MPMI Vol. 34, No. 12, 2021, pp. 1461–1464, https://doi.org/10.1094/MPMI-07-21-0173-A

Complete Genome Sequence of the Plant-Pathogenic Fungus *Colletotrichum lupini*

Riccardo Baroncelli,[1,†] Flora Pensec,[2] Daniele Da Lio,[2] Thais Boufleur,[3] Isabel Vicente,[4] Sabrina Sarrocco,[4] Adeline Picot,[2] Elena Baraldi,[1] Serenella Sukno,[5] Michael Thon,[5] and Gaetan Le Floch[2]

[1] Department of Agricultural and Food Sciences (DISTAL), University of Bologna, 40127 Bologna, Italy
[2] Laboratoire Universitaire de Biodiversité et Ecologie Microbienne (LUBEM), Univ Brest, 29280 Plouzané, France
[3] Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Piracicaba, 13418-900, São Paulo, Brazil
[4] Department of Agriculture, Food and Environment (DAFE), University of Pisa, 56124 Pisa, Italy
[5] Institute for Agribiotechnology Research (CIALE), University of Salamanca, 37185 Villamayor, Spain

Abstract

Colletotrichum is a fungal genus (Ascomycota, Sordariomycetes, Glomerellaceae) that includes many economically important plant pathogens that cause devastating diseases of a wide range of plants. In this work, using a combination of long- and short-read sequencing technologies, we sequenced the genome of *Colletotrichum lupini* RB221, isolated from white lupin (*Lupinus albus*) in France during a survey in 2014. The genome was assembled into 11 nuclear chromosomes and a mitochondrial genome with a total assembly size of 63.41 Mb and 36.55 kb, respectively. In total, 18,324 protein-encoding genes have been predicted, of which only 39 are specific to *C. lupini*. This resource will provide insight into pathogenicity factors and will help provide a better understanding of the evolution and genome structure of this important plant pathogen.

Colletotrichum has been reported as one of the 10 most important plant-pathogenic fungal genera worldwide based on its scientific and economic impact (Dean et al. 2012). Anthracnose disease caused by *Colletotrichum* spp. can affect a wide range of plants in agricultural and natural ecosystems. Lupin anthracnose outbreaks began in the 1980s and rapidly spread worldwide, becoming a destructive disease affecting all lupin species. Today, this disease can cause substantial yield losses as high as 100% and is the major limiting factor for lupin production (Talhinhas et al. 2016). Lupin anthracnose is caused by *Colletotrichum lupini*, a species of the acutatum complex (Damm et al. 2012) that contrasts with other members of the latter by its host specificity (Talhinhas et al. 2016). In addition to its economic significance, *C. lupini* is a model for evolutionary research due to its peculiar host-association pattern (Baroncelli et al. 2017).

C. lupini RB221 (also known as IMI 504893 and UBOCC-A-117274) was isolated from symptomatic white lupin (*Lupinus albus*) in Brittany (France) in 2014 during a survey of *Colletotrichum* spp. associated with lupin anthracnose. This isolate was chosen because it belongs to the most representative population of *C. lupini* worldwide and because it has been previously used in pathogenicity and molecular assays (Dubrulle et al. 2020a, b).

[†]Corresponding author: R. Baroncelli; riccardo.baroncelli@unibo.it

The author(s) declare no conflict of interest.

Accepted for publication 13 August 2021.

Funding

This research and the APC were supported by the European Agricultural Fund for Rural Development PROGRAILIVE project, grant number RBRE160116CR0530019, funded by the regions of Bretagne and Pays de la Loire (France) and the European Union (FEADER); by Ascochyta Colletotrichum Lupin Pois chiche CASDAR (AsCoLuP) number 19AIP5913; The National Council for the Improvement of Higher Education (PROEX/CAPES 8887.508683/2020-00); and the Ministerio de Ciencia and Innovación of Spain (grant RTI2018-093611-B-I00).

Keywords

anthracnose, comparative genomics, complete genome, fungus-plant interactions, genomics, *Lupinus* sp., SMRT sequencing

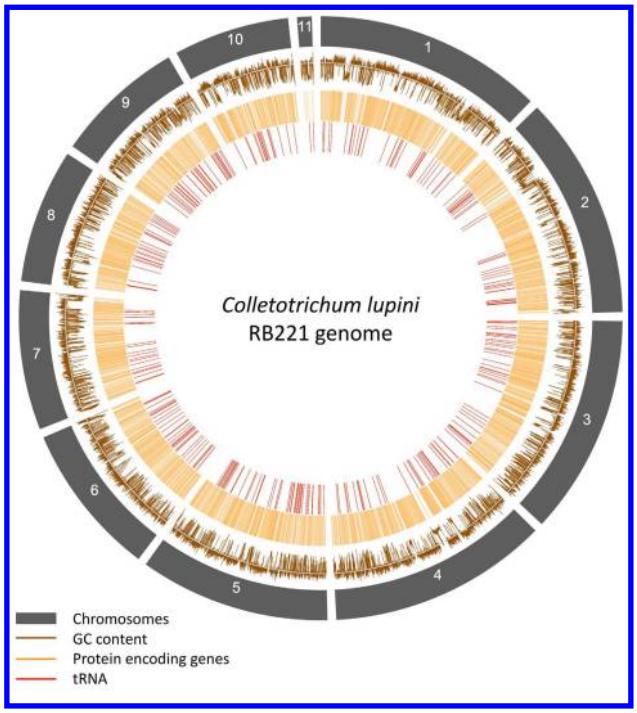

Fig. 1. Circos plot displaying *Colletotrichum lupini* RB221 genomic features. From outside to inside: chromosomes (gray blocks), GC content (brown line graph), protein encoding genes (orange lines), and transfer RNA (tRNA) genes (red lines).

High molecular weight DNA was extracted using a cetyltrimethylammonium-bromide protocol (Saghai-Maroof et al. 1984). DNA (17 μl at 242.0 ng/μl) was used for library preparation and sequencing. Illumina sequencing libraries were prepared using the Nextera DNA Library Prep kit. Samples were sequenced with an Illumina HiSeq 4000 platform (250 base, paired end). The *C. lupini* RB221 genome was also sequenced using six lanes of Pacific Biosciences (PacBio) RS II SMRT cells.

C. lupini RB221 PacBio sequences were assembled using CANU v1.7.1 (Koren et al. 2017). Illumina sequences were analyzed with FastQC (Babraham Bioinformatics) and trimmed with Trimmomatic v0.33 (Bolger et al. 2014). The trimmed sequences were used to correct the PacBio assembly using LoRDEC v0.5 (Salmela and Rivals 2014). The contigs corresponding to the mitochondrial DNA (mtDNA) genome were identified by local BLASTN v2.9.0 (Camacho et al. 2009) searches using the *C. graminicola* mitochondrial genome GenBank CM001021.1 (O'Connell et al. 2012) as the query sequence. The completeness of the assembly was assessed using BUSCO v3.1 (Simão et al. 2015) while statistics were evaluated with QUAST v5.0.2 (Gurevich et al. 2013).

The MAKER2 v2.0 annotation pipeline (Holt and Yandell 2011) was used to annotate the genome of *C. lupini* RB221. Transcriptomic data of the same isolate grown in liquid culture and during plant infection are available in the European Nucleotide Archive, bioproject accession number PRJEB40331 (Dubrulle et al. 2020b). All libraries were merged and assembled using rnaSPAdes v3.8.2. (Bushmanova et al. 2019) and aligned to the genome with HISAT v2.1.0 (Kim et al. 2019) to select only those belonging to *C. lupini* RB221. The transcript sequences were used as biological evidence in MAKER2. Three different ab initio gene annotation programs were trained for use with MAKER2. GeneMark-ES v4.10 (Borodovsky and Lomsadze 2011) was self trained for each of the four genomes. AUGUSTUS v3.3 (Stanke et al. 2006) was trained using the transcript sequences. Transfer RNA (tRNA) were predicted with tRNAscan-SE v1.3.1 (Lowe and Eddy 1997). Putative functions were assigned to the annotations using BLASTP v2.9.0 (Camacho et al. 2009) to identify homologs in a database constructed of proteins from the UniProt database (release 2013_12). Secreted proteins were identified using SignalP v.5.0b (Nielsen 2017).

Table 1. Summary of the *Colletotrichum lupini* RB221 genome assembly and annotation statistics[a]

Sequence	Accession	Topology	Telomeric repeats	Length	GC (%)	Protein encoding genes	rRNA	tRNA
Chr_1	CP019471	Linear	9/9	8,187,092	46.70	2,352	NA	33
Chr_2	CP019474	Linear	9/8	8,001,128	46.80	2,347	NA	37
Chr_3	CP019475	Linear	8/10	7,882,614	47.70	2,358	NA	38
Chr_4	CP019476	Linear	9/6	7,871,922	47.10	2,327	NA	41
Chr_5	CP019477	Linear	10/6	6,820,665	47.30	2,033	NA	51
Chr_6	CP019478	Linear	8/NA	5,502,741	46.70	1,584	23	39
Chr_7	CP019479	Linear	10/9	4,999,002	46.00	1,431	NA	24
Chr_8	CP019480	Linear	7/9	4,816,188	47.50	1,408	NA	27
Chr_9	CP019481	Linear	8/9	4,657,399	45.20	1,294	NA	38
Chr_10	CP019472	Linear	10/6	4,145,444	45.50	1,133	NA	17
Chr_11	CP019473	Linear	8/9	523,226	40.20	57	NA	2
Whole nuclear genome	–	–	–	63,407,421	46.70	18,324	23	345
Chr_mt	CP019482	Circular	NA	36,554	29.90	17	2	29

[a] rRNA = ribosomal RNA, tRNA = transfer RNA, and NA = not available.

The genome of *C. lupini* RB221 is 63.407 Mb, divided into 11 nuclear chromosomes and 1 circular mtDNA. The $_{N50}$ of the final assembly was 7,871,922 while the L_{50} was 4. Of the 11 nuclear chromosomes, 10 have telomeric repeats (TTAGGG) on both ends, and 1 has telomeric repeats on one end and gene clusters coding for ribosomal RNA on the other end. BUSCO predicted the genome to be 98.90% complete. Further genomic information can be found in Figure 1 and Table 1.

The gene annotation includes 18,324 proteins of which 1,767 (9.64%) are predicted to have signal peptides and, therefore, are predicted to be transported out of the cell into the extracellular space. Protein clustering analysis of all available *Colletotrichum* proteomes (Baroncelli et al. 2016, 2018; Huo et al. 2021) revealed that 47 clusters (48 proteins) were unique to *C. lupini*; among those, 39 lacked similarity with any other sequence available in the NCBI nonredundant protein sequence database (e-value cutoff = 10^{-3}) and, therefore, were classified as species specific. These genes may be associated with the peculiar capability of *C. lupini* to infect lupins and, therefore, selected for further analyses.

The genome sequence of *C. lupini* presented here will be useful for further research into the biology and evolution of these destructive pathogens.

Data Availability

The strain sequenced in this work has been deposited in the CABI culture collection (IMI 504893) and in the UBO Culture Collection (UBOCC-A-117274). The data generated in this study are publicly available from the NCBI GenBank database at Bioproject ID PRJNA360503 and Biosample ID SAMN06211573. The genome sequences have been deposited in GenBank under the accession CP019471-CP019482. The genome can also be accessed on the JGI MycoCosm website.

Author-Recommended Internet Resources

Babraham Bioinformatics FastQC:
https://www.bioinformatics.babraham.ac.uk/projects/fastqc
JGI MycoCosm: https://mycocosm.jgi.doe.gov/Collup1/Collup1.home.html

Literature Cited

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., Sukno, S. A., Sarrocco, S., Cafà, G., Le Floch, G., and Thon, M. R. 2018. Whole-Genome sequence of the orchid anthracnose pathogen *Colletotrichum orchidophilum*. Mol. Plant-Microbe Interact. 31:979-981.

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.

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Borodovsky, M., and Lomsadze, A. 2011. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Curr. Protoc. Bioinf. 4:4.6.1-4.6.10.

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A. D. 2019. rnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. Gigascience 8:giz100.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. 2009. BLAST+: Architecture and applications. BMC Bioinf. 10:421.

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.

Dubrulle, G., Pensec, F., Picot, A., Rigalma, K., Pawtowski, A., Gironde, S., Harzic, N., Nodet, P., Baroncelli, R., and Le Floch, G. 2020a. Phylogenetic diversity and temperature effect on growth and pathogenicity of *Colletotrichum lupini.* Plant Dis. 104:938-950.

Dubrulle, G., Picot, A., Madec, S., Corre, E., Pawtowski, A., Baroncelli, R., Zivy, M., Balliau, T., Le Floch, G., and Pensec, F. 2020b. Deciphering the Infectious Process of *Colletotrichum lupini* in Lupin through Transcriptomic and Proteomic Analysis. Microorganisms 8:1621.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.

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

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.

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37:907-915.

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.

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.

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., 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., 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.

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., and Allard, R. W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U.S.A. 81:8014-8018.

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.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. 2006. AUGUSTUS: Ab initio prediction of alternative transcripts. Nucleic Acids Res. 34:W435-W439.

Talhinhas, P., Baroncelli, R., and Le Floch, G. 2016. Anthracnose of lupins caused by *Colletotrichum lupini*: A recent disease and a successful worldwide pathogen. J. Plant Pathol. 98:5-14.