

# Genome Resources for the Endophytic Fungus *Paraphaeosphaeria sporulosa*

Riccardo Baroncelli,<sup>1,†</sup> Daniele Da Lio,<sup>2</sup> Giovanni Vannacci,<sup>2</sup> and Sabrina Sarrocco<sup>2,†</sup>

<sup>1</sup> Instituto Hispano-Luso de Investigaciones Agrarias (CIALE), University of Salamanca, Calle del Duero 12, 37185 Villamayor (Salamanca), Spain

<sup>2</sup> Department of Agriculture, Food and Environment (DAFE), University of Pisa, Via del Borghetto 80, 56124, Pisa, Italy

## Abstract

*Paraphaeosphaeria* genus includes plant pathogens or biocontrol agents as well as bio-remediators and endophytic fungi. *Paraphaeosphaeria sporulosa* 10515 was isolated in 2013 as an endophyte of *Festuca* spp. collected on Mount Etna at 1,832 meters above sea level. Here, we present the first-draft whole-genome sequence of a *P. sporulosa* endophytic isolate. This data will be useful for future research on understanding the genetic bases of endophytism.

## Funding

The research by R. Baroncelli is supported by the project Escalera de Excelencia co-funded by the P.O. FEDER of Castilla y León, Spain (CLU-2018-04). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Keywords

endophytism, *Festuca* spp., genome resource, *Paraphaeosphaeria sporulosa*

## Resource Announcement

Genus *Paraphaeosphaeria* generally includes plant pathogens or biocontrol agents as well as bioremediators and endophytic fungi able to produce antifungal or cytotoxic compounds (de Amorim et al. 2019; Liu et al. 2017; Verkley et al. 2014). First described as *Coniothyrium fuckelii* var. *sporulosum* (Gams and Domsch 1969), *Paraphaeosphaeria sporulosa* (W. Gams & Domsch) was reassessed in 2014 by Verkley et al. (2014) on the basis of internal transcribed spacer, large subunit,  $\gamma$ -actin, and  $\beta$ -tubulin gene sequences.

In 2013, *Paraphaeosphaeria sporulosa* 10515 was isolated as an endophyte from a root of *Festuca* sp. collected in an open field on Mount Etna (Italy), at 1,832 meters above sea level (upper belt of the mountain-Mediterranean zone). As a large, active volcano, Mount Etna presents a complex assemblage of plant species modulated by volcanic soil, climate, and human activity. The sampling area was characterized by a young lava substrate with sparse vegetation of herbaceous plants. When plated on PDA (potato dextrose agar) with streptomycin (300 mg per liter), only colonies morphologically referable to as *Paraphaeosphaeria* spp. developed from *Festuca* sp. roots surface-sterilized by an aqueous solution of NaClO (1% active chlorine) in 50% EtOH. The isolate is currently deposited in the fungal collection of the Department of Agriculture, Food and Environment of the University of Pisa (Italy), stored at 4°C on PDA under mineral oil.

The genome of *P. sporulosa* 10515 was sequenced using Illumina MiSeq paired-end sequencing technology and results are summarized in Table 1. Paired reads of 300 bp (2.95 Gbp, average nuclear coverage 30x) were assembled using SPAdes v3.13.1 (Bankevich et al. 2012). The nuclear genome of *P. sporulosa* consists of 565 scaffolds, with a total assembly length of 43.65 Mbp ( $N_{50}$  = 503,765 and  $L_{50}$  = 29), 52.35% GC-content, and a

<sup>†</sup>Corresponding authors: R. Baroncelli; [riccardobaroncelli@gmail.com](mailto:riccardobaroncelli@gmail.com) and S. Sarrocco; [sabrina.sarrocco@unipi.it](mailto:sabrina.sarrocco@unipi.it)

The author(s) declare no conflict of interest.

Accepted for publication 25 May 2020.

**Table 1.** Summary statistics of the *Paraphaeosphaeria sporulosa* 10515 genome

Variables	Statistics
Number of scaffolds	565
Assembly length (bp)	43,647,829
Maximum scaffold length	1,503,490
N <sub>50</sub>	503,765
L <sub>50</sub>	29
GC (%)	52.35
BUSCO completeness	98.9%
Number of predicted genes	16,682
Number of predicted secreted proteins	1,882

maximum scaffold size of 1,503,490 bp. Genome completeness was assessed through the presence of conserved single-copy fungal genes, using BUSCO v3 (Simão et al. 2015; Waterhouse et al. 2018). We used the *fungi\_odb9* data set, identifying 287 of 290 (98.9%) of these genes as present in the final assembly. The nuclear genome was annotated using the MAKER v3.01.02 pipeline (Campbell et al. 2014; Holt and Yandell 2011). Publicly available transcriptomic data (the *P. sporulosa* strain AP3s5-JAC2a; sequence reads archive accession number SRX2052362) were used to train Augustus v. 2.5.5 (Stanke and Morgenstern, 2005) and as biological evidence, while GeneMark-ES v. 4.48 (Borodovsky and Lomsadze 2011) was used for *ab initio* gene prediction. Overall, 16,682 protein-coding gene models were predicted, and analysis with SignalP v5.0 (Petersen et al. 2011) revealed that 1,882 predicted proteins (11.28% of the proteome) contain a secretion signal peptide.

Another strain of *P. sporulosa* (AP3s5-JAC2a) has been fully sequenced, with the genome sequences publicly available. *P. sporulosa* AP3s5-JAC2a was isolated from a natural freshwater lake historically polluted with elevated concentrations of phosphate and metals (Zeiner et al. 2016). To the best of our knowledge the genome here described is the second available of its species but the first related to a plant endophytic strain.

In the present study, we provide the draft genome sequence of *P. sporulosa* 10515, a strain endophytic of *Festuca* sp. from Mount Etna. The sequence represents a new useful source for future research in plant endophytism and in the evolution of Pleosporales.

## Nucleotide sequence accession numbers

This Whole-Genome Shotgun project has been deposited in GenBank under accession number JAAQOO000000000 (BioProject PRJNA612332; BioSample SAMN14368166). The version described in this paper is JAAQOO000000000.1.

## Literature Cited

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulkov, 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.
- Borodovsky, M., and Lomsadze, A. 2011. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr. Prot. Bioinform.* 35:4.6.1-4.6.10.
- Campbell, M. S., Holt, C., Moore, B., Yandell, M. 2014. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinform.* 48:4.11.1-4.11.39.
- de Amorim, M. R., Hilário, F., Dos Santos, F. M., Junior, Batista, J. M., Junior, Bauab, T. M., Araújo, A. R., Carlos, I. Z., Vilegas, W., and Dos Santos, L. C. 2019. New benzaldehyde and benzopyran compounds from the endophytic fungus *Paraphaeosphaeria* sp. F03 and their antimicrobial and cytotoxic activities. *Planta Med.* 85:957-964.
- Gams, W., and Domsch, K. H. 1969. Bemerkungen zu einigen schwer bestimmbarren Bodenpilzen. *Nova Hedwigia* 18:1-29.
- Holt, C., and Yandell, M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
- Liu, T., Greenslade, A., and Yang, S. 2017. Levels of rhizome endophytic fungi fluctuate in *Paris polyphylla* var. *yunnanensis* as plants age. *Plant Divers.* 39:60-64.
- 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.
- Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33:W465-W467.
- Verkley, G. J. M., Dukik, K., Renfurm, R., Göker, M., and Stielow, J. B. 2014. Novel genera and species of coniothyrium-like fungi in *Montagnulaceae* (Ascomycota). *Persoonia* 32:25-51.
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., and Zdobnov, E. M. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35:543-548.
- Zeiner, C. A., Purvine, S. O., Zink, E. M., Paša-Tolić, L., Chaput, D. L., Haridas, S., Wu, S., LaButti, K., Grigoriev, I. V., Henriessat, B., Santelli, C. M., and Hansel, C. M. 2016. Comparative analysis of secretome profiles of manganese(ii)-oxidizing ascomycete fungi. *PLoS One* 11:e0157844.