Genome Sequence Resource for the Ramie Oomycete Pathogen *Phytophthora vexans* HF1

Yongting Yu,¹ Chunsheng Gao,¹ Tuhong Wang,¹ Yikun Chen,¹ Yi Cheng,¹ Zhimin Li,¹ Jia Chen,¹ Litao Guo,¹ Xiangping Sun,¹ and Jianping Xu^{1,2,†}

¹ Institute of Bast Fiber Crops and Center for Southern Economic Crops, Chinese Academy of Agricultural Science, Changsha 410205, China

² Department of Biology, McMaster University, Hamilton, L8S 4K1, Canada

Abstract

The oomycete *Phytophthora vexans* is a causative agent of patch canker, damping-off, and crown, stem, and root rot in many economically important plants. *P. vexans* HF1 was isolated in China, where it caused brown root rot of ramie, a fiber crop broadly cultivated in Asia. The genome of HF1 was sequenced by a combination of technologies producing short (Illumina HiSeq X) and long (PacBio RS) reads. The genome is 41.73 Mbp long, assembled into 44 contigs. It has a GC content of 58.17% and contains 13,051 predicted coding genes, including 1,461 putative virulence genes and 220 putative antimicrobial resistance genes. This genome sequence provides a resource for determining the molecular mechanisms of disease development in this pathosystem.

Funding

This study was supported by the Chinese Academy of Agricultural Sciences Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2015-IBFC), the Project for Monitoring and Prevention of Crop Pests, Disease, and Mice from the Ministry of Agriculture (S158), and the Natural Science Foundation of Hunan Province (2017JJ2298).

Keywords

brown root rot disease, oomycete pathogen, ramie

Genome Announcement

Phytophthora vexans (de Barry) Abad, de Cock, Bala, Robideau, Lodhi & Lévesque [syn. *Pythium vexans* (de Barry)] is an important oomycete phytopathogen that can cause patch canker, damping-off, and crown rot, stem rot, and root rot in many economically important plants, such as durian, rubber trees, and citrus fruits (Adhikari et al. 2013; de Cock et al. 2015). Yu et al. (2016) reported that *P. vexans* strain HF1 caused brown root rot disease of ramie (*Boehmeria nivea* L. Gaud), a perennial flowering plant in the nettle family *Urticaceae*. Native to Asia, ramie is one of the oldest textile fiber crops and currently among the most important long-fiber crops in the world (Liu et al. 2013b). For example, it was used in mummy cloths in Egypt during the period from 5000 to 3300 B.C. Ramie has been cultivated in China for over 6,000 years and is now commercially grown in many countries, including Brazil, Japan, the Philippines, India, Thailand, and China (Liu et al. 2013b). Aside from its long stem fiber, other components of ramie are also of significant economic value, including a high level of crude protein content in leaves and shoots that make it an ideal source of feed for domesticated animals such as beef cattle, pigs, and geese (Kipriotis et al. 2015). Since 2013, brown root rot by *P. vexans* has caused up to 40% crop loss in ramie plantations in several regions in southern China (Yu et al. 2016). At present, the molecular mechanism of pathogenesis of

†Corresponding author: J. Xu; jpxu@mcmaster.ca

Yongting Yu is deceased.

*The *e-Xtra* logo stands for “electronic extra” and indicates there are supplementary tables published online.

The author(s) declare no conflict of interest.

Accepted for publication 9 August 2020.

Table 1. Summary statistics of the *Phytophthium vexans* HF1 genome assembly and comparison with strain CBS 119.80^a

Genome assembly characteristics	Strain HF1	Strain CBS 119.80
Genome size (Mbp)	41.73	33.85
No. of contigs	44	3,685
Largest contig (bp)	3,173,098	185,880
Contig N50 (bp)	1,512,378	29,240
Contig N90 (bp)	634,848	4,392 ^b
GC content	58.17%	58.85% ^b
BUSCO completeness	92.20%	90.60% ^b
No. of predicted genes	13,051	11,957
No. of predicted secreted proteins	1,261	834
No. of CAzyme genes	352	203, 373 ^c
No. of virulence genes	1,461	616
No. of effectors	239	12

^a CBS 119.80 statistics were previously published by Adhikari et al. (2013).

^b These data were not reported in the original study by Adhikari et al. (2013), but were obtained when we reanalyzed their assembly.

^c Reported by Adhikari et al. (2013) in their Supplementary Table S6 and by Zerillo et al. (2013) in their Table 2, respectively, based on the same dataset of strain CBS 119.80.

P. vexans on host ramie plants is not known. Having the genomic sequence of *P. vexans* HF1 will facilitate our understanding of this pathogen, including its interactions with host ramie plants.

A previous study reported the genome sequence of the *P. vexans* strain CBS 119.80 isolated from soil in Iran (Adhikari et al. 2013). Its genome was sequenced by the Illumina Genome Analyzer II platform, resulting in a total of 55,618,622 clean reads with a trimmed length of 51 bp per read. With a cutoff at 1.4× coverage and a minimum contig length of 500 bp as the assembly threshold, 1.93 Gb of their 2.83-Gb clean sequence was usable to assemble the genome into 3,685 contigs, with the longest contig length at 185.88 kb and an N50 of 29.24 kb. The genome size for strain CBS 119.80 was estimated at 33.85 Mb, containing 11,957 protein-coding genes. This estimated genome size and gene number for *P. vexans* CBS 119.80 were the smallest among all the oomycete phytopathogens sequenced so far (Adhikari et al. 2013; Yang et al. 2018). These and additional genome annotation statistics of CBS 119.80 are shown in Table 1.

At present, it is not known whether the small genome size and gene number of strain CBS 119.80 is typical of *P. vexans*. In addition, we are interested in the pathogenicity-related genes in a *P. vexans* strain pathogenic to ramie to help further research on the *P. vexans*–ramie pathosystem. To achieve the objectives, we sequenced strain *P. vexans* HF1, originally isolated as the causal agent of ramie root rot in Yuanjiang, Hunan province, China in 2013 (Yu et al. 2016). For genomic DNA extraction, strain HF1 was first cultured on potato dextrose agar medium amended with PARP (Manici et al. 2004). Cells were harvested and genomic DNA was extracted using the Rapid Extraction Kit for Fungi Genomic DNA (Aidlab, Beijing, China). Two genomic libraries were prepared, one containing insert fragments of approximately 440 bp and the second containing fragments of approximately 10 kb. The approximately 440-bp insert library was used for paired-end sequencing (2 × 150 bp) on an Illumina HiSeq X machine, from which we obtained 17,817,768 (×2) paired-end reads for a total of 5,368,817,374 bp sequences. The approximately 10-kb insert library was sequenced by the PacBio RS, resulting in 416,311 reads with an average read length of 10,081 bp and a summed length of 4,197,028,446 bp. The long-read length from the PacBio sequencing coupled with the high coverage of Illumina sequencing have become the standard for assembling accurate and high-quality genome sequences (Maio et al. 2019). Here in our genome sequencing of HF1, library construction and sequencing were performed at Shanghai Majorbio Co., Ltd., China, following standard protocols. After quality assessment and data filtering, we obtained 4.31 Gb (103× coverage) and 3.01 Gb (72× coverage) of clean data, respectively, from the Illumina and PacBio sequencing. The PacBio data were used to generate the initial assembly, using Canu v2.0 with adaptive K-mer weighting and repeat separation (Koren et al. 2017). The initial assembly, including the nucleotide sequence, were confirmed and corrected, if necessary, using the Illumina HiSeq data assembled by SOAPdenovo v2.04 (Luo et al. 2012). Our final assembled genome for strain HF1 was 41.73 Mb long, distributed among 44 contigs. The estimated genome size and contig number based on kn, ksp, and Genomeye methods at the optimal K-mer 17 were consistent with each other (detailed data not shown) (Liu et al. 2013a). The genome-wide heterozygosity was estimated

at approximately 0.04% by using KmerFreq_v1.0 of SOAPdenovo2 (Luo et al. 2012), and we were unable to identify any influence of the observed heterozygosity on our estimates of genome size and contig number. The largest assembled contig was 3,173,098 bp, contig N50: 1,512,378 bp, N90: 634,848 bp, and GC content: 58.17%. The assembly completeness was estimated to be 92.20% according to the BUSCO v4.0 (Simão et al. 2015) program. Further genetic crosses and linkage mapping are needed to combine the assembled contigs into chromosomes.

A total of 13,051 protein-coding genes were predicted by the MAKER2 pipeline for the HF1 genome. The predicted protein-coding genes were compared with the following databases: Nr, Genes, STRING, and gene ontology (GO) via BLASTp (BLAST 2.2.28+) to obtain their GO terms, and EuKaryotic Orthologous Groups and Kyoto Encyclopedia of Genes and Genomes functional classifications (von Mering et al. 2003). Pfam domains were assigned to 8,770 of these genes, and a total of 1,261 secreted-protein signals were identified by using SignalP v5.0. All genes with transmembrane domains were identified via TMHMM v2.0. We identified seven ribosomal RNA and 777 transfer RNA genes via barrnap v0.8 and tRNAscan v2.0, respectively (Lowe and Chan 2016). A total of 352 genes showed significant matches (E -value $< 1E-5$) to those in the CAzyme database (Park et al. 2010), including those with auxiliary activities, carbohydrate-binding modules, carbohydrate esterases, glycoside hydrolases, glycosyl transferases, and polysaccharide lyases (Supplementary Table S1). Comparisons with the fungal and oomycete virulence factors database via Diamond V0.8.35 (Lu et al. 2012) revealed 1,461 putative virulence genes. By comparing with the published literature, 28 of the 1,461 putative virulence genes have been confirmed as virulence factors in oomycetes (Yang et al. 2018) (Supplementary Table S2). Based on features related to signal peptide, transmembrane domain, cysteine number, and protein length, 239 of these 1,461 were identified as candidate effectors (Supplementary Table S3). In oomycete pathogens, effectors of the RxLR, YxSL[RK], CRN, and NLP types have attracted wide attention (Adhikari et al. 2013; Chen et al. 2018). Here, we investigated putative effectors of these four categories in the *P. vexans* HF1 genome. While no candidate effector of the RxLR type was found in strain HF1, our analyses revealed nine putative effectors containing the YxSL[RK] motif, seven of which were individually associated with a signal peptide. Of the CRN type, 17 were predicted in the HF1 genome, one of which contained a signal peptide. Finally, two effectors of the NLP type were predicted (Supplementary Table S4). In addition, comparison with the Comprehensive Antibiotic Resistance Database identified 220 putative antibiotic resistance genes for *P. vexans* HF1 (Supplementary Table S5).

The obtained HF1 genome is about 7.88 Mb larger and contains 1,094 more predicted genes than the genome previously assembled for *P. vexans* CBS 119.80 (Table 1). The corresponding annotated protein information between these two genomes was analyzed and presented in Supplementary Table S6. The obtained genome size and gene content of HF1 are similar to those of other oomycete phytopathogens reported earlier (Adhikari et al. 2013; Yang et al. 2018). Overall, due to the combined high-coverage Illumina HiSeq and PacBio sequencing of strain HF1, we believe that the HF1 genome assembly presented here is a more complete and accurate description of the *P. vexans* genome than that published previously for strain CBS 119.80. However, we cannot exclude the possibility that the smaller genome size and the fewer number of predicted genes for strain CBS 119.80 may be due to the intrinsic differences between the two strains, potentially related to differences in their geographic (Iran versus China) and ecological (soil versus ramie root) origins and evolutionary divergence (Supplementary Table S6). Indeed, the average amino acid sequence identity between the two strains for the 11,847 homologous proteins was 87.54%, suggesting that these two strains could be classified into different genome species (Xu 2020). Among the 11,847 homologous proteins between the two strains, 7023 (59.3%) showed $>95\%$ sequence identity (E -values were all $< E-28$), 3,728 (31.4%) showed between 50 and 95% sequence identity (E -values were all $< E-11$), and 1,096 (9.3%) had between 20 and 50% sequence identity (E -values were all $< E-5$). The details of these three categories and individual genes are shown in Supplementary Table S6. However, a higher depth of sequencing for strain CBS 119.80 using both the Illumina platform and a long-read sequencing platform such as PacBio RS could help reveal one or more potential sources for the observed genomic differences between strains HF1 and CBS 119.80. Regardless, our data provide a rich resource to help guide future investigations into the diversity and interaction between this pathogen and its host ramie plants.

The genome sequence reported here has been deposited at DDBJ/ENA/GenBank (International Nucleotide Sequence Database Collaboration) under the accession number QLOC00000000.1 (BioProject PRJNA454536, BioSample SAMN09010073). The version described in this paper is version QLOC01000000 (Table 1), with accession numbers QLOC01000001.1 to QLOC01000004.1 corresponding to the 44 contigs.

Acknowledgments

We thank H. Yoell for proofreading.

Author-Recommended Internet Resources

barnap v0.8: <https://github.com/tseemann/barnap>

NCBI whole-genome shotgun sequencing project: <https://www.ncbi.nlm.nih.gov/nuccore/1461202298>

SignalP server: <http://www.cbs.dtu.dk/services/SignalP>

TMHMM server: <http://www.cbs.dtu.dk/services/TMHMM>

Literature Cited

- Adhikari, B. N., Hamilton, J. P., Zerillo, M. M., Tisserat, N., Lévesque, C. A., and Buell, C. R. 2013. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. *PLoS One* 8:e75072.
- Chen, X. R., Huang, S. X., Zhang, Y., Sheng, G. L., Li, Y. P., and Zhu, F. 2018. Identification and functional analysis of the NLP-encoding genes from the phytopathogenic oomycete *Phytophthora capsici*. *Mol. Genet. Genomics* 293: 931-943.
- de Cock, A. W. A. M., Lodhi, A. M., Rintoul, T. L., Bala, K., Robideau, G. P., Abad, Z. G., Coffey, M. D., Shahzad, S., and Lévesque, C. A. 2015. *Phytophthora*: Molecular phylogeny and systematics. *Persoonia* 34:25-39.
- Kipriotis, E., Xiong, H. P., Vafeiadakis, T., Kipriotis, M., and Alexopoulou, E. 2015. Ramie and kenaf as feed crops. *Ind. Crops Prod.* 68:126-130.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:gr.215087.
- Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., Li, Z., Chen, Y., Mu, D., Fan, W. 2013a. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *Quant. Biol.* 35:62-67.
- Liu, T., Zhu, S., Tang, Q., Chen, P., Yu, Y., and Tang, S. 2013b. De novo assembly and characterization of transcriptome using Illumina paired-end sequencing and identification of CesA gene in ramie (*Boehmeria nivea* L. Gaud). *BMC Genomics* 14:125.
- Lowe, T. M., and Chan, P. P. 2016. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44:W54-W57.
- Lu, T., Yao, B., and Zhang, C. 2012. DFVF: Database of fungal virulence factors. *Database (Oxford)* 2012:bas032.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S. M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T. W., and Wang, J. 2012. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18.
- Maio, N. D., Shaw, L. P., Hubbard, A., George, S., Sanderson, N. D., Swann, J., Wick, R., AbuOun, M., Stubberfield, E., Hoosdally, S. J., Crook, D. W., Peto, T. E. A., Sheppard, A. E., Bailey, M. J., Read, D. S., Anjum, M. F., Walker, A. S., and Stoesser, N. 2019. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microbial Genomics* 5: 000294.
- Manici, L. M., Caputo, F., and Babini, V. 2004. Effect of green manure on *Pythium* spp. population and microbial communities in intensive cropping systems. *Plant Soil* 263:133-142.
- Park, B. H., Karpinets, T. V., Syed, M. H., Leuze, M. R., and Uberbacher, E. C. 2010. CAZymes analysis toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZY database. *Glycobiology* 20:1574-1584.
- 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.
- von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. 2003. STRING: A database of predicted functional associations between proteins. *Nucleic Acids Res.* 31:258-261.
- Xu, J. 2020. Fungal species concepts in the genomics era. *Genome*:1-10.
- Yang, M., Duan, S., Mei, X., Huang, H., Chen, W., Liu, Y., Guo, C., Yang, T., Wei, W., Liu, X., He, X., Dong, Y., and Zhu, S. 2018. The *Phytophthora cactorum* genome provides insights into the adaptation to host defense compounds and fungicides. *Sci. Rep.* 8:6534.
- Yu, Y. T., Chen, J., Gao, C. S., Zeng, L. B., Li, Z. M., Zhu, T. T., Sun, K., Cheng, Y., Sun, X. P., Yan, L., Yan, Z., and Zhu, A. G. 2016. First report of brown root rot caused by *Pythium vexans* on ramie in Hunan, China. *Can. J. Plant Pathol.* 38: 405-410.
- Zerillo, M. M., Adhikari, B. N., Hamilton, J. P., Buell, C. R., Lévesque, C. A., and Tisserat, N. 2013. Carbohydrate-active enzymes in *pythium* and their role in plant cell wall and storage polysaccharide degradation. *PLoS One* 8: e72572.