BMC Genomic Data

DATA NOTE Open Access

A highly contiguous genome sequence of *Alternaria porri* isolate *Apn*-Nashik causing purple blotch disease in onion

Richa Sharma[1], Rukmini Mishra[2] and Raj Kumar Joshi[1*]

Abstract

Objectives Purple blotch, caused by the necrotrophic pathogen *Alternaria porri*, is one of the most economically significant diseases of onion and allied crops. While the virulent nature of many *Alternaria* spp. has been identified, the pathogenic repertoire of *A. porri* is still unknown. The objective of this work was to sequence the genome of *A. porri* using the PacBio SMRT sequencing strategy and analyse the repertoire of CAZymes, secondary metabolites, secretome and effectors in *A. porri*. Our research group is working to identify onion germplasm with purple blotch resistance and to understand the genetics of the pathogen. The reported *de-novo* assembly will contribute to the analysis of potential variants and the gene repertoire contributing to the virulence and pathogenicity of the purple blotch pathogen.

Data description Long-read sequencing on a PacBio Sequel II system resulted in a 32.98 Mb (20 contigs) assembly with an N50 of 2, 657, 264 bp, the longest contig length of 5.05 Mb, and a GC content of 51.06%. The Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis resulted in 99.7% genome completeness at the Dothideomycetes lineage, representing a high-quality genome assembly. AUGUSTUS ab initio analysis resulted in 9875 protein-coding genes. Of the 6776 pathogenicity-related genes, 537 genes with effector functions were identified. Likewise, the glycoside hydrolases (434) were the most dominant group of the total 837 predicted CAZymes. The assembled genome of *A. porri* showed distinctive similarities to the genomes of *A. alternata* and *A. brassicicola*, the causal agents of leaf blight of onion and leaf spot of *Brassica* crops, respectively.

Keywords *Alternaria porri*, Onion purple blotch, PacBio Sequel II, Pathogen

*Correspondence:
Raj Kumar Joshi
rkjoshi@rdwu.ac.in; rajkumar.joshi@yahoo.co.in
[1]Department of Biotechnology, Rama Devi Women's University, Vidya Vihar, Bhubaneswar, Odisha 751022, India
[2]School of Applied Sciences, Centurion University of Technology and Management, Bhubaneswar, Odisha, India

Objective

Purple blotch (PB), caused by *A. porri* (Ellis) Ciferri, is one of the most destructive diseases of onion and allied *Allium spp.* [1]. The disease often affects the leaves and bulbs of the plants leading to heavy yield losses ranging from 76 to 97% [2]. Under favourable conditions, the small sunken lesions enlarge and produce brownish-purple necrotic lesions in the leaves, and subsequently wilt, delaying the creation and maturation of the bulb. Purple Blotch disease is especially severe in onion grown in high humidity and moderate temperature of 80–90% and 25–30 °C, respectively. Management of PB through the application of fungicides is only partially effective besides accruing environmentally hazardous residues in the crop and ecosystem [2]. Therefore, the identification of resistant cultivars through host resistance breeding is the most efficient method to control PB disease. However, wide-range screening for resistance to *A. porri* in cultivated onion germplasms has not resulted in any potential source of resistance [3].

The molecular determinants responsible for the pathogenicity of *A. porri* are fairly unknown. Similar to other pathogenic fungi, *Alternaria* spp secretes a broad range of host-specific (HSTs) and non-host-specific (nHSTs) phytotoxins that act as significant virulence factors during pathogenesis [4]. These are mostly the low molecular weight secondary metabolites that causes leaf necrosis of susceptible cultivars. The HSTs have been reported in at least 12 *A. alternata* pathotypes [5]. Likewise, nHSTs like tenuazonic acid, brefeldin A, zinniol and tentoxin have been reported from many other *Alternaria* species [4]. *Alternaria porri* has been reported to produce secondary metabolites such as altersolanol A, alterporriol, macrosporin, erythroglaucin and tentoxin [6]. However, their role in the pathogenicity of PB fungus is yet to be established.

The advent of genome sequencing and comparative analysis has largely contributed to the identification of species-specific pathogenicity factors in closely related species [7]. Currently 169 *Alternaria* genomes of 31 species and nine unclassified species including *A porri* is available in the GenBank, that has contributed immensely to the taxonomy and molecular physiology of the genus [4]. However, the species-specific pathogenicity factors that could confer the ability to infect onion and allied crops are yet to be carried out. The currently available genomic information for *A. porri* has been generated through short-read sequencing and thus is highly fragmented and laconically inconclusive [4]. The third generation long-read sequencing approaches, including the single molecule real time (SMRT) sequencing and Oxford nanopore sequencing, has facilitated the generation of high quality and complete genome assemblies with greater accuracy [8]. Therefore, the objective of this work was to sequence the genome of *A. porri* using the PacBio SMRT sequencing strategy and analyse the repertoire of pathogenic effectors in *A. porri.*

Data description

The detail description of the methodology is represented in data file 1. The *Alternaria porri* super virulent isolate *Apn*-Nashik was obtained as a hyphal culture (Table 1, Data file 2) from infected leaves of onion variety Agrifound Rose (Table 1, Data file 3) through the tissue segment approach was used for whole genome sequencing. The purified genomic DNA isolated from the fungal mycelium (diploid) was used for generating long read sequencing libraries through the PacBio single molecule real time (SMRT) sequencing technology. The sequencing resulted in 199 GB of data comprising 2,115,917 polymerase raw reads with 113x coverage. A total of ~30 Gb filtered polymerase read of unique single molecules were assembled through Improved Phased Assembler (IPA) that produced 33.5 Mb primary (33 contigs) assembly. After the elimination of low-quality reads and adapter sequences using the SMRTlink tool, an assembly of 20 contigs was generated. The final assembly consisted of 32,989,564 bp with N50 of 26,57,264 bp, longest contig length of 5.05 Mb, and GC content of 51.06% (Table 1, Data file 4). These data were found well within the predicted genome size but more than those reported by Dang et al. [4]. A total of 10,699,75 circular consensus sequence (CCS) reads were aligned against the nearest *Alternaria* genome (Accession ID: GCA_014751505.1_ASM1475150v1) using the minimap2 program and resulted in 100% mapped reads. The BUSCO analysis resulted in 99.7% genome completeness (including 99.5% complete+0.2% duplicated) at the Dothideomycetes lineage (Table 1, Data file 5), representing a high-quality genome assembly. A total of 3786 BUSCO genes were searched against the Dothideomycetes lineage, from which 3768 were categorized as complete and single copy, six genes as complete and duplicated, and one fragmented, while 13 genes were classified as missing.

Estimation of repeat content using the RepeatMasker revealed 2784 repetitive elements consisting of retroelement (209), LTRs (186), DNA transposons (52), Gypsy/DIRS1 (140), Ty1/Copia (44) and simple repeats (3065) (Table 1, Data file 6). Further, the simple repeats included 1536-di, 932- tri, 618-tetra, 249-penta and 103- hexanucleotide repeats. While a majority of filamentous plant pathogens have higher repeat content, *A. porri* genome accounted for only ~1.5% repeats, which is significantly lower compared to all the other *Alternaria* species sequenced to date [4, 9]. In addition, tRNAscan identified 289 tRNAs in the *A. porri* genome. This suggests that *A. porri* is a relatively new species in the evolutionary development of the order Pleosporales [9]. Also, the lower

Table 1 Overview of data files/data sets

Label	Name of the data files/data sets	File types	Data repository and identifiers
Data file 1	Detailed methodology	Microsoft word (.doc)	https://doi.org/10.6084/m9.figshare.26207534 [16]
Data file 2	Characterization of *A. porri* isolate *Apn* Nashik	Microsoft word (.doc)	https://doi.org/10.6084/m9.figshare.26212271 [17]
Data file 3	Onion plant infected with *A. porri Apn*-Nashik	Microsoft word (.doc)	https://doi.org/10.6084/m9.figshare.26213117 [18]
Data file 4	Read and Assembly statistics of *A. porri Apn*-Nashik	Microsoft word (.doc)	https://doi.org/10.6084/m9.figshare.26213180 [19]
Data file 5	BUSCO dothideomycetes completeness statistics	Microsoft word (.doc)	https://doi.org/10.6084/m9.figshare.26213216 [20]
Data file 6	Statistics on Genome Repeat masking	Microsoft word (.doc)	https://doi.org/10.6084/m9.figshare.26213237 [21]
Data set 1	Predicted genes of *A. porri* using AUGUSTUS	General feature format (.gff)	https://doi.org/10.6084/m9.figshare.24906318.v1 [22]
Data file 7	Summary of protein coding genes	Microsoft word (.doc)	https://doi.org/10.6084/m9.figshare.26213297 [23]
Data file 8	Functional annotation from GO and KEGG databases	Spreadsheet (.xlsx)	https://doi.org/10.6084/m9.figshare.26213552 [24]
Data file 9	Genes encoding pathogenicity related factors derived from PHI database	Spreadsheet (.xlsx)	https://doi.org/10.6084/m9.figshare.26213564 [25]
Data set 2	Annotated genes of *A. porri* using dbCAN2 program	Text format (.text)	https://doi.org/10.6084/m9.figshare.24906378.v1 [26]
Data file 10	Genes encoding Cazymes	Spreadsheet (.xlsx)	https://doi.org/10.6084/m9.figshare.26213570 [27]
Data file 11	Genome comparison of *A. porri* Apn Nashik with other *Alternaria* species	Microsoft word (.doc)	https://doi.org/10.6084/m9.figshare.26213636 [28]
Data file 12	Clusters of orthologous genes shared and unique among the *Alternaria spp*	Microsoft word (.doc)	https://doi.org/10.6084/m9.figshare.26213702 [29]
Data set 3	Sequencing reads of *A. porri* genomic DNA	Fastq file (.fq.gz)	NCBI SRA ID: SRX26191501 https://identifiers.org/ncbi/insdc.sra:SRX26191501[30]
Data set 4	Genomic assembly of *A. porri*	Fasta file	NCBI Genbank database: JAPZQB000000000.1 https://identifiers.org/nucleotide:JAPZQB000000000.1 [31]

repeat content could be attributed to repeat induced point mutations (RIP) which prevent the development of highly repeated sequences [10, 11].

AUGUSTUS ab initio analysis resulted in the identification of 9875 protein-coding genes (Table 1, Data set 1). Among them, 6031 (61%) genes and 3438 (34.8%) genes exhibited homologies with known functions in the GO and KEGG databases, respectively (Table 1, Data file 7). The short-read sequence assembly of *A. porri* BMP0178 as reported by Dang et al. [4] could assign GO terms and KEGG IDs to only 5238 and 3176 genes, respectively. The categorization of functional annotation in the predicted genes showed that 4419 genes (73.3%) were assigned with molecular functions, followed by 3915 (64.9%) genes with cellular components and 3970 (65.8%) genes with biological processes (Table 1, Data file 8). Catalytic (25.3%) and binding (21.6%) activities dominated the molecular function category, whereas, under the category of cellular component, cells (17.3%) and cell components (14.4%) formed the major groups. Likewise, KEGG enrichment showed that the majority of enzymes belong to hydrolases (1148), followed by transferases (1064) and oxidoreductases (588) (Table 1, Data file 7). Previous studies have shown that hydrolases are a major factor of pathogenicity as they help the fungus to break down plant cell walls, acquire nutrients from plant tissues, and evade plant immune responses [11, 12].

The pathogen-host interactions (PHI)-database identified a total of 6776 (74.03%) pathogenicity-related genes in *A. porri* genome, which could be crucial for the virulence of the purple blotch pathogen (Table 1, Data file 9). Among them, 3605 genes were associated with reduced virulence, 554 with loss of pathogenicity, 537 genes were identified with effector functions (plant avirulence determinant), 479 genes related to increased virulence, 157 related to lethal factors, 17 genes associated with chemical resistance, and nine genes related to chemical sensitivity. A majority of them belong to secreted proteases, CAZymes and transporters that play an important role in the interaction of the fungal pathogen with its host [13]. The enrichment of CAZymes using three algorithms, viz., Diamond, HMMER, and Hotpep in the dbCAN2 Meta server resulted in the identification of 837 genes (Table 1, Data set 2) in different CAZymes families including 434 glycoside hydrolases (GHs), 157 enzymes with auxiliary activities (AAs), 145 glycosyltransferases (GTs), 49 carbohydrate esterases (CEs), 25 polysaccharide lyases (PLs), and 27 carbohydrate-binding modules (CBMs) (Table 1, Data file 10). The higher number of CAZymes suggests that many of them might be involved in the degradation of plant cell wall components such as cellulose and hemicellulose, while at the same time utilizing the plant polysaccharides during colonization of the host plant [14]. Interestingly, the auxiliary activity enzyme 9 (AA9) was significantly represented (30 genes) in the *A. porri* genome. AA9, formerly known as GH61 (copper-dependent lytic polysaccharide monooxygenases) has been reported as over-represented in multiple *Alternaria*

spp [9]. This family of CAZymes exhibited dual role as a degrader of cell-wall polysaccharides and contributed to host necrosis through the generation of reactive oxygen species (ROS) [15].

Alignment of *A. porri* genome with the genomic sequences of *A. alternata*, *A. brassicae* and *A. brassicicola* showed distinctive inversions (flipped genomic segments), duplication (similar copy segments) and translocation (DNA fragments moving to different locations) events across the four genomes (Table 1, Data file 11). The presence of continuous coloured blocks showed that conserved sequences are shared among the compared genomes. *A. porri* was found evolutionarily closer to *A. alternata* followed by *A. brassicicola.* The gaps or mismatches within the aligned regions represented the insertion, deletions and substitutions among the closely related species. While, there was no significant gene insertion or deletion of large regions in *A. porri* genome, the location of homologous genes was different. Orthology analysis reported 8914 orthologous genes among the four *Alternaria* spp. (Table 1, Data file 12). *A. porri* displayed 207 unique genes, while it shared 604 genes with *A. alternata*, 3558 genes with *A. brassicicola*, and only nine genes with *A brassicae.*

Limitations

Using the latest genome sequencing platform, we have developed the nearly complete and contiguous genome assembly of *A. porri* in the present study. A greater variability has been represented in the PB isolates from different onion-growing regions of the world. As such, the assembled data from this study might not be fully accurate. If and when more funds are available, multiple isolates of PB could be sequenced using long-read sequencing technologies with greater coverage to increase the completeness and accuracy of the PB genome.

Abbreviations

BUSCO	Benchmarking Universal SingleCopy Orthologs
CCS	Circular Consensus Sequence
IPA	Improved Phased Assembler
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
NCBI	National Center for Biotechnology Information
PHI	Pathogen-Host Interactions
SMRT	Single Molecule Real Time

Acknowledgements

The work is supported by Dept. of Biotechnology, Rama Devi Women's University, Bhubaneswar. The facilities utilized developed under the DBT, Govt. of India project (BT/PR23412/BPA/118/284/2017) and DST-CURIE grant (DST/CURIE/01/2019) is duly acknowledged. RS is thankful to University Grants Commission (UGC), New Delhi for award of the SJSGC fellowship (grant no. UGCES-22-GE-ORI-F-SJSGC-8079).

Author contributions

RS and RKJ conceived and designed the research work. RS and RM conducted the research experiments including data curation, functional analysis and gene prediction and stress treatment. RS and RM wrote the manuscript and RKJ critically reviewed the manuscript. All the authors have read, revised and approved the final manuscript.

Funding

No funds have been received from any agencies to carry out this study.

Data availability

This whole genome shotgun project has been deposited at DDBJ/ENA/GenBankunder the accession id JAPZQB000000000. The version described in this paperis version JAPZQB010000000. The NCBI GenBank makes the sequencing reads used in this investigation publicly accessible under the BioProject accession number PRJNA916367 and the BioSample accession number SAMN32422955. The genome sequences are accessible in GenBank as contigs with the accession numbers JAPZQB01000001-JAPZQB010000020. Please see Table 1 for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 9 July 2024 / Accepted: 28 October 2024
Published online: 05 November 2024

References

1. Dar AA, Sharma S, Mahajan R, Mushtaq M, Salathia A, Ahamad S, Sharma JP. Overview of purple blotch disease and understanding its management through chemical, biological and genetic approaches. J Integr Agric. 2020;19(20):3013–24. https://doi.org/10.1016/S2095-.
2. Yadav RK, Singh A, Dhatt AS, Jain S. Field assessment of onion genotypes for resistance against purple blotch complex (*Alternaria Porri* and *Stemphylium vesicarium*) under artificial epiphytotic conditions in Indian Punjab. Int J Appl Sci Biotechnol. 2017;5:498–504.
3. Nanda S, Chand SK, Mandal P, Tripathi P, Joshi RK. Identification of novel source of resistance and differential response of Allium genotypes to purple blotch pathogen, Alternaria Porri (Ellis) Ciferri. Plant Pathol J. 2016;32:519–27. https://doi.org/10.5423/PPJ.OA.02.2016.0034.
4. Dang HX, Pryor B, Peever T, Lawrence CB. The *Alternaria* genomes database: a comprehensive resource for a fungal genus comprised of saprophytes, plant pathogens, and allergenic species. BMC Genom. 2015;16:239. https://doi.org/10.1186/s12864-015-1430-7.
5. Meena M, Gupta SK, Swapnil P, Andleeb Z, Dubey MK, Upadhyay RS. *Alternaria* toxins: potential virulence factors and genes related to pathogenesis. Front Microbiol. 2017;8:1451.
6. Anderson B, Dongo A, Pryor BM. Secondary metabolite profiling of *Alternaria dauci, A. porri, A. solani*, and *A. tomatophila*. Mycology Res. 2007; 112: 241–250. https://doi.org/10.1016/j.mycres.2007.09.004
7. Henry RJ. Progress in plant genome sequencing. Appl Biosci. 2022;1:113–28. https://doi.org/10.3390/applbiosci1020008.
8. Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A. Third-generation sequencing: the spearhead towards the radical transformation of modern genomics. Life. 2021;12(1):30. https://doi.org/10.3390/life12010030.
9. Rajarammohan S, Paritosh K, Pental D, Kaur J. Comparative genomics of *Alternaria* species provides insights into the pathogenic lifestyle of *Alternaria brassicae*–a pathogen of the Brassicaceae family. BMC Genom. 2019;20:1–13.
10. Hood ME, Katawczik M, Giraud T. Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum Violaceum*. Genetics. 2005;170:1081–9.
11. Chandra A, Singh D, Joshi D, Pathak AD, Singh RK, Kumar SA. Highly contiguous reference genome assembly for *Colletotrichum falcatum* pathotype *Cf*08 causing red rot disease in sugarcane. 3Biotech. 2021;11:148. https://doi.org/10.1007/s13205-021-02695-x.

12. Rafiei V, Vélëz H, Tzelepis G. The role of glycoside hydrolases in phytopathogenic fungi and omycetes virulence. Int J Mol Sci. 2021;22:9359. https://doi.org/10.3390/ijms22179359.
13. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nuc Acids Res. 2015;43:W78–84. https://doi.org/10.1093/nar/gkv487.
14. Bradley EL, Ökmen B, Doehlemann G, Henrissat B, Bradshaw RE, Mesarich CH. Secreted glycoside hydrolase proteins as effectors and invasion patterns of plant-associated fungi and oomycetes. Front Plant Sci. 2022;13:853106. https://doi.org/10.3389/fpls.2022.853106.
15. Dadwal A, Sharma S, Satyanarayana T. Thermostable cellulose saccharifying microbial enzymes: characteristics, recent advances and biotechnological applications. Int J Biol Macromol. 2021;188:226–44. https://doi.org/10.1016/j.ijbiomac.2021.08.024.
16. Joshi RK. Detailed description of the methodology. 2024. Figshare dataset, https://doi.org/10.6084/m9.figshare.26207534.v1
17. Joshi RK. Characterization of Alternaria porri isolate Apn Nashik. 2024. Figshare. figure, https://doi.org/10.6084/m9.figshare.26212271.v1
18. Joshi RK. Onion plant infected with Alternaria porri Apn-Nashik. 2024. Figshare figure, https://doi.org/10.6084/m9.figshare.26213117.v1
19. Joshi RK. Read and assembly statistics. 2024. Figshare dataset, https://doi.org/10.6084/m9.figshare.26213180.v1
20. Joshi RK. BUSCO dothideomycetes completeness statistics. 2024. Figshare dataset, https://doi.org/10.6084/m9.figshare.26213216.v1
21. Joshi RK. Statistics on Genome Repeat masking. 2024. Figshare dataset. https://doi.org/10.6084/m9.figshare.26213237.v1
22. Joshi RK. Predicted gene file of Alternaria porri using AUGUSTUS. 2024. Figshare dataset. https://doi.org/10.6084/m9.figshare.24906318.v1
23. Joshi RK. Summary of Protein coding genes. 2024. Figshare dataset. https://doi.org/10.6084/m9.figshare.26213297.v1
24. Joshi RK. Functional annotation from GO and KEGG databases. 2024. Figshare dataset. https://doi.org/10.6084/m9.figshare.26213552.v1
25. Joshi RK. Genes encoding pathogenicity related factors derived from PHI database. 2024. Figshare dataset. https://doi.org/10.6084/m9.figshare.26213564.v1
26. Joshi RK. Output files showing annotated genes of Alternaria porri using the dbCAN2 program. Figshare Dataset. 2024. https://doi.org/10.6084/m9.figshare.24906378.v1.
27. Joshi RK. Gene encoding Cazymes in Alternaria porri. 2024. Figshare dataset. https://doi.org/10.6084/m9.figshare.26213570.v1
28. Joshi RK. Genome comparison of Alternaria Porri Apn Nashik with other Alternaria species. 2024. Figshare figure. https://doi.org/10.6084/m9.figshare.26213636.v1
29. Joshi RK. Clusters of orthologous genes shared with and among the Alternaria spp. Figshare Fig. 2024. https://doi.org/10.6084/m9.figshare.26213702.v1.
30. Joshi RK. NCBI SRA database of the Alternaria Porri Apn Nashik genome. NCBI; 2024. https://identifiers.org/ncbi/insdc.sra:SRX26191501
31. Joshi RK. Datasets of the genomic assembly of *A. Porri*. NCBI; 2024. https://identifiers.org/nucleotide:JAPZQB000000000.1.

Publisher's note