

Whole-Genome Sequence of the Orchid Anthracnose Pathogen *Colletotrichum orchidophilum*

Riccardo Baroncelli,^{1,2,†} Serenella A. Sukno,² Sabrina Sarrocco,³ Giovanni Cafà,⁴ Gaetan Le Floch,¹ and Michael R. Thon^{2,†}

¹ Laboratoire Universitaire de Biodiversité et Ecologie Microbienne, IBSAM, ESIAB, Université de Brest, EA 3882, Technopôle Brest-Iroise, 29280 Plouzané, France

² Instituto Hispano-Luso de Investigaciones Agrarias (CIALE), University of Salamanca, Calle del Duero, 12; 37185 Villamayor (Salamanca), Spain

³ Department of Agricultural, Food and Environment, University of Pisa, Via del Borghetto 80, Pisa, Italy

⁴ CABI Europe-UK, Bakeham Lane, Egham, Surrey TW20 9TY, U.K.

Funding

This research was supported by “The Prograilive” project, the regions of Bretagne and Pays de la Loire, and FEADER grants, France (grant RBRE160116CR0530019), and by the Ministerio de Economía y Competitividad, Spain (grant AGL2015-66362-R).

Abstract

Colletotrichum orchidophilum is a plant-pathogenic fungus infecting a wide range of plant species belonging to the family Orchidaceae. In addition to its economic impact, *C. orchidophilum* has been used in recent years in evolutionary studies because it represents the closest related species to the *C. acutatum* species complex. Here, we present the first-draft whole-genome sequence of *C. orchidophilum* IMI 309357, providing a resource for future research on anthracnose of Orchidaceae and other hosts.

Most *Colletotrichum* spp. are associated with anthracnose diseases of plants and virtually every plant cultivated in the world is susceptible to one or more species. The genus *Colletotrichum* is considered both scientifically and economically to be one of the most important groups of plant-pathogenic fungi (Dean et al. 2012). *Colletotrichum orchidophilum* causes disease on several species of the family Orchidaceae. According to data collected from culture collections and GenBank, *C. orchidophilum* is restricted to the family Orchidaceae in countries such as the United States, the United Kingdom, Panama, and China (Damm et al. 2012). Recently described as a new species of *Colletotrichum*, *C. orchidophilum* does not belong to any major clade or species complex (Marin-Felix et al. 2017). Because *C. orchidophilum* is the most closely related species to the *C. acutatum* species complex, it has been widely used as a reference outgroup to build *C. acutatum* species phylogenetic trees (Baroncelli et al. 2017). *C. orchidophilum* strain IMI 309357 was collected on 27 August 1986 from orchid (*Phalaenopsis* sp.) in the United Kingdom.

The genome of *C. orchidophilum* strain IMI 309357 was sequenced using Illumina paired-end sequencing technology by the McGill University and Genome, Quebec Innovation Centre (Canada). Paired reads of 250 bp (10.02 Gbp; average predicted coverage 200×) were assembled using SPAdes v3.8.2 (Bankevich et al. 2012). The mitochondrial genome was assembled manually using Geneious R8.1.4 (Kearse et al. 2012). The nuclear genome of *C. orchidophilum* consists of 321 sequence scaffolds with a total assembled length of 48.56 Mbp (N50 = 284,473 and L50 = 57), 51.07% GC content, and a maximum scaffold size of 844,029 bp (Table 1). The completeness of the assembly was assessed using BUSCO v12

G. L. Floch and M. R. Thon contributed equally to this work.

†Corresponding authors: R. Baroncelli; E-mail: riccardobaroncelli@gmail.com; and M. R. Thon; E-mail: mthon@usal.es

Accepted for publication 30 March 2018.

Table 1. Summary statistics of the *Colletotrichum orchidophilum* strain IMI309357 genome assembly

Variables	Statistics
Number of scaffolds	321
Assembly length (bp)	48,556,462
N50 (bp)	284,473
L50 (bp)	57
BUSCO completeness	99.79%
Number of predicted genes	14,496
Number of predicted secreted proteins	1,488

(Simão et al. 2015), which estimated the genome sequence to be 99.8% complete. The nuclear genome was annotated using the MAKER2 pipeline (Holt and Yandell 2011). Overall, 14,496 protein-coding gene models were predicted. Analysis with SignalP 4.1 (Petersen et al. 2011) revealed that 1,564 predicted proteins (10.8% of the proteome) contain a secretion signal peptide. A comparative analysis with other *Colletotrichum* genomes available (Alkan et al. 2013; Baroncelli et al. 2014, 2016; Gan et al. 2013, 2016; Hacquard et al. 2016; O'Connell et al. 2012) confirmed that *C. orchidophilum* is closely related to *C. acutatum* species, sharing 83.6% (12,116) of the proteomes. Among all the secreted proteins encoded by *C. orchidophilum* IMI 309357, 197 are *Colletotrichum*-specific proteins, of which 11 are lineage-specific because they do not have any sequence similarity to proteins in public databases, based on BLAST searches (*e*-value threshold of 1e-10). Such features are characteristic of fungal effectors, which are proteins that are involved in the suppression or evasion of host defense.

The *C. orchidophilum* sequence represents a new resource for future research on anthracnose and *Colletotrichum* comparative genomics. The genome sequence has been deposited in GenBank under the accession number MJBS00000000.1 (BioProject: PRJNA342923; BioSample: SAMN05771038). The genome sequence and gene models of *C. orchidophilum* IMI 309357 are also available from the MycoCosm fungal genomics portal, developed by the U.S. Department of Energy Joint Genome Institute (Grigoriev et al. 2014).

Acknowledgments

We thank I. Grigoriev and the Joint Genome Institute staff for adding the genome sequence to the MycoCosm portal.

Literature Cited

- Alkan, N., Meng, X., Friedlander, G., Reuveni, E., Sukno, S., Sherman, A., Thon, M., Fluhr, R., and Prusky, D. 2013. Global aspects of pacC regulation of pathogenicity genes in *Colletotrichum gloeosporioides* as revealed by transcriptome analysis. *Mol. Plant-Microbe Interact.* 26:1345-1358.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Pribelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455-477.
- Baroncelli, R., Amby, D. B., Zapparata, A., Sarrocco, S., Vannacci, G., Le Floch, G., Harrison, R. J., Holub, E., Sukno, S. A., Sreenivasaprasad, S., and Thon, M. R. 2016. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics* 17:555.
- Baroncelli, R., Sanz-Martín, J. M., Rech, G. E., Sukno, S. A., and Thon, M. R. 2014. Draft genome sequence of *Colletotrichum sublineola*, a destructive pathogen of cultivated sorghum. *Genome Announc.* 2:e00540-e14.
- Baroncelli, R., Talhinhas, P., Pensec, F., Sukno, S. A., Le Floch, G., and Thon, M. R. 2017. The *Colletotrichum acutatum* species complex as a model system to study evolution and host specialization in plant pathogens. *Front. Microbiol.* 8:2001.
- Damm, U., Cannon, P. F., Woudenberg, J. H. C., and Crous, P. W. 2012. The *Colletotrichum acutatum* species complex. *Stud. Mycol.* 73:37-113.
- Dean, R., Van Kan, J. A. L., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., Rudd, J. J., Dickman, M., Kahmann, R., Ellis, J., and Foster, G. D. 2012. The top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 13:414-430.
- Gan, P., Ikeda, K., Irieda, H., Narusaka, M., O'Connell, R. J., Narusaka, Y., Takano, Y., Kubo, Y., and Shirasu, K. 2013. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol.* 197:1236-1249.
- Gan, P., Narusaka, M., Kumakura, N., Tsushima, A., Takano, Y., Narusaka, Y., and Shirasu, K. 2016. Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol. Evol.* 8:1467-1481.
- Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otilar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., and Shabalov, I. 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42:D699-D704.
- Hacquard, S., Kracher, B., Hiruma, K., Münch, P. C., Garrido-Oter, R., Thon, M. R., Weimann, A., Damm, U., Dallery, J.-F., Hainaut, M., Henrissat, B., Lespinet, O., Sacristán, S., Ver Loren van Themaat, E., Kemen, E., McHardy, A. C., Schulze-Lefert, P., and O'Connell, R. J. 2016. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat. Commun.* 7: Article 11362.

- Holt, C., and Yandell, M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinf.* 12:491.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647-1649.
- Marin-Felix, Y., Groenewald, J. Z., Cai, L., Chen, Q., Marincowitz, S., Barnes, I., Bensch, K., Braun, U., Camporesi, E., Damm, U., de Beer, Z. W., Dissanayake, A., Edwards, J., Giraldo, A., Hernández-Restrepo, M., Hyde, K. D., Jayawardena, R. S., Lombard, L., Luangsa-Ard, J., McTaggart, A. R., Rossman, A. Y., Sandoval-Denis, M., Shen, M., Shivas, R. G., Tan, Y. P., van der Linde, E. J., Wingfield, M. J., Wood, A. R., Zhang, J. Q., Zhang, Y., and Crous, P. W. 2017. Genera of phytopathogenic fungi: GOPHY 1. *Stud. Mycol.* 86:99-216.
- O'Connell, R. J., Thon, M. R., Hacquard, S., Amyotte, S. G., Kleemann, J., Torres, M. F., 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.-H., Lee, Y.-H., 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. L., Schwartz, D. C., Shenoy, N., Shirasu, K., Sikkakolli, 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., Ver Loren van Themaat, E., Ma, L.-J., and Vaillancourt, L. J. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat. Genet.* 44:1060-1065.
- 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.
- 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.

Author-Recommended Internet Resources

MycoCosm fungal genomics portal: <https://genome.jgi.doe.gov/programs/fungi/index.jsf>