In Silico Characterization of Mate Genes in Wheat

Deepika Mohanta (✉ deepikamohnata1995@gmail.com)
  Visva-Bharati University: Visva-Bharati

Sandip Debnath
  Visva-Bharati University: Visva-Bharati

Abstract

Background: Multidrug and toxic compound extrusion (MATE) genes are a group of multidrug efflux transporters that widely exists in all living organisms and play a major role in the detoxification of heavy metals, metalloids, exogenous xenobiotics and endogenous secondary metabolites out of the cells. However, insilico analysis of MATE gene family in plant species is very limited and thus such analysis need to be elucidated in wheat.

Results: We have identified forty-four MATE genes in wheat and categorized into seven families based on their phylogenetic analysis. Further, 43 genes were found to exhibit protein-protein interaction at the protein level by using STRING software. We observed that the maximum number of exons i.e., 14 was identified in genes TraesCS6A02G418800.1 and TraesCS6D02G407900.1. We employed MEME software to find protein motifs associated with the MATE genes where maximum number of motifs were set to 22. Here, the protein motifs among the families 1,2 and 3 were significantly different from the rest. We found that the majority of MATE genes were showing expressions during biotic stress conditions due to disease infestations and the highest level of expression was shown by the gene TraesCS5B02G326600.1 belonging to family 1 which got expressed during *Fusarium* head blight infestation by *Fusarium graminearum* after 4 days of inoculation by using Wheat expression browser tool. A total of 39 ternary plots consisting of homoeologous genes for 39 MATE genes, showing different level of expressions during biotic and abiotic stress conditions were composed, where we found 44 % of the triads tend to show non balanced expressions (extreme values) due to their higher tissue- specificity and greater intensity.

Conclusion: The results obtained from this study indicated that total 44 MATE genes were found to be directly involved in the metabolism of wheat and were expressed during different biotic and abiotic stress conditions. So such genes can be further evaluated for their interaction with heavy toxic metal elements and sequestration from the cells.

Background

Bread wheat (*Triticum aestivum* L.) is one of the three most essential cereal crops globally other than rice and maize, accounting for about 35% of the world's staple food (Han *et al* .,2019). It was the first domesticated crop and the youngest polyploid species amongst all the field crops (Gill *et al*.,2004). The hexaploid wheat consists of A, B and D genomes, each comprising 7 chromosomes with 6 copies of each chromosome (Han et al., 2019) which were derived from three diploid progenitor genomes viz., AA from *T. urartu*, BB from *Ae. speltoides*, and DD from *Ae. tauschii*, and finally originated from hybridization between cultivated tetraploid emmer wheat (AABB, *T. dicoccoides*) and diploid goat grass (DD, *Aegilops tauschii*) (Brenchley *et al*.,2012).

The complex wheat genome offers a unique opportunity to make variations in gene density and evolution within and between plant chromosomes (Lazo *et al*.,2004). The hexaploid wheat genome comprises about 17 Gb genome size, which is five times larger than human, approximately eight times larger than that of maize and forty fold larger than rice (Gill *et al*.,2004).

Plants have advanced detoxification mechanisms to survive against a large number of potentially toxic compounds, which include heavy metals, metalloids, exogenous xenobiotics and endogenous metabolites, mainly secondary metabolites such as, alkaloids, flavonoids, terpenoids, terpenoid-derived phytohormones, cuticle lipids, and mono lignols (Shoji.,2014).

Such compounds get transported following enzymatic modification or synthesis inside the cell and get accumulated in apoplastic cell walls or central vacuoles of the plant cells. Then the membrane transporters widely known as Multidrug and toxic compound extrusion (MATE) proteins actively translocate these diverse range of compounds across various membranes within and outside the cells (Shoji.,2014). MATE transporters belong to a family of cation antiporters and xenobiotic transmembrane transporters occurring in most organisms from prokaryotes to eukaryotes. This family constitutes one of the largest transporter families in plants, with more than 50 identified MATE genes in the Arabidopsis (Takanashi et al.,2014). Furthermore, MATE transporters are involved in a wide variety of physiological functions throughout plant development, transporting a broad range of substrates such as organic acids, plant hormones, which aids in categorization of plant MATE transporters according to their physiological roles and summarizes their tissue specificity, membrane localization, and transport substrates (Takanashi et al.,2014).

These MATE transporters were identified for the first time in *Vibrio parahaemolyticus* and *Escherichia coli* as multidrug efflux proteins and were entitled MATE due to their lack of sequence homology with other transporters (Brown *et al*.,1999; Morita *et al*.,1998). The first plant MATE transporter, i.e, AtALF5 (*Arabidopsis thaliana*) was isolated in 2001 and found to be involved in multidrug resistance (Diener

et al.,2001). The recent reports suggest that the study of MATE gene family and their diverse roles extend to numerous other plant species such as in rice and Arabidopsis (Wang et al.,2016), tomato (Santos *et al.*,2017), soybean (Liu *et al.*,2016), cotton (Lu *et al.*,2018), *Sorghum bicolor* (Sivaguru *et al.*,2013), blueberry (Chen *et al.*,2015), *Camellia sinensis* (Chen *et al.*,2020) etc. but in wheat the study of MATE gene family needs to be elaborated.

Completion of wheat whole genome sequencing in 2018 has opened up the area of gene annotation especially in silico in order to characterize the MATE gene family as well as their application in making gene based molecular markers for plant breeding purposes. In the current study, MATE genes in wheat (*Triticum aestivum* L) were identified and interaction among the genes at protein level was performed. In addition to this, the insilico expression analysis of MATE genes along with their homoeologous candidates in different genomes for various biotic and abiotic stress conditions was studied as well as protein motifs were analyzed.

Result And Discussions

Selection of mate genes and their chromosomal location

The bread wheat genome comprises of approximately 107,891 coding genes and 12,853 non-coding genes (Appels *et al.*,2018), out of which 44 MATE genes were identified with help of Uniprot and Uniprac databases and corresponding nucleotide and protein sequences were obtained (Additional file 1 & 2). The details of the 44 MATE genes including genome and protein ID's, size of the gene and protein, number of exons in each gene sequence as well as their source were presented in the table 2. Three different diploid wheat parents viz., *T. urartu, Ae. speltoides*, and *Ae. tauschii* (contributed A, B, and D genome, respectively) had contributed to the genomic evolution of modern hexaploid wheat (Liu *et al.*,2019). Out of the total MATE genes, 33 genes were displayed based on their descriptive chromosomal locations (Fig. 1) as karyotype view, retrieved from Gramene server (https://ensembl.gramene.org/Triticum_aestivum/Location/Genome?time=1600500223).

Phylogenetic analysis of mate genes in wheat

Studies have revealed that the phylogenetic analyses of membrane transporters were generally inaccurate to condemn specific substrates (Santos *et al.*,2017). However, phylogeny of the MATE family has been represented relatively useful to predict the affinities with potential molecule groups, such as organic acids (citrate), alkaloids (nicotine), and flavonoids (anthocyanin, proanthocyanidin etc.) (Santos *et al.*,2017). Multiple sequence alignments of MATE protein sequences were generally carried out by using ClustalX 2.1 software with its default settings (Peng *et al.*,2012). From that we employed Clustal-Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and phylogeny.fr:TreeDyn (http://www.phylogeny.fr/one_task.cgi?task_type=treedyn) tools to construct a multiple sequence alignment of the identified 44 full-length protein sequences. The complete multiple alignment profiles of protein and nucleotide sequences were used to establish a phylogenetic tree. The Neighbour-joining (N-J) tree topography has been employed to study the MATE gene family (Fig. 2.). Here, we have estimated seven major groups or families of MATE genes by increasing the scale of 0.1 to 0.12 and categorised the results in Table 1.

From the earlier study by Liu *et al., in* 2016, it was confirmed through the genome wide analysis of MATE genes in soybean that 117 MATE genes could be classified into four primary clades or families by using MEGA 6.0 tool and a maximum likelihood (ML) tree was constructed. In maize, a total of 49 recognized ZmMATE genes were found and grouped into seven clusters in the form of Neighbour-joining tree by using Mega5.0 software (Zhu *et al.*,2016). Similarly, to study the comparative evolutionary history as well as relationship of MATE gene family among two species of cotton and Arabidopsis, total 196 putative MATE genes were analysed, out of which 68 genes were of *G. arboreum* (GaMATE), 70 genes of *G. raimondii* (GrMATE) and 58 genes of *Arabidopsis*. Then, based on phylogenetic classification, these MATE genes were categorised into three subfamilies by using NJ method of MEGA6.0 software, where M1 subfamily was the largest one having 124 genes, followed by M2 having 48 genes M3 having 24 genes (Lu *et al.*,2018). From this, it may be concluded that the classification implemented in this study could be accepted like prior findings.

Further, the multiple sequence alignment of 44 genes were represented where variation at the nucleotide level has been exhibited (additional file 3). The nucleotide sequence alignment was performed by using Multalin database (http://multalin.toulouse.inra.fr/multalin/) where 44 MATE genes have been displayed in three colours which were red, blue and black signifying high consensus, low consensus and neutral colours respectively. Here, we noticed deletion regions at several places in the nucleotide sequence alignments which were highlighted in the Fig. 3. These evolutionary footprints are of immense importance for understanding the basic structure of wheat genes. Similarly, according to Ma *et al.*,2011, in rice and *Arabidopsis*; the multiple peptide sequence alignment of the plastocyanin-like domains (PCLDs) determined the conserved amino acids involved in copper binding. In rice,

all the MAPKKKs (Mitogen activated protein kinase kinase kinase) genes, exclusively involved in signal transduction pathways, were classified under Raf, ZIK and MEKK subfamilies and thus the analysis of rice MAPKKKs along with that of *Arabidopsis* for all the Raf, ZIK and MEKK subfamilies was carried out by creating multiple sequence alignments of kinase domains through Multalin program in order to detect specific conserved signature sequences. It was revealed that in subfamily Raf, 43 MAPKKKs in rice and 48 in *Arabidopsis*, in subfamily ZIK, about 10 MAPKKKs in rice were identified whereas in MEKK, 22 MAPKKKs from rice and 21 from *Arabidopsis* were found to be conserved (Rao *et al.*,2010). Also, in blueberry (*Vaccinium corymbosum*), out of total 33 MATE genes, the multiple sequence alignments of 08 novel VcMATE protein sequences along with the selected MATE transporter orthologs were analysed using ClustalX software. Here, it was found that VcMATE 2 shared highest level of identity with the known flavonoid transporters while VcMATE 1 and VcMATE 4 exhibited lowest similarities to the MATE-type flavonoid transporters (Chen *et al.*,2015).

PROTEIN-PROTEIN INTERACTION

It has been documented that, the cells communicate with each other through protein-protein interactions and perform all the physiological processes of life through interactions of various proteins (Szklarczyk *et al.*,2017). We have constructed the association network of protein interactions among 43 genes belonging to *Triticum aestivum* out of total 44 protein sequences (remaining 1 protein sequence i.e, A0A3B6AXC7 comes under *Triticum utaru* and didn't participate in the interaction) by using STRING online program. The primary interaction unit in STRING is mainly the 'functional association', and a link between two proteins that contribute mutually to a specific biological function (Szklarczyk *et al.*,2017). Here, the network (Fig. 4) has been stretched by an additional 20 proteins (through MORE button in STRING Interface) so as to get an extra clear image of the interaction, and the confidence cut-off for screening interaction links has been set to "medium confidence" at 0.4.

This (Fig. 5) determines the results retrieved after entering the set of 43 protein sequences projected to be involved in efflux of toxic multidrug compounds from the cells and tissues of the wheat plant. The network statistics of the set of proteins identified in functional subsystems revealed that number of nodes were 53 with average node degree being 1.55 and PPI enrichment p-value at 3.9e-12. Further, we have also noticed that the protein sequences from the same phylogenetic group were interacted closely and clustered together in the association network.

Similar studies using this approach was seen in *Brassica rapa*, where proteins involved in biosynthesis of camalexin (involved in resistance against *Botrytis cinerea* and *Alternaria brassicae*) were found to be interacted through functional proteins association networks and analysed using STRING database version 10.5. Here, a phytoalexin deficient 3 (PAD3) gene was identified as a key functional node along with CYP71A12 gene as a potential functional partner with which all other multi proteins were found to be associated (Gaur *et al.*,2018). In case of in silico analysis of functional linkage among arsenic induced MATE genes in rice by Seth *et al.*,2019, it was found that 37 MATE genes were found to be interacting at the protein. Also, in rice (*Oryza sativa ssp. japonica*), around 30 genes were identified from AbS (Abiotic Stress) responsive gene family, involved in stress responsive signalling during various abiotic conditions like drought, submergence, cold, salinity, metal toxicity etc., and out of these 30 seed −proteins, 22 genes were found to be extensively involved in protein protein interaction network along with the extra 34 derived neighbours by using String software, showing closely related functional modules and complexity of AbS (Muthuramalingam *et al.*,2017).

STRUCTURAL ANALYSIS OF MATE GENES IN WHEAT

The phases and structures of exons / introns of the MATE genes was examined by using EnsemblPlant software (https://plants.ensembl.org/Triticum_aestivum/Transcript/Summary). This analysis provides more insight into evolution of gene structures in wheat which provides detailed information regarding the transcript and translation length, number of coding exons, amino acids and the base pairs along with the transcript diagram. There were 44 exon intron transcripts that have been identified (Additional file 4).The maximum number of exons present in the following gene sequences were 14 which exist in genes TraesCS6A02G418800.1 and TraesCS6D02G407900.1 while the least number of exons that is only 1 was available in the genes sequences viz., TraesCS1A02G188100.1.cds1,TraesCS5B02G562500.1.cds1, TraesCS6A02G256400.1.cds1 and TraesCS6D02G384300.1.cds1.

Fan *et al.*,2014, suggested that on the basis of phylogenetic tree analysis, MATE genes were mainly grouped in to three subfamilies and the intron-exon structures were subfamily-specific, indicating the cotton MATE genes were considerably conserved and functionally diversified.

PROTEIN MOTIFS

Generally protein structures have conserved elements called motifs, which have a sufficient influence on the function of proteins (Conklin.,1995). The function of proteins usually imposes tight constraints on the evolution of specific regions of protein structure residues directly or indirectly in a function and often clustered in a short sequence motif (signature, pattern, framework or fingerprint) that is conserved across the various proteins sharing that function (Manning *et al.*,1998).The online software Multiple EM for Motif Elicitation (MEME) was employed to analyse the motifs in MATE proteins (Bailey et al., 2015). These conservative protein motifs in wheat MATE protein gene sequences were identified and predicted using MEME tool where the maximum number of motifs was set at 22. These were then arranged according to their four families. The motifs of wheat MATE protein were shown as coloured boxes, each motif represented as a number in the coloured box. They were listed according to the families 1 to 7 from the phylogenetic tree (Additional file 5). Zhu *et al.*,2016, observed that, usually, it is perceived that most of the closely related members within the same family were having common motif compositions, indicating their functional similarities. In this case, a total maximum of 22 conserved motifs were identified and represented as the different coloured boxes as symbols for different motif consensus. The types and sequences of the protein motifs among the families 1, 2 and 3 were significantly different from the rest. Further, we have noticed that the family 3 and 5 were having 19 and 14 number of protein motifs respectively while the rest have 22. Here, it may be concluded that the interacting MATE genes or protein sequences within a family are also having similar protein motifs.

An early study had defined that motifs were the short DNA or protein sequence which contribute towards the biological functions of the sequences in which they resides where they become one of the basic functional units of molecular evolution (Grant *et al.*,2011). Similar work has been done by Liu *et al.*,2016, in genome-wide analysis of MATE transporters in soybean where they have analysed maximum 12 conserved motifs in which identical type of motif sequences were present in the first three families and significantly different were in the fourth family with very less number of motifs by using MEME software.

IN SILICO EXPRESSION ANALYSIS AND THEIR HOMOEOLOGOUS CANDIDATES

Wheat is having MATE transporter proteins which are responsible for controlling different expressions and functions during vegetative growth, reproductive development, senescence as well as resistance to biotic and abiotic stresses similar to the other plants. The possible roles of 44 homologous MATE genes in plant growth and development in wheat has been constructed by heatmaps using wheat expression browser (http://www.wheat-expression.com/). The MATE genes showed their highest level of transcript accumulation in leaves, roots, shoots, flowers, grain, spikes etc. as shown in the Fig. 5 and thus indicated that they were involved in the development of all tissues or organs under normal conditions. The heatmaps with constitutive gene expressions of 44 MATE genes as in Fig. 6.were generated under different biotic and abiotic stress conditions taking into account such as-

1. Phosphate starvation in roots and shoots
2. Heat and drought stress time course in seedlings
3. Spikes with water stress
4. *Fusarium* head blight infected spikelets
5. Stripe rust infected seedlings
6. *Septoria tritici* infected seedlings
7. Powdery mildew time course of infection in seedlings
8. PAMP inoculation of seedlings
9. Time course of spikelets inoculated with *Fusarium* head blight/ABA/GA
10. Coleoptile infected with *Fusarium pseudograminearum* (crown rot)
11. *Zymoseptoria tritici* infected seedlings
12. Seedlings with peg (polyethylene glycol) to manage stress
13. Shoots from NILs segregating for crown rot resistance

The degree of expressions of genes in the tissues at a particular stage was depicted by the extent of their intensities which was expressed through expression units – tpm log2 (transcripts per million).

The heat maps of gene expressions of these 7 families were obtained where all the genes were displayed according to their phylogenetic associations (Additional file 6). Family 1 contained two genes with constitutive, high expressions which were TraesCS5B02G326600.1 and TraesCS4A02G245300.1 with tpm values 6.99 and 5.4 out of 7 Log☐ (tpm) which were highly expressed in fifth leaf blade, spikes, rachis, anthers for disease infestation such as Fusarium head blight and stripe rust respectively. While the genes

TraesCS3B02G563400.1 and TraesCS5B02G245500.1 had lowest expression in the group as they were showing least or even no expression in high as well as intermediate level of biotic and abiotic stress conditions. Similarly, the family 2 contains five genes showing higher expressions in mainly reproductive stages at spikes, rachis and during *Fusarium* infestation. These were TraesCS2B02G296000.1, TraesCS2D02G277400.1, TraesCS5D02G378200.1, TraesCS5D02G378300.1 and TraesCS2B02G296100.1 having tpm values 6.26, 6.35, 5.35, 6.45 and 5.13 respectively. The negligible expressions were displayed by the gene TraesCS5D02G413800.1. In family 4, two members out of three were constitutively expressed, although at mid to low levels in most of the tissues such as grain, spike, leaves and other shoot parts mainly at reproductive stage and during disease infestation except the gene TraesCS1A02G188100.1, showing no expression at all. In family 5, only one gene (TraesCS5B02G562500.1) is present which shows maximum degree of expressions at 24 days after sowing during 10th day of phosphorus starvation and other abiotic stress particularly at roots. Here, in family 6, we can notice that out of the two, only TraesCS4B02G244400.1 is highly expressive during reproductive stage at anthers and spikelets with maximum tmp value of 4.52. Unlike other families, all the MATE gene members in family 3 and 7 were constitutively expressed with varying transcriptional intensities with their tmp values ranging between 3 to 4 at both vegetative and reproductive stages in various tissues of the plant during stress conditions.

As all these MATE genes were located in the integral part of the plasma membrane, they were exclusively involved in molecular functions like solute - solute antiporter activity, xenobiotic transmembrane transporter activity and transmembrane transporter activity (Table 5). The results obtained suggested that different MATE genes were showing expressions during various biotic and abiotic stress conditions but majority of the genes were exhibiting expressions during biotic stress conditions due to disease infestations (Table 4), where the overall highest level of expressions has been shown by gene TraesCS5B02G326600.1 belonging to family 1, expressing during disease infestation of *Fusarium* head blight by *Fusarium graminearum* after 4 days of inoculation.

Lu *et al.*, in 2018, had analysed the expression profiles of *GaMATEs* and *GrMATEs* genes belonging to two species of cotton viz., *G. arboreum* and *G. raimondii*. The study was carried out in root tissues in order to examine the expressions levels of genes in roots tissues under abiotic stress conditions of drought, salinity and Cd stress. Out of the total MATE genes, *GrMATE54, GrMATE53 and GaMATE21 were found to be highly expressed during these three abiotic stress conditions* and were also involved in vacuolar sequestration and toxin effluxers while *GrMATE34* and *GaMATE54 found significantly expressed during stress conditions were involved in ABA transporting*. Similarly, in Soybean, out of 117 MATE genes, expression profiles of 113 MATE genes were constructed through heatmap by using MeV 4.9 software, which were differentially expressed in nine tissues viz., leaf, stem, flower, pod, seed, root, root hair, nodule and shoot apical meristem. These GmMATE genes exhibited tissue specific expressions such as GmMATE107 and GmMATE27 showed highest expression level in roots, root hairs and nodules while least expressions in above ground tissues. Similarly, GmMATE44, GmMATE81 and GmMATE36 showed high level of expressions in pods and developing seed while GmMATE62 and GmMATE7 were expressed in leaf tissues (Liu *et al.*,2016). In *Medicago truncatula*, out of total all the MATE genes, UGT78G1, MaT4, MaT5, and MATE2 were found to be expressed in various parts of the plant such as leaves, roots and flower but MATE2 gene had shown highest level expressions in flowers, followed by roots, vegetative buds, leaves, and seeds and was associated with the transport of glycosylated flavonoids. It has been reported that the gene was exclusively involved in the pigmentation of anthocyanin compound and thus lack of this pigment resulted in discolouration in leaves and flowers (Zhao *et al.*,2011). Tiwari *et al.*,2014, had also revealed that in genome-wide expression analysis of rice MATE genes, two arsenic responsive genes OsMATE1 and OsMATE2 were taken for functional study in transgenic lines of Arabidopsis, where majority of the expressions were shown in leaf, seed and flower morphology, pattern of rosette arrangement and flowering time. These OsMATEs were found to regulate plant growth and development in transgenic lines but their expressions were showing more susceptibility to the biotic and abiotic stresses as compared to the wild types.

We know that wheat is an allopolyploid having two (tetraploid wheat with two homoeoalleles) or three (hexaploid wheat with three homoeoalleles) homologous sub genomes. The homoeoalleles of a gene in polyploid wheat having higher affinity in DNA sequence and function, makes the gene cloning and functional analysis a challenging task (Zhang *et al.*,2018). The polyploidy that arises from whole genome interspecific hybridisation or duplication is present ubiquitously across the plant and fungal kingdom and thus the existence of extremely related genes in polyploids known as homoeologs has promoted the domestication and adaptation of many major polyploidy crops like hexaploid bread wheat (*Triticum aestivum*; AABBDD sub genome), cotton, coffee etc., (González *et al.*,2018).

The factors showing probable reasons for the variation in gene density in wheat chromosomes within homologous and sets could be due to the variation in genetic density or number of loci is the evolutionary history of the specific chromosome genomes, variation in individual size of chromosome and the structural aberrations comprising unequal exchanges of genetic material within and among chromosomes (Qi *et al.*,2004). By understanding in what way these homoeologous genes interactions effect the gene expressions, will

ultimately help to build strategies so as to improve the crops by targeting and manipulating individual or multiple homoeologs to quantitatively modulate trait responses (Borrill *et al.*,2015).

All the possible homologous genes for the 44 MATE genes were found with help of EnsemblPlant database and displayed in the form of ternary plots through wheat expression browser software and listed in table 3. Here, we have found that the ternary plot shows two homologous genes from different species like *Azhumaya* wheat (TraesCS1A02G188100.1) and Chinese spring wheat (TraesCS1B02G195900.1) of *Triticum aestivum* (TraesCS1D02G188200.1) as shown in Fig. 7.and their level of expressions indicates their transporting roles of alkaloids in tissues such as leaves, roots, rachis, spikes, coleoptiles etc. in different biotic and abiotic stress conditions like stripe rust, powdery mildew, heat and cold stress etc.

Similarly, from Table 4 and Additional file 7, we can analyze that the homoeologous genes of the identified wheat MATE genes TraesCS1A02G305200.1, TraesCS2B02G247700, TraesCS2D02G277400.1, TraesCS3B02G298700.1, TraesCS4B02G244400.1, TraesCS5B02G326600.1 and TraesCS2B02G296000.1 were exhibiting low to medium level of expressions of Abiotic stress in case of High level stress-disease, ranging from 20–50% where the maximum expression was displayed by the homoeologous gene TraesCS3A02G265100.1 i.e., 63.62% of the MATE gene TraesCS3B02G298700.1, while the in case of biotic stress like disease under High level stress-disease, the maximum expressions were shown by the gene TraesCS5D02G355500.1 among other homologous gene with 69.79 %.

Further, we have noticed the maximum level of expressions of genes i.e., 100% at stress disease condition. For example, the homoeologous genes such as TraesCS1B02G315900.1, TraesCS7D02G488000.1, TraesCS2A02G222300, TraesCS2D02G277400.1, TraesCS5B02G371200.1, TraesCS7D02G488000.1and TraesCS2D02G277400 belonging to MATE genes TraesCS1A02G305200.1, TraesCS7A02G500700.1, TraesCS2B02G247700, TraesCS2D02G277400.1, TraesCS5D02G378300.1, TraesCS7D02G488000.1 and TraesCS2B02G296000.1 respectively were highly expressed during stripe rust mixture 6/14 days. In addition to this, some other homologous genes such as TraesCS5B02G371200.1, TraesCS1A02G305200.1, TraesCS5D02G378200.1 and TraesCS3A02G499200 of MATE genes TraesCS5D02G378300.1, TraesCS1A02G305200.1, TraesCS5D02G378200.1and TraesCS3B02G562400.1 were also expressed during *Zymoseptoria tritici* inoculation 4 days, *Septoria tritici* 10 days, Stripe rust pathogen 87/66 9days and Flg22 500nM respectively.

Moreover, in case of intermediate stress, the homologous genes have revealed higher level of expressions such as the homologous genes TraesCS2B02G247700 of MATE gene TraesCS2B02G247700 exhibits 94.10 % expression level during stripe rust. Similarly, other homologous genes TraesCS5B02G371200.1, TraesCS5D02G378200.1 and TraesCS5D02G378200.1 of MATE genes TraesCS5D02G378300.1, TraesCS5D02G378300.1 and TraesCS6B02G383300 were expressed during intermediate stress at PAMP Chitin, *Fusarium pseudograminearum* study 2 and cold 2 weeks by 99.24%, 96.39% and 97.98 % respectively. Whereas the maximum level of expression at PAMP flg22 intermediate stress was shown by the homologous gene TraesCS3A02G499200 of MATE gene TraesCS3B02G562400.1 with 100% value.

In this study, out of the total 44 MATE genes, 39 ternary plots consisting respective homologous genes were traced out showing relatively different expression levels during biotic and abiotic stress for 39 genes where 56% triads (A, B and D homoeologs) show balanced expressions and 44 % of triads exhibit nonbalanced expression, being more tissue specific as well as of greater expressivity towards stress conditions. Here, each circular dot signifies a gene triad with an A, B, and D coordinates comprising relative contribution of each homoeolog to the overall triad expression.

Similar type of work had been done by Medina *et al.*,2014, in *Aspergillus flavus*, where the effect of climate change was studied on *Aspergillus flavus* and its aflatoxin B1 production by obtaining ternary plot diagrams. Besides this, according to the study of transcriptional landscape of polyploid wheat, ternary plots were showing relative expression abundance of 16,746 syntenic triads for 50,238 genes in hexaploid wheat during combined analysis of 15 tissues from Chinese spring. Further, it has been noticed that 70% of triads possessing A, B and D homoeologs showed balanced expressions among other homoeologs and were universally expressed and 30 % exhibited non balanced expressions and were more tissue specific (González *et al.*,2018).

Conclusions

In the present investigation, we have concluded that a total 44 MATE genes of *Triticum aestivum* were analysed for phylogenetic classification, protein-protein interaction among the genes, structural and functional analysis of genes, protein motifs as well as in silico expression analysis. The 44 MATE genes were further classified into seven families and a representative phylogenetic tree was

constructed using Clustal omega and Phylogeny.fr:TreeDyn tools. Out of these 44 genes, 43 genes were found to be interacting at the protein level by using STRING software with a medium confidence value at the protein level indicating that these genes were moderately interacted at the protein level. The maximum numbers of exons of 14 were found to be present in genes TraesCS6A02G418800.1 and TraesCS6D02G407900.1. We employed MEME software to find protein motifs associated with the MATE genes and a total of 22 conserved motifs were identified where the protein motifs among the families 1, 2 and 3 are significantly different from the rest and also family 3 and 5 were having 19 and 14 number of protein motifs respectively while the rest had 22. We found that the majority of MATE genes were showing expressions during biotic stress conditions due to disease infestations and the highest level of expression was shown by the gene TraesCS5B02G326600.1 belonging to family 1 which got expressed during *Fusarium* head blight infestation by *Fusarium graminearum* after 4 days of inoculation by using Wheat expression browser tool. A total of 39 ternary plots consisting of homoeologous genes for 39 MATE genes were constructed using Wheat expression browser tool showing different level of expressions during biotic and abiotic stress conditions. We further found that 44 % of the triads tend to show non balanced expressions (extreme values) due to their higher tissue- specificity and greater intensity.

Methods

Identification of MATE genes and their chromosomal location in wheat genome

The MATE nucleotide as well as protein sequences of wheat were obtained and downloaded from the UniProt and UniPrac databases (https://www.uniprot.org/). The UniProt is a universal protein resource which is comprehensive, high-quality and freely accessible database of protein sequence and functional information and UniPrac is a non-redundant database which stores each unique sequence only once and provides a stable and unique identifier (ID), thus making it possible to identify the same protein from different database sources (Apweiler *et al.,*2004). The peptide sequences were converted to their corresponding nucleotide sequences by using EMBOSS backtranseq tool (https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/). The chromosomal position of the 33 MATE genes of wheat were obtained and the karyotypic view displayed (fig.1) with relative distances from Gramene database (https://www.gramene.org/) through genome browser of IWGSC (*Triticum aestivum*).

Phylogenetic analysis and classification of MATE genes

The full-length nucleotide and amino acid sequences of MATE genes (Additional file 1 & 2) were used for phylogenetic analysis. The sequences were subjected to multiple sequence alignments by Clustal Omega software (https://www.ebi.ac.uk/Tools/msa/clustalo/) with the default parameters, in order to retrieve the newick tree data format. An unrooted neighbour-joining phylogenetic tree was constructed using TreeDyn-Phylogeny.fr tool (http://www.phylogeny.fr/one_task.cgi?task_type=treedyn) at a scale value of 0.12 with the default parameters. The variation in nucleotide sequence was examined through multalin browser (http://multalin.toulouse.inra.fr/multalin/) where the conserved nucleotide residues were displayed in red, blue and blaack colour according to their degree of conservation (Additional file 3).

Protein – protein interaction

The protein –protein interactions are essential to each and every aspect of cellular functions and are characterised as transient or stable. Therefore, a comprehensive knowledge of protein interactions is a crucial source of information to functionally interpret the proteins and to understand the model cellular processes on a genome-wide level (Uhrig.,2006). The Information of functional interactions between the expressed proteins of MATE genes in wheat is achieved by using STRING (Search tool for retrieval of interacting genes/proteins) database which helps to collect and integrate the information by consolidating the already known as well as predicted protein–protein link data for a large number of organisms. These interactions could be direct interactions as physical or as indirect functional interactions (Szklarczyk *et al.,* 2017). The network view of this database determines the network of predicted interaction for a specific cluster of proteins where the nodes are indicated as proteins and the edges signify predicted functional connections The number of links at each node is called its "degree" (Uhrig.,2006). Moreover, in evidence mode, an edge can be drawn with up to seven contrarily coloured lines which symbolize the existence of seven types of evidence used in predicting the connections (Seth *et al.,*2019).The confidence mode shows thickness of the indicated line as the degree of confidence prediction of the interaction while the action mode represents the additional information about the prediction (Seth *et al.,*2019).In addition to this, the confidence score denotes the estimated probability that a predicted link occurs between two enzymes in the same metabolic map in KEGG database. The lower score indicates more interaction along with more false positives. Here, the interactions which are above the minimum required score are only included in the predicted network while the maximum number of interactions can be chosen, however, the output limit has been set to the 10 best – scoring hits by default (Seth *et al.,*2019).The analysis component gives information of the inferred network, such as the number of nodes

and edges where the average node degree determines the number of interactions (at score threshold) that a protein has on average in the network. Here, the clustering coefficient determines the measure of connected nodes in the network where the highly connected networks show higher values.

Structural organisation and protein motifs

The exon-intron organisation of the *wheat MATE* family genes were retrieved based on their nucleotide transcript gene ID, and a diagram was obtained using EnsemblPlant database (https://plants.ensembl.org/Triticum_aestivum/Transcript/Summary?db=core;g=TraesCS1A02G305200;r=1A:497861545-497869619;t=TraesCS1A02G305200.1). The transcript determines number of exons, number of domains and features, associated variant alleles and oligo probe maps.

The motifs of MATE proteins were retrieved by using the Multiple Expectation Maximization for Motif Elicitation (MEME) (http://meme-suite.org/), and a representative diagram of protein motifs of each MATE protein was presented according to the default parameters, and the maximum number of motifs was set at 22.The MEME Suite delivers a large number of proteomic and genomic sequence databases for motif scanning and various motif databases for the motif comparison (Bailey *et al.,* 2015). It is composed of web - based integration of tools and database sets for executing motif-based amino acid sequence analyses. Moreover, the Suite having unified web server interface enables the users to implement four sorts of motif analysis viz. motif discovery, motif-motif database searching, motif-sequence database searching and assignment of function.

In Silico Expression Analysis and their homoeologous candidates

The understanding of expression patterns in specific tissues and organs suggests the molecular clues for their role and support in functionality of plants (Bhati *et al.,*2015).
The widely accessible wheat RNA-Sequence datasets which are utilised for producing gene expression levels were obtained from the Wheat Expression Browser software (http://www.wheat-expression.com). It is an expression database for polyploid wheat which helps us to visualize identified MATE gene expression profiles in wheat. This tool is helpful to analyse and relate homologous-specific transcript profiles across a wide range of tissues from different developmental stages in polyploid wheat which could be further queried by Heatmap, BLAST or by searching for an identified gene of interest by name or functional domain (Borrill *et al.,*2016).

Different studies taken into consideration that are based on parameters like study, tissues, high level tissues, age, high level age, stress - disease, high level stress - disease, variety, high level variety, intermediate and intermediate stress. Here, the results were concised according to high expression levels in tissues as grains, roots, leaves or shoots, rachis and spikes. The heat map was drawn in Log_2 (tpm) expression values according to the standardization. Different levels of expressions involves low (0-3 tpm), medium (4 tpm), and high (6-7 tpm) (Borrill *et al.,*2016)

The genome of hexaploid wheat (*T. aestivum*) is the result of multiple series of genome hybridization (containing three sub genomes i.e. A, B and D, which come from three ancestral grasses (*T. utaru*, *A. speltoids* and *A. tauschii* respectively), for which homoeologous gene distribution and rearrangement can contribute to the understanding of evolution of hexaploid species (Zhou *et al.,*2020).

The homoeologous genes of wheat MATE genes were identified by using EnsemblPlant and wheat expression browser software. The EnsemblPlant database determines the target % id which signifies the percentage of the homoeologous sequence matching the source sequence as well as query % id which signifies the percentage of the source sequence matching the homoeologous sequence of each homologous gene (table 2), while wheat expression browser software is helpful for identifying ternary plots of homologous genes associated with each MATE gene (Additional file 7). The comparative expressions of each homoeolog were determined by the triad's position in the ternary plot for the analysis of individual tissues as well as for different biotic and abiotic stress conditions.

Variation in homoeologs expressions across tissues (stable and dynamic triads)

The preference of variation in homoeologs expression in each ternary plot across the different tissues, was calculated and each individual tissue in which the triad was measured is expressed (González *et al.,*2018). The average of distances was defined as the "triad mean distance" and the triads were ranked by their triad mean distance and the percentile was calculated by
$$percentile_i = truncate\left(\frac{rank\,(cmd_i)}{length\,(CMD)}\right),$$ where CMD is vector containing all the triad mean distance and the first and last deciles were categorized as stable 10% and dynamic 10% triads, respectively (González *et al.,*2018).

Relative expression levels of the A, B, and D sub genome homoeologs across triads

The analysis is mainly focused exclusively on the triad which have a 1:1:1 correspondence across the three homoeologous sub genomes, a triad can be defined as expressed when the summation of the A, B, and D sub genome homoeologs was > 0.5 TPM. For standardizing the relative expression of each homoeolog across the ternary plot, the absolute TPM for each gene within the triad is normalised as follows (González *et al.,*2018) –

$$\text{expression}_A = \frac{\text{TPM}(A)}{\text{TPM}(A) + \text{TPM}(B) + \text{TPM}(D)}$$

$$\text{expression}_B = \frac{\text{TPM}(B)}{\text{TPM}(A) + \text{TPM}(B) + \text{TPM}(D)}$$

$$\text{expression}_D = \frac{\text{TPM}(D)}{\text{TPM}(A) + \text{TPM}(B) + \text{TPM}(D)}$$

where A, B, and D represent the gene corresponding to the A, B, and D homoeologs in the triad and the normalized expression was calculated for each one of the intermediate tissues and for the average across all expressed tissues ("combined analysis"). The values of the comparative contributions of each sub genome per triad were used to plot the ternary diagrams using the R package ggtern (Hamilton and Ferry*.,* 2018).

Abbreviations

ABA: Abscisic acid; GA: BLAST: Basic local alignment search tool; EMBOSS: European Molecular biology open software suite; Exp: Expression; expVIP: Expression Visualisation and Integration Platform; Gibberellic acid; HS: Heat stress; Id: Identification or identifier; IWGSC: The International Wheat Genome Sequencing Consortium; MAPKKKs: Mitogen activated protein kinase kinase kinase; MATE: Multidrug and toxic compound extrusion; MEGA: Molecular evolutionary genetics analysis; MEME: Multiple expectation maximization for motif elicitation; MeV: Multiple experiment viewer; ML: Maximum likelihood; Multalin: Multiple sequence alignment; N-J: Neighbour-joining; PAMP: Pathogen- associated molecular pattern molecules; PCLDs: Plastocyanin-like domains; PPI: Protein-protein interaction; STRING: Search tool for retrieval of interacting genes/proteins Tpm: Transcripts per million; UniProt: Universal protein resource

Declarations

Not applicable.

Author details

[1]sandip.debnath@visva-bharati.ac.in

Department of Genetics & Plant Breeding, PSB (Institute of Agriculture)Visva-Bharati University, Sriniketan, 731236,India

References

1. Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., IWGSC whole-genome assembly principal investigators:, Pozniak, C. J., Stein, N., Choulet, F., Distelfeld, A., Eversole, K., Poland, J., Rogers, J., Ronen, G., Sharpe, A. G., Whole-genome sequencing and assembly:, Pozniak, C., … Uauy, C. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science (New York, N.Y.), 361(6403), eaar7191. https://doi.org/10.1126/science.aar7191

2. Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., & Yeh, L. S. (2004). UniProt: the Universal Protein knowledgebase. Nucleic acids research, 32(Database issue), D115–D119. https://doi.org/10.1093/nar/gkh131

3. Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. Nucleic acids research, 43(W1), W39–W49. https://doi.org/10.1093/nar/gkv416

4. Bhati, K. K., Sharma, S., Aggarwal, S., Kaur, M., Shukla, V., Kaur, J., Mantri, S., & Pandey, A. K. (2015). Genome-wide identification and expression characterization of ABCC-MRP transporters in hexaploid wheat. Frontiers in plant science, 6, 488. https://doi.org/10.3389/fpls.2015.00488

5. Borrill, P., Adamski, N., & Uauy, C. (2015). Genomics as the key to unlocking the polyploid potential of wheat. The New phytologist, 208(4), 1008–1022. https://doi.org/10.1111/nph.13533

6. Borrill, P., Ramirez-Gonzalez, R., & Uauy, C. (2016). expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. Plant physiology, 170(4), 2172–2186. https://doi.org/10.1104/pp.15.01667

7. Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L., D'Amore, R., Allen, A. M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M. C., Sehgal, S., Gill, B., Kianian, S., … Hall, N. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature, 491(7426), 705–710. https://doi.org/10.1038/nature11650

8. Brown, M. H., Paulsen, I. T., & Skurray, R. A. (1999). The multidrug efflux protein NorM is a prototype of a new family of transporters. Molecular microbiology, 31(1), 394–395. https://doi.org/10.1046/j.1365-2958.1999.01162.x

9. Chen, G., Liang, H., Zhao, Q., Wu, A. M., & Wang, B. (2020). Exploiting MATE efflux proteins to improve flavonoid accumulation in Camellia sinensis in silico. International journal of biological macromolecules, 143, 732–743. https://doi.org/10.1016/j.ijbiomac.2019.10.028

10. Chen, L., Liu, Y., Liu, H., Kang, L., Geng, J., Gai, Y., & Li, Y. (2015). Identification and expression analysis of MATE genes involved in flavonoid transport in blueberry plants. PloS one, 10(3), e0118578.

11. Conklin, D. (1995). Macine discovery of protein motifs. Machine learn, 21, 1-2, 125-150. https://doi.org/10.1007/BF00993382

12. Diener, A. C., Gaxiola, R. A., & Fink, G. R. (2001). Arabidopsis ALF5, a multidrug efflux transporter gene family member, confers resistance to toxins. The Plant cell, 13(7), 1625–1638. https://doi.org/10.1105/tpc.010035

13. Fan, K., Wang, M., Miao, Y., Ni, M., Bibi, N., Yuan, S., Li, F., & Wang, X. (2014). Molecular evolution and expansion analysis of the NAC transcription factor in Zea mays. PloS one, 9(11), e111837. https://doi.org/10.1371/journal.pone.0111837

14. Gaur, M., Tiwari, A., Chauhan, R. P., Pandey, D., & Kumar, A. (2018). Molecular modeling, docking and protein-protein interaction analysis of MAPK signalling cascade involved in Camalexin biosynthesis in Brassica rapa. Bioinformation, 14(4), 145–152. https://doi.org/10.6026/97320630014145

15. Gill, B. S., Appels, R., Botha-Oberholster, A. M., Buell, C. R., Bennetzen, J. L., Chalhoub, B., Chumley, F., Dvorák, J., Iwanaga, M., Keller, B., Li, W., McCombie, W. R., Ogihara, Y., Quetier, F., & Sasaki, T. (2004). A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium. Genetics, 168(2), 1087–1096. https://doi.org/10.1534/genetics.104.034769

16. González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., Khedikar, Y., Robinson, S. J., Cory, A. T., Florio, T., Concia, L., Juery, C., Schoonbeek, H., Steuernagel, B., Xiang, D., Ridout, C. J., … Uauy, C. (2018). The transcriptional landscape of polyploid wheat. Science (New York, N.Y.), 361(6403), eaar6089. https://doi.org/10.1126/science.aar6089

17. Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics (Oxford, England), 27(7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064

18. Han, Z., Liu, Y., Deng, X., Liu, D., Liu, Y., Hu, Y., & Yan, Y. (2019). Genome-wide identification and expression analysis of expansin gene family in common wheat (Triticum aestivum L.). BMC genomics, 20(1), 101. https://doi.org/10.1186/s12864-019-5455-1

19. Lazo, G. R., Chao, S., Hummel, D. D., Edwards, H., Crossman, C. C., Lui, N., Matthews, D. E., Carollo, V. L., Hane, D. L., You, F. M., Butler, G. E., Miller, R. E., Close, T. J., Peng, J. H., Lapitan, N. L., Gustafson, J. P., Qi, L. L., Echalier, B., Gill, B. S., Dilbirligi, M., … Anderson, O. D. (2004). Development of an expressed sequence tag (EST) resource for wheat (Triticum aestivum L.): EST generation, unigene analysis, probe selection and bioinformatics for a 16,000-locus bin-delineated map. Genetics, 168(2), 585–593. https://doi.org/10.1534/genetics.104.034777

20. Liu, J., Li, Y., Wang, W., Gai, J., & Li, Y. (2016). Genome-wide analysis of MATE transporters and expression patterns of a subgroup of MATE genes in response to aluminum toxicity in soybean. BMC genomics, 17, 223. https://doi.org/10.1186/s12864-016-2559-8

21. Liu, X., Liu, Z., Niu, X., Xu, Q., & Yang, L. (2019). Genome-Wide Identification and Analysis of the NPR1-Like Gene Family in Bread Wheat and Its Relatives. International journal of molecular sciences, 20(23), 5974. https://doi.org/10.3390/ijms20235974

22. Lu, P., Magwanga, R. O., Guo, X., Kirungu, J. N., Lu, H., Cai, X., Zhou, Z., Wei, Y., Wang, X., Zhang, Z., Peng, R., Wang, K., & Liu, F. (2018). Genome-Wide Analysis of Multidrug and Toxic Compound Extrusion (MATE) Family in Gossypium raimondiiand Gossypium arboreum and Its Expression Analysis Under Salt, Cadmium, and Drought Stress. G3 (Bethesda, Md.), 8(7), 2483–2500. https://doi.org/10.1534/g3.118.200232

23. Ma, H., Zhao, H., Liu, Z., & Zhao, J. (2011). The phytocyanin gene family in rice (Oryza sativa L.): genome-wide identification, classification and transcriptional analysis. PloS one, 6(10), e25184. https://doi.org/10.1371/journal.pone.0025184

24. Manning, C. G., Wu, T. D., & Brutlag, D. L. (1998). Highly specific protein sequence motifs for genome analysis. Proceedings of the National Academy of Sciences of the United States of America, 95(11), 5865–5871. https://doi.org/10.1073/pnas.95.11.5865

25. Medina, A., Rodriguez, A., & Magan, N. (2014). Effect of climate change on Aspergillus flavus and aflatoxin B1 production. Frontiers in microbiology, 5, 348. https://doi.org/10.3389/fmicb.2014.00348

26. Morita, Y., Kodama, K., Shiota, S., Mine, T., Kataoka, A., Mizushima, T., & Tsuchiya, T. (1998). NorM, a putative multidrug efflux protein, of Vibrio parahaemolyticus and its homolog in Escherichia coli. Antimicrobial agents and chemotherapy, 42(7), 1778–1782. https://doi.org/10.1128/AAC.42.7.1778

27. Muthuramalingam, P., Krishnan, S. R., Pothiraj, R., & Ramesh, M. (2017). Global Transcriptome Analysis of Combined Abiotic Stress Signaling Genes Unravels Key Players in Oryza sativa L.: An In silico Approach. Frontiers in plant science, 8, 759. https://doi.org/10.3389/fpls.2017.00759

28. Peng, X., Zhao, Y., Cao, J., Zhang, W., Jiang, H., Li, X., Ma, Q., Zhu, S., & Cheng, B. (2012). CCCH-type zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments. PloS one, 7(7), e40120. https://doi.org/10.1371/journal.pone.0040120

29. Qi, L. L., Echalier, B., Chao, S., Lazo, G. R., Butler, G. E., Anderson, O. D., Akhunov, E. D., Dvorák, J., Linkiewicz, A. M., Ratnasiri, A., Dubcovsky, J., Bermudez-Kandianis, C. E., Greene, R. A., Kantety, R., La Rota, C. M., Munkvold, J. D., Sorrells, S. F., Sorrells, M. E., Dilbirligi, M., Sidhu, D., … Gill, B. S. (2004). A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics, 168(2), 701–712. https://doi.org/10.1534/genetics.104.034868

30. Rao, K. P., Richa, T., Kumar, K., Raghuram, B., & Sinha, A. K. (2010). In silico analysis reveals 75 members of mitogen-activated protein kinase kinase kinase gene family in rice. DNA research : an international journal for rapid publication of reports on genes and genomes, 17(3), 139–153. https://doi.org/10.1093/dnares/dsq011

31. Santos, A., Chaves-Silva, S., Yang, L., Maia, L., Chalfun-Júnior, A., Sinharoy, S., Zhao, J., & Benedito, V. A. (2017). Global analysis of the MATE gene family of metabolite transporters in tomato. BMC plant biology, 17(1), 185. https://doi.org/10.1186/s12870-017-1115-2

32. Seth, S., Debnath, S., & Chakraborty, N. R. (2019). In silico analysis of functional linkage among arsenic induced MATE genes in rice. Biotechnology reports (Amsterdam, Netherlands), 26, e00390. https://doi.org/10.1016/j.btre.2019.e00390

33. Shoji T. (2014). ATP-binding cassette and multidrug and toxic compound extrusion transporters in plants: a common theme among diverse detoxification mechanisms. International review of cell and molecular biology, 309, 303–346. https://doi.org/10.1016/B978-0-12-800255-1.00006-5

34. Sivaguru, M., Liu, J., & Kochian, L. V. (2013). Targeted expression of SbMATE in the root distal transition zone is responsible for sorghum aluminum resistance. The Plant journal : for cell and molecular biology, 76(2), 297–307. https://doi.org/10.1111/tpj.12290

35. Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J., & von Mering, C. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic acids research, 45(D1), D362–D368. https://doi.org/10.1093/nar/gkw937

36. Takanashi, K., Shitan, N.,Yazaki, K. (2014). The multidrug and toxic compound extrusion (MATE) family in plants. Plant Biotechnology, 31, 417–430. https://doi.org/10.5511/plantbiotechnology.14.0904a

37. Tiwari, M., Sharma, D., Singh, M., Tripathi, R. D., & Trivedi, P. K. (2014). Expression of OsMATE1 and OsMATE2 alters development, stress responses and pathogen susceptibility in Arabidopsis. Scientific reports, 4, 3964. https://doi.org/10.1038/srep03964

38. Uhrig J. F. (2006). Protein interaction networks in plants. Planta, 224(4), 771–781. https://doi.org/10.1007/s00425-006-0260-x

39. Wang, L., Bei, X., Gao, J., Li, Y., Yan, Y., & Hu, Y. (2016). The similar and different evolutionary trends of MATE family occurred between rice and Arabidopsis thaliana. BMC plant biology, 16(1), 207. https://doi.org/10.1186/s12870-016-0895-0

40. Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., Liu, J., Chen, K., Zhang, H., Gao, C., & Wang, D. (2018). Analysis of the functions of TaGW2 homoeologs in wheat grain weight and protein content traits. The Plant journal : for cell and molecular biology, 94(5), 857–866. https://doi.org/10.1111/tpj.13903

41. Zhao, J., Huhman, D., Shadle, G., He, X. Z., Sumner, L. W., Tang, Y., & Dixon, R. A. (2011). MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in Medicago truncatula. The Plant cell, 23(4), 1536–1555. https://doi.org/10.1105/tpc.110.080804

42. Zhou, C., Dong, Z., Zhang, T., Wu, J., Yu, S., Zeng, Q., Han, D., & Tong, W. (2020). Genome-Scale Analysis of Homologous Genes among Subgenomes of Bread Wheat (Triticum aestivum). International journal of molecular sciences, 21(8), 3015. https://doi.org/10.3390/ijms21083015

43. Zhu, H., Wu, J., Jiang, Y., Jin, J., Zhou, W., Wang, Y., Han, G., Zhao, Y., & Cheng, B. (2016). Genomewide analysis of MATE-type gene family in maize reveals microsynteny and their expression patterns under aluminum treatment. Journal of genetics, 95(3), 691–704. https://doi.org/10.1007/s12041-016-0686-2

Tables

Table 1
List of 44 MATE genes belonging to different families

Sr.No.	Family	Number of Genes	Available Genes
1.	MATE I	17	TraesCS3B02G562400.1
			TraesCS3B02G563400.1
			TraesCS1D02G030400.1
			TraesCS1A02G029900.1
			TraesCS5B02G326600.1
			TraesCS2B02G247700.2
			TraesCS6D02G384300.1
			TraesCS2B02G612900.1
			TraesCS4D02G068800.1
			TraesCS4A02G245400.1
			TraesCS4B02G070000.1
			TraesCS4B02G070100.1
			TraesCS4A02G245300.1
			TraesCS3B02G336100.1
			TraesCS5D02G355500.1
			TraesCS5D02G150100.1
			TraesCS5B02G245500.1
2.	MATE II	13	TraesCS3B02G298700.1
			TraesCS2A02G202000.1
			TraesCS7D02G541200.1
			TraesCS7A02G554500.1
			TraesCS7B02G479000.1
			TraesCS2B02G296000.1
			TraesCS2D02G277400.1
			TraesCS5D02G378200.1
			TraesCS5D02G378300.1
			TraesCS2B02G296100.1
			TraesCS7D02G488000.1
			TraesCS7A02G500700.1
			TraesCS5D02G413800.1
3.	MATE III	5	TraesCS6A02G350400.1
			TraesCS2A02G139200.1
			TraesCS1A02G032900.1
			TraesCS2B02G280500.1
			TraesCS2A02G261700.1

Sr.No.	Family	Number of Genes	Available Genes
4.	MATE IV	3	TraesCS6A02G256400.1
			TraesCS1A02G188100.1
			TraesCS4A02G059200.1
5.	MATE V	1	TraesCS5B02G562500.1
6	MATE VI	2	TraesCS1A02G305200.1
			TraesCS4B02G244400.1
7	MATE VII	3	TraesCS5B02G150400.2
			TraesCS6A02G418800.1
			TraesCS6D02G407900.1

Table 2 Transcript IDs

Sr. No.	Transcript ID	Homoeologues genes	Target %id	Query %id
1	TraesCS1A02G305200.1	TraesCS1B02G315900	96.38%	96.38%
		TraesCS1D02G304800	98.99%	98.99%
2	TraesCS1A02G188100.1.cds1	TraesCS1B02G195900	96.55%	96.55%
		TraesCS1D02G188200	96.55%	96.55%
3	TraesCS1A02G032900.1	TraesCS1D02G034300	93.14%	93.14%
4	TraesCS1A02G029900.1	TraesCS1B02G037000	94.74%	94.74%
		TraesCS1D02G030400	95.70%	95.52%
5	TraesCS1D02G030400.1	TraesCS1A02G029900	95.52%	95.70%
		TraesCS1B02G037000	95.32%	95.51%
6	TraesCS2A02G261700.1	TraesCS2B02G280500	98.37%	98.37%
		TraesCS2D02G262300	98.57%	98.57%
7	TraesCS2A02G202000.1	TraesCS2B02G229100	98.56%	93.55%
		TraesCS2D02G214300	99.18%	94.14%
8	TraesCS2A02G139200.1	TraesCS2B02G163300	92.77%	93.15%
		TraesCS2D02G142100	96.10%	86.93%
9	TraesCS2B02G296100.1	TraesCS2A02G278500	97.88%	94.87%
		TraesCS2D02G277600	97.30%	96.30%
10	TraesCS2B02G280500.1	TraesCS2A02G261700	98.37%	98.37%
		TraesCS2D02G262300	98.57%	98.57%
11	TraesCS2B02G612900.1			
12	TraesCS2B02G247700.2	TraesCS2A02G222300	98.99%	98.99%
		TraesCS2D02G228200	99.19%	99.19%
13	TraesCS2D02G277400.1	TraesCS2A02G278400	98.33%	98.33%
		TraesCS2B02G296000	98.33%	98.33%
14	TraesCS3B02G563400.1	TraesCS3A02G500000	96.78%	96.39%
		TraesCS3D02G506500	97.39%	97.39%
15	TraesCS3B02G298700.1	TraesCS3A02G265100	97.93%	97.93%
		TraesCS3D02G265400	98.34%	98.34%
16	TraesCS4A02G245300.1	TraesCS4B02G070100	97.04%	97.04%
		TraesCS4D02G068900	95.56%	95.56%
17	TraesCS4A02G245400.1	TraesCS4B02G070000	97.24%	96.42%
		TraesCS4D02G068800	97.89%	97.89%
18	TraesCS4A02G059200.1	TraesCS4B02G234800	95.16%	95.61%
19	TraesCS4B02G244400.1	TraesCS4D02G236000	95.29%	95.14%
		TraesCS4A02G066800	98.37%	98.37%
		TraesCS4D02G245000	98.73%	98.73%

Expression levels of homoeologous genes in 39 MATE genes

Sr. No.	Transcript ID	Homoeologues genes	Target %id	Query %id
20	TraesCS4B02G070000.1	TraesCS4A02G245400	96.42%	97.24%
		TraesCS4D02G068800	97.05%	97.88%
21	TraesCS4B02G070100.1	TraesCS4A02G245300	97.04%	97.04%
		TraesCS4D02G068900	97.04%	97.04%
22	TraesCS4D02G068800.1	TraesCS4A02G245400	97.89%	97.89%
		TraesCS4B02G070000	97.88%	97.05%
23	TraesCS5B02G245500.1	TraesCS5A02G247900	95.17%	94.97%
		TraesCS5D02G254500	98.06%	63.52%
24	TraesCS5B02G150400.2	TraesCS5A02G151800	97.28%	97.11%
		TraesCS5D02G157000	96.75%	96.75%
25	TraesCS5B02G326600.1	TraesCS5A02G326300	80.70%	89.84%
		TraesCS5D02G332300	98.82%	98.24%
26	TraesCS5B02G562500.1.cds1	TraesCS4A02G316600	92.10%	91.75%
		TraesCS5D02G566900	93.88%	94.24%
27	TraesCS5D02G413800.1	TraesCS5B02G408500	98.94%	98.11%
28	TraesCS5D02G150100.1	TraesCS5A02G142000	98.81%	98.81%
29	TraesCS5D02G355500.1	TraesCS5A02G349100	94.35%	93.41%
		TraesCS5B02G350500	95.15%	94.01%
30	TraesCS5D02G378300.1	TraesCS5A02G368600	88.71%	91.29%
		TraesCS5B02G371200	88.57%	94.81%
31	TraesCS5D02G378200.1	TraesCS5A02G368600	91.13%	93.78%
		TraesCS5B02G371200	90.12%	96.47%
32	TraesCS6A02G256400.1.cds1	TraesCS6B02G269000	96.15%	96.15%
		TraesCS6D02G237600	97.03%	97.03%
33	TraesCS6A02G418800.1	TraesCS4B02G322300	94.79%	55.28%
		TraesCS6D02G407900	81.34%	86.58%
34	TraesCS6A02G350400.1	TraesCS6B02G383300	97.52%	97.52%
		TraesCS6D02G332800	92.58%	92.96%
35	TraesCS6D02G407900.1	TraesCS4B02G322300	94.48%	51.76%
		TraesCS6A02G418800	86.58%	81.34%
36	TraesCS6D02G384300.1.cds1	TraesCS6A02G400100	98.03%	98.03%
		TraesCS6B02G440300	96.43%	95.86%
37	TraesCS7A02G500700.1	TraesCS7B02G407800	96.00%	96.00%
		TraesCS7D02G488000	96.63%	96.63%
38	TraesCS7A02G554500.1	TraesCS7B02G477900	96.35%	78.72%

Expression levels of homoeologous genes in 39 MATE genes

Sr. No.	Transcript ID	Homoeologues genes	Target %id	Query %id
		TraesCS7B02G478700	94.47%	94.47%
		TraesCS7B02G479000	94.04%	94.04%
		TraesCS7D02G541800	95.32%	95.32%
39	TraesCS7B02G479000.1	TraesCS7A02G554500	94.04%	94.04%
		TraesCS7D02G541200	94.47%	94.47%
		TraesCS7D02G541800	92.34%	92.34%
40	TraesCS7D02G488000.1	TraesCS7A02G500700	96.63%	96.63%
		TraesCS7B02G407800	97.05%	97.05%
41	TraesCS7D02G541200.1	TraesCS7B02G477900	96.88%	79.15%
		TraesCS7B02G478700	94.47%	94.47%
		TraesCS7B02G479000	94.47%	94.47%
42	TraesCS2B02G296000.1	TraesCS2A02G278400	98.75%	98.75%
		TraesCS2D02G277400	98.33%	98.33%
43	TraesCS3B02G562400.1	TraesCS3A02G499200	99.20%	99.20%
		TraesCS3D02G506100	98.80%	98.80%
44	TraesCS3B02G336100.1	TraesCS3A02G300400	95.19%	95.59%
		TraesCS7D02G014200	97.06%	97.06%
Expression levels of homoeologous genes in 39 MATE genes				

Table 3 List of homoeologous genes

Sl.No.	Homologous Genes (A,B and D)	Stress Conditions			
		High level stress-disease	Stress disease		Intermediate stress
1.	TraesCS1A02G305200.1	Abiotic stress	Stripe rust mixture 6/14 days	Septoria tritici 10 days	Drought & heat
	TraesCS1A02G305200.1	16.36%	0%	100%	8.34%
	TraesCS1B02G315900.1	29.96%	100%	0%	24.92%
	TraesCS1D02G304800.1	53.69%	0%	0%	66.74%
2.	TraesCS1A02G188100.1	No expressions	Fusarium graminearum inoculation 4 days		No expressions
	TraesCS1A02G188100.1		14.36%		
	TraesCS1B02G195900.1		74.86%		
	TraesCS1D02G188200.1		10.78%		
3.	TraesCS1A02G032900.1				
	TraesCS1A02G032900.1				
	TraesCS1D02G034300.1 (no ternary plot)				
4.	TraesCS1A02G029900.1	No expressions	Phosphorus starvation 10 days		Fusarium pseudograminearum
	TraesCS1A02G029900.1		4.32%		1.91%
	TraesCS1B02G037000.1		77.88%		64.07%
	TraesCS1D02G030400.1		17.80%		34.02%
5.	TraesCS1D02G030400.1	No expressions	Phosphorus starvation 10 days		Fusarium pseudograminearum
	TraesCS1A02G029900.1		4.32%		1.91%
	TraesCS1B02G037000.1	77.88%			64.07%
	TraesCS1D02G030400.1	17.80%			34.02%
6.	TraesCS2A02G261700.1	Disease	Chitin(1g per L)		PAMP Chitin
	TraesCS2A02G261700.1	11.92%	14.25%		14.25%
	TraesCS2B02G280500.1	29.44%	disease		3.75%
	TraesCS2D02G262300.1	58.63%	11.92%		82.00%
7.	TraesCS2A02G202000.1	Disease	Mock inoculation 50 hrs		Unknown
	TraesCS2A02G202000.1	13.07%	6.72%		9.68%
	TraesCS2B02G229100.1	24.04%	17.71%		24.51%
	TraesCS2D02G214300.1	62.88%	75.57%		65.81%
8.	TraesCS2A02G139200.1	Disease	6hr of heat stress		PEG 6000
	TraesCS2A02G139200.1	46.68%	93.53%		84.02%
	TraesCS2B02G163300.1	29.75%	0.00%		0.00%
	TraesCS2D02G142100.1	23.57%	6.47%		15.98%

Sl.No.	Homologous Genes (A,B and D)	Stress Conditions		
9.	TraesCS2B02G296100.1	Disease	Septoria tritici 13 days	Zymoseptoria tritici
	TraesCS2A02G278500.1	56.41%	76.15%	62.57%
	TraesCS2B02G296100.1	19.58%	14.19%	18.26%
	TraesCS2D02G277600.1	24.01%	9.66%	19.17%
10	TraesCS2B02G280500.1	Disease	Chitin	PAMP chitin
	TraesCS2A02G261700.1	11.92%	14.25%	14.25%
	TraesCS2B02G280500.1	29.44%	3.75%	3.75%
	TraesCS2D02G262300.1	58.63%	82.00%	82.00%
11	TraesCS2B02G612900.1 No homologous genes			
12.	TraesCS2B02G247700	Abiotic	Stripe rust mixture 6/14 days	Stripe rust
	TraesCS2A02G222300	22.06%	0.00%	2.34%
	TraesCS2B02G247700	24.95%	100.00%	94.10%
	TraesCS2D02G228200	52.99%	0.00%	3.56%
13.	TraesCS2D02G277400.1	Abiotic	Stripe rust mixture 6/14 days	Cold two weeks
	TraesCS2A02G278400.2	11.56%	0.00%	7.25%
	TraesCS2B02G296000.1	27.32%	0.00%	7.35%
	TraesCS2D02G277400.1	61.12%	100.00%	85.40%
14	TraesCS3B02G563400.1	No expressions	Fusarium pseudograminearum 5 days	Fusarium pseudograminearum
	TraesCS3A02G500000.1		5.35%	8.36%
	TraesCS3B02G563400.1		0.00%	0.66%
	TraesCS3D02G506500.1		94.65%	90.98%
15	TraesCS3B02G298700.1	Abiotic	Stripe rust pathogen CYR31 72 hrs	Stripe rust 2
	TraesCS3A02G265100.1	63.62%	93.20%	91.39%
	TraesCS3B02G298700.1	31.39%	5.44%	7.24%
	TraesCS3D02G265400.1	4.99%	1.36%	1.37%
16	TraesCS4A02G245300.1	Disease	Zymoseptoria tritici inoculation 1 day	Cold 2 weeks
	TraesCS4A02G245300.1	32.79%	82.29%	33.66%
	TraesCS4B02G070100	15.89%	8.12%	6.96%
	TraesCS4D02G068900	51.32%	9.58%	59.38%
17	TraesCS4A02G245400.1	Disease	Stripe rust pathogen 87/66 1 day	PAMP flg22
	TraesCS4A02G245400.1	8.35%	2.59%	2.14%
	TraesCS4B02G070000	49.65%	87.38%	65.46%
	TraesCS4D02G068800	42.00%	10.02%	32.40%
18	TraesCS4A02G059200.1	Disease	Magnaporthe oryzae asymptomatic	PAMP chitin

Sl.No.	Homologous Genes (A,B and D)	Stress Conditions		
	TraesCS4A02G059200.1	41.89%	69.48%	67.43%
	TraesCS4B02G234800	31.78%	9.67%	4.81%
	TraesCS4D02G236000.1	26.32%	20.86%	27.76%
19	TraesCS4B02G244400.1	Abiotic	ti_co Fusarium pseudograminearum study 2	
	TraesCS4A02G066800	28.77%	71.28%	7.29%
	TraesCS4B02G244400.1	28.12%	13.14%	62.07%
	TraesCS4D02G245000	43.10%	15.58%	30.63%
20	TraesCS4B02G070000.1	Disease	Stripe rust 1 day	PAMP flg22
	TraesCS4B02G070000.1	8.35%	2.59%	2.14%
	TraesCS4A02G245400	49.65%	87.38%	65.46%
	TraesCS4D02G068800	42.00%	10.02%	32.40%
21	TraesCS4B02G070100.1	Disease	Zymoseptoria tritici inoculation 1 day	Cold 2 weeks
	TraesCS4A02G245300	32.79%	82.29%	33.66%
	TraesCS4B02G070100.1	15.89%	8.12%	6.96%
	TraesCS4D02G068900	51.32%		59.38%
			9.58%	
22	TraesCS4D02G068800.1	Disease	Stripe rust pathogen 87/66 1 day	PAMP flg22
	TraesCS4A02G245400	8.35%	2.59%	2.14%
	TraesCS4B02G070000	49.65%	87.38%	65.46%
	TraesCS4D02G068800.1	42.00%	10.02%	32.40%
23	TraesCS5B02G245500.1	No expressions	No expressions	No expressions
	TraesCS5A02G247900			
	TraesCS5B02G245500.1			
	TraesCS5D02G254500			
24	TraesCS5B02G150400.2	Disease	Fusarium 12 hr	Stripe rust 2
	TraesCS5A02G151800	39.23%	21.26%	42.79%
	TraesCS5B02G150400	24.74%	21.80%	25.18%
	TraesCS5D02G157000	36.04%	56.93%	32.04%
25	TraesCS5B02G326600.1	Abiotic	Mock inoculation 3 days	Fusarium pseudograminearum
	TraesCS5A02G326300.1	0.95%	0.39%	0.53%
	TraesCS5B02G326600.1	59.88%	16.77%	25.32%
	TraesCS5D02G332300	39.18%	82.84%	74.15%
26	TraesCS5B02G562500.1.cds1	disease	Flg22 (500nM)	PAMP flg22

Sl.No.	Homologous Genes (A,B and D)	Stress Conditions			
	TraesCS4A02G316600	31.11%	29.43%		29.43%
	TraesCS5B02G562500.1.cds1	32.17%	11.12%		11.12%
	TraesCS5D02G566900	36.72%	59.45%		59.45%
27	TraesCS5D02G413800.1				
	TraesCS5B02G408500				
	No ternary plot				
28	TraesCS5D02G150100.1				
	TraesCS5A02G142000				
	No ternary plot				
29	TraesCS5D02G355500.1	Disease	Stripe rust pathogen 86/66 11days		Fusarium pseudograminearum study 2
	TraesCS5A02G349100	6.69%	2.36%		1.20%
	TraesCS5B02G350500	23.52%	4.48%		15.93%
	TraesCS5D02G355500.1	69.79%	93.17%		82.87%
30	TraesCS5D02G378300.1	Disease	Stripe rust mixture 6/14 days	Zymoseptoria tritici inoculation 4 days	PAMP chitin
	TraesCS5A02G368500.1	0.18%	0.00%	0.00%	0.11%
	TraesCS5B02G371200.1	48.19%	100.00%	100.00%	99.24%
	TraesCS5D02G378300.1	51.63%	0.%00	0.00%	0.65%
31	TraesCS5D02G378200.1	Disease	Stripe rust pathogen 87/66 9days		Fusarium pseudograminearum study 2
	TraesCS5A02G368500.1	0.14%	0.00%		0.00%
	TraesCS5B02G371200.1	39.77%	0.00%		3.61%
	TraesCS5D02G378200.1	60.08%	100.00%		96.39%
32	TraesCS6A02G256400.1.cds1	Disease	Chitin(1g per L)		PAMP chitin
	TraesCS6A02G256400.1.cds1	37.96%	73.65%		73.65%
	TraesCS6B02G269000	37.69%	14.34%		14.34%
	TraesCS6D02G237600	24.35%	12.01%		12.01%
33	TraesCS6A02G418800.1	Disease	Stripe rust pathogen 87/66 11days		Stripe rust 2
	TraesCS6A02G418800.1	44.56%	71.19%		56.56%
	TraesCS4B02G322300	19.33%	2.00%		7.29%
	TraesCS6D02G407900	36.11%	26.80%		36.15%
34	TraesCS6B02G383300	Disease	Cold 2 weeks		Cold 2 weeks
	TraesCS6A02G350400.1	36.37%	0.00%		0.00%
	TraesCS6B02G383300	55.99%	97.98%		97.98%

Sl.No.	Homologous Genes (A,B and D)	Stress Conditions		
	TraesCS6D02G332800	7.64%	2.02%	2.02%
35	TraesCS6D02G407900.1	Disease	Stripe rust pathogen 87/66 11days	Stripe rust 2
	TraesCS6A02G418800.1	44.56%	71.19%	56.56%
	TraesCS6B02G470700.2	19.33%	2.00%	7.29%
	TraesCS6D02G407900.1	36.11%	26.80%	36.15%
36	TraesCS6D02G384300.1.cds1	No expressions	Mock inoculation 6h	Fusarium pseudograminearum
	TraesCS6A02G400100		88.21%	57.96%
	TraesCS6B02G440300		5.57%	22.63%
	TraesCS6D02G384300.1.cds1		6.22%	19.42%
37	TraesCS7A02G500700.1	Disease	Stripe rust mixture 6/14 days	PAMP flg22
	TraesCS7A02G500700.1	29.25%	0.00%	10.45%
	TraesCS7B02G407800.1	1.73%	0.00%	1.98%
	TraesCS7D02G488000.1	69.02%	100.00%	87.57%
38	TraesCS7A02G554500.1	No expressions	Phosphorus starvation 10 days	Not aavailable
	TraesCS7A02G554500.1		1.00%	
	TraesCS7B02G477900		1.80%	
	TraesCS7D02G541200.1		97.21%	
39	TraesCS7B02G479000.1			
	TraesCS7B02G479000.1			
	TraesCS7D02G541200			
	TraesCS7D02G541800			
	No ternary plot			
40	TraesCS7D02G488000.1	Disease	Stripe rust mixture 6/14 days	PAMP flg22
	TraesCS7A02G500700	29.25%	0.00%	10.45%
	TraesCS7B02G407800	1.73%	0.00%	1.98%
	TraesCS7D02G488000.1	69.02%	100.00%	87.57%
41	TraesCS7D02G541200.1	No expressions	Phosphorus starvation 10 days	Not aavailable
	TraesCS7A02G554500.1		1.00%	
	TraesCS7B02G477900		1.80%	
	TraesCS7D02G541200.1		97.21%	
42	TraesCS2B02G296000.1	Abiotic	Stripe rust mixture 6/14 days	Cold 2 weeks
	TraesCS2A02G278400	11.56%	0.00%	7.25%
	TraesCS2B02G296000.1	27.32%	0.00%	7.35%
	TraesCS2D02G277400	61.12%	100.00%	85.40%
43	TraesCS3B02G562400.1	No expressions	Flg22 500nM	PAMP flg22

Sl.No.	Homologous Genes (A,B and D)	Stress Conditions		
	TraesCS3B02G562400.1		0.00%	0.00%
	TraesCS3A02G499200		100.00%	100.00%
	TraesCS3D02G506100		0.00%	0.00%
44	TraesCS3B02G336100.1	Disease	Zymoseptoria tritici 1 day	Zymoseptoria tritici
	TraesCS3B02G336100.1	9.65%	0.36%	1.95%
	TraesCS3A02G300400	27.89%	7.67%	22.32%
	TraesCS7D02G014200	62.46%	91.96%	75.73%

Table 4 Expression levels of homoeologous genes in 39 MATE genes

SR.NO.	GENE ID	LOCATION	FUNCTION	SOURCE
1	TraesCS1A02G305200.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W4ZRS6
2	TraesCS1A02G188100.1.cds1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W4ZTW3
3	TraesCS1A02G032900.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W4ZWM1
4	TraesCS1A02G029900.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W4ZYL7
5	TraesCS1D02G030400.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5ANP1
6	TraesCS2A02G261700.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5APY7
7	TraesCS2A02G202000.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5B067
8	TraesCS2A02G139200.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5B3C9
9	TraesCS2B02G296100.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5B6D4
10	TraesCS2B02G280500.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5BB68
11	TraesCS2B02G612900.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5BC47
12	TraesCS2B02G247700.2	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5BL55
13	TraesCS2D02G277400.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5C2A7
14	TraesCS3B02G563400.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB – W5CS70
15	TraesCS3B02G298700.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - A0A077S2G0
16	TraesCS4A02G245300.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5DNN7
17	TraesCS4A02G245400.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5DUS3
18	TraesCS4A02G059200.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5DZZ7
19	TraesCS4B02G244400.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5E135
20	TraesCS4B02G070000.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB – W5EB08
21	TraesCS4B02G070100.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5ECN1
22	TraesCS4D02G068800.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5ETZ0
23	TraesCS5B02G245500.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB – W5F8N2
24	TraesCS5B02G150400.2	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB – W5FAQ4

SR.NO.	GENE ID	LOCATION	FUNCTION	SOURCE
25	TraesCS5B02G326600.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5FBV9
26	TraesCS5B02G562500.1.cds1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5FBY2
27	TraesCS5D02G413800.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5FQH5
28	TraesCS5D02G150100.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB – A0A341VVS7
29	TraesCS5D02G355500.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5FT30
30	TraesCS5D02G378300.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5FV95
31	TraesCS5D02G378200.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5FVV9
32	TraesCS6A02G256400.1.cds1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5G6V1
33	TraesCS6A02G418800.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5G7Y4
34	TraesCS6A02G350400.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5GAI6
35	TraesCS6D02G407900.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB – W5GZP8
36	TraesCS6D02G384300.1.cds1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5GZR4
37	TraesCS7A02G500700.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - :A0A341Y439
38	TraesCS7A02G554500.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5H9N8
39	TraesCS7B02G479000.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5HKP8
40	TraesCS7D02G488000.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5HXX9
41	TraesCS7D02G541200.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - W5HZB8
42	TraesCS2B02G296000.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB – W5BHT8
43	TraesCS3B02G562400.1	Transmembrane	Transmembrane transporter activity Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - A0A341SW85
44	TraesCS3B02G336100.1	Transmembrane	Solute:solute Antiporter Activity, Xenobiotic transmembrane transporter activity	UniProtKB - A0A077RVF7

Table 5 **Location and function of different MATE genes**

Sr.No	Transcript ID	Exons	bp	Protein	Protein ID	Source
1	TraesCS1A02G305200.1	13	1846	497aa	A0A2Z4GX53	UniProtKB - W4ZRS6
2	TraesCS1A02G188100.1.cds1	1	2379	551aa	W4ZTW3	UniProtKB - W4ZTW3
3	TraesCS1A02G032900.1	7	1953	481aa	W4ZWM1	UniProtKB - W4ZWM1
4	TraesCS1A02G029900.1	8	1963	515aa	W4ZYL7	UniProtKB - W4ZYL7
5	TraesCS1D02G030400.1	8	1843	512aa	W5ANP1	UniProtKB - W5ANP1
6	TraesCS2A02G261700.1	7	1976	491aa	A0A1D5T6R7	UniProtKB - W5APY7
7	TraesCS2A02G202000.1	7	2167	512aa	A0A1D5TRE7	UniProtKB - W5B067
8	TraesCS2A02G139200.1	7	1673	482aa	W5B3C9	UniProtKB - W5B3C9
9	TraesCS2B02G296100.1	10	1982	487aa	A0A1D5TDG1	UniProtKB - W5B6D4
10	TraesCS2B02G280500.1	7	1778	491aa	W5BB68	UniProtKB - W5BB68
11	TraesCS2B02G612900.1	7	1659	480aa	W5BC47	UniProtKB - W5BC47
12	TraesCS2B02G247700.2	7	3009	493aa	A0A1D5UFC0	UniProtKB - W5BL55
13	TraesCS2D02G277400.1	9	1835	480aa	A0A1D5TX06	UniProtKB - W5C2A7
14	TraesCS3B02G563400.1	5	1431	447aa	A0A1D5VBS5	UniProtKB −W5CS70
15	TraesCS3B02G298700.1	7	1834	483aa	A0A077S2G0	UniProtKB - A0A077S2G0
16	TraesCS4A02G245300.1	8	1797	473aa	W5DNN7	UniProtKB - W5DNN7
17	TraesCS4A02G245400.1	8	1842	475aa	W5DUS3	UniProtKB - W5DUS3
18	TraesCS4A02G059200.1	2	2555	638aa	W5DZZ7	UniProtKB - W5DZZ7
19	TraesCS4B02G244400.1	14	2682	553aa	W5E135	UniProtKB - W5E135
20	TraesCS4B02G070000.1	8	1997	471aa	A0A1D6RZ03	UniProtKB −W5EB08
21	TraesCS4B02G070100.1	9	1997	473aa	W5ECN1	UniProtKB - W5ECN1
22	TraesCS4D02G068800.1	8	1928	475aa	W5ETZ0	UniProtKB - W5ETZ0
23	TraesCS5B02G245500.1	8	1476	477aa	A0A1D6D9H5	UniProtKB − W5F8N2
24	TraesCS5B02G150400.2	12	1560	519	A0A341VAI3	UniProtKB − W5FAQ4
25	TraesCS5B02G326600.1	7	2162	512aa	A0A1D5ZMF2	UniProtKB - W5FBV9
26	TraesCS5B02G562500.1.cds1	1	1566	521aa	W5FBY2	UniProtKB - W5FBY2
27	TraesCS5D02G413800.1	7	1607	476aa	W5FQH5	UniProtKB - W5FQH5
28	TraesCS5D02G150100.1	9	1768	505aa	A0A1D5ZZU1	UniProtKB −A0A341VVS7
29	TraesCS5D02G355500.1	8	2035	501aa	W5FT30	UniProtKB - W5FT30
30	TraesCS5D02G378300.1	8	1827	482aa	W5FV95	UniProtKB - W5FV95
31	TraesCS5D02G378200.1	8	1449	482aa	W5FVV9	UniProtKB - W5FVV9
32	TraesCS6A02G256400.1.cds1	1	2095	572aa	W5G6V1	UniProtKB - W5G6V1
33	TraesCS6A02G418800.1	14	2111	559aa	A0A1D6A8I3	UniProtKB - W5G7Y4
34	TraesCS6A02G350400.1	7	1971	483aa	A0A2X0S2A9	UniProtKB - W5GAI6
35	TraesCS6D02G407900.1	14	2238	595aa	A0A143DH78	UniProtKB − W5GZP8
36	TraesCS6D02G384300.1.cds1	1	1524	507aa	W5GZR4	UniProtKB - W5GZR4

Sr.No	Transcript ID	Exons	bp	Protein	Protein ID	Source
37	TraesCS7A02G500700.1	7	1651	475aa	A0A1D6BJ61	UniProtKB - :A0A341Y439
38	TraesCS7A02G554500.1	8	2024	470aa	W5H9N8	UniProtKB - W5H9N8
39	TraesCS7B02G479000.1	8	1791	470aa	W5HKP8	UniProtKB - W5HKP8
40	TraesCS7D02G488000.1	7	1428	475aa	W5HXX9	UniProtKB - W5HXX9
41	TraesCS7D02G541200.1	8	2014	470aa	W5HZB8	UniProtKB - W5HZB8
42	TraesCS2B02G296000.1	9	1969	471aa	A0A341QSM2	UniProtKB − W5BHT8
43	TraesCS3B02G562400.1	7	2109	498aa	A0A341SW85	UniProtKB - A0A341SW85
44	TraesCS3B02G336100.1	8	1716	476aa	A0A077RVF7	UniProtKB - A0A077RVF7

Figures
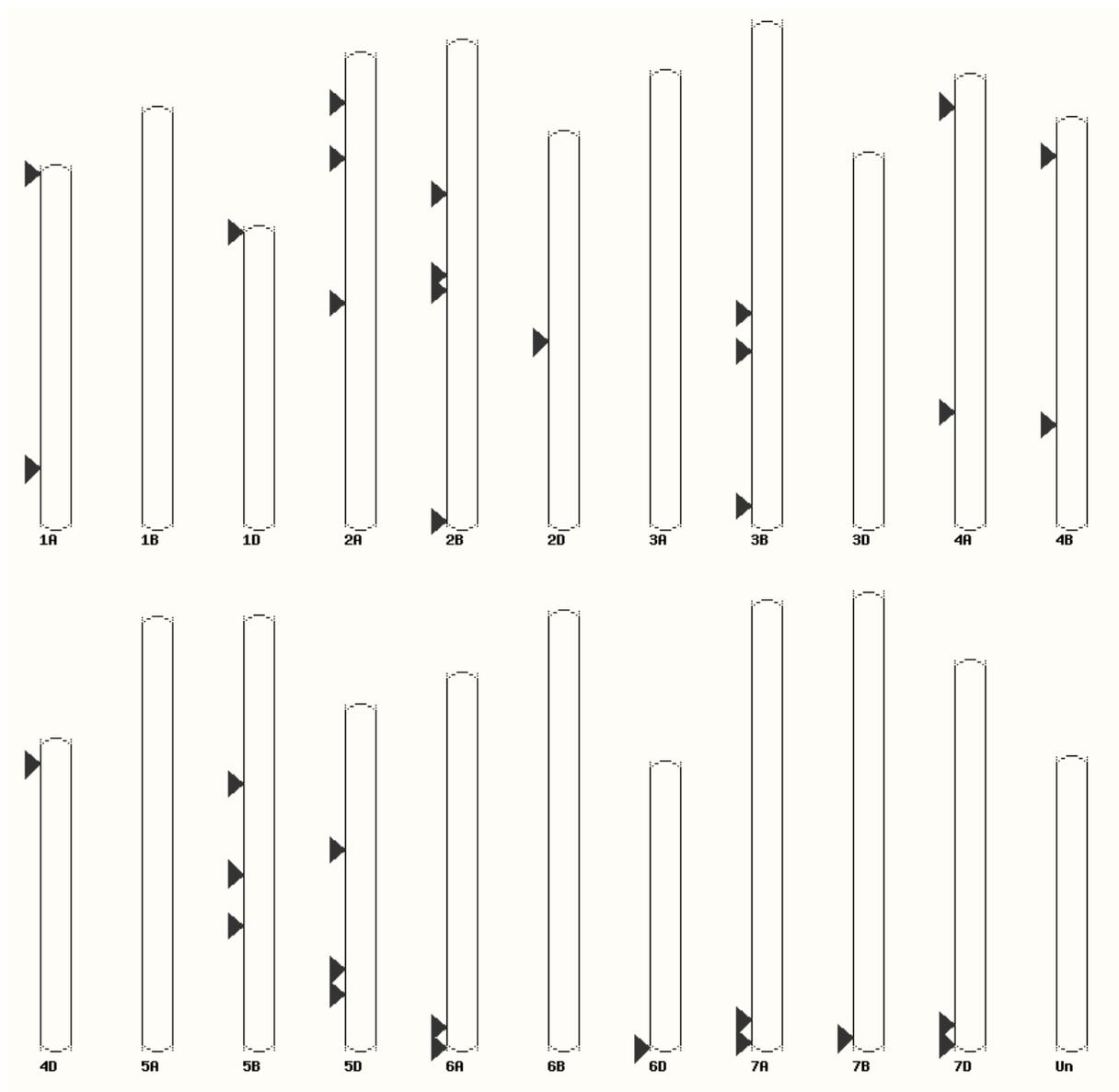
Figure 1

Chromosomal locations of MATE genes under study
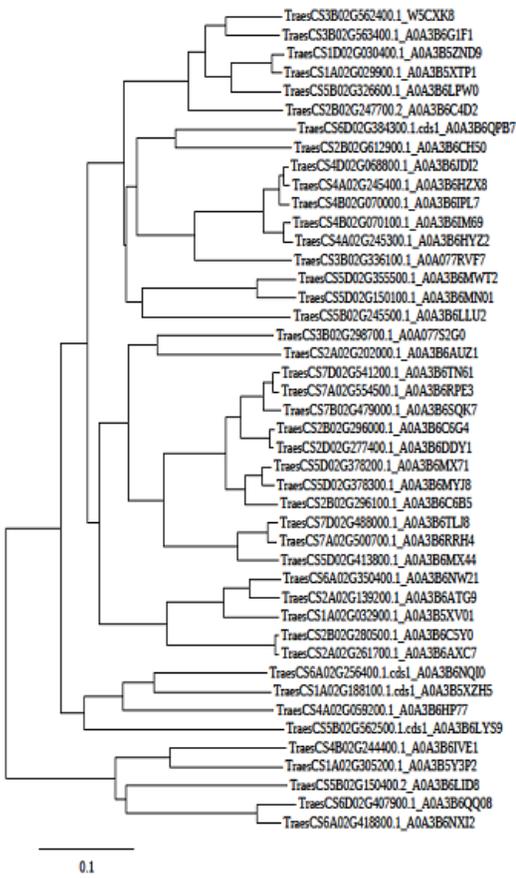

Figure 2

Neighbour-joining (N-J) Phylogenetic tree of the predicted nucleotide and protein sequences of the wheat multidrug and toxic compound extrusion (MATE) under study.

Figure 3

Deletion in the MATE gene structure

Figure 4

Representation of protein association network of 43 MATE genes using STRING software
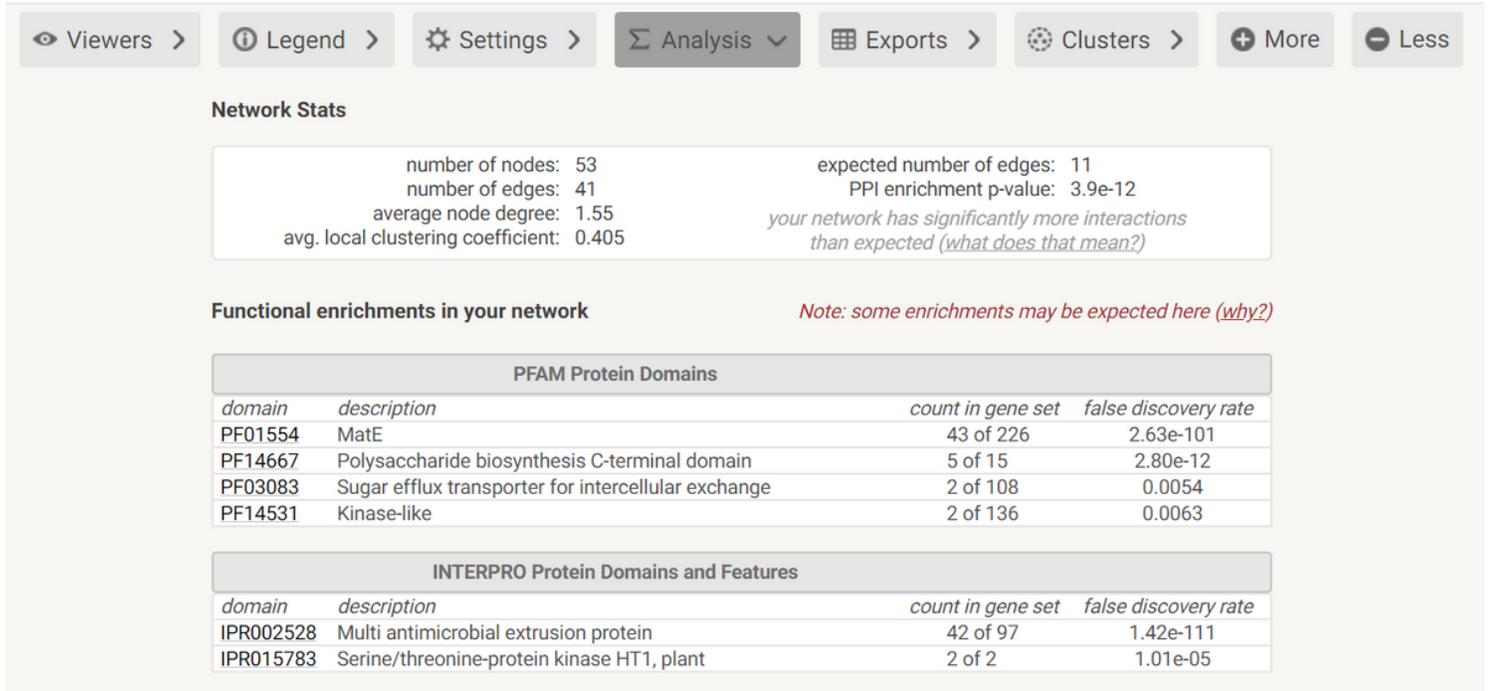
Figure 5

Network and enrichment analysis of MATE genes through STRING software

Figure 6

Heatmap representation of MATE genes expressing different degree of colour intensities across various tissues and developmental stages during stress conditions by using wheat expression browser software. The bottom scale represents values obtained in log2 (tpm) where 0, 4, 7 represent low, intermediate and high expression respectively.
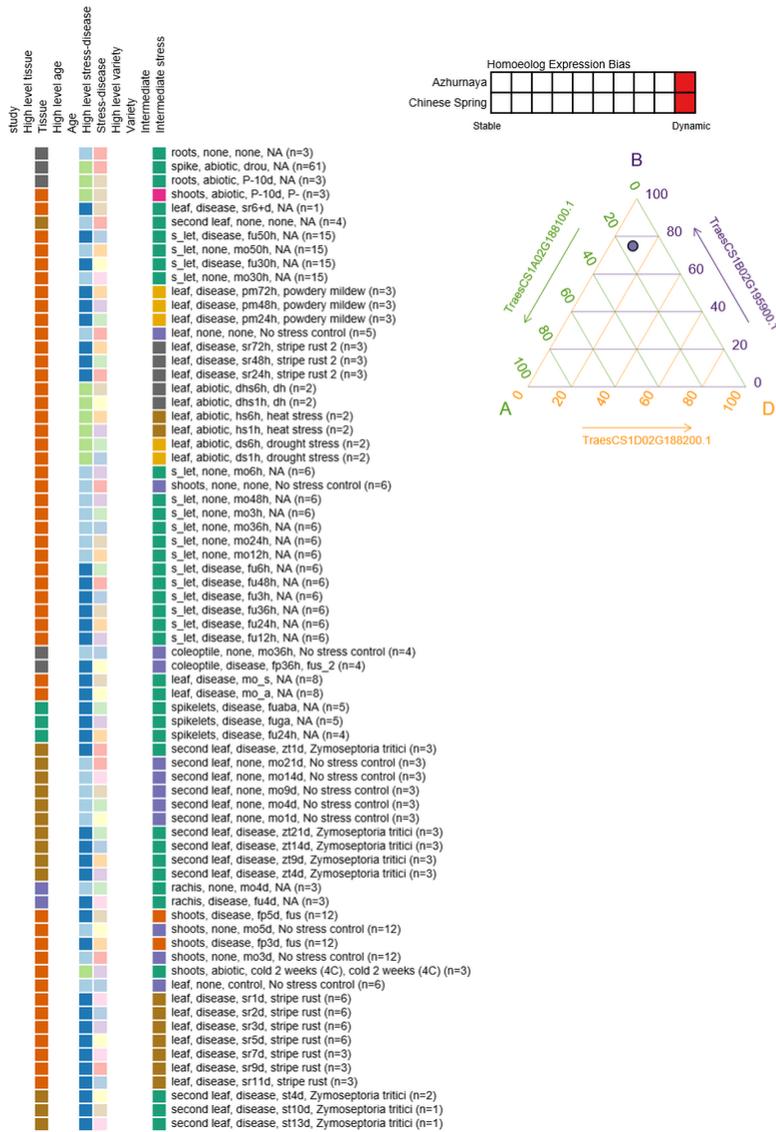
Figure 7

Ternary plot showing homologous gene

Supplementary Files

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

- Additionalfile1.docx
- Additionalfile2.docx
- Additionalfile3.docx
- Additionalfile4.docx
- Additionalfile5.docx
- Additionalfile6.docx
- Additionalfile7.docx