Genome Sequence Resource of *Coniella vitis*, a Fungal Pathogen Causing Grape White Rot Disease

Shanyue Zhou,[†] and Baohua Li[†]

College of Plant Health and Medicine, The Key Lab of Integrated Crop Pests Management of Shandong Province, Qingdao Agricultural University, Qingdao, Shandong 266109, China

Funding

This research was supported by China Agriculture Research System (CARS-28).

Abstract

Several fungal pathogens cause grape white rot disease and *Coniella vitis* is a predominant pathogen in Chinese vineyards. The disease occurs on leaves, vines, and fruit berries, leading to considerable yield losses and even to total destruction of vineyards. Here, we present the first PacBio and Illumina-sequenced draft genome assembly of *C. vitis* QNYT13637 and its annotation. This genome sequence provides a unique resource that will be a powerful foundation for future research on exploring virulence-related genes, investigating the pathogenicity mechanism of the pathogen, and, finally, improvement of white rot disease management strategies.

Keywords

Coniella vitis, genome, grape white rot

Genome Announcement

White rot disease on grape (*Vitis vinifera* L. and *Vitis* spp.) occurs commonly worldwide. The disease occurs mostly in warm and rainy regions. The causal pathogens of the disease were reported as *Coniella diplodiella* (Speg.) Petr. & Syd. (Bisiach 1988; Crous and Carstens 2000; Crous et al. 2000; Liu et al. 1999; Van Niekerk et al. 2004), *C. fragariae* (Oudem) B. Sutton (Crous and Carstens 2000; Tiedemann 1985), *Pilidiella castaneicola* (He et al. 2017), and *C. vitis* (Chethana et al. 2017). In China, white rot disease is one of the main fungal diseases of grape, with at least 16% yield loss annually (Li et al. 2008), and *C. vitis* has been reported as the predominant causing agent (Chethana et al. 2017).

In this study, we present a high-quality draft genome sequence of a *C. vitis* strain, QNYT13637, providing preliminary reference genomic information of the fungus. The genomic DNA of the strain was extracted and sequenced using long-read PacBio (for a 20-kb DNA library) platforms with single molecule, real-time (SMRT) technology (Beijing Novogene Bioinformatics Technology Co., Ltd.). For the high-quality genome, the short-read Illumina PE150 (for the 350-bp fragments DNA library) sequencing was performed as well. Finally, the error-corrected genome sequences were generated by mapping Illumina short reads to the PacBio-generated assembly result.

The low-quality reads were filtered with SMRT Link v5.0.1 (Ardui et al. 2018; Reiner et al. 2018) and the filtered reads were assembled to contigs without gaps. A total of 192,376 reads were generated with a coverage of 46x. The reads were assembled into 22 contigs with $N_{50} = 3.20$ Mb, maximum contig length = 5.11 Mb. The final draft genome of the fungus is 41.54 Mb with 49.59% G+C content (Table 1). We used the Augustus 2.7 program to retrieve the related protein-coding genes for the fungus *C. vitis*. In total 7,985 protein-coding genes were

[†]Corresponding authors: S. Zhou; zhoushanyao@126.com and B. Li; baohuali@qau.edu.cn

The author(s) declare no conflict of interest.

Accepted for publication 13 March 2020.

Table 1. Genomics features of *Coniella vitis* QNYT13637

Variables	Statistics
Sequence coverage	46×
Assembly Size	41.54 Mb
Number of contigs	22
N ₅₀	3.20 Mb
GC content	49.59
Number of coding genes	7,985
Repeat rate	5.76%
BUSCO completeness	99.3%

identified, with a total length of 11,595,087 bp, constituting 27.91% of the total genome. The prediction of the interspersed repetitive sequences was performed with Repeat Masker v4.0.5 (Saha et al. 2008), and tandem repeats were detected with a tandem repeats finder program (TRF v4.07b) (Benson 1999). The repetitive sequences are a total 2.39 Mb in length, 5.76% of the whole genome, consisting of interspersed repetitive sequence (1.43 Mb, 3.44%) and the tandem repeats (0.96 Mb, 2.32%). Noncoding RNAs (ncRNAs) in the genome were predicted as follows: transfer RNAs (tRNA) were predicted with tRNAscan-SE (v1.3.1) (Lowe and Eddy 1997); ribosomal RNAs (rRNAs) were analyzed with rRNAmmer (v1.2) (Lagesen et al. 2007); small RNA, small nuclear RNA, and microRNA were analyzed by BLAST against the Rfam database (Finn et al. 2009; Gardner et al. 2009; Nawrocki et al. 2009). In total 32,693bp in the genome were transcribed to ncRNAs.

To annotate the genome, a whole-genome BLAST search was performed against protein databases. Among the predicted coding genes, 7,256 (90.87%), 6,814 (85.34%), 5,314 (66.55%), 2,906 (36.39%), and 1,931 (24.18%) genes had homologs with known functions in the nonredundant protein database, Kyoto Encyclopedia of Genes and Genomes (KEGG), gene ontology (GO), Swiss Uni-Prot, and COG (clusters of orthologous groups) databases, respectively. As a fungal plant pathogen, the genes that might be involved in the pathogenicity were analyzed. According to the best-matched proteins from the PHI (pathogen-host interactions) and DFVF (database of fungal virulence factors) databases, 1,120 (14.03%) proteins were considered to be involved in pathogen-host interactions and 400 (5.01%) proteins were predicted to act as virulence factors. Furthermore, 417 (5.22%) proteins were categorized as carbohydrate-active enzymes, of which 243 were identified as glycoside hydrolases. With the cut-off E value $\leq 1e^{-5}$ and alignment length percentage $\geq 40\%$, the annotations of the protein-coding genes were determined.

It is well-known that secretory proteins of pathogens, especially the effectors, play critical roles in fungal pathogenicity and virulence (González-Fernández et al. 2010; Stergiopoulos and de Wit 2009). For secretory protein prediction, the proteins were analyzed with SignalP v4.1 (Petersen et al. 2011) for the signal peptide and were scanned with TMHMM v. 2.0 for the transmembrane (TM) domains. The proteins with no TM or with one TM were then further analyzed with WoLF PSORT v0.2 software (Horton et al. 2007) for their subcellular locations. The proteins with a predicted extracellular score >15 were divided into secretomes. A total of 536 predicted proteins in the genome were thought to be credible secretory proteins. Secondary metabolites are also important virulence factors of fungal pathogens, such as mycotoxins (Goswami and Kistler 2005; Proctor et al. 1995). As such, putative genes and clusters for the secondary metabolites were predicted using the web-based tool antiSMASH 2.0 (Blin et al. 2013). As the result, 271 genes in 28 clusters were identified in the genome (32 genes in three nonribosomal peptide synthetase [NRPS] clusters, 18 genes in three terpene clusters, 133 genes in 13 polyketide synthase (PKS) clusters, 15 genes in one PKS-NRPS cluster, and 73 genes in eight other clusters).

This is the first report of the genome of *C. vitis*. It will be a preliminary foundation for investigating the mechanism of the pathogenicity of the fungus. This whole-genome shotgun project has been deposited in the DDBJ/ENA/GenBank database under accession number WUJK00000000 (PRJNA597685 for BioProject and SAMN13679937 for BioSample). The version described in this paper is version WUJK01000000.

Acknowledgments

We thank postgraduates J. Qin and L. Yin for their preparation for the fungal strain. We would also thank L. Wang in University of Saskatchewan and T. Passey in East Malling Research for the language editing.

Author-Recommended Internet Resources

COG databases: <https://www.ncbi.nlm.nih.gov/COG>
GO database: <http://geneontology.org>
KEGG database: <https://www.genome.jp/kegg>
Repeat Masker v4.0.5: <http://www.repeatmasker.org>
SMRT Link v5.0.1: <https://www.pacb.com/support/software-downloads>
TMHMM server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM>)
Uni-Prot resource: <https://www.ebi.ac.uk/uniprot>

Literature Cited

- Ardui, S., Ameur, A., Vermeesch, J. R., and Hestand, M. S. 2018. Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. *Nucleic Acids Res.* 46:2159-2168.
- Benson, G. 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27:573-580.
- Bisiach, M. 1988. White rot. Pages 22-23 in: *Compendium of grape diseases*. R. C. Pearson, and A. C. Gohén, eds. APS Press, St. Paul, MN, U.S.A..
- Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., and Weber, T. 2013. antiSMASH 2.0—A versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* 41 (W1): W204-W212.
- Chethana, K. W. T., Zhou, Y., Zhang, W., Liu, M., Xing, Q. K., Li, X. H., Yan, J. Y., Chethana, K. W. T., and Hyde, K. D. 2017. *Coniella vitis* sp. nov. is the common pathogen of white rot in Chinese vineyards. *Plant Dis.* 101:2123-2136.
- Crous, P. W., and Carstens, E. 2000. Confusion surrounding white rot disease of grapevines. *Winelands* 9:89-90.
- Crous, P. W., Phillips, A. J. L., and Baxter, A. P. 2000. *Phytopathogenic Fungi from South Africa*. University of Stellenbosch, Department of Plant Pathology Press, Western Cape South Africa.
- Finn, R., Mistry, J., Tate, J., Coghill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, S. K., Holm, L., Sonnhammer, E. L. L., Eddy, S. R., and Bateman, A. 2009. The Pfam protein family's database. *Nucleic Acids Res.* 38(Database):D211-D222.
- Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., Wilkinson, A. C., Finn, R. D., Griffiths-Jones, S., Eddy, S. R., and Bateman, A. 2009. Rfam: Updates to the RNA families database. *Nucleic Acids Res.* 37 (Database):D136-D140.
- González-Fernández, R., Prats, E., and Jorrín-Novo, J. V. 2010. Proteomics of plant pathogenic fungi. *J. Biomed. Biotechnol.* 2010:932527.
- Goswami, R. S., and Kistler, H. C. 2005. Pathogenicity and in planta mycotoxin accumulation among members of the fusarium graminearum species complex on wheat and rice. *Phytopathology* 95:1397-1404.
- He, Z., Cui, C. Y., and Jiang, J. X. 2017. First report of white rot of grape caused by *Piliidiella castaneicola* in China. *Plant Dis.* 101:1673.
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., and Nakai, K. 2007. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* 35 (Web Server):W585-W587.
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. 2007. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100-3108.
- Li, D., Wan, Y., Wang, Y., and He, P. 2008. Relatedness of resistance to anthracnose and to white rot in Chinese wild grapes. *Vitis* 47:213-215.
- Liu, C., Zhao, K., Wang, K., and Bai, J. 1999. Redesignation of the causal agent of grape white rot in China. *Acta Phytopathol. Sin.* 29:174-176.
- Lowe, T. M., and Eddy, S. R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955-964.
- Nawrocki, E. P., Kolbe, D. L., and Eddy, S. R. 2009. Infernal 1.0: Inference of RNA alignments. *Bioinformatics* 25:1335-1337.
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8:785-786.
- Proctor, R. H., Hohn, T. M., and McCormick, S. P. 1995. Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. *Mol. Plant-Microbe Interact* 8:593-601.
- Reiner, J., Pisani, L., Qiao, W., Singh, R., Yang, Y., Shi, L., Khan, W. A., Sebra, R., Cohen, N., Babu, A., Edelmann, L., Jabs, E. W., and Scott, S. A. 2018. Cyto-genomic identification and long-read single molecule real-time (SMRT) sequencing of a *Bardet-Biedl Syndrome 9 (BBS9)* deletion. *NPJ Genom. Med.* 3:3.
- 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.
- Stergiopoulos, I., and de Wit, P. J. G. M. 2009. Fungal effector proteins. *Ann. Rev. Phytopathol.* 47:233-263.
- Tiedemann, A. 1985. Untersuchungen zur Pathogenität des Erregers der Weistauke (*Coniella petrakii* SUTT.) an Amerikaner-und Europaroben und Verbreitung und Bedeutung des Pilzes in den deutschen Weinbaugebieten. Ph.D. dissertation. Göttingen University, Germany.
- Van Niekerk, J. M., Groenewald, J. Z. E., Verkley, G. J. M., Fourie, P. H., Wingfield, M. J., and Crous, P. W. 2004. Systematic reappraisal of *Coniella* and *Piliidiella*, with specific reference to species occurring on Eucalyptus and Vitis in South Africa. *Mycol. Res.* 108:283-303.