BMC Genomic Data

DATA NOTE Open Access

First fine mapping of a strain of *Rhizoctonia Solani* AG-3, causing tobacco target spot

Pan Ma[1], Mengjuan Qiu[1], Rubing Xu[2], Zhao Wang[1], Tom Hsiang[3], Junbin Huang[1], Lu Zheng[1] and Yanyan Li[2*]

Abstract

Objectives *Rhizoctonia solani* AG-3 is the casual pathogen of tobacco target spot, a serious fungal disease of tobacco that severely decreases yield and quality. To examine the pathogenic mechanisms of this fungus, it is crucial to understand its genetics. The objective of this work was to generate the first fine mapping of a *R. solani* AG-3 strain from tobacco and to explore potential virulence genes, which will lay the foundation for genetic characterization and its interaction with tobacco. The functional genes involved in this study can be used as the candidates for follow-up experimental analyses.

Data description *Rhizoctonia solani* AG-3 strain XEMS25-1 was isolated from disease leaves of tobacco target spot in Enshi, Hubei Province, China. The DNA was sequenced using Pacific Biosciences Sequel II (PacBio) and Illumina NovaSeq PE150 (Nova). Data from both sequencing platforms were combined, and the de novo assembly yielded an estimated 39.4 Mb genome. Completeness of the genome examined using Benchmarking Universal SingleCopy Orthologs (BUSCO) showed that the assembly had 93.7% of the 758 genes in fungi_odb10. PHI (Pathogen Host Interactions) database analysis revealed 519 reduced virulence genes, 91 loss of pathogenicity genes, 28 hypervirulence genes and 18 effectors might be the pathogenicity-related genes in *R. solani* AG-3 strain XEMS25-1. These genes could be selected as the RNA-silencing targets for exploring the molecular mechanisms of *R. solani* AG-3 pathogenicity on tobacco.

Keywords PacBio, Illumina, Genome, Tobacco target spot, *Rhizoctonia Solani* AG-3

Objectives

Tobacco is an important economic crop cultivated worldwide. In recent years, tobacco target spot disease, caused by the fungal pathogen *Rhizoctonia solani* AG-3, has emerged as a severe foliar disease [1, 2], widely affecting tobacco-growing regions in China [3–6]. The pathogen primarily overwinters in the soil and plant debris in the form of sclerotia or mycelia, serving as the main source of infection [7]. However, the molecular pathogenesis of *R. solani* AG-3 on tobacco is not thoroughly understood, nor the genomics of the pathogen.

A *R. solani* AG-3 strain, XEMS 25 – 1, was isolated from tobacco leaves in Enshi, Hubei Province, China in 2022. This strain was identified as the causal agent of tobacco target spot and verified as *R. solani* AG-3 fusion group [5]. To date, the draft genome of *R. solani* AG-3 from potato has been sequenced based on second generation sequencing [8], but an isolate from tobacco has not been reported and fine mapping of *R. solani* AG-3 strain also has not been generated. In this study, whole

*Correspondence:
Yanyan Li
yanyanli0025@126.com
[1]Hubei Key Laboratory of Plant Pathology, Huazhong Agricultural University, Wuhan 430070, China
[2]Tobacco Research Institute of Hubei Province, Wuhan 430030, China
[3]School of Environmental Sciences, University of Guelph, Guelph N1G 2W1, Canada

Table 1 Overview of all data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	all_sample.Cleandata.stat	Excel file(.xls)	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 2	all_kmer.stat	Excel file(.xls)	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 3	Cleandata.stat	Excel file(.xls)	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 4	Assembly.final.stat	Excel file(.xls)	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 5	BUSCO.stat	Excel file(.xls)	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 6	all_sample.gene.stat	Excel file(.xls)	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 7	all_sample.trf.stat	Excel file(.xls)	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 8	all_sample.annoSummary.stat	Excel file(.xls)	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 9	R.S.S.Secretory-Protein	.gff	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 10	R.S.S.Secretory-Protein	.pep	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data file 11	R.S.S.phi.anno	Excel file(.xls)	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Figure 1	R.S.S.Overview	.jpg	Figshare: https://doi.org/10.6084/m9.figshare.27960000.v4 [10]
Data set 1	SRP549198	.fasta; .gff	NCBI: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1193463/ [9]
Data set 2	Bioproject PRJCA034663	.fasta	CNCB: https://ngdc.cncb.ac.cn/gwh/Assembly/88088/show [16]

genome sequencing using hybrid second and third generation sequencing techniques on *R. solani* AG-3 was performed, and the sequences assembled. This study provides a reference for detailed genome-wide mapping and genome annotation of *R. solani* AG-3 for researchers. It also provides a theoretical basis for screening the pathogenic genes and related pathways of *R. solani* AG-3, and clarifying the pathogenic mechanisms of tobacco target spot disease.

Data description

The tobacco target spot strain, XEMS 25 – 1, collected from Enshi, China (29°07′10″N, 108°23′12″E) was used for genomic DNA extraction. Isolation, pathogenicity test and identification of the strain XEMS 25 – 1 were performed based on previous work [5]. Hyphae were transferred onto potato dextrose agar and cultured at 25 °C for 72 h. Hyphae were collected, and DNA was extracted using NEBNext®Ultra™ DNA Library Prep Kit-for Illumina (NEB, USA) kit. A sample of 2 μg/ml DNA was sequenced specifying ~ 350 bp fragments on an Illumina PE150, to obtain 6,049 Mb of raw data (Data file 1, Data set 1, Table 1) [9, 10]. After filtering the raw data, there were 5,609 Mb of clean data. An analysis based on K-mer statistics was used to estimate genome size of 48.14 Mb before genome assembly (Data file 2, Data set 1, Table 1) [9, 10]. Illumina data were used for initial assessment and correction of the genome. The genomic DNA of XEMS 25 – 1 was then extracted using an in-house method called GT1 [11], after which the purity and integrity of the DNA was measured by agarose gel electrophoresis and quantified using Qubit. A library was constructed using the SMRT bell TM Template kit (version 2.0) [12], and sequenced using the PacBio platform. This generated 12.05 Gb of data with 738,951 subreads, and average subread length of 16,301 bp (N50 = 18,183 bp, N90 = 13,636 bp) (Data file 3, Data set 1, Table 1) [9, 10]. Based on the clean data of the third-generation sequencing data after quality control of each sample, the reads were assembled using Falcon software (https://github.com/PacificBiosciences/FALCON/) [13], and then Racon (version: 1.4.13) [14] software for three rounds of error correction based on the third-generation sequencing data, and then three rounds of Pilon (version: 1.22) [15] error correction with second-generation read data to obtain the final assembly results. The cumulative assembly resulted in 26 contigs, with an N50 of 2,548,437 bp (Data file 4, Data set 2, Table 1) [10, 16]. The completedness of the assembled genome was assessed using BUSCO software [14], and the genome was found to contain 710 of the 758 BUSCO genes (93.7%), with 705 complete single copies (93%), 5 duplicates (0.7%), and 5 fragmented (0.7%), with 42 BUSCO genes missing (5.6%) (Data file 5, Table 1) [10, 17]. The XEMS 25 – 1 genome was then annotated and genome-wide mapped against *R. solani* AG-3 Rhs1AP as reference [8]. The predicted number of genes was 10,317 using De-novo Augustus prediction (Data file 6, Data set 1, Table 1) [9, 10, 18]. There were 3,367 tandem repeats in the assembly (Data file 7, Table 1) [10, 19]. SignalP (Version 4.1) [20] and TMHMM (Version 2.0c) were used for detection of signal peptides and transmembrane structures, respectively, identifying 741 secreted proteins (Data file 8, 9, 10, Table 1) [10]. Genome basic annotation was performed by alignment and functional annotation against five major databases (NR [21], Swiss-Prot [22], Pfam, GO, and KEGG) [23]. In the gene function analysis, 10,076 genes were annotated by comparison with the NR database; 2,485 with the Swiss-Prot database; 6,049 with the PFAM database; 6,049 by GO database annotation; and 6,005 by alignment annotation with the KEGG database (Data file 8, Table 1) [10]. The assembled genome sequence

for *Rhizoctonia solani* AG-3 strain XEMS 25 – 1, combined with the prediction results of coding genes, were depicted using a visualization tool Circos software (Table 1) [10, 24]. PHI database was then used to predict potential pathogenicity genes and a total of 983 PHI-related genes were identified in the *R. solani* XEMS25-1 strain, including 519 reduced virulence genes, 91 loss of pathogenicity genes, 28 hypervirulence genes, 18 effectors and others (Data file 11, Table 1) [10]. These candidate genes might be the RNA-silencing targets using Host-induced Gene Silencing technology for exploring the molecular mechanisms of *R. solani* AG-3 pathogenicity on tobacco. The XEMS25-1 genome was used to compare with the potato Phs1AP genome, and the results revealed that the XEMS25-1 genome contains 5,306 core genes and 3,330 specific genes, indicating genome variation might be exist between *R. solani* AG-3 strains from different host. The clean data of the XEMS25-1 genome were deposited in NCBI with Bioproject PRJNA1193463 and Biosample SAMN45140045 (Data set 1, Table 1) [9]. The assembly data were deposited in CNCB with Bioproject PRJCA034663 and Biosample SAMC4533655 [16]. Other results files are available on the figshare website https://doi.org/10.6084/m9.figshare.27960000.v4 [10].

Limitations

The data generated in this study is limited to a single genome sequence. We can not get more information of host specificity or adaptation of the pathogen. To overcome this limitation, other strains of tobacco target spot fungus *R. solani* AG-3 from different locations should be sequenced. Particularly, *R. solani* AG-3 strains with different pathogenic characteristics could be used for whole genome sequencing. Furthermore, transcriptome data of key infection stages of *R. solani* AG-3 on tobacco would be helpful to elucidate the pathogenic mechanism.

Abbreviations

Nova	Illumina NovaSeq PE150
PacBio	Pacific Biosciences Sequel II
NCBI	National center for biotechnology information
SMRT	Single molecule real time
NR	Non-redundant protein database
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
BUSCO	Benchmarking universal single-copy orthologs
PHI	Pathogen host interactions database
CIRCOS	Circular genome data visualization

Acknowledgements

The authors appreciate financial support from the Tobacco Research Institute of Hubei Province and thank Novogene company (Beijing, China) for providing high-quality sequencing.

Authors' contributions

PM: performed DNA extraction, analysis of sequencing data, and manuscript writing. MJQ and RBX: performed strain collection and preservation. ZW: assisted in assembling the genome and aiding in genome annotation. TH: proof-read and revised the manuscript. LZ and YYL: designed the project and revised manuscript. All authors have read and approved the final version of the manuscript.

Funding

This work was financially supported by the Science & Technology Project of Hubei Tobacco Company (027Y2022-020, 027Y2021-004), and Pests and Diseases Green Prevention and Control Major Special Project (110202101045 (LS-05)).

Data availability

The data described in this Data note can be freely and openly accessed on NCBI (Bioproject PRJNA1193463, Biosample SAMN45140045) (Data sets 1), CNCB (Bioproject PRJCA034663) and figshare website (https://doi.org/10.6084/m9.figshare.27960000).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 9 January 2025 / Accepted: 20 March 2025
Published online: 02 April 2025

References

1. Zhong J, Sui WW, Bai XY, Qiu ZL, Li XG, Zhu JZ. Characterization and biocontrol mechanism of *Streptomyces olivoreticuli* as a potential biocontrol agent against *Rhizoctonia Solani*. Pestic Biochem Phys. 2023;197:105681.
2. Wu YH, Zhao YQ, Fu Y, Zhao XX, Chen JG. First report of target spot of flue-cured tobacco caused by *Rhizoctonia Solani* AG-3 in China. Plant Dis. 2012;96(12):1824.
3. Wu YH, WANG ZB, LIU ZH, Zhao XX, Liang JY. Tobacco target spot disease-a new tobacco disease in China. Acta Taba Sin. 2006;12:22–51.
4. Xia B, Xu CT, Xu JK, Wu YH, Xie Q, Jiang S, Xie YB. First report of target leaf spot on flue-cured tobacco by *Rhizoctonia Solani* AG-3 in Sichuan, China. Plant Dis. 2019;103(3):581.
5. Qiu MJ, Li YY, Xu TT, Liu HF, Wu CF, Zheng L, Huang JB, Li XH. Identification and anastomosis group study of the tobacco target spot pathogen in Hubei Province. Chin Tob Sci. 2022;43(5):50–5.
6. Wang Y, Guo Y, Guo SP, Qi L, Li B, Jiang LQ, Xu CT, An MN, Wu YH. RNA interference-based exogenous double-stranded RNAs confer resistance to *Rhizoctonia Solani* AG-3 on *Nicotiana tabacum*. Pest Manag Sci. 2024;80(4):2170–8.
7. Zachow C, Grosch R, Berg G. Impact of biotic and a-biotic parameters on structure and function of microbial communities living on sclerotia of the soil-borne pathogenic fungus *Rhizoctonia Solani*. Appl Soil Ecol. 2011;48(2):193–200.
8. Cubeta MA, Thomas E, Dean RA, Jabaji S, Neate SM, Tavantzis S, Toda T, Vilgalys R, Bharathan N, Fedorova-Abrams N, Pakala SB, Pakala SM, Zafar N, Joardar V, Losada L, Nierman WC. Draft genome sequence of the plant-pathogenic soil fungus *Rhizoctonia Solani* anastomosis group 3 strain Rhs1AP. Genome Announc. 2014;2(5):10–1128.
9. National Center for Biotechnology Information. Genome Assembly Bioproject PRJNA1193463. 2024. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1193463/.
10. Figshare. Genome sequencing and assembly of *Rhizoctonia Solani* AG-3 strain XEMS25-1. 2024. https://doi.org/10.6084/m9.figshare.27960000.
11. Pahlich E, Gerlitz C. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry. 1980;19:11–3.
12. Reiner J, Pisani L, Qiao W, Singh R, Yang Y, Shi L, Khan WA, Sebra R, Cohen N, Babu A, Edelmann L, Jabs EW, Scott SA. Cytogenomic identification and long-read single molecule real-time (SMRT) sequencing of a Bardet-Biedl syndrome 9 (BBS9) deletion. NPJ Genom Med. 2018;3(1):3.
13. Chin CS, Peluso P, Sedlazeck F, Nattestad M, Concepcion G, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Figueroa-Balderas R, Morales-Cruz A, Cramer

GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods. 2016;13(12):1050–4.
14. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de Novo genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737–46.
15. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9(11):e112963.
16. China National Center for Bioinformation. Genome Assembly Bioproject PRJCA034663. 2025. https://ngdc.cncb.ac.cn/gwh/Assembly/88088/show.
17. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. CCF TCBI. 2015;31(19):3210–2.
18. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de Novo gene finding. CCF TCBI. 2008;24(5):637–44.
19. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
20. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785–6.
21. Li W, Jaroszewski L, Godzik A. Tolerating some redundancy significantly speeds up clustering of large protein databases. CCF TCBI. 2002;18(1):77–82.
22. Amos B, Rolf A. The SWISS-PROT protein sequence database and its supplement trembl in 2000. Nucleic Acids Res. 2000;28(1):45–8.
23. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for Deciphering the genome. Nucleic Acids Res. 2004;32(suppl1):D277–80.
24. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. Sep; 2009;19(9):1639–45.

Publisher's Note