Genome Sequence Resource for *Colletotrichum viniferum*, the Cause of Grapevine Ripe Rot in China

Mengru Dou,^{1,2} Yu Hao,^{1,2} Jing Yang,^{1,2} Xiaojian Yuan,^{1,2} Xiao Yin,^{1,2} Yuntong Jiao,^{1,2} Jiahe Zhao,³ Tingting Chen,^{1,2} Yuejin Wang,^{1,2} and Yan Xu^{1,2,†}

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

² Key Laboratory of Horticultural Plant Biology and Germplasm Innovation in Northwest China, Ministry of Agriculture, Northwest A&F University, Yangling, Shaanxi 712100, China

³ State Key Laboratory of Crop Stress Biology in Arid Areas, College of Plant Protection, Northwest A&F University, Yangling, Shaanxi 712100, China

Abstract

Grape ripe rot is an important disease that has seriously damaged the yield and quality of grape worldwide. The disease is caused by *Colletotrichum viniferum*, a hemibiotrophic fungus that belongs to the Glomerellaceae family of Sordariomycetes class. Here, we present the genome of *C. viniferum* CvYL2a from grape, based on Illumina HiSeq 2500 and PacBio RS II. The high-quality genome consists of 70 contigs with a 73.41 Mb genome size and encodes 14,668 protein-coding genes. These genes were annotated using Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, EuKaryotic Orthologous Groups, Nonredundant Protein, and Swiss-Prot databases. In addition, we identified a series of genes involved in pathogenicity, including 909 carbohydrate-active enzymes, 67 secondary metabolite gene clusters, and 307 cytochrome P450 enzymes. This genome sequence provides a valuable reference for research on grape–*C. viniferum* interactions, the pathogenesis of *C. viniferum*, and comparative genome analyses.

Genome Announcement

Colletotrichum spp. invade all parts of the plant, including leaves, stems, young shoots, tendrils, and berries (Wang et al. 2021), thereby causing serious diseases on crop plant species, such as grapevine, chili, and lentil (Bhadauria et al. 2019; Huo et al. 2021; Lei et al. 2016).

Data from the International Organization of Vine and Wine showed that the total vineyard surface area in China was 875,000 hectares with a yield of 11,700,000 tons in 2018. Most grape-growing areas in China experience high temperature and precipitation, which causes grape ripe rot and results in substantial yield losses. Grape ripe rot occurs in most viticulture regions of China and damages grape any time, mainly during the berry mature stage. As a result, red-brown spots are formed on the surface of the fruit and seriously influence the yield and quality of grapes (Alleweldt and Possingham 1988). *Colletotrichum gloeosporioides* (Penz) has been traditionally considered as the only pathogen causing grape ripe

Funding

National Key Research and Development Program of China number: 2018YFD1000300; Shaanxi Province Key Research and Development Program number: 2018ZDXMNY053-1; Shaanxi province Agriculture Research Systems for the grape industry; Key Industry Chain of Innovation Project in Shaanxi Province number: 2021ZDLNY04-08; Flexible Introduction Team Project of Ningxia Hui Autonomous Region number: 2020RXTDLX08.

Keywords

Colletotrichum viniferum, genome, grape, pathogenesis

†Corresponding author: Y. Xu; yan.xu@nwfau.edu.cn

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

The author(s) declare no conflict of interest.

Accepted for publication 20 September 2021.

Copyright © 2021 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

Table 1. Genome features of the *C. viniferum* YL-2a

Features	Statistics
Genome size (Mb)	73.40
Gene number	14,668
Gene total length(bp)	21,757,859
Gene length/Genome (%)	29.64
GC content of coding DNA (%)	56.91
Gene average length (bp)	1,483
Gene internal length (bp)	51,648,482
Gene internal GC Content (%)	40.07
Gene internal (% of total genome)	70.36
Transfer RNA number	482
Ribosomal RNA number	131
Small nuclear RNA number	26
Repetitive sequences (%)	4.42
Contig	70
Max contig length (Mb)	6.4
N ₅₀ contig length (Mb)	2.0

rot. To date, *C. acutatum*, *C. aenigma*, *C. capsici*, *C. fruticola*, and *C. viniferum* have been reported as the ripe rot pathogens of grapes (Hong et al. 2008; Peng et al. 2013; Whitelaw-Weckert et al. 2007; Yan et al. 2015). Fungal genome sequencing provides a theoretical foundation for understanding host-pathogen interactions and developing effective pathogen defense measures.

C. viniferum was collected from *Vitis vinifera* cv. Kyoho in the grapevine germplasm resources orchard at Northwest A&F University in Yangling, Shaanxi Province, China. The specimen was isolated by the single-spore method, and the strain was identified as *C. viniferum* and was named CvYL2a (Supplementary Fig. S1), based on the multigene phylogenetic analyses of internal transcribed spacer (ITS) regions (Gardes and Bruns 1993; White et al. 1990), GADPH (Templeton et al. 1992), CHS-1 (Carbone and Kohn 1999), ACT (Carbone and Kohn 1999), and TUB2 (O'Donnell and Cigelnik 1997). This report presents the whole-genome sequence of *C. viniferum* CvYL2a. Highly purified total genomic DNA was extracted from fungal mycelia from 5-day-old cultures grown on potato dextrose agar at 25°C in darkness, using the cetyl trimethyl ammonium bromide method (Weiland 1997). Its genome was sequenced with a combination of Illumina short-read and PacBio long-read sequencing technologies (Novogene, Beijing). A short-insert library (350 bp) was constructed and sequenced on the Illumina HiSeq2500 platform. A long-read SMRT (single-molecule real-time) Bell library (20 Kb) was constructed and sequenced with PacBio RS II SMRT cells. A total of 19.27 Gb of PacBio long reads, with 262-fold sequence depth, were generated. The final assembled genome was 73.41 Mb in length and consisted of 70 contigs, with the largest length being 6.45 M, with an N₅₀ of 2.10 Mb (Table 1). Its assembly size was longer than that of other published *Colletotrichum* sp. genomes, such as *C. gloeosporioides* SMCG1#C (59.04 Mb), *C. gloeosporioides* Lc1 (61.9 Mb), and *C. asianum* ICMP 18580 (64.73 Mb) (Fu et al. 2020; Huang et al. 2019; Meng et al. 2019). Compared with strain *C. viniferum* CGW01, which was sequenced with Illumina and was submitted to the National Center for Biotechnology Information in 2018, *C. viniferum* CvYL2a has a longer genome, fewer contigs, and a longer N₅₀ length (Supplementary Table S1). Repeat sequences in the genome were predicted and analyzed, using Repeat Masker v4.0.5 and Tandem Repeats Finder v4.07b (Saha et al. 2008; Benson 1999). Approximately 4.42% of these sequences were identified as repeat sequences (Supplementary Table S2A). In summary, 14,668 protein-coding genes were predicted, with the average gene length being 1,483 bp. The general feature format of these protein-coding genes is listed in Supplementary Table S2B. Only 9,003 (61.37%), 12,111 (82.56%), 2,229 (15.19%), 1,3881 (94.63%), and 3,748 (25.55%) of these predicted genes were respectively mapped to the Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, EuKaryotic Orthologous Groups, Nonredundant Protein, and Swiss-Prot databases, respectively (Ashburner et al. 2000; Bairoch and Apweiler 2000; Kanehisa et al. 2006; Li et al. 2002). These proteins were classified using the Pfam v 32.0 databases (El-Gebali et al. 2019). The results identified 9,003 genes containing Pfam protein domain (Supplementary Figs. S2 and S3; Supplementary Table S3A to F).

Carbohydrate-active enzymes (CAZymes) potentially degrade the plant cell wall and modify the fungal cell wall (O'Connell et al. 2012). CAZymes have been annotated using the CAZy database (Cantarel et al. 2009). A total of 909 CAZy genes include 129 carbohydrate-binding modules, 413 (45.43%) glycoside hydrolases, 121 (13.31%) glycosyl transferases, 44 (4.84%) polysaccharide lyases, 68 (7.48%) carbohydrate esterases, and 134 (14.74%) auxiliary activities (Supplementary Fig. S3B; Supplementary Table S3G). Secondary metabolites are correlated to the pathogenicity of many phytopathogenic fungi in the infection process (Wang et al. 2020). AntiSMASH v 2.0.2 identified 67 secondary metabolite biosynthesis clusters, namely, one indole-t1pks cluster, nine nrps clusters, one t3pks cluster, seven indole clusters, 14 terpene clusters, 23 t1pks clusters, four t1pks-nrps clusters, and either other clusters (Medema et al. 2011) (Supplementary Fig. S3D). Cytochrome P450 (CYP) enzymes catalyze the conversion of fungal hydrophobic intermediates for primary and secondary metabolism (Yin et al. 2015). Using the P450 database, a total of 307 (2.09%) CYP enzymes were identified in *C. viniferum* CvYL2a (Supplementary Table S3H).

This study provides a high-quality genome sequence of *C. viniferum* CvYL2a, a fungal pathogen causing grape ripe rot. This sequence was markedly better than that of *C. viniferum* CGW01. This whole-genome sequence will aid comparative genomic studies and provide theoretical guidance to further explore the molecular mechanism of plant pathogen interactions.

Data Availability

This whole-genome shotgun project of *C. viniferum* CvYL2a was deposited in the DDBJ/ENA/GenBank database under accession number JAGFBE000000000 (BioProject PRJNA714705; BioSample SAMN18316538).

Literature Cited

- Alleweldt, G., and Possingham, J. V. 1988. Progress in grapevine breeding. *Theor. Appl. Genet.* 75:669-673.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* 25:25-29.
- Bairoch, A., and Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28:45-48.
- Benson, G. 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27:573-580.
- Bhadauria, V., MacLachlan, R., Pozniak, C., Cohen-Skalie, A., Li, L., Halliday, J., and Banniza, S. 2019. Genetic map-guided genome assembly reveals a virulence-governing minichromosome in the lentil anthracnose pathogen *Colletotrichum lentis*. *New Phytol.* 221:431-445.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* 37 (Database):D233-D238.
- Carbone, I., and Kohn, L. M. 1999. A method for designing primer sets for speciation studies in filamentous ascomycetes. *Mycologia* 91:553-556.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., and Finn, R. D. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res.* 47 (D1):D427-D432.
- Fu, F. F., Hao, Z., Wang, P., Lu, Y., Xue, L. J., Wei, G., Tian, Y., Hu, B., Xu, H., Shi, J., Cheng, T., Wang, G., Yi, Y., and Chen, J. 2020. Genome sequence and comparative analysis of *Colletotrichum gloeosporioides* isolated from *Liriodendron* Leaves. *Phytopathology* 110:1260-1269.
- Gardes, M., and Bruns, T. D. 1993. ITS primers with enhanced specificity for basidiomycetes—Application to the identification of mycorrhizae and rusts. *Mol. Ecol.* 2:113-118.
- Hong, S. K., Kim, W. G., Yun, H. K., and Choi, K. J. 2008. Morphological variations, genetic diversity and pathogenicity of *Colletotrichum* species causing grape ripe rot in Korea. *Plant Pathol. J.* 24:269-278.
- Huang, L., Kim, K. T., Yang, J. Y., Song, H., Choi, G., Jeon, J., Cheong, K., Ko, J., Xu, H., and Lee, Y. H. 2019. A high-quality draft genome sequence of *Colletotrichum gloeosporioides* sensu stricto SMC61#C, a causal agent of anthracnose on *Cunninghamia lanceolata* in China. *Mol. Plant-Microbe Interact.* 32:139-141.
- Huo, J., Wang, Y., Hao, Y., Yao, Y., Wang, Y., Zhang, K., Tan, X., Li, Z., and Wang, W. 2021. Genome sequence resource for *Colletotrichum scovillei*, the cause of anthracnose disease of chili. *Mol. Plant-Microbe Interact.* 34:122-126.
- Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. 2006. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* 34:D354-D357.
- Lei, Y., Tang, X., Jayawardena, R. S., Yan, J., Wang, X., Liu, M., Chen, T., Liu, X., Wang, X., Wang, J., and Chen, Q. 2016. Identification and characterization of *Colletotrichum* species causing grape ripe rot in southern China. *Mycosphere* 7:1177-1191.
- Li, W., Jaroszewski, L., and Godzik, A. 2002. Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics* 18:77-82.
- Medema, M. H., Blin, K., Cimermancic, P., Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. 2011. antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39:W339-W346.
- Meng, Y., Gleason, M. L., Zhang, R., and Sun, G. 2019. genome sequence resource of the wide-host-range anthracnose pathogen *Colletotrichum siamense*. *Mol. Plant-Microbe Interact.* 32:931-934.
- O'Connell, R. J., Thon, M. R., Hacquard, S., Amyotte, S. G., Kleemann, J., Torres, M., Damm, U., Buiate, E. A., Epstein, L., Alkan, N., Altmüller, J., Alvarado-Balderrama, L., Bauser, C. A., Becker, C., Birren, B. W., Chen, Z., Choi, J., Crouch, J. A., Duvick, J. P., Farman, M. A., Gan, P., Heiman, D., Henrissat, B., Howard, R. J., Kabbage, M., Koch, C., Kracher, B., Kubo, Y., Law, A. D., Lebrun, M., Lee, Y., Miyara, I., Moore, N., Neumann, U., Nordström, K., Panaccione, D. G., Panstruga, R., Place, M., Proctor, R. H., Prusky, D., Rech, G., Reinhardt, R., Rollins, J. A., Rounsley, S., Schardl, C. I., Schwartz, D. C., Shenoy, N., Shirasu, K., Sikhakolli, U. R., Stüber, K., Sukno, S. A., Sweigard, J. A., Takano, Y., Takahara, H., Trail, F., van der Does, H. C., Voll, L. M., Will, I., Young, S., Zeng, Q., Zhang, J., Zhou, S., Dickman, M. B., Schulze-Lefert, P., van Themaat, E. V. L., Ma, L., and Vaillancourt, L. J. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat. Genet.* 44:1060-1067.

- O'Donnell, K., and Cigelnik, E. 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol. Phylogenet. Evol.* 7:103-116.
- Peng, L. J., Sun, T., Yang, Y. L., Cai, L., Hyde, K. D., Bahkali, A. H., and Liu, Z. Y. 2013. *Colletotrichum* species on grape in Guizhou and Yunnan provinces, China. *Mycoscience* 54:29-41.
- 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.
- Templeton, M. D., Rikkerink, E. H. A., Solon, S. L., and Crowhurst, R. N. 1992. Cloning and molecular characterization of the glyceraldehyde-3-phosphate dehydrogenase-encoding gene and cDNA from the plant pathogenic fungus *Glomerella cingulata*. *Gene* 122:225-230.
- Wang, J., Shi, D., Wei, L., Chen, W., Ma, W., Chen, C., and Wang, K. 2020. Mutations at sterol 14 α -demethylases (CYP51A&B) confer the DMI resistance in *Colletotrichum gloeosporioides* from grape. *Pest Manag. Sci.* 76:4093-4103.
- Wang, M., Luo, X., and Wang, H. 2021. First draft genome sequence resource of *Colletotrichum liriopes* causing leaf anthracnose on *Ophiopogon japonicus*. *Plant Dis.* 105:1179-1182.
- Weiland, J. J. 1997. Rapid procedure for the extraction of DNA from fungal spores and mycelia. *Fungal Genet. Rep.* 44:60-63.
- White, T. J., Bruns, T., Lee, S., and Taylor, J. W. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-322 in: *PCR Protocols: A Guide to Methods and Applications*. D. H. Gelfand, J. J. Sninsky, and T. J. White, eds. Academic, New York.
- Whitelaw-Weckert, W. M. A., Curtis, S. J., Huang, R., Steel, C. C., Blandchard, C. L., and Roffey, P. E. 2007. Phylogenetic relationships and pathogenicity of *colletotrichum acutatum* isolates from grape in subtropical australia. *Plant Pathol.* 56:448-463.
- Yan, J. Y., Jayawardena, M. M. R. S., Goonasekara, I. D., Wang, Y., Zhang, W., Liu, M., Huang, J. B., Wang, Z. Y., Shang, Y. L., Peng, Y. L., Bankali, A., Hyde, K. D., and Li, X. H. 2015. Diverse species of *colletotrichum* associated with grapevine anthracnose in china. *Fungal Divers.* 71:233-246.
- Yin, Z., Liu, H., Li, Z., Ke, X., Dou, D., Gao, X., Song, N., Dai, Q., Wu, Y., Xu, J. R., Kang, Z., and Huang, L. 2015. Genome sequence of *Valsa* canker pathogens uncovers a potential adaptation of colonization of woody bark. *New Phytol.* 208:1202-1216.