

Genome Sequences of *Verticillium dahliae* Defoliating Strain XJ592 and Nondefoliating Strain XJ511

Haiyuan Li,¹ Jichen Dai,¹ Jun Qin,¹ Wenjing Shang,¹ Jieyin Chen,² Li Zhang,³ Xiaofeng Dai,² Steven J. Klosterman,⁴ Xiangming Xu,⁵ Krishna V. Subbarao,^{6,†} Sanhong Fan,^{1,†} and Xiaoping Hu^{1,†}

¹ State Key Laboratory of Crop Stress Biology for Arid Areas, College of Plant Protection, Northwest A&F University, Yangling, Shaanxi 712100, China

² Laboratory of Cotton Disease, Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences, Beijing 100193, China

³ The Key Laboratory of Prevention and Control for Oasis Crop Disease, Shihezi University, Shihezi 832003, China

⁴ United States Department of Agriculture, Agricultural Research Service, Salinas, CA 93905, U.S.A.

⁵ Pest & Pathogen Ecology, NIAB EMR, East Malling, Kent ME19 6BJ, U.K.

⁶ Department of Plant Pathology, University of California Davis, c/o United States Agricultural Research Station, Salinas, CA 93905, U.S.A.

Funding

The study was partially sponsored by the National Natural Science Foundation of China (31371888) and the National Key Research and Development Program of China (2018YFE0112500).

Keywords

defoliation, genomics, fungus-plant interactions, nondefoliation, secreted protein

Abstract

Verticillium dahliae is a widely distributed soilborne pathogen that causes vascular wilt in more than 200 plant species. Defoliating and nondefoliating symptoms caused by the disease that result in either the loss or retention of leaves in infected plants, respectively, in hosts such as cotton, olive, and okra, divide the causal agent into defoliating and nondefoliating pathotypes. Our goal in this current work was to generate genome resources for the defoliating strain XJ592 and the nondefoliating strain XJ511 of *V. dahliae* isolated from cotton in China.

Verticillium dahliae Kleb. can infect more than 200 plant species including many high-value annual and perennial crops and landscape, fruit, and ornamental trees and shrubs (Bhat and Subbarao 1999; Inderbitzin and Subbarao 2014; Klosterman et al. 2009; Pegg and Brady 2002). *Verticillium* wilt results in billions of dollars in annual losses worldwide (Klosterman et al. 2009). Typical symptoms include wilting and vascular discoloration that typically appear only near the end of the disease cycle and often coincide with host maturity (Inderbitzin and Subbarao 2014). *V. dahliae* can survive in soil without hosts in the form of microsclerotia for more than 14 years. Control of *Verticillium* wilt is difficult and costly in all economically important crops, including cotton (Inderbitzin et al. 2011; Klosterman et al. 2009; Subbarao et al. 2007). Several *V. dahliae* genome sequences (approximately 35 Mb) have been publicly released (National Center for Biotechnology Information [NCBI] website). Detailed structural and comparative analyses of some of these *V. dahliae* genomes suggested eight chromosomes (Chen et al. 2018; de Jonge et al. 2012; Faino et al. 2015; Klosterman et al. 2011).

H. Li and J. Dai contributed equally to this research.

†Corresponding authors: K. V. Subbarao; kvsbarao@ucdavis.edu, S. Fan; shfan@nwsuaf.edu.cn, and X. Hu; xphu@nwsuaf.edu.cn

The author(s) declare no conflict of interest.

Accepted for publication 17 January 2020.

Table 1. Genome statistics of *Verticillium dahliae* strains XJ592 and XJ511

| Statistics | XJ592 | XJ511 |
|--|---------------|---------------|
| Assembly | | |
| Depth of PacBio sequence | 85x | 75x |
| Depth of Illumina sequence | 129x | 133x |
| Assembly size | 35,670,849 bp | 34,896,600 bp |
| Number of contigs | 12 | 11 |
| N50 length | 3,716,643 bp | 3,584,248 bp |
| Size of top eight contigs | 32,405,901 bp | 31,030,758 bp |
| GC content | 53.28% | 53.98% |
| BUSCO completeness | 98.90% | 98.96% |
| Repeat rate | 13.92% | 9.36% |
| Annotation | | |
| Predicted protein-coding genes | 10,305 | 10,462 |
| Annotated proteins with the non-redundant database | 10,290 | 10,439 |
| Annotated proteins with Swiss-Prot | 6,887 | 6,941 |
| Unique proteins | 417 | 551 |
| Secreted proteins | 793 | 794 |
| Unique secreted proteins | 47 | 44 |
| Secreted carbohydrate active enzymes | 267 | 267 |
| Pathogen-host interaction (PHI) proteins | 1,136 | 1,134 |
| Unique PHI proteins | 29 | 27 |

Verticillium wilt caused by *V. dahliae* is a major impediment to cotton production. Based on the symptoms caused by *V. dahliae* on cotton, strains can be classified as defoliating or nondefoliating (Hu et al. 2015). While defoliating strains cause total defoliation in infected plants, leaves on plants infected by nondefoliating strains remain intact (Hu et al. 2015).

To generate genome resources for defoliating and nondefoliating strains of *V. dahliae*, we sequenced and assembled genomes of two *V. dahliae* strains (XJ592 and XJ511), isolated from diseased cotton plants in Xinjiang, China. Koch's postulates conducted previously demonstrated that both were pathogens on cotton, and the strains showed typical defoliating and nondefoliation symptoms, respectively (Hu et al. 2015).

Both strains were cultivated on PDA (potato dextrose agar) medium at 25°C in the dark for 7 days. DNA was extracted with the CTAB method (Möller et al. 1992). DNA for sequencing was sheared, using a Covaris g-TUBE (Covaris, Woburn, MA, U.S.A.). The DNA template prep kit 3.0, DNA/polymerase binding kit, and DNA sequencing reagent 4.0 were used for library construction and sequencing on a PacBio RS II platform. To ensure high-quality reads for use in assembling the genome, the adapter sequence, low-quality reads, and N bases were initially filtered out from sequences before genome assembling. Sequence filtering and de novo assembling were carried out with MECAT v1.0. To improve PacBio assemblies, further sequences were obtained on the Illumina HiSeq platform, following manufacturer protocols (Illumina, San Diego, CA, U.S.A.). Illumina HiSeq results were used to correct errors in the assembled PacBio sequences with BCFtools v1.9 (23,520 loci from PacBio-assembled genome of XJ592 were corrected, and 30,142 loci from PacBio-assembled genome of XJ511 were corrected).

The completeness of assembly was verified with BUSCO v3.1.0, based on lineage-specific profile library fungi_odb9 (Simão et al. 2015), which is based on a set of 1,438 common fungal genes as benchmark universal single-copy orthologs. Results showed that the genome assembly completeness of XJ592 and XJ511 were predicted as 98.90 and 98.96%, respectively (Table 1). The final genome assemblies of XJ592 and XJ511 consist of 12 contigs (N50 length of 3,716,643 bp) and 11 contigs (N50 length of 3,584,248 bp), respectively. Total length of the eight longest contigs was about 32.41 Mb (90.85% of genome size) for XJ592 and about 31.03 Mb (88.92% of genome size) for XJ511. These results suggest the assembly for both XJ592 and XJ511 is close to the chromosome level. Repetitive sequences of both strains were analyzed by de novo-based and homology-based methods. To detect de novo repeats, the RepeatModeler open-1.0.11 pipeline was used with default settings (Saha et al. 2008). After filtering the RepeatModeler repeat library by removal of repetitive gene sequences, the filtered RepeatModeler repeat library was used for repeat sequence prediction and categorization with RepeatMasker open-4.0.6. In parallel, tandem repeat sequences were predicted de novo with Tandem Repeats Finder. A homology-based repeat element search was conducted by RepeatMasker and RepeatProteinMask based on

the RepBase fungal library (built on January 27, 2017). After the repeat sequence analyses were completed, the results were combined and analyzed. About 13.92% of the XJ592 sequences and about 9.36% of the XJ511 sequences were predicted as repetitive sequences. Lineage-specific region G-LSR2, which is diagnostic of the defoliating *V. dahliae* strains (Chen et al. 2018), is also present in the genome of defoliating strain XJ592 at the contig named `tig00000126_len = 3716643_reads = 24420`.

For protein-coding gene prediction, GeneMark-ES v4.46 (Ter-Hovhannisyan et al. 2008) and Augustus v3.3.2 (Stanke and Morgenstern 2005) were trained for use with MAKER v2.31.10 (Holt and Yandell 2011). GeneMark-ES was trained on the genomes of XJ592 and XJ511 and the outputs were then used to train Augustus. The gene model of VdLs.17 obtained from NCBI was also used by MAKER. Total of 10,305 and 10,462 protein-coding genes were predicted in the XJ592 and XJ511 assemblies, respectively. Functional annotation of predicted genes was conducted based on Swiss-Prot and non-redundant proteins (NR), Swiss-Prot annotation identified 6,887 genes in XJ592 and 6,941 genes in XJ511 with functional information; the corresponding values were 10,290 and 10,439 for NR annotation. For predicted protein-coding genes, there were 417 unique genes in XJ592 and 551 unique genes in XJ511 with Blastp v2.9.0+ (pident < 50%, pcovs < 50%, e-value > 1e-7). Putative extracellular proteins were identified as secreted proteins if they contain a signal peptide, lack transmembrane domains, and locate at the extracellular subcellular. Secreted proteins were predicted with SignalP v5.0, transmembrane domains with TMHMM v2.0, and subcellular localizations with WoLF PSORT. A total of 793 and 794 genes were predicted to encode secreted proteins in XJ592 and XJ511, respectively. BlastP with those predicted secreted proteins was used to identify unique secreted proteins between XJ592 and XJ511, resulting in 47 unique secreted proteins in XJ592 and 44 in XJ511. The putative secreted carbohydrate active enzymes (CAZymes) were annotated using dbCAN2 (Zhang et al. 2018), revealing a total of 267 genes encoding CAZymes in both XJ592 and XJ511. Homologs of known pathogen-host interaction (PHI) genes were predicted using the PHI database v4.8 (Urban et al. 2017). A total of 1,136 genes from XJ592 and 1,134 genes from XJ511 were predicted as PHI genes (pident > 50%, pcovs > 50%, e-value < 1e-7). BlastP with these protein sets identified 29 unique PHI proteins in XJ592 and 27 in XJ511.

Since XJ592 and XJ511 represent the typical defoliating and nondefoliating pathotypes of *V. dahliae* on cotton, respectively, the current genome assemblies together with corresponding annotation information are helpful resources for research to understand the mechanisms underlying disease symptom development due to *V. dahliae* infection of cotton plants.

The Whole Genome Shotgun project of XJ592 has been deposited at DDBJ/ENA/GenBank under accession number WJDY00000000 (BioProject PRJNA577186; BioSample SAMN13021461); the version described here is version WJDY01000000. The Whole Genome Shotgun project of XJ511 has been deposited at DDBJ/ENA/GenBank under accession number WJDV00000000 (BioProject PRJNA577198; BioSample SAMN13021591); the version described here is version WJDV01000000.

Author-Recommended Internet Resources

NCBI Verticillium genome assembly: <https://www.ncbi.nlm.nih.gov/genome/genomes/832>
PHI database v4.8: <http://www.phi-base.org>
SignalP v5.0 server: <http://www.cbs.dtu.dk/services/SignalP>
TMHMM v2.0 server: <http://www.cbs.dtu.dk/services/TMHMM>
WoLF PSORT prediction tool: <https://wolfsort.hgc.jp>

Literature Cited

- Bhat, R. G., and Subbarao, K. V. 1999. Host range specificity in *Verticillium dahliae*. *Phytopathology* 89:1218-1225.
- Chen, J. Y., Liu, C., Gui, Y. J., Si, K. W., Zhang, D. D., Wang, J., Short, D. P. G., Huang, J. Q., Li, N. Y., Liang, Y., Zhang, W. Q., Yang, L., Ma, X. F., Li, T. G., Zhou, L., Wang, B. L., Bao, Y. M., Subbarao, K. V., Zhang, G. Y., and Dai, X. F. 2018. Comparative genomics reveals cotton-specific virulence factors in flexible genomic regions in *Verticillium dahliae* and evidence of horizontal gene transfer from *Fusarium*. *New Phytol.* 217:756-770.
- de Jonge, R., van Esse, H. P., Maruthachalam, K., Bolton, M. D., Santhanam, P., Saber, M. K., Zhang, Z., Usami, T., Lievens, B., Subbarao, K. V., and Thomma, B. P. H. J. 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 109:5110-5115.
- Faino, L., Seidl, M. F., Datema, E., van den Berg, G. C. M., Janssen, A., Wittenberg, A. H. J., and Thomma, B. P. H. J. 2015. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. *MBio* 6:e00936-15.

- Holt, C., and Yandell, M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
- Hu, X. P., Gurung, S., Short, D. P. G., Sandoya, G. V., Shang, W. J., Hayes, R. J., Davis, R. M., and Subbarao, K. V. 2015. Nondefoliating and defoliating strains from cotton correlate with races 1 and 2 of *Verticillium dahliae*. *Plant Dis.* 99: 1713-1720.
- Inderbitzin, P., Bostock, R. M., Davis, R. M., Usami, T., Platt, H. W., and Subbarao, K. V. 2011. Phylogenetics and taxonomy of the fungal vascular wilt pathogen *Verticillium*, with the descriptions of five new species. *PLoS One* 6:e28341.
- Inderbitzin, P., and Subbarao, K. V. 2014. *Verticillium* systematics and evolution: How confusion impedes *Verticillium* wilt management and how to resolve it. *Phytopathology* 104:564-574.
- Klosterman, S. J., Atallah, Z. K., Vallad, G. E., and Subbarao, K. V. 2009. Diversity, pathogenicity, and management of *Verticillium* species. *Annu. Rev. Phytopathol.* 47:39-62.
- Klosterman, S. J., Subbarao, K. V., Kang, S., Veronese, P., Gold, S. E., Thomma, B. P. H. J., Chen, Z., Henrissat, B., Lee, Y.-H., Park, J., Garcia-Pedrajas, M. D., Barbara, D. J., Anchieta, A., de Jonge, R., Santhanam, P., Maruthachalam, K., Atallah, Z., Amyotte, S. G., Paz, Z., Inderbitzin, P., Hayes, R. J., Heiman, D. I., Young, S., Zeng, Q., Engels, R., Galagan, J., Cuomo, C. A., Dobinson, K. F., and Ma, L. J. 2011. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog.* 7:e1002137.
- Möller, E. M., Bahnweg, G., Sandermann, H., and Geiger, H. H. 1992. A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Res.* 20:6115-6116.
- Pegg, G. F., and Brady, B. L. 2002. *Verticillium* wilts. CABI, Wallingford, U.K.
- Saha, S., Bridges, S., Magbanua, Z. V., and Peterson, D. G. 2008. Empirical comparison of ab initio repeat finding programs. *Nucleic Acids Res.* 36: 2284-2294.
- 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.
- Stanke, M., and Morgenstern, B. 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33 (Web Server):W465-W467.
- Subbarao, K. V., Kabir, Z., Martin, F. N., and Koike, S. T. 2007. Management of soilborne diseases in strawberry using vegetable rotations. *Plant Dis.* 91: 964-972.
- Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* 18:1979-1990.
- Urban, M., Cuzick, A., Rutherford, K., Irvine, A., Pedro, H., Pant, R., Sadanadan, V., Khamari, L., Billal, S., Mohanty, S., and Hammond-Kosack, K. E. 2017. PHI-base: A new interface and further additions for the multi-species pathogen-host interactions database. *Nucleic Acids Res.* 45 (D1):D604-D610.
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. 2018. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46 (W1):W95-W101.