

# Telomere to Telomere Genome Assembly of *Fusarium musae* F31, Causal Agent of Crown Rot Disease of Banana

Luca Degradi, Valeria Tava, Andrea Kunova, Paolo Cortesi, Marco Saracchi, and Matias Pasquali<sup>†</sup>

Department of Food, Environmental and Nutritional Sciences, Università degli Studi di Milano, via Celoria 2, 20133 Milano, Italy

## Abstract

*Fusarium musae* causes crown rot of banana and it is also associated to clinical fusariosis. A chromosome-level genome assembly of *F. musae* F31 obtained combining Nanopore long reads and Illumina paired-end reads resulted in 12 chromosomes plus one contig with overall N<sub>50</sub> of 4.36 Mb, and is presented together with its mitochondrial genome (58,072 bp). The F31 genome includes telomeric regions for 11 of the 12 chromosomes representing one of the most complete genomes available in the *Fusarium fujikuroi* species complex. The high-quality assembly of the F31 genome will be a valuable resource for studying the pathogenic interactions occurring between *F. musae* and banana. Moreover, it represents an important resource for understanding the genome evolution in the *F. fujikuroi* species complex.

*Fusarium musae* (Van Hove et al. 2011) is a filamentous fungus (class Sordariomycetes, phylum Ascomycota) able to infect banana but it is also found in clinical settings (Tava et al. 2021; Triest et al. 2016). This fungus is one of the causal agents of crown rot of banana, a major cause of postharvest losses in banana (Van Hove et al. 2011). It belongs to the *Fusarium fujikuroi* species complex (FFSC), which includes at least 50 distinct species (O'Donnell et al. 1998; Yilmaz et al. 2021). Improving the quality of the available assembly in the FFSC will provide a valuable resource for the identification of pathogenic genes and for comparative genomic studies. The complete genome of *F. musae* will contribute to study genomes evolution within the FFSC. Actual reference for the FFSC (Niehaus et al. 2016) is *F. verticillioides* 7600 (AN: GCA\_000149555.1). Only one assembly of *F. musae* species is available to date (AN: GCA\_013623345.1).

To obtain a complete genome, a hybrid sequencing approach was used. The F31 strain was originally isolated from infected banana from the Dominican Republic and proved to be pathogenic on banana (Kamel et al. 2016) and deposited in the German Collection of Microorganisms and Cell Cultures GmbH (Leibniz, Germany) under the accession number DSM 112727.

In our work, high molecular weight DNA was extracted from strain F31 lyophilized tissue using the cetyltrimethylammonium bromide extraction method (Breakspear et al. 2011) followed by the Qiagen genomic tip (Qiagen) cleaning procedure according to the

<sup>†</sup>Corresponding author: M. Pasquali; [matias.pasquali@unimi.it](mailto:matias.pasquali@unimi.it)

L. Degradi and V. Tava contributed equally to this work.

The author(s) declare no conflict of interest.

Accepted for publication 12 August 2021.

## Funding

Support was provided by Università degli Studi di Milano (APC contribution).

## Keywords

cross-kingdom pathogen, crown rot disease, fungus–plant interactions, fusariosis, *Fusarium fujikuroi* species complex, genomics, metabolomics, proteomics, *Musa* spp., reference genome

**Table 1.** Summary of the genome assembly and annotation statistics of *Fusarium musae* strain F31, compared with a previous assembly of *F. musae* NRRL strain 25059 and with the best reference available in the *Fusarium fujikuroi* species complex (FFSC), *F. verticillioides* 7600

Strains	F31 (this study)	NRRL 25059 (GCA_013623345.1) <sup>a</sup>	<i>F. verticillioides</i> 7600 (GCA_000149555.1) <sup>b</sup>
Number of contigs	14	1,149	37 (scaffold)
Number of chromosomes	12	NA	11
Mitochondrial DNA (size)	Yes (58,072 bp)	Yes (57,891 bp)	Yes (53,743 bp)
Genome size (Mb)	44.07	42.5	41.8
GC content (%)	47.22	48.70	48.68
Contig N <sub>50</sub> (Mb)	4.36	0.09	1.96 (scaffold)
Number of protein coding genes	13,963	13,931	13,699
BUSCO completeness <sup>c</sup>	4,485/4,494	4,463/4,494	4,372/4,494

<sup>a</sup> NA = not available.

<sup>b</sup> Nearest species belonging to FFSC with chromosome level assembly.

<sup>c</sup> Completed BUSCO (single and multiple copy)/total *hypocreales\_odb10* BUSCO.

manufacturer's instructions. The MiSeq Reagent Kit v2 (Illumina) led to a total of 11,812,123 paired-end short reads (2 × 151) and SQK-LSK109 sequencing kit (Nanopore) led to 2,221,411 reads of single-molecule long-read sequencing using the MinION MIN101B platform, R9.4.1 flow cell (Nanopore). Obtained coverage was 60× by Illumina and 250× by Nanopore reads.

Assembly was performed on NanoLyse processed sequences (to remove phage λ reads) using Canu v.2.1.1 + galaxy0 (Koren et al. 2017) with default settings. Autopolishing was performed using Medaka v.1.0.3 + galaxy2. Minimap2 v.2.17 + galaxy4 (Li and Durbin 2010) was used to align short reads on the obtained assembly and Pilon v.1.20.1 (Walker et al. 2014) was used to extract the consensus sequence, giving 0.48% of corrected erroneous positions ( $n = 214,296$ ).

Manual correction was done using Geneious Prime software v.11 (Biomatters). The Minimap2 plugin (default settings) tool aligning long reads on the obtained assembly validated seven contiguous contigs. Others had errors corrected using a short-reads assembly obtained with Shovill v1.1.0 + galaxy0 with default settings. Manual correction allowed the connection of different contigs. Telomeric regions, based on Telomerase DB, were annotated to validate chromosomes.

Final assembly statistics were evaluated using Quast tool v.5.0.2 + galaxy1 (Mikheenko et al. 2018). The ultimate genome size of *F. musae* F31 was 44.07 Mb (1.5 Mb more than the reference), divided into 12 chromosomes, the circular mitochondrial DNA (mtDNA), and one unplaced contig. N<sub>50</sub> of the final assembly was 4.36 Mb, confirming high contiguity of the obtained assembly. Of the 12 chromosomes, 11 had both telomeres (TTAGGG) included, while 1 had only one telomeric region. Compared with previous assembly of *F. musae* (GCA\_013623345.1) and *F. verticillioides* (GCA\_000149555.1) present in NCBI, this genome substantially improves completeness within the FFSC. Completeness measured using BUSCO v.5.0.0 + galaxy0 tool (Simão et al. 2015) on the *hypocreales\_odb10* database was higher for F31 compared with the NRRL25059 strain of *F. musae*. Moreover, contig number and N<sub>50</sub> results were significantly improved for F31 genome assembly (Table 1).

Annotation was performed using Funannotate software v.1.8.7. After the first step of repeat sequence masking, done with repeat mask v.4.0.7, Augustus tool v.3.3.3 (Keller et al. 2011), trained with the *F. graminearum* protein database, was used for nuclear ab initio annotation. Functional annotation of the predicted coding regions was performed using Antismash v.5.0, Interproscan v.5.51-85.0, and EggNOG mapper v.5.0.0 databases. For mtDNA annotation, Mfannot and RNAWeasel were used with NOVOplasty v.4.2 + galaxy0 (Dierckxsens et al. 2020) assembly using the *F. verticillioides* 7600 strain (AN: GCA\_000149555.1) mtDNA as a reference and a sequence of 1,700 bp in length as a seed sequence.

After the masking sequence process, which masked 4.32% of sequences, ab initio gene prediction generated the prediction position of the coding region. Functional annotation finalized to assign function to predicted sequences led to a total of 13,963 annotated genes, of which 13,661 were proteins and 302 transfer RNA (tRNA) for nuclear DNA. For mtDNA 15 proteins, two open reading frames, the small subunit ribosomal RNA, and 27 tRNA were annotated.

The genome assembly of F31 will be a useful resource for comparative analysis of *F. musae* species and represents a reference for completeness in the FFSC.

## Data Availability

The F31 genome project can be found at GenBank under BioProject accession number PRJNA718489. The chromosomal sequence and genes annotation of F31 can be found at GenBank with accession number GCA\_019915245.1, BioProject accession number JAHBCI000000000. Nanopore reads and Illumina paired-end reads are available in the NCBI Sequenced Read Archive under the accession numbers SRR14117444 and SRR14117445, respectively.

## Acknowledgments

The work is part of the activity of the HUPLANT Cost Action.

## Author-Recommended Internet Resources

Antismash: <https://fungismash.secondarymetabolites.org#!/start>  
EggNOG mapper: <http://eggnog-mapper.embl.de/>  
Funannotate: <https://github.com/nextgenusfs/funannotate>  
Galaxy: <https://usegalaxy.eu>  
Interproscan: <https://github.com/ebi-pf-team/interproscan>  
Medaka: <https://github.com/nanoporetech/medaka>  
Mfannot: <https://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>  
NanoLyse: <https://github.com/wdecoster/nanolyse>  
Repeat-mask: <https://github.com/rmhubble/RepeatMasker>  
RNAWeasel: <https://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl>  
Shovill: <https://github.com/tseemann/shovill>  
Telomerase DB: <http://telomerase.asu.edu/sequences.html>

---

## Literature Cited

- Breakspear, A., Pasquali, M., Broz, K., Dong, Y., and Kistler, H. C. 2011. Npc1 is involved in sterol trafficking in the filamentous fungus *Fusarium graminearum*. *Fungal Genet. Biol.* 48:725-730.
- Dierckxsens, N., Mardulyn, P., and Smits, G. 2020. Unraveling heteroplasmy patterns with NOVOPlasty. *NAR Genom. Bioinf.* 2:lqz011.
- Hubble, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F., and Wheeler, T. J. 2016. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* 44:D81-D89.
- Kamel, M. A. M., Cortesi, P., and Saracchi, M. 2016. Etiological agents of crown rot of organic bananas in Dominican Republic. *Postharvest Biol. Technol.* 120:112-120.
- 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.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* 27:722-736.
- Li, H., and Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589-595.
- Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., and Gurevich, A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142-i150.
- Niehaus, E.-M., Münsterkötter, M., Proctor, R. H., Brown, D. W., Sharon, A., Idan, Y., Oren-Young, L., Sieber, C. M., Novák, O., Pěncík, A., Tarkowská, D., Hromadová, K., Freeman, S., Maymon, M., Elazar, M., Youssef, S. A., El-Shabrawy, E. S., Shalaby, A. B., Houterman, P., Brock, N. L., Burkhardt, I., Tsavkelova, E. A., Dickschat, J. S., Galuszka, P., Güldener, U., and Tudzynski, B. 2016. Comparative “Omics” of the *Fusarium fujikuroi* species complex highlights differences in genetic potential and metabolite synthesis. *Genome Biol. Evol.* 8:3574-3599.
- O'Donnell, K., Cigelnik, E., and Nirenberg, H. I. 1998. Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* 90:465-493.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Tava, V., Prigitano, A., Cortesi, P., Esposto, M. C., and Pasquali, M. 2021. *Fusarium musae* from diseased bananas and human patients: Susceptibility to fungicides used in clinical and agricultural settings. *J. Fungi* 7:784.
- Triest, D., Piérard, D., De Cremer, K., and Hendrickx, M. 2016. *Fusarium musae* infected banana fruits as potential source of human fusariosis: May occur more frequently than we might think and hypotheses about infection. *Commun. Integr. Biol.* 9:e1162934.
- Van Hove, F., Waalwijk, C., Logrieco, A., Munaut, F., and Moretti, A. 2011. *Gibberella musae* (*Fusarium musae*) sp. nov., a recently discovered species from banana is sister to *F. verticillioides*. *Mycologia* 103:570-585.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963.
- Yilmaz, N., Sandoval-Denis, M., Lombard, L., Visagie, C. M., Wingfield, B. D., and Crous, P. W. 2021. Redefining species limits in the *Fusarium fujikuroi* species complex. *Persoonia* 46:129-162.