Chromosome-scale genome assembly of *Fusarium oxysporum* strain Fo47, a fungal endophyte and biocontrol agent

Bo Wang[1], Houlin Yu[2,3], Yanyan Jia[1], Quanbin Dong[4], Christian Steinberg[5], Claude Alabouvette[5], Veronique Edel-Hermann[5], H. Corby Kistler[6], Kai Ye[1,4,†], Li-Jun Ma[2,3,†], Li Guo[1,4,†]

[1] *MOE Key Laboratory for Intelligent Networks & Network Security, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China*

[2] *Graduate Program of Plant Biology, University of Massachusetts Amherst, Amherst MA 01003 USA*

[3] *Department of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst MA 01003 USA*

[4] *School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049 China*

[5] *Agroécologie, AgroSup Dijon, INRA, University of Bourgogne Franche-Comté, F-21000 Dijon, France.*

[6] *USDA ARS Cereal Disease Laboratory, University of Minnesota, St. Paul, MN, 55108, USA*

† Corresponding authors:

kaiye@xjtu.edu.cn (KY)

lijun@biochem.umass.edu (LJM)

guo_li@xjtu.edu.cn (LG)

Abstract: Here, we report a chromosome-level genome assembly of *Fusarium oxysporum* strain Fo47 (12 pseudomolecules; contig N50: 4.52Mb), generated using a combination of PacBio long-read, Illumina pair-ended and Hi-C sequencing data. Although *F. oxysporum* causes vascular wilt to over 100 plant species, the strain Fo47 is classified as an endophyte and widely used as a biocontrol agent for plant disease control. The Fo47 genome carries a single accessory chromosome of 4.23 Mb, compared to the reference genome of *F. oxysporum* f.sp. *lycopersici* strain Fol4287. The high-quality assembly and annotation of the Fo47 genome will be a valuable resource for studying the mechanisms underlying the endophytic interactions between *F. oxysporum* and plants, as well as deciphering the genome evolution of the *F. oxysporum* species complex.

Keywords: *Fusarium oxysporum*, genome assembly, fungal endophyte, plant-fungal interactions, biocontrol

Genome announcement

A filamentous ascomycete fungus, *Fusarium oxysporum* represents a species complex of diverse members that cause soil-borne vascular wilt diseases to over 100 plant species in a host-specific fashion (Kistler et al. 1997; Ma et al. 2013; Edel-Hermann and Lecomte 2019). Strains pathogenic to specific host species are classified as *formae speciales* (f. sp.), and, together with related non-pathogenic strains, they are collectively known as the *F. oxysporum* species complex (FOSC) (Kistler et al. 1997; Michielse and Rep 2009). Although notorious as plant pathogens, commonly *F. oxysporum* strains are plant endophytes often beneficial to host plants (Alabouvette et al. 2009). Fo47, a strain isolated from plant disease suppressive soils (Alabouvette 1986), grows endophytically and is widely used as biological control agent for preventing vascular wilt of tomato caused by *F. oxysporum* f.sp. *lycopersici* (Fol) (Alabouvette et al. 2009) and other plant diseases as well (He et al. 2002; Paparu et al. 2007; Veloso and Díaz 2012). In general, beneficial microbes contribute to the fitness of host plants potentially through diverse mechanisms including induced systemic resistance, promoting nutrient assimilation and conferring antagonisms against microbial pathogens (Alabouvette et al. 2009; Van der Ent et al. 2009; Aimé et al. 2013; Pieterse et al. 2014). The biocontrol effect of Fo47 is widely exploited in agriculture but the molecular mechanisms behind it remain poorly understood. Decoding the genome of Fo47 will provide a valuable resource for not only facilitating dissection of such mechanisms, but also contributing to comparative genomic studies for understanding the genome evolution of FOSC in terms of endophytism, pathogenesis and host specificity. The genome assembly of the same strain was previously deposited at NCBI under the accession number GCA_000271705.2.

High molecular weight genomic DNA was extracted from four days old Fo47 mycelia using CATB (Cetyl Trimethyl Ammonium Bromide) method. A total of 6.85 Gb paired-end short sequence reads (2x150bp) and 18.09 Gb single-molecule long-read sequence reads were generated using Illumina NovaSeq 6000 platform and PacBio SMRT RS II platform, respectively. K-mer (17mer) frequency analysis of 6.85 Gb Illumina reads (138X coverage) using Jellyfish-GenomeScope pipeline (Marcais and Kingsford 2011; Vurture et al. 2017) revealed an estimated genome size of 49.54 Mb for Fo47. The 18.09 Gb (341X coverage) PacBio SMRT data (1.24 million reads) has a read length N50 of 18,156 bp after filtering. We then assembled Fo47 genome from the PacBio reads using the assembler *NextDenovo* v2.0 (https://github.com/Nextomics/NextDenovo). The PacBio assembly was further polished by Illumina PE reads using *Pilon* v. 1.22 (Walker et al. 2014) to correct 2,040 (0.004%) erroneous positions. To make the chromosome-scale assembly, the contigs were scaffolded using chromatin contact maps produced by Hi-C (high throughput chromosome conformation capture) reads. Basically, a single Hi-C library prepared from cross-linked chromatins of fungal cells using a standard Hi-C protocol (Belton and Dekker 2015) were sequenced by NovaSeq 6000 to yield 5.76Gb (115x coverage) Illumina PE reads. The Hi-C data was used to anchor all contigs on 12 pseudomolecules (chromosomes) using Juicer v. 1.5 (Durand et al. 2016a) and Juicebox v. 1.11.08 (Durand et al. 2016b), followed by manual checks. The number of chromosomes was determined by centromeric interaction regions detected in Hi-C contacted map figures (Marbouty et al. 2014).

The final assembly of the strain Fo47 genome is 50.36 Mb placed in 12 chromosomes, proximate to its estimated genome size. Each chromosome is represented in a single contig,

with a contig N50 of 4.52Mb, demonstrating the high continuity of this assembly (Table 1). Out of 12 chromosomes, nine had more than three copies of the telomeric repeat unit TTAGGG on both ends, whereas the other three contained those copies on only one end, indicating a high-quality Fo47 genome assembly of chromosome level. Compared to a previous Fo47 genome assembly (GenBank accession: GCA_000271705.2) at NCBI (Table 1), the major improvement happens at chromosome Chr7, as 73 contigs (3.96Mb) of the initial assembly can be aligned to the single 4.23Mb Chr7 of the new assembly using *Mummer* v. 3.1 (Kurtz et al. 2004).

For genome annotation, *GeneMark-ES* (Borodovsky and Lomsadze 2011) was first used for *de novo* gene structure prediction. A *Fusarium*-specific gene model was then used to train *Augustus* v. 3.1 (Mario and Burkhard 2005). *MAKER2* pipeline (Cantarel et al. 2008) was used to find protein-coding genes integrating gene models predicted from *GeneMark-E*S and *Augustus* and protein sequences of the *F. oxysporum* reference genome of a tomato pathogenic isolate Fol4287 (GenBank assembly accession: GCA_000149955.2) (Ma et al. 2010), with RepeatMasker v. 4.0.7 (Tarailo-Graovac and Chen 2009) option on to find and mask repetitive elements. In total, the Fo47 genome has 17,234 protein coding genes and 3.42% repeat content (Table 1). The genome assembly completeness was assessed by BUSCO (Benchmarking Universal Single Copy Ortholog) (Simão *et al.* 2015) using both genome and proteome sequences, showing BUSCO score of 98% (genome) and 99% (proteome) based on the 'fungi_odb9' library (Table 1), suggesting a high degree of completeness for this Fo47 genome assembly.

Compared to the reference genome Fol4287 (Ma et al. 2010), Fo47 has a smaller genome size, fewer genes and repeat elements (Table 1). A whole genome sequence alignment between Fo47 and Fol4287 using *Mummer* v. 3.1 (Kurtz et al. 2004) reveals that the size of sequence mapped to core genome of Fol4287 (11 chromosomes) (Ma et al. 2010) was 46.10Mb, suggesting a set of 11 core chromosomes in Fo47. A single Fo47 chromosome Chr7 (4.23Mb) did not align to any Fol4287 chromosome, indicative of a lineage-specific chromosome in Fo47. Chr7 is highly enriched for transposable elements (TE), with 51% of total TE sequences located on Chr7 and 49% located on the rest of genome. Over representation of repetitive sequences in this chromosome explains the fragmentation in the initial assembly. Chr7 has a smaller average gene density (290 per Mb) compared to core chromosomes (347 per Mb). The 1,247 genes on Chr7 are significantly enriched (P-value: 1.5e-4) for regulators of G-protein signaling functions suggesting their likely role in external sensing and organism response to external nutrient sources and other environmental signals (Bayram and Braus 2012; Kim et al. 2017). This genome assembly of Fo47 will be a useful resource for comparative genomics of *F. oxysporum*, and future functional genomic studies of Fo47 and host endophytic interactions. The assembled genome and gene annotations have been deposited at GenBank under the accession numbers CP052038-CP052049. PacBio reads, Illumina reads and Hi-C Illumina data are available in the NCBI Sequenced Read Archive under the accession numbers SRR11526760, SRR11523115, and SRR11523120, respectively.

Acknowledgment

We would like to thank the anonymous reviewers for their kind and helpful comments on

the original manuscript.

Funding

Literature Cited

Alabouvette, C. 1986. Fusarium-wilt suppressive soils from the Châteaurenard region: Review of a 10-year study. Agronomie 6:273-284.

Alabouvette, C., Olivain, C., Migheli, Q., and Steinberg, C. 2009. Microbiological control of soil-borne phytopathogenic fungi with special emphasis on wilt-inducing *Fusarium oxysporum*. New Phytol. 184:529-544.

Aimé S., Alabouvette C., Steinberg C. and Olivain C. 2013. The endophytic strain *Fusarium oxysporum* Fo47: a good candidate for priming the defence responses in tomato roots. Mol. Plant Microbe In. 26:918-926.

Bayram, O., and Braus, G. H. 2012. Coordination of secondary metabolism and development in fungi: the velvet family of regulatory proteins. FEMS Microbiol. Rev. 36:1-24.

Belton, J. M., and Dekker, J. 2015. Hi-C in budding yeast. Cold Spring Harb. Protoc. 2015:649-661.

Borodovsky, M., and Lomsadze, A. 2011. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Curr. Protoc. Bioinf. 4:4.6.1–4.6.10.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18:188–196.

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S. and Aiden, E. L. 2016a. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3:99-101.

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., and Aiden, E. L. 2016b.

Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 3:95-98.

Edel-Hermann V. and Lecomte, C. 2019. Current status of *Fusarium oxysporum* formae speciales and races. Phytopathology. 109:512-530.

He, C. Y., Hsiang, T., and Wolyn, D. J. 2002. Induction of systemic disease resistance and pathogen defense responses in *Asparagus officinalis* inoculated with nonpathogenic strains of *Fusarium oxysporum*. Plant Pathol. 51:225-230.

Kim, Y., Heo, I. B., Yu, J. H., and Shin, K. S. 2017. Characteristics of a Regulator of G-Protein Signaling (RGS) rgsC in *Aspergillus fumigatus*. Front. Microbiol. 8:2058.

Kistler, H. C. 1997. Genetic Diversity in the Plant-Pathogenic Fungus *Fusarium oxysporum*. Phytopathology 87:474-479.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S. L. 2004. Versatile and open software for comparing large genomes. Genome Biology 5:R12

Ma, L.-J., Geiser, D. M., Proctor, R. H., Rooney, A. P., O'Donnell, K., Trail, F., Gardiner, D. M., Manners, J. M., and Kazan, K. 2013. *Fusarium Pathogenomics*. Annu. Rev. Microbiol. 67:399-416.

Ma, L.-J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P. M., Kang, S., Shim, W.-B., Woloshuk, C., Xie, X., Xu, J.-R., Antoniw, J., Baker, S. E., Bluhm, B. H., Breakspear, A., Brown, D. W., Butchko, R. A. E., Chapman, S., Coulson, R., Coutinho, P. M., Danchin, E. G. J., Diener, A., Gale, L. R., Gardiner, D. M., Goff, S., Hammond-Kosack, K. E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C. D., Koehrsen, M., Kumar, L., Lee, Y.-H., Li, L., Manners, J. M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.-Y., Proctor, R. H., Regev, A., Ruiz-Roldan, M. C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D. C., Turgeon, B. G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C. A., Kistler, H. C., and Rep, M. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464:367–373.

Marbouty, M., Cournac, A., Flot, J. F., Marie-Nelly, H., Mozziconacci, J., and Koszul, R. 2014. Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. eLife 3:e03318.

Marcais, G., and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764-770.

Mario, S., and Burkhard, M. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33:W465-467.

Michielse, C. B., and Rep, M. 2009. Pathogen profile update: *Fusarium oxysporum*. Mol. Plant Pathol. 10:311-324.

Paparu, P., Dubois, T., Coyne, D., and Viljoen, A. 2007. Defense-related gene expression in susceptible and tolerant bananas (*Musa* spp.) following inoculation with non-pathogenic *Fusarium oxysporum* endophytes and challenge with *Radopholus similis*. Physiol. Mol. Plant Pathol. 71:149-157.

Pieterse, C. M., Zamioudis, C., Berendsen, R. L., Weller, D. M., Van Wees, S. C., and Bakker, P. A. 2014. Induced systemic resistance by beneficial microbes. Annu. Rev. Phytopathol. 52:347-375.

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.

Tarailo-Graovac, M., and Chen, N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinf. 5: 4.10.1–4.10.14

Van der Ent, S., Van Wees, S. C. M., and Pieterse, C. M. J. 2009. Jasmonate signaling in plant interactions with resistance-inducing beneficial microbes. Phytochemistry 70:1581-1588.

Veloso, J., and Diaz, J. 2012. *Fusarium oxysporum* Fo47 confers protection to pepper plants against *Verticillium dahliae* and *Phytophthora capsici*, and induces the expression of defense genes. Plant Pathol. 61:281-288.

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., and Schatz, M. C. 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33:2202-2204.

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

Table 1. Summary of the genome assembly and annotation statistics of *Fusarium oxysporum* Fo47 compared to a previous assembly of Fo47 (GenBank accession: GCA_000271705.2) and *F. oxysporum* f. sp. *lycopersici* Fol4287 (GenBank accession: GCA_000149955.2) (Ma *et al.* 2010).

Strains	Fo47 (this study)	Fo47 (GCA_000271705.2)	Fol4287[b]
Number of contigs	12	124	88
Number of chromosomes	12[a]	NA[c]	15
Number of lineage-specific chromosomes	1	NA[c]	4
Genome size (Mb)	50.36	49.67	61.47
GC content (%)	47.71	47.68	48.48
Contig N50 (bp)	4,523,040	3,844,136	4,589,937
Number of protein coding genes[d]	17,234	17,055	17,637
Repeat content (%)	3.42	2.77	5.43
Genome assembly completeness (%) [e]	98	96	99
Predicted proteome completeness (%) [e]	99	99	98

a. Number of putative centromeres was determined using Hi-C contact map.

b. Reported by Ma *et al.* (2010).

c. NA: unavailable

d. Protein-coding genes were identified by the same pipeline (this study) to ensure a fair comparison.

e. Completeness = Total BUSCOs - Missing BUSCOs / Total BUSCOs * 100