Differential gene expression analysis in germinating and dormant teliospores of *Tilletia indica* using RNA seq approach

Haritha Mohan M1, Malkhan Singh Gurjar1*, Prachi Jain1, Naresh Kumar2, Mahender Singh Saharan1 & Rashmi Aggarwal1

1Fungal Molecular Biology Laboratory, Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, New Delhi-110012 India
2Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi-110012, India

Received 28 December 2021; revised 10 February 2023

Karnal bunt of wheat is an important quarantine disease that interrupts India’s wheat trade in the international market. The whole transcriptome of germinating and dormant teliospores of *Tilletia indica* was performed using the RNA Seq approach to identify germination-related genes. Approximately 63 million reads were generated using the RNA sequencing by the Illumina NextSeq500 platform. The high-quality reads were deposited in NCBI SRA database (accession: PRJNA522347). The unigenes from the pooled teliospores were 16,575 having unigenes length of 28,998,753 bases. The high-quality reads of germinating teliospores mapped on to 21,505 predicted CDSs. 9,680 CDSs were common between dormant and germinating teliospores of *T. indica*. 11,825 CDSs were found to be in germinating teliospores while only 91 were unique in dormant spores of pathogen. The pathway analysis showed the highest number of pathways was found in germinating spores than dormant spores. The highest numbers of CDSs were found to be associated with translation (431 in number), transport and catabolism (340), signal transduction (326), and carbohydrate metabolism (283). The differential expression analysis (DESeq) of germinating and dormant teliospores showed that 686 CDS were up-regulated and 114 CDS were down-regulated in the germinating teliospores. Significant germination-related genes in the spores were validated using qPCR analysis. Ten genes viz. Ti3931, Ti6828, Ti7098, Ti7462, Ti7522, Ti 9289, Ti 8670, Ti 7959, Ti 7809 and Ti10095 were highly up-regulated in germinated teliospores which may have role in germination of spores. Further, these differentially expressed genes provide insights into the molecular events. This first study of transcriptome will be helpful to devise better management strategies to manage Karnal bunt disease.

**Keywords:** Coding sequence, Gene ontology, Karnal bunt, Transcriptome

Wheat (*Triticum aestivum* L.) is an important staple food crop across the world. India is the second largest producer of wheat in the world. In the world, wheat cultivation covers an area of 222.2 million hectares, with a production of 779.03 million metric tons. In India, it has an acreage of 30.54 million hectares with an average production of 106.84 million tonnes yielding on an average 3484 kg/ha (2021-22). Among biotic challenges, Karnal bunt is a disease of quarantine importance limits the country's trade in wheat exports and usually occurs sporadically. The disease is endemic to the north-western plain zone of India. It was first reported by Manoranjan Mitra in 1931. The changes in wheat varieties have significantly affected the occurrence of the disease, so the onset and severity of the disease differ over the years. Presently it is re-emerging disease in changing climate scenario. The disease had been reported in the states of Punjab, Haryana, Jammu, lower Himachal Pradesh, Uttar Pradesh, Delhi, Rajasthan and Madhya Pradesh in India. It causes yield loss of 0.01% to 1%. It is a major biosecurity concern for wheat exporting countries. It is also reported in few other countries viz. Afghanistan, Pakistan, Nepal, Mexico, Iraq, Iran, Lebanon, Syria, Sweden, Turkey, and United States and South Africa. The pathogen (*Tilletia indica* Mitra) is seed, soil and air-borne in nature. The teliospores live for several years in the soil that is the main source of inoculum and can resist adverse environmental conditions. Seed- or soil-borne teliospores seem to play only a starting role in disease epidemics. Newly...
formed teliospores have dormancy that can be broken by exposing them to high temperatures of 40-43°C for 18 days or more under direct sunlight. One of the important distinctions of teliospores of *T. indica* is that immediately after its production spores undergo dormancy for 9 months. Temperature and moisture are critical elements in determining whether there will be a disease outbreak. During the process of germination, the diploid nucleus of the spore undergoes meiosis, which will be followed by mitosis. The teliosporic population of *T. indica* Mitra was detected based on qPCR diagnostic marker. Recently, the full genome of *T. indica* Mitra has been sequenced and virulence genes were identified to understand the mechanism of pathogenesis.

The mystery of spore dormancy and germination is a key question, which remains unresolved until now. The proposed hypothesis that is highly expressed genes induced during spore germination would most likely encode proteins that regulate teliospores formation and/or germination. So far, the mechanism of teliospores germination of *T. indica* Mitra and threshold level of teliospores to cause disease is not understood. The transcriptome is the analysis of the transcripts of the complete set of RNA transcripts that are produced by the genome using high-throughput methods. Comparative transcriptome analysis were carried out in *Aspergillus* for revealing the dormancy and germination-related genes and the role of the *AtfA* gene in the dormancy of conidia has been revealed. The proteins involved in glycometabolism, spore wall proteins, and ricin B lectin spores of *Nosema bombycis* were upregulated in phosphatase-associated gene was indicated that dephosphorylation may be having a role in the germination of spore in *Ustilago maydis* (DC.) Corda inciting smut of maize. Keeping this in view, here, we have made an attempt to identify the gene expression profiles in the teliospores, differential gene expression and genes related to germination using transcriptome analysis (RNA-seq approach) of teliospores of *Tilletia indica*.

**Materials and Methods**

*T. indica* Mitra teliospores sample preparation

Infected Karnal bunt samples collected and maintained in Fungal Molecular Biology Laboratory, Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, New Delhi were used in the present study. The diseased samples (wheat grains filled with mass of teliospores of *T. indica*) were surface disinfected with 0.1%. Black teliospores powder was harvested on butter paper under aseptic conditions from the single infected seed, microscopically analysed and was stored in −20°C for further use. One-year-old spores were taken as dormant teliospores. For germination, 50 mg teliospores were taken in the Eppendorf tube, soaked in water and were kept for germination at 16±1°C in the incubator for 15 days. The germination of spores was assessed under the microscope at 40X magnification (Nikon H600L).

**RNA extraction, quantification and tape-station profiles of library**

Total RNA was isolated from dormant and germinating teliospores using the TRizol method with slight modifications. The quality and quantity of isolated RNA was checked on 1% denaturing RNA agarose gel and Nano drop, respectively. The qualities of the isolated RNA were checked on high sensitivity RNA Screen Tape and the quantifications were done using Qubit Fluorometer. The RNA-seq paired-end libraries were prepared from RNA samples using the TruSeq stranded mRNA Library Prep Kit. The means of the paired end library fragment size distribution were 416 bp and 422 bp, respectively for germinating and dormant teliospores. The libraries were sequenced on NextSeq 500 using 2 × 75 bp chemistry to generate ~5 GB of data per sample.

**cDNA Library construction and RNA sequencing**

For RNA-Seq, paired-end(PE) sequencing libraries of cDNAs were prepared from the isolated total RNA using the Illuma TruSeq stranded mRNA sample preparation kit. Briefly, mRNA was enriched from the total RNA using poly-T attached magnetic beads, followed by enzymatic fragmentation and conversion to the first-strand cDNA. The second strand of cDNA was then synthesized using second strand mix and Act-D mix to facilitate RNA dependent synthesis. The double-stranded cDNA samples were then purified using Ampure XP beads followed by A-tailing, adapter ligation and then enriched by the limited number of PCR cycles. The PCR enriched libraries were quantified using the Qubit 3.0 and analyzed in 4200 Tape Station system (Agilent Technologies, India) using high sensitivity D1000 Screen tape as per the manufacturer’s instructions. After obtaining the Qubit concentration for the libraries and the mean peak size from Agilent Tape Station profile, the PE Illuma libraries were loaded onto the NextSeq 500
for cluster generation and sequencing. The kit reagents were used for binding of samples to complementary adapter oligos on the paired-end flow cell. The adapters were designed to allow selective cleavage of the forward strands after re-synthesis of the reverse strand during sequencing. The copied reverse strand was then used to sequence from the opposite end of the fragment.

Sequence data generation and quality control
The sequenced raw data was processed to obtain high-quality clean reads using the Trimmomatic v0.35 to remove adapter sequences, ambiguous reads (reads with unknown nucleotides “N” larger than 5%), and low-quality sequences (reads with more than 10% quality threshold (QV) and < 20 Phred score). A minimum length of 50 nucleotides (nt) after trimming was applied. After removing the adapter and low-quality sequences from the raw data, 34,341,621 (2x75 bp) and 29,335,701 (2x75 bp), high quality reads were obtained for dormant and germinating teliospores respectively. The high quality (QV>20) paired-end reads were used for de-novo assembly.

De-novo transcriptome assembly, unigenes and coding sequence prediction
The filtered high-quality reads of the teliospores were pooled together and assembled into transcripts using the velvet v1.2.10 and Oases v0.2.09 with a k-mer of 29. Large amounts of misassembled transcripts, erroneous and poorly supported transcripts arised during the assembly. Therefore, all high quality transcripts were mapped on the assembled transcripts using the BWA v0.7.12. Transcripts with poor coverage and mapping were filtered out using an in–house Perl script. The validated transcripts were further clustered into unigenes based on global sequence identify threshold of 90% using CD-HIT-EST version 4.6 to reduce redundancy. The UniGene is an NCBI database of the transcriptome. Each entry is a set of transcripts that appear to stem from the same transcription locu. Information on protein similarities, gene expression, cDNA clones, and the genomic location was included with each entry. The TransDecoder version 2.0 was used to predict coding sequences from the Unigenes Database. TransDecoder identified candidate-coding regions within unigene sequences. The TransDecoder version 2.0 was used to predict coding sequences from unigenes.

Functional annotation and sample wise coding sequences
The predicted CDS were searched against NCBI non-redundant protein database (Nr) using the basic local alignment search tool (BlastX) (E-value: 1e-05). The high-quality reads for each sample were mapped on to the predicted 21,596 CDS using a BWA aligner for read count calculation.

Gene ontology analysis
Gene ontology (GO) annotations of the CDS were determined by the Blast2GO program. GO assignments were used to classify the functions of the predicted CDS. The GO mapping also provided ontology of defined terms representing gene product properties which were grouped into three main domains: Biological process (BP), Molecular function (MF), and Cellular component (CC). GO mapping was carried out in order to retrieve GO terms for all the BLASTx functionally annotated CDS viz. (i) BLASTX result accession IDs were used to retrieve gene names or symbols, identified gene names or symbols then searched in the species-specific entries of the gene product table of GO database. (ii) BLASTX result accession IDs were used to retrieve UniProt IDs making use of PIR which includes PSD, UniprotKB, TrEMBL, GenPept and PDB database. (iii) Accession IDs were searched directly in the dbxref table of the GO database. BLASTX result accession IDs searched directly in the gene product table of the GO database.

Pathway analysis
The pathways analyses were performed using the KAAS–KEGG automatic annotation server. To identify the potential involvement of the predicted CDS of dormant teliospore and germinating teliospore in biological pathways, the CDS were mapped to reference canonical pathways in KEGG (Kyoto Encyclopedia of Genes and Genomes) (https://www.kegg.jp/). All the CDS were classified mainly under five categories: metabolism, cellular system, genetic information processing, environmental information processing and organismal systems. The output of KEGG analysis included KEGG orthology (KO) assignments and corresponding enzyme commission(ES) numbers and metabolic pathways of predicted CDS using KEGG automated annotation serve KASS (http://www.genome.jp/kaas main).

Differential gene expression analysis (DESeq)
Differential expression analysis was performed on the coding sequences common between dormant and germinating teliospores by employing a negative binomial distribution model (DESeq v1.26.0 package http://www-huber.embl.de/users/anders/DESeq/). Dispersion values were estimated with the following Parameters:method = blind, sharingMode = fit-only and
fittype = local.Log2fold change(FC) value calculated using the formula, $FC = \log_2 \left( \frac{\text{Treated}}{\text{Control}} \right)$. FC value greater than zero were considered up-regulated whereas less than zero were down-regulated. The $p$-value threshold of 0.05 was used to filter statistical results. A complete linkage hierarchical cluster analysis was performed on the top 100 differentially expressed genes using multiple experiment Viewer (MEV v4.8.1). A heatmap showing level of transcript abundance provided. The differentially expressed genes identified in control and experimental conditions were analyzed by hierarchical clustering.

**Identification of germination-related genes in T. indica from differentially expressed genes**

The DESeq data for germinating teliospores were used for candidate gene selection. The up-regulated and down-regulated genes were selected from the data based on significant fold changes of more than 2, differential expression, gene ontology and pathway analysis. Although, many genes were found up- and down-regulated, only few significant genes were selected for further analysis. These genes were expected to be having role in the process of germination and dormancy of teliospores of T. indica.

**Relative expression analysis of germinated teliospores of T. indica**

Ten germination-related genes were taken for expression analysis (Table 1). Total RNA and cDNA synthesis were performed as described in a previous study. The qPCR conditions were standardized. The PCR was performed in a 20 μL reaction volume for each gene. The SYBR green was used as a detection dye. The non-treated control samples were also amplified to calculate the fold-change expression levels. The PCR conditions used were: pre-incubation at 95°C for 10 min, denaturation at 95°C for 15 s, annealing and extension at 60°C for 1 min with 40 cycles of amplification. The threshold ($C_T$) values were automatically calculated. After 40 cycles, the specificity of the amplification was verified by melting from 60 to 95°C, resulting in melting curves. The gene expression in T. indica was normalized by subtracting the mean $C_T$ values for a reference gene GAPDH from $^{\Delta C_T}$ values of DEGs. The fold change values were calculated where $^{\Delta \Delta C_T}$ represent $^{\Delta C_T}$ condition of target gene - $^{\Delta C_T}$ control gene. The fold expression was calculated according to the $2^{\Delta \Delta C_T}$ method.

**Results**

**RNA sequence and De novo transcriptome assembly**

RNA-seq generated more than 5 GB transcriptome data from each library constructed out of dormant and germinating teliospores. After removing the adapter and low-quality sequences from the raw data 34,341,621 (2×75 bp) and 29,335,701 (2×75 bp), high quality reads were obtained for dormant and germinating teliospores respectively. Approximately 63 million reads were generated using RNA sequencing by Illumina NextSeq500 platform (Table 2). The high-quality reads were deposited in NCBI SRA database (accession: PRJNA522347). The

| Table 1 — List of selected genes and primers used for expression analysis |
|-----------------------------|----------|--------------------------|
| Sequence IDs | Genes | Product length | Primer sequences (5'-3') |
| Unigene_3931 | GTPase domain containing | 159 | G TTTGACGCTAAGACCACTAT  |
| | | | R GTGTTGGCCATGACGTTGGA |
| Unigene_6828 | 60S acidic ribosomal P0 | 193 | ATCATCCTGACCAGCAGTG |
| | | | R CCTTGGACGATCTTGGTGGA |
| Unigene_7098 | ribosomal s RNA binding domain containing | 158 | GCATGCACACTGGTCAGTTC |
| | | | R GTAGTGTGGAGGTACGGG |
| Unigene_7462 | 3-oxoacid –transferase | 199 | CCTGAGGAGATGCTGTCAT |
| | | | R GCCCAAGTGGAGAAGAAGA |
| Unigene_7522 | histone H2A-like-1 | 171 | ATCTTATGTACCACCGCGCA |
| | | | R ATCCTGCAACATCACAATGTC |
| Unigene_7959 | cytochrome c551 | 219 | CAGGTCCCGCAGGAAGAGAG |
| | | | R TGGCAAGTGCACCTCAGATGTC |
| Unigene_8670 | ABC transporter G family member 20 | 239 | CGACGACAGACAAAGACGGGA |
| | | | R GATGTTGGAGGTACGGG |
| Unigene_7959 | Fatty acid desaturase | 222 | TCAGGATGCGACAGCGGAG |
| | | | R CGAGGATGCGTCTACAGT |
| Unigene_7809 | Pleiotropic drug resistance 3 | 191 | TGTTTGTCAACAGGATCGAC |
| | | | R ATGGAGGTGAGACACGAGA |
| Unigene_10095 | NAD-dependent succinate-semialdehyde dehydrogenase | 167 | AACTTCCCAACCGCAGATGAT |
| | | | R TCACCCGGTCAAGATGTTGAG |
validated transcripts were further clustered into unigenes based on global sequence identity threshold of 90% using the CD-HIT-EST version 4.6 to reduce redundancy.

Unigenes and coding sequence (CDS) prediction

On the basis of length, 16,575 unigenes were found from the pooled (dormant and germinating) teliospores of *T. indica* having 28,998,753 bases (Table 3) of unigene length. The maximum unigene length was found to be 13,476 bases and the minimum unigene length was found to be 200 bases, which gave a mean unigene length of 1749.5 bases. Further, unigenes containing bases between 1000 and 2000 were highest in number (5,696). The unigenes with less than 500 bases were 2,295, unigene between bases 500 to 1000 were 3,385. The unigenes with more than 3000 bases were 2,384.

The total number of CDS from pooled teliospores was 21,596 having a CDS length of 28,955,355 bp (Table 4). The maximum and minimum CDS length were 11,169 bases and 297 bases, respectively. The mean CDS length was found to be 1340.8 bases. CDS with less than 500 bases were 2,433 and between 2000 and 3000 bases were 2,433 CDS. Numbers of CDS with more than 3000 bases were found to be less in number (1,252). Highest number of CDS (21,505) were found in germinating teliospores as compared to dormant teliospores (16,575) of *Tilletia indica* (Table 5). The high-quality reads of each teliospores data were mapped based on CDSs. 9,680 CDSs were found to be common between dormant and germinating teliospores. 11,825 CDSs were found to be in germinating teliospores while only 91 were unique in dormant teliospores (Fig. 1).

**Table 2 — Statistical summary of RNA sequence analysis**

<table>
<thead>
<tr>
<th>Samples</th>
<th>Library</th>
<th>Number of reads</th>
<th>Number of bases</th>
<th>Total data in GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dormant teliospores of <em>T. indica</em></td>
<td>Paired end</td>
<td>34,341,621</td>
<td>5,174,467,005</td>
<td>5.17</td>
</tr>
<tr>
<td>Germinating teliospores of <em>T. indica</em></td>
<td>Paired end</td>
<td>29,335,701</td>
<td>4,415,443,070</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3 — Distribution of the unigenes according to their length and ranges**

<table>
<thead>
<tr>
<th>Particulars</th>
<th>Pooled teliospores</th>
<th>Ranges of Unigenes (bases)</th>
<th>Pooled teliospores</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of unigenes</td>
<td>16,575</td>
<td>Unigene &lt;500</td>
<td>2,295</td>
</tr>
<tr>
<td>Total unigene length (bases)</td>
<td>28,998,753</td>
<td>500 ≤ Unigene &lt;1000</td>
<td>3,385</td>
</tr>
<tr>
<td>Maximum unigene length (bases)</td>
<td>13,476</td>
<td>1000 ≤ Unigene &lt;2000</td>
<td>5,696</td>
</tr>
<tr>
<td>Minimum unigene length</td>
<td>200</td>
<td>2000 ≤ Unigene &lt;3000</td>
<td>2815</td>
</tr>
<tr>
<td>Mean unigene length (bases)</td>
<td>1749.5</td>
<td>Unigene ≥3000</td>
<td>2,384</td>
</tr>
</tbody>
</table>

**Table 4 — Summary of coding sequences (CDS) statistics**

<table>
<thead>
<tr>
<th>Particulars</th>
<th>Pooled teliospores</th>
<th>Ranges of CDS</th>
<th>Pooled teliospores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of CDS</td>
<td>21,596</td>
<td>CDS &lt;500</td>
<td>3,065</td>
</tr>
<tr>
<td>Total CDS length (bases)</td>
<td>28,955,355</td>
<td>500 ≤ CDS &lt;1000</td>
<td>6,229</td>
</tr>
<tr>
<td>Maximum CDS length</td>
<td>11,169</td>
<td>1000 ≤ CDS &lt;2000</td>
<td>8,619</td>
</tr>
<tr>
<td>Minimum CDS length</td>
<td>297</td>
<td>2000 ≤ CDS &lt;3000</td>
<td>2,433</td>
</tr>
<tr>
<td>Mean CDS length</td>
<td>1340.8</td>
<td>CDS ≥3000</td>
<td>1,252</td>
</tr>
</tbody>
</table>

**Table 5 — Coding sequences in dormant and germinating teliospores of *Tilletia indica***

<table>
<thead>
<tr>
<th>Particulars</th>
<th>Dormant teliospores</th>
<th>Germinating teliospores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of CDSs</td>
<td>16,575</td>
<td>21,505</td>
</tr>
<tr>
<td>Total CDS length(base)</td>
<td>14324370</td>
<td>28884753</td>
</tr>
<tr>
<td>Maximum CDS length</td>
<td>9,744</td>
<td>11,169</td>
</tr>
<tr>
<td>Minimum CDS length</td>
<td>297</td>
<td>297</td>
</tr>
<tr>
<td>Mean CDS length</td>
<td>1,466</td>
<td>1,343.16</td>
</tr>
</tbody>
</table>

2000 bases were 8,619, and between 2000 and 3000 bases were 2,433 CDS. Numbers of CDS with more than 3000 bases were found to be less in number (1,252). Highest number of CDS (21,505) were found in germinating teliospores as compared to dormant teliospores (16,575) of *T. indica* (Table 5). The high-quality reads of each teliospores data were mapped based on CDSs. 9,680 CDSs were found to be common between dormant and germinating teliospores.11,825 CDSs were found to be in germinating teliospores while only 91 were unique in dormant teliospores (Fig. 1).

**Functional annotation**

The predicted CDS were searched against NCBI non-redundant protein database (Nr) using Basic local alignment search tool (BlastX) (E-value-1e-05). It suggested that 14,784 CDS were found in annotation using BLASTX whereas, 6,812 CDS were no BLAST hit significant. The majority of hits were found to be against the fungi *Anthracocystis flocculosa* PF – (woody smut fungi) (Fig. 2).

**Gene ontology analysis**

Gene ontology was performed using the BLAST2GO program. The genes were categorized into different categories (Table 6). Further, GO terms of germinating...
spores of *T. indica* were divided into functions viz. biological process, including cellular metabolic process (35), organic substance metabolic process (33), primary metabolic process (29), nitrogen compound metabolic process (24), establishment of localization (19), single organism cellular process (17), single organism metabolic process (12), biosynthetic process (9), and regulation of cellular process (9) (Fig. 3A). Molecular functions consisted organic cyclic compound binding (241), heterocyclic compound binding (216), transferase activity (158), small molecule binding (119), hydrolase activity (108), and oxido-reductase activity (103) (Fig. 3B). Cellular component included intracellular component (711), intracellular part component (663), intracellular organelle (440), membrane-bounded organelle (302), intracellular organelle part (198), non-membrane-bounded organelle (160), protein complex (136), intrinsic component of membrane (132), and ribonucleoprotein complex (120) (Fig. 3C).

**Pathway analysis**

The pathway analysis suggested the involvement of predicted CDS in several processes in dormant and germinating teliospores of *T. indica* (Table 7). Overall, the maximum numbers of pathways were involved in germinating spores. The pathways were carbohydrate metabolism (283), energy metabolism (167), lipid metabolism (176), nucleotide metabolism (125), amino acid metabolism (253), metabolism of other amino acids (83), glycan biosynthesis and metabolism (64), metabolism of cofactors and vitamins (123), metabolism of terpenoids and polyketides (46), biosynthesis of other secondary metabolites (39), xenobiotics biodegradation and metabolism (53), transcription (171), translation (431), folding, sorting and degradation (282), replication and repair (119), membrane transport (7), signal transduction (263), transport and catabolism (340), cell growth and death (159), cellular community – eukaryote (63), cellular community – prokaryote (90), cell motility (34), environmental adaptation (35). Where as in dormant teliospores there were less number of pathways included viz. carbohydrate metabolism (204), energy metabolism...
(131), lipid metabolism (143), nucleotide metabolism (102), Amino acid metabolism (212), glycan biosynthesis and metabolism (52), metabolism of cofactors and vitamins (105), metabolism of terpenoids and polyketides (30), Biosynthesis of other secondary metabolites (32) and xenobiotics biodegradation and metabolism (35).

Differential gene expression analysis (DESeq)

As many as 9,680 genes were found to be commonly expressed in both the germinating and dormant teliospores. Among them, 686 were found to be upregulated and 114 were downregulated significantly using the DESeq analysis.

In silico identified genes viz. GTPase domain-containing, 60S acidic ribosomal P0, ribosomal sRNA binding domain, 3-oxoacid–transferase, histone H2A-like-1, cytochrome c551, ABC transporter G family member 20, fatty acid desaturase, NAD-dependent succinate-semialdehyde dehydrogenase, methionine aminopeptidase, zinc transporter 5, protein disulfide-isomerase, a ribosomal component of cytosolic 80S ribosome and 60S large subunit, ribosomal protein S17, succinate dehydrogenase iron-sulfur precursor, glycosyl hydrolase involved in high up-regulation in germinating teliospores of *Tilletia indica*. Hypothetical protein TRAVEDRAFT_33922, AAA family CDC48 subfamily, ATP synthase subunit alpha, hypothetical protein A1Q1_04525, alcohol dehydrogenase, hypothetical protein MAJ_10967, partial, potential hexose transporter, hypothetical protein TRAVEDRAFT_33922 were showed down regulation in dormant teliospores of *Tilletia indica*.

Validation of germination related genes in teliospores of *Tilletia indica* using qPCR analysis

Ten genes were used for expression analysis in teliospores of *Tilletia indica*. The gene GTPase domain-containing, designated as Ti 3931 was found high up-regulated with a fold change of 7.88 and 11.15 in both the years of 2015 and 2018 germinating spores, respectively (Fig. 4A). The gene 60S acidic ribosomal
The gene ABC transporter G family member 20, designated as \( Ti\ 7098 \) showed 6.16- and 9.91-fold changes in both years, respectively. The gene for 3-oxoacid –transferase designated as \( Ti7462 \) showed 5.81 and 7.43 fold from the germinating teliospores. Similarly, histone H2A-like-1, designated as \( Ti\ 7522 \) expressed showing about 7 fold changes.

The gene cytochrome c551, chosen as \( Ti\ 9289 \) was upregulated with a fold change of 10 and 8 in 2015 and 2018 germinating teliospores, respectively. The gene ABC transporter G family member 20, designated as \( Ti\ 8670 \) found to have significant upregulation with a fold change of 4.95 and 8.28 in the 2015 and 2018 germinating teliospores, respectively. The gene for fatty acid desaturase, designated as \( Ti\ 7959 \) showed expression with a fold change of 6.34 and 6.52 in the both years, respectively. Similarly, NAD-dependent succinate-semialdehyde dehydrogenase, designated as \( Ti10095 \) expressed a fold change of 6.34 and 7.21 in both the years, respectively (Fig. 4B).

**Discussion**

*Tilletia indica* Mitra is a heterothallic fungi causing Karnal bunt of wheat. In recent times, pathogenomics is playing a significant role in understanding the dynamics of pathogens and to devise management strategies. The teliospores are acting as a primary source of inoculum. Germination of spores is complex in nature. The spores are playing major role for sporadic occurrence of the disease in the north-western plain zone of India. Till now, extraction of teliospores and the understanding mechanism of germination spores is not clearly implicit. Genome wide association mapping revealed the presence of 13 SNPs associated with virulence of Karnal bunt. Very little information is available regarding teliospores germination and dormancy. Keeping in view, the whole transcriptome of germinating and dormant teliospores of *T. indica* Mitra was performed using the RNA seq approach to identify specific genes related to germination of spores.

The present investigation revealed 14,784 CDS were found in the annotation of pooled teliospores whereas 6,812 CDS had no significant BLAST hits. The maximum similarity was matched to be against the fungi *Anthracocystis flocculosa* PF for teliospores, which is a wood smut fungus that belongs to basidiomycetes. In the present study, the KEGG pathway analysis revealed the major role of the carbohydrate metabolism and energy metabolism in germination of teliospores. In earlier reports, carbon flows between glucose and glycogen are playing a crucial role in carbohydrate metabolic reactions. The regulations of carbon and energy metabolism was during the early stages development of *Coprinopsis cinerea*.

Using the differential gene expression analysis, 986 were highly upregulated and 114 were down regulated significantly. The \( Ti3931 \) gene for GTPase domain was found to be significantly upregulated germinating spores. GTPases regulate many essential cellular processes, including actin dynamics, gene transcription, cell-cycle progression so activation of GTPase is most essential step during germination of spores, which in turn act as a molecular switch for regulation of several genes. It was also found to be involved in lipid metabolism and activation of the actin cytoskeleton. \( Ti6828 \) gene related to 60S acidic ribosomal P0 was up regulated (with fold change of 7.06, 9.61 in 2015 and 2018 respectively) in the germinating spores. It is involved in the synthesis of several components of active cell metabolism. Several types of 60S ribosomal proteins were up-regulated in *Puccinia striiformis* f.sp. *tritici* (Pst). Ribosomal protein synthesis germinating and dormant spores of *Dictyostelium discoideum* revealed that ribosomal protein were lacking in dormant spores and, these mRNAs were discovered in germinating spores after 1 h of germination of spores. These findings were consistent with the hypothesis that there was transcriptional regulation of the synthesis of most ribosomal proteins in the process of germination of spores. \( Ti7098 \) related to involvement of small RNA binding domain, in the germinating spores showed higher expression, with higher fold change (6.16 & 9.91 in 2015 and 2018, respectively). Earlier studies revealed that the activity of the enzyme in un-germinated spores was low or absent. It has increased radically when germination started. Small RNAs participate in gene expression regulation at the post-transcriptional level by blocking the translation of target mRNA through antisense binding. \( Ti7462 \) generelated to 3-oxoacid–transferase (fold change 5.81 & 7.43 in 2015 and 2018, respectively), \( Ti7809 \) related to pleiotropic drug resistance (fold change 5.77 & 4.4 in 2015 and 2018, respectively), were found to be up-regulated in the germinating spores.
Until now, no study was available in relation to involvement of these two genes (Ti7462, Ti7809) in the process of germination of spores. Ti7522 related to histone H2A was found to be upregulated significantly in the germinating spores of T. indica. Active cell division is a major activity during the germination process. In Saccharomyces cerevisiae, histone H2A was essential for appropriate centromere–kinetochore function during chromosome segregation in cell division. Hence, the upregulation of histone H2A is significant in T. indica. Ti9289 gene related to cytochrome c551 showed higher expression in germinating spores of T. Indica of the year 2015 and 2018 with fold change of 10.05 and 7.67, respectively. Earlier reports suggested that the upregulation of cytochrome c551 in germinating spores of Schizosaccharomyces pombe was increased rate of spore germination. When the cytochrome c551 gene was disrupted, there was delayed sporulation and spore germination. The significant genes expressed in transcriptional analysis during spore germination of Schizosaccharomyces pombe were identified.

**Conclusion**

This is first study to identify germination-related genes in the teliospores of Tilletia indica using RNA Seq approach. KEGG pathway analysis suggested that maximum number of pathways were involved in germinating spores viz. carbohydrate metabolism, energy metabolism and lipid metabolism as compared to dormant spores. Identified candidate genes viz. Ti3931 gene for GTPase domain; Ti6828 gene related to 60S acidic ribosomal; Ti7098 related to involvement of small RNA binding domain; Ti7462 gene related to 3-oxoacid–transferase and Ti7522 related to histone H2A) having role in the spore germination and evidenced through qPCR analysis. Further, mechanism of teliospores germination and dormancy in T. indica will be understood by functional genomics approach, which could be helpful in developing anti-sporelation strategies. It provides the basis for better strategies to manage Karnal bunt disease of wheat.

**Acknowledgement**

Authors thank the Indian Council of Agricultural Research, New Delhi for financial support in the form of CRP genomics project (ICAR-G/CRP Genomics/2015-2720/IARI-12-151). The first author (HMM) acknowledges PG School, ICAR-IARI for providing fellowship.

**Data availability statements**

Data available in a publicly accessible repository. The data presented in this study are openly available in https://www.ncbi.nlm.nih.gov. The transcriptome data project has been deposited at DDBJ/ENA/GenBank under the accession numbers PRJNA 522347.

**Conflict of interest**

Authors declare no competing interests.

**References**

31 Pietrek M, Lutz M & Yorou NS, A molecular phylogenetic framework for Anthracocystis (Ustilaginales), including five new combinations (inter alia for the asexual Pseudozyma flocculosa), and description of Anthracocystis grodzinskae sp. nov. Mycol Progress, 14 (2015) 88.