Genome-Wide Characterization and Expression Profiling of Plant-Specific PLATZ Transcription Factor Family Genes in *Brassica rapa* L.

Jaber Bin Azim, Md. Fahim Hassan Khan, Lutful Hassan, Arif Hasan Khan Robin*

Department of Genetics and Plant Breeding, Bangladesh Agricultural University, Mymensingh 2202, Bangladesh

**ABSTRACT** Plant AT-rich sequence and zinc-binding (PLATZ) proteins constitute a plant-specific transcription factor family with two conserved zinc-dependent DNA-binding domains. The PLATZ proteins operate significant functions in regulating plant development and resistance. To date, PLATZ genes have been studied only in a few model plants, including *Arabidopsis*, rice, maize and soybean, but not yet in any *Brassica* species. We identified 24 *Brassica rapa* PLATZ gene family (*BrPLATZ*) genes through genome-wide characterization and profiled their expression using available RNA-Sequencing data. We divided 153 PLATZ proteins from eight plant species into seven groups based on sequence alignment and phylogeny. The *BrPLATZ* genes were generally conserved in groups with similar motif and exon-intron distribution. The 24 *BrPLATZ* genes were located in eight of the ten *B. rapa* chromosomes, with segmental duplication detected in 20 paralogs. Analysis of Ka/Ks ratios revealed that the duplicated genes were under purifying selection. *Cis*-elements analysis implied that *BrPLATZ* genes are diverse in functions including tissue-specific, stress and hormone responsive expressions. Furthermore, expression profiling based on RNA-sequencing data revealed that the *BrPLATZ* genes were expressed in various tissues, with most genes preferentially expressed in flower and silique compared to other tissues. Systematic analysis revealed structural and functional diversity among *BrPLATZ* proteins, which indicated the possibility of diverse functions of *BrPLATZ* genes in development and stress resistance. The characterization of PLATZ gene family members may aid in the selection of appropriate candidate genes responsible for biological functions in *B. rapa* and relevant species.

**Keywords** PLATZ, *Brassica rapa*, DNA-binding domain, Duplication, Tissue-specific expression

**INTRODUCTION**

Transcription factors (TFs) perform a unique role in varied biological mechanisms, such as developmental control, signal activation and stress response cascades and are essential in modulating gene expression (Singh *et al.* 2002). For instance, NAC and GRF regulate root formation, flower, and seed development and SPL TFs are involved in plant transition from juvenile to adult while NF-Y, MYB, and WRKY TFs play significant role in multiple stress tolerance (Samad *et al.* 2017). Transcriptional regulators (TRs) carefully regulate downstream transcription targets, with the interplay of TFs (Singh *et al.* 2002). TFs activate or disable transcription of a gene by getting attached into a certain sequence of its *cis*-elements within the target promoter. TFs carry out their regulating activity by interacting with proteins or by re-shaping chromatin structure (Jing and Lin 2015; Jin *et al.* 2017). Usually, TF has one or more DNA-binding domains and therefore can govern multiple gene expression. Conversely, several TFs may control one gene (Liu *et al.* 2018). So, the comprehensive identification and characterization of TF genes are necessary to understand the historical evolution, biological functions and regulating systems.
Zinc finger (ZF), a major structural motif, comprises a zinc ion at the center, which is enclosed by a number of amino acids, most of them being cysteines or histidines (Krishna et al. 2003). Approximately 15% of TFs in plants bear ZF motifs (Kielbowicz-Matuk 2012). The Cys and His residues, which are imperative for the secondary zinc-binding framework, are categorized into several distinct kinds according to numbers and order such as C3H, C2H2, C2C2, C2HC, C2C2 and C2HC2C2 (Sánchez-García and Rabbits 1994; Takatsuji 1998; Ciftci-Yilmaz and Mittler 2008). For example, TFIIIA and GATA types ZF motifs directly participate in particular DNA sequences recognition (Takatsuji 1998) whereas LIM- and RING-finger types are mainly associated with the interaction among proteins (Takatsuji 1998). Likewise, several new plant-specific ZF motifs were also reported in WRKY and Dof proteins (Takatsuji 1998; Ciftci-Yilmaz and Mittler 2008). For example, TFIIIA and GATA types ZF motifs directly participate in particular DNA sequences recognition (Takatsuji 1998) whereas LIM- and RING-finger types are mainly associated with the interaction among proteins (Takatsuji 1998). Nevertheless, the features of the many ZF-TFs remain unfamiliar in many plant species.

Plant AT-rich sequence and zinc-binding (PLATZ) TF family is a novel class of plant-based zinc ion and DNA-binding proteins (Nagano et al. 2001). Substantial cellular mechanisms like replication of DNA and regulation of gene expression are facilitated by DNA-binding proteins (Nagano et al. 2001). PLATZ1 gene separated from pea (Pisum sativum), was the first reported member of this family (Nagano et al. 2001). PLATZ1 and its paralogs from Arabidopsis and other species have two conserved regions, viz. C-X2-H-X11-C-X2-C-X4-5-C-X2-C-X3-7-H-X2-H (CHC5H3) and C-X2-C-X10-11-C-X3-C (Nagano et al. 2001; Zhang et al. 2018) which are different from other zinc-binding motifs, for instance, RING (C2HC4), LIM (C2HC3) (Schwabe and Klug 1994), GATA finger (C-C-X17-18-C-C) (Teakle et al. 2002), COSTANS/CONSTANS-like (CO/COLs) (Griffiths et al. 2003), and the DNA-binding finger (Dof) etc. (Noguero et al. 2013). The PLATZ1 of pea protein has been reported to tie up with the upstream DE1 component at the A/T-rich DNA sequence and it is necessary for transcriptional inhibition (Nagano et al. 2001). The 12-bp cis regulatory DE1 element, usually found in the developing region of pea epicotyl, was needed for the down-regulation of the pea prl2 gene (Inaba et al. 1999). Multiple members of PLATZ genes exist in plants. For instance, the Arabidopsis genome contains 12 PLATZ genes. Among them AtPLATZ1 and AtPLATZ2 have been reported to regulate desiccation tolerance in vegetative tissues and seeds (Gonzalez-Morales et al. 2016). The maize genome contains 15 PLATZ genes (ZmPLATZ) which are involved in RNAPII-mediated transcriptional regulation (Li et al. 2017; Wang et al. 2018). GmPLATZ1 gene of soybean is associated with the germination process under osmotic stress (So et al. 2015). GL6 encodes a plant-specific PLATZ TF in rice that regulates grain length and spikelet number (Wang et al. 2019). In transgenic Arabidopsis, osmotic stress tolerance increases during germination and seedling stage due to transfer of cotton gene GhPLATZ1 (Zhang et al. 2018). Nevertheless, literatures related to biological functions or physiological mechanisms regulated by PLATZ genes are still limited.

Brassica crops are commonly utilized as vegetables, oils, fodder and condiments. Chinese cabbage (Brassica rapa ssp. pekinensis) is one of the few subspecies of B. rapa (Song et al. 2014) of which the genome of variety Chiifu-401-42 has been sequenced (Wang et al. 2011). The sequences were arranged based on their important economic value and close relationship to Arabidopsis whereas more than 40,000 proteins have been characterized (Wang et al. 2011). Thus, B. rapa has a great potential as a model for genomic and evolutionary research of Brassica species. Moreover, the release of the entire B. rapa genome sequence, as well as others, including Arabidopsis, potato, and tomato, offers us an opportunity for comparative genomic research on PLATZ transcription factor.

In this study, we systematically and comprehensively performed a genome-wide analysis of the PLATZ TF proteins in B. rapa. The objective of this study was identification and characterization of BrPLATZ proteins, sequence analysis, structural organization, conserved motif analysis, evolutionary analysis through phylogenetic classification, chromosomal position, gene duplication, cis-elements and expression profiling in different tissues based on available RNA-sequencing data. This study will provide a useful
resource for further studies to understand the regulation of development and stress resistance in *B. rapa* by the PLATZ proteins.

**MATERIALS AND METHODS**

**Identification of PLATZ family genes and sequence analysis of their proteins**

We retrieved 24 PLATZ genes of *B. rapa* (*BrPLATZs*) to conduct genome-wide characterization of PLATZ gene family. PLATZ coding DNA sequences (CDS), protein sequences and genomic sequences (Supplementary data file 1) of *B. rapa* were collected from BRAD database (Cheng et al. 2011). HMM test was done to confirm their identity using HMMER web server (Potter et al. 2018). The identified sequences were verified in iTAK (Plant Transcription factor and Protein Kinase Identifier and Classifier) (Zheng et al. 2016) and Ensembl Plants databases (Kersey et al. 2016). ProtParam tool (Gasteiger et al. 2005) was used to determine the length, molecular weight (Mw), iso-electric point (pI) and grand average of hydropathicity (GRAVY) of PLATZ proteins. Ensembl Plants database (Kersey et al. 2016) was used to identify open Reading Frame (ORF) of PLATZ genes. Cell-PLoc 2.0 (Chou and Shen 2010) was used to predict the sub-cellular localization of the identified proteins. The protein sequences of *Arabidopsis* (12), rice (15), bread wheat (27), maize (15), tomato (21), potato (9) and soybean (30) were collected from the plant transcription factor and protein kinase identifier and classifier (iTAK) databases (Zheng et al. 2016). These sequences were then confirmed through comparisons with the *Arabidopsis* information resource (TAIR) (Berardini et al. 2015), the plant transcription factor database (PlnTFDB) (Pérez-Rodríguez et al. 2010), grass regulatory information services (GRASSIUS) (Yilmaz et al. 2009) and SoyBase (Grant et al. 2009) databases. The complete protein sequences of *BrPLATZ* proteins were aligned with other plant proteins using DNAMAN software (Lynnon Biosoft, USA).

**Evolutionary analysis of *BrPLATZ* proteins through phylogenetic classification**

Multiple sequence alignments of the full-length protein sequences of *B. rapa, Arabidopsis, rice, bread wheat, maize, tomato, potato* and *soybean* were performed using MEGA7.0 (Kumar et al. 2016) with default parameters. A phylogenetic tree based on the alignment was constructed using MEGA7.0 and the Neighbor-Joining (NJ) method (Saitou and Nei 1987) with the following parameters: poisson correction, pair-wise deletion and bootstrap values in percentages with 1000 replicates. Only groups with bootstraps value higher than 50 were selected for the consensus tree.

**Conserved motif analysis, exon-intron distribution and domain prediction**

MEME suite 5.1.0 was employed to analyze the *BrPLATZ* protein motifs (Bailey et al. 2009). The preset analysis conditions included maximum number of motifs 10, minimum width 6, and maximum width 50. The CDSs of PLATZ genes were aligned with the gene structure display server (GSDS) web tool (Hu et al. 2015) to analyze their exon-intron distribution. The 24 amino acid sequences of 24 *BrPLATZ* genes were used to predict conserved domains using the Pfam database (Finn et al. 2015) and the SMART conserved domain search tool (Letunic and Bork 2018). Protein domain structures were drawn from the start to the end positions using DOG Illustrator (Ren et al. 2009).

**Chromosomal locations and gene duplication of *B. rapa* PLATZ genes**

The chromosomal locations (i.e., start-end positions, chromosome number and length) of the 24 *BrPLATZ* genes were identified using the Ensembl Plants database (Kersey et al. 2016) and their positions in the 10 *B. rapa* chromosomes were mapped using the MapGene2Chrom web v2 web tool (Jiangtao et al. 2015). An NCBI BLAST search (Stephen 1997) was performed to identify gene duplication based on the percentage of query cover to identity of the *BrPLATZ* genes against each other.
Calculation of Ka/Ks ratios to date the duplication events

The duplicated gene pairs obtained from BLAST search were used to calculate the duplication events from synonymous rate (Ks), non-synonymous rate (Ka) and evolutionary constraint (Ka/Ks ratio). At first, protein sequences of the gene pairs were aligned using Clustal Omega (Sievers and Higgins 2014), then the sequence alignments of proteins and the corresponding cDNA sequences were converted to codon alignments using PAL2NAL web tool (Suyama et al. 2006). Finally, the resulting codon alignment was used to calculate Ks and Ka using the CODEML program of PAML (Yang 2007). Ks can be used as the proxy for time and these values were then translated into divergence time in millions of years using the equation $T = \frac{Ks}{2\lambda}$, assuming $1.5 \times 10^{-8}$ substitutions/synonymous site/year as the clock-like rates (Koch et al. 2000).

Cis-acting elements and functional prediction of BrPLATZ genes

The cis-acting regulatory elements (approximately 5 to 10 bp) of 24 BrPLATZ genes were detected using the PlantCARE web-based tool (Lescot et al. 2002). The molecular functions, biological processes and cellular localization of the 24 BrPLATZ proteins were assessed using the Blast2GO functional annotation and genomics software (Conesa and Gotz 2008). The amino acid sequences were loaded in FASTA format into the Blast2GO program and QBlast from NCBI was performed. Subsequently, mapping, annotation and interproscan of GO terms associated with each query were carried out sequentially to predict protein function.

Brassica rapa RNA-sequencing data analysis

For the expression profiling of BrPLATZ genes, we utilized the RNA-sequencing data taken from the Expression Atlas database (https://www.ebi.ac.uk/gxa/home) (Petryszak et al. 2016) that was previously generated and analyzed by Tong et al. (2013). Five tissues of B. rapa accession Chifu-401-42, including root, stem, leaf, flower and silique were analyzed. The transcript abundance is expressed as fragments per kilobase of exon model per million mapped reads (FPKM) values. Using these values, a heat map was generated for the BrPLATZ genes in Microsoft office excel.

RESULTS

Identification of PLATZ family genes and sequence analysis of their proteins

The 24 PLATZ genes those were identified in this study were designated as BrPLATZ1–BrPLATZ24 (Br for Brassica rapa) according to their gene id (Table 1). Considerable variation in length was found in the CDS of PLATZs: from 498 bp (BrPLATZ1) to 978 bp (BrPLATZ17) (Table 1). Among the 24 genes, BrPLATZ17 encoded the longest protein (325 aa), while the shortest (165 aa) one was encoded by BrPLATZ1 (Table 1). In addition, the theoretical pI values of the two proteins (BrPLATZ10 and BrPLATZ24) were below 7, indicating that they were acidic, while the proteins encoded by other PLATZ genes were basic (> 7) (Table 1). Furthermore, the molecular weights of these proteins ranged from 19.158 kDa (BrPLATZ1) to 36.917 kDa (BrPLATZ17) (Table 1). The GRAVY of the BrPLATZs varied from −0.202 to −0.819 indicating that the proteins were hydrophilic in nature (Table 1). The proteins were anticipated to be nuclear in the subcellular location (Table 1).

Multiple protein sequence alignment of B. rapa proteins with other seven plant proteins in the PLATZ family disclosed two conserved regions, one at the N-terminus region having five cysteine and three histidine residues and other at the central region having four cysteine residues (Fig. 1). The consensus signature of these two conserved regions were C-x2-H-x11-C-x2-C-x(4–5)-C-x2-C-x(3–7)-H-x2-H and C-x2-C-x(10–11)-C-x3-C. The intermediate sequence between these two regions was between 56 and 65 amino acids in length. Therefore, these two distant regions could be two distinct domains.

Evolutionary analysis of BrPLATZ proteins through phylogenetic classification

A phylogenetic comparison of the PLATZ proteins has been carried out for the first time in B. rapa. The evolution-
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene id</th>
<th>ORF(^2)</th>
<th>Chromosome location</th>
<th>Length</th>
<th>Mol. Wt.</th>
<th>pI(^3)</th>
<th>GRAVY(^4)</th>
<th>Exons</th>
<th>Introns</th>
<th>Subcellular localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>BrPLATZ1</td>
<td>Bra003715</td>
<td>498</td>
<td>A07</td>
<td>17934666 - 17935230</td>
<td>165</td>
<td>19.158</td>
<td>-0.243</td>
<td>5</td>
<td>2</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ2</td>
<td>Bra007550</td>
<td>693</td>
<td>A09</td>
<td>31260895 - 31263520</td>
<td>286</td>
<td>32.963</td>
<td>-0.143</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ3</td>
<td>Bra007562</td>
<td>570</td>
<td>A07</td>
<td>12934473 - 12935030</td>
<td>189</td>
<td>22.053</td>
<td>-0.013</td>
<td>3</td>
<td>2</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ4</td>
<td>Bra011992</td>
<td>600</td>
<td>A07</td>
<td>11219939 - 11220523</td>
<td>197</td>
<td>22.633</td>
<td>-0.359</td>
<td>3</td>
<td>2</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ5</td>
<td>Bra012632</td>
<td>645</td>
<td>A03</td>
<td>23069898 - 23071556</td>
<td>214</td>
<td>24.261</td>
<td>-0.620</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ6</td>
<td>Bra013149</td>
<td>834</td>
<td>A03</td>
<td>20208675 - 20210224</td>
<td>277</td>
<td>31.436</td>
<td>-0.294</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ7</td>
<td>Bra013266</td>
<td>684</td>
<td>A01</td>
<td>4628713 - 4630353</td>
<td>227</td>
<td>25.763</td>
<td>-0.645</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ8</td>
<td>Bra014478</td>
<td>747</td>
<td>A04</td>
<td>896990 - 898431</td>
<td>248</td>
<td>27.962</td>
<td>-0.246</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ9</td>
<td>Bra014899</td>
<td>741</td>
<td>A07</td>
<td>7646546 - 7649781</td>
<td>246</td>
<td>28.126</td>
<td>-0.427</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ10</td>
<td>Bra015731</td>
<td>717</td>
<td>A07</td>
<td>24406251 - 24407588</td>
<td>238</td>
<td>26.799</td>
<td>-0.518</td>
<td>3</td>
<td>2</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ11</td>
<td>Bra016441</td>
<td>732</td>
<td>A08</td>
<td>17716632 - 17717585</td>
<td>243</td>
<td>27.413</td>
<td>-0.567</td>
<td>3</td>
<td>2</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ12</td>
<td>Bra017531</td>
<td>690</td>
<td>A09</td>
<td>12999803 - 13001075</td>
<td>229</td>
<td>25.845</td>
<td>-0.704</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ13</td>
<td>Bra020129</td>
<td>675</td>
<td>A08</td>
<td>9592044 - 9593365</td>
<td>224</td>
<td>25.227</td>
<td>-0.627</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ14</td>
<td>Bra020555</td>
<td>672</td>
<td>A02</td>
<td>18591737 - 18592977</td>
<td>223</td>
<td>25.184</td>
<td>-0.604</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ15</td>
<td>Bra020588</td>
<td>690</td>
<td>A02</td>
<td>18613651 - 18614916</td>
<td>229</td>
<td>25.823</td>
<td>-0.577</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ16</td>
<td>Bra023155</td>
<td>978</td>
<td>A09</td>
<td>21232031 - 21237321</td>
<td>325</td>
<td>36.917</td>
<td>-0.819</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ17</td>
<td>Bra023280</td>
<td>852</td>
<td>A09</td>
<td>20046533 - 20049414</td>
<td>283</td>
<td>32.138</td>
<td>-0.441</td>
<td>5</td>
<td>4</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ18</td>
<td>Bra025003</td>
<td>681</td>
<td>A06</td>
<td>23568509 - 23569652</td>
<td>226</td>
<td>25.622</td>
<td>-0.586</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ19</td>
<td>Bra025865</td>
<td>735</td>
<td>A06</td>
<td>8282561 - 8283617</td>
<td>244</td>
<td>27.452</td>
<td>-0.545</td>
<td>3</td>
<td>2</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ20</td>
<td>Bra034379</td>
<td>756</td>
<td>A07</td>
<td>12599091 - 12600158</td>
<td>191</td>
<td>22.383</td>
<td>-0.791</td>
<td>3</td>
<td>2</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ21</td>
<td>Bra035494</td>
<td>660</td>
<td>A08</td>
<td>7858260 - 7860325</td>
<td>219</td>
<td>24.891</td>
<td>-0.548</td>
<td>3</td>
<td>2</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ22</td>
<td>Bra037203</td>
<td>768</td>
<td>A09</td>
<td>4960864 - 4962668</td>
<td>255</td>
<td>28.947</td>
<td>-0.420</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
<tr>
<td>BrPLATZ23</td>
<td>Bra038403</td>
<td>723</td>
<td>A08</td>
<td>9161833 - 9164041</td>
<td>240</td>
<td>27.732</td>
<td>-0.522</td>
<td>4</td>
<td>3</td>
<td>Nuclear</td>
</tr>
</tbody>
</table>

\(^2\)Open reading frame. bp: base pair, aa: amino acid, MW: molecular weight, KDa: kilo Dalton.
\(^3\)Iso-electric point.
\(^4\)Grand average of hydropathy.
Genome-Wide Analysis of PLATZ Genes

Fig. 1. Sequence alignment of BrPLATZ proteins and PLATZ proteins from Arabidopsis, rice, bread wheat, maize, tomato, potato and soybean: (a) C-x_2-H-x_11-C-x_2-C-x_4(5)-C-x_2-C-x_3-H-x_2-H region is indicated by the red box; and (b) C-x_2-C-x_10-C-x_3-C region is indicated by the green box. Identical amino acids are indicated by a dark black background and amino acids with > 50% similarity is indicated by a light black background.

ary relationships of the PLATZ proteins among different plant species were studied to explore how the PLATZ gene family was evolved. The tree was constructed from 153 amino acid sequences of B. rapa (24), Arabidopsis (12), soybean (30), rice (15), bread wheat (27), maize (15), potato (9) and tomato (21) (Fig. 2). Because of the low
Fig. 2. Phylogenetic tree showing the relatedness of BrPLATZ proteins to those of Arabidopsis, rice, bread wheat, maize, tomato, potato and soybean. Full-length polypeptide sequences were used to make the phylogenetic tree. The sequences were aligned using MEGA 7.0. The tree was constructed by Neighbor-Joining (NJ) method in MEGA 7.0. The numbers at the branches represented the percent bootstraps values based on 1000 replications. The tree was divided into seven groups (A-G) according to bootstrap support values and evolutionary distances. A species acronym was added before each PLATZ protein name. Br: Brassica rapa, At: Arabidopsis thaliana, Os: Oryza sativa, Ta: Triticum aestivum, Zm: Zea mays, Sl: Solanum lycopersicum, St: Solanum tuberosum, Gm: Glycine max. Only groups with bootstraps value higher than 50 were selected for the consensus tree.

identity among species, the bootstrap values of some nodes were low; nevertheless, most of the nodes had more credible bootstrap values. The PLATZ protein family was divided into seven groups (A-G). Among these, group A contained 39 members and group G contained 38 members, accounting for 25% of the total PLATZ genes, group B contained 25 members and accounting for 16% of the total PLATZ genes, followed by group D (25), group E (14), group C (7) and group F (6). Groups C and F contained PLATZs from monocot species only, whereas other groups contained clusters of PLATZs from both monocot and dicot species. The 24 BrPLATZ proteins were distributed into four out of seven groups except for the groups B, C and F. The highest number of BrPLATZ was clustered in group G containing ten members, and lowest number of BrPLATZ was clustered in group D containing two members. Most BrPLATZ proteins in the tree were represented by paralogous pairs (i.e. BrPLATZ4/BrPLATZ21, BrPLATZ7/BrPLATZ23 and BrPLATZ6/BrPLATZ14). Some orthologous pairs (i.e. BrPLATZ18/AtPLATZ3, BrPLATZ3/ZmPLATZ15) were also observed.

Conserved motif analysis, exon-intron distribution and domain prediction

To further investigate the diversity of the BrPLATZ genes, we analyzed the motifs of the BrPLATZ proteins using the MEME online server. Ten conserved motifs were identified, i.e., motif 1 to 10. An overview of these protein
motifs is presented in Supplementary Fig. S1. Motif 7, the most common motif was present in all the 24 gene products. Motif 1 was absent in BrPLATZ3 protein. Besides, motif 2, 4 and 6 were present in 21 proteins, while motif 3 and 5 were present in 20 proteins. As expected, most of the closely related members in the phylogenetic tree had common motif compositions. Motif 8 was mainly found in group G proteins (Fig. 2, 3). Motif 9 was mainly present in group A and Motif 10 was mainly present in group E (Fig. 2, 3). The differences of motif distribution in different groups of BrPLATZ proteins were also observed. The motif distribution also confirmed that the BrPLATZ proteins were conserved during evolution.

Structural variation among genes is well known to drive the evolution of multigene families. To gain further insights into the structural diversity of PLATZ genes in B. rapa, we deduced the exon-intron distribution of individual BrPLATZ genes (Fig. 4). All BrPLATZ genes contained introns in their coding sequences, which varied from one to four. Most of the genes (14 out of 24) contained three introns, seven genes contained two introns, one gene contained one intron and two genes contained four introns. Most closely related members in the same group shared almost identical exon-intron organization, either in their intron numbers or in exon lengths (Fig. 2, 4). For example, BrPLATZ genes in group A had three introns and BrPLATZ genes in group D had two introns. However, BrPLATZ1, BrPLATZ3, BrPLATZ18 and BrPLATZ22 genes had different numbers of exons and introns, compared to those of other members in the same group (Fig. 2, 4).

PLATZ proteins were categorized as TFs with a conserved PLATZ domain, although the components of other domains have not been recognized. The protein sequences of 24 BrPLATZ genes were subjected to prediction of conserved domains using the Pfam and SMART databases. It was predicted that all BrPLATZ members (except BrPLATZ3) contained a PLATZ domain (Pfam accession number PF04640) (Supplementary Table S1, Fig. 5). Additionally, many genes (15 out of 24) also had BBOX (B-Box-type zinc finger) domain located before the PLATZ domain (Supplementary Table S1, Fig. 5). The PLATZ domain was highly conserved between BrPLATZs. Besides, BrPLATZ3 and BrPLATZ4 had a CC (coiled coil) domain, and BrPLATZ3 and BrPLATZ24 had a signal peptide (SP) domain. BrPLATZ3 had homeobox (H-BOX) domain and transmembrane (TM) region too (Fig. 5).

Chromosomal locations and gene duplication of BrPLATZ genes

We mapped the 24 BrPLATZ genes onto the 10 B. rapa chromosomes and physical location of each gene was determined (Fig. 6). BrPLATZ genes were unevenly distributed on eight of the ten chromosomes. Chromosome A09 contained the highest number of BrPLATZ genes (six genes) followed by chromosome A07 (five genes) and A08 (four genes). Chromosomes A02, A03, A04 and A06 contained two genes each, while chromosome A01 possessed only one gene. There were no PLATZ genes on the two remaining chromosomes (A05 and A10) of B. rapa.

Fig. 3. Schematic representation of the 10 conserved motifs in BrPLATZ proteins. BrPLATZ protein motifs were identified using the online MEME program. Different colored boxes represent different motifs, where the number in center of each boxes indicates their name (Motif 1 to 10). The colored boxes were drawn and ordered manually according to the results of MEME analysis. The length of each box in the figure does not represent the actual motif size in the proteins.
Fig. 4. Schematic presentation of the exon-intron distribution in the *BrPLATZ* genes. The dark blue boxes represent the exons and the red lines represent the introns. Untranslated region is indicated by green box.

Fig. 5. Schematic diagram of BrPLATZ proteins with domains drawn by DOG Illustrator. The domains were identified using the Pfam and SMART databases with the default parameters. PLATZ: PLATZ domain, BBOX: B-Box-type zinc finger, SP: signal peptide, CC: coiled coil, H-BOX: homeobox domain, TM: transmembrane region.

The sequence identity among the 24 *B. rapa* PLATZ proteins ranged from 31% (between BrPLATZ2 and BrPLATZ12) to 95% (between BrPLATZ5 and BrPLATZ20) (Supplementary Table S2). Paralogs were considered as tandemly duplicated genes if two genes were situated in a
100-kb region on a chromosome and separated by five or fewer genes (Wang et al. 2015). If the percentage of query cover and identity of the genes was ≥ 80, they were regarded as segmentally duplicated genes (Kong et al. 2013). We detected 20 pairs of segmentally duplicated paralogous BrPLATZ genes based on the percentage of query cover and identity (more than 80%) (Supplementary Table S3). These genes were located on chromosomes A01, A02, A03, A04, A06, A07, A08 and A09 either on the same chromosome (4 pairs) or on separate chromosomes (16 pairs) (Fig. 6). All duplicated gene pairs belong to the same group according to the phylogenetic relationships shown in Fig. 2. There was no tandem duplication in any of the genes. Combined with chromosomal distribution analysis, four pairs of paralogs were involved in regional duplication within the same chromosome, whereas others were between chromosomes.

Calculation of Ka/Ks ratios to date the duplication events

Almost the entire PLATZ gene family of B. rapa was expanded by the large-scale segmental gene duplication. To determine the selection constraints acting on this gene family, we estimated the Ka/Ks ratio for 20 pairs of paralogous genes in the network of duplicated regions of B. rapa (Supplementary Table S3). The mode of selection was identified by evaluating the Ka/Ks values. In this case, the results may be Ka/Ks < 1, = 1 or > 1 in which Ka/Ks less than 1 indicates the functional constraint with negative or purifying selection of the genes, a Ka/Ks ratio of 1 means that the genes are drifting neutrally, and Ka/Ks more than one 1 suggests accelerated evolution with positive selection (Nekrutenko et al. 2002). All the Ka/Ks ratios from the twenty BrPLATZ paralogous gene pairs were less than 1.
indicating that there was a strong purifying/negative selection pressure in these genes, with a little variation after duplication and they were slowly evolving at the protein level. The duplication events were estimated to occur between 3.51 (Ks = 0.105) to 39.17 (Ks = 1.175) million years ago (Mya) (Supplementary Table S3). Therefore, we concluded that the large-scale duplication events involving BrPLATZ genes occurred within the last 3.51-39.17 million years.

**Cis-acting elements and functional prediction of BrPLATZ genes**

The Cis-acting elements have a vital role in determining patterns of gene expression to particular tissues, stresses or hormones in different environmental situations (Yamaguchi-Shinozaki and Shinozaki 2005). Therefore, a web search was performed using PlantCare database to identify possible stress and hormone responsive cis-acting elements of BrPLATZ Genes (Supplementary Table S4). We found three abiotic-stress responsive cis-elements: MBS (MYB binding site) present in 5 different BrPLATZ genes were responsive to drought, LTR (low-temperature responsiveness) in 7 genes and TC-rich repeats in 11 genes were responsive to defense and stress (Supplementary Table S4). We also found five hormone responsive cis-elements: gibberellin in 14 genes, methyl jasmonate (MeJA) in 12 genes, abscisic acid (ABA) in 12 genes, auxin in 8 genes and salicylic acid in 10 genes (Supplementary Table S4). In addition, several other cis-acting elements that were related to tissue-specific expression (meristem, endosperm and mesophyll), circadian regulation, anaerobic induction, zein metabolism and light responsiveness were found in the BrPLATZ genes (Supplementary Table S4). Furthermore, most of the BrPLATZ genes within the same group in the phylogenetic tree have similar cis-acting elements in their promoter regions.

Functional prediction based on gene ontology (GO) classifications placed all BrPLATZ genes in some common groups including: zinc ion binding (21 genes), protein binding (4 genes) and lipid binding (1 gene) in the molecular function category; leaf senescence (3 genes) and regulation of transcription (1 gene) in the biological process category; and nucleus (4 genes) and membrane (4 genes) in the cellular component category (Supplementary Table S5). Although the 24 BrPLATZ genes have some common functions based on GO category, however, due to the variation of cis-elements, there might have some functional variations among the genes.

**Differential expression profiling of BrPLATZ genes in various tissues**

The tissue-specific expression profiling of a gene family can provide clues about its possible functional roles in developmental processes. To understand which PLATZ genes may be involved in regulating specific tissue or organ growth in B. rapa, the expression patterns of the BrPLATZs in five tissues (leaf, stem, flower, silique, root) were explored based on FPKM values obtained from RNA-sequencing data which were previously analyzed by Tong et al. (2013) in B. rapa (Fig. 7).

The results indicated that, nine genes (BrPLATZ6,
BrPLATZ8, BrPLATZ11, BrPLATZ12, BrPLATZ14, BrPLATZ18, BrPLATZ19, BrPLATZ20 and BrPLATZ22) were expressed at variable levels in all tissues (Fig. 7). Among the 24 genes, BrPLATZ5, BrPLATZ6, BrPLATZ11, BrPLATZ12, BrPLATZ14, BrPLATZ20, BrPLATZ21 and BrPLATZ22 were highly expressed in flowers compared to other tissues. Four genes (BrPLATZ2, BrPLATZ10, BrPLATZ17 and BrPLATZ24) showed the highest transcript accumulation in the stems, six genes in silique (BrPLATZ1, BrPLATZ3, BrPLATZ8, BrPLATZ9, BrPLATZ15 and BrPLATZ18), three genes in leaves (BrPLATZ13, BrPLATZ16 and BrPLATZ19) and two genes (BrPLATZ7 and BrPLATZ23) in root (Fig. 7).

From the phylogenetic analysis, we found that most of the members within the same phylogenetic group shared a similar expression profile in B. rapa tissues. For instance, only group E and G showed expression in all tissues at variable levels (Fig. 2, 7). Also, BrPLATZ4 and BrPLATZ21 belonging to group D showed higher expression in flower while BrPLATZ10, BrPLATZ17 and BrPLATZ24 belonging to group A showed higher expression in stem (Figs. 2, 7). However, BrPLATZ2 and BrPLATZ9 belonging to group A also showed divergence in expression patterns in different tissues.

**DISCUSSION**

The faster development of genome sequencing projects is opening the windows to configure the functions and the mechanisms of genes related to plant growth, development and stress resistance. The study of the gene family has recently become a valuable technique for analyzing the gene function, structure and evolution. Studies related to PLATZ genes have been reported in the higher model plant species such as Arabidopsis (González-Morales et al. 2016), maize (Li et al. 2017), soybean (So et al. 2015) and rice (Wang et al. 2019). However, there is no extensive study on the PLATZ gene family in B. rapa. The species B. rapa (2n = 20, AA) includes several subspecies which are providing human nutrition in the form of leaves, roots, stem vegetables and edible oils. It also represents the origin of the Brassica ‘A’ genome of other cultivated oilseed crops of Brassica allopolyoids: B. napus (AACC) and B. juncea (AABB). Therefore, genome-wide identification and characterization of the PLATZ genes in B. rapa is viable and vital.

In this study, 24 PLATZ genes from B. rapa were identified and characterized. The number of PLATZ genes in B. rapa was somewhat higher than that of potato (9), Arabidopsis (12), rice (15), maize (15) and tomato (21) but lower than that of bread wheat (27) and soybean (30). However, no correlation exists between the number of PLATZ genes and the size of various crop genomes. Compared with the differences in overall genome size between B. rapa (283.8 Mb) (Wang et al. 2011) and larger-genome plants such as rice (441 Mb) (Eckardt 2000), potato (726 Mb) (Xu et al. 2011), tomato (900 Mb) (Sato et al. 2012), soybean (1115 Mb) (Schmutz et al. 2010), maize (2300 Mb) (Schnable et al. 2009) and bread wheat (14.5 Gb) (Appels et al. 2018), the number of PLATZ genes in B. rapa is relatively large. This enormous divergence suggests that genome duplications could have contributed to the expansion of the PLATZ gene families in the B. rapa genome.

In general, TFs are located in the cytoplasm and translocated into the nucleus where they interact with cis-acting elements after receiving a signal from the cell membrane signal transduction (Liu et al. 2018). The BrPLATZ proteins were also anticipated to be located in the nuclear region. The two distantly located conserved zinc finger regions in PLATZ organized by cysteine and histidine residues and some variable regions were required for DNA binding activities dependent on zinc ion binding (Nagano et al. 2001). The N-terminus region is similar to some ‘double-zinc-finger’ domains, such as RING (C3HC4) and LIM (C2HC5) (Schwabe and Klug 1994) and central region seems nearly identical to the GATA finger (C2C2) (Teakle et al. 2002). It states that nucleotide substitution and minor insertions/deletions contributed in the evolution and diversification of different conserved regions (Purugganan and Wessler 1994). The zinc finger protein domain is one of the most versatile domains of intracellular proteins that mediate protein interactions with other biomolecules, such as DNA, RNA, other proteins or lipids. In BrPLATZ proteins, B-Box and PLATZ are the two main domains, which
are involved in DNA binding and zinc ion binding. The B-Box domain is a cysteine rich protein domain found in all eukaryotes whereas the PLATZ domain is only found in plants. This domain is considered as functional as well as an evolutionary unit of the protein whose coding sequence can be duplicated and undergo recombination. Furthermore, based on functional prediction, it is confirmed that BrPLATZ proteins interact with lipid, DNA and other proteins.

Our phylogenetic analysis revealed the evolutionary relationships between the BrPLATZ and AtPLATZ proteins were closer compared to other crop PLATZ proteins. This finding suggested that they may have evolved from common ancestors with functional similarities. The presence of seven distinct groups of PLATZ proteins and the presence of both monocots and dicots members in most of them indicated that PLATZ genes were diversified before the monocot-dicot split. In addition, many paralogous and orthologous pairs were observed in the tree. Paralogs are the homologous genes that have diverged within same species and created by the gene duplication events giving rise to a new gene with a new function, though the function is often related to the role of the ancestral gene (Gabaldón and Koonin 2013). On the contrary, orthologs have diverged after a speciation event and maintain a similar function to that of the ancestral gene (Gabaldón and Koonin 2013). The orthologs between monocot and dicot PLATZ proteins in the tree imply that there might be some ancestor of PLATZ genes before monocot and dicot divergence, while the paralogs suggest that PLATZ gene family expanded after the monocot and dicot divergence. The relationships of BrPLATZ indicated by the phylogenetic tree were further supported by the similar gene structure and protein-motif patterns within each group.

Divergence in exon-intron structures plays a significant role in the evolution of many gene families. Therefore, the exon/intron structure of each member of the B. rapa PLATZ family along with their phylogenetic relationships were analyzed. The different exon-intron structures suggested that there are structural variations in this gene family. Exon/intron gain or loss, exonization or pseudo-exonization, and insertion or deletion of amino acids, are responsible for variations in the coding regions; particularly those that can alter gene function (Xu et al. 2012). However, the similar exon-intron distribution in different groups suggested that these genes were highly conserved during evolution. We further analyzed the conserved motifs by the MEME server, which supports the results of phylogenetic analysis. The majority of the BrPLATZ proteins in the same group shared similar motifs indicating that they were probably generated by gene expansion. The differences in motif distribution in different groups indicated sources of functional divergence in PLATZ genes over the evolutionary history.

Gene duplication is a significant process of evolution that allows organisms to adapt in diverse environments by creating new genes (Li et al. 2014). The major reasons for the expansion of gene family in plants are tandem and segmental duplication along with transposition phenomena (Cannon et al. 2004; Li et al. 2014). Among these, segmental duplication is more common, as most of the plants are diploidized-polyploids and have a number of chromosomal duplicates in their genomes (Cannon et al. 2004). Duplication of genes among various chromosomes is referred to as segmental duplication, while two or more genes on a single chromosome confirm a tandem duplication event (Liu et al. 2011). In the current study, no evidence of tandem duplication was spotted for any gene pair, indicating that segmental duplication rather than tandem duplication has played a leading role in the expansion of the B. rapa PLATZ gene family. Duplicated genes belonged to the same groups in the phylogenetic tree suggest group specific gene duplication events during evolution. Furthermore, combined with exon-intron distribution, duplicated genes shared almost identical structures suggesting that structural divergences have played a more important role during the evolution of duplicated genes rather than non-duplicated genes.

The Ka/Ks ratio is considered as an indicator for determining the type of selection pressure. In this study, the ratios showed that during the course of evolution, duplicated genes underwent a strong purifying selection. We estimated that the segmental duplication event of the BrPLATZ genes occurred within 3.51-39.17 Mya. Arabidopsis and Brassica perhaps were split from each other between 23.4 and 33.5 Mya ago (Kumar et al. 2017). Therefore, the
duplication of BrPLATZ genes frequently occurred during the evolution of B. rapa before and after the split of B. rapa from the Arabidopsis lineage. These duplicated genes may face three alternative outcomes: (i) one copy may become silenced within a few million years; (ii) one copy may acquire a unique, beneficial function and play a significant role in the passive origin of new species, while the other copy may retain the original function or (iii) both copies may become partially compromised by mutations (Lynch and Conery 2000).

The functional diversity of genes can be predicted from their differential expression in different tissues. The differential expression of BrPLATZ genes in different tissues at various levels implies that functional variation might have arisen during evolution and suggests that they might have various regulatory functions in growth and development of B. rapa. The predominant expression of most of the genes (8 genes) in flowers implies that these genes might function in flower development through the association with other flower-specific genes (Fig. 7). The root is the main organ responsible for absorption of water and minerals. BrPLATZ7 and BrPLATZ23 were predominately expressed in roots suggesting that they might play a role in root formation and/or water and nutrient uptake. In addition, higher expression of three genes in leaves compared to other tissues indicated that they are involved in leaf growth and development. The role of PLATZ genes in plant growth and development has been confirmed in other species. For instance, in pea, the expression of PLATZ1 gene is more abundant in the root tip and terminal buds rather than in the mature leaf, stem, and root tissues (Nagano et al. 2001). ZmPLATZ12 (Fl3) is specifically expressed in the starchy cells of endosperm in maize (Li et al. 2017). In addition, AtORE15, a PLATZ TF, was found to express in young leaves and was involved in regulation of leaf growth and suppression of senescence (Kim et al. 2018). The development of fruit is regulated by numerous TFs such as NAC, MADS-box, and EIN3/EIL gene families (Feng et al. 2016). In this study, the transcript levels of six BrPLATZ genes were generally higher in silique compared to other tissues, suggesting that they might function in growth and development of B. rapa silique. The similar expressions of duplicate paralog gene pairs (i.e., BrPLATZ13 and BrPLATZ19) indicate that even after duplication, their functions may have been retained and they can play redundant role in the regulation of tissue formation. However, the diverse expression patterns of several paralogous pairs suggest that they play different roles in development of B. rapa.

Cis-acting regulatory elements can also regulate tissue-specific or stress-responsive expression patterns in multi-stimulus responsive genes. They are important molecular switches that participate in gene transcriptional regulation and control the large network of genes involved in different biological phenomena including stress responses and developmental processes (Zhang et al. 2005). In the current study, several stress-responsive cis-elements were found in the promoter regions of most of the BrPLATZ genes suggesting their role in abiotic stress and environmental adaptation. Phytohormones also function as intracellular messengers and assemble different pathways of signal transduction to respond against stresses (Wolters and Jürgens 2009). All the BrPLATZ genes have one or more than one of the following phytohormone-responsive cis-elements in their promoter regions such as auxin-, methyl jasmonate-, gibberellin-, abscisic acid- and salicylic acid-responsive cis-elements supporting the likelihood of stress tolerance-related functions for the BrPLATZ genes.

In summary, this study systematically characterized BrPLATZ family genes using different tools of bioinformatics and RNA-sequencing data obtained from public database. We performed comprehensive genome-wide analysis of the PLATZ gene family in B. rapa to explore their potential roles in organ development and adaptive responses to stresses. We also analyzed the predicted gene structures, conserved domains, chromosomal distributions, duplication events, and evolutionary divergence of the BrPLATZ genes and classified them based on phylogenetic analysis. Finally, we predicted the functions of BrPLATZ genes based on their expression profiles and the presence of cis-elements in their upstream promoter regions. Our results will lay the foundation for further studies aimed at uncovering the important biological functions of PLATZ genes in plants.
AUTHOR CONTRIBUTIONS

AHKR and LH planned this study. JBA and MFHK retrieved data. JBA conducted in silico analysis and wrote the manuscript. AHKR critically edited the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

This research was supported by the Bangladesh Agricultural University Research Systems (Grant no. 2018/604/BAU).

REFERENCES

cleic Acids Res. 45: 1040-1045.


