Short communication

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

S. Chandra Nayaka[a], H. Shekar Shetty[a], C.Tara Satyavathi[b], Rattan S. Yadav[c], P.B.Kavi Kishor[d], M. Nagaraju[d], T.A. Anoop[e], M. Milner Kumar[e], Boney Kuriakose[e], Navajeet Chakravartty[e], A.V.S.K. Mohan Katta[e], V.B. Reddy Lachagari[e], Om Vir Singh[b], Pranav Pankaj Sahu[c], Swati Puranik[c], Pankaj Kaushal[f], Rakesh K. Srivastava[g,*]

[a] Downy Mildew Research Laboratory, DOS in Biotechnology, University of Mysore, Mysuru, Karnataka, India
[b] PC Unit ARS Mandore, Jodhpur, Rajasthan, India
[c] Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Gogerddan, SY23 3EB, United Kingdom, UK
[d] Osmania University, Hyderabad, Telangana, India
[e] AgriGenome Labs Pvt. Ltd, Hyderabad, Telangana, India
[f] ICAR-National Institute of Biotic Stress Management, Raipur India
[g] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, Telangana, 502324, India

A R T I C L E I N F O

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 23D$_2$B$_1$P1-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., San Diego, CA, USA) platform at 2 × 100 bp length. Mate pair (MP) libraries were prepared using the Nextera mate pair library preparation kit (Illumina Inc., USA). The libraries were normalized, pooled and sequenced on Illumina MiSeq

(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 Kb 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, N$_{50}$ 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

Table 1
Comparative statistics of the promising genome assemblers.

Assembler	Minimum	Maximum	Mean	N$_{50}$	No. of Contigs	Sum of Contigs	CEGMA Complete	CEGMA Partial
Abyss_DBG2OLC Scaffold	2730	235,195	27,432	22,557	5126	140,615,056	56.45	62.5
SOAP_DBG2OLC Scaffolds	2079	194,864	28,748	26,386	5404	155,354,949	65.73	71.37
MaSuRCA_Scaffolds	1000	238,843	11,196	17,909	26,786	299,901,251	89.52	93.95

Table 2
Important biological process identified using GO annotation.

Biological process	Number of GO terms
DNA integration [GO:0015074]	699
NADH dehydrogenase (ubiquinone) activity [GO:0008137]	95
Cytochrome-c oxidase activity [GO:0004129]	81
ATP binding [GO:0005524]	79
Heme binding [GO:0020037]	68
Cytochrome-c oxidase activity [GO:0004129]	68
Hydrogen ion transmembrane transporter activity [GO:0015078]	61
Proton-transporting ATP synthase complex, coupling factor F(o) [GO:0045263]	58
Microtubule motor activity [GO:0003777]	35
Intracellular signal transduction [GO:0035556]	34
Small-subunit processome [GO:0032040]	29
Unfolded protein binding [GO:0051082]	24
Magnesium ion binding [GO:0000287]	20
Intracellular protein transport [GO:0006886]	20

Table 3
Sclerospora graminicola pathotype 1 genome characteristics and resources.

Name	Genome characteristic/resource
NCBI bioproject ID	PRJNA325098
NCBI biosample ID	SAMN05219233
NCBI SRA accession No.	SRP076363 with accession numbers SRR3658180 and SRR3658181
Sequence type	Illumina HiSeq2500 and Illumina MiSeq, PacBio RSII
Total number of reads	73,889,924 from PE Library, 3,851,788 from MP Library
Read length	2 × 100 bp for PE and 2 × 300 bp for MP
Overall coverage	40×
Estimated genome size	299.9 Mb
Predicted protein coding genes	65,404
Annotated Genes	38,120

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 MIQA00000000. This version of the project (02) has the accession number MIQA02000000, and consists of sequences MIQA02000001-MIQA02026786, 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, VBRL, ATA, MKM, BK, NC, MKAVSK performed research. RKS, CNS, VBRL, SSH, RSY, 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

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 were observed with a significant BLASTX match with E-value cutoff of 1e-5 and 40% identity percentage. Out of 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, INF1 Elicitin-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, Endo-1,3(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 diphosphoinositol-pentakisphosphate 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

References

[1] A. Supriya, S. Senapathy, C.T. Hash, N. Thirunavukkarasu, R. Vengaldas Kankanti, R. Sharma, R.P. Thakur, P.R. Veeranki, R.C. Yadav, R.K. Srivastava, QTL mapping of pearl millet rust resistance using an integrated DArT- and SSR-based linkage map, Euphytica 209 (2) (2016) 461–476.

[2] S.D. Singh, S.B. King, J. Werder, Downy Mildew Disease of Pearl Millet. Information Bulletin No. 37, Patancheru, Andhra Pradesh 502324, International Crops Research Institute for the Semi-Arid Tropics, India, 1993, p. 36.

[3] S.D. Singh, Downy mildew of pearl millet, Plant Dis. 79 (1995) 545–550.

[4] C.T. Hash, S.D. Singh, R.P. Thakur, B.S. Talukdar, Breeding for disease resistance, in: I.S. Khairwal, K.N. Rai, D.J. Andrews, G. Harinarayana (Eds.), Pearl Millet Breeding, Oxford & IBH, New Delhi, India, 1999, pp. 337–379.

[5] D.E. Hess, R.P. Thakur, C.T. Hash, P. Sérémé, C.W. Magill, Pearl millet downy mildew: problems and control strategies for a new millennium, in: J.F. Leslie (Ed.), Sorghum and Millets Diseases Ames, Iowa State Press, Iowa, USA, 2002, pp. 37–42.

[6] O.P. Yadav, K.N. Rai, Genetic improvement of pearl millet in India, Agric. Res. 2 (2013) 275–292.

[7] C.P. Anup, K.R. Kini, Analysis of dynamics of proteome in resistant cultivar of pearl millet seedlings during *Sclerospora graminicola* infection, J. App. Biol. Biotechnol. 4 (2016) 067–071.

[8] J. Sudisha, K.S. Ananda, H.S. Shetty, Characterization of Downy Mildew Isolates of *Sclerospora graminicola* by using differential cultivars and molecular markers, J. Cell Mol. Biol. 7 (2008) 41–55.

[9] A.V. Zimin, M. Guillaume, P. Daniela, R. Michael, L.S. Steven, A.Y. James, The MaSuRCA genome assembler, Bioinformatics 29 (2013) 2669–2677.