



# De novo genome sequencing and secretome analysis of *Tilletia indica* inciting Karnal bunt of wheat provides pathogenesis-related genes

Malkhan Singh Gurjar<sup>1</sup> · Rashmi Aggarwal<sup>1</sup> · Abhimanyu Jogawat<sup>1</sup> · Deepika Kulshreshtha<sup>1</sup> · Sapna Sharma<sup>1</sup> · Amolkumar U. Solanke<sup>1,2</sup> · Himanshu Dubey<sup>1,2</sup> · Rakesh Kumar Jain<sup>1</sup>

Received: 26 November 2018 / Accepted: 8 May 2019 / Published online: 20 May 2019  
© King Abdulaziz City for Science and Technology 2019

## Abstract

*Tilletia indica* is an internationally quarantined fungal pathogen causing Karnal bunt of wheat. The present study carried out that the whole genome of *T. indica* was sequenced and identified transposable elements, pathogenicity-related genes using a comparative genomics approach. The *T. indica* genome assembly size of 33.7 MB was generated using Illumina and Pac Bio platforms with GC content of 55.0%. A total of 1737 scaffolds were obtained with N<sub>50</sub> of 58,667 bp. The ab initio gene prediction was performed using *Ustilago maydis* as the reference species. A total number of 10,113 genes were predicted with an average gene size of 1945 bp out of which functionally annotated genes were 7262. A total number of 3216 protein-coding genes were assigned in different categories. Out of a total number of 1877 transposable elements, gypsy had the highest count (573). Total 5772 simple sequence repeats were identified in the genome assembly, and the most abundant simple sequence repeat type was trinucleotide having 42% of total SSRs. The comparative genome analysis suggested 3751 proteins of *T. indica* had orthologs in five fungi, whereas 126 proteins were unique to *T. indica*. Secretome analysis revealed the presence of 1014 secretory proteins and few carbohydrate-active enzymes in the genome. Some putative candidate pathogenicity-related genes were identified in the genome. The whole genome of *T. indica* will provide a window to understand the pathogenesis mechanism, fungal life cycle, survival of teliospores, and novel strategies for management of Karnal bunt disease of wheat.

**Keywords** Karnal bunt · Wheat · *Tilletia indica* · Whole genome sequencing · Secretome · Pathogenesis-related genes

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13205-019-1743-3>) contains supplementary material, which is available to authorized users.

Malkhan Singh Gurjar and Rashmi Aggarwal contributed equally towards this paper.

✉ Rashmi Aggarwal  
rashmi.aggarwal2@gmail.com

Malkhan Singh Gurjar  
malkhan\_iari@yahoo.com

Abhimanyu Jogawat  
abmanu.jnu@gmail.com

Deepika Kulshreshtha  
deepikakul12@gmail.com

Sapna Sharma  
teenaiari1@gmail.com

Amolkumar U. Solanke  
amolsgene@gmail.com

Himanshu Dubey  
hemu.bt@gmail.com

Rakesh Kumar Jain  
rakeshjain56@yahoo.co.in

## Introduction

Wheat is one of the most important cereal crops in the world. The world acreage under wheat crop is 218 million hectares with the production of 740 million tones and an average yield of 3394 kg/ha. In India, wheat is grown in an area of about 31.78 million hectares, production of 98.40 million

<sup>1</sup> Fungal Molecular Biology Laboratory, Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, Pusa Campus, New Delhi 110012, India

<sup>2</sup> ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi 110012, India

tons, and average productivity of 3095 kg/ha (Anonymous 2017–2018). Among biotic stresses, *Tilletia indica* is a quarantined fungal pathogen worldwide causing Karnal bunt disease of wheat which affects commercial seed trading as well as the quality and quantity of grains. It was first discovered from Karnal, Haryana, India (Mitra 1931). Presently, it is re-emerging disease in wheat growing areas of the north-western plain zone of India and coefficient of infection was recorded up to 14.25% (Gurjar et al. 2016). In India, the pathogen is considered widespread in Delhi, Haryana, Punjab, Uttar Pradesh, Himachal Pradesh, Rajasthan, Jammu and Kashmir, West Bengal, Madhya Pradesh, and Gujarat (Singh et al. 1985). The occurrence of Karnal bunt disease is in sporadic in nature. The pathogen has been reported from Pakistan, Nepal, Afghanistan, Mexico, Syria, Brazil, USA, Iran, and South Africa (Tan et al. 2013). The wheat importing countries have imposed strict quarantine measures and insist on a zero tolerance limit (Singh and Gogoi 2011). The disease poses an economic threat to the wheat industry due to a reduction in grain quality rather than yield (Tripathi et al. 2013).

*Tilletia indica* is a heterothallic fungus with bipolar incompatibility, controlled by multiple alleles at one locus (Duran and Cromarty 1977) belonging to the phylum Basidiomycota, order Ustilaginales, and family Ustilaginaceae (Nagarajan et al. 1997). *T. indica* infects naturally on bread wheat (*Triticum aestivum*) (Mitra 1931), durum wheat (*T. durum*), and triticale (*Triticosecale*) (Agarwal et al. 1977). However, it can be artificially induced on ryes (*Secale cereale*) and several other wild and related species of *Triticum*, *Aegilops*, *Bromus*, *Lolium*, and *Oryzopsis* (Warham 1986; Gill et al. 1993). The *T. indica* is hemibiotrophic and partial systemic in nature (Carris et al. 2006). It is soil, seed, and air-borne fungal pathogen. Teliospores can survive up to 5 years under severe environmental stress conditions (Agarwal and Verma 1983; Bonde et al. 2004). Soil-borne inoculum (teliospores) is the primary source of annual recurrence of the Karnal bunt disease. Fresh teliospores have a period of dormancy. A dormancy period of 1–6 months is needed prior to spore germination (Prescott 1984). In comparison with freshly harvested spores, old teliospores germinate favourably. The highest germination occurs with year-old teliospores (Bansal et al. 1983). The germination of teliospores is sensitive in terms of temperature and light conditions. A qPCR based diagnostic marker has been developed to detect teliospores of *T. indica* in soil (Gurjar et al. 2017). The fungus shows high pathogenic and genetic variability because of its genetic recombination or mating behavior between two compatible allantoid sporidia just before infection (Kumar et al. 2009; Aggarwal et al. 2010; Tripathi et al. 2011; Thirumalaisamy and Singh 2012). It possesses a distinguished pathogenesis mechanism which starts to infect the host plant at the specific spike development stage

of wheat, i.e., boot stage (S2, Zadok's stage 49) by dikaryotization between compatible mating types. At the time of maturity of wheat grain, teliospores are liberated during harvesting/threshing to the soil surface or dispersed on or in the grain and the disease cycle begins again (Dhaliwal and Singh 1989). Recently, few putative pathogenicity-related genes in *T. indica* were analyzed using real-time PCR (Gurjar et al. 2018).

It is difficult to manage the disease effectively through cultural practices and fungicide applications because of its complex infection behavior (Pandey et al., 2018; Gurjar et al. 2018). The best approach to manage the disease is through resistant cultivars (Brar et al. 2018). Until now, wheat varieties with durable Karnal bunt resistance have not been developed as the fungus interaction with its host to render wheat susceptible is still elusive. Screening the wheat varieties using different pathogenic isolates of *T. indica* is difficult; thereby, identification of resistance and susceptibility is difficult, although a combination of two stable QTL [*QKb.cim-2B* and *QKb.cim-3D* (Pop1), and *QKb.cim-3B1* and *QKb.cim-5B2* (Pop2)] in each population was identified in RILs' population reducing Karnal bunt disease infection by 24–33%, respectively (Brar et al. 2018). Recently, the whole genome *Tilletia indica* has been sequenced with 26.7 Mb (Kumar et al. 2017), but the size of genome was low and transposable elements (TEs) were not identified in the genome of *T. indica*.

Therefore, we report detailed analysis of de novo whole genome assembly and secretome analysis of *T. indica*. Genes were predicted, annotated, and comparison of orthologous gene families led to the identification of pathogenesis-related genes. The first time we have also identified transposable elements in the genome. The predicted candidate effector proteins of *T. indica* will be helpful for understanding their pathogenesis, resistance mechanism and to devise new approaches for the management of Karnal bunt disease of wheat.

## Materials and methods

### Collection, isolation of *T. indica* fungus, growth conditions, and DNA isolation

The Karnal bunt disease sample was collected from Uttar Pradesh, India. The monoteliosporic culture of *T. indica* (RAKB\_UP\_1) was established from the single teliospore of Karnal bunt-infected seeds. Then, fungus was grown in 100 mL potato dextrose broth and kept in a shaker incubator at  $18 \pm 1$  °C under light and dark conditions for 15 days. Then, the fungal mat was harvested, followed by washing with sterilized distilled water and immediately stored in deep freezer ( $-80$  °C). Total genomic DNA was isolated

using NucleoSpin<sup>®</sup> Tissue kit following the manufacturer's instructions. Genomic DNA quality and integrity was checked on 0.8% agarose gel (loaded with 2  $\mu$ l). The gel was run at 120 V for 30 min. 1  $\mu$ l of the sample was used to determine the A260/280 ratio (Nanodrop 2000) and concentration (Qubit<sup>®</sup> 3.0 Fluorometer) of gDNA.

### Genome sequencing and hybrid assembly

The genome of *T. indica* RAKB\_UP\_1 isolate was sequenced using Illumina HiSeq 2500 and PacBio RSII platforms. The paired-end DNA libraries of average 359 bp inserts were prepared using the TruSeq Nano DNA Library sample preparation kit and sequenced using 2 $\times$  125 bp chemistry to generate ~10 Gb data for sequencing. The PacBio library (5–8 kb size) was prepared using a Hairpin adaptor protocol for ultra-long read sequencing on the PacBio RS II platform to generate ~400 Mb data using one SMRT (Single-Molecule Real Time) cell. The reads were obtained from both the platforms of Illumina HiSeq 2500 and PacBio RSII. The Illumina reads were filtered using Trimmomatic v 0.35 (Bolger et al. 2014) with quality value QV > 30 and other contaminants such as adapters were trimmed. Quality filtration of PacBio data is not required, so data were filtered for adapter sequences. The draft genome was assembled using hybrid approach by SPAdes, a de Bruijn-based assembler, Version: 1.5.2 (Bankevich et al. 2012) using high-quality paired-end reads of Illumina HiSeq 2500 and long reads of PacBio RS II with default parameter, while SPAdes itself optimize the kmers and give best genome assembly.

### Transposable elements and simple sequence repeats

Transposable elements were identified in the genome of *T. indica* using TransposonPSI (<http://transposonpsi.sourceforge.net>). Transposon PSI is an analysis tool to identify protein or nucleic acid sequence homology to proteins encoded by diverse families of transposable elements. PSI-Blast is used with a collection of (retro-) transposon ORF homology profiles to identify statistically significant alignments. Simple sequence repeats were identified through MicroSatellite identification tool (MISA) (<http://pgrc.ipk-gatersleben.de/misa/download/misa.pl>) with default parameters.

### Gene prediction

AUGUSTUS is a program that predicts genes in eukaryotic genome sequences. It is the most accurate ab initio program for gene prediction. Protein-coding genes in *T. indica* were predicted using AUGUSTUS Version: 3.2.1 (Stanke et al. 2004) with default parameters (<http://bioinf.uni-greifswald.de/augustus>). A total of 10,113 genes were predicted on the

basis of *Ustilago maydis* as reference model fungus (Basse and Steinberg 2004).

### Gene annotation, functional prediction, and pathway analysis

Functional annotation of the predicted genes was performed using the BLASTx search program, which is a part of ncbi-blast-2.3.0+ standalone tool (Altschul et al. 1990). BLASTx searched the homologous sequences for the genes against NCBI non-redundant database with cut-off *E* value of  $\leq 1e-06$  and similarity  $\geq 34\%$ . Gene ontology (GO) analysis was performed using Blast2GO PRO 4.1.5 (Conesa and Gotz 2008). B2G performed in three different mappings as follows: (1) blast result accessions are used to retrieve gene names (symbols) making use of two mapping files provided by NCBI (gene info, gene 2 accessions). Identified gene names are searched in the species-specific entries of the gene product table of the GO database; (2) blast result GI identifiers were used to retrieve UniProt IDs making use of a mapping file from PIR (non-redundant reference protein database) including PSD, UniProt, Swiss-Prot, TrEMBL, RefSeq, GenPept, and PDB. Pathways analyses were performed using KAAS-KEGG Automatic Annotation Server. KAAS (KEGG Automatic Annotation Server) provides a functional annotation of genes by BLAST or GHOST comparisons against the manually curated KEGG database (Moriya et al. 2007). Blast result accessions were searched directly in the DBXRef table of the GO database.

Pathogen–host interaction database (PHI-base) is a web-accessible database (Winnenburg et al. 2006) that catalogues experimentally verified pathogenicity, virulence, and effectors genes from fungal, oomycete, and bacterial pathogens, which infect an animal, plant, fungal, and insect hosts. To identify the putative pathogenicity-related genes, BLASTP was used against the pathogen–host interaction database with an *E* value of  $\leq 1e-06$ .

### Phylogenetic analysis of *T. indica* genome

The nine basidiomycetes fungal genomes viz. *Ustilago maydis* 521, *Tilletia horrida* QB-1, *Tilletia controversa* DAOM 236426, *Tilletia caries* DAOM 238032, *Tilletia indica* DAOM236416, *Tilletia indica* PSWKBGD 11, *Tilletia indica* PSWKBGD 13, and *Tilletia indica* PSWKBGD 12 were downloaded from NCBI database (<http://www.ncbi.nlm.nih.gov/Traces/wgs>). Scaffolds of *T. indica* RAKB\_UP\_1 were aligned against the downloaded genome of basidiomycetes fungi using the Mauve version with default parameters. The phylogenetic relationships among the genomes were then reconstructed from the pairwise distance matrix (Darling et al. 2004).

## Comparative analysis of orthologous gene families

The orthologous groups among *T. indica* RAKB\_UP\_1, *Tilletia caries* DAOM 238032, *Tilletia horrida* QB-1, *Tilletia indica* PSWKBGD 1 1, *Tilletia walkeri* DAOM 236422, and *Ustilago maydis* were identified using OrthoVenn (Wang et al. 2015) with parameter (BlastP, threshold  $E$  value  $\leq 1e-5$ , inflation value = 1.5).

## Secretome prediction and analysis of secretory effectors

The 10,113 predicted proteins from *T. indica* genome assembly were analyzed in SignalP v4.1 (Petersen et al. 2011) (<http://www.cbs.dtu.dk/services/SignalP/>) as well as TargetP v1.1 (<http://www.cbs.dtu.dk/services/TargetP/>) for prediction of the secretory signal peptides. After this, the protein sets were scrutinized for the presence of transmembrane domain using TMHMM v2.0. (<http://www.cbs.dtu.dk/services/TMHMM/>) and simultaneously, for the presence of GPI (glycosylphosphatidyl inositol)-anchor with PredGPI (<http://gpcr.biocomp.unibo.it/predgpi/>). Proteins having no transmembrane domain and one transmembrane domain within the N-terminal signal peptide were selected. Cysteine content in the predicted secretory proteins was analyzed. The predicted secretome was functionally annotated by assigning GO terms using BLAST2GO (Altschul et al. 1990). The dbCAN (dbCAN HMMs 5.0) (Yin et al. 2012) was used to detect carbohydrate metabolism active enzymes (CAZymes) based on the CAZY database in the *T. indica* secretome.

## Results

### Genome sequencing, assembly, and annotation of *Tilletia indica*

An Indian isolate of *T. indica* RAKB\_UP\_1 was used for whole genome sequencing. The *T. indica* pathogen confirmed using ITS specific region and sequence was submitted in NCBI Genbank (KX369242). Full genome sequencing was performed using Illumina Hi Seq 2500 and Pac Bio RSII platforms. Paired-end libraries with average insert size of 359 bp were sequenced to generate shorter sequence reads (2 × 125 bp) using Illumina and PacBio ultra-long sequence reads (5–8 kb). High-quality paired-end reads of Illumina HiSeq and long reads of PacBio RSII were assembled using hybrid approach by SPAdes (a de Bruijn-based assembler, Version: 1.5.2) with default parameter and kmer ranges in between 53 and 67. Assembly size of *T. indica* was 33.7 Mb with average GC content of 55.0%. An average gene density in the genome of *T. indica* was 405 genes per Mb (Table 1). High coverage of 214× and 23× was achieved on paired-end

**Table 1** Genome features of *Tilletia indica* using Illumina HiSeq 2500 and Pac Bio RSII

Characteristics	<i>T. indica</i> RAKB_UP_1
Size (Mb)	33.7 MB
Coverage (Illumina)	214×
Coverage (PacBio)	23×
Coverage	99.91%
Protein-coding genes	9977
% GC content	55%
Gene density (number of genes per Mb)	405
Number of Scaffolds	1737
Average size of scaffolds	19,443 bp
Scaffold N <sub>50</sub>	58,667 bp
Maximum size of scaffold	462,479 bp
Minimum size of scaffold	500 bp
Secreted proteins	1014

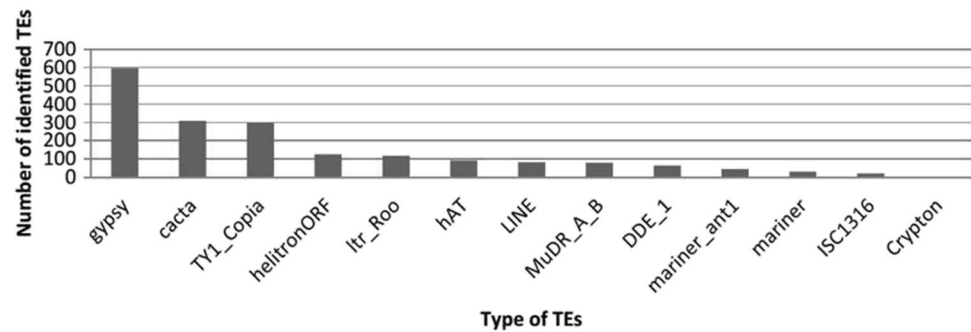
reads and PacBio reads, respectively. Genome assembly consisted of 1737 scaffolds, with N<sub>50</sub> scaffold size approximately 58,667 bp. A total of 1,737 scaffolds were generated with the N<sub>50</sub> scaffolds of 58,667 bp and an average scaffold of 19,443 bp. The ab initio gene prediction from the assembled scaffolds was performed using AUGUSTUS (Version: 3.2.1) with default parameters and *Ustilago maydis* used as the reference model fungus. A total number of 10,113 protein-coding genes were predicted with an average gene size of 1945 bp. The maximum and minimum sizes of the genes were 36,447 bp and 201 bp, respectively. In functional annotation, a total of annotated genes were 7262 out of 10,113 genes in the genome. In gene ontology (GO), 3216 protein-coding genes (Online Resource 1) were assigned to different categories including ‘Biological Process’ (1148 genes), ‘Cellular Component’ (833 genes), and ‘Molecular Function’ (1235 genes). In addition, a total of 3,216 genes related to 154 pathways were annotated through the KEGG database (Online Resource 2).

### Identification of transposable elements and simple sequence repeats

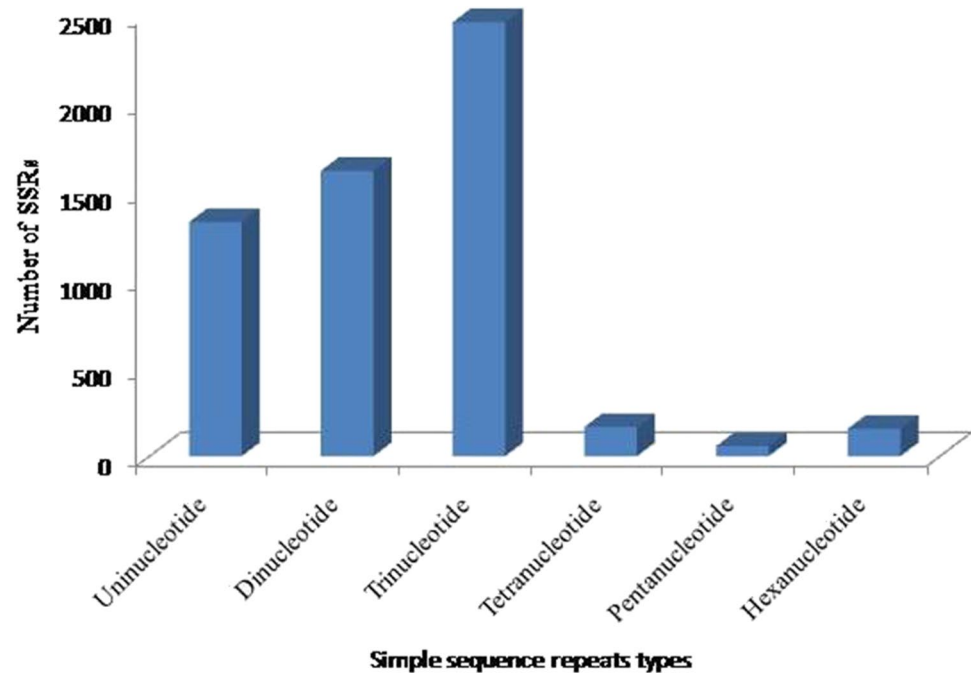
Transposable elements were analyzed using TransposonPSI. A total number of 1877 transposable elements were identified out of which gypsy is having the highest count of 573 followed by cacta with 309 times occurrence (Fig. 1; Online Resource 3). Microsatellite or simple sequence repeats create and maintain genetic variation and play an active role in genome evolution. SSRs were identified as mono, di, tri, tetra, penta, and hexanucleotide in genome of *T. indica*. In total, 5772 simple sequence repeats were identified in the genome assembly (Fig. 2). The most abundant simple



**Fig. 1** Identified transposable elements in *T. indica* genome showing highest count of gypsy



**Fig. 2** Number of simple sequence repeats with mono, di, tri, tetra, penta, and hexanucleotide in the *Tilletia indica* genome



sequence repeats type was trinucleotide in *T. indica* genome with 2456 in number (42% of all SSRs).

### Comparative study with other phytopathogenic basidiomycetes fungal genomes

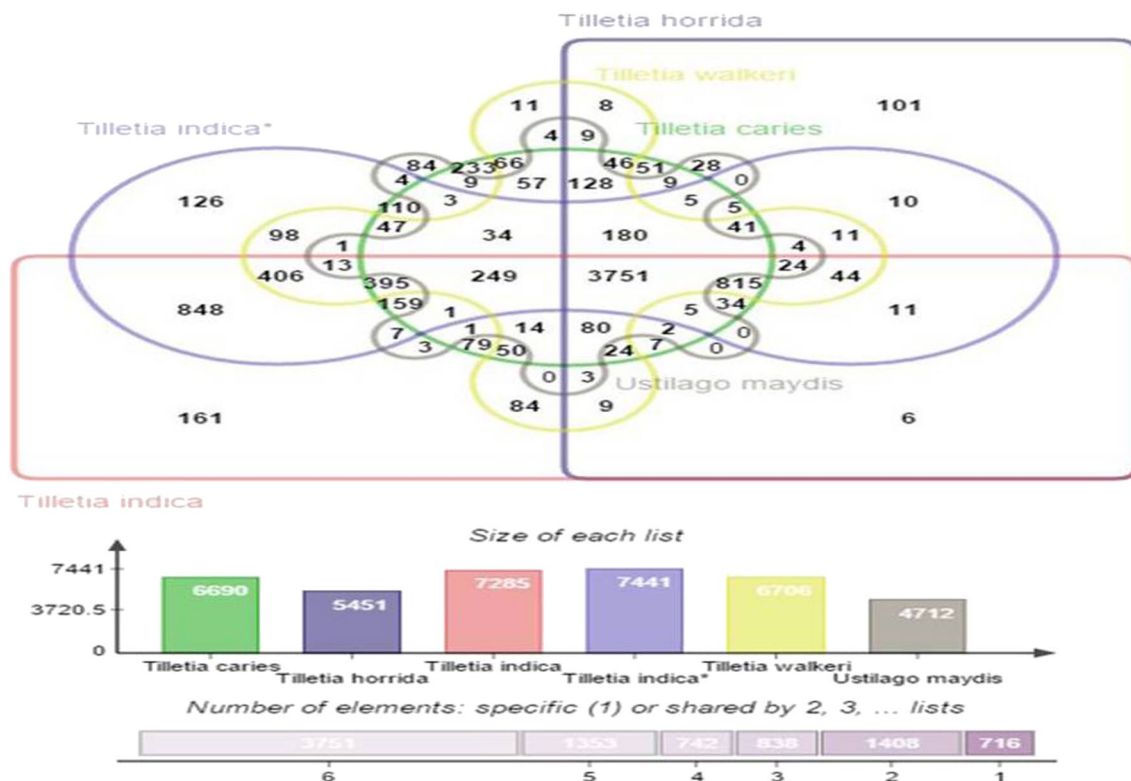
The sequenced genome of *T. indica* RAKB\_UP\_1 was compared with closely related basidiomycetes fungi viz. *Tilletia caries* DAOM 238032, *Tilletia horrida* QB-1, *Tilletia indica* PSWKBGD 1 1, *Tilletia walkeri* DAOM 236422, and *Ustilago maydis* 521 using OrthoVenn (Wang et al. 2015). A total number of 7441 clusters and 1817 singletons in *T. indica* RAKB\_UP\_1 genome identified. It showed that 3751 protein families of *T. indica* were orthologs in five phytopathogenic fungi causing bunt and smut diseases, whereas 126 protein families were unique to *T. indica* only (Fig. 3).

The evolutionary relationship of *T. indica* with other phytopathogenic smut and bunt fungi was determined through phylogenetic analysis. The phylogenetic tree revealed two

clades; one is grouped to smut fungi (*U. maydis*) separately as an out group and second is grouped to bunt fungi (all *Tilletia* sp.). The genome of *Tilletia indica* RAKB\_UP\_1 was found closely related to *Tilletia indica* DAOM236416 from Pakistan (30.38 Mb) followed by *Tilletia indica* PSWKBGD 11, Karnal, India (37.46 Mb) on the basis on whole genome sequencing (Fig. 4).

### Secretome prediction and analysis of *T. indica* genome

Most of the plant pathogenic fungi mainly depend on secreted proteins, especially effectors. Secretome analysis was performed to know the secretory proteins in the genome. A computational pipeline was used for the prediction of secretory effectors proteins. 772 proteins of SignalP and 1307 of targetP proteins were predicted to have secretory signals (Online Resource 4). Predicted secretory proteins were merged and duplicates were removed. In total, 1337 unique



**Fig. 3** Comparison of orthologous genes with basidiomycetes fungi. Venn diagram showing the distribution pattern of shared and unique orthologous gene families. The orthologous gene families viz. *Tilletia caries* DAOM 238032, *Tilletia horrida* QB-1, *Tilletia indica* PSWK-

BGD 1 1, *Tilletia walkeri* DAOM 236422, and *Ustilago maydis* 521 were suggested using Ortho Venn. A total of 3, 751 proteins of *T. indica* were orthologs in five fungi, whereas 126 proteins were unique

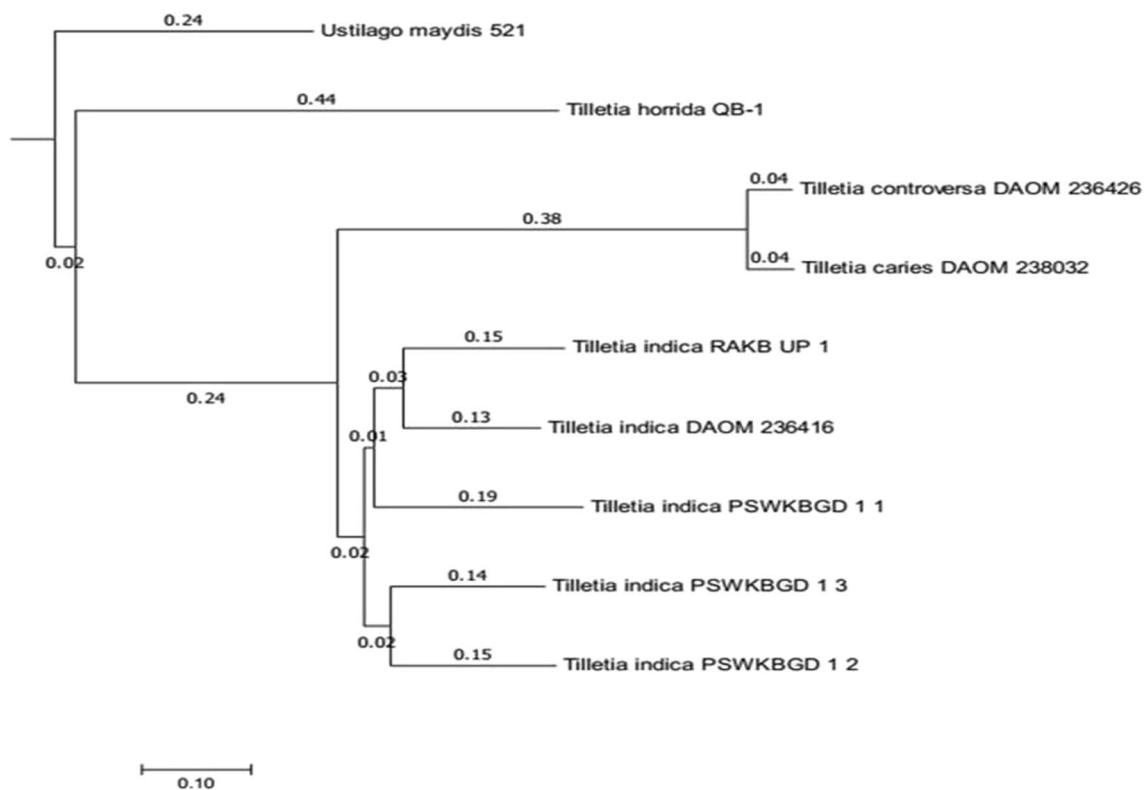
proteins were having secretory signatures. Furthermore, these secretory proteins were analyzed for the presence of a transmembrane domain using TMHMMv2.0. It suggested that 829 secretory proteins had zero transmembrane domain (TmHmm 0), and 185 proteins had one transmembrane domain (TmHmm 1) including 34 highly probable glycosylphosphatidylinositol (GPI) anchor containing sequences. In the predicted secretome, the functionally annotated were 646 secreted proteins and 368 were having no-hit in non-redundant database. 360 secretory proteins were annotated as hypothetical or putative proteins. For analysis of cysteine residue, predicted secretory proteins were divided into two parts; first part  $\leq 200$  amino acids (small-secreted proteins) and second part containing proteins with  $> 200$  amino acids (large-secreted proteins). 2.66% of small-secreted proteins were identified having cysteine content  $> 4\%$  and 3.25% of large-secreted proteins having cysteine content  $> 4\%$  (Fig. 5; Online Resource 5).

Enzymes required for degrading plant cell walls are crucial for pathogen invasion. The growth efficiency and aggressiveness of plant pathogens are often associated with their carbohydrate metabolism active enzymes. In addition, 315 predicted secretory proteins were annotated with

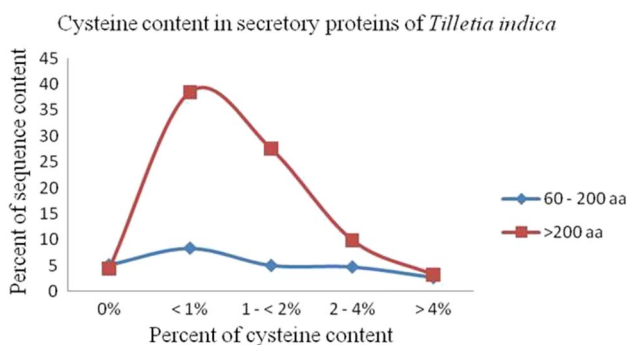
the CAZyme database (Fig. 6a, Online Resource 6). The secreted carbohydrate enzymes were consisting of 105 glycosyl hydrolase (GH) families, 85 of glycosyl transferase (GT) families, 83 of carbohydrate esterase (CE) families, 8 of carbohydrate-binding module (CBM) families, 30 of auxiliary activity (AA) families, and 4 of polysaccharide lyase (PL) families (Fig. 6 b–g). Glycosyl hydrolase, glycosyl transferase, and carbohydrate esterase families were highly prevalent which are required for degradation of the plant cell wall. The most prevalent CAZyme of glycosyl hydrolases classes were GH16 and GH5.

### Pathogenesis-related genes in *T. indica*

To identify the putative pathogenicity-related genes, 803 genes were annotated against the pathogen–host interaction database (PHI database). Based on the homology of pathogenicity proteins, 97 genes were related to effectors (plant avirulence determinant), 25 genes to increased virulence, 63 genes to loss of pathogenicity, and 7 genes resistance to chemicals (Fig. 7; Online Resource 7). Furthermore, these pathogenicity-related genes can be analyzed for functional characterization.



**Fig. 4** Phylogenetic tree was constructed with Basidiomycetes fungi. The phylogeny represents *T. indica* RAKB\_UP\_1 is more closely related to *T. indica* DAOM236416



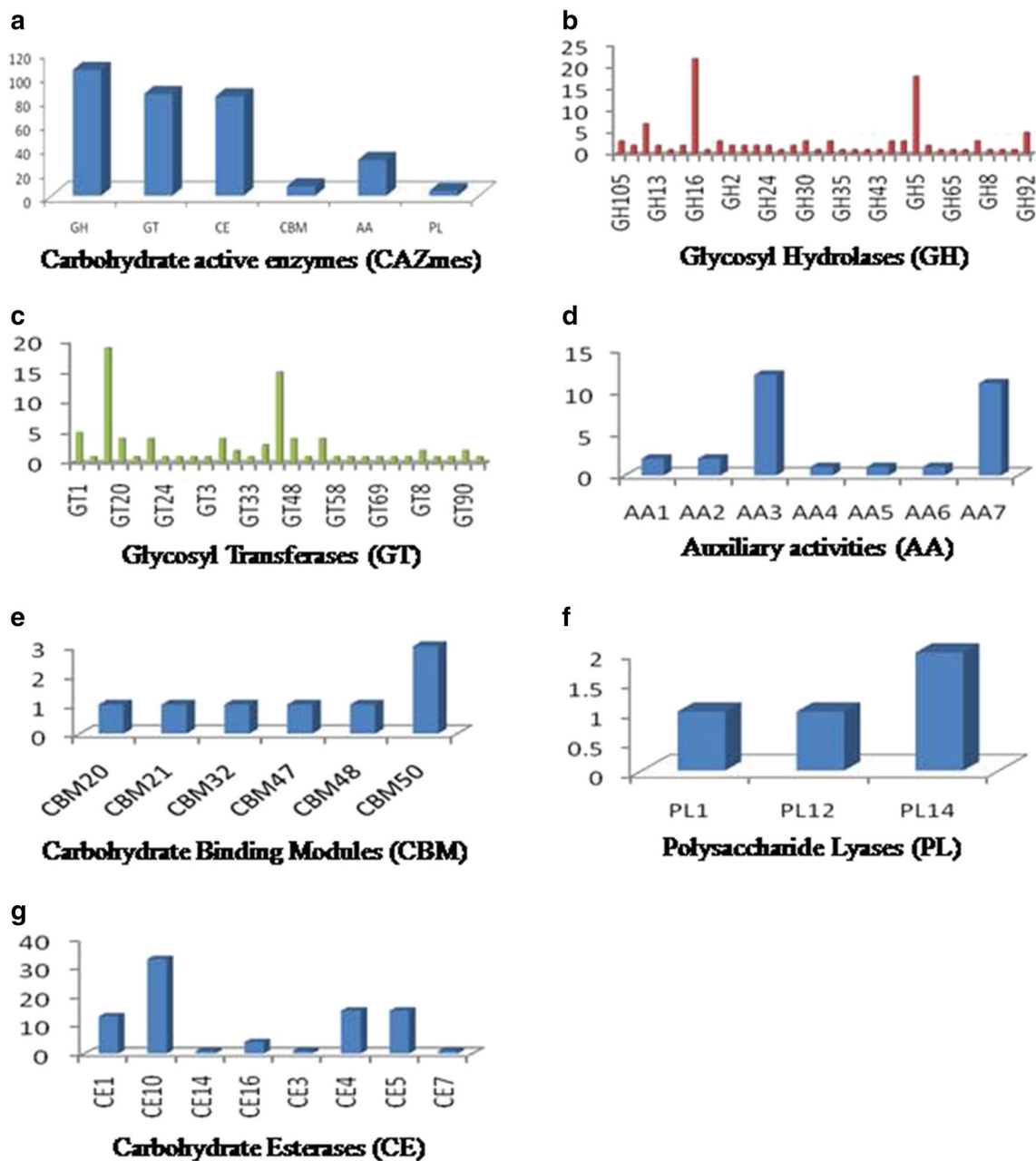
**Fig. 5** Analysis of cysteine content in secretory proteins, and it suggested that 2.66% of small-secreted proteins having cysteine content >4%

## Discussion

India's wheat production has reached 98.40 million tones, and now, the country is in a position to export wheat to other countries, but Karnal bunt disease is a major constraint for

wheat export. The disease was first reported from Karnal, India. *T. indica* is a hemibiotroph fungus that has a unique life style to cause the Karnal bunt disease at the boot stage of wheat (Zadoks 49 (Singh and Gogoi 2011)). It is an important quarantined fungal pathogen of wheat and having more attention to climate-based occurrence of Karnal bunt disease of wheat. The fungus is very slow growing in media. The pathogenesis mechanism is complex due to survival of teliospores in soil and genetic recombination between two compatible allantoid sporidia just before infection (Gurjar et al. 2017). However, there are very few reports available on the functional characterization of pathogenicity genes in *T. indica* fungus (Gupta et al. 2013).

In this study, genome sequence data were generated using long- and short-sequence read strategies to accomplish the complete genome sequence of Karnal bunt pathogen. Based on a hybrid assembly approach, we established a full genome of *T. indica*, a quarantined fungal pathogen. The assembly size was 33.7 Mb with 55% of GC content. The genome of *T. indica* is highly compact and the genome size was found comparable with other smut fungi. Earlier, *Tilletia indica* genome has been sequenced with 26.7 Mb, 31.38 Mb



**Fig. 6** **a** Functional annotation of CAZmes. CAZmes were consisted of **b** glycosyl hydrolases (GH) families, **c** glycosyl transferases (GT), **d** carbohydrate esterases (CE), **e** carbohydrate-binding modules (CBM), **f** auxiliary activities (AA), **g** polysaccharide lyases (PL)

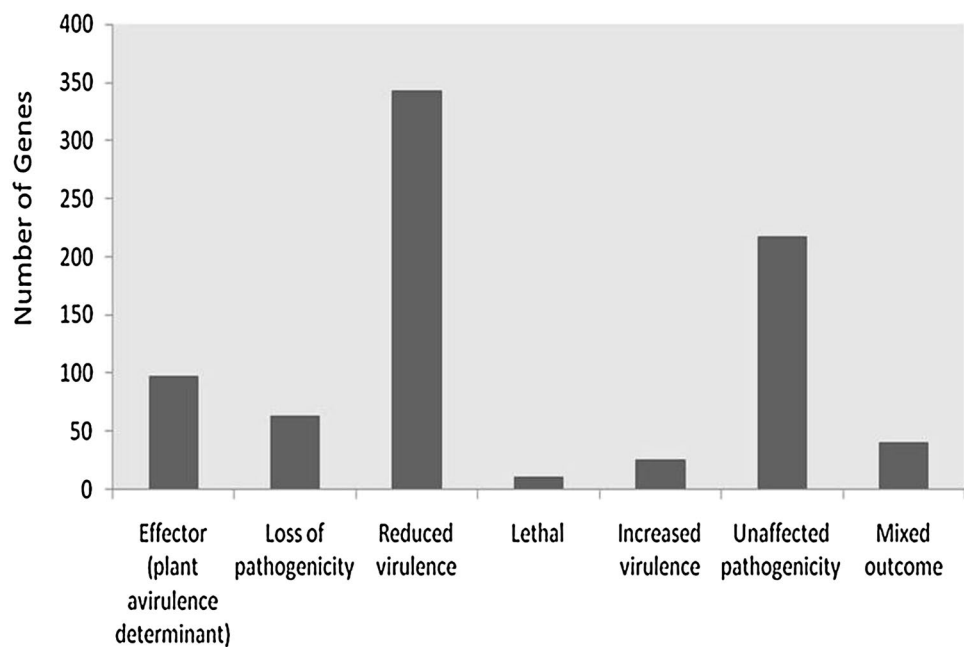
assemblies, and GC content of 53.99% (Kumar et al. 2017; Kumar et al. 2018). The  $N_{50}$  scaffold length was 3 kb in 26.7 Mb genome assembly of *T. indica* (Kumar et al. 2017), but transposable elements were not analyzed. Our present study revealed that the  $N_{50}$  scaffold length was 58,667 bp in 33.7 Mb of *T. indica* genome assembly. Comparatively, our genome assembly was more accurate and near to complete genome assembly of *T. indica*.

The biotrophic fungus *Ustilago maydis* is arguably one of the best model pathogens for the study of host–pathogen

interactions and molecular mechanisms involved in pathogenesis (Schirawski et al. 2010). In this study, a total number of 10,113 genes were predicted based on *U. maydis* as the reference fungus. Genomes of smut fungi harbor less in mean repetitive elements compared to other fungi. Transposable elements (TEs) cause genetic changes and contributions to the evolution of pathogens (Castanera et al. 2016). In the present investigations, genome of *T. indica* possess 1877 transposable elements (1.65% TEs) out of which gypsy



**Fig. 7** Functional annotation of predicted genes using pathogen–host interaction (PHI) database



was having the highest count (573 in number) followed by cacta with 309 times occurrence. The addition of TEs may increase the size of the fungal genomes, but it could also aid the species adaptation in new or adverse environmental conditions (Laurie et al. 2012). TEs may increase the size of *T. indica* genome variation. Only 1.1% of TEs were found in *Ustilago maydis* genome causing smut of maize (Kamper et al. 2006).

Comparative genome analysis suggested that proteins described for *Tilletia indica* RAKB\_UP\_1, *Tilletia caries* DAOM 238032, *Tilletia horrida* QB-1, *Tilletia indica* PSWKBGD 11, *Tilletia walkeri* DAOM 236422, and *Ustilago maydis* 521 genomes were 10,113, 10,204, 6716, 10,226, 7970, and 9038, respectively. 3751 protein families of *T. indica* were orthologs in five phytopathogenic fungi causing bunt and smut fungi, whereas 126 protein families were unique to *T. indica* genome. Comparative genomics of closely related species is the best approach for the identification of virulence determinants (Kamper et al. 2006). The secreted carbohydrates enzymes consisted of 105 of GH (glycosyl hydrolases) families, 85 of GT (glycosyl transferases) families, and 83 of CE (carbohydrate esterases) families. CAZymes play a vital role in degrading plant biomass, belonging to the family glycosyl hydrolases, carbohydrate esterases, and polysaccharide lyases also serve as cell-wall degrading enzymes (Ospina-Giraldo et al. 2010; Zhao et al. 2013). Oxalic acid and few novel virulence factors such as suppression of host defense, lignin degradation, penetration, etc. were identified in *T. indica* using proteomics approach under host factor influence (Pandey et al. 2018, 2019)

Secretory proteins play a crucial role during infection and pathogenesis (Verma et al. 2016). Of 1014 predicted secretory

proteins identified in genome assembly, 829 secretory proteins were having zero transmembrane domain (TmHmm 0) and 185 with one transmembrane domain (TmHmm 1) including 34 highly probable GPI anchor. Pathogen–host interaction database analysis suggested that 97 genes were related effector (plant avirulence determinant), 25 genes related to increased virulence, 63 genes related to loss of pathogenicity, and 7 genes related to resistance to chemicals. Some genes could play a role in pathogenesis and development of Karnal bunt disease. These pathogenicity-related genes could be utilized for functional characterization to understand the infection mechanism of *T. indica* causing the Karnal bunt of wheat.

To conclude, high coverage of the whole genome of *T. indica* using Illumina and Pac Bio platforms allowed to the successful analysis of comparative genomics and pathogenicity-related genes. Higher numbers of transposable elements were occurred in the genome. The addition of TEs may increase the size of the fungal genomes. Comparative genome analysis allowed for the identification of secretory proteins, carbohydrate-active enzymes and pathogenesis-related gene clusters. The present investigation shall provide new prospects for functional genomics of different kinds of biological processes in *T. indica* hemibiotroph pathogen. Furthermore, functional analysis of putative candidate's genes is necessary for determining functions in disease development and pathogenesis mechanism. The whole genome and putative candidate genes will be useful in devising novel disease management strategies for the Karnal bunt of wheat and other smut diseases.

**Acknowledgements** The authors are highly thankful to ICAR-Consortium Research Platform (CRP) on Genomics (ICAR-G/

CRP-Genomics/2015-2720/IARI-12-151) for funding this work. We are also highly thankful to Director, Joint Director (Research), ICAR-Indian Agricultural Research Institute, New Delhi for providing guidance and facilities for carry out this study. Authors are also thankful to Eurofins Genomics India Pvt. Ltd., Bangalore, India for whole genome sequencing.

**Author contributions** MSG and RA were involved in conceptualization of project, whole genome sequencing, and data compilation. AJ, DK, SS, and HD performed DNA extraction, assembly, gene prediction, annotation, phylogenetic analysis, and bioinformatics analysis. RA and MSG supervised the genome sequencing data and drafted the manuscript. AKS and RKJ were involved in critical inputs and finalization of manuscript. All authors have read and contributed for final manuscript.

## Compliance with ethical standards

**Conflict of interest statement** The authors declare that they have no conflict of interest in the publication.

**Availability of data and materials** The whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers MBSW00000000. The version described in this paper is version MBSW01000000.

## References

- Agarwal VK, Verma HS (1983) A simple technique for the detection of karnal bunt infection in wheat seed samples. *Seed Res* 11(1):100–102
- Agarwal VK, Verma HS, Khetarpal RK (1977) Occurrence of partial bunt on triticale. *FAO Plant Prot Bull* 25:210–211
- Aggarwal R, Tripathi A, Yadav A (2010) Pathogenic and genetic variability in *Tilletia indica* monosporial culture lines using universal rice primer-PCR. *Eur J Plant Pathol* 128:333–342. <https://doi.org/10.1007/s10658-010-9655-4>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Anonymous (2017–2018) Production of major crops. *Econ Surv* 2:A35. (<http://mofapp.nic.in:8080/economicsurvey/appendixPDF/tab1.12.pdf>)
- Bankevich A, Nurk S, Antipov D (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Bansal R, Singh DV, Joshi LM (1983) Germination of teliospores of Karnal bunt of wheat. *Seed Res* 11:258–261
- Basse CW, Steinberg G (2004) *Ustilago maydis*, model system for analysis of the molecular basis of fungal pathogenicity. *Mol Plant Pathol* 5(2):83–92
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bonde MR, Nester SE, Olsen MW, Berner DK (2004) Survival of teliospores of *Tilletia indica* in Arizona field soils. *Plant Dis* 88:804–810. <https://doi.org/10.1094/PDIS.2004.88.8.804>
- Brar GS, Fuentes-Dávila G, He Xinyao, Sansaloni CP, Singh RP, Singh PK (2018) Genetic mapping of resistance in hexaploid wheat for a quarantine disease: Karnal bunt. *Front Plant Sci* 9:1497. <https://doi.org/10.3389/fpls.2018.01497>
- Carris LM, Castlebury LA, Goates BJ (2006) Nonsystemic bunt fungi- *Tilletia indica* and *T. horrida*: a review of history, systematics, and biology. *Annu Rev Phytopathol* 44:113–133. <https://doi.org/10.1146/annurev.phyto.44.070505.143402>
- Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J et al (2016) Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genet* 12(6):e1006108. <https://doi.org/10.1371/journal.pgen.1006108>
- Conesa A, Gotz S (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* 2018:619832. <https://doi.org/10.1155/2008/619832>
- Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14(7):1394–1403. <https://doi.org/10.1101/gr.2289704>
- Dhaliwal HS, Singh DV (1989) Production and inter-relationship of two types of secondary sporidia of *Neovossia indica*. *Curr Sci* 58:614–618
- Duran R, Cromarty R (1977) *Tilletia indica*: a heterothallic wheat bunt fungus with multiple alleles controlling incompatibility. *Phytopathology* 67:812–815
- Gill KS, Sharma I, Aujla SS (1993) Karnal and wheat production. Punjab Agriculture University, Ludhiana, p 153
- Gupta AK, Joshi GK, Seneviratne JM, Pandey D, Kumar A (2013) Cloning, in silico characterization and induction of *TiKpp2* MAP kinase in *Tilletia indica* under the influence of host factor (s) from wheat spikes. *Mol Biol Rep* 40(8):4967–4978. <https://doi.org/10.1007/s11033-013-2597-0>
- Gurjar MS, Jogawat A, Kulshreshtha D, Sharma S, Gogoi R, Aggarwal R (2016) Intraspecific variation of *Tilletia indica* isolates causing Karnal bunt of Wheat in India. *Indian Phytopath* 69:352–356
- Gurjar MS, Aggarwal R, Sharma S, Kulshreshtha D, Gupta A, Gogoi R, Thirumalaisamy PP, Saini A (2017) Development of real time PCR assay for the detection and quantification of teliospores of *Tilletia indica* causing Wheat Karnal bunt in soil. *Indian J Exp Biol* 55(6):549–554. <http://nopr.niscair.res.in/handle/123456789/42556>
- Gurjar MS, Jogawat A, Saharan MS, Aggarwal R (2018) Response of putative pathogenicity-related genes in *Tilletia indica* inciting Karnal bunt of wheat. *Cereal Res Commun* 46(1):89–103. <https://doi.org/10.1556/0806.45.2017.067>
- Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ et al (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444:97–101. <https://doi.org/10.1038/nature05248>
- Kumar J, Shoran J, Sharma I, Aggarwal R, Saharan MS, Sharma AK (2009) Distribution of heterothallic alleles of *Tilletia indica* in India. *Indian Phytopath* 62(3):285–294
- Kumar A, Pandey V, Singh M, Pandey D, Saharan MS, Marla SS (2017) Draft genome sequence of Karnal bunt pathogen (*Tilletia indica*) of wheat provides insights into the pathogenic mechanisms of quarantined fungus. *PLoS One* 12(2):e0171323. <https://doi.org/10.1371/journal.pone.0171323>
- Kumar A, Mishra P, Maurya R, Mishra AK, Gupta VK, Ramteke PW, Marla SS (2018) Improved draft genome sequence of a monoteliosporic culture of the Karnal bunt (*Tilletia indica*) pathogen of wheat. *Genome Announc* 6(20):e0001518. <https://doi.org/10.1128/genomeA.00015-18>
- Laurie JD, Ali S, Linning R, Mannhaupt G, Wong P, Guldener U et al (2012) Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. *Plant Cell* 24(5):1733–1745. <https://doi.org/10.1105/tpc.112.097261>
- Mitra M (1931) A new bunt on wheat in India. *Ann App Biol* 18(2):178–179. <https://doi.org/10.1111/j.1744-7348.1931.tb02294.x>

- Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185. <https://doi.org/10.1093/nar/gkm321>
- Nagarajan S, Aujla SS, Nanda GS, Sharma I, Goel LB, Kumar J, Singh DV (1997) Karnal bunt (*Tilletia indica*) of wheat—a review. *Rev Plant Pathol* 76(12):1207–1214
- Ospina-Giraldo MD, Griffith JG, Laird EW, Mingora C (2010) The CAZyome of *Phytophthora* spp: a comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus *Phytophthora*. *BMC Genomics* 11:525–541. <https://doi.org/10.1186/1471-2164-11-525>
- Pandey V, Singh M, Pandey D, Kumar A (2018) Integrated proteomics, genomics, metabolomics approaches reveal oxalic acid as pathogenicity factor in *Tilletia indica* inciting Karnal bunt disease of wheat. *Sci Rep* 8(1):7826. <https://doi.org/10.1038/s41598-018-26257-z>
- Pandey V, Gupta AK, Singh M, Pandey D, Kumar A (2019) Complementary Proteomics, Genomics approaches identifies potential pathogenicity/virulence factors in *Tilletia indica* induced under the influence of host factor. *Sci Rep* 9(1):553. <https://doi.org/10.1038/s41598-018-37810-1>
- Petersen TN, Brunak S, von Gunnar H, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions Thomas Nordahl. *Nature Methods* 8(10):785–786. <https://doi.org/10.1038/nmeth.1701>
- Prescott JM (1984) Overview of CIMMYT's Karnal bunt programme, Proceedings of the conference on Karnal bunt, 1984, Obregon, Sonora, Mexico, p 24
- Schirawski J, Mannhaupt G, Munch K, Brefort T, Schipper K, Doeblemann G et al (2010) Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* 12(330):1546–1548. <https://doi.org/10.1126/science.1195330>
- Singh DV, Gogoi R (2011) Karnal bunt of wheat (*Triticum* sp.): a global scenario. *Indian J Agril Sci* 81:3–14
- Singh DV, Srivastava KD, Joshi LM, Verma BR (1985) Evaluation of some fungicides for control of Karnal bunt of wheat. *Indian Phytopath* 38:571–573
- Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 32(2):W309–W312. <https://doi.org/10.1093/nar/gkh379>
- Tan MK, Brennan JP, Wright D, Murray GM (2013) A review of the methodology to detect and identify Karnal bunt—a serious biosecurity threat. *Austr Plant Pathol* 42:95–102. <https://doi.org/10.1007/s13313-012-0176-9>
- Thirumalaisamy PP, Singh DV (2012) Variability of Indian isolates of *Tilletia indica* assessed by pathogenicity and molecular markers. *J Phytopathol* 160:525–531. <https://doi.org/10.1111/j.1439-0434.2012.01940.x>
- Tripathi A, Aggarwal R, Yadav A (2011) Determination of variability in monosporial lines of *Tilletia indica* by RAPD analysis. *Arch Phytopathol Plant Protect* 44:1312–1321. <https://doi.org/10.1080/03235408.2010.496560>
- Tripathi A, Aggarwal R, Yadav A (2013) Differential expression analysis of defense related genes responsive to *Tilletia indica* infection in wheat. *Turk J Biol* 37:606–613. <https://doi.org/10.3906/biy-1301-16>
- Verma S, Gazara RK, Nizam S, Parween S, Chattopadhyay D, Verma PK (2016) Draft genome sequencing and secretome analysis of fungal phytopathogen *Ascochyta rabiei* provides insight into the necrotrophic effector repertoire. *Sci Rep* 6:24638. <https://doi.org/10.1038/srep24638>
- Wang Y, Coleman-Derr D, Chen G, Gu YQ (2015) OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 43:W78–W84. <https://doi.org/10.1093/nar/gkv487>
- Warham EJ (1986) Karnal bunt disease of wheat: a literature review. *Trop Pest Manag* 32(3):229–242
- Winnenburg R, Baldwin TK, Urban M, Rawlings C, Kohler J, Hammond-Kosack KE (2006) PHI-base: a new database for pathogen-host interactions. *Nucleic Acids Res* 34:D459–D464. <https://doi.org/10.1093/nar/gkj047>
- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <https://doi.org/10.1093/nar/gks479>
- Zhao Z, Liu H, Wang C, Xu J (2013) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* 23(14):274–289. <https://doi.org/10.1186/1471-2164-14-274>