaceum* strain F156N33, an ice-nucleation-active strain isolated from an atmospheric sample collected by drone 100 m above the ground at Kentland Farm, a Virginia Tech research farm in Virginia, United States (Kunert et al. 2019). The strain was isolated on *Fusarium*-selective medium and identified as *F. avenaceum* based on the translation elongation factor 1-α gene (Lin 2013).

Genomic DNA of *F. avenaceum* strain F156N33 was extracted from mycelium grown on potato dextrose agar (PDA) using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research).

[†]Corresponding author: B. A. Vinatzer; vinatzer@vt.edu

The author(s) declare no conflict of interest.

Accepted for publication 21 July 2021.

Funding

Funding was provided by the National Science Foundation, Division of Integrative Organismal Systems grant IOS-1754721. Funding to B. A. Vinatzer was also provided in part by the Virginia Agricultural Experiment Station and the Hatch Program of the United States Department of Agriculture–National Institute of Food and Agriculture.

Keywords

atmosphere, cereals and grains, field crops, fungi, *Fusarium avenaceum*, Fusarium head blight, ice nucleation activity, pathogen diversity

Table 1. Assembly summary and annotation features of *Fusarium avenaceum*, strain F156N33

Features	Strain F156N33
Assembly size (bp)	41,175,306
Number of contigs	214
Maximum contig length (bp)	3,233,628
Minimum contig length (bp)	210
Average contig length (bp)	192,487
Median contig length (bp)	1075
N_{50} contig length (bp)	1,472,944
GC content (%)	48.44
Assembly BUSCO coverage (%)[a]	C:97.8; F:0.5; M:1.7
Annotation BUSCO coverage (%)	C:96.5; F:0.3; M:3.2
Number of predicted coding genes	11,233
Mean gene length (bp)	2,478

[a] For Benchmarking Universal Single-Copy Ortholog (BUSCO) coverage, C stands for complete BUSCOs, F stands for fragmented BUSCOs, and M stands for missing BUSCOs.

Whole-genome sequencing was performed on an Illumina Nova Seq 6000 Platform at Novogene Corporation Inc. (Sacramento, CA, U.S.A.), and low-quality reads and adapters were removed by the company. FastQC v0.11.9 was used to check the quality of reads prior to the genome assembly (Andrews et al. 2010). De novo assembly was performed by SPAdes v3.13.0 with "automatic k-mer" selection, "careful" option, and "automatic coverage cutoff value" setting (Prjibelski et al. 2020). The completeness of the assembly was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCOs) v5.0.0 (Seppey et al. 2019).

The genome was confirmed as *F. avenaceum* by BLASTN v2.10.0+ (Camacho et al. 2009) using a custom database containing *F. avenaceum* assemblies from GenBank and the *Fusarium*-ID database (Geiser et al. 2004). It was determined to be 41,175,306 bp in size, with a G+C content of 48.44%. The BUSCO assessment, based on the lineage-specific profile library hypocreales_odb10 (4,494 genes), revealed the presence of 4,393 genes (97.8%). Among them, 4,381 BUSCOs (97.5%) were single copy, a higher number compared with the other *F. avenaceum* genomes in GenBank. The assembly statistics are shown in Table 1.

Total RNA of *F. avenaceum* strain F156N33 was extracted from mycelium grown on PDA using the RNeasy Plant Mini Kit (Qiagen) and ice nucleation activity of the mycelium was confirmed to ensure that genes for ice nucleation activity were expressed. RNA-seq was performed on an Illumina Nova Seq 6000 Platform at Novogene Corporation Inc., and low-quality reads and adapters were removed by the company. FastQC v0.11.9 was used for quality control, followed by alignment of RNA-seq reads to the genome assembly with STAR v2.7.8a (Dobin et al. 2013). The aligned RNA-seq reads in BAM format and the genome assembly in FASTA format were used to assemble the transcriptome in a genome-guided de novo assembly approach by Trinity v2.12.0, and a FASTA file was generated (Haas et al. 2013). Meanwhile, StringTie v2.1.5 was used to assemble transcripts (Pertea et al. 2015) and generate a GTF file that was later converted to a GFF file by GffRead v0.12.1 (Pertea and Pertea 2020).

The MAKER annotation pipeline (v3.01.03) was used for genome annotation in four consecutive rounds (Campbell et al. 2014). The previously assembled transcriptome in both FASTA format and GFF format was incorporated into the MAKER pipeline to provide transcript evidence. The pipeline also incorporated protein evidence from *F. graminearum* (ID: UP000070720), the most closely related species available. In the first round, the assembled genome was first soft masked using RepeatMasker v4.1.0 (Institute for Systems Biology); transcript evidence and protein evidence were aligned to the genome with BLASTN and BLASTX from BLAST v2.10.0+ (Camacho et al. 2009) and exonerate v2.2.0 (Slater and Birney 2005) ("est2genome" and "protein2genome" options in MAKER). In the second and third rounds, training with SNAP v2013-02-16 was conducted using the GFF file generated from the previous run (Korf 2004) before MAKER was performed for ab initio gene prediction ("snaphmm" option in MAKER). The resulting GFF file from the third round was used to train the AUGUSTUS model for *F. avenaceum* strain F156N33 with AUGUSTUS v3.4.0 (Stanke et al. 2008). In the final round, MAKER was performed with the "augustus_species"

option, which generated the final GFF file containing annotation data. Based upon the final GFF file, functional annotations were performed using InterProScan v5.46-81.0 for the presence of Pfam domains (Jones et al. 2014) and using BLASTP from BLAST v2.10.0+ for searching against the February 2021 release of the Swiss-Prot database (E-value, $<1 \times 10^{-6}$) (Camacho et al. 2009; UniProt Consortium 2020). The quality of the genome annotation was assessed by evaluating the completeness of annotated gene sets by BUSCO v5.0.0 in the protein mode (Seppey et al. 2019). The BUSCO assessment was run against the lineage-specific profile library hypocreales_odb10 (4,494 genes).

The BUSCO assessment indicated a good quality of genome annotation. In all, 4,336 genes (96.5%) were found to be complete and 15 genes (0.3%) to be fragmented (Table 1). In total, 11,233 protein-coding genes were predicted and 8,079 proteins (71.9%) were annotated with BLASTP. In addition, 8,691 proteins (77.4%) were annotated with InterProScan with the Pfam database. In total, 9,111 of the predicted proteins (81.1%) received a functional annotation.

In conclusion, this article reports the first-draft genome of *F. avenaceum* using both whole-genome sequencing and RNA-seq data from the same strain. In addition, customized training was used to improve the quality of the genome annotation. The genome can help to provide insights into the molecular aspects of this important plant pathogen. The genome also represents the first draft genome of an experimentally confirmed ice-nucleation-active fungal isolate. Thus, it will also be valuable to the scientific community studying ice nucleation activity.

The draft genome has been deposited at DNA Data Bank of Japan/European Nucleotide Archive/GenBank under accession number JAGPUO000000000, Biosample SAMN18673054, Bioproject PRJNA720629. The version described in this article is version JAGPUO010000000.

Acknowledgments

We thank D. G. Schmale III for sharing *F. avenaceum* strain F156N33.

Literature Cited

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Institute, Babraham, U.K.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. 2009. BLAST+: Architecture and applications. BMC Bioinf. 10:421.

Campbell, M. S., Holt, C., Moore, B., and Yandell, M. 2014. Genome annotation and curation using MAKER and MAKER-P. Curr. Protoc. Bioinf. 48:4.11.11-14.11.39.

Desjardins, A. E. 2003. *Gibberella* from *A* (*venaceae*) to *Z* (*eae*). Annu. Rev. Phytopathol. 41:177-198.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21.

Doohan, F. M., Brennan, J., and Cooke, B. M. 2003. Influence of climatic factors on *Fusarium* species pathogenic to cereals. Eur. J. Plant Pathol. 109:755-768.

Geiser, D. M., del Mar Jiménez-Gasco, M., Kang, S., Makalowska, I., Veeraraghavan, N., Ward, T. J., Zhang, N., Kuldau, G. A., and O'Donnell, K. 2004. FUSARIUM-ID v. 1.0: A DNA sequence database for identifying Fusarium. Eur. J. Plant Pathol. 110:473-479.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., and Regev, A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494-1512.

Hasegawa, Y., Ishihara, Y., and Tokuyama, T. 1994. Characteristics of ice-nucleation activity in *Fusarium avenaceum* IFO 7158. Biosci. Biotechnol. Biochem. 58: 2273-2274.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., and Hunter, S. 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30:1236-1240.

Korf, I. 2004. Gene finding in novel genomes. BMC Bioinf. 5:59.

Kunert, A. T., Pöhlker, M. L., Tang, K., Krevert, C. S., Wieder, C., Speth, K. R., Hanson, L. E., Morris, C. E., Schmale, D. G., III, Pöschl, U., and Fröhlich-Nowoisky, J. 2019. Macromolecular fungal ice nuclei in *Fusarium*: Effects of physical and chemical processing. Biogeosciences 16:4647-4659.

Lin, B. 2013. Movement and structure of atmospheric populations of *Fusarium*. Doctoral dissertation, Virginia Polytechnic Institute and State University. https://vtechworks.lib.vt.edu/handle/10919/23203

Nalim, F. A., Elmer, W. H., McGovern, R. J., and Geiser, D. M. 2009. Multilocus phylogenetic diversity of *Fusarium avenaceum* pathogenic on lisianthus. Phytopathology 99:462-468.

Parry, D. W., Jenkinson, P., and McLeod, L. 1995. *Fusarium* ear blight (scab) in small grain cereals—A review. Plant Pathol. 44:207-238.

Pertea, G., and Pertea, M. 2020. GFF utilities: GffRead and GffCompare. F1000 Res. 9:J-304.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290-295.

Pouleur, S., Richard, C., Martin, J. G., and Antoun, H. 1992. Ice nucleation activity in *Fusarium acuminatum* and *Fusarium avenaceum*. Appl. Environ. Microbiol. 58:2960-2964.

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. 2020. Using SPAdes *de novo* assembler. Curr. Protoc. Bioinf. 70:e102.

Richard, C., Martin, J.-G., and Pouleur, S. 1996. Ice nucleation activity identified in some phytopathogenic *Fusarium* species. Phytoprotection 77:83-92.

Seppey, M., Manni, M., and Zdobnov, E. M. 2019. BUSCO: Assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962:227-245.

Slater, G. S. C., and Birney, E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinf. 6:31.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics 24:637-644.

Uhlig, S., Jestoi, M., and Parikka, P. 2007. *Fusarium avenaceum*—The North European situation. Int. J. Food Microbiol. 119:17-24.

UniProt Consortium. 2020. UniProt: The universal protein knowledgebase in 2021. Nucleic Acids Res. 49:D480-D489.