Draft genome sequence of *Sclerospora graminicola*, the pearl millet downy mildew pathogen

S. Chandra Nayak a,⁎, H. Shekar Shetty a, C.Tara Satyavathi b, Rattan S. Yadav c, P.B.Kavi Kishor d, M. Nagaraju e, T.A. Anoop f, M. Milner Kumar g, Boney Kuriakose h, Navajeet Chakravartty e, A.V.S.K. Mohan Katta i, V.B. Reddy Lachagarie j, Om Vir Singh k, Pranav Pankaj Sahu l, Swati Puranik c, Pankaj Kaushal m, Rakesh K. Srivastava n, r.k.srivastava@cgiar.org (R.K. Srivastava).

⁎ Corresponding author.

A R T I C L E  I N F O

Keywords:
*Sclerospora graminicola*

Pathotype 1

Downy mildew

Whole genome sequence

A B S T R A C T

*Sclerospora graminicola* pathogen is the most important biotic production constraints of pearl millet in India, Africa and other parts of the world. We report a *de novo* whole genome assembly and analysis of pathotype 1, one of the most virulent pathotypes of *S. graminicola* from India. The draft genome assembly contained 299,901,251 bp with 65,404 genes. This study may help understand the evolutionary pattern of pathogen and aid elucidation of effector evolution for devising effective durable resistance breeding strategies in pearl millet.

Pearl millet (*Pennisetum glaucum* (L.) R. Br.), is an important crop of the semi-arid and arid regions of the world. It is capable of growing in harsh and marginal environments with the highest degree of tolerance to drought and heat stresses among cereals [1].

Downy mildew is the most devastating disease of pearl millet caused by *Sclerospora graminicola* (sacc. Schroet), particularly on genetically uniform hybrids. Estimated annual grain yield loss due to downy mildew is approximately 10–80% [2–7].

Pathotype 1 has been reported to be the highly virulent pathotype of *Sclerospora graminicola* in India [8]. We report a *de novo* whole genome assembly and analysis of *S. graminicola* pathotype 1 from India.

A susceptible pearl millet genotype Tift 23D3P1-P5 was used for obtaining single-zoospore isolates from the original oosporic sample. The library for whole genome sequencing was prepared according to the instructions by NEB ultra DNA library kit for Illumina (New England Biolabs, USA). The libraries were normalized, pooled and sequenced on Illumina HiSeq 2500 (Illumina Inc., USA) platform at 2 × 300 bp length. One SMRTbell library was prepared with 20Kb insert size sequenced on PACBIO RSII platform.

The whole genome sequencing was performed by sequencing of 7.38 Gb with 73,889,924 paired-end reads from the paired-end library, and 1.15 Gb with 3,851,788 reads from the mate pair library generated from Illumina HiSeq2500 and Illumina MiSeq, respectively. Illumina reads were filtered with a quality score of at least 20 and read duplicates were removed before the assembly. A total 597,293 filtered sub reads with average read length of 6.39 bp was generated on PACBIO RSII with P6-C4 chemistry. Approximately 51% of data generated from the reads with more than 10Kb length with a maximum read length of 49,261 bp. The sequences were assembled using various genome assemblers like ABySS, MaSuRCA, Velvet, SOAPdenovo2, and ALLPATHS-LG. The hybrid assembly generated by MaSuRCA [9] algorithm was found to be superior over other algorithms (Table 1). Assembled draft genome sequence of *S. graminicola* pathotype 1 was 299,901,251 bp in length, N50 of 17,909 bp with a minimum of 1 Kb scaffold size. The GC content was 47.2% consisting of 26,786 scaffolds with longest scaffold size of 238,843 bp. The overall coverage was 40 × . The draft genome
sequence was used for gene prediction using AUGUSTUS which resulted in 65,404 genes using Saccharomyces cerevisiae as a model. The completeness of the assembly was investigated through CEGMA and revealed 92.7% proteins completely present and 95.6% proteins partially present, while BUSCO v3 fungal dataset indicated 64.9% complete, 12.4% fragmented, 22.7% missing out of 290 BUSCO groups. A total of 52,285 predicted genes found homology using BLASTX against nr database and 38,120 genes annotated a set of 11,873 genes had UniProt entries, while 7248 were GO terms and 9686 with KEGG IDs. Of the 7248 GO terms, 2724 were associated with the biological processes. Some important GO terms are listed in Table 2. During the annotation, we are observed many protein molecules which have known role in pathogenicity. Some of these include Crinkler (CRN) family protein, Glucanase inhibitor, Serine protease inhibitor, Cystiene inhibitor, Infl1 Elicitor-like protein, SWI4 1, Peter Pan-like protein suppressor, Sterol binding protein, PexRD2, Glyceraldehyde-3-phosphate dehydrogenase, Ribonuclease, HECT E3 ubiquitin ligase, Alpha-1,2-Mannosidase, Endo1,3(4)-beta-glucanase putative, Palmitoyltransferase, Serine/threonine-protein phosphatase 2A activator, Protein kinase, putative, NAD-dependent histone deacetylase sir2-like protein, rpp 13-like proteins, rpm, Glycoside hydrolase, Pre-mRNA-splicing factor SF2, NADH dehydrogenase flavoprotein 1, Mitochondrial Aldehyde dehydrogenase, Deoxyhypusine hydroxylase, DEAD/DEAH box RNA helicase, putative, CAMK protein kinase, Alpha-1,2-Mannosidase, Ornithine aminotransferase, mitochondrial Phosphatidate cytidylyltransferase, Acetolactate synthase, Inositol hexakisphosphate and diphosphoinositolpentakisphosphate kinase.

Repetitive element analysis with Repbase revealed 115 Ty1/Copia, 50 Gypsy, 419 small RNA, 23,618 simple repeats and 3365 low complex repeats. Microsatellite analysis with misa tool revealed 8179 mononucleotide repeats, 1082 low complexity repeats and 5562 dinucleotide to hexanucleotide repeats. S. graminicola pathotype 1 genome characteristics and resources are mentioned in Table 3.

The S. graminicola pathotype 1 sample has been deposited at the National fungal herbarium facility with accession number 52052 at the Herbarium Cryptogamae Indiae Orientalis (HCIO), Division of Plant Pathology, Indian Agricultural Research Institute (IARI), New Delhi, India.

Information on deposited data

The genome information of downy mildew pathogen is available in the NCBI GenBank database. The Sclerospora graminicola whole genome shotgun (WGS) project has the project accession MQA00000000, and consists of sequences MQA02000001-MQA020026786, with BioProject ID PRJNA325098 and BioSample ID SAMN05219233, and can be accessed at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA325098/.

Author contributions

RKS, HSS, CNS designed the experiment. RKS, CNS, SSH, CTS, RSY, VBL, ATA, MKM, BK, NC, MRAVSK performed research. RKS, CNS, VBL, SSH, RYS, CTS, PPS, SP, PK, OVS wrote the manuscript.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

The authors gratefully acknowledge the support received from the All India Coordinated Research Project on Pearl Millet (AICRP-PM), Mandor; and the Indian Institute of Millets Research (IIMR), Hyderabad in collecting and maintaining downy mildew inoculum. The financial assistance received from ICAR and AgriGenome Labs is gratefully acknowledged. The authors also acknowledge the Herbarium Cryptogamae Indiae Orientalis (HICO), Division of Plant Pathology, Indian Agricultural Research Institute (IARI), for undertaking conservation of the deposited fungal specimen.


