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

Nayaka, S. Chandra; Shetty, H. Shekar; Tara Satyavathib, C.; Yadav, Rattan; Kishor, P. B. Kavi; Nagaraju, M.; Anoop, T. A.; Kumar, M. Milner; Kuriakose, Boney; Chakravarty, Navajeet; Katta, A. V. S. K. Mohan; Lachagan, V. B. Reddy; Singh, Om Vir; Sahu, Pranav Pankaj; Puranik, Swati; Srivastava, Rakesh; Kaushal, Pankaj

Published in:
Biotechnology Reports
DOI:
10.1016/j.btre.2017.07.006
Publication date:
2017

Citation for published version (APA):
Nayaka, S. C., Shetty, H. S., Tara Satyavathib, C., Yadav, R., Kishor, P. B. K., Nagaraju, M., ... Kaushal, P. (2017). Draft genome sequence of Sclerospora graminicola, the pearl millet downy mildew pathogen: Genome sequence of pearl millet downy mildew pathogen. Biotechnology Reports, 16, 18-20. https://doi.org/10.1016/j.btre.2017.07.006
Draft genome sequence of Sclerospora graminicola, the pearl millet downy mildew pathogen

S. Chandra Nayaka, H. Shekar Shetty, C.Tara Satyavathi, Rattan S. Yadav, P.B.Kavi Kishor, M. Nagaraju, T.A. Anoop, M. Milner Kumar, Boney Kuriakose, Navajeet Chakravarty, A.V.S.K. Mohan Kattae, V.B. Reddy Lachagarie, Om Vir Singh, Pranav Pankaj Sahu, Swati Puranik, Pankaj Kaushal, Rakesh K. Srivastava

ABSTRACT

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.

Keywords:
Sclerospora graminicola
Pathotype 1
Downy mildew
Whole genome sequence

Short communication

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 23D3bP1-P5 was used for obtaining single-zooospore 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 and Illumina MiSeq, respectively. Illumina 2 × 100 bp length paired-end 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

Received 19 May 2017; Received in revised form 13 July 2017; Accepted 17 July 2017
Available online 24 August 2017
2215-017X/ © 2017 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).
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, 50 Gypsy, 419 small RNA, 23,618 simple repeats and 3365 low complexity repeats. Microsatellite analysis with misa tool revealed 8179 mono-nucleotide repeats, 1082 low complexity repeats and 5562 dinucleotide repeats. ATP binding [GO:0005244] 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:0005108] 24
Magnesium ion binding [GO:0002878] 20
Intracellular protein transport [GO:0006886] 20

Table 1 Comparative statistics of the promising genome assemblers.

| Assembler                  | Minimum | Maximum | Mean   | N50   |
|----------------------------|---------|---------|--------|-------|
| Abyss_DBG2OLC Scaffold     | 2730    | 235,195 | 27,432 | 22,557|
| SOAP_DBG2OLC Scaffolds     | 2079    | 194,864 | 28,748 | 26,386|
| MaSuRCA_Scaffolds          | 1000    | 238,843 | 11,196 | 17,909|
|                            |         |         |        | 5126  |
|                            |         |         |        | 140,615,056 |
|                            |         |         |        | 56.45  |
|                            |         |         |        | 62.5   |
|                            |         |         |        | 5404  |
|                            |         |         |        | 155,354,949 |
|                            |         |         |        | 65.73  |
|                            |         |         |        | 71.37  |
|                            |         |         |        | 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:0005244]                                | 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:0005108]                   | 24                 |
| Magnesium ion binding [GO:0002878]                      | 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 x 100 bp for PE and 2 x 300 bp for MP |
| Overall coverage            | 40 x |
| 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 Cryptogamiae Indicae 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/](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, 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 Agri Genome Labs is gratefully acknowledged. The authors also acknowledge the Herbarium Cryptogamiae Indicae Orientalis (HCIO), Division of Plant Pathology, Indian Agricultural Research Institute (IARI), for undertaking conservation of the deposited fungal specimen.
References

[1] A. Supriya, S. Senapathy, C.T. Hash, N. Thirunavukkarasu, R. Vengaladas 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. Sébé-Ogée, 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. Rev. 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.