Journal of Plant Pathology
https://doi.org/10.1007/s42161-023-01353-w

GENOME RESOURCE PAPER

Draft genome sequence of the apple pathogen *Colletotrichum chrysophilum* strain M932

Greice Amaral Carneiro[1] · Martina Calì[1] · Eleonora Cappelletti[1] · Riccardo Baroncelli[1,2] · Antonio Prodi[1]

Received: 14 February 2023 / Accepted: 26 February 2023

Keywords Malus domestica · Apple Bitter Rot · Glomerella Leaf spot · Gloeosporioides · Comparative genomics

Colletotrichum chrysophilum (Ascomycota, Sordariomycetes, Glomerellaceae) is a species belonging to the *C. gloeosporioides* complex. Described in 2017 as responsible for anthracnose on *Musa acuminata* (banana plants; Vieira et al. 2017), *C. chrysophilum* has been associated with *Persea americana* (avocado) and *Prunus persica* (peach) (Talhinhas and Baroncelli 2021). Moreover, together with *Colletotrichum fructicola* and *C. noveboracense*, it is considered one of the major causal agents of Glomerella leaf spot (GLS) and Apple bitter rot (ABR) diseases on *Malus domestica* (apple) (Astolfi et al. 2022; Khodadadi et al. 2020). Originally, *C. chrysophilum* was presumed to be limited to the American and Asian continents (Astolfi et al. 2022; Talhinhas and Baroncelli 2021), however, reports of GLS and ABR caused by this pathogen in European apple orchards, such as in Italy and Spain, start emerging in 2022 (Cabrefiga et al. 2022; Deltedesco and Oettl 2022).

Colletotrichum chrysophilum was isolated in September 2021 from symptomatic leaves showing GLS symptoms from an apple orchard with a disease incidence close to 50%, in northern Italy (Province of Ferrara, Emilia-Romagna). The monosporic strain M932 was transferred onto fresh PDA medium (supplemented with 200 ml/L streptomycin and 200 ml/L neomycin) and incubated at 20 °C for 10 days to obtain mycelium for genomic DNA extraction using a modified CTAB method (Prodi et al. 2011).

The DNA of *C. chrysophilum* strain M932 was sequenced using the Illumina NovaSeq 6000 150bp paired-end sequencing system. NovaSeq 6000 adapters were trimmed using Trimmomatic v0.39 (Bolger et al. 2014) and low-quality reads were removed using TrimGalore v0.6.4 (Krueger 2015). The quality of the reads was assessed and compared using FastQC v0.11.9 (Andrews 2010). Illumina reads were assembled using SPAdes v3.15.1 (Bankevich et al. 2012). The first draft of the nuclear genome of *C. chrysophilum* consists of 1497 scaffolds with a total length of 55.56 Mbp (N50= 86538 bp and N75= 44545 bp). BUSCO v5.2.2 (Seppey et al. 2019) software was used to assess the integrity of the fungal genome assembly while assembly statistics were evaluated with QUAST v5.0.2 (Gurevich et al. 2013). Results are reported in Table 1.

A total of 20,041 protein-coding genes were predicted to be encoded by the nuclear using MAKER v3.01.02 pipeline (Holt and Yandell 2011) with self-trained GeneMark-ES v4.10 (Borodovsky and Lomsadze 2011) and AUGUSTUS v3.3 prediction performed using the "Fusarium" model (Stanke et al. 2008). SignalP v5.0 (Almagro Armenteros et al. 2019) revealed that 2,350 proteins in *C. chrysophilum* are secreted and among those 991 have been predicted to be candidate effectors by EffectorP v3.0 (Sperschneider and Dodds 2022). A comparative analysis of the newly sequenced genome with those publicly

Riccardo Baroncelli and Antonio Prodi contributed equally to this work.

✉ Antonio Prodi
antonio.prodi@unibo.it

[1] Department of Agricultural and Food Sciences (DISTAL), University of Bologna, 40126 Bologna, Italy

[2] Center for Studies on Bioinspired Agro-Enviromental Technology, University of Naples Federico II, 80055 Portici, Italy

Published online: 28 March 2023

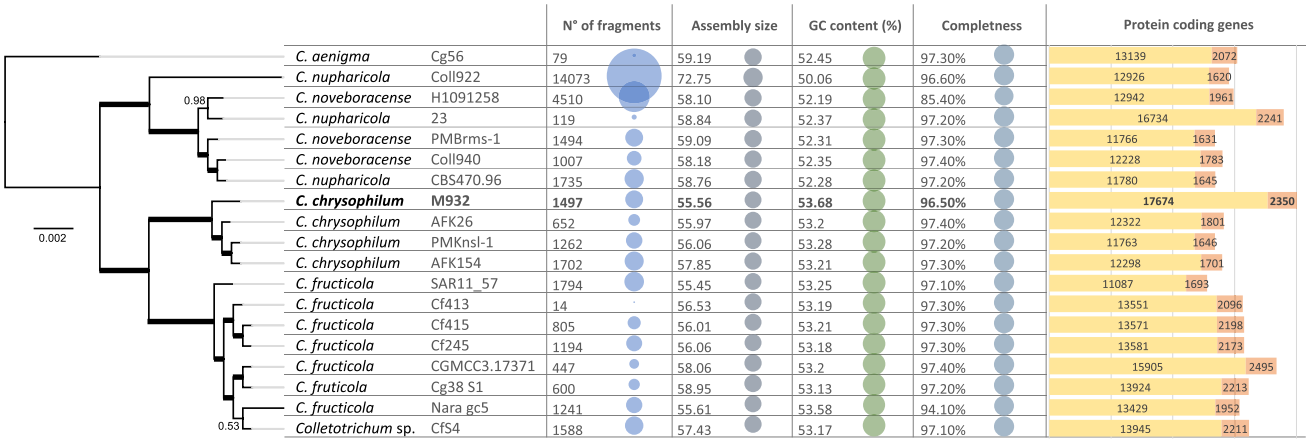

Fig. 1 Comparative analysis of the newly sequenced genome with those of closely related species publicly available. The genome sequenced in the present study is highlighted in bold. On the left side a phylogenomic tree showing the evolutionary relationships between genomes; number next to the nodes represent Bayesian posterior probability (BPP) values while thicker branches indicate a support value of BPP = 1.00. In the center four bubble plots illustrating assembly fragmentation, size, GC content and completeness. The bubble sizes have been scaled to each panel and are not comparable across panels. The bar diagram on the right reports the size of not secreted (in yellow) and secreted (in orange) predicted protein encoding genes

available (Gan et al. 2013; Armitage et al. 2020; Gan et al. 2021; Baroncelli et al. 2022) showed similar genomic features in terms of genome size and GC% but a high diversity in gene content within strains of *C. chrysophilum* and with closely related species (Fig. 1). A phylogenomic approach, performed as described in Baroncelli et al. 2022 did also highlight incongruence in the taxonomic designation of deposited data as strains C. *nupharicola* and *C. noveboracense* do not form distinct clusters (Figure 1); further analyses are needed to fully understand the diversity and the taxonomy of this group.

The availability of the genome of *C. chrysophilum* M932 offers the possibility to perform further comparative analyses, to fully understand species boundaries within the *Colletotrichum gloeosporioides* species complex and to develop molecular diagnostic methods.

Table 1 Summary statistics of the *Colletotrichum chrysophilum* M932 genome

Assembly Variables	Statistics
Assembly length (Mbp)	59,19
Number of scaffolds	1497
Largest scaffold size (bp)	406526
N50	86538
N75	44545
L50	210
L75	428
GC (%)	53,68
BUSCO completeness	96,50%
Complete and single-copy	96,20%
Complete and duplicated	0,30%
Fragmented	1,00%
Missing	2,50%
Protein encoding genes	
Number of predicted genes	20024
Number of predicted secreted proteins	2350
Number of predicted effector proteins	991
Number of predicted cytoplasmic effectors	458
Number of predicted apoplastic effectors	533

Nucleotide sequence accession numbers

This whole-genome shotgun project has been deposited in GenBank under the accession no. JAQOWY000000000 (BioProject: PRJNA928458; BioSample: SAMN32933927).

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement. This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 103217/06/2022, CN00000022).

Data Availability The "data availability statement" is reported in the "Nucleotide sequence accession numbers" section.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37(4):420–423. https://doi.org/10.1038/s41587-019-0036-z

Armitage AD, Nellist CF, Bates HJ, Zhang L, Zou X, Gao QH, Harrison RJ (2020) Draft genome sequence of the strawberry anthracnose pathogen *Colletotrichum fructicola*. Microbiol Resour Announc 9(12):e01598-19. https://doi.org/10.1128/mra.01598-19

Astolfi P, Velho AC, Moreira V, Mondino PE, Alaniz SM, Stadnik MJ (2022) Reclassification of the main causal agent of glomerella leaf spot on apple into *Colletotrichum chrysophilum* in southern Brazil and Uruguay. Phytopathology 112(9):1825–1832. https://doi.org/10.1094/PHYTO-12-21-0527-SC

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477. https://doi.org/10.1089/cmb.2012.0021

Baroncelli R, Cobo-Díaz JF, Benocci T, Peng M, Battaglia E, Andreopoulos W, LaButti K, Pangilinan J, Lipzen A, Koriabine M, Bauer D, le Floch G, Mäkelä MR., Drula E, Henrissat B, Anne Crouch J, de Vries RP, Sukno SA, Thon MR (2022) Genome evolution and transcriptome plasticity associated with 1 adaptation to monocot and eudicot plants in Colletotrichum fungi. bioRxiv 2022.09.22.508453t https://doi.org/10.1101/2022.09.22.508453

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borodovsky M, Lomsadze A (2011) Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Curr Protoc Bioinf 35:4.6.1–4.6.10. https://doi.org/10.1002/0471250953.bi0406s35

Cabrefiga J, Piz D, Vilardell P, Luque J (2022) First report of *Colletotrichum chrysophilum* causing apple bitter rot in Spain. Plant Dis 106(6). https://doi.org/10.1094/PDIS-07-21-1578-PDN

Deltedesco E, Oettl S (2022) First report of preharvest decay caused by *Colletotrichum chrysophilum* on apples in Italy (South Tyrol). Plant Dis First look. https://doi.org/10.1094/PDIS-11-21-2453-PDN

Gan P, Hiroyama R, Tsushima A, Masuda S, Shibata A, Ueno A, Kumakura N, Narusaka M, Hoat TX, Narusaka Y, Takano Y, Shirasu K (2021) Telomeres and a repeat-rich chromosome encode effector gene clusters in plant pathogenic *Colletotrichum* fungi. Environ Microbiol 23(10):6004–6018. https://doi.org/10.1111/1462-2920.15490

Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, Takano Y, Kubo Y, Shirasu K (2013) Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol 4:1236–1249. https://doi.org/10.1111/nph.12085

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29(8):072–1075. https://doi.org/10.1093/bioinformatics/btt086

Holt C, Yandell M (2011) MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinform 12:491. https://doi.org/10.1186/1471-2105-12-491

Khodadadi F, González JB, Martin PL, Giroux E, Bilodeau GJ, Peter KA, Doyle VP, Aćimović SG (2020) Identification and characterization of *Colletotrichum* species causing apple bitter rot in New York and description of *C. noveboracense* sp. nov. Sci Rep 10:11043. https://doi.org/10.1038/s41598-020-66761-9

Krueger F (2015) Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore.

Prodi A, Purahong W, Tonti S, Salomoni D, Nipoti P, Covarelli L, Pisi A (2011) Difference in chemotype composition of Fusarium graminearum populations isolated from durum wheat in adjacent areas separated by the Apennines in Northern-Central Italy. Plant Pathol J 27(4): 354–359). https://doi.org/10.5423/PPJ.2011.27.4.354

Seppey M, Manni M, Zdobnov EM (2019) BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol 1962:227–245. https://doi.org/10.1007/978-1-4939-9173-0_14

Sperschneider J, Dodds PN (2022) EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. Mol Plant-Microbe Interact 35(2):146–156. https://doi.org/10.1094/MPMI-08-21-0201-R

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24(5):637–644. https://doi.org/10.1093/bioinformatics/btn013

Talhinhas P, Baroncelli R (2021) *Colletotrichum* species and complexes: geographic distribution, host range and conservation status. Fungal Divers 110(1):109–198. https://doi.org/10.1007/s13225-021-00491-9

Vieira WAS, Lima WG, Nascimento ES, Michereff SJ, Câmara MPS, Doyle VP (2017) The impact of phenotypic and molecular data on the inference of *Colletotrichum* diversity associated with Musa. Mycologia 109(6):912–934. https://doi.org/10.1080/00275514.2017.1418577