

Whole Genome Sequencing and Expression Analysis of *ToxA* in *Bipolaris Sorokiniana* Provides Discernment of Pathogenicity Causing Spot Blotch of Wheat

Rashmi Aggarwal (✉ rashmi.aggarwal2@gmail.com)

ICAR-Indian Agricultural Research Institute

Shweta Agarwal

ICAR-Indian Agricultural Research Institute

Sapna Sharma

ICAR-Indian Agricultural Research Institute

Malkhan Singh Gurjar

ICAR-Indian Agricultural Research Institute

Bishnu Maya Bashyal

ICAR-Indian Agricultural Research Institute

Atmakuri Ramakrishna Rao

ICAR- Indian Council of Agricultural Research

Sarika Sahu

ICAR- IASRI

Prachi Jain

ICAR-Indian Agricultural Research Institute

Mahender Singh Saharan

ICAR-Indian Agricultural Research Institute

Research Article

Keywords: *Bipolaris sorokiniana*, *ToxA*, *Tsn1*, whole genome sequence, qPCR, spot blotch, wheat

Posted Date: August 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-677095/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Spot blotch disease of wheat caused by *Bipolaris sorokiniana* Boerma (Sacc.) is an emerging problem in South Asian countries. In this study, whole genome of highly virulent isolate of *Bipolaris sorokiniana* (BS112) was sequenced, pathogenicity related gene(s) were identified and role of *ToxA* gene in spot blotch disease development was established.

Results: *Bipolaris sorokiniana* isolate BS112 infecting wheat was sequenced using hybrid assembly approach. The assembly size of the genome was 35.64Mb (GenBank accession number RCTM000000000) with GC content of 50.2%, providing coverage of 97.6% on reference ND90Pr genome. Average gene density predicted was 250-300 genes/Mb. A total of 235 scaffolds were obtained using pyScaf assembler with N_{50} of 16,54,800 bp. In addition, 152 transcription factors involved in various biological processes were identified and a total of 682 secretory proteins were predicted using secretome analysis. *ToxA* gene (535bp) was analyzed and identified in the genome of *B. sorokiniana* which revealed 100% homology with *ToxA* gene of *Pyrenophora tritici repentis*. Further, *ToxA* gene was amplified, sequenced and validated in the 39 isolates of *B. sorokiniana* which confirmed the presence of *ToxA* gene in all the isolates of *B. sorokiniana*. All these *ToxA* sequences were submitted in NCBI database (MN601358-MN601396). As *ToxA* gene interacts with *Tsn1* gene of host, 13 wheat genotypes were evaluated for the *Tsn1* gene and five genotypes (38.4%) were found to be *Tsn1* positive with more severe necrotic lesions compared to *Tsn1* negative wheat genotypes. *In vitro* expression analysis of *ToxA* gene in *B. sorokiniana* isolate (BS112) using qPCR revealed maximum upregulation (14.67 fold) at 1st day after inoculation (DAI). Further, *in planta* expression analysis of *ToxA* gene in *Tsn1* positive and *Tsn1* negative genotypes, Agra local and Chiriya 7 respectively was also conducted. Results revealed maximum expression (7.89 fold) of *ToxA* gene in *Tsn1* positive genotype, Agra local at 5th DAI compared to *Tsn1* negative genotype Chiriya 7 which showed minimum expression (0.048 fold) at 5th DAI.

Conclusions: Full genome of *B. sorokiniana* was sequenced; secreted proteins and virulence genes were identified in the genome. *ToxA* gene was validated in thirty nine isolates of *B. sorokiniana*. *In planta ToxA-Tsn1* interaction studies established that spot blotch disease is more severe in *Tsn1* positive genotypes. This genomic resource will provide a new insight into better understanding and management of spot blotch disease and *B. sorokiniana* of wheat.

Background

Wheat is considered as staple and a leading source of a protein in human food. Wheat production throughout North and South Asia is affected by several abiotic and biotic stresses. Among biotic stresses, the important wheat disease throughout North and South Asia is spot blotch, caused by *Bipolaris sorokiniana*. Various other diseases are also caused by this fungus like foliar blight, common root rot, black point and seedling blight of barley and wheat [1, 2]. However, spot blotch of wheat is majorly the most important diseases which is prevalent in high temperature and humidity areas [3]. Almost 9 million ha of land is affected by spot blotch in India [4] causing 14.9% (approx.) of yield loss [5, 6]. *B. sorokiniana*

is an aggressive fungal pathogen that causes small, dark brown colored lesions on leaves with the length of 1–2 mm without chlorotic margin. These lesions become dark brown in 4 to 5 days and extend to several centimeters in susceptible genotypes inducing leaf abscission. Environmental factors contribute an important role in disease severity. In India, late sown wheat crop is highly vulnerable to this disease due to temperature which ranges between 26–30°C [7].

In the past, it is believed that spot blotch is caused due to compound of three species i.e. *Alternaria triticina*, *Helminthosporium sativum* and *Pyrenophora tritici-repentis* [8]. But later study revealed that it is mainly caused by a hemibiotroph aggressive pathogen *B. sorokiniana*. At the seedling stage this disease appears and increases as plant grows with more severity. As this pathogen infects the plant, conidium germination was observed on the leaf surfaces followed by a germ tube to form a rudimentary appressorium. This infection hyphae either directly penetrate the cuticle and epidermal cell wall or through stomata [9, 10]. Many non-selective toxins and hydrolytic enzymes are produced by *B. sorokiniana* which form lesions on the leaf, leading to leaf abscission [11, 12].

Metabolites produced by *B. sorokiniana* are considered as toxins which play an important role in disease development. During 1990, *ToxA* was discovered in *Parastagonospora nodorum*, causing septoria nodorum blotch (SNB) which horizontally got transferred to *Pyrenophora tritici-repentis* causing tan spot [13, 14]. *ToxA* was reported to be unique to *Pyrenophora tritici-repentis* but not conserved in its genome [15]. Recently, *ToxA* gene was discovered in Australian *B. sorokiniana* isolates [16] and it also exists in the *B. sorokiniana* population in the winter wheat region of the United States [17]. According to Gupta *et al.* [18] effector triggered susceptibility was identified in wheat-*B. sorokiniana* pathosystem where *ToxA* (virulence gene) utilizes *Tsn1* (sensitivity gene) in the host, which helps fungus to cause disease by invading the host. Necrosis was induced on wheat leaves of *ToxA* sensitive wheat genotypes (possessing the *Tsn1* susceptibility gene) [19]. According to McDonald *et al.* [16] also presence of sensitivity gene *Tsn1* in wheat generally helps a *ToxA* positive pathogen to cause spot blotch disease and *ToxA* functions in a *Tsn1* dependent manner. In general, the *ToxA-Tsn1* system is a prototype of an inverse gene-for-gene relationship [20].

Present study was undertaken to analyze the genome, its function and to provide new insights into the genome of *B. sorokiniana* Indian isolate BS112 (collected from BHU, Uttar Pradesh; Accession number: MN601379). Here, three genome sequences of *Bipolaris* species were compared with the genome sequence of *B. sorokiniana* strain BS112. Further, genes involved in carbohydrate-active enzymes, pathogenesis and secondary metabolism have been identified and validated. The genome was also analyzed for *ToxA* and its copy numbers in the genome. Till now six genomes of *B. sorokiniana* were sequenced, including Indian isolate BS112 of *B. sorokiniana*. Recently, Aggarwal *et al.* [21] announced the draft genome of highly virulent strain of *B. sorokiniana* (BS112) (RCTM00000000) with a genome assembly size of 35.64 Mb. However, comparative genomic information along with other *Bipolaris* species is not available. This whole genome *de novo* sequence of Indian wheat isolate of *B. sorokiniana* represents the first genome scale assembly for *B. sorokiniana*. The identification of virulence gene *ToxA*

in wheat isolates of *B. sorokiniana*, led us to further characterize its expression in different Indian wheat genotypes.

Results

Genomic features of *B. sorokiniana*

Whole genome sequencing of *B. sorokiniana* (BS112) was performed using Illumina Hiseq (SRX6673824), Oxford Nanopore Sequencing (SRX6100476) and Ion-Torrent platform (SRX6673825) technologies. For the Illumina library preparation TruSeq DNA PCR free kit was used with an insert size of 350bp and 150bp read size from both the ends. As a result of which 36,915,848 paired end reads were found (Table 1).

Table 1 Illumina Sequencing Statistics	
Sample name	BS 112
Raw read	36915848
Trimmed read	35218897
Chemistry	101*2
Platform	Illumina Hiseq
Library type	Paired end
Total_Raw_Data (Nucleotide size)	3728500648
Total_Trimmed_data (Nucleotide size)	3728500648

Ion Xpress plus fragment library kit was used for fungal DNA shearing enzymatically into suitably sized fragments for Ion torrent library. A target peak of 330bp was selected for adapter ligated library using E-gel seize select 2% agarose gel (Invitrogen Corporation, USA). Finally, Ion plus fragment library kit (Life technologies, USA) was used for adapter ligation, size selected library was amplified and the run was performed on an Ion PGM system version 1. As a result, 8,45,589 reads were obtained. Oxford Nanopore Platform library preparation was done using a ligation sequencing kit (SQK-LSK 108) from Oxford Nanopore Technologies; as a result of which 5,60,150 reads were generated, including 1,14,369 as maximum read length and 171 as minimum read length (Table 2).

Table 2 Nanopore Sequencing Statistics	
Summary	Count
Reads Generated	560150
Maximum Read length	114369
Minimum read length	171
Average read length	4804.4 ± 5450.4
Median Read length	8617.5
Total read length	2691210065

All three types of sequencing reads were incorporated in hybrid genome assembly pipeline using the MaSuRCA tool version 3.2.2 (7) to produce a draft genome assembly of 35.64 Mb with a GC content of 50.2% (Table 3) and the contigs were organized into scaffolds using pyScaf version 3 (https://pypi.org/search/?q_PyScaf_version_3). A total of 235 scaffolds were produced with the N₅₀ scaffolds of 3501266 bp and an average scaffold of 1,51,674.3 ± 4,98,379.7bp (Table 4).

Table 3 Hybrid genome Assembly Statistics using MaSuRCA tool	
Assembler	MaSuRCA
Contigs generated	627
Maximum Contig length	4,93,103
Minimum Contig length	1518
Average contig length	51,819.0 ± 70,246.9
Total Contigs length	3,24,90,483
N ₅₀ value	1,20,114

Table 4	
Scaffolding using pyScaf (final hybrid assembly)	
Assembler	pyScaf
Scaffolds generated	235
Maximum scaffold length	35,01,266bp
Minimum scaffold length	1553bp
Average scaffold length	1,51,674.3 ± 4,98,379.7bp
Total scaffolds length	3,56,43,459bp
Total number of non –ATGC characters	31,79,647
Percentage of non-ATGC characters	8.921
N ₅₀ value	16,54,800

Gene prediction and gene ontology annotation

A total number of 10,460 protein-coding genes were identified with an average gene size of 435-545bp. Total proteins annotated against fungal database was 10,141 and unannotated proteins were 391 (Additional file 1). An average gene density was 250–300 genes/ Mb in the genome of *B. sorokiniana* and 8,506 bp and 35 bp were the maximum and minimum sizes of the genes, respectively. To classify the putative pathogenicity-related genes, 3,627 genes were annotated against the pathogen–host interaction database (PHI database). Based on the homology of pathogenicity proteins, 1,475 genes were related to reduced virulence, 1,281 genes to unaffected pathogenicity, 264 genes to loss of pathogenicity, 174 genes related to lethal and 117 genes related to unaffected pathogenicity reduced virulence (Fig. 1).

In gene ontology (GO) from a total of 2,788 protein-coding genes 1,024 genes were assigned to 'Biological Process' which included 217 genes for transmembrane transport, 187 genes for carbohydrate metabolic process, 182 genes for metabolic process (Fig. 2a). Whereas, 1,274 genes assigned for 'Molecular Function', 432 genes were related to ATP binding, 356 genes to Zinc ion binding, 245 to DNA binding, 204 to oxido reductase activity (Fig. 2b). A total of 1,814 genes were assigned for 'Cellular Component' which included 493 genes for integral component of membrane, 394 for nucleus, 188 for cytoplasm, 93 for intracellular component (Fig. 2c) (Additional file 2).

Phylogenetic inferences

The evolutionary relationship of five species of *Bipolaris*, two isolates of *B. sorokiniana* along with one isolate of *Pyrenophora tritici-repentis* (Accession number MH017419.1) determined through phylogenetic analysis (Fig. 3a) revealed 95–99% similarity in all the species of *Bipolaris* which were grouped in a single major cluster. *B. sorokiniana* query sample (BS112), *B. sorokiniana* (ND90Pr), *B. maydis*

(Accession no. NW_014024938, Strain- ATCC 48331) are grouped together in subcluster 1 showing closer similarity. Whereas, *B. oryzae*, *B. zeicola* and *B. victoriae* (Accession no. GCA_000527765, Strain- F13) were grouped together in subcluster 2.

Comparative genome annotation between the *Bipolaris* species

The orthologous gene families among four *Bipolaris* species including *B. victoriae* (Accession no. GCA_000527765.1, Strain- F13), *B. oryzae* (Accession no. GCA_000523455.1, Strain- ATCC 44560), *B. sorokiniana* and *B. zeicola* (Accession no. GCA_000523435.1, Strain- 26-R-13) were identified using OrthoVenn (Fig. 3b). A total number of 11,938 clusters and 8,370 single copy gene clusters in *B. sorokiniana* genome were identified in four species which revealed that 8,674 gene families are common in *Bipolaris* species. Orthologous cluster study reveals 98 common clusters between *B. sorokiniana* and *B. victoriae*, 115 common clusters between *B. sorokiniana* and *B. oryzae* and 74 common clusters between *B. sorokiniana* and *B. zeicola*. Further, 21 unique clusters were identified in *B. sorokiniana*, 25 in *B. victoriae*, 13 in *B. oryzae* and 5 in *B. zeicola* (Additional file 3).

Identification of SSRs (Simple Sequence Repeats) and SNPs (Single Nucleotide Polymorphism)

A total of 5996 SSRs were identified with the total size of 35.64Mb (Fig. 4). Out of these, 617 compound SSRs were identified. These SSRs were classified as mono, di, tri, tetra, penta, and hexa nucleotide in the genome of *B. sorokiniana*. The most abundant SSR type was tetra nucleotide in *B. sorokiniana* genome with 2,248 in number (63% of all SSRs) (Additional file 4). Total number of variants detected through Sam-tools were 93,122. Additionally, a total of 88,672 SNPs and 4,450 Indels were also identified in the genome (Additional file 5).

Protein domain analysis

Pfam were scanned to the predicted proteins of *B. sorokiniana* protein domain collection. A total of 12,082 proteins domains were estimated from derived protein sequences. 3,357 total protein family was identified from which top Pfam domain "WD domain, G-beta repeat" is involved in transcription factors (422 proteins). Super family MSF domain (203 proteins) is the second top most Pfam domain associated with transportation of solutes involved in chemiosmotic ion gradients. In addition, Ankyrin repeat contains 138 proteins of Pfam domain, is highly involved in diverse set of cellular functions (Fig. 5) (Additional file 6).

Prediction and analysis of the attributes of *B. sorokiniana* secretome

Among 10,460 predicted proteins, 1,078 proteins were annotated as classical secretory proteins by SignalP v4.1 [22] (<http://www.cbs.dtu.dk/services/SignalP/>), while 1,134 proteins were classified as secretory proteins by TargetP v1.11 (<http://www.cbs.dtu.dk/services/TargetP/>). To predict the total secretome SignalP v4.1 and TargetP v1.1 was combined initially, removing all duplicate segments and 1,184 unique proteins having secretory signatures were generated using TMHMM software. Within the gene set, 935 secretory proteins containing zero transmembrane domain (TmHMM0) having 92 highly probable glycosylphosphatidylinositol (GPI) anchor containing sequences and 47 proteins containing one transmembrane domain (TmHMM1) were identified. A total of 982 transmembrane proteins from the protein data set screened through WoLFPSORT using “run WoLFPSORT v. 0.2” for the confirmation of secretory protein, resulted in identification of 682 secretory proteins (Table 5). From this total 682 proteins, 573 were assigned to GO terms in 3 categories namely; molecular function (343), biological process (88) and cellular component (137) (Fig. 6).

Genome	Number of proteins
SignalP v5.0	1078
TargetP	1134
Uniq merged (SignalP v5.0 & targetP)	1184
TMHMMv2.0 with 0 transmembrane	935
TMHMMv2.0 with 1 transmembrane	47
Zero transmembrane domain with GPI	92
WoLFPSORT	682

Using CAZy database, 216 CAZy enzymes were predicted specifically, with GH (Glycoside hydrolases) families being the most copious (Fig. 7). Forty five GH families were spotted having 78 total genes, from which 7 families showed presence of more than 2 genes including GH5 having maximum number of genes (6 genes) followed by GH3 (5 genes), GH7 (5 genes), GH43 (5 genes), GH10 (4 genes), GH11 (4 genes) and GH35 (3 genes). Seventy six genes from 18 CBM (Carbohydrate Binding Modules) family were identified, from which 8 CBM families' showed more than 2 genes including CBM1 (30 genes), CBM2 (8 genes), CBM18 (8 genes), CBM13 (5 genes), CBM6 (3 genes), CBM32 (3 genes), CBM35 (3 genes) and CBM60 (3 genes). Sixteen genes were identified from 6 CE (Carbohydrate esterase) families, from which maximum number of genes were identified for CE5 (5 genes) followed by CE1 and CE4 3 genes each. Further, 41 genes from 10 AA (Auxiliary activities) family were identified, from which maximum number of genes were present in AA3 (11 genes) followed by AA9 (10 genes), AA1 (5 genes), AA5 (4 genes) and AA1 (3 genes). Further, 9 genes were identified from 4 families of PL (Polysaccharide

lyases), which includes maximum number of genes in PL3 (3 genes) and PL4 (3 genes). From all of them, CBM1, GH5 and AA3 families were most ample (Additional file 7).

ToxA gene analysis in the genome of *B. sorokiniana*

Screening of *ToxA* gene in the genome of *B. sorokiniana* (RCTM01000005.1) using BLAST_n revealed three copies of *ToxA* gene in this pathogen which was identified in different scaffolds i. e. (a) scaffold 7180000000228.1880955; start- 1236214, end- 1236802, (b) scaffold00139 99.755; start-408, end-747 and (C) scaffold00139 100.000; start-133, end-246.

ToxA gene sequencing and phylogenetic analysis in Indian population of *B. sorokiniana*

ToxA gene got amplified in thirty nine *B. sorokiniana* isolates collected from different locations in India (Table 6), which confirmed the presence of *ToxA* in all the isolates studied. The coding length of *ToxA* gene was 535bp (excluding intron) which was exactly same in all the *B. sorokiniana* isolates (Fig. 8a). All these *ToxA* sequences were submitted in NCBI database under the accession numbers (MN601358-MN601396) (Table 6).

Table 6
Details of *Bipolaris sorokiniana* isolates from different regions of India

S.No	Agroclimatic zone	Location/state	Centre name	Designated name of isolate	Accession no.
1	North Hills zone	Pantnagar (Uttarakhand)		BS-1	MN601358
2	North Western plains zone	Asandh, panipat (HR)	Asandh - 1	BS-3	MN601359
3	North Eastern plains zone	Faziabad, Ayodhya (UP)	Faizabad-1	BS-4	MN601360
4	North Western plain zone	Karnal (HR)	DWR-1,BS-2	BS-6	MN601361
5	North Hills zone	Almora (Uttarakhand)	Almora UPAS-1	BS-11	MN601362
6	North Western plains zone	Karnal (HR)	Jaipur-6	BS-14	MN601363
7	North Eastern plains zone	Faziabad, Ayodhya (UP)	Faizabad-1	BS-27	MN601364
8	North Eastern plains zone	Faziabad, Ayodhya (UP)	Faizabad-3	BS-29	MN601365
9	North Eastern plains zone	Faziabad, Ayodhya (UP)	Faizabad-4	BS-30	MN601366
10	North Eastern plains zone	Faziabad, Ayodhya (UP)	Faizabad-6	BS-32	MN601367
11	North Eastern plains zone	Faziabad, Ayodhya (UP)	Faizabad-7	BS-50	MN601368
12	North Hills zone	Almora (Uttarakhand)	Almora UPAS-1	BS-54	MN601369
13	Peninsular zone	Mahabaleshwar (MS)	Mahabaleshwar	BS-65	MN601370
14	Peninsular zone	Mahabaleshwar (MS)	Mahabaleshwar - 1a	BS-66	MN601371
15	Peninsular zone	Pune (MS)	IARI,Pune-1	BS-68	MN601372
16	Peninsular zone	Pune (MS)	IARI,Pune-1	BS-69	MN601373
17	Peninsular zone	Dharwad (KN)	Dharwad-1	BS-72	MN601374
18	Southern Hill Zone	Wellington, Tamil nadu	Wellington (Nilgiri-2)	BS-75	MN601375

S.No	Agroclimatic zone	Location/state	Centre name	Designated name of isolate	Accession no.
19	Southern Hill Zone	Wellington, Tamil nadu	K8	BS-88	MN601376
20	Southern Hill Zone	Wellington, Tamil nadu	11 K-7	BS-91	MN601377
21	Southern Hill Zone	Wellington, Tamil nadu	R-3	BS-94	MN601378
22	North Eastern Plains Zone	Varanasi, Uttar pradesh	BHU	BS-112	MN601379
23	North Eastern Plains Zone	Varanasi, Uttar pradesh		BS-124	MN601380
24	North Western plains zone	New Delhi	IARI	BS-125	MN601381
25	North Western plains zone	New Delhi	IARI	BS-126	MN601382
26	North Western plains zone	New Delhi	IARI	BS-127	MN601383
27	North Western plains zone	New Delhi	IARI	BS-128	MN601384
28	North Eastern Plains Zone	Masodha, Uttar pradesh	Masodha	BS-129	MN601385
29	North Eastern Plains Zone	Kanpur, Uttar pradesh	Kanpur	BS-130	MN601386
30	North Western plains zone	New Delhi	IARI	BS-131	MN601387
31	North Western plains zone	Karnal (HR)		BS-132	MN601388
32	Southern Hill Zone	Wellington, Tamil nadu		BS-133	MN601389
33	North Western plains zone	New Delhi	IARI	BS-134	MN601390
34	Southern Hill Zone	Wellington, Tamil nadu		BS-135	MN601391
35	North Western plains zone	New Delhi	IARI	BS-136	MN601392
36	North Western plains zone	New Delhi	IARI	BS-137	MN601393

S.No	Agroclimatic zone	Location/state	Centre name	Designated name of isolate	Accession no.
37	Southern Hill Zone	Wellington, Tamil nadu		BS-138	MN601394
38	North Western plains zone	New Delhi	IARI	BS-139	MN601395
39	North Western plains zone	New Delhi	IARI	BS-140	MN601396

Phylogenetic relationship of *ToxA* gene amongst *B. sorokiniana* (RCTM00000000), *Pyrenophora tritici repentis* (MH017419.1), *Parastagonospora nodorum* (CP022834) and *Phaeosphaeria avenaria* (JX997418) was concluded using the Neighbor-Joining method in MEGA X (Fig. 8b). The nucleotide sequences of *B. sorokiniana ToxA* from India (*BSToxA*) and Australia (*AusBSToxA1* and *AusBSToxA2*) aligned to *Pyrenophora tritici repentis* (*ptr_ToxA*) and *Parastagonospora nodorum* (*Sn4_ToxA*). All the isolates revealed 100% homology with *Pyrenophora tritici repentis* sequence and differing by a single nucleotide from one of the two *BSToxA* haplotype identified in Australia (Fig. 9).

ToxA – Tsn1 gene relationship in wheat genotypes

In the host, the presence of *Tsn1* (*xfcp623*), a sensitivity gene; helps a *ToxA* positive pathogen to cause more severe necrotic lesions on wheat genotypes. To predict allelic status at *Tsn1* (400bp amplicon) locus using PCR amplification, thirteen genotypes of wheat were screened; from which 5 wheat genotypes (WH 542, WL 711, HD 29, PBW 343 and Agra local) (38.4%) were found to be *Tsn1* positive, whereas remaining 8 lacked *Tsn1* and instead carry *tsn1* (Fig. 10a). Further, pathogenicity assay of *B. sorokiniana* (BS112) was tested on these thirteen wheat genotypes to identify correlation between *Tsn1* (presence or absence) and necrotic lesions formation. Average disease index was calculated on the basis of number and length of lesions on the leaf of different wheat genotypes (Fig. 10b). Genotypes viz. WH 542, WL 711, HD 29, PBW 343 and Agra local showed severe necrotic lesions on leaves when inoculated with *B. sorokiniana* while *Tsn1* negative wheat genotype showed no necrotic lesion formation. This confirmed that *ToxA* has susceptible reaction phenotype on wheat genotypes harboring *Tsn1* gene [Additional file 8: Fig. S1].

Relative expression analysis ToxA gene under in vitro and in planta conditions

To evaluate host gene expression during fungal infection, relative qPCR with an *EFN-1 alpha* gene as internal control was performed to determine time-dependent effect of *ToxA* on wheat genotypes. *In vitro* expression analysis of *ToxA* gene in *B. sorokiniana* isolate (BS112) collected at different time intervals, using qPCR revealed maximum upregulation (14.67 fold) at 1st day after inoculation (DAI), followed by 2nd DAI (11.83 fold), then gradually expression decreased at 3, 4 and 5 DAI in minimal basal medium

(Fig. 11a). *In planta* expression analysis *via* inoculated leaf collected at different time intervals from two genotypes was performed using qPCR. The maximum expression (7.89 fold) was observed at 5th DAI in susceptible cultivar (Agra local), while minimum expression (0.048 fold) was found in resistant cultivar (Chiriyā 7) at 5th DAI (Fig. 11b).

Discussion

Wheat is the most vital food-grain in the Northern and North-western parts of India. After China, India is the second chief producer of wheat in the world and accounts for 8.7 per cent of the world's total production of wheat [23]. It is rich in proteins, carbohydrates and vitamins and provides balanced food. High yield loss in wheat is mainly due to many fungal diseases like wheat rusts, spot blotch, head blight etc. Over the next several decades, due to predicted high global population growth rate, it would be a challenge to meet this rocketing demand of wheat. Management of fungal diseases and development of resistant genotypes of wheat is the key component to meet this challenge [24]. Fungal disease management and to understand the zestful nature of genome of fungal pathogen, whole genome sequencing is of great significance. Genome sequencing of fungal pathogens show extensive variations in genome structure, its composition between species and also between isolates of same species. Genomic sequencing is antecedent for many strategies and alternatives for plant defense reactions against fungal pathogens [25].

Using next generation sequencing technology we have produced a draft genome of *B. sorokiniana* and a total sequence assembly size of 35.64Mb was predicted [21], which is very similar to *B. sorokiniana* isolate ND90Pr (34.41Mb) and Shoemaker (34.33 Mb). The genome size of *B. sorokiniana* is notably larger than other species of *Bipolaris* like *B. maydis* (32.93 Mb), *B. victorae* (32.83 Mb), *B. zeicola* (31.26 Mb), *B. oryzae* (31.51 Mb) and *B. papendorfii* (33.4 Mb).

In this approach, pathogen host database systematically compiles pathogenicity genes involved in pathogen host interactions. Disrupted genes was also reflected by PHI database which reveals functional redundancy in fungi. Endopolygalacturonases (disruption by a multigene family in *B. sorokiniana*), a plant cell wall degrading enzymes did not affect virulence because other bare type enzymes are reserved to degrade plant cell wall. To better understand gene functions, 2,788 protein coding genes were assigned to different categories. The top categories included transmembrane transport, carbohydrate metabolism, ATP binding etc. which is mainly involved in carbohydrate mechanism such as gluconeogenesis, glycolysis and citric acid cycle (TCA cycle) [26]. This mechanism can facilitate for the use of various substrates of *B. sorokiniana* exposed to various carbohydrate sources in the environment.

Further, study revealed 138 proteins for ankyrin repeat in Pfam domain, which is highly involved in diverse set of cellular functions. Four Ankyrin repeats were present in *ToxE*, which is responsible for the expression of *Tox2* gene present in *Bipolaris zeicola* (highly virulent on maize genotypes) [27].

To penetrate the rigid barriers of plant cell wall, many types of CAZymes (Carbohydrate active enzymes) were produced by fungi. To identify genes responsible for CAZymes, the *B. sorokiniana* genome was mapped to CAZy database. From total CAZymes, 45 glycoside hydrolases (GH) were identified. The number of GH family in other *Bipolaris* species is also reported higher compared to other CAZymes. Interestingly, the *B. sorokiniana* genome contains a lesser number of GH family (45) compared with that of the reported *Bipolaris* species as *Bipolaris maydis* has 276 families. Our findings revealed that *B. sorokiniana* has 6 carbohydrate esterase (CE) family. The highest number of CE₅ (5 copies) and CE₁ (3 copies) indicated that *B. sorokiniana* produces many plant degrading enzymes containing cutinase and carboxylesterase activity. From 18 carbohydrate binding modules (CBM) family CBM₁ and CBM₁₈ has the highest number of genes indicating chitin binding activity. High number of CBM and GH family may have the ability to degrade lignocellulose (main structural component of plant) collectively [28]. This indicate that *B. sorokiniana* has wide variety of CAZymes which is involved in the deprivation of plant cell wall complex carbohydrates.

In this study, we reported the functional identification of *ToxA* gene in different Indian populations of *B. sorokiniana*. *ToxA* gene in a global population dataset was found to be drastically different depending upon the area from which population was sampled. Earlier, for Australian populations of *P. nodorum* and *P. tritici-repentis*, *ToxA* was present in all the tested isolates [29, 30]. Recently, Navathe *et al.* [20] reported 70% of *ToxA* positive isolates in tested Indian population of *B. sorokiniana*, whereas our results revealed that *ToxA* gene existed (100%) in all the Indian isolates of *B. sorokiniana* taken for the study. In the present study, the gene *ToxA* was PCR amplified in all the 39 isolates of *B. sorokiniana* and the length of the amplicon in all the isolates was uniform (535bp). The sequencing data suggested that, there is a high degree of similarity of *ToxA* gene present in *B. sorokiniana*, *Pyrenophora tritici* and *Phaeosphaeria* species. The similarity between these three species can be best illustrated by horizontal gene transfer [13]. The horizontal gene transfer may be due to hybridization between the various fungal species [16]. The exact sequence match of *B. sorokiniana* (BS112) and other two species reveals a recent global migration of these strains.

The expression of *ToxA* gene investigated using qPCR, measure the level of *B. sorokiniana* in leaf tissue at different time intervals displaying disease symptoms. Gene expression levels were compared by C_T (threshold) values for each runner gene. The stability of gene expression was reflected through coefficient of variation of C_T values, wherein higher the C_T values lower will be the gene expression levels and *vice versa*. Pathogenicity assay and PCR amplification confirmed that the isolate carrying *ToxA* caused severe necrotic lesions on wheat genotypes having the sensitivity gene *Tsn1*. In contrast, it caused less or no lesions to wheat genotypes lacking *Tsn1*. Quantitative gene expression analysis and pathogenicity assay revealed that necrotic symptoms observed on Agra local genotype leaves was due to high expression of *ToxA* gene, which gradually increased and caused leaf abscission, whereas, on chiriya 7 its effect is reverse. *Tsn1* is the part of complex network of resistance / susceptibility to spot blotch. According to Navathe *et al.*, *ToxA-Tsn1* have gene to gene relationship; similarly our study also confirms *ToxA* positive isolates showing more severe necrosis on wheat genotype harboring *Tsn1* gene [20]. Recently, Lei *et al.*

also confirmed the *ToxA-Tsn1* relationship through culture filtrate infiltration method. However, the appearance of necrosis on the wheat leaf is not only induced by *ToxA* but can be through some other toxins present in culture filtrate or crude extract [31]. Study of *ToxA-Tsn1* interaction is very useful for breeding program, which are highly prone to spot blotch. Wheat breeders may avoid the genotype with *Tsn1* allele and retain genotypes with *tsn1* allele.

Conclusions

In this study, detailed analysis of genome of *B. sorokiniana* (BS112) was conducted, which lead to the identification of unique SSRs, protein domains, secretory proteins, establishment of phylogenetic relationship among different *Bipolaris* species and understanding the genes involved in host-pathogen interactions at molecular level. This study, further stated the *ToxA-Tsn1* interaction in Indian wheat genotypes and confirmed that spot blotch is more severe in *Tsn1* positive wheat genotypes.

Methods

Fungal- collection, isolation and DNA extraction

Wheat infected leaves were collected from different region of India. For the isolation of the pathogen, infected leaves were surface sterilized and necrotic lesions were cut into smaller fragments. These fragments were washed with 0.1% sodium hypo chlorite (NaOCl_2) solution, followed by washing with water 2 times. To induce sporulation, these fragments were placed on PDA (potato dextrose agar) under a 12-h photoperiod at room temperature (RT). Several single spores from each of the PDA plates transferred to another PDA plates and allow to grow at 25°C. Meanwhile, these spores are observed in microscope and fungus was identified based on conidia and its morphology [32].

Thirty nine isolates of *B. sorokiniana* were established from infected wheat leaf samples collected from different regions of India. These cultures were grown in potato dextrose broth (PDB) and incubated at 25°C with periodical shaking. After 25 days the mycelial mats (0.2 g) from each isolates was harvested aseptically and snap frozen in liquid nitrogen. Frozen mycelia were subjected to DNA extraction using CTAB method [33]. Then, the quantity and quality were assessed using a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Wilmington, USA) and agarose gel electrophoresis respectively. Genomic DNA of BS112 (ITS NCBI accession- KU201275) was sent to Genotypic Technology (P) Ltd. for genome sequencing using hybrid genome assembly through Illumina, Nano pore and Ion-torrent platforms.

Genome sequencing and hybrid assembly

The genome of *B. sorokiniana* was sequenced by using Illumina Hiseq, Oxford Nanopore sequencing and Ion-Torrent platform technologies. The fungal hybrid assembly is the main aim of the project. The Illumina reads were pre-processed using Trim-galore [34], and Nano pore fast5 data were base called

using Albacore [35]. Hybrid genome assembly was created using these three types of sequenced reads. Kmer analysis will not require in hybrid assembly because the internal long read algorithm was based on overlap layout consensus (OLC). For hybrid assembly MaSuRCA tool [36] is used and output of the hybrid assembly of genome is 35.64Mb. Later using pyScaf, the contigs were processed [37]. The paired end library data was generated using Illumina HiSeq platform.

Gene prediction, annotation and pathway analysis

Genes was predicted in the genome of *B. sorokiniana* using AUGUSTUS annotation tool [38]. A total of 10,460 genes were projected with an average gene length of 435-545 bp on the basis of *B. sorokiniana* ND90Pr as reference genome. Maximum gene length of 8,506 and minimum gene length of 35 was identified.

Functional annotation of genes were performed using BLASTx tool [39]. The predicted proteins were similarity searched against fungal protein database using ncbi-blast-2.2.29+ blastp program with an e-value of 1e-5 for gene ontology and annotation. Around 98% of predicted genes were annotated against the protein database. The proteins were annotated against all Viridi plantae kingdom protein sequences (from Uniprot Protein Database). Further analysis, those proteins with more than 30% identity as cut off were taken. The gene annotation was done by fungal hybrid assembly in biological, chemical and molecular functions. The pathway analysis of the predicted genes was performed through KEGG database [40].

Phylogenetic analysis of *B. sorokiniana* genomes

Based on whole-genome alignment, Progressive Mauve version 20150213 build 0 with default parameters was used to execute phylogenetic analysis [41]. Five species of *Bipolaris* namely *B. sorokiniana*/ *Cochliobolus sativus*, *Bipolaris maydis*/ *Cochliobolus heterostrophus*, *B. oryzae*/ *Cochliobolus miyabeanus*, *B. zeicola*/ *Cochliobolus carbonum* and *B. victoriae*/ *Cochliobolus victoriae*; two isolates of *B. sorokiniana* namely *B. sorokiniana*/ *Cochliobolus sativus* (ND90Pr), *B. sorokiniana*/ *Cochliobolus sativus* query sample (BS112) and one isolate of *Pyrenophora tritici-repentis* (outgroup) were taken for the multiple genome alignment. Through neighbor joining tree, comparative analysis was performed on all six genomes. The guide tree was an output of multiple sequence alignment in Mauve. The phylogenetic trees were constructed using the guide tree in phylogeny.

Comparative genome annotation of orthologous gene families

The orthologous gene families among the *Bipolaris* species GCA_000527765.1, GCA_000523455.1, GCA_000523435.1 were identified using OrthoVenn. Comparison among these species revealed 8674

gene families in common.

SNP (Single Nucleotide Polymorphism) and SSR (Simple Sequence Repeat) analysis

SNP prediction was processed using snpEff tool with Illumina Hiseq data [42]. Alignment was done using minimap2 [43] and variant detection was done using Sam-tools [44]. Through MicroSatellite identification tool (MISA) tool with default parameters simple sequence repeats were identified [45].

Protein family classification

Protein domain analysis was widely done using Pfam database [46]. HMMER 3.0 was used to infer Pfam domains by removing overlapping clans. From derived protein sequences, in total 12082 proteins domain were found to be associated.

Secretome analysis

In the current study, secretome prediction of *B. sorokiniana* genome was analyzed using SignalP v4.1 in combination with TargetP v1.1 for all possible secreted proteins. These secreted proteins then examined for transmembrane (TM) spanning regions using TMHMM (TMHMM v2.0; http://www.cbs.dtu.dk/cgi-bin/nph-sw_request?tmhmm). The proteins located in the predicted N-terminal signal peptide is retained as 0 or 1 transmembrane region. Big-PI (http://mendel.imp.ac.at/gpi/cgi-bin/gpi_pred_fungi.cgi) is used to predict GPI- anchor proteins. Localization of the remaining proteins was predicted using ProtComp by LocDB and PotLocDB databases (ProtComp v8.0; <http://www.softberry.com>). After secretion, location of proteins was estimated using WoLFPSORT, analysis done by “run WolfPsort summary fungi” in the WoLFPSORT V0.2 package (70% specificity and sensitivity) [47]. Blast2GO [48] (<http://www.blast2go.com/b2glaunch>) was used to explore whether any additional gene annotation existed from protein set within the predicted refined secretome. From refined secretome, genes were identified which was further classified into six carbohydrate-active enzymes (CAZy) classes; polysaccharide lyases (PL), glycoside hydrolases (GH), carbohydrate esterases (CE), carbohydrate binding modules (CBM), auxiliary activities (AA) and glucosyl transferases (GT) using CAZy database (<http://www.cazy.org/>).

Primer designing and amplification of *ToxA* gene

ToxA gene was extracted from a genomic region of *Pyrenophora tritici-repentis* isolate PTR-01, complete coding sequence (Accession number- MH017419). The sequence specific primers were designed using Integrated DNA Technologies (IDT) primer designing tool. Amplification of *ToxA* gene from genomic DNA

was ensured using primer pair *ToxA* BS F 5'TCATGCGTTCTATCCTCGTA3' and *ToxA* BS R 5'CTAATTTTCTAGCTGCATTCTCCA3'. The PCR condition include initial denaturation at 95°C for 5 min; followed by 39 cycles of amplification, each cycle with the following plan: denaturation at 94°C for 30s, annealing at 52°C for 30s, and extension at 72°C for 1 min; followed by final extension at 72°C for 10 min. By gel electrophoresis (1.2%), PCR products were visualized. Amplified product was gel purified, sequenced and analyzed against NCBI database. The amplicon sequencing was outsourced by Eurofins Genomics India Pvt. Ltd. (Bengaluru, India). The sequences obtained were aligned using BioEdit Sequence Alignment Editor. Phylogenetic analysis was done using Molecular Evolutionary Genetics Analysis (MEGA-X), amongst plant pathogen *viz.* *B. sorokiniana*, *Pyrenophora tritici repentis*, *Parastagonospora nodorum*, *phaeosphaeria avenaria* species was inferred using the Neighbor-Joining method. In this study, the sequences of *ToxA* used were deposited in NCBI *GenBank*.

***Tsn1* allelic status in Indian wheat genotype and its pathogenicity assay**

ToxA, a necrotropic effector was mapped to wheat chromosome arm 5BL, reveals the inverse gene to gene relationship with a sensitivity gene called as *Tsn1*. *B. sorokiniana* isolate carrying *ToxA* gene causes severe necrotic lesions on wheat genotypes harboring *Tsn1* sensitivity gene. Thirteen wheat genotypes were taken from different regions of India which includes WH 542, WL 711, Agra Local, Suzoe, HD 29, PBW 343, HD 3086, Chiriya 7, HD 30, Milan, HD 2329, Ning 8139, HD 2967. These genotypes are further classified based on presence or absence of *Tsn1*. To test *ToxA-Tsn1* relationship, the collected wheat genotypes was grown in a moist chamber maintaining the temperature of approx. 25°C; at two leaf stage BS112 (positive for *ToxA* gene) was inoculated on leaves. Genomic DNA was isolated from above mentioned wheat genotypes using CTAB method. PCR amplification was carried out using dominant marker *Xfcp623* (F 5'CTATTTCGTAATCGTGCCTTCCG3' and R 5'CCTTCTCTCTCACCGCTATCTCATC3') procured from intron five of *Tsn1* gene at annealing temperature of 57°C. PCR results, amplicon presence infer *Tsn1* (presence) and its absence infer *tsn1* (absence).

Pathogenicity assay were tested on 3 week old plant of thirteen different wheat genotypes as mentioned above. The inocula was prepared from BS112, cultured in sorghum grain under aseptic conditions and after 25 days, the inoculum was crushed in water and filter through muslin cloth having a spore density of 1×10^6 spores/mL, for uniform spraying Tween 20 was added to spore suspension and the inoculum was prepared. Plants were kept in poly house in glass chamber where the temperature was maintained at 25°C, followed by >80% of relative humidity. Spore suspension of BS112 was sprayed on the flag leaf of all genotypes in four plant each. Photographs were taken at 4 days of post-inoculation and disease severity (%) was calculated as per the scale; 0=free of spots; 1= up to 5% leaf area covered with necrotic spots; 2= 6-20% of the leaf area covered; 3= 21-40 % of the leaf area covered; 4= 41-60 % of the leaf area covered; 5= spots inclusion more than 60 % of the leaf area tangled.

Average Disease Index (ADI) was calculated using the formula: $ADI = ((\text{sum of rating of each leaf}) / (\text{total leaf} * 5)) / 100$. ADI was converted into disease responses *viz.* 0 = No infection; 0-10= resistant response (R); 10.1-20 = moderately resistance (MR); 20.1-30 = moderately susceptible (MS); 30.1- 50= susceptible(S); and more than 50= highly susceptible (HS) [49, 50].

Expression analysis of *ToxA* under *in vitro* and *in planta* conditions

For *in planta* gene expression analysis using qPCR, BS112 (most aggressive isolate) was inoculated to highly resistant cultivar (Chiriya 7) and susceptible cultivar (Agra local) of wheat. The pots were kept in moist chamber maintaining the temperature of approx. 25°C. After inoculation of BS112 (as described earlier), leaf samples with disease symptoms (photographs were taken) were collected at 1st, 2nd, 3rd, 4th and 5th day, kept in aluminum foil and immediately frozen in liquid nitrogen. *In vitro* expression analysis was also done using qPCR of *ToxA* gene in *B. sorokiniana* isolate BS112 grown in potato dextrose broth (PDB); harvested at same time points as mentioned above.

Total RNA was isolated from leaf samples inoculated with BS112 and fungal mycelia (BS112) using Trizol (Invitrogen) reagent method as per the manufacturer' guidelines. RNA quantification was done using nanodrop. First strand cDNA was synthesized using Revertaid cDNA synthesis kit (Thermo fisher scientific, Wilmington, USA) as per the manufacturer' guidelines and was stored at -20°C for further use.

Quantitative gene expression was performed using primer pair *ToxA*_{RP}_FTGCGTTCTATCCTCGTACTTC and *ToxA*_{RP}_RGTGATTGACATGCAGCTTCC (designed using Integrated DNA Technologies (IDT) primer designing tool) to amplify the *ToxA* gene. A housekeeping *EFN-1 alpha* gene (*EFN-1* FCTTCTCGCCTACACCCTTG and *EFN-1* RCCTTCTCCCAACCCTTGAC) was also used. The designed primers were synthesized (Eurofins Genomics India Pvt. Ltd., Bengaluru, Karnataka, India).

The quantitative PCR was performed on 48- well plates mini opticon real time PCR system (Biorad, Hercules, California, USA) with the following conditions: an initial activation step at 95 °C for 7 min; followed by 35 cycles of denaturation at 95 °C for 15 s, annealing at 53 °C for 30 s, and extension at 72 °C for 30 s; melt curve analysis of the PCR product was carried out at 72 °C for 1 min and ramped from 75 to 95 °C with a rise by 1 °C every 5 s. qPCR was done using SYBR Green PCR master mix (Fermentas) which includes 10 µl of SYBR Green PCR master mix , 10nM of the appropriate primer (both *EFN-1 alpha* and *ToxA*) and 100ng of template cDNA with a total of 20 µl reaction. Each experiment was carried out in triplicates with two technical replicates, relative gene expressions were calculated in terms of fold changes using the $\Delta\Delta C_t$ method [51].

Abbreviations

SSR: Simple Sequence Repeat; GO: Gene ontology; DAI: Days after inoculation; SNP: Single nucleotide polymorphism; GH: Glycoside hydrolases; CBM: Carbohydrate Binding Modules; CE: Carbohydrate esterase; PL: Polysaccharide lyases; AA: Auxiliary activities; GT: Glucosyl transferases; qPCR: quantitative real time PCR; TCA: Citric acid cycle; CaZymes: Carbohydrate active enzymes; OLC: Overlap layout consensus; ADI: Average disease index; R: resistance; MR: Moderately resistance; HR: Highly resistance; S: susceptible; HS: Highly susceptible; MS: Moderately susceptible; IDT: Integrated DNA technologies; SNB: Septoria nodorum blotch

Declarations

Acknowledgements

Authors are highly thankful to the ICAR-Consortium Research Platform (CRP) on Genomics (ICAR-G/CRP-Genomics/2015-2720/IARI-12-151) for funding this work. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. We are very thankful to the Director and Joint Director (Research) of the ICAR-Indian Agricultural Research Institute, New Delhi, the Head, Division of Plant Pathology, ICAR-IARI, New Delhi, and the Director of ICAR-NIPB, New Delhi for providing facilities.

Funding

This work is financially supported by the Indian Council of Agricultural Research—Consortium Research Project (CRP) on Genomics (ICAR-G/CRP-Genomics/2015-2720).

Authors' contributions

RA and SA wrote the manuscript; RA, BB, MG and MS contributed corrections and suggestions; AR and SSa performed the bioinformatics analysis; RA, SA and PJ analyzed the data and conducted the wet lab experiments; RA, SSh and MG conceived and designed the experiment. All the authors read and approved the final manuscript.

Availability of data and materials

This whole-genome sequencing project has been deposited in DDBJ/ENA/GenBank under the accession number RCTM00000000, BioProject number PRJNA489551, and BioSample identification (ID) SAMN09901300. The version described in this paper is the first version (GenBank accession number RCTM01000000). The raw sequencing data of all the three platforms were deposited in the DDBJ SRA database under the accession numbers SRX6673825 (Ion Torrent), SRX6673824 (Illumina HiSeq), and SRX6100476 (Oxford Nanopore).

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

ICAR- Indian Agriculture Research Institute, Pusa campus, New Delhi.

ICAR- Centre for Agricultural Bioinformatics, IASRI, New Delhi

Indian Council of Agricultural Research, ICAR, New Delhi

References

1. Zillinsky F. In: Common diseases of small grain cereals, a guide to identification. CIMMYT, Mexico, DF; 1983. pp141.
2. Wiese MV. Compendium of wheat diseases. In: Duveiller E, Dubin HJ, Reeves J, McNab A (eds) Proc. Int. Workshop Helminthosporium Disease of Wheat: Spot Blotch and Tan Spot. 9–14 Feb. 1997, CIMMYT, El Batan, Mexico, DF. 1998;pp 114–118.
3. Van Ginkel M, Rajaram S. Breeding for durable resistance to diseases in wheat: an additional perspective. In: Jacobes T, Parlevliet JE (eds) Durability of disease resistance. Kluwer Academic Publishers, Dordrecht, Netherlands. 1993; pp259–272.
4. Nagarajan S, Kumar J. Foliar blights of wheat in India: germplasm improvement and future challenges for sustainable high yielding wheat production. In: Duveiller E, Dubin HJ, Reeves J and McNab A (eds) Proc. Int. Workshop Helminthosporium Diseases of Wheat: Spot Blotch and Tan Spot, 9–14 February 1997, CIMMYT, El Batan, Mexico, DF. 1998;pp 52–58

5. Alam KB, Banu SP, Shaheed MA. The occurrence and significance of spot blotch disease in Bangladesh. In: Duveiller E, Dubin HJ, Reeves J and McNab A (eds) Proc. Int. Workshop on Helminthosporium Disease of Wheat: Spot Blotch and Tan Spot, CIMMYT, El Batan, Mexico. 9–14 February 1997. 1998;pp 63–66
6. Kumar S, Roder MS, Tripathi SB, Kumar S, Chand R, Joshi AK, Kumar U. Mendelization and fine mapping of a bread wheat spot blotch disease resistance QTL. Mol. Breed. 2015;35:218.
7. Chaurasia S, Chand R, Joshi AK. Relative dominance of *Alternaria triticina* Pras. et Prab and *Bipolaris sorokiniana* (Sacc.) Shoemaker, in different growth stages of wheat (*T. aestivum* L.). J. Plant Dis. Prot. 2000;107:176–181.
8. Maraite H, Zinno TDI, Longree H, Daumerie V, Dueveiller E. Fungi association with foliar blight of wheat areas. In: Duveiller E, Dubin HJ, Reeves J, McNab A (eds.), Helminthosporium Diseases of Wheat: Spot Blotch and Tan Spot. CIMMYT, Mexico. 1998;pp. 293–300.
9. Kumar J, Schafer P, Huckelhoven R, Langen G, Baltruschat H, Stein E, Nagarajan S, Kogel KH. *Bipolaris sorokiniana*, a cereal pathogen of global concern: cytological and molecular approaches towards better control. Molecular Plant Pathology. 2002;3: 185–195.
10. Aggarwal R, Das S, Jahani M, Singh D. Histopathology of spot blotch disease of wheat caused by *Bipolaris sorokiniana* (Teleomorph: *Cochliobolus sativus*). Acta Phytopathologica et Entomologica Hungarica. 2008;43(1):23–30.
11. Gayad SK. Production of symptoms of barley leaf spot disease by cultural filtrates of *Helminthosporium sativum*. Nature. 1961;191:725–726.
12. Jahani M, Aggarwal R, Gupta S, Sharma S, Dureja P. Purification and characterization of a novel toxin from *Bipolaris sorokiniana*, causing spot blotch of wheat and analysis of variability in the pathogen. Cereal Research Communications. 2014;42(2): 252–261.
13. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP. Emergence of a new disease as a result of interspecific virulence gene transfer. Nat. Genet. 2006;38:953–956.
14. Stukenbrock EH, McDonald BA. Geographical variation and positive diversifying selection in the host-specific toxin *SnToxA*. Mol. Plant Pathol. 2007;8:321–332.
15. Moolhuijzen PM, See PT, Oliver RP, Moffat CS. Genomic distribution of a novel *Pyrenophora tritici-repentis* *ToxA* insertion element. PLoS ONE. 2018;13(10):e0206586.
16. McDonald MC, Ahren D, Simpfendorfer S, Milgate A, Solomon PS. The discovery of the virulence gene *ToxA* in the wheat and barley pathogen *Bipolaris sorokiniana*. Mol Plant Pathol. 2018;19(2):432–439.
17. Friesen TL, Holmes DJ, Bowden RL, Faris JD. *ToxA* Is Present in the U.S. *Bipolaris sorokiniana* Population and Is a Significant Virulence Factor on Wheat Harboring *Tsn1*. Plant Disease. 2018;102(12):2446–2452.
18. Gupta P, Vasistha N, Aggarwal R and Joshi A. Biology of *B. sorokiniana* (syn. *Cochliobolus sativus*) in genomics era. Journal of Plant Biochemistry and Biotechnology. 2017;27(10):1007/s13562-017-

0426-6.

19. Faris JD, Zhang Z, Lu H, et al. A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proc. Natl. Acad. Sci. U S A.* 2010;107(30):13544–13549.
20. Navathe S, Yadav PS, Chand R, Mishra VK, Vasistha NK, Meher PK, Joshi AK, and Gupta PK. *ToxA–Tsn1* Interaction for Spot Blotch Susceptibility in Indian Wheat: An Example of Inverse Gene-for-Gene Relationship. *Plant Disease.* 2020;104(1):71 – 8.
21. Aggarwal R, Sharma S, Singh K, Gurjar MS, Saharan MS, Gupta S, Bashyal BM and Gaikwad K. First Draft Genome Sequence of Wheat Spot Blotch Pathogen *Bipolaris sorokiniana* BS_112 from India, Obtained Using Hybrid Assembly. *Microbiol. Resour. Announc.* 2019;8(38):e00308-19.
22. Peterson TN, Brunak S, von HG, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods.* 2011;8:785–786.
23. Ramadas S, Kiran Kuma, TM, Pratap Singh G. Wheat Production in India: Trends and Prospects. *Recent Advances in Grain Crops Research.* 2020; doi:10.5772/intechopen.86341.
24. Figueroa M, Hammond-Kosack KE, Solomon PS. A review of wheat diseases-a field perspective. *Mol. Plant Pathol.* 2018;19(6):1523–1536.
25. Aylward J, Steenkamp ET, Dreyer LL, Roets F, Wingfield BD, Wingfield MJ. A plant pathology perspective of fungal genome sequencing. *IMA Fungus.* 2017;8(1):1–15.
26. Kuan CS, Yew SM, Toh YF, Chan CL, Ngeow YF, Lee KW, Na SL, Yee WY, Hoh CC, Ng KP. Dissecting the fungal biology of *Bipolaris papendorffii*: from phylogenetic to comparative genomic analysis. *DNA Res.* 2015;22(3):219–32.
27. Pedley KF, Walton JD. Regulation of cyclic peptide biosynthesis in a plant pathogenic fungus by a novel transcription factor. *Proceedings of the National Academy of Sciences.* 2001;98(24):14174–14179.
28. Lam MQ, Oates NC, Thevarajoo S, Tokiman L, Goh KM, Mason SJM, Bruce NC, Chong CS. Genomic analysis of a lignocellulose degrading strain from the underexplored genus *Meridianimaribacter*. *Genomics.* 2020;112(1).
29. Antoni EA, Rybak K, Tucker MP, Hane JK, Solomon PS, Drenth A, Shankar M, Oliver RP. Ubiquity of *ToxA* and absence of *ToxB* in Australian populations of *Pyrenophora tritici-repentis*. *Austral. Plant Pathol.* 2010;39:63–68.
30. McDonald MC, Oliver RP, Friesen TL, Brunner PC, McDonald BA. Global diversity and distribution of three necrotrophic effectors in *Phaeosphaeria nodorum* and related species. *New Phytol.* 2013;199: 241–251.
31. Wu L, He X, Lozano N, et al. *ToxA*, a significant virulence factor involved in wheat spot blotch disease, exists in the Mexican population of *Bipolaris sorokiniana*. *Trop. plant pathol.* 2020; pp46; doi:10.1007/s40858-020-00391-4.
32. Aggarwal R, Singh VB, Shukla R, Gurjar MS, Gupta S, Sharma TR. URP-based DNA fingerprinting of *Bipolaris sorokiniana* isolates causing spot blotch of wheat. *J. Phytopathol.* 2010;158: 210–216.

33. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 1980;8(19): 4321–4325.
34. Krueger F. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. 2015.
35. Sahoo N. Sequence Base-calling through Albacore software: A part of the OxfordNanopore Technology (Doctoral dissertation). 2017.
36. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. *Bioinformatics.* 2013;29(21):2669–2677.
37. Bryant JM, Baumgarten S, Lorthiois A, Scheidig-Benatar C, Claës A, Scherf A. *De Novo* Genome Assembly of a *Plasmodium falciparum* NF54 Clone Using Single- Molecule Real-Time Sequencing. *Genome announcements.* 2018;6(5):e01479-17.
38. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 2005;W465-7. doi: 10.1093/nar/gki458.
39. Altschul, Stephen F, et al. "Basic local alignment search tool." *Journal of molecular biology.* 1990;215(3):403–410.
40. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research.* 2017;D1(45):353–361.
41. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLOS ONE.* 2010;5(6):e11147.
42. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118. iso-2;iso-3. *Fly.* 2012;6(2):80–92.
43. Li H. Minimap2: fast pairwise alignment for long DNA sequences. *Bioinformatics.* 2018;34:3094–3100.
44. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R. 1000 Genome Project Data Processing Subgroup, The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9 [19505943].
45. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;15:33(16):2583–2585.
46. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. Pfam: the protein families database. *Nucleic Acids Res.* 2014;42:222–30.
47. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Collier CJA, Nakai K. WoLF PSORT: protein localization predictor. *Nucleic Acids Research.* 2007;35(2):85–87.
48. Conesa A, and Götz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International journal of plant genomics.* 2008;619832.

49. Adlakha KL, Wilcoxson HD and Raychaudhury SP. Resistance of wheat to leaf spot caused by *Bipolaris sorokiniana*. Plant Disease. 1984; 68:320–321.
50. Aggarwal R, Singh VB, Gurjar MS, Gupta S, Srinivas P. Intraspecific variations in Indian isolates of *Bipolaris sorokiniana* infecting wheat based on morphological, pathogenic and molecular characters. Indian Phytopath. 2009;62(4):449–460.
51. Livak KJ, Thomas S. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods. 2001;25:402–408.

Figures

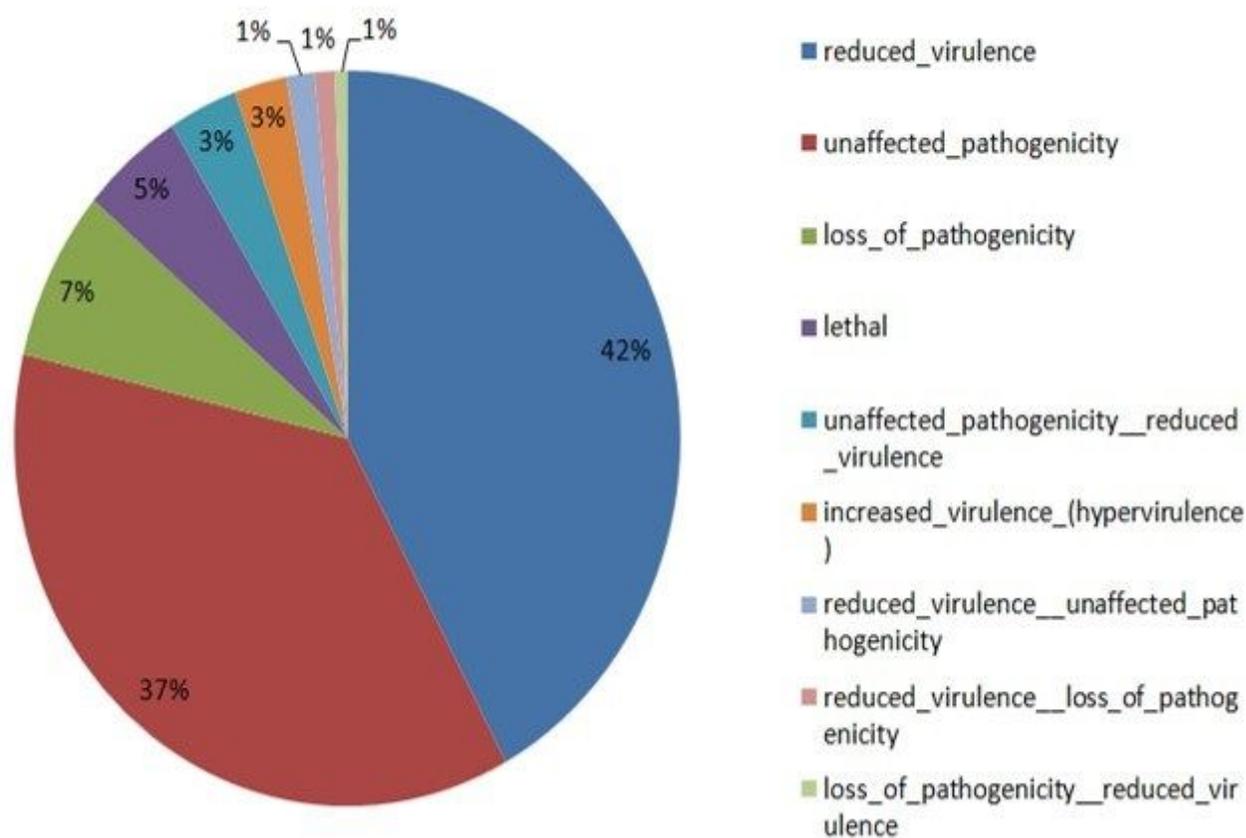


Figure 1

Whole genome functional annotation of *Bipolaris sorokiniana* predicted genes using pathogen–host interaction (PHI) database.

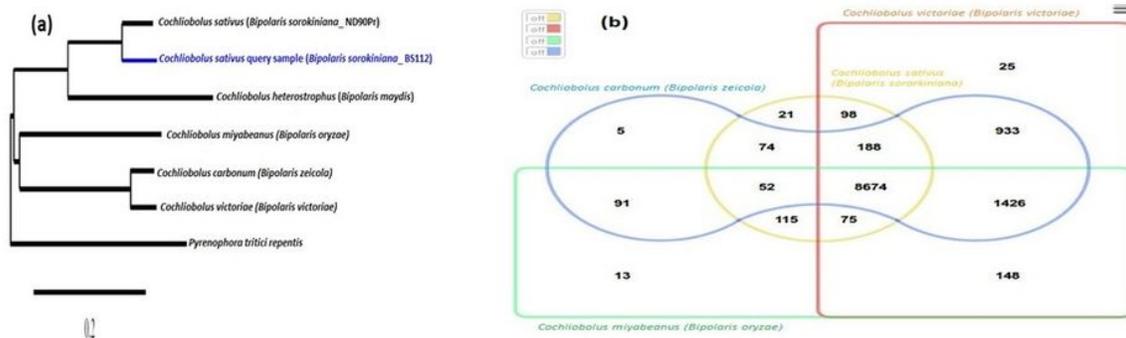


Figure 3

(a) Whole genome phylogenetic analysis between different species of Bipolaris and *Pyrenophora tritici repentis* used as outgroup; (b) Venn diagram showing unique and shared orthologous gene families between Bipolaris species.

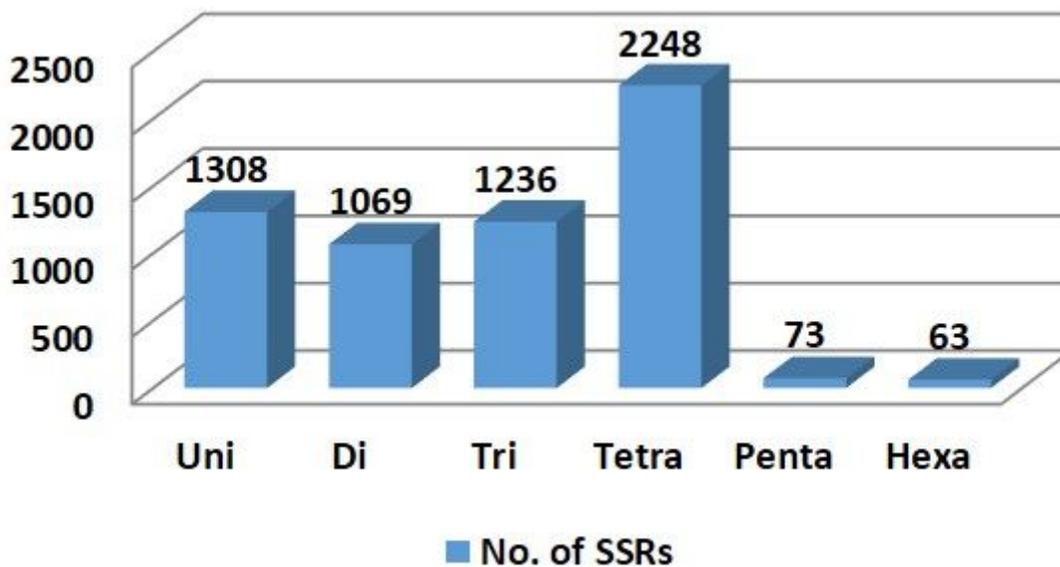


Figure 4

Number of simple sequence repeats with uni, di, tri, tetra, penta, and hexanucleotide in the *B. sorokiniana* genome.

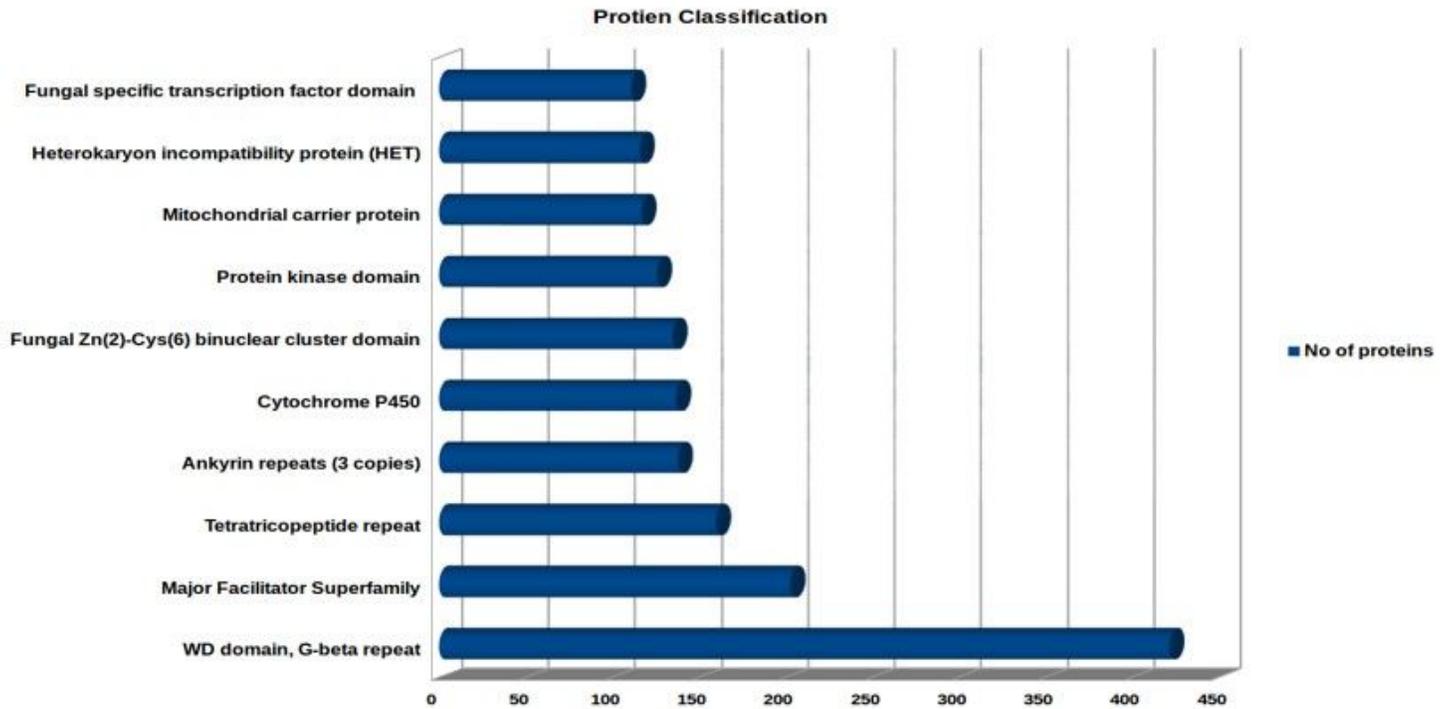


Figure 5

Top 10 PFam domain distribution in the genome of *Bipolaris sorokiniana*.

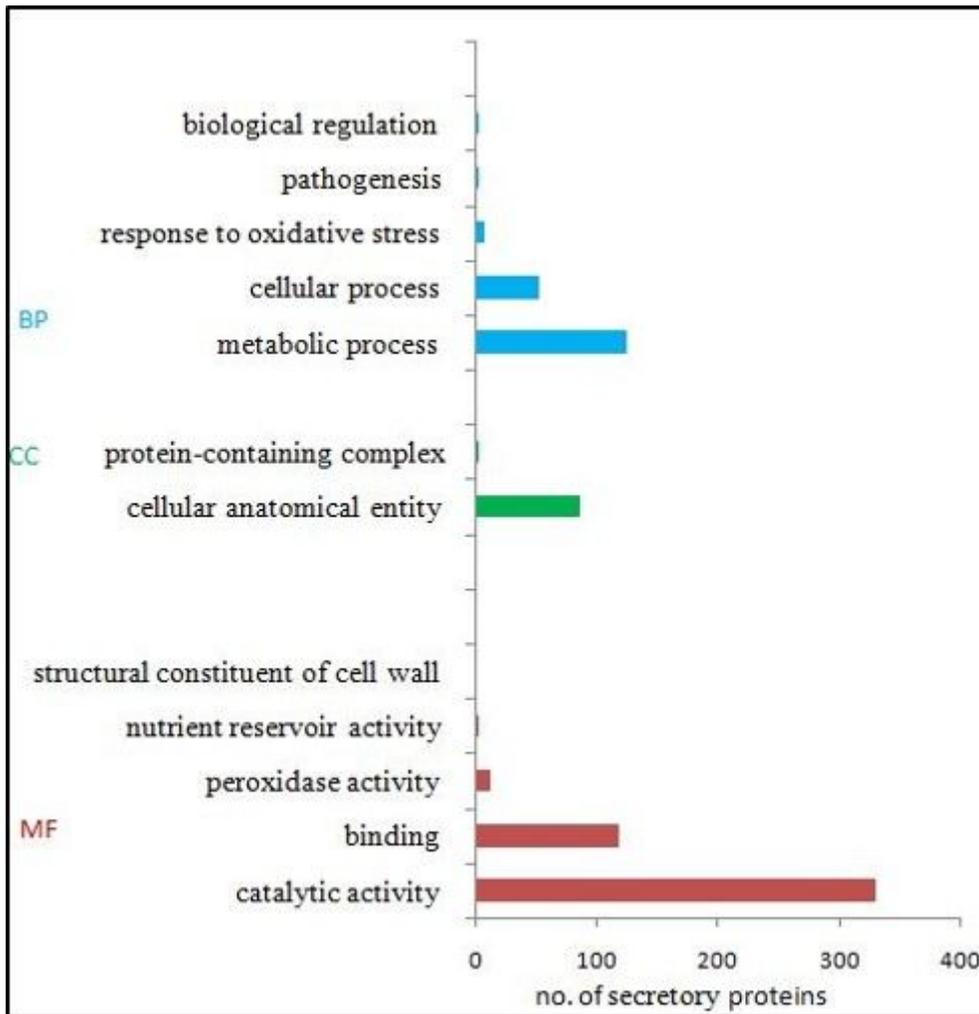


Figure 6

Functional annotation of the *Bipolaris sorokiniana* secretome showing topmost category Viz. MF- molecular function; CC- cellular component; BP- biological process.

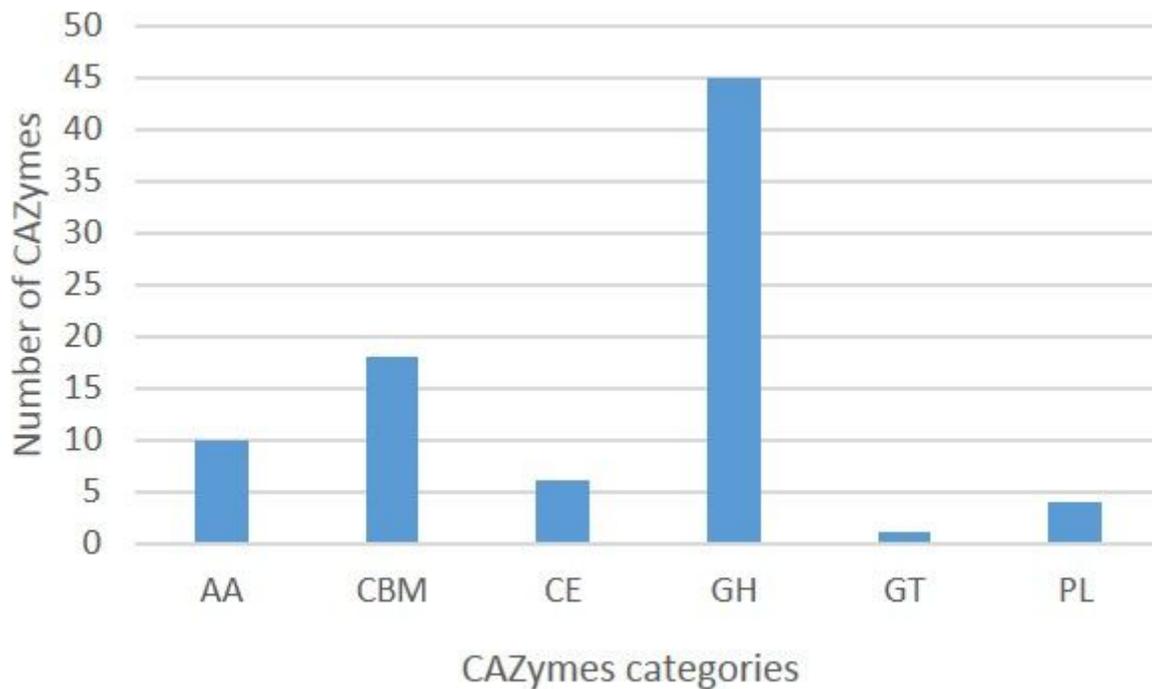


Figure 7

CAZymes class annotation distribution of *Bipolaris sorokiniana* BS112 genome. AA- Auxiliary activities; CBM- Carbohydrate Binding Modules; CE- Carbohydrate esterase; GH- Glycoside hydrolases; GT- Glcosyl transferases; PL- Polysaccharide lyases.

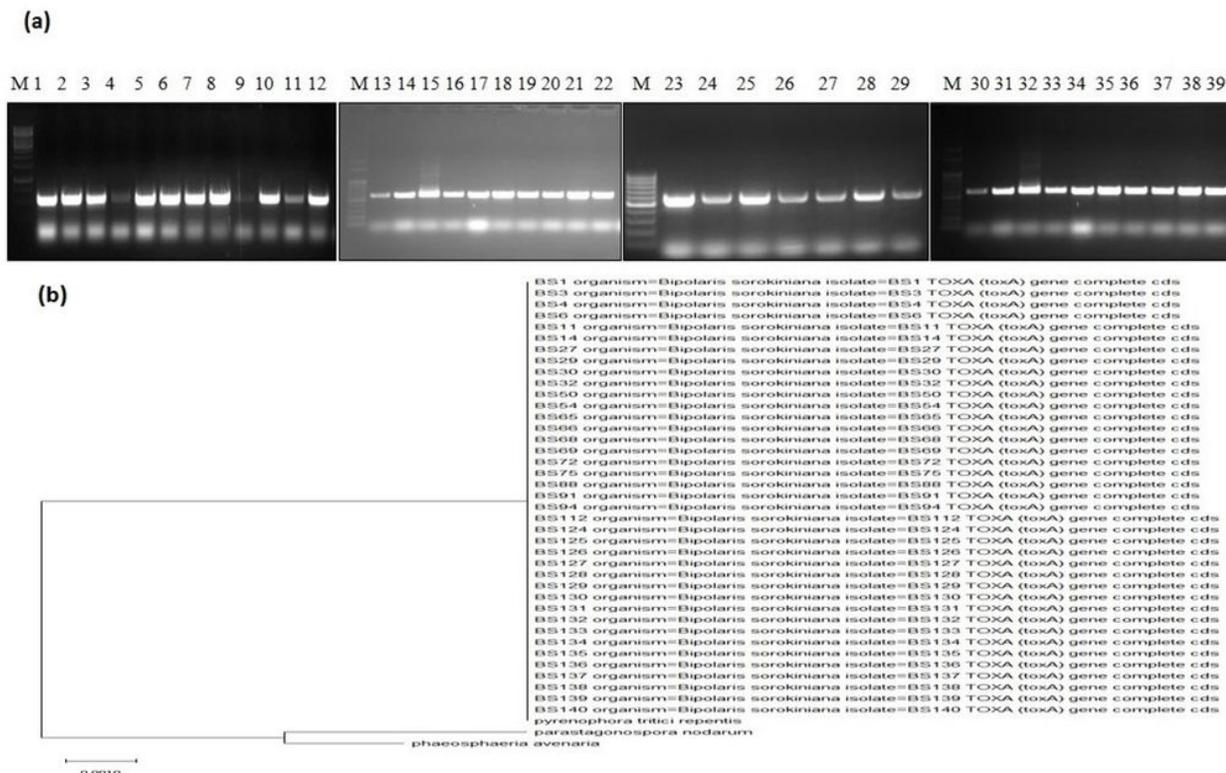


Figure 8

(a) Amplification profile of 39 isolates of *B. sorokiniana* harbouring ToxA gene. M (100bp bioprep marker), 1-BS1, 2-BS3, 3-BS4, 4-BS6, 5-BS11, 6-BS14, 7-BS27, 8-BS29, 9-BS30, 10-BS32, 11-BS50, 12-BS54, 13-BS65, 14-BS66, 15-BS68, 16-BS69, 17-BS72, 18-BS75, 19-BS88, 20-BS91, 21-BS94, 22-BS112, 23-BS124, 24-BS125, 25-BS126, 26-BS127, 27-BS128, 28-BS129, 29-BS130, 30-BS131, 31-BS132, 32-BS133, 33-BS134, 34-BS135, 35-BS136, 36-BS137, 37-BS138, 38-BS139, 39-BS140; (b) Phylogenetic tree of different isolates of *B. sorokiniana*, *Pyrenophora tritici repentis*, *Phaeosphaeria avenaria* and *Parastagonospora nodorum*.

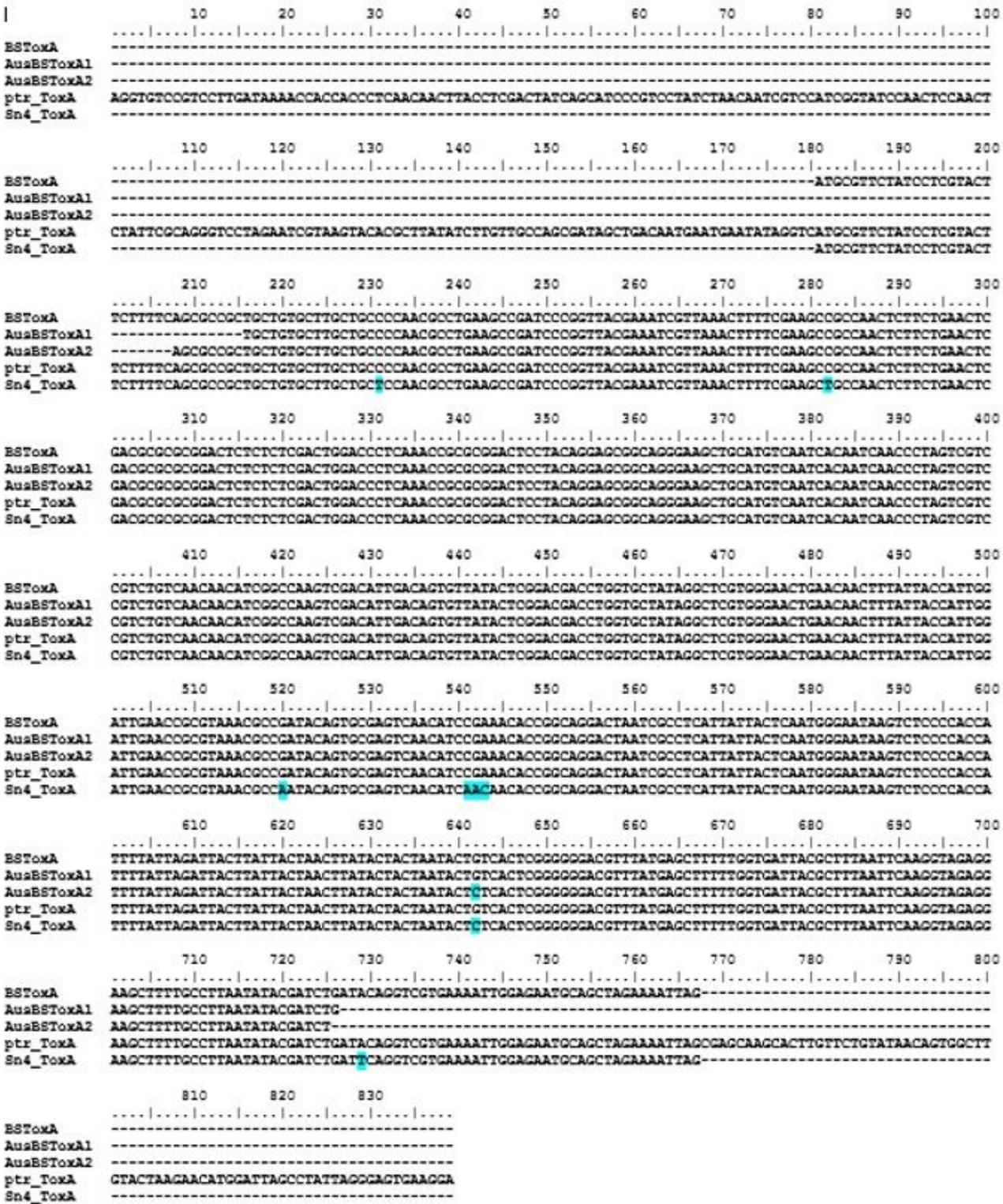


Figure 9

Nucleotide sequences of *Bipolaris sorokiniana* ToxA gene from India (BSToxA) and Australia (AusBSToxA1 and AusBSToxA1) aligned to *Pyrenophora tritici-repentis* (ptr_ToxA) and *Parastagonospora nodorum* (Sn4_ToxA).

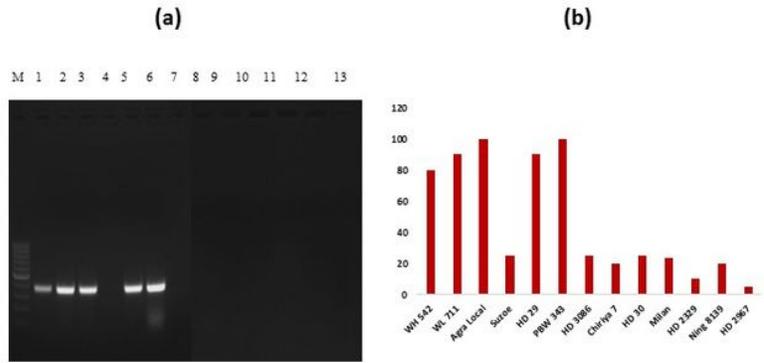


Figure 10

(a) Amplification profile of 13 wheat genotypes harboring Tsn1 gene. 1-WH 542, 2-WL 711, 3-Agra Local, 4-Suzoe, 5- HD 29, 6- PBW 343, 7-HD 3086, 8-Chiriya 7, 9-HD 30, 10-Milan, 11-HD 2329, 12- Ning 8139, 13- HD 2967; (b) Average Disease Index (ADI) of 13 wheat genotypes harboring Tsn1 gene.

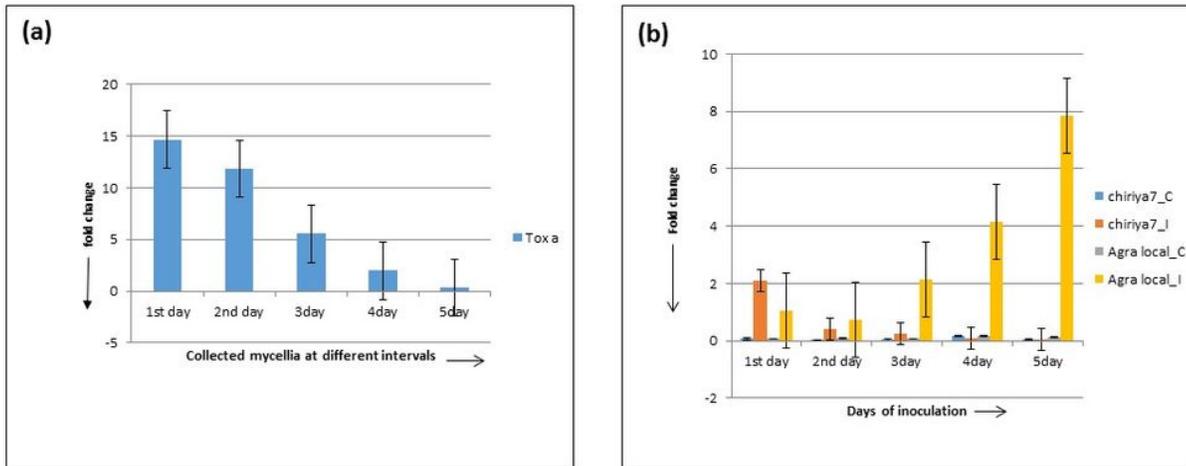


Figure 11

(a) In vitro expression of ToxA gene in BS112 at different time intervals; (b) In vivo expression of ToxA gene after inoculation on two different genotypes at different time intervals.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xlsx](#)
- [Additionalfile2.xls](#)
- [Additionalfile3.xls](#)
- [Additionalfile4.xls](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile6.xls](#)
- [Additionalfile7.xls](#)
- [Additionalfile8fig.S1.jpg](#)