Draft genome sequence of *Cercospora canescens*: a leaf spot causing pathogen

Ramesh Chand[1,*], Chhattar Pal[1], Vineeta Singh[1], Manoj Kumar[1,2], Vinay Kumar Singh[3] and Pallem Chowdappa[4]

[1]Department of Mycology and Plant Pathology, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi 221 005, India
[2]Department of Biochemistry, Faculty of Science, Banaras Hindu University, Varanasi 221 005, India
[3]Centre for Bioinformatics, School of Biotechnology, Banaras Hindu University, Varanasi 221 005, India
[4]ICAR-Central Plantation Crops Research Institute, Kudlu, P.O. Kasargod 671 124, India

***Cercospora canescens* (Ellis and Martin) is a hemibiotrophic pathogen causing leaf spot disease on mungbean (*Vigna radiata* L). Genome sequence (~33.97 Mb) assembled in 8239 contigs with 10627 protein coding genes. A total of 2842 proteins were identified as homologous of 223 predicted and 7562 putative uncharacterized involved in biological processes, molecular functions. The identified proteins are mainly involved in infection process used to compromise nutrients or destroy host tissues gycosidases, transposases, cytochrome P450s, genes codes to signal transduction, cell wall breakdown, transporters, host stomata perception, adhesion, polyketide synthase and cercosporin. A total of 528 simple sequence repeats were also identified from the genome sequence assembly of *C. canescens*. This study provides insights into pathogenic mechanism and a better understanding of virulence differentiation of *C. canescens*. It will also help in identification of similarity and differences in regions among the genomes of different species of *Cercospora*.**

Keywords: *Cercospora canescens*, functional annotation, gene prediction, sequencing and assembly, *Vigna radiata*.

MUNGBEAN (*Vigna radiata* L.) is grown over 2.31 m ha in India with a total annual production of around 1.19 million tonnes[1,2]. *Cercospora canescens*; MTCC-10835 (Eliis & Martin) is a haploid, hemibiotrophic fungus belonging to Ascomycota (Dothideomycetes; Mycosphaerellaceae), one of the most destructive groups of plant pathogens that cause cercospora leaf spot (CLS) of mungbean (Figure 1). *C. canescens* infection is a serious problem as it can lead to significant loss (46–61%) in crop yield[3–5]. The disease spreads rapidly in susceptible varieties causing premature defoliation and reduction in the size of pods and grains[6]. Due to the evolution of new pathogenic variants of *C. canescens* (NFCCI 2370), many mungbean genotypes have become susceptible or show moderate response to the pathogenic nature of this fungi[7,8]. CLS is of particular concern in South East Asian countries, viz. India, Pakistan, Thailand, Philippines, Sri Lanka, Myanmar and Bangladesh; where the infection is widespread and mungbean is among the principal legumes[9,10]. In India, CLS was reported[11] to be an issue in majority of mungbean-growing states and even rose to epidemic proportions in Uttar Pradesh from 1934 and 1941. The isolation and detection of *C. canescens* is difficult due to its slow growth and poor sporulating nature in culture media[12]. The latter can be explained as sporulation of *C. canescens* is based on isolates rather than influenced by culture conditions[13]. The pathogen produces a photo-activated perylinquinone toxin called cercosporin, which helps the fungus to obtain its nourishment by killing host cells[14].

The *Cercospora* species are classified by their morphological characteristics and have been sorted into more than 5000 anamorphic species. The present DNA sequence data available in gene banks are partial and insufficient for species identification. However, recent phylogenetic studies have recognized several species complexes within the teleomorph genus, *Mycosphaerella* Deighton. Such studies on partial sequences were done using the internal transcribed spacer (ITS), though the ITS region lacks resolution to distinguish most *Cercospora* species. Phylogenetic studies based on partial sequences from the intervening 5.8S rRNA, actin, calmodulin, histone H3 and translation elongation factor 1-alpha genes used in multi-locus sequences have been one of the most effective ways to genetically recognize different species of *Cercospora*. However, this approach is inappropriate to recognize the host range of *Cercospora* species without extensive pathogenicity tests, though it does provide a basic understanding of the identity of a particular strain in question[15].

In light of these difficulties for the proper classification of *Cercospora*, the elucidation of the draft genome sequence will provide insight to better understanding the various biosynthesis pathways. Genes involved in various molecular functions, including cercosporin biosynthesis need to be explored in order to understand the role of secondary metabolites in its pathogenesis[16]. Another

*For correspondence. (e-mail: rc_vns@yahoo.co.in)

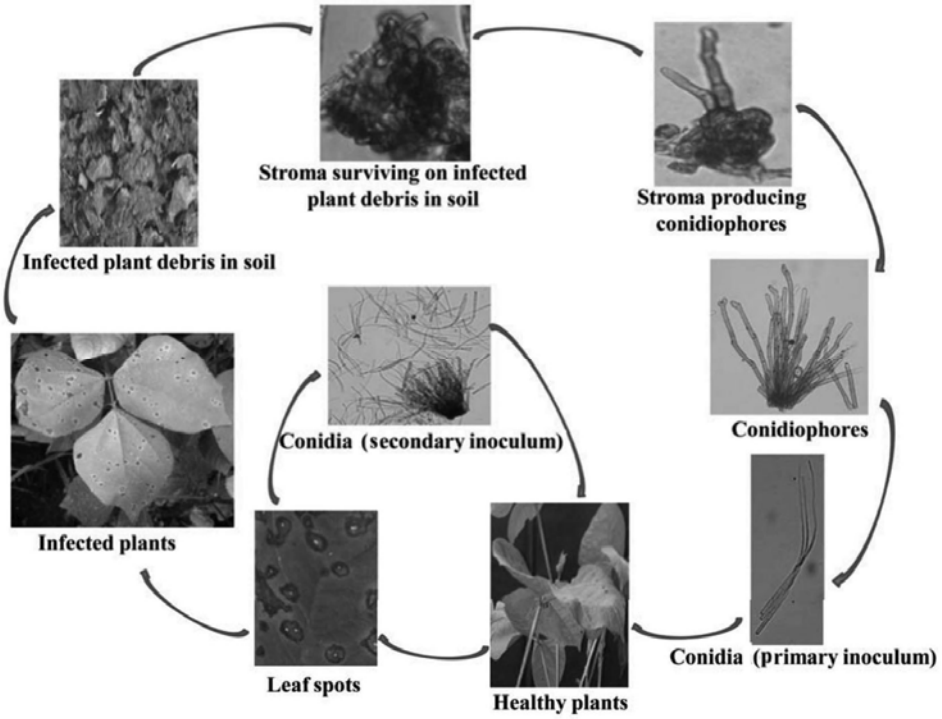

Figure 1. Life cycle of *Cercospora canescens*.

difficulty in controlling the spread of this pathogenic fungus is the development of gradual resistance against the applied fungicides (benzimidazoles: benomyl, carbendazim, thiophanate-methyl; demethylation inhibitors: tebuconazole, propiconazole, myclobutanil and triflumizole)[17,18]. To resolve the problem of fungicide resistance and for screening of new fungicides against pathogen infections, it is important to study the genetic expression involved in pathogenesis and virulence.

The genomic sequence of *C. canescens* will enable the detection of species and open avenues to several thousands of genes that play an important role in the production of secondary metabolites, mechanisms of pathogenicity, virulence and other important molecular functions. It will also provide an opportunity for comparative analyses of *C. canescens* with other species of *Dothideomycetes*.

Materials and methods

Strain, growth conditions and genomic DNA isolation

C. canescens (MTCC-10835) was isolated from an infected leaf of a mungbean plant at Banaras Hindu University (BHU), Varanasi, India[8]. The fungus was cultured at 25°C for 12 days in liquid potato dextrose medium. DNA was extracted by CTAB method, as previously described[19].

Sequencing and assembly

The sequencing was performed by a commercial service provider (Genotypic Technology, Bangalore, India) using the Genome Analyzer GAIIx (Illumina Inc., San Diego, CA). The sample was processed individually for short and long inserts. Sequence assembly was performed with Velvet (v. 1.2.07) tool. Raw reads were subjected to quality control using SeqQC (http://genotypic.co.in/SeqQC.html).

Gene prediction and annotation

The optimal hash length contigs were selected for gene prediction using the Augustus tool[20]. The predicted proteins were subjected to BLASTp analysis against non-redundant (nr) protein sequences of the *Ascomycota* group using NCBI database[21]. The non-redundant gene bank databases, including coding sequence (CDS) translations, protein data bank (PDB), Swiss-prot, protein information resource (PIR), protein research foundation (PRF) and excluding environmental samples from whole genome sequence (WGS) projects were used as default parameters. The functional annotation was performed using the gene ontology database[22] for molecular function, biological processes and cellular components.

Identification of simple sequence repeats

The simple sequence repeats (SSRs) were mined in the genome sequence using the micro-satellite identification

tool (MISA) available under http://pgrc.ipk-gatersleben.de/misa and the SSR Locator[23,24] programs at least ten repeats for mono, six repeats for di, five repeats for tri, tetra, penta and hexa-nucleotides. The Primer3 program[25] was used to design the primers pair for the identification of SSRs, annealing temperature (T_m) between 57°C and 60°C, product size between 102 and 500 bp, primer length between 18 and 24 bp, and GC content between 40% and 60%.

Phylogenetic relationship of candidate genes

The phylogenetic relationship was performed with mitogen activated protein kinase (MAPK) and isocitrate lyase (ICL) gene sequences, US. The homologous sequences were retrieved from the NCBI database and Fungal Genome portal available at the Joint Genome Institute, US, Department of Energy (JGI, DOE; http://jgi.doe.gov). The amino acid sequences were aligned with ClustalW[26] and an UPGMA-based phylogenetic tree was constructed using MEGA5.01 with a bootstrap value of 1000 (ref. 27). A relationship was established with *C. zeae-maydis* (Tehon and Daniels) as a closely related species to classify *C. canescens* and another two species of *Mycosphaerella*. Other members of the class Dothideomycetes: *Zymoseptoria tritici* (Fuckel) J. Schröt, *Phaeosphaeria nodorum* (E. Müll.) Hedjar and *Leptosphaeria maculans* (Sowerby) P. Karst, were also used for comparison. Similar homologs of MAPK and ICL genes from eight different fungal species were individually subjected to protein signature testing using the Inter ProScan server[28]. The protein sequences for each gene were used for multiple sequence alignments and phylogenetic classification.

Results and discussion

Genome sequencing and assembly

The shotgun sequence of the *C. canescens* genome (~33.97 Mb) consists of 8239 curated contigs. These contigs were collectively deposited to the National Centre for Biotechnology Information (NCBI) under accession number ANSM00000000 (Table 1). The raw reads were deposited in the Sequence Read Archive (SRA) database of NCBI with accession number SRX209601. A total of 22.14 million reads (2.8 Gb) were generated by the Genome Analyzer IIx system (Illumina Inc., San Diego, CA). The high quality scores (also known as the *Q*-score) with greater than 20 bases were numbered more than 93.06% in the forward and 85.45% in the reverse paired end reads. The percentage of unresolved bases (Ns) was observed to be minimal (0.162% in forward read and 0.345% in reverse reads). The results also show that the average *Q*-scores are above 30 bases and up to the 76 base positions (Figure 2 *a*). The sequence assembly of the processed reads provided the minimum contig length of 105 bp and maximum of 79323 bp using Velvet (V2.1.07)[29]. These high-quality processed paired end reads were used to assemble the contigs. The number of contigs with >200 bp was 6606 and contained 59.90% of the sequence, whereas the number of contigs with >10 kb was 1119 and comprised of 10.15% of the total assembly (Figure 2 *b*). The GC content observed was 26% and 25% respectively (Figure 2 *c*). The higher N50 value represents better assembly. Hence our sequence assembly can be assumed to be of high quality, as 13,944 N50 bases were present[30].

Functional annotation

Within the *C. canescens* genome, 10,627 protein coding genes were predicted (see Additional file 1 online). For the comparative analysis, 10,082 predicted genes from *Neurospora crassa* Shear & Dodge and 9457 genes from the distantly related organism *Aspergillus nidulans* G. Winter were taken. Both species are non-pathogenic in nature, but the pathogenic *Magnaporthe grisea* (Hebert) Barr genome contains 11,109 predicted genes[31]. The *C. canescens* proteins show significant similarity to *Ascomycota* proteins, as indicated by the UniProt database[32]. From these, 2842 were homologous to characterized proteins, 223 showed predicted proteins and 7562 were found similar to putative uncharacterized proteins. The putative orthologous genes were found to be involved in several different pathways, including cellular and molecular processes, seemingly conserved in *C. canescens.* In addition, many *C. canescens* proteins showed homology to uncharacterized, predicted and annotated proteins. Gene ontology (GO) terms were assigned to the putative *C. canescens* proteins, which showed significant similarity to *Ascomycota.* A total of 12,027 proteins were assigned at least one GO term, among which 3798 were assigned in the biological process category, 5582 in molecular function and 2647 in the cellular component category (see Additional file 1 online). For biological processes, top ten GO annotation proteins were taken to generate a functional annotation graph, including metabolic (723), transport

Table 1. *Cercospora canescens* whole genome assembly features

Genome characteristics	Value
Size (bp)	33,967,224
Number of contigs	8,239
Contigs >10 kb	1,119
Maximum contig length	79,323
N50 bases	13,944
Protein coding genes	10,627
G + C content (%)	51%

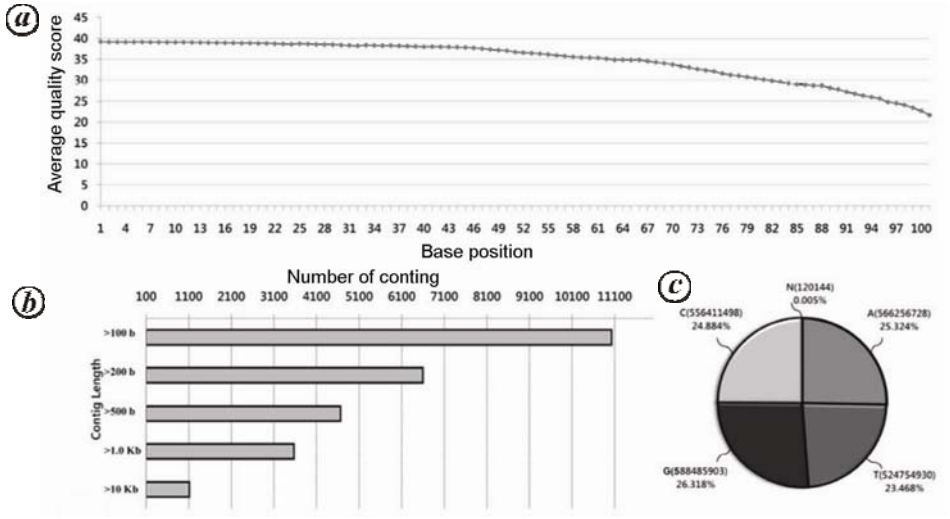

Figure 2. Assembly information. ***a***, Quality score; ***b***, Contig length distribution; ***c***, ATGC composition.

(578), biosynthetic (434), transcription (401) and catabolic (334) processes. The genes involved in other important biological processes such as oxidation–reduction (264), carbohydrate metabolism (200), proteolysis (166), signal transduction (115) and nucleotide and nucleic acid process (60) were also identified through GO annotations. Similarly, reductase (832), zinc metal binding (685), ATP binding (622), transferase (524), hydrolase (433) and catalytic (326) activity were highly represented. Genes that were important to other molecular functions such as kinase activity (243) and nucleic acid binding (238) were also identified. The genes that were highly represented among the cellular component category were integral to membrane (805), nuclear (537), cytoplasmic (249), mitochondrial (178), ribosomal (116) and intracellular regions (115) (Figure 3).

Pathogenicity and virulence associated genes in C. canescens

C. canescens is a highly pathogenic plant fungus. Therefore, it has a number of genes associated with the host–pathogen interaction. We have identified the pathogen–host interaction (PHI) genes from pathogen–host interaction database (see http://www.insilicogenomics.in/ccvir/ccvir.html; Additional file 2 online). The role of these genes in pathogenesis, breakdown of cell wall, signal transduction, adhesion and cercosporin production is well studied in other plant pathogenic fungi like *Cercospora nicotianae and C. kikuchii*. The MAP kinases (EC 2.7.11.24) are a family of serine/threonine-specific protein kinases, which are evolutionarily conserved in eukaryotic organisms. These kinases are activated through protein kinase cascades and respond to diverse external stimuli[33]. Notably, the MAPKs are well established in regulating morphology and virulence in *C. zeae-maydis*[34]. Targeted disruption of CZK3 suppressed gene expression was predicted to participate in cercosporin biosynthesis and consequently abolish-cercosporin production in *C. zeae-maydis*. The homologs of CZK3 (ANSM01001723.1) were found in *C. canescens*[35]. Similarly, in *M. grisea*, deletion of the MAPK homolog, OSM1, induces pleiotropic effects, including osmotic sensitivity, reduced conidiation and over production of appressoria[36]. Host invasion begins with the emergence of a germ tube from the conidium, followed by appressorium formation[37]. A PMK1-deficient mutant of *M. grisea* is unable to develop the appressoria in rice plants and hence becomes nonpathogenic. Therefore, appressorium formation was found to be specifically regulated by PMK1 in response to surface signals and is also necessary for invasive growth. The later stages of plant penetration are regulated by MAP kinase and PMK1 (ANSM01002982.1)[38]. Further, these data suggest that the *C. canescens* genome encodes a large repertoire of pathogenicity-associated genes. A fungal pathogen requires cell wall-degrading enzymes and toxins to enable their pathogenicity and virulence in host plants. The genes encoding cell wall-degrading enzymes such as cellulase, xylanase, catalase, cutinase, glucosidase, extracellular lipase and cellobiohydrolase were found in the WGS data (see Additional file 1 online). About 127 families of glycoside hydrolases (GHs) were reported based on the amino acid sequences in the CAZy database[39]. GHs were reported to be involved in the breakdown of glycosidic bonds between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety[40]. The *C. canescens* genome encodes 44 CAZymes, including 20 GHs from 11 families, 20 carbohydrate esterases (CE), one carbohydrate binding module (CBM), one glycosyl transferase (GT) and two isocitrate lyases (ICL). Notably, there is a comparative number of GHs in

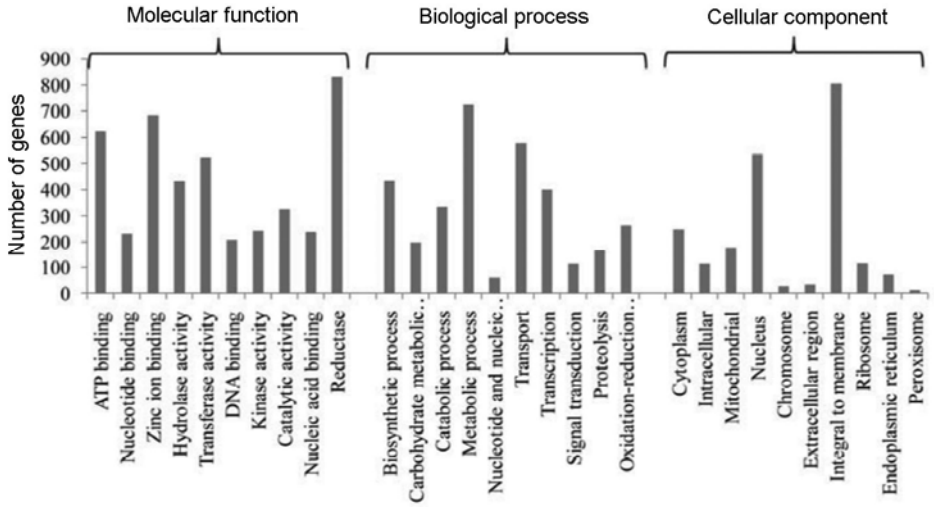

Figure 3. Functional annotations of *C. canescens* proteins in different categories i.e. biological processes, molecular functions and cellular components.

C. canescens and other plant pathogenic fungi (Table 2). There are three endoglucanases (GH5, GH17, GH61-putative), one cellobiohydrolases (GH7), 23 β-glucosidases (GH3, GH5) and nine α-glucosidases (GH 13, GH 31, GH 71, CBM20) for the hydrolysis of cellulose. Six CE5 candidate cutinases have been found in the whole genome, indicating that these are critical for determining initial entry.

Signal transduction

Signal transduction is an important part of fungal development, pathogenicity and virulence in many plant pathogenic fungi[41]. The pathogenicity related signalling pathways are essential for relaying information through cell surface receptors[42]. In *M. grisea* appressorium, differentiation takes place by the PHI protein PTH11-like GPCR (G-Protein-Coupled Receptor), which has been reported to regulate responses on the plant surface[43]. Similarly, 44 PTH11-like GPCRs were reported in *M. phaseolina* (Tassi) Goid[41], though no putative PTH11 like GPCRs were found in the *C. canescens* genome. A total of 158 protein kinases were found in *C. canescens*, which is greater than the average (131) as reported in other ascomycetes fungi (Table 2). The protein kinases in *C. canescens* may be playing a functional role in PHI and pathogenicity[16]. The identification of CRP1 was accomplished as light is required by *C. zeae-maydis* in order to perceive the stomata and reorient growth during leaf infection. This gene encodes a putative blue light photoreceptor homologous to white collar-1 (WC-1) of *Neurospora crassa*[16]. Surface recognition and appressoria initiation are dependent on adenylate cyclase *MAC1* (ANSM01005037.1). Early research revealed that deletion of the *MAC1* gene blocked appressorium formation in *M. grisea*. The catalytic subunit of cAMP-dependent protein kinase A (ANSM01006797.1) is also required for appressoria formation and penetration. Furthermore, the *C. canescens* genome contains hetero-trimeric G-proteins (ANSM01004170.1), which could also contribute to the pathogenic signalling pathways[44].

Secondary metabolic pathways

A diverse range of secondary metabolites is produced by many plant pathogenic fungi, which are required for pathogenicity such as host non-selective toxins[14]. We identified 25 putative secondary metabolites in *C. canescens* compared to 75 in the *M. phaseolina* genome, 32 in *M. grisea*, 37 in *Botrytis cinerea* Pers., 29 in *Sclerotinia sclerotiorum* (Lib.) de Bary and 37 in *Fusarium graminearum* (Schwein.) Petch. There are 24 genes predicted to encode polyketide synthases (PKS) compared to 23 PKS genes in *M. grisea* and 35 in *M. phaseolina.* In addition, several β ketoacyl synthase, acyl carrier protein, acyltransferase and thioesterases (IPR000794) involved in biosynthesis of polyketide have been identified[45]. Ten non-ribosomal peptide synthetases (NRPS) were found in *C. canescens*, which may catalyse the production of cyclic peptides, including numerous toxins; however in *M. grisea*, only six NRPS genes and eight hybrid PKS-NRPS are present (Table 2).

Transport and detoxification of secondary metabolites

Several detoxification-related homologs such as cytochrome P450 (IPR001128), Cof protein (IPR000150) and Cu/Zn superoxide dismutase (IPR001424) are present in the *C. canescens* genome. The MFS transporter (63) acts

Table 2. Size of selected protein families of *C. canescens* and other fungi

Protein family[a]	CC[b]	MP	FO	FG	MO	BC	SS	NC	AN	AF	PCH	PP
Fungal-specific transcription factors	45	156	101	192	95	118	90	89	209	169	65	63
C_2H_2 zinc finger transcription factors	3	66	73	85	58	48	54	63	58	51	77	42
Zn2/Cys6 transcription factors	20	113	370	376	155	142	108	110	307	230	146	118
Major facilitator superfamily	63	270	352	274	198	225	167	110	279	232	141	184
Cytochrome P450	41	256	178	112	137	129	93	40	116	74	155	236
Pth11-like G-protein coupled receptor	1	44	55	51	60	22	23	28	39	15	14	26
Protein kinases	158	140	160	129	129	124	164	111	127	131	106	56
Histidine kinase	11	1	37	20	6	3	5	8	12	6	19	24
Heterokaryon incompatibility	4	65	82	88	41	59	34	45	7	8	3	2
Serine proteases	0	1	12	60/150[c]	56/91	19/34	20/33	32/74	53/136	29/46	0	2
Subtilisin	6	19	36	16/24	26/29	4/7	4/6	6/10	3/4	3/7	12	33
Trypsin	1	2	3	2/3	3/3	1/1	1/1	0/2	1/2	0/0	0	0
Carboxypeptidase	10	19	31	12/21	7/8	7/9	8/11	6/9	5/12	14/15	24	22
Aspartic protease	0	4	0	15/18	14/19	11/14	9/21	15/19	7/16	7/9	38	18
Threonine protease	0	0	0	3/18	2/18	2/13	2/13	2/20	0/20	1/17	0	0
Cysteine protease	0	3	0	5/57	4/31	3/24	1/27	4/41	6/57	3/31	0	0
Metalloprotease	2	8	26	32/111	38/91	6/50	7/48	21/81	22/105	20/77	0	0
All proteases	19	113	261	354	250	135	142	235	334	180	228	325
Lipase	20	53	61	4/31	2/23	3/28	2/25	0/16	2/27	3/25	23	40
Esterase/thioesterase	6	108	95	70	64	70	58	42	63	52	74	69
Glycoside hydrolase-related	20	219	168	159	198	120	126	137	200	165	180	144
Transposases	0	101	19	17	15	73	426	15	15	109	12	11
Cutinase	6	10	12	12	18	11	8	3	4	5	0	0
Polysaccharide lyase	0	16	23	25	9	25	20	5	24	27	4	6
Secondary metabolite backbone genes	25	75	34	37	32	37	29	15	58	40	51	39

[a]Corresponding UniProtKB/TrEMBL codes are listed in Additional file 1 (see online). [b]Fungal species are CC, *Cercospora canescens*; MP, *Macrophomina phaseolina*; FO, *Fusarium oxysporum;* FG, *Fusarium graminearum*; MO, *Magnaporthe oryzae*; BC, *Botrytis cinerea*; SS, *Sclerotinia sclerotiorum*; NC, *Neurospora crassa*; AN, *Aspergillus nidulans*; AF, *A. fumigatus*; PCH, *Phanerochaete chrysosporium*; PP, *Postia placenta.* Fractions indicate the number of total proteins in each family that are secreted.

to transport cercosporin from the inside of cells to the outside, which helps the fungus to cope with the self-produced toxin[46]. The homologous gene, *CRG1* (ANSM01006634.1) encodes the cercosporin toxin resistance protein, which also defends against self-produced toxin present in *C. canescens*. The *C. canescens* genome encodes 534 transporter genes, and majority of them are similar and catalogued in the PHI-database; however, 41 ABC superfamily members have been found. For example, dehydrogenases (306), monooxygenases (226) and cytochrome P450s (41) were found in *C. canescens*. The *C. canescens* genome contains five glucose and galactoside transporters, whereas the *M. phaseolina* genome has only one *F. graminearum* has 12, suggesting that these genes could be generally important for establishing a plant–fungus relationship (see Additional file 1 online).

Phylogenetic classification

For phylogenetic classification, two important candidate genes (ICL and MAP kinase) were studied to infer an evolutionary relationship among members of *Dothideomycetes* and other related *Ascomycetes*. Eight closely related fungal species were used for a homology search to infer a phylogenetic relationship. Based on the MAP kinase sequence, the phylogenetic classification of the fungal species was made into two major clusters (I and II). *C. canescens* and *C. zeae-maydis*, the two anamorphic species of *Mycosphaerella*, show a close relationship to each other. Five fungal species were grouped together in a major cluster (I) and further divided into two sub-clusters 1.1 and 1.2. *C. zeae-maydis* (jgi Cerzm1116595), *C. canescens* (ANSM01005206.1) and *M. fijiensis* Morelet (jgi Mycfi2211094) were grouped in sub-cluster 1.1, whereas *M. graminicola* (Fuckel) Schröt (jgi Mycgr3 34612) and *Z. tritici* (XP003855965.1) were grouped together into sub-cluster 1.2. In major cluster II, *L. maculans* (Sowerby) Karst (XP0086ZF1.1), *Pyrenophora teres f. teres* Drechsler (XP 003296534.1) and *P. Nodorum* (Müll.) Hedjar (XP001796025.1) were grouped with each other. *C. canescens* (ANSM01005206.1) was closely similar to *C. zeae-maydis* (jgi Cerzm1 116595) with 70% tree topology and both species were closely related to *M. fijiensis* (jgi Mycfi2 211094) with 99% tree topology and falling in cluster I (sub-cluster 1.1). *M. graminicola* (jgi Mycgr334612) and *Z. tritici* (XP 003855965.1) were grouped together with 100% tree topology and falling in cluster I (sub-cluster 1.2). In cluster II, three species, *L. maculans* (XP 0086ZF1.1), *P. teres f. teres* (XP 003296534.1) and *P. nodorum* (XP 001796025.1) were grouped together. *P. teres f. teres* (XP 003296534.1) and *P. nodorum* (XP 001796025.1) depicted 71% tree topology, whereas *L. maculans* (XP 0086ZF1.1) depicted 100% tree topology (Figure 4).

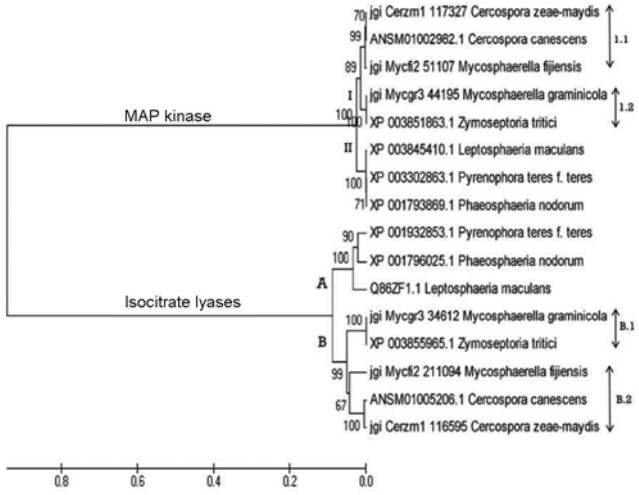

Figure 4. Phylogenetic analyses of eight different species for the two genes, MAP kinase and isocitrate lyase.

Identification of SSRs

SSRs or microsatellites are short repeat sequences of 2–6 bases that are important molecular markers in a wide range of genetic and genomic applications. A total of 528 SSRs were identified in 11,028 contigs. Tri-nucleotide SSRs were the most abundant, accommodating 56.93% of the total identified SSRs, followed by di-nucleotide (15.49%), penta-nucleotide (11.58%), tetra-nucleotide (10.28%) and hexa-nucleotide (5.22%). Pentanucleotide repeats were found with a maximum threshold of seven, whereas di-nucleotide repeats were given a maximum threshold of 27. Tetra-nucleotide and hexa-nucleotide repeats were given a minimum threshold of 3 (ref. 47; see Additional file 3 online).

Availability of supporting data

The supporting datasets to the results of this article are included within the article (with its additional files online) and using the data available at http://www.insilico-genomics.in/vigna/vigcc.html.

1. Raturi, A., Singh, S. K., Sharma, V. and Pathak, R., Stability and environmental indices analyses for yield attributing traits in Indian *Vigna radiata* genotypes under arid conditions. *Asian J. Agric. Sci*., 2012, **4**, 126–133.
2. Project coordinator's report (mungbean and urdbean), All India coordinated research project on MULLaRP, Indian Institute of Pulses Research, Kanpur, 2013–2014.
3. Iqbal, S. M., Jubair, M. and Malik, B. A., Control of cercospora leaf spot of mungbean by foliar fungicides. *Pak. J. Phytopathol*., 1995, **7**, 84–85.
4. Kapadiya, H. J. and Dhruj, I. U., Management of mungbean cercospora leaf spot through fungicides. *Indian Phytopathol*., 1999, **52**(1), 96–97.
5. Sompong, C., Somta, P., Sorjjapinum, W. and Srinives, P., Quantitative trait loci mapping of Cercospora leaf spot resistance in mungbean, *Vigna radiata*, *Mol. Breed*., 2011, **28**, 255–264.
6. Grewal, J. S., Pal, M. and Kulshrestha, D. D., Control of cercospora leaf spot of greengram by spraying bavistin. *Indian J. Agric. Sci*., 1980, **50**, 707–711.
7. Shanmugasundram, S., *Mungbean varietal improvement*, AVRDC, Taiwan, 1987, pp. 1–49.
8. Chand, R., Singh, V., Pal, C., Kumar, P. and Kumar, M., First report of a new pathogenic variant of *Cercospora canescens* on mungbean (*Vigna radiata*) from India. *New Dis. Rep.*, 2012, **26**, 6; doi: org/10.5197/j.2044-0588.2012.026.006.
9. Gupta, S. K., Bansal, R. and Gopalakrishna, T., Development and characterization of genic SSR markers for mungbean (*Vigna radiata* (L.) Wilczek) *Euphytica*, 2014, **195**, 245–258; doi: 10.1007/s10681-013-0993-0.
10. Poehlman, J. M., What we have learned from the international mungbean nurseries. In Proceedings of the First International mungbean Symposium, University of Phillipines, Los Banos, 1977, pp. 97–100.
11. Agrawal, S. C., *Diseases of Greengram and Blackgram*, International Book Distributors, Dehradun, 1993, p. 321.
12. Goode, M. J. and Brown, G. R., Detection and characterization and *Cercospora citrullina* isolates that sporulate readily in culture. *Phytopathology*, 1970, **30**, 1502–1503.
13. Chand, R., Kumar, P., Singh, V. and Pal, C., Technique for spore production in *Cercospora canescens*. *Indian Phytopthol.*, 2013, **66**, 159–163.
14. Daub, M. E., Cercosporin, a photosensitizing toxin from *Cercospora* species *Phytopathology*, 1982, **72**, 370–374.
15. Groenewald, J. Z. *et al.*, Species concepts in *Cercospora*: spotting the weeds among the roses. *Stud. Mycol*., 2013, **75**, 115–170.
16. Chen, H., Lee, M. H., Daub, M. E. and Chung, K. R., Molecular analysis of the cercosporin biosynthetic gene cluster in *Cercospora nicotianae*. *Mol. Microbiol*., 2007, **64**, 755–770.
17. Zhang, G. R., Newman, M. A. and Bradley, C. A., First report of the soybean frogeye leaf spot fungus (*Cercospora sojina*) resistant to quinone outside inhibitor fungicides in North America. *Plant Dis*., 2012, **96**, 767.
18. Karaoglanidis, G. S. and Bardas, G., Control of benzimidazole- and DMI-resistant strains of *Cercospora beticola* with strobilurin fungicides. *Plant Dis*., 2006, **90**, 419–424.
19. Murray, M. G. and Thompson, W. F., Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res*., 1980, **8**, 4325.
20. Stanke, M. and Morgenstern, B., AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*., 2005, **33**, W465–W467.
21. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., 1997, **25**, 3389–3402.
22. Ashburner, M. *et al*., Gene ontology: tool for the unification of biology. *Gene Ontol. Consort. Nature Genet*., 2000, **25**, 25–29.
23. Thiel, T., Michalek, W., Varshney, R. K. and Graner, A., Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet*., 2003, **106**, 411–422.
24. da Maia, L. C., Palmieri, D. A., de Souza, V. Q., Kopp, M. M., de Carvalho, F. I. and Costa de Oliveira A., SSR Locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation,. *Int. J. Plant Genomics*, 2008; doi: 10.1155/2008/412696
25. Koressaar, T. and Remm, M., Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 2007, **23**, 1289–1291.
26. Thompson, J. D., Higgins, D. G. and Gibson, T. J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position- specific gap penalties and weight matrix choice. *Nucleic Acids Res*., 1994, **22**, 4673–4680.

27. Kumar, S., Tamura, K. and Nei, M., MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput. Appl. Biosci*., 1994, **10**, 189–191.
28. Hunter, S. *et al.*, InterPro: the integrative protein signature database. *Nucleic Acids Res*., 2009, **37**, D211–D215.
29. Zerbino, D. R. and Birney, E., Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.*, 2008, **18**, 821–829.
30. Annadurai, R. S. *et al.*, Next generation sequencing and *de novo* transcriptome analysis of *Costus pictus* D. Don, a non-model plant with potent anti-diabetic properties. *BMC Genomics*, 2012, **13**, 663.
31. Dean, R. A. *et al.*, The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*, 2005,**434**, 980–986.
32. Magrane, M. and Consortium, U., UniProt Knowledgebase: a hub of integrated protein data. *Database* (*Oxford*). 2011; doi: 10.1093/database/bar009.
33. H., Xu, X., Wang, Y., Bajpai, V. K., Huang, L., Chen, Y. and Baek, K. H., The mitogen-activated protein kinase signal transduction pathways in *Alternaria* species. *Plant Pathol. J*., 2012, **28**, 227–238.
34. Chen, H. Q., Lee, M. H. and Chung, K. R., Functional characterization of three genes encoding putative oxidoreductases required for cercosporin toxin biosynthesis in the fungus *Cercospora nicotianae*. *Microbiology*, 2007, **153**, 2781–2790.
35. Shim, W. B. and Dunkle, L. D., CZK3, a MAP kinase kinase kinase homolog in *Cercospora zeaemaydis*, regulates cercosporin biosynthesis, fungal development, and pathogenesis. *Mol. Plant Microbe Interact*., 2003, **16**, 760–768.
36. Dixon, K. P., Xu, J. R., Smirnoff, N. and Talbot, N. J., Independent signaling pathways regulate cellular turgor during hyperosmotic stress and appressorium-mediated plant infection by *Magnaporthe grisea*. *Plant Cell*, 1999, **11**, 2045–2058.
37. Beckman, P. M. and Payne, G. A., External growth, penetration and development of *Cercospora zea-maydis* in corn leaves. *Phytopathology*, 1982, **72**, 810–815.
38. Xu, J. R. and Hamer, J. E., MAP kinase and cAMP signaling regulate infection structure formation and pathogenic growth in the rice blast fungus *Magnaporthe grisea*. *Genes Dev*., 1996, **10**, 2696–2706.
39. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B., The Carbohydrate-Active Enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res*., 2009, **37**, D233–D238.
40. Davies, G. and Henrissat, B., Structures and mechanisms of glycosyl hydrolases. *Structure*, 1995, **3**, 853–859.
41. Ma, D. and Li, R., Current understanding of HOG-MAPK pathways in *Aspergillus fumigatus. Mycopathology*, 2013, **175**, 13–23.
42. Islam, M. S. *et al.*, Tools to kill: genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*. *BMC Genomics*, 2012, **13**, 493.
43. Kulkarni, R. D., Thon, M. R., Pan, H. and Dean, R. A., Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea. Genome Biol*., 2005, **6**, R24.
44. Choi, W. and Dean, R. A., The adenylate cyclase gene *MAC1* of *Magnaporthe grisea* controls appressorium formation and other aspects of growth and development. *Plant Cell*, 1997, **9**, 1973–1983.
45. Beck, J., Ripka, S., Siegner, A., Schiltz, E. and Schweizer, E., The multi-functional 6-methylsalicylic acid synthase gene of *Penicillium patulum:* its gene structure relative to that of other polyketide synthases. *Eur. J. Biochem*., 1990, **192**, 487–498.
46. Daub, M. E. and Chung, K.-R., Cercosporin: a phytoactivated toxin in plant disease. APS*net*Features., 2007, doi: 10.1094/APSnetFeature/2007-0207.
47. Karaoglu, H., Lee, C. M. Y. and Meyer, W., Survey of simple sequence repeats in completed fungal genomes. *Mol. Biol. Evol*., 2005, **22**, 639–649.

ACKNOWLEDGEMENTS. We thank the Indian Council of Agricultural Research (ICAR), New Delhi, for providing financial support through the networking project ‘Diagnosis and Management of Leaf Spot Diseases of Field and Horticultural Crops’. The funding agency had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We also thank ICAR for providing financial support through the Out-Reach Programme on ‘Diagnosis and Management of Field and Horticultural Crops’. The authors have no conflict of interests.

Received 23 January 2015; revised accepted 1 August 2015

doi: 10.18520/v109/i11/2103-2110