International Journal of Chemical Studies

P-ISSN: 2349-8528
E-ISSN: 2321-4902
www.chemijournal.com
IJCS 2021; 9(4): 392-401

Received: 27-05-2021
Accepted: 30-06-2021

Manoj V Parakhia
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

Piyush Ukani
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

Corresponding Author:
Manoj V Parakhia
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

Draft genome sequencing and analysis of plant pathogenic fungi *Macrophomina phaseolina* MRf1 isolated from Castor

Manoj V Parakhia and Piyush Ukani

Abstract
Castor is a plant that is commercially very important to the world. Castor seed oil cake is very useful manure to crops. Macrophomina root rot is one of the most destructive diseases of dry land castor crop in India and this disease cause by fungi *Macrophomina phaseolina*. Draft genome sequencing of plant pathogenic fungi *Macrophomina phaseolina* using next generation sequencer that showed the size of genome is 98.6 Mb. The draft genome having 3061 contings, 30756 genes, 183303 exon, 28096 SSR and 13947 repeat region present in the genome. In genome 24.30% of genes involve in molecular function, 34.27% of genes involve in cellular component and 41.43% of genes involve in biological process. pathogenicity related genes identified in this study have high relevance in future fungicide designing and primers will be used for the specific identification of pathogenic fungi *Macrophomina phaseolina*.

Keywords: *Macrophomina phaseolina*, Draft genome, functional genes

Introduction
Castor (*Ricinus communis* L.) is cultivated around the world because of the commercial importance of its oil. Castor is a small annual plant. It ranges from 1 to 7 meters in height. The castor plant appears to have originated in eastern Africa, especially around Ethiopia. It now grows throughout the warm-temperate and tropical regions and flourishes under a variety of climatic conditions. India is the main producer in castor production in the world. The Indian variety of castor has an oil content of 48% and 42% can be extracted, while the cake retains the rest. Castor grows under tropical conditions. A hot and humid climate is required for its production. Castor oil obtained from castor seed is inedible but is of great industrial importance. Castor is a plant that is commercially very important to the world. Castor seed oil cake is very useful manure to crops. It is a very good fertilizer alternative containing optimum levels of Nitrogen, Phosphorous and Potassium which is suitable for cultivation of Paddy, Wheat, Maize and Sugarcane. Castor oil and its derivatives find major application in soaps, lubricants, grease, hydraulic brake fluids and polymers and perfumery products.

Castor is cultivated in 30 different countries on commercial scale - of which India, China and Brazil are the major castor growing countries - accounting for 90% of the world's production. Historically, Brazil, China and India have been the key producing countries meeting global requirement India is the leader in global Castor production and dominates in international Castor oil trade. Gujarat is the chief producing state, having a share of 75% of domestic production, followed by Rajasthan and Andhra Pradesh. A bag packed of castor seed contains 75 kg by weight.

India is the world's largest producer of castor contributing to around 85% of world's total production and dominating the global trade with a share of more than 9% from the country. India produces around 10 lakh tons of Castor seed and around 5.5 lakh tons of Castor oil.

Macrophomina root rot is one of the most destructive diseases of dry land castor crop in India. A long dry spell with high temperature favours this disease. Maiti and Raoof (1984) [9] reported severe aerial infection of *Macrophomina phaseolina* causing die-back symptoms in castor. The fungal infection and disease development are favoured by high temperatures and drought (Dhingra and Sinclair 1978; Grezes-Besset *et al*. 1996; Gupta *et al*. 2012) [5, 7, 8] The disease usually starts before maturation of the primary spikes as spike blight as a result of aerial infections; the stem and twig blights follow. By this time, the crop experiences two types of moisture stress - a drought spell in September-October and moisture depletion

physiologically due to spike maturation. Both the moisture depletions are unfavourable for plant growth but favourable for the pathogen and the disease develops. Based on the major symptoms, the disease is called stem and twig blight.

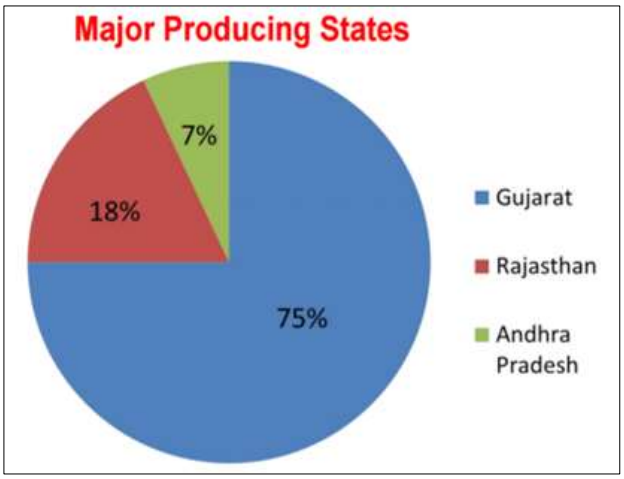


Fig: Major Producing States

M. phaseolina (Tassi) Goid. (syns. *M.* phaseolina (Maubl.) Ashby, *Rhizoctonia bataticola* (Taub.) Briton-Jones, Sclerotium bataticola Taub., and *Botryodiplodia phaseoli* (Maubl.) Thrium.), is a soil borne plant pathogen belonging to the phylum Deuteromycetes and class Coelomycetes. It is highly variable, with isolates differing in microsclerotial size and presence or absence of pycnidia. The pycnidial stage is not common on soybean, but is on peanut. Pycnida are initially immersed in host tissue, then erumpent at maturity. They are 100-200 µm in diameter; dark to grayish, becoming black with age; globose or flattened globose; membranous to subcarbonaceous with an inconspicuous or definite truncate ostiole. The pycnida bear simple, rod-shaped conidiophores, 10-15 µm long. Conidia (14-33 x 6-12 µm) are single celled, hyaline, and elliptic or oval.

Colonies in culture range in color from white to brown or gray and darken with age. Hyphal branches generally form at right angles to parent hyphae, but branching is also common at acute angles. Aerial mycelium with completely or partially appressed growth may or may not be produce in culture. Some isolates may form concentric growth rings.

M. phaseolina survives as microsclerotia in the soil and on infected plant debris. The microsclerotia serve as the primary source of inoculum and have been found to persist within the soil up to three years (Dhinga, O. D. and Sinclair, J. B. 1977) [4]. The microsclerotia are black, spherical to oblong structures that are produced in the host tissue and released in to the soil as the infected plant decays. These multi-celled structures allow the persistence of the fungus under adverse conditions such as low soil nutrient levels and temperature above 30 C. Microsclerotial survival is greatly reduced in wet soils surviving no more than 7 to 8 weeks and mycelium no more than 7 days. Seeds may also carry the fungus in the seed coat. Infected seed do not germinate or produce seedlings that die soon after emergence.

Germination of the microsclerotia occurs throughout the growing season when temperatures are between 28 and 35 C. Microsclerotia germinate on the root surface, germ tubes form appresoria that penetrate the host epidermal cell walls by mechanical pressure and enzymatic digestion or through natural openings (Bowers, G. R). The hyphae grow first intercellularly in the cortex and then intracellulary through the xylem colonizing the vascular tissue. Once in the vascular tissue *M. phaseolina* spreads through the taproot and lower stem of the plant producing microsclerotia that plug the vessels (Wyllie, T. D. 1988) [11]. The rate of infection increases with higher soil temperatures and low soil moisture will further enhance disease severity.

Hot, dry weather promotes infection and development of charcoal rot (Wyllie, T. D. 19). In soybean charcoal rot is a greater problem after anthesis and often occurs when the plant is under drought stress (Dhinga, O. D. and Sinclair, J. B. 1977) [4]. *M. phaseolina* can grow and produce large amounts of microsclerotia under relatively low water potentials allowing this disease to be recognized as favoring drought (Olaya, G. and Abawi, G. S. 1996) [10]. The mechanical plugging of the xylem vessels by microsclerotia, toxin production, enzymatic action, and mechanical pressure during penetration lead to disease development. The population of *M. phaseolina* in soil will increase when susceptible hosts are cropped in successive years and can be redistributed by tillage practices (Wyllie, T. D. 19).

Genome sequencing can reveal role of each individual gene and their networks responsible for plant pathogen interaction, growth, evolutionary relationship and genes for pathogenicity. Whole genome sequencing of *M. phaseolina* is imperative not only to study the host-pathogen (HP) interaction but such knowledge discovery may lead to more effective disease combating strategy. Annotated genes/ predicted proteins can be directly used as new targets in fungicides development using computational approach (Fernandez Acero, F. J. *et al.* 2011) [6]

Materials and Methods

DNA Isolation: *Macrophomina phaseolina* MRf1 was isolated from a root of infected castor from Castor farm, Oilseeds research station, Junagadh Agricultural University, Junagadh, Gujarat, India (21.51° N, 70.45° E). Potato Dextrose Agar (PDA) medium was used for growth and maintenance of isolates. These cultures were stored at 4 °C for long-term storage. *Macrophomina phaseolina* MRf1 strains were grown in PDA broth medium for 6-days at 200 rpm at 28 °C. Mycelia were filtered and genomic DNA was extracted using Hipur ATM Fungal DNA purification kit (Himedia) as per manufacturer's protocol.

DNA library construction and Ion torrent PGM sequencing: Library construction was conducted as per the Ion plus fragment library kit (Invitrogen) for whole genome libraries. Total genomic DNA input was 100 ng which was fragmented using Ion shearTM enzyme mix II enzyme with an average of 400 bp DNA fragment sizes. The fragmented genomic library was cleaned using Agencourt Ampure XP Reagent (Beckman Coulter). The fragmented DNA was quantified using the Qubit DSDNA HsAssay Kit with the Qubit Fluorometer (Invitrogen). Purified DNA fragments were ligated with adapters specific to cleavage site of endonuclease enzyme, followed by the size selection of genomic library using E-gel 2% in order to get fragment size of 350–400 bp. The library was amplified using 10 cycles of PCR for enrichment of adapter ligated fragments and purified with Agencourt Ampure XP Reagent (Beckman Coulter). Template preparation for sequencing was conducted according to the OneTouch Ion™ Template Kit (Life Technologies). The selected PCR products were again used for emulsion PCR, followed by positive bead recovery. Ion

Torrent sequencing was conducted using the Ion PGMTM 400 Sequencing Kit (Life Technologies) on an Ion Torrent Personal Genome Machine (PGMTM, Life Technologies) using a 318V2-chip (Ion 318TM chip, Life Technologies).

Data pre-processing and genome assembly

Reads were processed using Fastx toolkit v0.013 (http://hannonlab.cshl.edu/fastx_toolkit/) and bases having phred score (Q) less than 20 were trimmed. High quality processed reads were used for estimation of genome size using kmergenie v1.69972 tool (Chikhi and Medvedev, 2014) [3]. *De novo* assembly were performed using MIRA v4.0.247 (Chevreux B, *et al*, 2004) [2] assembler followed by CAP3 software to obtain final draft genome assembly of *Macrophomina phaseolina*.

Gene prediction and annotation

Repeat Masker (http://www.repeatmasker.org/) was used to mask the repeats in contigs by selecting fungi as species and rmblastn as the search engine with slow search option.

To search the genes in draft genome, Genemark-ES suite v4.2149 was used with following parameters –ES (self-training to predict genes); –fungus; –min-contig size: 500; –max-intron size: 3000; –min_gene_prediction:120. Cufflinks was used for extracting coding and exon sequence in fasta format using –J option50. Presence of full length genes from the assembly was confirmed by using ORFfinder tool51. CEGMA v2.552 was used to assess the completeness of the genome assembly and functional annotation of identified genes was done using Blast2GO Pro Ver. 4.0.7 software53. To access the domains in the genome of *Macrophomina phaseolina MRf1*, the assembly was analysed against Pfam54 database at e-value cut-off 1e-05. The assembled genome was further used for searching genes involved in secondary metabolite biosynthesis using antiSMASH webserver 55.

Secretory protein prediction and its analysis

All predicted proteins of *Macrophomina phaseolina* MRf1 was used as input for identification and analysis of secretory proteins. SignalP v4.156 was used for the prediction of the signal peptide and those lacking the signal peptide were analyzed by SecretomeP v1.02757. Further, the output obtained from both tools were analysed using TargetP v1.158. Subsequently, transmembrane domains and GPI (glycosylphosphatidyl inositol)-anchor were predicted using TMHMM v2.059 and big-PI Fungal Predictor tool60, respectively.

Proteins having no transmembrane and one transmembrane domain within N-terminal Signal peptides were enlisted. Further, the proteins obtained from both classical and non-classical pipeline were subjected to sub-cellular localization prediction using ProtComp v9.0 based on LocDB and PotLocDB databases (http://www. softberry.com). Finally, GPI-anchor proteins were filtered out from the final set and rest of the sequences were the designated as secretome.

Identification of the pathogenicity related genes

PHI-base (a database of Host-Pathogen gene interactions) 61 was used to search the pathogenicity related gene in query sequence using Blastp. We also used dbCAN webserver62 with default parameter to search the genes for their function as carbohydrate active enzyme (CAZY). This search was based on CAZY database which classified the hit into different families according to the type of reaction being catalyzed.

Phylogenetic relationships

Whole genome sequence of 13 fungal species belonging to same class, Agaricomycetes were downloaded from NCBI (http://ncbi.nlm.nih.gov/). Phylogenetic tree was constructed using RealPhy software63 and visualized using the Drawgram module of Phylip tool v3.69564. This tool takes number of genomic sequences as input, uses bowtie2 for read mapping and constructs a phylip file containing the input genomic sequences 65.

Search for Orthologous gene family

The complete proteome of the three fungal species i.e. *A. rolfsii, G.luxuricans, C. subvermispora, M. roreri* were downloaded from NCBI database. Orthovenn server66 was used for identification of shared and unique orthologous gene families clusters in the above mentioned fungal genomes.

Data Deposition

The whole-genome sequence and annotation of *Macrophomina phaseolina MRf1* isolate MR10 have been deposited at NCBI (https://www.ncbi.nlm.nih.gov/) with accession JZWR00000000; BioSample SAMN03388249.

Result and discussion

Morphological and microscopically observation of fungi.

The root of infected plants drying and death starts from apex and progress. Infected capsules discoloured and drop easily. Sudden wilting of plants in patches under high moisture stress coupled with high soil temperature. Plant exhibit symptoms of drought and drooping of leaves. At ground level black lesions are formed on the stem. Young leaves curl inwards with black margins and drop off later, such branches Die-back. Entire branch and top of the plant withers.

Root rot fungal pathogen showed Colonies in culture range in color from white to brown or gray and darken with age. Hyphal branches generally form at right angles to parent hyphae, but branching is also common at acute angles. Aerial mycelium with completely or partially apprised growth may or may not be produce in culture. Some isolates may form concentric growth rings.

Fungi infected Castor plant's root Scanning Electron microscopy

Fig 2: Morphology and Microscopy images of *M. phaseolina*

B. DNA library construction and Ion torrent PGM sequencing: Library construction was conducted as per the Ion plus fragment library kit for whole genome libraries sequencing was conducted using the Ion PGMTM 400 Sequencing Kit (Life Technologies) on an Ion Torrent Personal Genome Machine (PGMTM, Life Technologies) using a 318V2-chip (Ion 318TM chip, Life Technologies) we got following results.

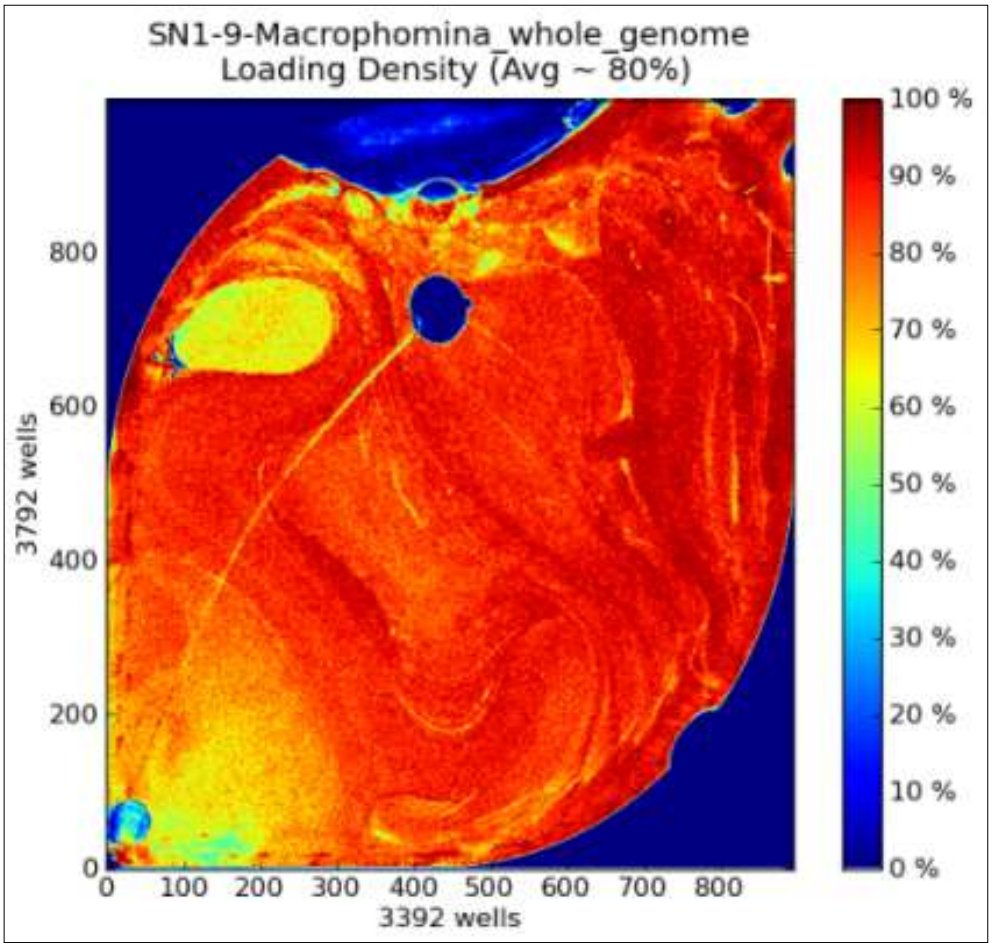


Fig 3: Bead density after the loading of sample

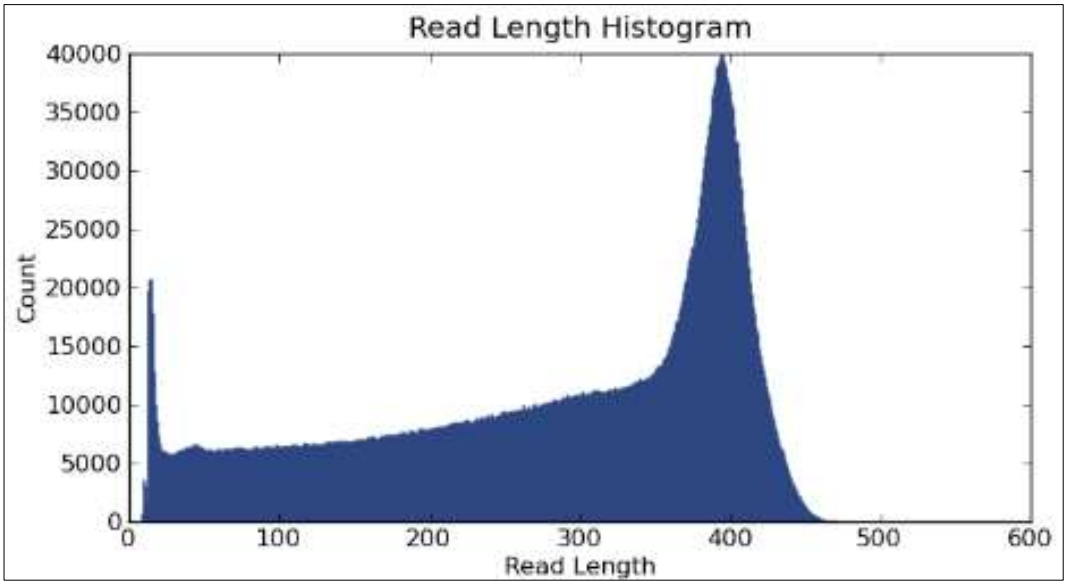


Fig 4: Read length histogram

Table 1: Result of Ion-torrent after sequencing

Total Number of Bases [Mbp]	1,330.66
Number of Q20 Bases [Mbp]	1,137.20
Total Number of Reads	4,792,383
Mean Length [bp]	277
Longest Read [bp]	634

After the analysis of using the CLC workbench software following results were obtained.

Table 3: Genome information after pre-cleaning

Total sequence length	**98,695,791**
Number of contigs	3,061
Contig N50	59,539
Contig L50	541
Genome Coverage	90.0x
GC content	54.70%

Size of Genome

The estimated genome size of ~98.6 Mb was obtained by *k*-mergenie software. An assembly size of 98.6 Mb was obtained using MIRA assembler having 3061 contigs with N_{50} value of 59539 bp and largest contig length being 335485 bp.

D. No. of contings and gene annotation

Genome information of *Macrophomina phaseolina*. CAP3 software.

Table 4: Genome information

Bioproject	PRJNA271369
Biosample	SAMN03274156
Total	98,695,791
Contings	3,061
N50	59, 539
Repeat region	13,947
Gene	30,756
Exon	183,303
Intron	71,469
tRNA	546
mRNA	44,613
ORF	1,273,150
CDS	183,303
SSR	28,096

Gene prediction and annotation

Genemark-ES suite and Augustu used for gene prediction and we got result in a total of 30756 genes present in draft genome while m-RNA systhesis genes were 44,613. Repeatmodeler analysis showed the repeat region present in the genome as follow.

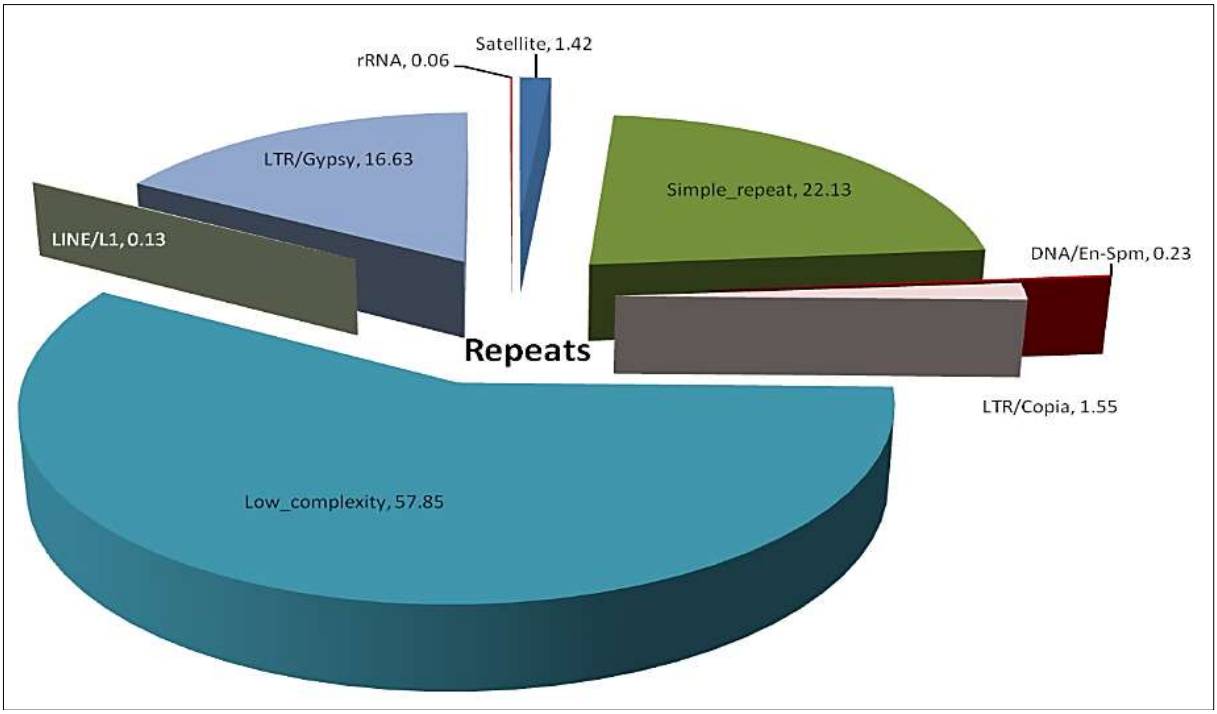


Fig 5: Repeats present in genome (Repeat masker)

Result of tRNA scan

The genome analysis was performed by using tRNA scan for the identification of genes for t-RNA, and identified as the 368 tRNA present in the genome of *M. phaseolina*

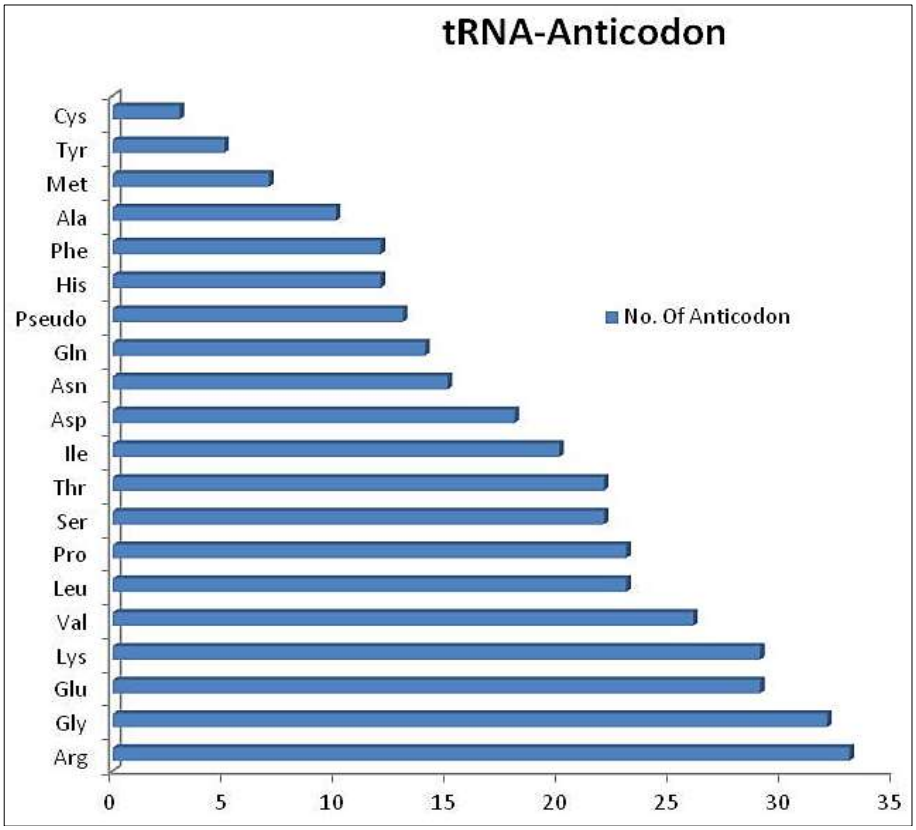


Fig 6: No. of tRNA anticodon present in genome

Function of Genes

GO level annotation assigned the genes into three groups, *viz*., molecular function, cellular component and biological process. Based on molecular function, 12.28% and 3% of genes were annotated for catalytic and transporter activities, respectively Figure 6).

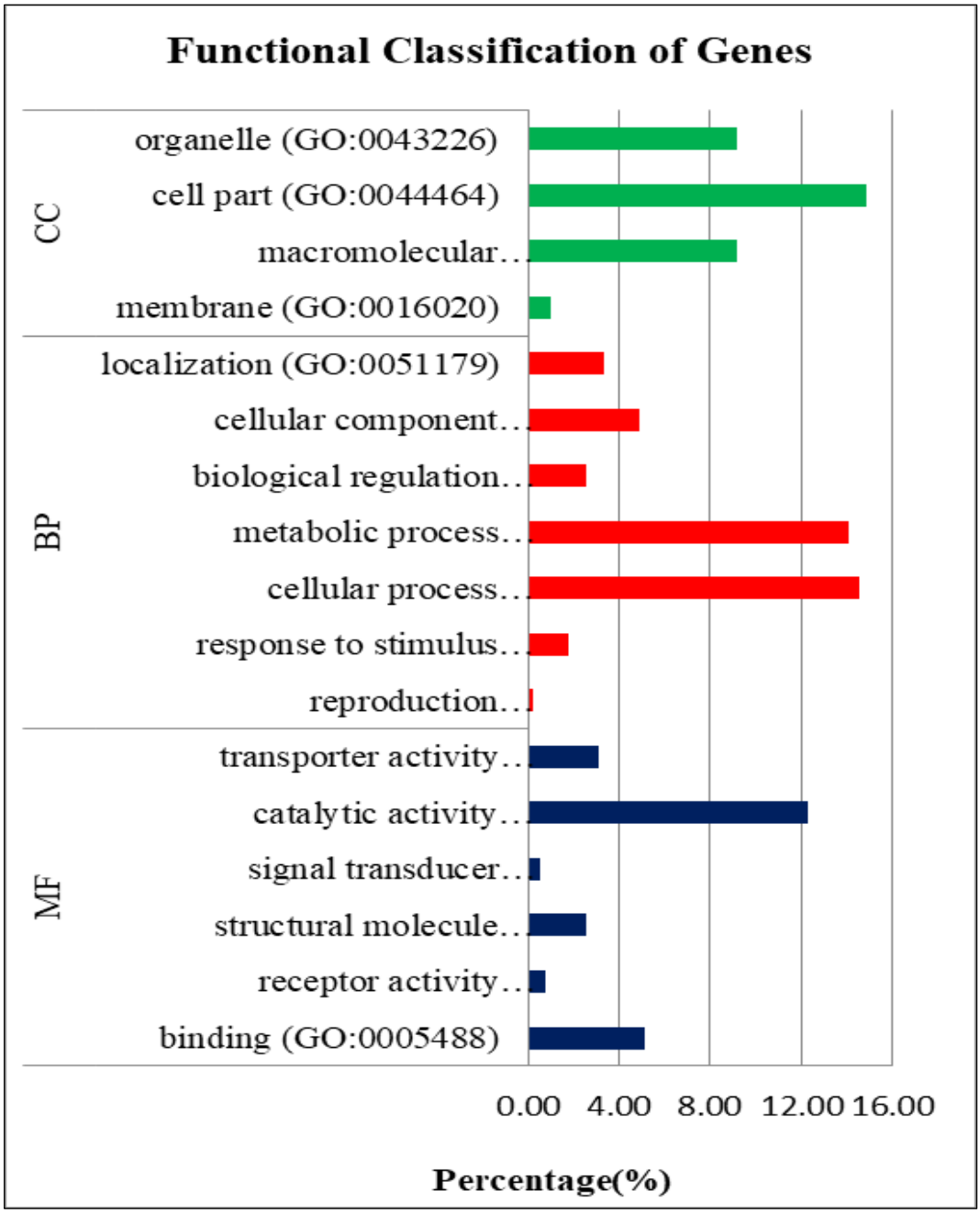


Fig 7: Function of genes

Similarly, biological process classified 41.43% genes involved in regulation and 1.79% and 14% genes in stimulus and metaboloc process, respectively. However, cellular location assigned 3.32% while genes belonged to organelle 9.21% and 1.02% genes were from the membrane. Fungi are the rich source of secondary metabolites performing many biological functions. These metabolites are directly or indirectly affecting the growth of fungus species and mediation of pathogenicity. Our search for the gene involved in secondary metabolite biosynthesis resulted in the identification of two gene clusters, *viz*., terpene cyclases (TCs) and non-ribosomal peptides synthases (NRPSs). Both ascomycota and basidiomycota are known to produce various terpenoid compounds as secondary metabolites. Fungal terpenoids play role as toxicant and also acts for defensive purpose in pathogenesis.

Protein Class analysis: The protein class identified using the panther database available online and predict the following result

Gene family analysis: *Macrophomina phaseolina* protein sequences with those of five other plant pathogenic species including the core *Sclerotina sclerotium, Aspergillus flavus, Fusarium oxysporum, Phaeosphaeria nodorum, Puccinia graminis*. In total, 2,07,057 sequences from the six species were grouped into 12,455 clusters. Of these, 2708 clusters contained at least two species (Figure 8). There were 1990 orthologous gene clusters shared across all six species indicating their conservation within all selected species, while 289 clusters containing 2280 genes were specific to *Macrophomina*. *Macrophomina* and Puccinia, *Sclerotina, Phaeosphaeria Fusarium,and Aspergillus* shared 15, 64, 321, 229, 83 gene clusters respectively, which indicates highest gene clusters shares with *Phaeosphaeria nodorum* and lowest with *Puccinia graminis*

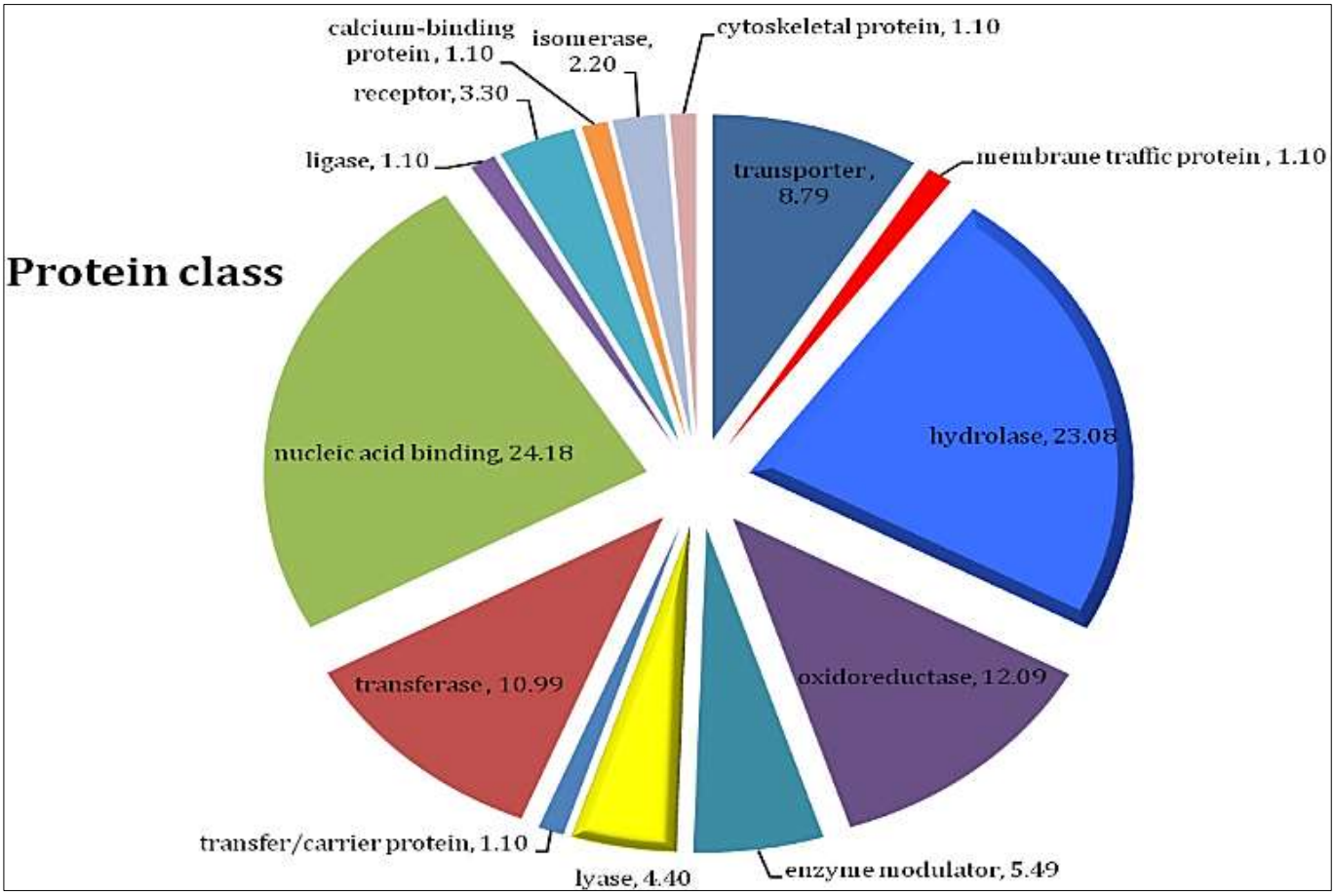
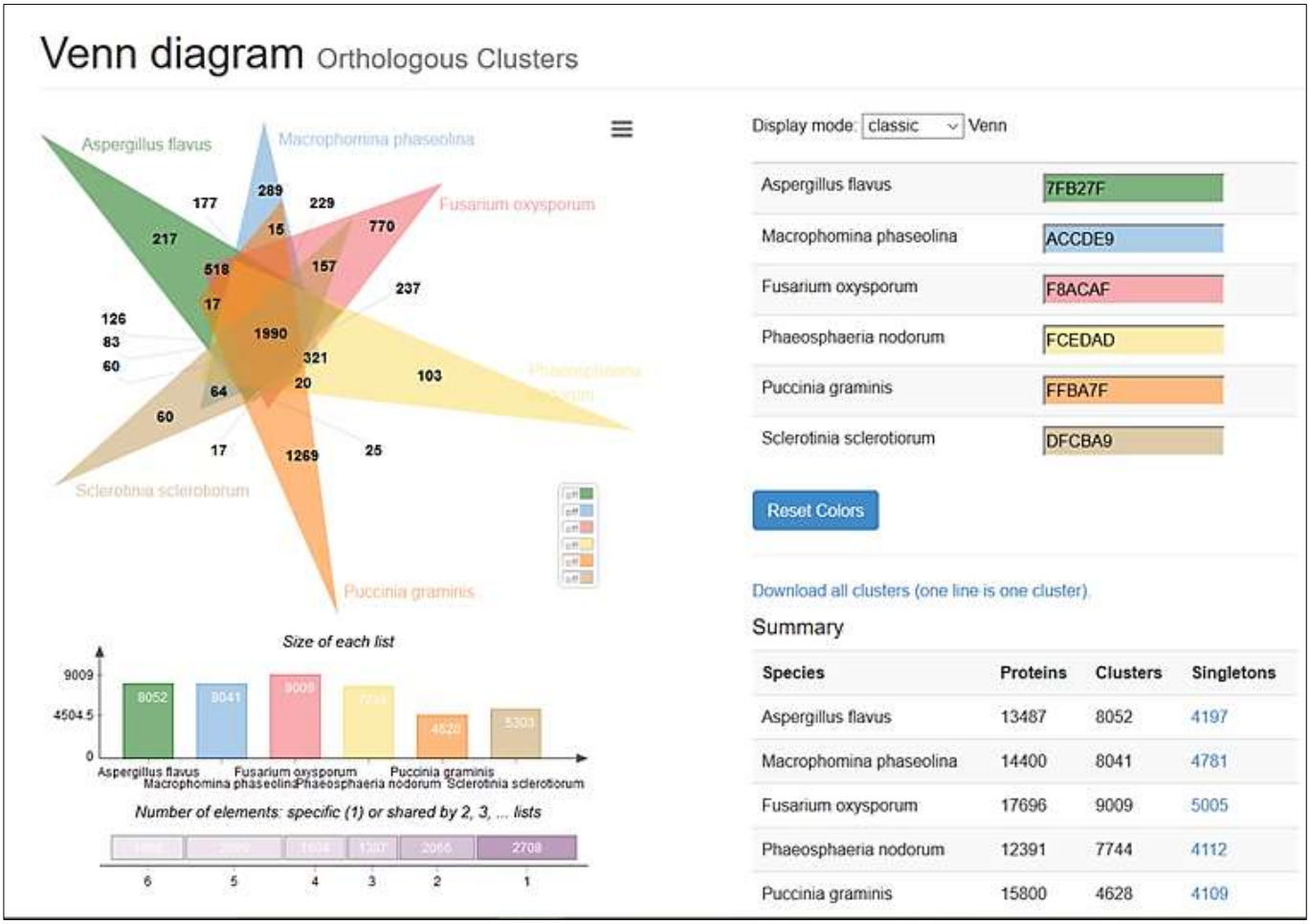


Fig 8: Class of protein present in genome



Species	Proteins	Clusters	Singletons
Aspergillus flavus	13487	8052	4197
Macrophomina phaseolina	14400	8041	4781
Fusarium oxysporum	17696	9009	5005
Phaeosphaeria nodorum	12391	7744	4112
Puccinia graminis	15800	4628	4109

Fig 9: Comparative gene family with other phytopathogenic fungi

SSR Mining

SSR were mine by using Simple sequence repeat finder its showed that total 609 SSR present in the genome which covers di- to hexa-repeats.

Pathogenicity gene analysis

Complete proteome of *M. phaseolina* was aligned to PHI database to reveal the pathogenicity related proteins. We observed a total of 3085(27.99%) PHI genes were classified into different classes such as chemical susceptibility, virulence, pathogenicity, effector, lethal and mixed. As shown in Figure 10, we observed 39 genes associated with chemical susceptibility having 11 resistant and 21 sensitive genes. Furthermore, 602 genes belonged to lethal and 480 to increased virulence class. Remaining three classes *viz*., pathogenicity, effector and mixed were having 1046, 567 and 80 genes, respectively.

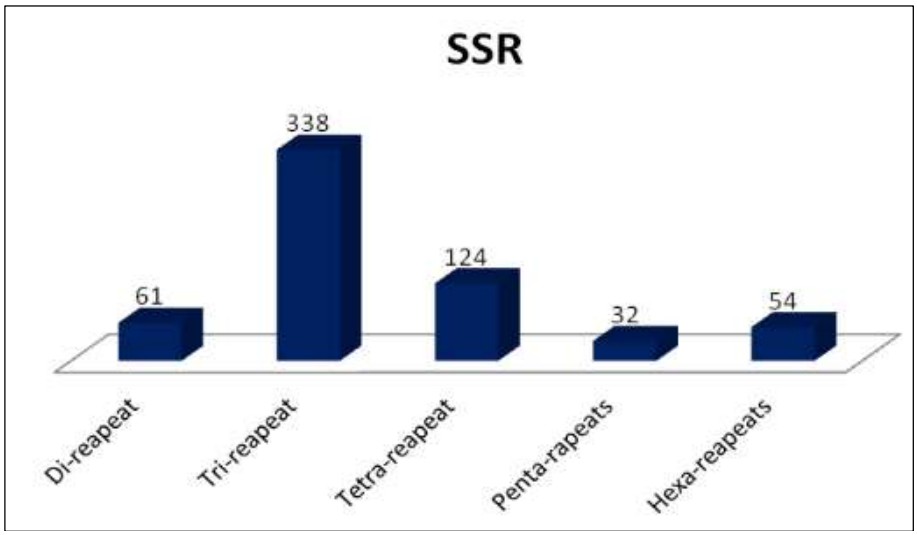


Fig 10: of SSR present in genome (di-to hexa repeats)

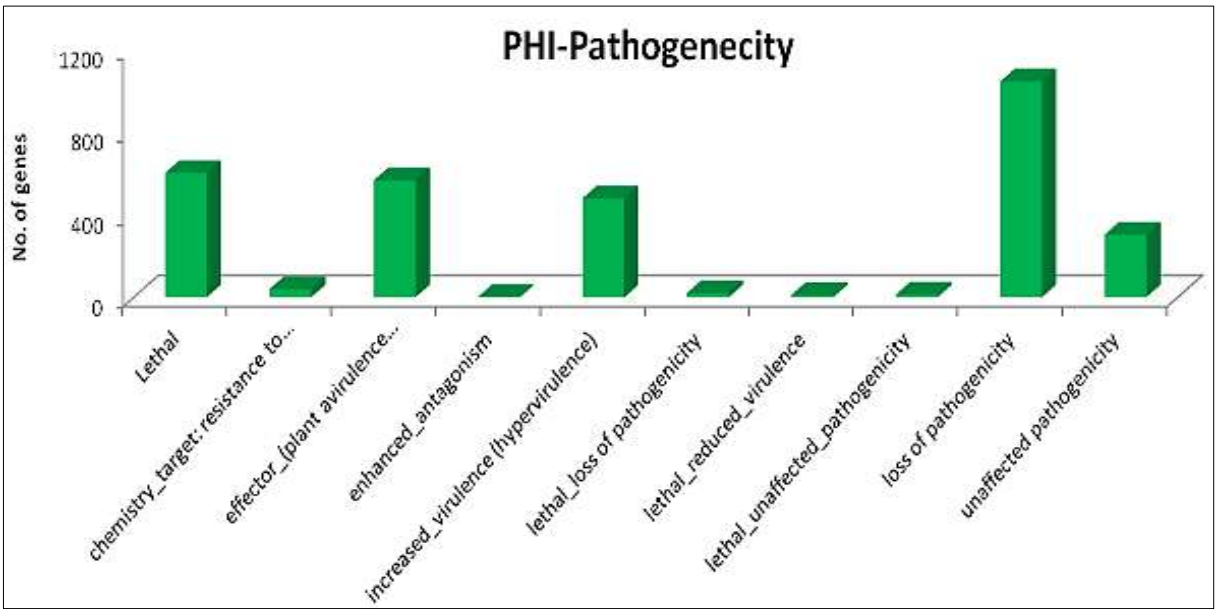


Fig 11: PHI-Pathogenecity

Table 5: Genes related to pathogenicity compare with PHI-database

Attributes	No. of Genes
Lethal	602
Chemistry_target: resistance to chemical	39
Effector_(plant avirulence determinant)	567
Enhanced_antagonism	3
Increased_virulence (hypervirulence)	480
Lethal_loss of pathogenicity	21
Lethal_reduced_virulence	11
Lethal_unaffected_pathogenicity	13
Loss of pathogenicity	1046
Unaffected pathogenicity	303
	3085

Conclusion

It is recommended to the scientific community involved in castor improvement that the sequencing of plant pathogenic fungi *Macrophomina phaseolina* showed the size of genome is 98.6 Mb. The draft genome having 3061 contings, 30756 genes, 183303 exon, 28096 SSR and 13947 repeat region present in the genome. In genome 24.30% of genes involve in molecular function, 34.27% of genes involve in cellular component and 41.43% of genes involve in biological process. pathogenicity related genes identified in this study have high relevance in future fungicide designing and following primers will be used for the specific identification of pathogenic fungi *Macrophomina phaseolina*

References

1. Bowers GR, Russin JS. Soybean disease management. In: Heatherly LG, Hodges HF, editors. Soybean production in the Mid-South. CRC Press; c1999.
2. Chevreux B, Wetter T, Suhai S. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 2004;14:1147–59. doi: 10.1101/gr.1917404.
3. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2014;30:31–37. doi: 10.1093/bioinformatics/btt3.
4. Dhingra OD, Sinclair JB. An annotated bibliography of Macrophomina phaseolina, 1905-1975. Universidade Federal de Viçosa, Minas Gerais, Brazil; c1977.
5. Dhingra OD, Sinclair JB. Biology and pathology of Macrophomina phaseolina. Viçosa/MG, Brasil: UFV, Imprensa Universitária; c1978.
6. Fernandez Acero FJ, Sanzani SM, Di Serio F, *et al*. Development of proteomics-based fungicides: New strategies for environmentally friendly control of fungal plant diseases. Int J Mol Sci. 2011;12:795–816. doi: 10.3390/ijms120100795.
7. Grzes-Besset B, Lucante N, Kelechian V, Dargent R, Muller H. Evaluation of castor bean resistance to sclerotial wilt disease caused by Macrophomina phaseolina. Plant Dis. 1996;80:842–846. doi: 10.1094/PD-80-0842.
8. Gupta GK, Sharma SK, Ramteke R. Biology, epidemiology, and management of the pathogenic fungus Macrophomina phaseolina (Tassi) Goid with special reference to charcoal rot of soybean (Glycine max (L.) Merrill). J Phytopathol. 2012;160:167–180. doi: 10.1111/j.1439-0434.2011.01827.x.
9. Maiti S, Raoof MA. Aerial infection of Macrophomina phaseolina on castor. J Oilseeds Res. 1984;1:232–233.
10. Olaya G, Abawi GS. Effect of water potential on mycelial growth and on production and germination of sclerotia of Macrophomina phaseolina. Plant Dis. 1996;80:1347–1350. doi: 10.1094/PD-80-1347.
11. Wyllie TD. Charcoal rot of soybean—current status. In: Wyllie TD, Scott DH, editors. Soybean Diseases of the North Central Region. APS Press; c1988.