Genome Sequence of *Fusarium oxysporum* f. sp. *matthiolae,* a *Brassicaceae* Pathogen

Houlin Yu,[1,3] Dilay Hazal Ayhan,[2,3] Andrew C. Diener,[4,†] and Li-Jun Ma[1,2,3,†]

[1]*Plant Biology Graduate Program, University of Massachusetts Amherst, Amherst, Massachusetts 01003, U.S.A.*

[2]*Molecular and Cellular Biology Graduate Program, University of Massachusetts Amherst, Amherst, Massachusetts 01003, U.S.A.*

[3]*Department of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, Massachusetts 01003, U.S.A.*

[4]*Department of Molecular Biology and Centre for Computational and Integrative Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, U.S.A.*

[†]Corresponding author: L.-J. Ma lijun@biochem.umass.edu; A. Diener diener@molbio.mgh.harvard.edu

Abstract:

The filamentous fungus *Fusarium oxysporum* is a soil-borne pathogen of many cultivated species and an opportunistic pathogen of humans. *F. oxysporum* f. sp. *matthiolae* is one of three formae speciales that are pathogenic to crucifers, including *Arabidopsis thaliana*, a premier model for plant molecular biology and genetics. Here, we report a genome assembly of *F. oxysporum* f. sp. *matthiolae* strain PHW726, generated using a combination of PacBio and Illumina sequencing

technologies. The genome assembly presented here should facilitate in-depth investigation of *F. oxysporum*-Arabidopsis interactions and shed light on the genetics of fungal pathogenesis and plant immunity.

Keywords

genome, *Fusarium oxysporum* f. sp. *matthiolae*, *Arabidopsis thaliana*, model, microbe-plant molecular interaction

Genome Announcement

Strains of the filamentous fungus *Fusarium oxysporum*, a notorious plant pathogen, can infect hundreds of cultivated species and at the same time have distinct host-specificity (Kistler 1997; Michielse and Rep 2009; Ma et al. 2013). This host-specificity is used to classify *F. oxysporum* into formae speciales, and a forma specialis typically represents one to a few monophyletic clonal lineages that cause disease in a narrow range of taxonomically related plants (Kistler 1997). At the genomic level, host-specificity corresponds to the presence of lineage-specific chromosomes (Ma et al. 2010). However, little is known about molecular mechanism involved in these host-specific plant-fungal interactions. One of the three formae speciales that are pathogenic to the crucifer *Arabidopsis thaliana* (Diener and Ausubel 2005; Provart et al. 2016), the genome sequence of *F. oxysporum* forma specialis *matthiolae* will enable the genetic analysis of fungal pathogenesis and host immunity using the model plant *Arabidopsis thaliana.*

Phylogenic analyses indicate that Fom isolates form a single clonal lineage (Bosland and Williams 1987; O'Donnell et al. 2009; Kistler and Benny 1989; Kistler et al. 1987; Kistler et al. 1991), although two races of Fom are distinguished by the differential susceptibility of varieties of *M. incana* (Bosland and Williams 1988). Natural variation of immunity is observed among wild accessions or ecotypes of *A. thaliana* toward Fusarium wilt (Diener and Ausubel 2005). Quantitative trait loci (QTLs) mapping in offspring of crosses between resistant and susceptible ecotypes has identified three *RESISTANCE TO F. OXYSPORUM* (*RFO*) genes, one receptor-like protein gene (*RFO2*) and two receptor-like kinase (RLK) genes (*RFO1* and *RFO3*) from different RLK gene subfamilies (Diener and Ausubel 2005; Diener 2013; Shen and Diener 2013; Cole and Diener 2013). As receptor-mediated immunity is reported to be the major determinant of disease resistance to Fom (Cole and Diener 2013), investigation of the interaction of Fom and *A. thaliana* should lead to a fundamental understanding of receptor-mediated plant immunity, especially against fungal pathogens. The genome sequence described here comes from DNA purified from Fom race 2, isolated from wilted garden stock (*Matthiola incana*), a cultivated plant in the crucifer or mustard (*Brassicaceae*) family, prized for its colorful flowers (Baker 1948; Tatsuzawa et al. 2012). This strain was previously deposited in American Type Culture Collection (ATCC 16603) by GM Armstrong and subsequently designated by PH Williams as PHW726 (Kistler et al. 1987).

The pipeline for genome assembly was adapted from Ayhan et al. 2018. Genomic DNA was purified from the mycelium of PHW726, and then sequenced by Illumina MiSeq and PacBio RS II platforms with 119× and 21× coverage, respectively. We used MiSeq paired-end sequencing with 150 cycles. The maximum size of the PacBio RS II reads was 59 kb while the mean size was 8.5 kb.

Trimmomatic version 0.32 (Bolger et al. 2014) was used to remove adaptors and trim ends of Illumina reads (parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). FastQC (version 0.11.5) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to check the quality of all reads. SPAdes version 3.9.1 (Antipov et al. 2016, using default parameters) was used to combine PacBio subreads and trimmed Illumina reads into an initial hybrid assembly. BWA version 0.7.12 (Li and Durbin 2009) was used to map the Illumina reads to the assembly. Further cleaning, fixing, and sorting of mapping reads was done with Picard version 2.0.1 (http://broadinstitute.github.io/picard/) and Samtools version 1.3 (Li et al. 2009). A structural variant (SV) caller, GRIDSS version 1.4.1 (Cameron et al. 2017) was used to identify links between scaffolds in the initial assembly. A custom script (available at github.com/d-ayhan/tools/scaffolding.m) was used for scaffolding. Minimap2 version 2.17 (Li 2018) was used to map PacBio subreads to new scaffolds, and links were manually inspected and, if necessary, fixed. Further polishing was performed by re-mapping Illumina reads to the assembly, during which FreeBayes v0.9.10-3-g47a713e (Garrison and Marth 2012) was used to identify base variants between reads and the assembly (specially, 70% support of minimal 10 alternate counts, with a minimal base mapping-quality greater than q30). Identified variants were used to correct the assembly by a custom script (available at github.com/d-ayhan/tools/FASTAeditWithVCF.m). RepeatMasker 4.0.5 (Tarailo-Graovac and Chen 2009) was used to screen the repeats. Mummer 3.22 (Kurtz et al. 2004) was used to align the assembly with the reference genome assembly for the tomato pathogen *F. oxysporum* f. sp. *lycopersici* Fol4287 (Ma et al. 2010).

As summarized in Table 1, the final assembly was 57.3 Mb in total length and comprised of 583 scaffolds with an N_{50} value of 0.77 Mb. The largest scaffold size was 3.6 Mb. The GC content was 47.4%. The size of total interspersed repeats was 3.1 Mb, which accounted for 5.4% of the assembly. A comparison with Fol4287 assembly (Ayhan et al. 2018) suggested a larger assembly size and higher interspersed repeat content of PHW726. The size of sequence mapped to the core chromosomes of Fol4287, which including 66 scaffolds (defined as core scaffolds), was 43.8 Mb. The assembly also included a scaffold of 52,365 bp that captured the entire mitochondrial DNA. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession WJXY00000000. The version described in this paper is version WJXY01000000. This assembly for the genome of PHW726 should facilitate future molecular genetics and genomic studies. Candidate Fom genes that promote pathogenesis or elicit immune response in *A. thaliana* and *M. incana* can now be predicted, subcloned and genetically characterized.

Acknowledgements

We thank Dr. Ravi Ranjan at the UMass Amherst Genomic Center and Dr. Quan Zeng, at the Connecticut Agricultural Experiment Station, for their consistent support for our genome projects, and the MGHPCC for providing high-performance computing capacity for the genome assembly process.

Funding Resources

This work was supported by the National Science Foundation (IOS-1652641). D.H.A. and L.-J.M. are also supported by the Burroughs Welcome Foundation (grant 1014893) and National Institute of Health (NIH/NEI R01 EY030150).

Literature Cited

Antipov, D., Korobeynikov, A., McLean, J. S., and Pevzner, P. A. 2016. HYBRIDSPADES: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32:1009–1015

Ayhan, D. H., López-Díaz, C., Di Pietro, A., and Ma, L.-J. 2018. Improved Assembly of Reference Genome *Fusarium oxysporum* f. sp. *lycopersici* Strain Fol4287. Microbiol Resour Announc. 7:e00910-18.

Baker, K. F. 1948. Fusarium wilt of garden stock (*Mathiola incana*). Phytopathology 38:399-403

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120

Bosland, P. W., and Williams, P. H. 1987. An evaluation of *Fusarium oxysporum* from crucifers based on pathogenicity, isozyme polymorphism, vegetative compatibility, and geographic origin. Can. J. Bot. 65:2067-2073

Bosland, P. W., and Williams, P. H. 1988. Pathogenicity of Geographic Isolates of *Fusarium oxysporum* from Crucifers on a Differential Set of Crucifer Seedlings. J Phytopathol. 123:63-68

Cameron, D. L., Schröder, J., Penington, J. S., Do, H., Molania, R., Dobrovic, A., Speed, T. P., and Papenfuss, A. T. 2017. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res. 27:2050-2060

Cole, S. J., and Diener, A. C. 2013. Diversity in receptor-like kinase genes is a major determinant of quantitative resistance to *Fusarium oxysporum* f.sp. *matthioli*. New Phytol. 200:172-184

Diener, A. C. 2013. Routine mapping of Fusarium wilt resistance in BC1 populations of *Arabidopsis thaliana*. BMC Plant Biol. 13:171

Diener, A. C., and Ausubel, F. M. 2005. *RESISTANCE TO FUSARIUM OXYSPORUM 1*, a Dominant Arabidopsis Disease-Resistance Gene, Is Not Race Specific. Genetics 171:305-321

Garrison, E., and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907v2 [q-bio.GN]

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

Kistler, H. C., and Benny, U. 1989. The Mitochondrial Genome of *Fusarium oxysporum*. Plasmid 22:86-89

Kistler, H. C., Bosland, P. W., Benny, U., Leong, S., and Williams, P. H. 1987. Relatedness of Strains of *Fusarium oxysporum* from Crucifers Measured by Examination of Mitochondrial and Ribosomal DNA. Phytopathology 77:1289-1293

Kistler, H. C., Momol, E. A., and Benny, U. 1991. Repetitive Genomic Sequences for Determining Relatedness Among Strains of *Fusarium oxysporum*. Phytopathology 81:331-336

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

Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25: 1754-1760

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079

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

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

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

O'Donnell, K., Gueidan, C., Sink, S., Johnston, P. R., Crous, P. W., Glenn, A., Riley, R., Zitomer, N. C., Colyer, P., Waalwijk, C., van der Lee, T., Moretti, A., Kang, S., Kim, H.-S., Geiser, D. M., Juba, J. H., Baayen, R. P., Cromey, M. G., Bithell, S., Sutton, D. A., Skovgaard, K., Ploetz, R., Corby Kistler, H., Elliott, M., Davis, M., and Sarver, B. A. J. 2009. A two-locus DNA sequence database for typing plant and human pathogens within the *Fusarium oxysporum* species complex. Fungal Genetics and Biology 46:936-948

Provart, N. J., Alonso, J., Assmann, S. M., Bergmann, D., Brady, S. M., Brkljacic, J., Browse, J., Chapple, C., Colot, V., Cutler, S., Dangl, J., Ehrhardt, D., Friesner, J. D., Frommer, W. B., Grotewold, E., Meyerowitz, E., Nemhauser, J., Nordborg, M., Pikaard, C., Shanklin, J., Somerville, C., Stitt, M., Torii, K. U., Waese, J., Wagner, D., and McCourt, P. 2016. 50 years of Arabidopsis research: highlights and future directions. New Phytol. 209:921-944

Shen, Y., and Diener, A. C. 2013. *Arabidopsis thaliana RESISTANCE TO FUSARIUM OXYSPORUM 2* Implicates Tyrosine-Sulfated Peptide Signaling in Susceptibility and Resistance to Root Infection. PLoS Genet. 9:e1003525

Tarailo-Graovac, M., and Chen, N. 2009. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Current Protocols in Bioinformatics 5: 4.10.1–4.10.14

Tatsuzawa, F., Saito, N., Toki, K., Shinoda, K., and Honda, T. 2012. Flower Colors and their Anthocyanins in *Matthiola incana* Cultivars (Brassicaceae). J. Japan. Soc. Hort. Sci. 81:91-100

194 **Table 1**. Summary of the *F. oxysporum* f. sp. *matthiolae* strain PHW726 genome assembly and a

195 comparison with Fol4287 (Ayhan et al. 2018)

Variables	**Statistics**	
	PHW726	**Fol4287**
Assembly size (bp)	57,270,650	53,912,367
Core sequence size (bp)	43,818,233	42,239,438
Number of scaffolds	583	499
Number of core scaffolds	66	55
Size of largest scaffold (bp)	3,557,637	5,733,288
Interspersed repeat content	5.35%	4.21%
N_{50} (bp)	774,050	1,338,693
N_{90} (bp)	47,752	49,310
L_{50} (bp)	18	11
GC content	47.44%	47.68%

196