MPMI Vol. 34, No. 7, 2021, pp. 870–873. https://doi.org/10.1094/MPMI-01-21-0016-A

ONT-Based Draft Genome Assembly and Annotation of *Alternaria atra*

e-Xtra*

Bhawna Bonthala,[1] Corinn S. Small,[1] Maximilian A. Lutz,[1] Alexander Graf,[2] Stefan Krebs,[2] German Sepúlveda,[3] and Remco Stam[1†]

[1] Chair of Phytopathology, School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
[2] Gene Centre Munich, Laboratory for Functional Genome Analysis, Ludwig-Maximilians-University, Munich, Germany
[3] Departmento de recursos Ambientales, Facultad de Ciencias Agronómicas, Universidad de Tarapacá, Arica, Chile

Funding
This project was funded by the German Science Foundation (Deutsche Forschungsgemeinschaft, grant 403835372) and the Technische Universität München Seed Fund.

Abstract

Species of *Alternaria* (phylum Ascomycota, family Pleosporaceae) are known as serious plant pathogens, causing major losses on a wide range of crops. *Alternaria atra* (previously known as *Ulocladium atrum*) can grow as a saprophyte on many hosts and causes Ulocladium blight on potato. It has been reported that it can also be used as a biocontrol agent against *Botrytis cinerea*. Here, we present a scaffold-level reference genome assembly for *A. atra.* The assembly contains 43 scaffolds with a total length of 39.62 Mbp, with scaffold N_{50} of 3,893,166 bp, L_{50} of 4, and the longest 10 scaffolds containing 89.9% of the assembled data. RNA-sequencing-guided gene prediction using BRAKER resulted in 12,173 protein-coding genes with their functional annotation. This first high-quality reference genome assembly and annotation for *A. atra* can be used as a resource for studying evolution in the highly complicated *Alternaria* genus and might help in understanding the mechanisms defining its role as pathogen or biocontrol agent.

The fungal genus *Alternaria* includes endophytic, pathogenic, and saprophytic species that are ubiquitous in nature. They can cause a wide variety of diseases to both fruit and vegetables in the field and at postharvest stage (Scott 2001; Thomma 2003). The phylogeny of *Alternaria* spp. is particularly complex and hampered by the highly similar morphology of closely related species and inability to resolve monophyletic trees with simple barcodes (Simmons 2007; Woudenberg et al. 2013, 2014, 2015). *Alternaria atra* (Preuss) Woudenb. & Crous. (previously known as *Ulocladium atrum*) (Woudenberg et al. 2013) is one of those globally occurring *Alternaria* spp. It causes Ulocladium blight on potato in large parts of the world (Esfahani 2018) yet it has been reported to have biocontrol potential as a saprophyte on different crops against *Botrytis cinerea* and *Sclerotinia sclerotiorum* (Boff et al. 2001; Elead et al. 1994; Li et al. 2003; Ronseaux et al. 2013). Moreover, the species can be found on many wild plant species with various degrees of symptoms. Genomic resources for the *Alternaria* genus are limited, and no good reference exists for *A. atra.* Here, we present a scaffold-level reference assembly and gene annotation for *A. atra* isolate CS162. These data

†Corresponding author: R. Stam; stam@wzw.tum.de

*The *e*-Xtra logo stands for "electronic extra" and indicates that supplementary materials are published online.

The author(s) declare no conflict of interest.

Accepted for publication 15 March 2021.

Table 1. Genome statistics of *Alternaria atra* (CS162)

Genomic features[a]	*A. atra* (CS162)
Genome sequencing	
Sequencing platform	Nanopore MinION
Number of reads	2,105,684
Size (gb)	4.06
Coverage (×)	104
Assembly statistics	
Total assembly size (Mbp)	39.62
Number of contigs	87
Largest contig	6,982,780
Number of scaffolds	43
GC content	50.87
N_{50} (bp)	3,893,166
L_{50}	4
N's per 100 kbp	854.45
BUSCO	
BUSCO completeness, genome (%)	98.8
BUSCO completeness, CDS (%)	98.8
Repeat annotation results (%)	
Retroelements	2.51
LINEs	0.18
LTR elements	2.33
DNA transposons	3.23
Unclassified	2.29
Total interspersed repeats	8.02
Simple repeats	0.34
Low complexity	0.05
Genome annotation	
Genes	12,173
Transcripts	12,228
Protein	12,228
Ribosomal RNA	15
Functional annotation	
Effector proteins	1,148
Signal-peptide containing protein	1,330
Secreted protein	1,479
CAZyme families	598
Destruxin	23
Transcription	1,872
Transmembrane	1,193
PfamScan annotation	
Total amount of input sequences	12,228
Average length	499
GO annotated sequences (*n*)	7,019
GO annotated sequences (%)	57.40%
Number of GO annotations	25,785
Average GOs per sequence	3.67
Pfam accession distribution	
Family	16,345
Domain	11,895
Repeat	1,762
Coiled-coil	1,298
Motif	163
Disordered	59

[a] BUSCO = benchmarking universal single-copy ortholog, CDS = coding sequences, and GO = gene ontology.

will help future studies on the phylogeny, genetic diversity, and biology of this intriguing fungal genus.

A. atra isolate CS162 was collected from the wild tomato species *Solanum chilense* in northern Chile near a canyon riverbed, close to the Bolivian border. We confirmed the identity of the *A. atra* isolate through the analysis of the sequences of multiple conserved genes *Alt1a* and *RPB2* (Woudenberg et al. 2015) using BLASTn. The isolate was purified and grown on potato dextrose agar medium, and whole genomic DNA was isolated from mycelia of 7-day

old cultures using phenol/chloroform based extraction. Purified high molecular weight genomic DNA (2 μg) was sent for Oxford Nanopore custom sequencing. The sequencing was then run on an MinION R9 flow cell; the run produced 2,105,684 reads amounting to 4.06 GB and corresponding to a coverage of 104×. The de novo assembly using wtdbg2 (Ruan and Li 2020) generated a total of 87 contigs, with the largest contig containing 6,982,780 bases.

BRAKER v2.1.5 (Brůna et al. 2020), a combination of the GeneMark-ET (Lomsadze et al. 2014) and AUGUSTUS (Stanke et al. 2006, 2008) annotation tools, was used for gene prediction. AUGUSTUS used 51.86 million high-quality paired end RNA-seq reads as extrinsic evidence into the gene prediction and identified 12,173 genes encoding 12,228 proteins, with a BUSCO score of 98.8%. Contigs were scaffolded into 43 scaffolds using SLR (Luo et al. 2019), and polished with 4.2 million trimmed PE Illumina reads obtained from the same genomic DNA, using Pilon (Walker et al. 2014). The final *A. atra* genome assembly had a size of 39.62 Mb. The quality of the assembled data was assessed using QUAST (Gurevich et al. 2013) (Table 1). The assembled sequence had an N_{50} of 3.89 Mb and L_{50} was 4, with an average GC content of 50.87%. The genome comprises of 8.02% transposable elements, as identified using the RepeatModeler pipeline (Flynn et al. 2020). The quality and completeness of the assembled genome was estimated using benchmarking universal single-copy ortholog (BUSCO) software v5.0.0 (Seppey et al. 2019). The completeness of this genome is 98.8% (Table 1). Out of 749 total BUSCO groups searched, the assembly contained 705 complete single-copy, 44 complete duplicated, and 9 missing orthologs, and no fragmented BUSCO was observed. Similar relatively high numbers of duplicated BUSCO groups were also reported in other *Alternaria* genomes (Bihon et al. 2016; Feng et al. 2021).

Gene ontology terms were associated with 1,872 transcriptional genes and 1,193 transmembrane proteins (Supplementary File S1). Protein domains were searched for in the Pfam database (Mistry et al. 2007), and 241 predicted proteins containing domains for fungal-specific transcription genes and 41 Laminin proteins used for adhesion of fungal conidia to host were also identified. The *A. atra* proteome contained 1,479 putative secreted proteins, 1,330 signal-peptide-containing proteins, and 1,148 potential effector proteins (Supplementary File S1), as predicted by TargetP-2.0 (Emanuelsson et al. 2007), SignalP-5.0 (Petersen et al. 2011), and EffectorP-2.0 (Sperschneider et al. 2016), respectively. Furthermore, 128 unique CAZy families were identified using conserved unique peptide patterns (CUPP). Barrett and Lange (2019) assigned 635 CUPP groups with 180 unique EC numbers to specify enzyme-catalyzed property. Using a blast (HSP) high-scoring segment pair, 23 gene clusters responsible for the production of the pathogenicity factor Destruxin-B, an important secondary metabolite produced by pathogenic *Alternaria* spp. (Rajarammohan et al. 2019), were also identified (Supplementary File S1). PFAM domain output showed that 17 TBC-associated domains and 100 fungus-specific HET domains were present. Proteins with these domains are used by fungi to inactivate specific membrane-trafficking processes and, ultimately, lead to the death of the host's cells (Gabernet-Castello et al. 2013; Paoletti and Clavé 2007). Six toxin gene families—namely, HicA toxin, HigB-like toxin, ParE-like toxin, Toxin YhaV, YafO toxin, and YdaT toxin—were identified by PFAM. The domains of some other toxins—namely, CDtoxinA, Chi-conotoxin, Conotoxin, Endotoxin_N, Fst_toxin, MazE_antitoxin, ParD_antitoxin, VapB_antitoxin, and Zeta_toxin—were also present in the *A. atra* proteome (Supplementary File S1). Identification of these toxins will facilitate genome comparisons within the species and enhance our understanding of the principle behind molecular mechanisms underlying the pathogenicity and host specificity of this fungal pathogen.

We generated the first scaffold-level genome of the ubiquitous plant-associated fungus *A. atra* using Oxford Nanopore long-read and Illumina short-read data, with RNA-sequencing-driven gene prediction and functional annotation of factors with high relevance for pathogenicity. This draft genome report will provide useful information for phylogenetic studies and functional genome comparisons among the most important plant pathogens, endophytes, and saprophytes belonging to the *Alternaria* genus.

Data Availability

The sequencing data sets produced for this study are deposited at the EBI European Nucleotide Archive (ENA) under the project reference PRJEB42493. Fasta files for the genome, coding sequences, and protein sequences and the GTF files, as well as all result files from the functional annotation, are also available at Zenodo.

Acknowledgments

We thank R. Salvatierra Martínez for support during pathogen sampling in Chile, F. Diaz and other members of the Rodriguez Guittierez lab (Ponteficia Universidad Catolica, Santiago) for helping us with essential sampling preparations in their labs, R. Dittebrandt for help with help for fungal propagation for DNA and RNA extractions, and T. Schmey for help with RNA extraction. Bioinformatics analyses were performed on the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, and 031A532B).

Author-Recommended Internet Resource

Zenodo: https://zenodo.org/record/4436555

Literature Cited

Barrett, K., and Lange, L. 2019. Peptide-based functional annotation of carbohydrate-active enzymes by conserved unique peptide patterns (CUPP). Biotechnol. Biofuels 12:102-123.

Bihon, W., Cloete, M., Gerrano, A. S., Oelofse, D., and Adebola, P. 2016. Draft genome sequence of *Alternaria alternata* isolated from onion leaves in South Africa. Genome Announce. 4:e01022-16.

Boff, P., Kraker, J. D., Van Bruggen, A. H. C., Gerlagh, M., and Köhl, J. 2001. Conidial persistence and competitive ability of the antagonist *Ulocladium atrum* on strawberry leaves. Biocontrol Sci. Technol. 11:623-636.

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., and Borodovsky, M. 2020. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. bioRxiv.

Elead, Y., Köhl, J., and Fokkema, N. J. 1994. Control of infection and sporulation of *Botrytis cinerea* on bean and tomato by saprophytic bacteria and fungi. Eur. J. Plant Pathol. 100:315-336.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2:953-971.

Esfahani, M. N. 2018. Identification of *Ulocladium atrum* causing potato leaf blight in Iran. Phytopathol. Mediterr. 57:112-114.

Feng, F., Hsiang, T., Liang, X., Zhang, R., and Sun, G. 2021. Draft genome sequence of cumin blight pathogen *Alternaria burnsii*. Plant Dis. 105:1165-1167.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., and Smit, A. F. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U.S.A. 117:9451-9457.

Gabernet-Castello, C., O'Reilly, A. J., Dacks, J. B., and Field, M. C. 2013. Evolution of Tre-2/Bub2/Cdc16 (TBC) Rab GTPase-activating proteins. Mol. Biol. Cell 24:1574-1583.

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

Li, G. Q., Huang, H. C., and Acharya, S. N. 2003. Antagonism and biocontrol potential of *Ulocladium atrum* on *Sclerotinia sclerotiorum.* Biol. Control 28:11-18.

Lomsadze, A., Burns, P. D., and Borodovsky, M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 42:e119.

Luo, J., Lyu, M., Chen, R., Zhang, X., Luo, H., and Yan, C. 2019. SLR: A scaffolding algorithm based on long reads and contig classification. BMC Bioinf. 20:539.

Mistry, J., Bateman, A., and Finn, R. D. 2007. Predicting active site residue annotations in the Pfam database. BMC Bioinf. 8:298.

Paoletti, M., and Clavé, C. 2007. The fungus-specific HET domain mediates programmed cell death in *Podospora anserina.* Eukaryot. Cell 6:2001-2008.

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.

Rajarammohan, S., Paritosh, K., Pental, D., and Kaur, J. 2019. Comparative genomics of *Alternaria* species provides insights into the pathogenic lifestyle of *Alternaria brassicae*—A pathogen of the *Brassicaceae* family. BMC Genomics 20:1036.

Ruan, J., and Li, H. 2020. Fast and accurate long-read assembly with wtdbg2. Nat. Methods 17:155-158.

Scott, P. M. 2001. Analysis of agricultural commodities and foods for *Alternaria* mycotoxins. J. AOAC Int. 84:1809-1817.

Ronseaux, S., Clément, C., and Barka, E. A. 2013. Interaction of *Ulocladium atrum*, a potential biological control agent, with *Botrytis cinerea* and grapevine plantlets. Agronomy (Basel) 3:632-647.

Seppey, M., Manni, M., and Zdobnov, E. M. 2019. BUSCO: Assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962:227-245.

Simmons, E. G. 2007. *Alternaria.* An identification Manual. CBS Biodiversity Series 6. CBS Fungal Biodiversity Centre, Utrecht, The Netherlands.

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.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24:637-644.

Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinf. 7:62.

Thomma, B. P. H. J. 2003. *Alternaria* spp.: From general saprophyte to specific parasite. Mol. Plant Pathol. 4:225-236.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.

Woudenberg, J. H. C., Groenewald, J. Z., Binder, M., and Crous, P. W. 2013. *Alternaria* redefined. Stud. Mycol. 75:171-212.

Woudenberg, J. H. C., Seidl, M. F., Groenewald, J. Z., de Vries, M., Stielow, J. B., Thomma, B. P., and Crous, P. W. 2015. *Alternaria* section *Alternaria*: Species, formae speciales or pathotypes? Stud. Mycol. 82:1-21.

Woudenberg, J. H. C., Truter, M., Groenewald, J. Z., and Crous, P. W. 2014. Large-spored *Alternaria* pathogens in section Porri disentangled. Stud. Mycol. 79:1-47.