Genome Sequence Resources of *Colletotrichum truncatum*, *C. plurivorum*, *C. musicola*, and *C. sojae*: Four Species Pathogenic to Soybean (*Glycine max*)

Flávia Rogério,¹ Thaís R. Bouffleur,^{1,2} Maisa Ciampi-Guillardi,¹ Serenella A. Sukno,² Michael R. Thon,² Nelson Sidnei Massola Júnior,¹ and Riccardo Baroncelli^{2,†}

¹ Department of Plant Pathology and Nematology, University of São Paulo (USP), Luiz de Queiroz College of Agriculture (ESALQ), Piracicaba, São Paulo, Brazil

² Instituto Hispano-Luso de Investigaciones Agrarias (CIALE), Universidad de Salamanca, Villamayor, Salamanca, Spain

Abstract

Colletotrichum is a large genus of plant pathogenic fungi comprising more than 200 species. In this work, we present the genome sequences of four *Colletotrichum* species pathogenic to soybean: *C. truncatum*, *C. plurivorum*, *C. musicola*, and *C. sojae*. While *C. truncatum* is globally considered the most important pathogen, the other three species have been described and associated with soybean only recently. The genome sequences will provide insights into factors that contribute to pathogenicity toward soybean and will be useful for further research into the evolution of *Colletotrichum*.

Resource Announcement

Colletotrichum is a large genus of plant pathogenic fungi, currently considered one of the top 10 most important fungal pathogens by plant pathologists (Crouch et al. 2014; Dean et al. 2012). Several species of *Colletotrichum* have been associated with soybean (*Glycine max*) anthracnose worldwide, with *C. truncatum* globally considered the most prominent (Rogério et al. 2017; Yang et al. 2014). In recent years, emerging species such as *C. plurivorum*, *C. sojae* (Barbieri et al. 2017; Damm et al. 2019), and *C. musicola* (Bouffleur et al. 2020) were associated with the disease for the first time. Such species belong to the *Orchidearum* species complex (s.c.) (Damm et al. 2019), a monophyletic clade of species closely related to the *Orbiculare* s.c. Here, we report the reference genome sequences of four species pathogenic to soybean: *C. sojae*, *C. musicola*, and *C. plurivorum* from the *Orchidearum* s.c., and *C. truncatum*, from the *Truncatum* s.c.

Currently, two other *C. truncatum* genomes are available: MTCC3414 isolated in India from *Capsicum annuum* (Rao and Nandineni 2017) and TYU isolated in South Korea from *Taxus cuspidata* (accession numbers NBAU00000000.1 and NOWE00000000.1, respectively), both of which have been released without gene annotation. Based on the nucleotide sequences of selected loci (ACT, CHS, and GAPDH) usually used for *Colletotrichum* species identification, isolate TYU has been erroneously assigned to the wrong species and was deposited as *C. gloeosporioides* (Rogério et al. 2020).

All isolates used in this study were cultivated in potato dextrose broth medium and genomic DNA was extracted from fresh mycelial tissue. For *C. truncatum*, DNA was extracted using a modified SDS-based protocol (Giraud et al. 2013) and the final extract was purified using a Genomic-tip 100/G-Blood & Cell Culture DNA Midi Kit (Qiagen Inc., Valencia, CA). Genomic

Funding

The project was funded by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2017/09178-8), National Science and Technology Development Council (CNPq Processes 153958/2016-2 and 70328/1997-2), and National Council for the Improvement of Higher Education (CAPES/PDSE 88881.133223/2016-01, PROEX/CAPES 330002037002P3). R. Baroncelli was supported by the project Escalera de Excelencia CLU-2018-04 co-funded by the P.O. FEDER of Castilla y León 2014-2020 Spain.

Keywords

anthracnose, comparative genomics, *Glomerella*, Illumina, single molecule real-time cells, sequencing

[†]Corresponding author: R. Baroncelli; riccardobaroncelli@gmail.com

First and second authors contributed equally to this work.

The author(s) declare no conflict of interest.

Accepted for publication 29 April 2020.

Table 1. Summary statistics of *Colletotrichum truncatum*, *C. plurivorum*, *C. musicola*, and *C. sojae* genomes

Variables	Statistics			
	<i>C. truncatum</i>	<i>C. plurivorum</i>	<i>C. sojae</i>	<i>C. musicola</i>
Isolate	CMES1059	LFN00145	LFN0009	LFN0074
Culture collection	IMI 507125	IMI 507127	IMI 507126	IMI 507128
Tissue sampled	Stem	Stem	Leaf	Stem
Sequencing technology	PacBio RSII & Illumina HiSeq2000	Illumina HiSeq2000	Illumina HiSeq2000	Illumina HiSeq2000
Average coverage	124x	280x	285x	315x
Number of scaffolds	126	984	1,140	2,463
Total assembly length (Mb)	56.10	49.70	49.35	52.73
Largest contig length (bp)	2,707,168	850,834	726,643	344,569
Scaffold N50 (bp)	831,709	130,438	104,192	46,381
Scaffold L50	23	104	125	232
N's per 100 kbp	0	35.82	21.33	59.80
GC (%)	50.12	55.86	55.92	54.97
BUSCO completeness (%)	99.0	98.8	98.6	99.0
Number of predicted genes	15,901	16,153	16,124	16,826
Secreted proteins	2,116	1,989	1,931	1,988
Genome accession	VUJX000000000	WIGO000000000	WIGN000000000	WIGM000000000

DNA of the other strains was extracted using a Wizard Genomic DNA Purification Kit (Promega, Madison, WI). The *C. truncatum* genome was sequenced using a combination of long and short reads; all other genomes were sequenced using short read technology. Long reads were generated by PacBio Technology (Pacific Biosciences of California, Inc.), using two single molecule real-time cells (P5-C3 chemistry). A large-insert library (15 to 20 kb) was prepared using a PacBio large insertion library prep kit. For short reads, genomic DNA libraries of all four strains were constructed using an NEBNext Ultra Prep Kit (paired-end 2 × 150 bp, with a 500-bp insert size) and sequenced on an Illumina HiSeq2000 sequencer (Illumina Inc., San Diego, CA). Illumina reads belonging to *C. truncatum* were trimmed with Trim Galore 0.3.6. (Krueger 2012) and used to correct PacBio sequences using LoRDEC v.0.9 (Salmela and Rivals 2014). Corrected PacBio sequences were assembled with CANU v.1.8 (Koren et al. 2017). Illumina reads belonging to *C. musicola*, *C. plurivorum*, and *C. sojae* were trimmed with Trim Galore v.0.4.5 and forward and reverse reads were merged using Flash v.1.2.7 (Magoč and Salzberg 2011). De novo assemblies of combined and uncombined reads were performed with SPAdes v3.13.1 (Bankevich et al. 2012). All genomic regions with deep read coverage (such as those belonging to rRNA clusters and the mtDNA) and low coverage scaffolds were identified and masked. Assembly completeness was estimated with the program BUSCO (v3.0.2) (Simão et al. 2015), based on the sordariomyceta_odb9 lineage dataset. Gene annotation was performed with MAKER3 v3.01.02 (Carson and Yandell 2011) as described by Baroncelli et al. (2016). Gene expression profiling using RNAseq for *C. truncatum* was performed using infected plant hypocotyls at four time points (0, 12, 48, and 120 h postinfection), assembled using rnaSPAdes v3.13.1 (Bushmanova et al. 2019) and used as biological evidence for MAKER3. Secreted proteins were identified using SignalP (v.5.0b) (Nielsen 2017). Comparative analysis of the newly sequenced genomes and closely related species such as *C. orbiculare*, *C. sidae*, *C. spinosum*, *C. trifolii* (Gan et al. 2019), *C. fructicola* (Gan et al. 2013), and *C. gloeosporioides* (Alkan et al. 2015) was performed with OrthoFinder2 (Emms and Kelly 2019).

All four newly sequenced species had similar genome sizes, GC contents, and number of genes (Table 1) compared with other *Colletotrichum* species (Alkan et al. 2013; Baroncelli et al. 2014; O'Connell et al. 2012). The percentage of similarity between the proteomes was higher between species belonging to the same complex (Rogério et al. 2020). However, *C. truncatum* shared more proteins with species belonging to the Orchidearum s.c. than with the closer species, *C. gloeosporioides* (Rogério et al. 2020). The proteomes were assigned to 25,387 orthogroups, of which 7,687 are shared among all genomes analyzed.

To the best of our knowledge, the *C. truncatum* genome presented in this work is the first representative of its species with available gene annotation, while the genome sequences of *C. musicola*, *C. plurivorum*, and *C. sojae* represent new resources. The data reported here may provide insights into pathogenicity factors toward soybean and will be a useful resource for further research into comparative genomics and evolutionary studies in *Colletotrichum*.

Data deposition. The genome assemblies and annotations have been deposited at DDBJ/ENA/GenBank: *Colletotrichum truncatum* (BioProject PRJNA563581, Biosample SAMN12683944, Genome VUJX000000000, Version VUJX010000000); *Colletotrichum plurivorum* (BioProject PRJNA577396, Biosample SAMN13025176, Genome WIGO000000000,

Version WIGO01000000); *Colletotrichum sojae* (BioProject PRJNA577398, Biosample SAMN13025178, Genome WIGN00000000, Version WIGN01000000); *Colletotrichum musicola* (BioProject PRJNA577394, Biosample SAMN13025166, Genome WIGM00000000, Version WIGM01000000).

Acknowledgments

We thank P. Glaudieux and S. Cros-Arteil from the Institut National de la Recherche Agronomique (INRA) for providing methodological support in the DNA extraction and purification for PacBio sequencing, and G. Cafà for the deposit of the isolates at CABI Centre.

Literature Cited

- Alkan, N., Friedlander, G., Ment, D., Prusky, D., and Fluhr, R. 2015. Simultaneous transcriptome analysis of *Colletotrichum gloeosporioides* and tomato fruit pathosystem reveals novel fungal pathogenicity and fruit defense strategies. *New Phytol.* 205:801-815.
- 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.
- Barbieri, M. C. G., Ciampi-Guillardi, M., Moraes, S. R., Bonaldo, S. M., Rogério, F., De Castro, R. R., and Massola Júnior, N. S. 2017. First report of *Colletotrichum cliviae* causing anthracnose on soybean in Brazil. *Plant Dis.* 101:1677.
- 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., Sreenivasaprasad, S., Sukno, S. A., Thon, M. R., and Holub, E. 2014. Draft Genome Sequence of *Colletotrichum acutatum* Ssensu Lato (*Colletotrichum fioriniae*). *Genome Announc.* 2:e00112-e00114.
- Bouffleur, T. R., Castro, R. R. L., Rogério, F., and Ciampi-Guillardi, M., Baroncelli, R., and Massola, N. S., Jr. 2020. First report of *Colletotrichum musicola* causing soybean anthracnose in Brazil. *Plant Dis.* 104:1858.
- Bushmanova, E., Antipov, D., Lapidus, A., and Pribelski, A. D. 2019. RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* 8: giz100.
- Carson, H., and Yandell, M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
- Crouch, J., O'Connell, R., Gan, P., Buiate, E., Torres, M. F., Beim, L., Shirasu, K., and Vaillancourt, L. 2014. The genomics of *Colletotrichum*. Pages 123-140 in: *Genomics of Plant-Associated Fungi: Monocot Pathogens*. Springer, Berlin, Heidelberg.
- Damm, U., Sato, T., Alizadeh, A., Groenewald, J. Z., and Crous, P. W. 2019. The *Colletotrichum dracaenophilum*, *C. magnum* and *C. orchidearum* species complexes. *Stud. Mycol.* 92:1-46.
- Dean, R., Kan Van, J. A., 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.
- Emms, D. M., and Kelly, S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.
- 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., Tsushima, A., Narusaka, M., Narusaka, Y., Takano, Y., Kubo, Y., and Shirasu, K. 2019. Genome sequence resources for four phytopathogenic fungi from the *Colletotrichum orbiculare* species complex. *Mol. Plant-Microbe Interact.* 32:1088-1090.
- Giraud, E., Xu, L., Chaintreuil, C., Gargani, D., Gully, D., and Sadowsky, M. J. 2013. Photosynthetic *Bradyrhizobium* sp. strain ORS285 is capable of forming nitrogen-fixing root nodules on soybeans (*Glycine max*). *Appl. Environ. Microbiol.* 79:2459-2462.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722-736.
- Krueger, F. 2012. Trim Galore! https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- Magoč, T., and Salzberg, S. L. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957-2963.
- Nielsen, H. 2017. Predicting secretory proteins with SignalP. *Methods Mol. Biol.* 1611: 59-73.
- O'Connell, R. J., Thon, M. R., Hacquard, S., Amyotte, S. G., Kleemann, J., Torres, M. F., Damm, U., and Buiate, E. A., Epstein, L., Alkan, N., Altmüller, J., Alvarado-Balderama, 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., 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., 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.
- Rao, S., and Nandineni, M. R. 2017. Genome sequencing and comparative genomics reveal a repertoire of putative pathogenicity genes in chilli anthracnose fungus *Colletotrichum truncatum*. *PLoS One* 12:e0183567.
- Rogério, F., Bouffleur, T. R., Ciampi-Guillardi, M., Sukno, S. A., Thon, M. R., Massola Júnior, N. S., and Baroncelli, R. 2020. Genome sequence resources of *Colletotrichum truncatum*, *C. plurivorum*, *C. musicola* and *C. sojae*: Four species pathogenic to soybean (*Glycine max* L. Merr.). *Figshare* 10.6084/m9.figshare.12018426
- Rogério, F., Ciampi-Guillardi, M., Barbieri, M. C. G., Bragança, C. A. D., Seixas, C. D. S., Almeida, A. M. R., and Massola Júnior, N. S. 2017. Phylogeny and variability of *Colletotrichum truncatum* associated with soybean anthracnose in Brazil. *J. Appl. Microbiol.* 122:402-415.
- Salmela, L., and Rivals, E. 2014. LoRDEC: Accurate and efficient long read error correction. *Bioinformatics* 30:3506-3514.
- 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.
- Yang, H.-C., Haudenschild, J. S., and Hartman, G. L. 2014. *Colletotrichum incanum* sp. nov., a curved-conidial species causing soybean anthracnose in USA. *Mycologia* 106:32-42.