3-15-2017

Draft Genome Sequence of *Cercospora sojina* Isolate S9, a Fungus Causing Frogeye Leaf Spot (FLS) Disease of Soybean

Fanchang Zeng
Shandong Agricultural University, China

Chaofan Wang
Shandong Agricultural University, China

Guirong Zhang
University of Illinois at Urbana-Champaign

Junmei Wei
Shandong Agricultural University, China

Carl A. Bradley
University of Kentucky, carl.bradley@uky.edu

See next page for additional authors

Follow this and additional works at: https://uknowledge.uky.edu/plantpath_facpub

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Part of the Diseases Commons, Fungi Commons, Genetics and Genomics Commons, and the Plant Pathology Commons

Repository Citation

Authors

Fanchang Zeng, Chaofan Wang, Guirong Zhang, Junmei Wei, Carl A. Bradley, and Ray Ming

Draft Genome Sequence of *Cercospora sojina* **Isolate S9, a Fungus Causing Frogeye Leaf Spot (FLS) Disease of Soybean**

Genomics Data

Data in Brief

Draft genome sequence of *Cercospora sojina* isolate S9, a fungus causing frogeye leaf spot (FLS) disease of soybean

Fanchang Zeng [a,b,1], Chaofan Wang [a,1], Guirong Zhang [b], Junmei Wei [a], Carl A. Bradley [c,d], Ray Ming [b,e,*]

[a] State Key Laboratory of Crop Biology, College of Agronomy, Shandong Agricultural University, Tai'an, Shandong 271018, China
[b] Department of Plant Biology, University of Illinois, Urbana, IL 61801, USA
[c] Department of Plant Pathology, University of Kentucky, Princeton, KY 42445, USA
[d] Department of Crop Sciences, University of Illinois, Urbana, IL 61801, USA
[e] FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China

ARTICLE INFO

ABSTRACT

Fungi are the causal agents of many of the world's most serious plant diseases causing disastrous consequences for large-scale agricultural production. Pathogenicity genomic basis is complex in fungi as multicellular eukaryotic pathogens. The fungus *Cercospora sojina* is a plant pathogen that threatens global soybean supplies. Here, we report the genome sequence of *C. sojina* strain S9 and detect genome features and predicted genomic elements. The genome sequence of *C. sojina* is a valuable resource with potential in studying the fungal pathogenicity and soybean host resistance to frogeye leaf spot (FLS), which is caused by *C. sojina*. The *C. sojina* genome sequence has been deposited and available at DDBJ/EMBL/GenBank under the project accession number AHPQ00000000.

Specifications

Organism/cell line/tissue	*Cercospora sojina*
Sex	–
Sequencer or array type	Illumina GA IIx
Data format	Processed
Experimental factors	DNA extracted from a field strain, no treatment
Experimental features	Whole genome sequence
Consent	n/a
Sample source location	Soybean field from Georgia, USA

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/nuccore/AHPQ00000000.

2. Materials and methods

2.1. C. sojina whole-genome sequencing and assembly

Genome of *C. sojina* strain S9 [1], from a soybean field in Georgia, was sequenced using the Illumina GA IIx next generation technology by paired-end sequencing method to a depth of 239× at the Keck Center at the University of Illinois Urbana-Champaign. The produced sequences had a read length of 124 base pairs (bp). A total of 29,619,123 reads from each end were produced for a total of 59,238,246 reads from one lane. *C. sojina* genomes were assembled using Velvet algorithm [2] to obtain optimized results with high quality assembly. The assembly contained 30,797,991 bases with 1804 scaffolds (N50 = 37.690 bp) and the G + C content was 53.80%.

2.2. C. sojina genome annotation.

The *C. sojina* genes were predicted with *ab initio* gene finders (FGENESH, FGENESH+, and GENEWISE). We referred to the gene models from *Zymoseptoria tritici* (*Mycosphaerella graminicola*) as the most closely related species with *C. sojina* to train the gene finding programs. BlastX against publicly available non-redundant protein and BlastN against ESTs databases are used to validate and curate predicted complete coding regions of the gene models. The entire DNA sequence was also compared against the nonredundant protein databases in all six reading frames, using BlastX with threshold E < 1e^{-5} to identify any possible coding sequences previously missed by using ARTEMIS to collate data and facilitate annotation [3]. Finally, a non-redundant set of gene models is produced, in which a single best gene model per locus is selected, preferring the candidate annotation with supporting evidence of homolog protein/EST sequence in public database and complete coding sequence region. rRNA genes were identified with RNAmmer [4]. And tRNA and ncRNA genes were identified by RFAM

* Corresponding author.
E-mail address: rming@life.illinois.edu (R. Ming).
[1] F. Z. and C. W. contributed equally to this work.

Table 1

C. sojina genome sequencing assembly and annotation statistics.

Assembly and annotation	No.
Total no. paired-end reads	59,238,246
Average depth-coverage of mapped reads	239
Average read length (bp)	124
Assembly size (bp)	30,797,991
Total contigs (>500 bp)	1804
N50	37,690
Protein-coding genes	9099
Protein coding genes >100 amino acids	8868
Average CDS size (bp)	1764
Average gene length (bp)	2742
% coding	64
Genes with known function	7542
Genes with unknown function	1557
rRNA	72
ncRNA	9
tRNA	63

and TRNASCAN-SE [5,6] in JGI Annotation system. The complete genome sequence contained 9099 protein-coding genes, 72rRNAs, 63tRNAs, 9ncRNAs. The average size of a gene was 2742 bp and the average CDS size was 1764 bp. The *C. sojina* sequencing and assembly statistics are summarized in Table 1.

2.3. Genomics functional annotation.

All predicted genes are annotated for Gene Ontology (GO) using Blast2Go function annotation system [7,8], according to sequence comparison with BlastP and domains/motif identification with interProScan [7] and PFAM [9]. The *C. sojina* genome encoded predominant potential genomic elements involved in mycelium related development. It is through the mycelium that a fungus absorbs nutrients from its environment. The major set of genes involved in the *C. sojina* life cycle suggests the fundamental systems for life sustainment. Additionally, the *C. sojina* genome contains extensive genetic factors involved in auxin biosynthetic process, and Ser/Thr protein kinases signal transduction and transcription factors regulation. It implies that considerable auxin regulation and signal transduction pathway involved in appropriate physiology

processes and pathogenesis. ABC transporters (ATP binding cassette protein) genes were highly redundant and assigned to the largest set of molecular function category in *C. sojina* genome. These results accord with the report that ABC transporters act as an essential virulence factors by mediating secretion of host-specific toxins compounds during pathogenesis [10].

References

[1] M.A.R. Mian, A.M. Missaoui, D.R. Walker, D.V. Phillips, H.R. Boerma, Frogeye leaf spot of soybean: a review and proposed race designations for isolates of *Cercospora sojina* Hara. Crop Sci. 48 (2008) 14–24.

[2] D.R. Zerbino, E. Birney, Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18 (2008) 821–829.

[3] K. Rutherford, J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.A. Rajandream, B. Barrell, Artemis: sequence visualization and annotation. Bioinformatics 16 (2000) 944–945.

[4] K. Lagesen, P. Hallin, E.A. Rodland, H.H. Staerfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35 (2007) 3100–3108.

[5] S. Griffiths-Jones, S. Moxon, M. Marshall, A. Khan-na, S.R. Eddy, A. Bateman, Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33 (2005) D121–D124.

[6] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25 (1997) 955–964.

[7] S. Götz, J.M. García-Gómez, J. Terol, T.D. Williams, S.H. Nagaraj, M.J. Nueda, M. Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36 (2008) 3420–3435.

[8] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (2005) 3674–3676.

[9] A. Bateman, E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S.R. Eddy, S. Griffiths-Jones, K.L. Howe, M. Marshall, E.L. Sonnhammer, The Pfam protein families database. Nucleic Acids Res. 30 (2002) 276–280.

[10] L.H. Zwiers, I. Stergiopoulos, J.G. Van Nistelrooy, M.A. De Waard, ABC transporters and azole susceptibility in laboratory strains of the wheat pathogen *Mycosphaerella graminicola*. Antimicrob. Agents Chemother. 46 (2003) 3900–3906.