MPMI Vol. 32, No. 9, 2019, pp. 1088–1090. https://doi.org/10.1094/MPMI-12-18-0352-A

RESOURCE ANNOUNCEMENT

Genome Sequence Resources for Four Phytopathogenic Fungi from the *Colletotrichum orbiculare* Species Complex

P. Gan,[1] A. Tsushima,[1,2] M. Narusaka,[3] Y. Narusaka,[3] Y. Takano,[4] Y. Kubo,[5] and K. Shirasu[1,2†]

[1] RIKEN Center for Sustainable Resource Sciences, Yokohama, Kanagawa, Japan
[2] Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo, Japan
[3] Research Institute for Biological Sciences Okayama, Okayama Prefectural Technology Center for Agriculture, Forestry, and Fisheries, Okayama, Japan
[4] Graduate School of Agriculture, Kyoto University, Kyoto, Japan
[5] Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, Kyoto, Japan

Abstract

Colletotrichum orbiculare species complex fungi are hemibiotrophic plant pathogens that cause anthracnose of field crops and weeds. Members of this group have genomes that are remarkably expanded relative to other *Colletotrichum* fungi and compartmentalized into AT-rich, gene-poor and GC-rich, gene-rich regions. Here, we present an updated version of the *C. orbiculare* genome, as well as draft genomes of three other members from the *C. orbiculare* species complex: the alfalfa pathogen *C. trifolii*, the prickly mallow pathogen *C. sidae*, and the burweed pathogen *C. spinosum*. The data reported here will be important for comparative genomics analyses to identify factors that play a role in the evolution and maintenance of the expanded, compartmentalized genomes of these fungi, which may contribute to their pathogenicity.

Colletotrichum is a genus of diverse plant-pathogenic fungi whose members can be subdivided into species complexes comprising closely related but distinct fungal species (Cannon et al. 2012). Among these, the *Colletotrichum orbiculare* species complex consists of hemibiotrophic pathogens of herbaceous plants, including agriculturally important crops such as cucurbits and alfalfa (Damm et al. 2013). Within the *C. orbiculare* species complex, *C. orbiculare,* which infects cucurbits, and *C. lindemuthianum*, the common bean pathogen, have been sequenced (de Queiroz et al. 2017; Gan et al. 2013). Remarkably, both of these genomes are considerably expanded relative to those of other *Colletotrichum* fungi, including *C. gloeosporioides*, *C. graminicola*, *C. chlorophyti*, *C. destructivum, C. acutatum, C. spaethianum*, and *C. orchidearum* species complex members. The *C. orbiculare* species complex genomes appear to be organized into AT-rich, gene-poor regions and GC-rich, gene-rich regions. In other plant pathogens, transposable element-rich, gene-poor regions are hypothesized to contribute to pathogenicity by promoting the rapid evolution of effector genes required for virulence (Dong et al. 2015; Rouxel et al. 2011). The AT-rich, gene-poor genomic regions of *C. orbiculare* species complex member genomes may play a similar role, highlighting their interest to studies on the evolution of virulence factors among this group of pathogens. To study this, genomic resources are required. Here, we report the updated genome assembly of *C. orbiculare* as well as assemblies of three additional species from the

†Corresponding author: K. Shirasu; ken.shirasu@riken.jp

The author(s) declare no conflict of interest.

Accepted for publication 19 March 2019.

Funding

This work was supported, in part, by the Japan Society for the Promotion of Science KAKENHI 17H06172 and 17J02983 to K. Shirasu; Agriculture, Forestry and Fisheries Research Council Science and Technology Research Promotion Program for Agriculture, Forestry, Fisheries, and Food Industry awarded to K. Shirasu, Y. Takano, and Y. Narusaka; the Project of the NARO Bio-oriented Technology Research Advancement Institution (research program on development of innovative technology) to Y. Takano and Y. Narusaka; and JSPS Grant-in-Aid for JSPS Research Fellow 17J02983 to A. Tsushima.

Keywords

genomics, metabolomics, proteomics

Table 1. Genome assembly statistics of *Colletotrichum orbiculare, C. trifolii, C. spinosum*, and *C. sidae*

	Species[a]			
Statistics	***C. orbiculare***	***C. trifolii***	***C. spinosum***	***C. sidae***
Culture collection	MAFF 240422, CBS 514.97	MAFF 305078	CBS 515.97	CBS 518.97
Host	*Cucumis sativus*	*Medicago sativa*	*Xanthium spinosum*	*Sida spinosa*
Sequencing technology	454, PacBio RSII	PacBio RSII, Illumina HiSeq2000	Illumina HiSeq2000	Illumina HiSeq2000
Number of scaffolds	355	10,473	10,715	14,826
Total assembly length (Mb)	89.74	109.66	82.73	85.83
Largest contig length (bp)	5,815,657	900,455	1,849,982	532,833
Scaffold N50 (bp)	2,078,754	36,142	101,238	41,033
Scaffold L50	13	425	139	352
N's per 100 kbp	2,012.85	3,789.84	11.66	6.51
GC (%)	37.58	36.66	38.90	38.24
BUSCO complete (%)	99.00	99.00	99.00	98.80
BUSCO fragmented (%)	0.30	0.60	0.50	0.40
Number of genes	13,253	12,292	12,540	12,442
Genome accession	AMCV00000000	RYZW00000000	QAPG00000000	QAPF00000000
Version accession	AMCV02000000	RYZW01000000	QAPG01000000	QAPF01000000

[a] CBS = Culture collection of the Centraalbureau voor Schimmelcultures, Fungal Biodiversity Centre, Utrecht, The Netherlands; and MAFF = MAFF Genebank Project, Ministry of Agriculture, Forestry and Fisheries, Tsukuba, Japan.

C. orbiculare species complex: *C. trifolii,* which infects alfalfa (*Medicago sativa*); *C. sidae*, which infects the shrub prickly mallow (*Sida spinosa*); and *C. spinosum,* which infects burweed (*Xanthium spinosum*).

The *C. trifolii, C. sidae*, and *C. spinosum* genomes were sequenced to 123×, 107×, and 121× coverages, respectively, on an Illumina HiSeq2000 sequencer (RIKEN Omics Science Center, Yokohama, Japan). For each sample, 100-bp paired-end libraries of 150- and 500-bp insert sizes were generated using Illumina TruSeq PCR-free DNA Sample Prep Kits according to the manufacturer's instructions. Low-quality reads were trimmed with TrimGalore with fastqc (v0.11.7) and cutadapt (v1.2.1) (Martin 2011). Contigs for *C. sidae* and *C. spinosum* were assembled separately using Megahit (v1.1.2) (Li et al. 2015), scaffolded using SSPACE-Standard-2.0 (Boetzer et al. 2011), and then corrected with SOAPdenovo GapCloser (v1.12) (Luo et al. 2012) using default settings for all three programs. On the other hand, *C. trifolii*-derived paired-end reads were assembled into contigs using CLC Genomics Workbench, version 8 (QIAgen), setting the word and bubble size automatically. Contigs from *C. trifolii* and the previously published assembly of *C. orbiculare* (Gan et al. 2013) were then scaffolded separately with SSPACE-LongRead (v1.1) (Boetzer and Pirovano 2014) using 12× and 32× reads generated from one and three PacBio RSII cells, respectively, with default settings. Gap closing was performed on the *C. trifolii* assembly using GapCloser (v1.12). The completeness of each assembly was estimated by searching for conserved genes in the sordariomyceta_odb9 lineage dataset using the BUSCO program (v3.0.2) (Simão et al. 2015) with the settings –fungal–long. *C. trifolii, C. sidae*, and *C. spinosum* assemblies were annotated using the MAKER2 pipeline (Holt and Yandell 2011) to incorporate predictions from Augustus (v3.3) (Stanke et al. 2004), trained using BUSCO-identified single-copy genes; and GeneMarkES (Lukashin and Borodovsky 1998), trained on the respective genome sequences using the settings –ES–fungus. Further, previously predicted proteins from *C. orbiculare* (Gan et al. 2013) were included as additional evidence using the Exonerate (v2.2.0) (Slater and Birney 2005) protein2gene function in MAKER2. For *C. orbiculare,* gene models were trained with the BRAKER1 (Hoff et al. 2016) pipeline using hints from *C. orbiculare*-derived RNA-seq reads mapped to the genome with hisat2 (Kim et al. 2015), setting 1,000 bp as the maximum intron size. Predicted genes were annotated using Annie the ANNotator (Tate et al. 2014) based on BLASTp hits against the UniProt SwissProt database (Boutet et al. 2007) with an E-value cut off of 1E-5. All sequences were deposited in DDBJ/ENA/GenBank and sequence accessions are shown in Table 1. All four genomes have GC contents of less than 40% and genome sizes greater than 85 Mb (Table 1), which are larger compared with other *Colletotrichum* spp. outside of the *C. orbiculare* species complex. Despite this, the number of genes encoded by each genome are not significantly expanded relative to other fungi in the *Colletotrichum* genus. Comparing the genomes reported here against other fungal genomes may provide insights into factors that contribute to the unique

genome organization of this group of fungi, as well as help identify genes that are important for host pathogenicity.

Literature Cited

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27: 578-579.

Boetzer, M., and Pirovano, W. 2014. SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. BMC Bioinf. 15:211.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., and Bairoch, A. 2007. Uniprotkb/swiss-prot. Pages 89-112 in: Plant Bioinformatics, Humana Press, Totowa, NJ, U.S.A.

Cannon, P. F., Damm, U., Johnston, P. R., and Weir, B. S. 2012. *Colletotrichum* – current status and future directions. Stud. Mycol. 73:181-213.

Damm, U., Cannon, P. F., Liu, F., Barreto, R. W., Guatimosim, E., and Crous, P. W. 2013. The Colletotrichum orbiculare species complex: Important pathogens of field crops and weeds. Fungal Divers. 61:29-59.

de Queiroz, C. B., Correia, H. L. N., Menicucci, R. P., Vidigal, P. M. P., and de Queiroz, M. V. 2017. Draft Genome Sequences of Two Isolates of *Colletotrichum lindemuthianum*, the Causal Agent of Anthracnose in Common Beans. Genome Announce. 5:e00214-17.

Dong, S., Raffaele, S., and Kamoun, S. 2015. The two-speed genomes of filamentous pathogens: Waltz with plants. Curr. Opin. Genet. Dev. 35:57-65.

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.

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. 2016. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767-769.

Holt, C., and Yandell, M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinf. 12:491.

Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: A fast spliced aligner with low memory requirements. Nat. Methods 12:357-360.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674-1676.

Lukashin, A. V., and Borodovsky, M. 1998. GeneMark.hmm: New solutions for gene finding. Nucleic Acids Res. 26:1107-1115.

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.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10-12.

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A. J., Ciuffetti, L. M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S. B., Gout, L., Glaser, N., Linglin, J., Kema, G. H. J., Lapalu, N., Lawrence, C. B., May, K., Meyer, M., Ollivier, B., Poulain, J., Schoch, C. L., Simon, A., Spatafora, J. W., Stachowiak, A., Turgeon, B. G., Tyler, B. M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R. P., Wincker, P., Balesdent, M.-H., and Howlett, B. J. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. Nat. Commun. 2:202.

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.

Slater, G. S. C., and Birney, E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinf. 6:31.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. 2004. AUGUSTUS: A web server for gene finding in eukaryotes. Nucleic Acids Res. 32: W309-W312.

Tate, R., Hall, B., DeRego, T., and Geib, S. 2014. Annie: The ANNotation Information Extractor (Version 1.0). http://genomeannotation.github.io/annie/