

1 **Telomere to telomere genome assembly of *Fusarium musae* F31, causal agent of**
2 **crown rot disease of banana.**

3 Luca Degradi*, Valeria Tava*, Andrea Kunova, Paolo Cortesi, Marco Saracchi, Matias
4 Pasquali.

5 Department of Food, Environmental and Nutritional Sciences, Università degli Studi di
6 Milano, via Celoria 2, 20131 Milano (Italy).

7

8 *equal contribution

9

10 Corresponding author: Pasquali M. matias.pasquali@unimi.it

11 **Abstract:**

12 *Fusarium musae* causes crown rot of banana and it is also associated to clinical fusariosis. A
13 chromosome-level genome assembly of *Fusarium musae* F31 obtained combining Nanopore
14 long reads and Illumina paired end reads resulted in 12 chromosomes plus one contig with
15 overall N₅₀ of 4.36 Mb, and is presented together with its mitochondrial genome (58072 bp).
16 F31 genome includes telomeric regions for 11 of the 12 chromosomes representing one of
17 the most complete genome available in the *Fusarium fujikuroi* species complex. The high-
18 quality assembly of the F31 genome will be a valuable resource for studying the pathogenic
19 interactions occurring between *F. musae* and banana. Moreover, it represents an important
20 resource for understanding the genome evolution in the *Fusarium fujikuroi* species complex.

21

22

23 **Key word:** *Fusarium fujikuroi* species complex, crown rot disease, *Musa* spp., cross-
24 kingdom pathogen, fusariosis, reference genome.

25

26 *Fusarium musae* (Van Hove et al., 2011) is a filamentous fungus (Class Sordariomycetes,
27 phylum Ascomycota) able to infect banana (Kamel et al., 2016), but it is also found in clinical
28 settings (Triest et al., 2016). This fungus is one of the causal agents of crown rot of banana,
29 a major cause of postharvest losses in banana (Van Hove et al., 2011). It belongs to the
30 *Fusarium fujikuroi* species complex (FFSC), which includes at least 50 distinct species
31 (O'Donnell et al., 1998; Yilmaz et al., 2021). Improving the quality of the available assembly
32 in the FFSC will provide a valuable resource for the identification of pathogenic genes and
33 for comparative genomic studies. The complete genome of *F. musae* will contribute to study
34 genomes evolution within the FFSC. Actual reference for the FFSC is *F. verticillioides* 7600
35 (AN: GCA_000149555.1). Only one assembly of *F. musae* species is available up to date
36 (AN: GCA_013623345.1).

37 To obtain a complete genome, a hybrid sequencing approach was used. The F31 strain was
38 originally isolated from infected banana from the Dominican Republic and proved to be
39 pathogenic on banana (Kamel et al., 2016) and deposited in the German Collection of
40 Microorganisms and Cell Cultures GmbH (Leibniz, GER) under the accession number DSM
41 112727.

42 In our work, high-molecular weight DNA was extracted from F31 strain lyophilized tissue
43 using the CTAB extraction method (Breakspear *et al.*, 2011) followed by Qiagen genomic tip
44 (Qiagen, USA) cleaning procedure according to manufacturer's instructions. MiSeq Reagent
45 Kit v2 (Illumina, USA) led to a total of 11.812.123 paired-end short reads (2 x 151) and SQK-
46 LSK109 sequencing kit (Nanopore, UK) led to 2.221.411 reads of single molecule long read
47 sequencing using MinION MIN101B platform, R9.4.1 flow cell (Nanopore, UK). Obtained
48 coverage was 60X by Illumina and 250X by Nanopore reads.

49 Assembly was performed on NanoLyse processed sequences (to remove lambda reads)
50 using Canu v.2.1.1+galaxy0 (Koren *et al.*, 2017) with default settings. Autopolishing was
51 performed using Medaka v.1.0.3+galaxy2. Minimap2 v.2.17+galaxy4 (Li *et al.*, 2010) was
52 used to align short reads on the obtained assembly and Pilon v.1.20.1 (Walker *et al.*, 2014)
53 was used to extract the consensus sequence giving 0.48% of corrected erroneous positions
54 (n=214.296).

55 Manual correction was done using Geneious Prime software v11 (Biomatters, NZ).
56 Minimap2 plugin (default settings) tool aligning long reads on the obtained assembly
57 validated 7 contiguous contigs. Others had errors corrected using short reads assembly
58 obtained with Shovill v1.1.0+galaxy0 with default settings. Manual correction allowed to
59 connect different contigs. Telomeric regions, based on Telomerase DB, were annotated to
60 validate chromosomes.

61 Final assembly statistics were evaluated using Quast tool v.5.0.2+galaxy1 (Mikheenko *et al.*,
62 2018). The ultimate genome size of *F. musae* F31 was 44.07 Mb (1.5 Mb more than the
63 reference), divided into 12 chromosomes, the circular mtDNA and 1 unplaced contig. N50 of
64 the final assembly was 4.36 Mb confirming high contiguity of the obtained assembly. Of the
65 12 chromosomes, 11 had both telomeres (TTAGGG) included, while one had only one
66 telomeric region. Compared with previous assembly of *F. musae* (GCA_013623345.1) and
67 *F. verticilloides* (GCA_000149555.1) present on NCBI, this genome improves substantially
68 completeness within the *F. fujikuroi* species complex. Completeness measured using
69 BUSCO v.5.0.0+galaxy0 tool (Simao *et al.*, 2015) on *hypocreales_odb10* database was
70 higher for F31 compared to NRRL25059 strain of *F. musae*. Moreover, contig number and
71 N50 results were significantly improved for F31 genome assembly (Table 1).

72 Annotation was performed using Funannotate software v. 1.8.7. After first step of repeat
73 sequences masking done with repeat mask v. 4.0.7, Augustus tool v.3.3.3, (Keller *et al.*,
74 2011) trained with *Fusarium graminearum* protein database, was used for nuclear "ab initio"
75 annotation. Functional annotation of the predicted coding region was performed using
76 Antismash v. 5.0, Interproscan v. 5.51-85.0, EggNOG mapper v. 5.0.0, database. For
77 mtDNA annotation, Mfannot and RNAWeasel were used with NOVOplasty v.4.2+galaxy0
78 (Dierckxsens *et al.*, 2019) assembly using the *F. verticilloides* 7600 (AN:
79 GCA_000149555.1) strain mtDNA as a reference and a sequence of 1700 bp length as a
80 seed sequence.

81 After masking sequences process, which masked 4.32% of sequences, ab initio gene
82 prediction generate the prediction position of coding region. Final functional annotation
83 finalized to assign function to predicted sequences lead to a total of 13963 annotated genes,
84 13661 of which are proteins and 302 are tRNA for nuclear DNA. For mtDNA 15 proteins, the
85 small subunit ribosomal RNA (rns) and 27 tRNA were annotated.

86 The genome assembly of F31 will be a useful resource for comparative analysis of *Fusarium*
 87 *musae* species and represents a reference for completeness in the *Fusarium fujikuroi*
 88 species complex.

89 The F31 genome project can be found at GenBank under BioProject accession number
 90 PRJNA718489. The chromosomal sequence and genes annotation of F31 can be found at
 91 GenBank with the accession number JAHBCI000000000. Nanopore reads and Illumina
 92 paired end reads are available in the NCBI Sequenced Read Archive under the accession
 93 numbers SRR14117444 and SRR14117445 respectively.

94 **Author-recommended internet resource:**

95 Antismash: <https://fungismash.secondarymetabolites.org/#!/start>

96 EggNOG mapper: <http://eggno-mapper.embl.de/>

97 Funannotate: <https://github.com/nextgenusfs/funannotate>

98 Galaxy: <https://www.galaxyeu.org>

99 Interproscan: <https://github.com/ebi-pf-team/interproscan>

100 Medaka: <https://github.com/nanoporetech/medaka>

101 Mfannot: <https://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl> NanoLyse:
 102 <https://github.com/wdecoster/nanolyse>

103 Repeat-mask: <https://github.com/rmhubble/RepeatMasker>

104 RNAWeasel: <https://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl>

105 Shovill: <https://github.com/tseemann/shovill>

106 Telomerase DB: <http://telomerase.asu.edu/sequences.html>

107

108 **Acknowledgements**

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

110

111 **References:**

112 Breakspear A, Pasquali M, Broz K, Dong Y, Kistler HC (2011). Npc1 is involved in sterol
 113 trafficking in the filamentous fungus *Fusarium graminearum*. Fungal Genetics and Biology
 114 48: 725–730. <https://doi.org/10.1016/j.fgb.2011.03.001>

115 Dierckxsens N, Mardulyn P, Smits G (2020) Unraveling heteroplasmy patterns with
 116 NOVOPlasty. NAR Genomics and Bioinformatics 2: 1-10.
 117 <https://doi.org/10.1093/nargab/lqz011>

118 Hove FV, Waalwijk C, Logrieco A, Munaut F, Moretti A, 2011. *Gibberella musae* (*Fusarium*
 119 *musae*) sp. nov., a recently discovered species from banana is sister to *F. verticillioides*.
 120 Mycologia 103: 570–585. <https://doi.org/10.3852/10-038>

121 Hubble R, Finn RD, Clements J, *et al.* (2016) The Dfam database of repetitive DNA
 122 families. Nucleic Acids Research 44: D81–D89. <https://doi.org/10.1093/nar/gkv1272>

- 123 Kamel MAM, Cortesi P, Saracchi M (2016) Etiological agents of crown rot of organic
 124 bananas in Dominican Republic. *Postharvest Biology and Technology* 120:112–120.
 125 <https://doi.org/10.1016/j.postharvbio.2016.06.002>
- 126 Koren S, Walenz BP, Berlin K, et al (2017) Canu: scalable and accurate long-read
 127 assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 27: 722–
 128 736. <https://doi.org/10.1101/gr.215087.116>
- 129 Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler
 130 transform. *Bioinformatics* 26: 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- 131 Mikheenko A, Prjibelski A, Saveliev V, et al. (2018) Versatile genome assembly evaluation
 132 with QUAST-LG. *Bioinformatics* 34: i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>
- 133 Niehaus E-M, Münsterkötter M, Proctor RH, et al. (2016) Comparative “Omics” of the
 134 *Fusarium fujikuroi* species complex highlights differences in genetic potential and metabolite
 135 synthesis. *Genome Biology and Evolution* 8: 3574–3599.
 136 <https://doi.org/10.1093/gbe/evw259>
- 137 O’Donnell K, Cigelnik E, Nirenberg HI (1998a) Molecular systematics and phylogeography
 138 of the *Gibberella fujikuroi* species complex. *Mycologia* 90: 465–493.
 139 <https://doi.org/10.1080/00275514.1998.12026933>
- 140 Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: assessing genome
 141 assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–
 142 3212. <https://doi.org/10.1093/bioinformatics/btv351>
- 143 Triest D, Hendrickx M (2016) Postharvest disease of banana caused by *Fusarium musae*: a
 144 public health concern? *PLOS Pathogens* 12: e1005940.
 145 <https://doi.org/10.1371/journal.ppat.1005940>
- 146 Triest D, Piérard D, De Cremer K, Hendrickx M (2016) *Fusarium musae* infected banana
 147 fruits as potential source of human fusariosis: may occur more frequently than we might
 148 think and hypotheses about infection. *Communicative and Integrative Biology* 9: e1162934.
 149 <https://doi.org/10.1080/19420889.2016.1162934>
- 150 Walker BJ, Abeel T, Shea T, et al (2014) Pilon: an Integrated tool for comprehensive
 151 microbial variant detection and genome assembly improvement. *PLOS ONE* 9: e112963.
 152 <https://doi.org/10.1371/journal.pone.0112963>
- 153 Yilmaz N, Sandoval-Denis M, Lombard L, Visagie CM, Wingfield BD, Crous, PW, 2020.
 154 Redefining species limits in the *Fusarium fujikuroi* species complex. *Persoonia*
 155 <https://doi.org/10.3767/persoonia.2021.46.05>

156
 157
 158
 159
 160
 161

162 **Table_1.** Summary of the genome assembly and annotation statistics of *F. musae* F31 strain,
 163 compared with a previous assembly of *F. musae* NRRL 25059 strain and with the best reference
 164 available in the FFSC, *F. verticillioides* 7600.

165

Strains	F31 (this study)	NRRL 25059 (GCA_013623345.1)	<i>F. verticillioides</i> 7600 ^a (GCA_000149555.1)
Number of contigs	14	1149	37 (scaffold)
Number of chromosomes	12	N.A. ^b	11
MtDNA (size)	Yes (58072 bp)	Yes (57891 bp)	Yes (53743 bp)
Genome size (Mb)	44.07	42.5	41.8
GC content (%)	47.22	48.70	48.68
Contig N50 (Mb)	4.36	0.09	1.96 (scaffold)
Number of protein coding genes	13963	13931	13699
BUSCO completeness ^c	4485/4494	4463/4494	4372/4494

166 ^anearest species belonging to FFSC with chromosome level assembly.

167 ^bNot Available

168 ^cCompleted BUSCO (single and multiple copy)/Total *hypocreales_odb10* BUSCO

169

170