MPMI Vol. X, No. X, XXXX, pp. X-X, https://doi.org/10.1094/MPMI-09-22-0185-A

RESOURCE ANNOUNCEMENT

Complete Genome Sequences of *Septoria linicola*: A Resource for Studying a Damaging Flax Pathogen

Nicolas Lapalu,[1] Adeline Simon,[1] Boris Demenou,[2] Delphine Paumier,[3] Marie-Pierre Guillot,[1] Lilian Gout,[1] Frederic Suffert,[1] and Romain Valade[3,†]

[1] Université Paris Saclay, INRAE, UR BIOGER, 78850 Thiverval-Grignon, France
[2] ARVALIS Institut du Végétal, 31450 Baziège, France
[3] ARVALIS Institut du Végétal, 91720 Boigneville, France

Abstract

Fungal genus *Septoria* causes diseases in a wide range of plants. Here, we report the first genome sequences of two strains of *Septoria linicola*, the causal agent of the pasmo disease of flax (*Linum usitatissimum*). The genome of the first strain, SE15195, was fully assembled in 16 chromosomes, while 35 unitigs were obtained for a second strain, SE14017. Structural annotations predicted 13,096 and 13,085 protein-encoding genes and transposable elements content of 19.0 and 18.1% of the genome for SE15195 and SE14017, respectively. The four smaller chromosomes 13 to 16 show genomics features of potential accessory chromosomes. The assembly of these two genomes is a new resource for studying *S. linicola* and improving management of pasmo.

Funding

This work was supported by the French Ministry for Agriculture and Food Cares through the C-2014-03 grant ('SeptoLIN' project 2014-2018) from the CasDAR program (the French Special Agricultural and Rural Development Account). The INRAE BIOGER (fungal Biology and Risk Management in Agriculture) laboratory benefits from the support of Saclay Plant Science-SPS (ANR-17-EUR-0007).

Keywords

flax, genome, pasmo, *Septoria linicola*

Septoria linicola (Speg.) Garassini (teleomorph *Mycosphaerella linicola*) is, as a plant pathogen of class Ascomycete (Verkley et al. 2013), the causal agent of the disease pasmo in flax (*Linum usitatissimum* L.), both in linen (fiber flax) and linseed (oil flax). This disease affects production in many flax-growing areas around the world, including France, with an increasing incidence over the last two decades (Paumier et al. 2021). Despite the economic impact of *S. linicola*, no complete sequence data of this fungal species is available in public databases, although this kind of data has proven to be very useful for studying plant-pathogen interactions and helping to control diseases.

A strain of *S. linicola* was chosen for each of the two types of flax crops, to take into consideration the potential adaptation to the host, including potential differences in aggressiveness, and thus maximize the genetic and genomic diversity covered in the sequences. The strain SE14017 was isolated in 2014 from linen (cultivar Melina at Boissy Sans Avoir, France) and the strain SE15195 in 2015 from linseed (cultivar Angora at Lavaur, France). DNAs were extracted, using phenol, with a method adapted from Zhong et al. (2017). Library preparation and sequencing were performed by the GeT-PlaGe core facility at INRAE Toulouse. Five PacBio RS II single-molecule real-time cells were loaded per strain, providing long reads with an average length of 12 Kb, with a coverage of 60× for SE14017 and 80× SE15195, for an expected genome size close to 40 Mb.

Genome assemblies were generated from several runs of the Canu v2.1.1 (Koren et al. 2017) and Flye v2.8.2 (Kolmogorov et al. 2019) assemblers. For strain SE15195, the best assembly, i.e., in complete chromosome molecules, was obtained from Canu,

[†]Corresponding author: R. Valade; r.valade@arvalis.fr

The author(s) declare no conflict of interest.

Accepted for publication 3 October 2022.

Table 1. Summary statistics of whole-genome sequencing from the two *Septoria linicola* strains (SE15195 and SE14017)

Feature	SE15195	SE14017
Assembly length (Mbp)	39.6	39.1
Chromosomes/unitigs	16	35[a]
Mitochondrion length (Kbp)	36.9	36.9
% Transposable elements	19.0	18.1
Number of predicted genes	13,096	13,085
% BUSCO completeness	99.5	99.4
Number of predicted effectors	281	274

[a] Unitigs.

with parameters minReadLength = 2,000, minOverlapLength = 1,000, and corrected ErrorRate = 0.035, while the best assembly for SE14017 was obtained from the combination of selected molecules reconciliating different assemblies. Draft genome assemblies and raw reads were respectively aligned with nucmer (Kurtz et al. 2004) and were mapped with Blasr to the best consensus assembly, to validate the final genome assembly. The draft genomes were polished with the Arrow algorithm. The mitochondrial genomes were assembled with Organelle_PBA (Soorni et al. 2017), using the *Zymoseptoria tritici* IPO323 mitochondrial genome as reference (Goodwin et al. 2011).

As no transcriptional data were available, gene models were predicted by GeneMark-EP+ (Brůna et al. 2020), with the ProtHint pipeline, from 20 sources of proteomes of different *Mycosphaerellales* genera (*Septoria*, *Zymoseptoria*, *Cercospora*, *Pseudocercospora*, and *Ramularia*). The same proteomes were aligned with Exonerate (Slater and Birney 2005) and were used as evidence to select and filter out gene models. The completeness of the assemblies was evaluated with BUSCO (Simão et al. 2015), using the Ascomycota (odb10) gene set as evidence. Genes encoding fungal effectors were detected with Predector (Jones et al. 2021), a bioinformatics pipeline composed of different versions of tools, such as EffectorP (Sperschneider 2016, 2018), SignalP (Almagro Armenteros et al. 2019; Bendtsen et al. 2004; Petersen et al. 2011), TMHMM (Krogh et al. 2001), Phobius (Käll et al. 2004), or TargetP (Armenteros et al. 2019). Results were then post-filtered with criteria: effector_score $>$ 0, manual_effector_score $>$ 2.5, manual_secretion_score $>$ 2, any_signal_peptide = 1, single_transmembrane = 0, residue_number $<$ 300.

Transposable elements (TEs) were annotated using the REPET package (Amselem et al. 2015). Briefly, the TEdenovo pipeline (Flutre et al. 2011) was used to detect repeated elements in the genome and to provide a consensus sequence for each family. Consensus sequences were then classified using the PASTEC tool (Hoede et al. 2014), based on the Wicker hierarchical TE classification system (Wicker et al. 2007). After manual correction, the resulting library of consensus sequences was used to annotate TE copies in the whole genome, using the TEannot pipeline (Quesneville et al. 2005).

Synteny between both strains was analyzed with SynChro (Drillon et al. 2014) with a delta parameter of 3.

Phylogeny was reconstructed with the 28S large subunit of the nrRNA (LSU) gene to determine whether the two strains belonged to the *Septoria* clade. The *Septoria azaleae* LSU sequence homologs were determined for SE14017 and SE15195 strains and for three reference genomes (*Cercospora zeae-maydis* SCOH1-5, *Z. tritici* IPO323, and *Sphaerulina musiva* SO2202), using blastn search (Altschup et al. 1990). The homolog sequences were aligned with LSU sequence data of 21 taxa (including eight *Septoria* species) representing most of the clades used by Quaedvlieg et al. (2011), using Geneious Prime 2022.1.1, resulting in alignment of 760 bp of sequence length. A consensus neighbor-joining phylogenetic tree was reconstructed, using FastTree 2.1.11 (Price et al. 2010), and was rooted with *Cladosporium bruhnei*.

The genome of *S. linicola* SE15195 is 39.6 Mb long with 16 complete chromosomes, all ended with telomere repeats (TTAGGG). The assembly of strain SE14017 is close to being complete with 35 unitigs for a total size of 39.1 Mb. Telomere repeats are present in 22 unitigs, among them, four are complete molecules with telomeres at both extremities. Mitochondrial genomes of both strains are 36.9 Kb in length (Table 1).

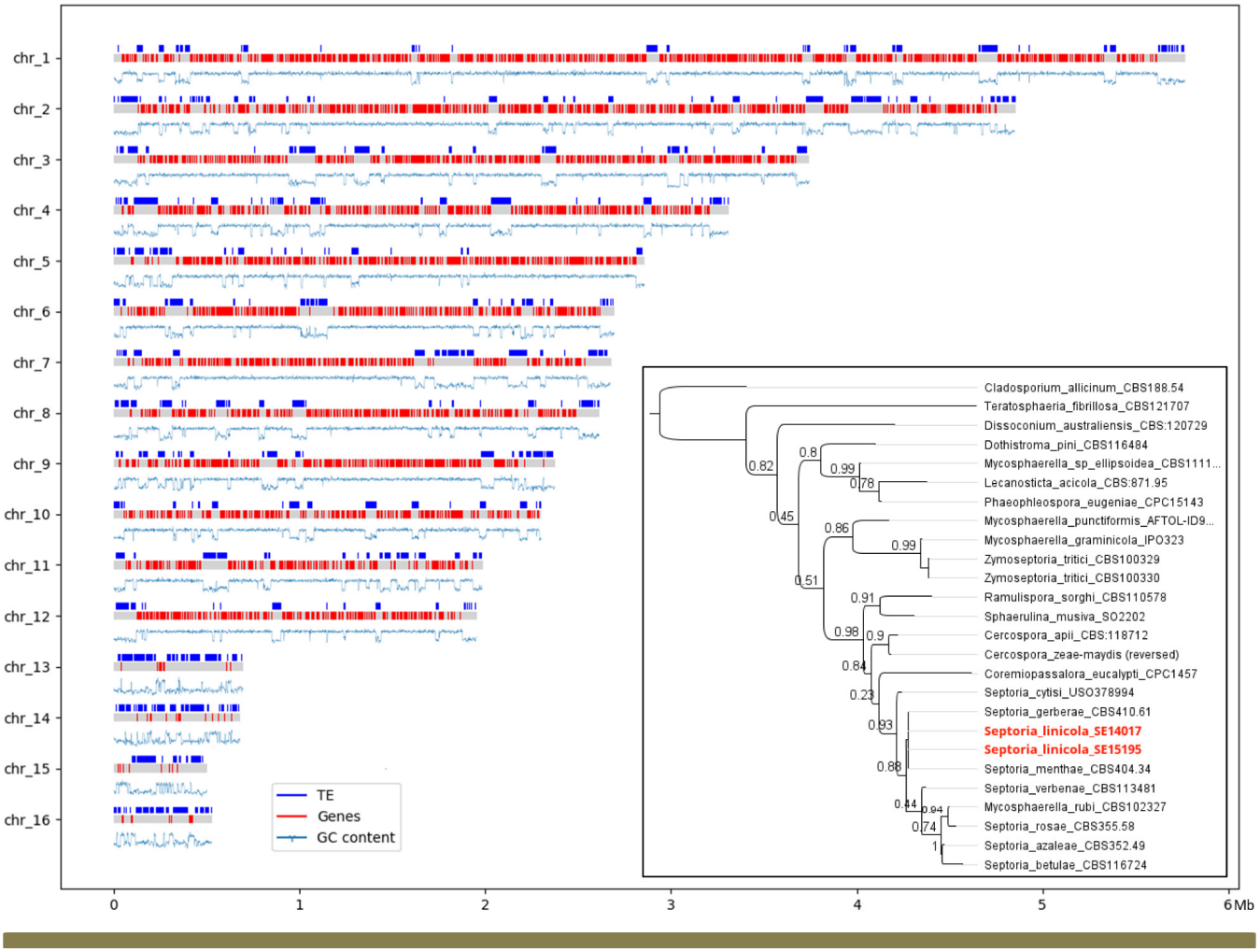

Fig. 1. Karyoplot of *Septoria linicola* SE15195, showing the 16 chromosomes with genes (red), transposable elements (blue), and the GC content (light-blue line). The insert shows the phylogenetic tree of the two *S. linicola* strains (SE15195 and SE14017) within the *Capnodiales* order (*Dothideomycetes*), with branches labeled by bootstrap support from 1,000 replicates.

The genome annotation provided 13,096 and 13,085 genes for strain SE15195 and SE14017, respectively, corresponding to a completeness of conserved Ascomycota genes close to 99.5%, according to BUSCO. Among them, 281 genes were identified as potential effectors for SE15195 and 274 genes for SE14017.

The TE annotation allowed to identification of 22 and 21 subfamily elements according to the Wicker classification, covering 19 and 18% of the genome of SE15195 and SE14017, respectively. The TE content ranged between 11 and 23% of SE15195's longest chromosomes (1 to 12) and 47 to 68% of the four smallest chromosomes (13 to 16).

The large TE content and the few genes mainly predicted by the *ab initio* method without evidence suggests that chromosomes 13 to 16 might be accessory chromosomes (Fig. 1) (Bertazzoni et al. 2018).

The synteny between both strains detected 13,042 pairs of orthologs involving 12,779 SE14017 genes and 12,862 SE15195 genes, revealing a potential gene split or duplication. The few synteny breaks observed were mainly detected around regions enriched in TEs, which may be a source of evolution or artefact in gene prediction.

The phylogenetic analysis with the LSU genetic marker confirmed the taxonomical classification of both strains as genus *Septoria* (Fig. 1).

The genome sequences of the two strains of *S. linicola*, SE15195 and SE14017, isolated from two flax varieties, should be considered as reference strains. They will be useful for further research on this fungal pathogen, especially for specific analysis of plant-pathogen interaction in fiber flax and oilseed flax.

Data Availability

The strains sequenced in this work have been deposited in the Arvalis culture collection.

The genome assemblies and annotations have been deposited at DDBJ/ENA/GenBank under the accessions CP099418 to CP099434 (SE15195) and JAMGVY000000000 (SE14017), with associated BioProject accessions PRJNA833952 and PRJNA833955, respectively.

The genomes can also be accessed on the BIOINFOBIOGER platform and browsed with dedicated genome browsers. Further details (e.g., gene lists, scaffolding and synteny map of SE14017 on SE15195) are also available on each genome page.

Acknowledgments

The authors are grateful to the GenoToul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul) for providing help amd computing and storage resources.

Author-Recommended Internet Resources

Bioinfo Genotoul: http://bioinfo.genotoul.fr
BIOINFOBIOGER platform: https://bioinfo.bioger.inrae.fr/portal/genome-portal

Literature Cited

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37:420-423.

Altschup, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Amselem, J., Lebrun, M. H., and Quesneville, H. 2015. Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. BMC Genomics16:141.

Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., and Nielsen, H. 2019. Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance 2:e201900429.

Bendtsen, J. D., Nielsen, H., Von Heijne, G., and Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340:783-795.

Bertazzoni, S., Williams, A. H., Jones, D. A., Syme, R. A., Tan, K. - C., and Hane, J. K. 2018. Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in plant-pathogenic fungi. Mol. Plant-Microbe Interact. 31:779-788.

Brůna, T., Lomsadze, A., and Borodovsky, M. 2020. GeneMark-EP+: Eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics Bioinform. 2:lqaa026.

Drillon, G., Carbone, A., and Fischer, G. 2014. SynChro: A fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. PLoS One 9:e92621.

Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. 2011. Considering transposable element diversification in de novo annotation approaches. PLoS One 6.

Goodwin, S. B., M'Barek, S. B., Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., Foster, A. J., van der Lee, T. A. J., Grimwood, J., Aerts, A., Antoniw, J., Bailey, A., Bluhm, B., Bowler, J., Bristow, J., van der Burgt, A., Canto-Canché, B., Churchill, A. C. L., Conde-Ferràez, L., Cools, H. J., Coutinho, P. M., Csukai, M., Dehal, P., de Wit, P., Donzelli, B., van de Geest, H. C., van Ham, R. C. H. J., Hammond-Kosack, K. E., Henrissat, B., Kilian, A., Kobayashi, A. K., Koopmann, E., Kourmpetis, Y., Kuzniar, A., Lindquist, E., Lombard, V., Maliepaard, C., Martins, N., Mehrabi, R., Nap, J. P. H., Ponomarenko, A., Rudd, J. J., Salamov, A., Schmutz, J., Schouten, H. J., Shapiro, H., Stergiopoulos, I., Torriani, S. F. F., Tu, H., de Vries, R. P., Waalwijk, C., Ware, S. B., Wiebenga, A., Zwiers, L. H., Oliver, R. P., Grigoriev, I. V., and Kema, G. H. J. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 7:1002070.

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., and Quesneville, H. 2014. PASTEC: An automatic transposable element classification tool. PLoS One 9:e91929.

Jones, D. A. B., Rozano, L., Debler, J. W., Mancera, R. L., Moolhuijzen, P. M., and Hane, J. K. 2021. An automated and combinative method for the predictive ranking of candidate effector proteins of fungal plant pathogens. Sci. Rep. 11: 19731.

Käll, L., Krogh, A., and Sonnhammer, E. L. L. 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338: 1027-1036.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37:540-546.

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 adaptivek-mer weighting and repeat separation. Genome Res. 27:722-736.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. L. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305:567-580.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S. L. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12.

Paumier, D., Bammé, B., Penaud, A., Valade, R., and Suffert, F. 2021. First report of the sexual stage of the flax pathogen *Mycosphaerella linicola* in France and its impact on pasmo epidemiology. Plant Pathol. 70:475-483.

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.

Price, M. N., Dehal, P. S., and Arkin, A. P. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One 5.

Quaedvlieg, W., Kema, G. H. J., Groenewald, J. Z., Verkley, G. J. M., Seifbarghi, S., Razavi, M., Mirzadi Gohari, A., Mehrabi, R., and Crous, P. W. 2011. *Zymoseptoria* gen. nov.: e9490 A new genus to accommodate *Septoria*-like species occurring on graminicolous hosts. Persoonia Mol. Phylogeny Evol. Fungi. 26: 57-69.

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., and Anxolabehere, D. 2005. Combined evidence annotation of transposable elements in genome sequences. PLoS Comput. Biol. 1:0166-0175.

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 Bioinformatics 6:31.

Soorni, A., Haak, D., Zaitlin, D., and Bombarely, A. 2017. Organelle_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. BMC Genomics 18:49.

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., and Taylor, J. M. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol. Plant Pathol. 19:2094-2110.

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., and Taylor, J. M. 2016. EffectorP: Predicting fungal

effector proteins from secretomes using machine learning. New Phytol. 210: 743-761.

Verkley, G. J. M., Quaedvlieg, W., Shin, H. D., and Crous, P. W. 2013. A new approach to species delimitation in Septoria. Stud. Mycol. 75:213-305.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., and Schulman, A. H. 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8:973-982.

Zhong, Z., Marcel, T. C., Hartmann, F. E., Ma, X., Plissonneau, C., Zala, M., Ducasse, A., Confais, J., Compain, J., Lapalu, N., Amselem, J., McDonald, B. A., Croll, D., and Palma-Guerrero, J. 2017. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the *Stb6* resistance gene. New Phytol. 214.