



OPEN

## A genome-wide comparative evolutionary analysis of zinc finger-BED transcription factor genes in land plants

Athar Hussain<sup>1,6</sup>, Jinbao Liu<sup>2,6</sup>, Binoop Mohan<sup>2,6</sup>, Akif Burhan<sup>3</sup>, Zunaira Nasim<sup>3</sup>, Raveena Bano<sup>3</sup>, Ayesha Ameen<sup>4</sup>, Madiha Zaynab<sup>5</sup>, M. Shahid Mukhtar<sup>2</sup>✉ & Karolina M. Pajeroska-Mukhtar<sup>2</sup>✉

Zinc finger (Zf)-BED proteins are a novel superfamily of transcription factors that controls numerous activities in plants including growth, development, and cellular responses to biotic and abiotic stresses. Despite their important roles in gene regulation, little is known about the specific functions of Zf-BEDs in land plants. The current study identified a total of 750 Zf-BED-encoding genes in 35 land plant species including mosses, bryophytes, lycophytes, gymnosperms, and angiosperms. The gene family size was somewhat proportional to genome size. All identified genes were categorized into 22 classes based on their specific domain architectures. Of these, class I (Zf-BED\_DUF-domain\_Dimer\_Tnp\_hAT) was the most common in the majority of the land plants. However, some classes were family-specific, while the others were species-specific, demonstrating diversity at different classification levels. In addition, several novel functional domains were also predicated including WRKY and nucleotide-binding site (NBS). Comparative genomics, transcriptomics, and proteomics provided insights into the evolutionary history, duplication, divergence, gene gain and loss, species relationship, expression profiling, and structural diversity of Zf-BEDs in land plants. The comprehensive study of Zf-BEDs in *Gossypium* sp., (cotton) also demonstrated a clear footprint of polyploidization. Overall, this comprehensive evolutionary study of Zf-BEDs in land plants highlighted significant diversity among plant species.

The zinc finger (Zf) proteins are a superfamily of proteins involved in a wide range of plant functions, such as growth, development, and resistance to various biotic and abiotic stressors<sup>1,2</sup>. Zinc fingers are small peptide fragments stabilized with zinc ions that interact with nucleic acids and act as regulatory proteins that control transcription<sup>3–5</sup>. The Zf domains are often associated with other domains that modulate their biological functions. One of such domains is the BED domain, consisting of about 50–60 amino acid residues that contain an organized motif with two highly conserved aromatic positions and a shared pattern of cysteines and histidine<sup>3,6</sup>. The BED domain was named after the *Drosophila* proteins BEAF and DREF<sup>7</sup>, and can be found in one or more copies in the genomes of various animals, plants, and fungi<sup>8</sup>. BED-domain proteins have been shown to have DNA binding activity, and some are thought to be involved in chromatin insulation and ethylene regulation<sup>9</sup>.

The zinc finger-homeodomain proteins are categorized into nine types based on their structural and functional variation including C2H2, C8, C6, C3HC4, C2HC, C2HC5, C4, C4HC3, and CCCH (where C and H represent cysteine and histidine, respectively)<sup>10–13</sup>. Various zinc finger domains-containing proteins such as Lesion Simulating Disease Resistance 1 (LSD1) (C2C2), LSD One Like (LOL1) (C2C2), ZAT12 (C2H2), ZAT7 (C2H2) and AtNFX1 (NF-X1) in *Arabidopsis*<sup>14,15</sup>, StZFP1 (Zinc Finger Protein) (C2H2) in *Solanum tuberosum*

<sup>1</sup>Genomics Lab, School of Food and Agricultural Sciences (SFAS), University of Management and Technology (UMT), Lahore 54770, Pakistan. <sup>2</sup>Department of Biology, University of Alabama at Birmingham, 1300 University Blvd, Birmingham, AL 35294, USA. <sup>3</sup>Department of Life Science, University of Management and Technology (UMT), Lahore 54770, Pakistan. <sup>4</sup>Office of Research Innovation and Commercialization, University of Management and Technology, Lahore 54770, Pakistan. <sup>5</sup>Shenzhen Key Laboratory of Marine Bioresource and Eco-Environmental Sciences, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 51807, Guangdong, China. <sup>6</sup>These authors contributed equally: Athar Hussain, Jinbao Liu and Binoop Mohan. ✉email: smukhtar@uab.edu; kmukhtar@uab.edu

(potato) and OsLSD1 (C2C2), OsLOL1 (C2C2), OsRING-1 (RING H2, RING HC), OsRFP1, OsDOS (CCCH), OsZFP (C2H2) & SRZ1 (C2C2) in *Oryza sativa* (rice)<sup>11,13,16</sup>. Recent reports indicate that zinc finger motifs have an important role in host–pathogen interactions<sup>17</sup>. For instance, multiple resistance proteins contain a Zf-BED domain, conferring immunity to various biotic stresses<sup>18</sup>. The Zf-BED domains also associate with the Nod-like receptor (NLR) proteins either at the N- or C- termini<sup>17,19–21</sup>. As such, NLR proteins play critical roles in plant immune responses<sup>22–24</sup>. Many defense proteins of *Arabidopsis* and rice containing zinc finger domains have also been shown to regulate programmed cell death (PCD)<sup>2,25</sup>. Likewise, Zf-BED proteins participate in numerous abiotic stresses including leaf senescence<sup>26,27</sup>.

In the current study, we have analyzed the evolutionary relationship of Zf-BED-containing genes in 35 land plant species including algae, bryophytes, gymnosperms, and flowering plants. Our study included identification, characterization, comparison, evolutionary relationship, orthologues, and species-based trees in all land plants. As a focal species for our analyses, we selected cotton (*Gossypium* sp.), which showed an unusually high number of Zf-BEDs and is an economically important crop plant worldwide that is also targeted by various deadly pathogens<sup>28–30</sup>. Here, we report a comprehensive genome-wide comparative study of Zf-BED among diverse plant species with a special focus on *Gossypium* sp. The presented results will provide a foundation for further functional analyses of the Zf-BED proteins in plants.

## Methods

**Plant genomes assemblies.** In the current study, we have selected 35 land plants for genome-wide identification and evolutionary analysis of Zf-BED-containing genes. The selected plant species span the evolutionary spectrum from algae to higher plants including *Chlamydomonas reinhardtii* (Algae), *Marchantia polymorpha* (Bryophytes), *Selaginella moellendorffii* (Lycophytes), *Picea abies* (Gymnosperms), *Ginkgo biloba* (Gymnosperms), *Spirodela polyrhiza* (Monocots), *Zea mays* (Monocots), *Sorghum bicolor* (Monocots), *Oryza sativa* (Monocots), *Brachypodium distachyon* (Monocots), *Hordeum vulgare* (Monocots), *Amborella trichopoda* (basal Angiosperms), *Nelumbo nucifera* (basal Angiosperms), *Aquilegia coerulea* (Dicots), *Solanum lycopersicum* (Dicots), *Solanum tuberosum* (Dicots), *Solanum melongena* (Dicots), *Vitis vinifera* (Dicots), *Citrus clementina* (Dicots), *Theobroma cacao* (Dicots), *Gossypium raimondii* (Dicots), *Gossypium arboreum* (Dicots), *Gossypium barbadense* (Dicots), *Gossypium hirsutum* (Dicots), *Gossypium turneri* (Dicots), *Gossypium herbaceum* (Dicots), *Gossypium thurberi* (Dicots), *Capsella rubella* (Dicots), *Arabidopsis thaliana* (Dicots), *Arabidopsis lyrata* (Dicots), *Prunus persica* (Dicots), *Populus trichocarpa* (Dicots), *Medicago truncatula* (Dicots), *Betula pendula* (Dicots) and *Cucumis sativus* (Dicots). We have downloaded publicly available latest genome assemblies from their respective genome databases as well as NCBI, Phytozom, Plaza, etc. (Table S1).

**Identification of Zf-BED and associated domains.** To screen the Zf-BED domain-containing genes, the Pfam (protein family) databases were downloaded and installed on the local server. The pfamScan.pl algorithm was used to search for Zf-BED domains with e-value (1.1e–50) in Pfam-A. The pfamscan.pl output file was processed. This led to the filtering of all genes containing Zf-BED domains and considered *Zf-BED* genes. To find additional Zf-BED associated domains, we followed Hussain et al.<sup>18</sup> protocol and used *domain.pl* to arrange domains at their respective position. Based on the additional associated domains, the identified *Zf-BED* containing genes were also classified.

**Evolution and diversity of Zf-BEDs in land plants.** To provide an understanding of evolution and diversity in Zf-BED proteins among land plants, we used an advanced comparative genomics tool, OrthoFinder<sup>31</sup>. The DIAMOND tool was used for fast sequence similarity searches. The graph clustering was done with the MCL clustering algorithm. The gene tree inference and distance matrix of the orthogroups were constructed with DendroBLAST<sup>32</sup>. A distance-based phylogeny tree was constructed using FastME 2.0. For multiple sequence alignments, MAFFT 7.0<sup>33</sup> was used. The maximum likelihood phylogenetic tree of large alignment was constructed using FastTreeMP<sup>34</sup> with 1000 bootstrap values. Species based phylogenetic tree was also constructed using the same method.

**Class-wise gene ontology and 3D structure prediction.** The ontology analysis was carried out using Gene Ontology Resource<sup>35</sup>. The class-wise ontology distribution was made in Microsoft PowerPoint. To predict the 3D structures of Zf-BED proteins, we have selected one representative protein sequence from each class. The 3D structures were predicted using the direct build model option with an open template. The server itself searches for the most similar template from PDB and predicts based on template information. The predicted structures were visualized through PyMol<sup>36</sup>. The conserved motifs were predicted through the MEME suite motif analysis tool<sup>37</sup> and graphically presented with TBtools.

**Chromosomal location, physical properties and cis-acting elements of Zf-BEDs.** To determine the chromosomal distribution of Zf-BED encoding genes in *Gossypium* sp., *A. thaliana* and *Z. mays*, the GFF3 files were downloaded from Cottongen, TAIR, and MaizeGDB database<sup>38–40</sup>, and plots of gene density on chromosomes were generated. To identify and compare the cis-acting regulatory element of *Zf-BED* genes of *G. arboreum*, *G. barbadense*, and *G. hirsutum*, upstream sequences of 2 kb sequences were retrieved from CottonFGD<sup>41</sup> and subjected to analyses in PlantCARE<sup>42</sup>. The physio-chemistry analysis of Zf-BED proteins was carried through the ExpASY ProtParam tool<sup>43</sup> including molecular weight (MW), theoretical pI (IEP), polarity, aliphatic index, and grand average of hydropathicity index (GRAVY).

**Expression profiling of *Zf-BED* genes.** To find the differential expression and response of *Zf-BED* genes in different tissues and stresses, we have retrieved FPKM values of *Zf-BED* genes from Cotton Functional Genomics Database (CottonFGD) under data-fetch and expression analysis with Bioproject PRJNA490626 and PRJNA594268. The collected data covered FPKM values of genes in the root, stems, leaf, oval fiber, and filament at different time intervals and under abiotic stresses including cold, heat, salt and drought<sup>44</sup>. For the biotic stresses, we downloaded RNA-seq SRA BioProject (PRJNA390823)<sup>25</sup> and PRJNA398803<sup>45</sup> SRA and processed them with transcriptomics pipelines. The FPKM values are used to generate a heatmap in TBTools under the heatmap graphics tool. The values were adjusted with log<sub>2</sub> base and row and column clusters options<sup>41</sup>.

For the expression profiling of *Arabidopsis thaliana* and *Zea mays Zf-BED* genes, the FPKM values were collected from *Arabidopsis* and maize RNA-seq Database<sup>46</sup> including tissue-specific expression (root, leaf, seedling, shoot, stem, meristem, flower, pollen, seed, embryo and endosperm), biotic stresses (*Acinetobacter radioresistens* strain SA188, *Blumeria graminis*, *Botrytis cinerea*, *Colletotrichum tofieldiae*, *Fusarium graminearum*, *Heterodera schachtii*, *Hyaloperonospora arabidopsidis*, *Microbacterium* sp, *Plutella xylostella*, *Pseudomonas syringae*, *Rhizobium* sp, *Rhizotonia solani*, *Scierotinia scierotiorum*, tobacco mosaic virus, *Verticillium dahliae*) and abiotic stresses (cold, dark, dehydration, drought, heat, hypoxia, irradiation, nutrient deficiency, osmotic, oxidative, ozone, salt, shade, UV, water, wounding).

**Plant material, growth conditions, and treatments.** All work was performed with *Arabidopsis thaliana* ecotype Col-0 and *Solanum lycopersicum* L. (tomato) cultivar Micro-Tom. Col-0 seeds, purchased from the Ohio State University's *Arabidopsis* Biological Resource Center, were directly sowed on Fafard Germination Mix in 90 mm × 68 mm Pöppelmann TEKU round pots and cold stratified in the cold room (4 °C) for three days. Plants were grown on growth racks at 23 °C with a 12 h light/12 h dark cycle with a light intensity of 120 μE/m<sup>2</sup>/s and watered every three days with tap water to keep soil moisture near field capacity. 10-day-old seedlings were transplanted to 12 × 6 Landmark Plastic flats with continuous irrigation till day 28. For biotic treatments, *Pseudomonas syringae* pv. tomato DC3000 (hereafter DC3000) or HrcC<sup>-</sup> (a type III secretion system mutant of DC3000) with OD<sub>600nm</sub> = 0.0002 in 10 mM MgCl<sub>2</sub> was syringe-infiltrated into the leaves of 28-day-old *Arabidopsis* plants as described<sup>47</sup>. Six plants with four leaves per plant were infiltrated and collected at 48 h for the following quantitative real-time polymerase chain reaction (qRT-PCR). For drought stress, six plants were grown without further input of water since day28 while six plants were supplied with 3 ml tap water per plant every day till day35. Heat stress was performed on 4-week-old *Arabidopsis* for 1 h at 37 °C while plants were kept at room temperature as a control group. Samples were collected at each corresponding time point and then quickly frozen in liquid nitrogen followed by storing at -80 °C for the next step. Three leaves were harvested for each replicate and three biological replicates were prepared for each experiment.

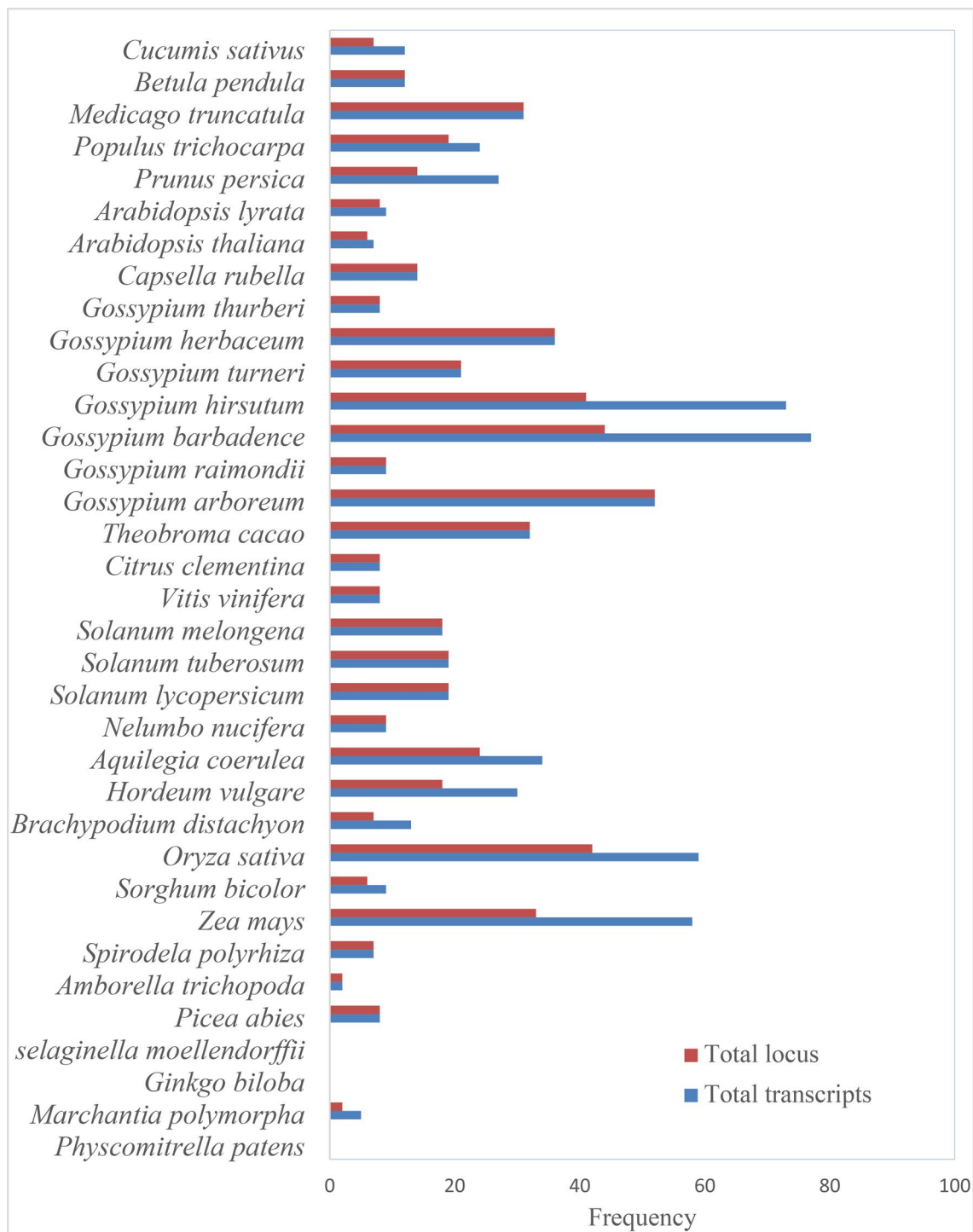
Micro-Tom seeds (Totally Tomatoes Seed Company, Randolph, WI, USA) were germinated on 150 mm Whatman filter papers saturated with tap water and placed on 150 mm × 15 mm Fisherbrand Petri Dishes. Plants were grown on growth racks at 23 °C with a 12 h light/12 h dark cycle with a light intensity of 120 μE/m<sup>2</sup>/s. Seven-day-old seedlings were transplanted to 12 × 6 Landmark Plastic flats with Fafard Germination Mix and watered every two days to keep moisture near field capacity till day 21. For biotic treatment, DC3000 or HrcC<sup>-</sup> with OD<sub>600nm</sub> = 0.0002 in 10 mM MgCl<sub>2</sub> was syringe-infiltrated into the leaves of 21-day-old Micro-Tom plants. Six plants with three leaves per plant were infiltrated and collected at 48 h post-inoculation for downstream RT-qPCR analyses. For drought stress, six plants were grown without further input of water since day21 while six plants were supplied with 3 mL of tap water per plant every day till day28. Samples were collected at each time point and then quickly frozen in liquid nitrogen followed by storage at -80 °C for the next step. Two leaves were harvested for each replicate and three biological replicates were prepared for each experiment.

**RNA extractions and qRT-PCR.** Leaf samples of Col-0 and Micro-Tom were fine ground with Bead Ruptor 96 Well Plate Homogenizer. Total RNA of *Arabidopsis* was extracted using the TRIzol (Invitrogen) according to the manufacturer's protocol<sup>48</sup>. 1 ml TRIzol was used per sample. The RNA pellet was dissolved in 20 μl DEPC-treated water and quantified by BioPhotometer Plus (Eppendorf). Total RNA of Micro-Tom was extracted using Direct-zol RNA MicroPrep kit (Zymo Research) according to the manufacturer's protocol. 10 μg of *Arabidopsis* and Micro-Tom RNA were then treated with DNase using the TURBO DNA-free Kit (Ambion) in a 20 μl reaction. cDNA was synthesized using SuperScript IV reverse transcriptase first-strand synthesis kit (Invitrogen) with 2 μg DNA-free RNA in a 10 μl reaction. PCR programs from both DNase treatment and reverse transcription reaction were performed by using Applied Biosystems 96 Well Thermal Cycler. qRT-PCR was performed on an ABI 7500 Fast PCR System (ThermoFisher Scientific) with 2X PowerUp SYBR green master mix (Applied Biosystems, ThermoFisher Scientific) using the following settings: 50 °C for 2 min and 95 °C for 10 min followed by 40 cycles of 95 °C for 15 s, 55 °C for 15 s and 72 °C for 1 min.

**Ethics approval and consent to participate.** The study is conducted in accordance with local and national regulations.

## Results

**Genome-wide identification and classification of *Zf-BED* encoding gene in land plants.** The whole-genome screening of *Zf-BED*-encoding genes in 35 plant species demonstrated that three lower land plants do not contain any sequences resembling the *Zf-BED* gene family. In total, we identified 750 *Zf-BED* encoding genes, which were distributed across the remaining 32 plant species. Interestingly, the highest number of genes were observed in the cotton plants including *G. barbadence* (77 genes), *G. hirsutum* (73 genes), *G. arboreum* (52 genes), *G. herbaceum* (36 genes), however *Oryza sativa* (59 genes) and *Zea mays* (58 genes) also



**Figure 1.** Distribution of Zf-BED encoding genes in land plant species.

had large number of genes followed by *Aquilegia coerulea* (34 genes), *Theobroma cacao* (32 genes), *Medicago truncatula* (31 genes), *Hordeum vulgare* (30 genes), *Prunus persica* (27 genes), *Populus trichocarpa* (24 genes), *G. turneri* (21 genes), *Solanum lycopersicum* (19 genes), *S. tuberosum* (19 genes), *Solanum melongena* (18 genes), *Capsella rubella* (14 genes), *Brachypodium distachyon* (13 genes), *Betula pendula* (12 genes), *Cucumis sativus* (12 genes), *Sorghum bicolor* (9 genes), *Nelumbo nucifera* (9 genes), *G. raimondii* (9 genes), *Arabidopsis lyrata* (9 genes), *Picea abies* (8 genes), *Vitis vinifera* (8 genes), *Citrus clementina* (8 genes), *G. thurberi* (8 genes), *Spirodela polyrhiza* (7 genes), *Arabidopsis thaliana* (7 genes), *Marchantia polymorpha* (5 genes) and *Amborella trichopoda* (2 genes) (Fig. 1, Tables S1–S2).

Based on the conserved domains, motifs, and their patterns in the primary sequence of proteins, we have classified all 750 genes into 22 major classes named class I–XXII. Of these 22 classes, class I

(Zf-BED–DUF-domain-Dimer\_Tnp\_hAT) and class II (Zf-BED) were more common in all plant species, containing 390 and 144 genes, respectively, followed by class III (Zf-BED-Zf-BED-DUF-domain-Dimer\_Tnp\_hAT), IV (Zf-BED–DUF-domain), V (Zf-BED-Dimer\_Tnp\_hAT), and VI (Zf-BED-PHD). The classification of Zf-BED-encoding genes reported several new domains associated with Zf-BED domains. The vast majority of these domains are involved in plant defense functions, such as the GRAS domain (gibberellin signaling regulator), PHD (Cys4-His-Cys3 motif in the plant homeodomain), WRKY (DNA-binding domain and biotic and abiotic stress regulator), NBS (Nucleotide-Binding Site; one of the major superfamilies of plant resistance genes), Sina domain (N-terminal RING finger domain), F-box with LRR-motif, etc. We have also observed diversity in the presence and absence of these 22 classes in 32 plant species such as the *Marchantia polymorpha*, which is a liverwort (bryophyte) and has only five genes i.e., one gene from class V (Zf-BED-Dimer\_Tnp\_hAT) and four genes from class XIII (GST\_N\_3–GST\_C\_3-Zf-BED; Glutathione S-transferase, C-terminal domain). Similarly, in the *Picea abies* genome, a gymnosperm plant, there were only eight genes, all of which belonged to class II (Zf-BED). The *Amborella trichopoda*, considered a basal species among the flowering plants, has only two genes (one from class I and another from class V). The *Spirodela polyrhiza*, a monocot near to basal of angiosperms, has seven Zf-BED genes i.e., three genes in class I and one in each class III, IV, V and XXII. In summary, a distinct pattern could be identified among monocots since all studied species, except *Oryza sativa*, contained Zf-BED genes that belonged to the first six classes (I–VI), while the dicot plants showed a more diverse sampling of the Zf-BED genes across various classes. In the dicot plants, especially in *Gossypium* sp., the *G. raimondii* genome exhibited only the first three classes (I–III), which is consistent with its conserved genome, while other *Gossypium* species showed a large number of genes in other classes. It was also observed that most of the classes were only specific to *Gossypium* sp. For example, the class VIII (GRAS-Zf-BED-DUF-domain-Dimer\_Tnp\_hAT), IX (GRAS-Zf-BED-DUF-domain-Dimer\_Tnp\_hAT-Peptidase\_C48) and XI (Zf-BED-DUF-domain-Dimer\_Tnp\_hAT-Peptidase\_C48) were found only in *Gossypium* sp. (Fig. 2, Figure S1, Table S3–S4). Overall, we found that diverse plant species possess different Zf-BED gene classes, highlighting the genomic diversity in land plants.

**Evolutionary study of Zf-BEDs in land plants.** To find the evolutionary relationships among the Zf-BED-encoding genes, we selected 25 land plants including *Aquilegia coerulea* (*Aqu*), *Arabidopsis thaliana* (*At*), *Betula pendula* (*Bpe*), *Brachypodium distachyon* (*Bra*), *Capsella rubella* (*Car*), *Citrus clementina* (*Cic*), *Cucumis sativus* (*Cuc*), *G. arboreum* (*Gar*), *Gossypium barbadense* (*Gba*), *Gossypium herbaceum* (*Ghe*), *G. hirsutum* (*Ghi*), *G. thurberi* (*Got*), *G. raimondii* (*Gra*), *Vitis vinifera* (*Gsv*), *Nelumbo nucifera* (*Nnu*), *Oryza sativa* (*OsR*), *Picea abies* (*Pab*), *Prunus persica* (*Pru*), *Solanum tuberosum* (*Sot*), *S. lycopersicum* (*Sol*), *S. melongena* (*Sme*), *Sorghum bicolor* (*Sob*), *Spirodela polyrhiza* (*Spi*), *Theobroma cacao* (*Tca*), and *G. turneri* (*Gtu*), and used their Zf-BED protein sequences to determine comparative genomics relationships, gene duplication events, gene trees, orthogroups, orthologs, putative xenologs, and species trees.

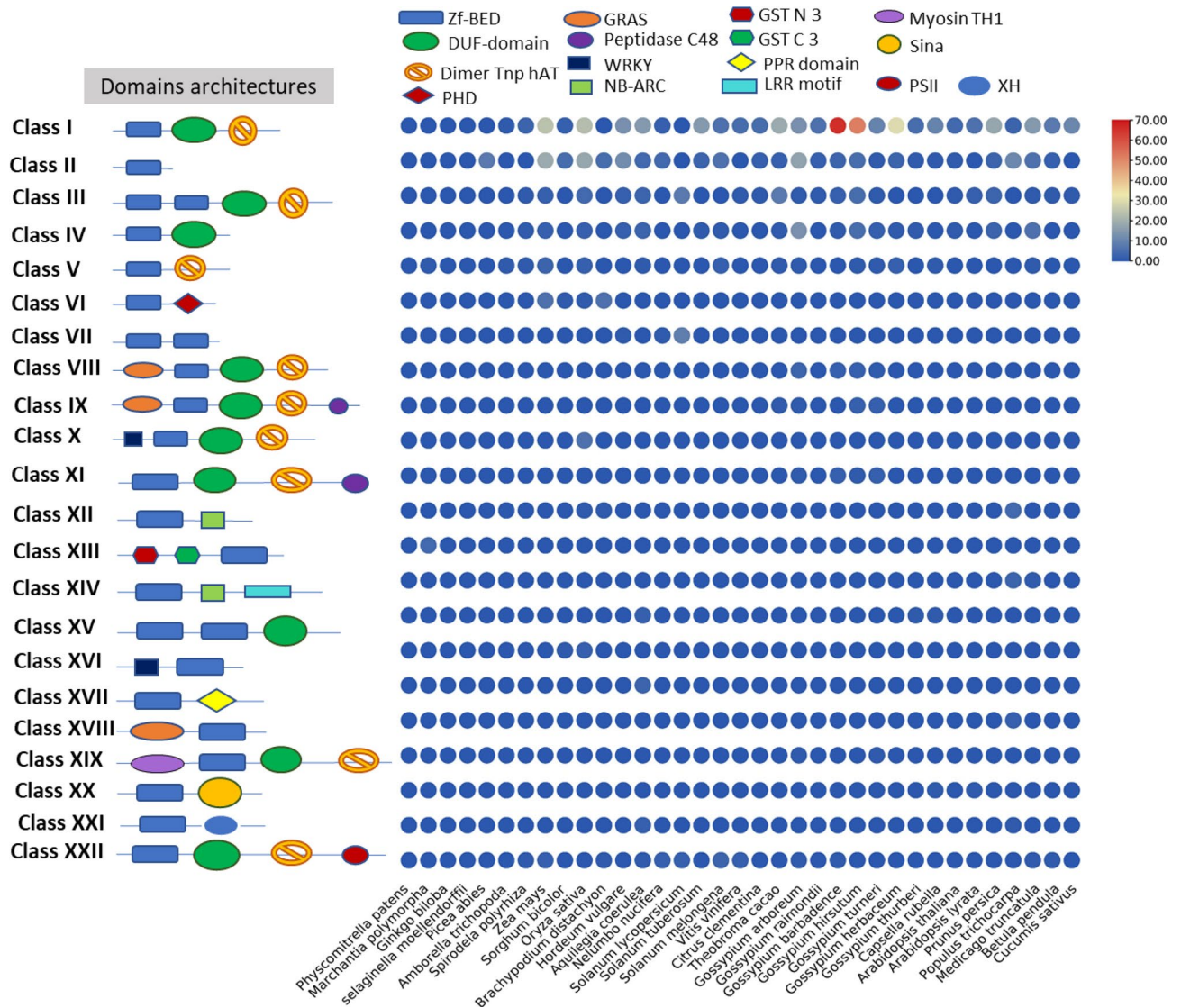
The comparative genomics analyses revealed that all Zf-BEDs were divided into 9 orthogroups (OG0; 330 genes, OG1; 168 genes, OG2; 23 genes, OG3; 20 genes, OG4; 15 genes and OG6 to OG8 each with 2 genes) (Figure S3, Table S7) covering 564 genes (99.8% of genes in orthogroups) with only one unassigned gene. Of these orthogroups, only one orthogroup was common to all 26 species, while three orthogroups were species-specific, containing only six genes each. Overall, the orthogroup mean and the median were computed as 62.7 and 15 genes, respectively (Figure S2, Table S5).

The species-wise orthogroups distribution showed that all species shared orthogroups (100% genes in orthogroups). Some orthogroups were species-specific; for instance, *O. sativa* had two species-specific orthogroups containing four genes (6.9% of genes). Similarly, *S. melongena* had one species-specific orthogroup containing two genes (13.3% of genes) (Figure S4, Table S6, Table S8, and Table S10–S13).

The orthogroup duplication events demonstrated that only five orthogroups (OG0 to OG4) passed through duplication events during the evolutionary time scale. The highest duplication numbers were recorded in OG0 (194 duplications) followed by OG1 (84 duplications), OG3 (12 duplications), OG4 (9 duplications), and OG2 (3 duplication events) (Fig. 3, Table S9).

For the species-based phylogenetic tree analysis, 26 representative species were included. The phylogenetic tree was divided into three major clades representing two clades of dicot plants and one clade of monocots. The gene duplication events at internal nodes and terminal nodes also demonstrated remarkable gene duplication in higher plants. For instance, in the case of the monocots clade, the highest gene duplication was observed in *O. sativa* (40 duplications) followed by *P. abies* (8 duplications). Similarly, in the dicot clades, the highest duplications are found in tetraploid species of cotton (*G. barbadense* and *G. hirsutum* with 29 and 31 gene duplication events, respectively). Furthermore, several gene duplication events were also found in the internal nodes, and the highest duplications occurred at N10 (49 duplications) (Fig. 3, Table S14). In summary, higher plants experienced more duplication events during evolutionary processes and different plant families carry species-specific orthogroups displaying their unique genetic makeup.

**Gene ontologies, conserved residues, and motifs in Zf-BEDs.** The gene ontology prediction of all Zf-BED genes indicated that all of them were involved in diverse molecular functions, mainly “DNA binding”, “nucleic acid binding”, and “protein dimerization” activities. However, different classes also exhibited additional diverse functions such as “hydrolase activity”, “regulation of transcription”, and “catabolic process”. The class-wise gene ontology demonstrated their putative roles in different molecular functions. For instance, class I is involved in protein dimerization and nucleic acid binding. Similarly, class II is implicated in DNA binding/nucleic acid binding activity. However, some classes were involved in more than one molecular function, such as class XIII, which is required for DNA binding, metabolic processes, aromatic amino acid metabolism, catalytic activity, and protein binding activity. Class XXII contained different domain architectures, and conse-

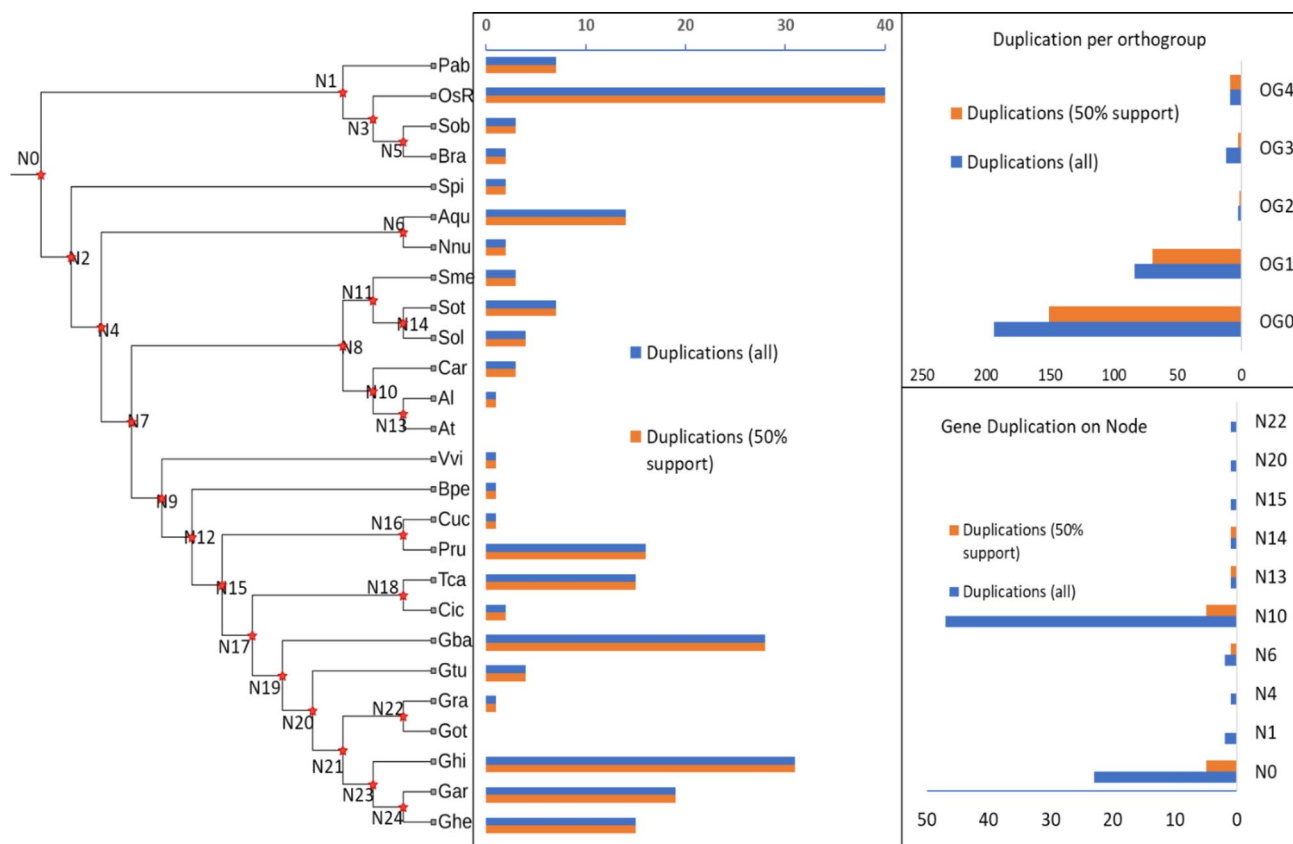


**Figure 2.** Classes of Zf-BED proteins based on conserved domains and their comparison among land plants.

quently its members engaged in diverse functions including photosynthesis, chlorophyll binding, RNA-DNA hybrid ribonuclease activity, hydrolase activity, etc. (Fig. 4, Table S15). In summary, the *Zf-BED* gene ontology varied with the addition of associated domains. Therefore, the associated domains appear to have a high impact on the function of the *Zf-BED* proteins.

The multiple sequence alignment of *Zf-BED* proteins demonstrated highly conserved functional residues in the active sites. However, the lower plants showed highly diverse residues at functional sites. For instance, W<sup>7</sup> was conserved in 87% sequences. Similarly, H<sup>9</sup>, F<sup>10</sup>, C<sup>20</sup>-C<sup>23</sup> (X<sup>2-4</sup> motif), G<sup>33</sup>, G<sup>38</sup>, T<sup>39</sup>, L<sup>42</sup>, K<sup>43</sup>, H<sup>45</sup> (His motif), L<sup>46</sup>, C<sup>52</sup>, D<sup>65</sup>, P<sup>83</sup>, F<sup>92</sup>, E<sup>124</sup>, W<sup>146</sup>, C<sup>247</sup>, D<sup>259</sup>, L<sup>318</sup>, W<sup>346</sup>, P<sup>372</sup>, K<sup>409</sup>, and L<sup>592</sup> sites were conserved in more than 60% of the *Zf-BED* proteins (Fig. S5). Most of the sequence variations at these sites were either genus-specific or species-specific. Conserved motif analysis also demonstrated the insertion or deletion of a specific motif in different classes of *Zf-BED* proteins. We also observed similar motif patterns in the same class members in the same genome. For instance, Class\_I of *GbaZf-BEDs* had similar motifs with the same pattern. The CX2-4C motif (motif-5) was conserved in all classes and all land plants. WX[YH]F (motif\_14) at the C-terminus was also found in most of the classes except classes XIII, XIV, XV, and XXI. Similarly, motif GTXXLXXH[LT] (motif\_13) was also conserved in the majority of the proteins. In summary, most of the functional motifs were conserved in more than 90% of the analyzed sequences. However, some motifs were species-specific, while the others were *Zf-BED* classes-specific (Figures S5–S6).

**Homology-based modeling of *Zf-BED* classes.** The homology-based 3D structures of different *Zf-BED* classes displayed different associations with their ligands. Most of the classes' structures were associated with zinc ions in addition to DNA. Some other ligands exhibited a tendency to interact with different molecules. We have determined representative 3D structures with their associated ligands for each class. The comparative



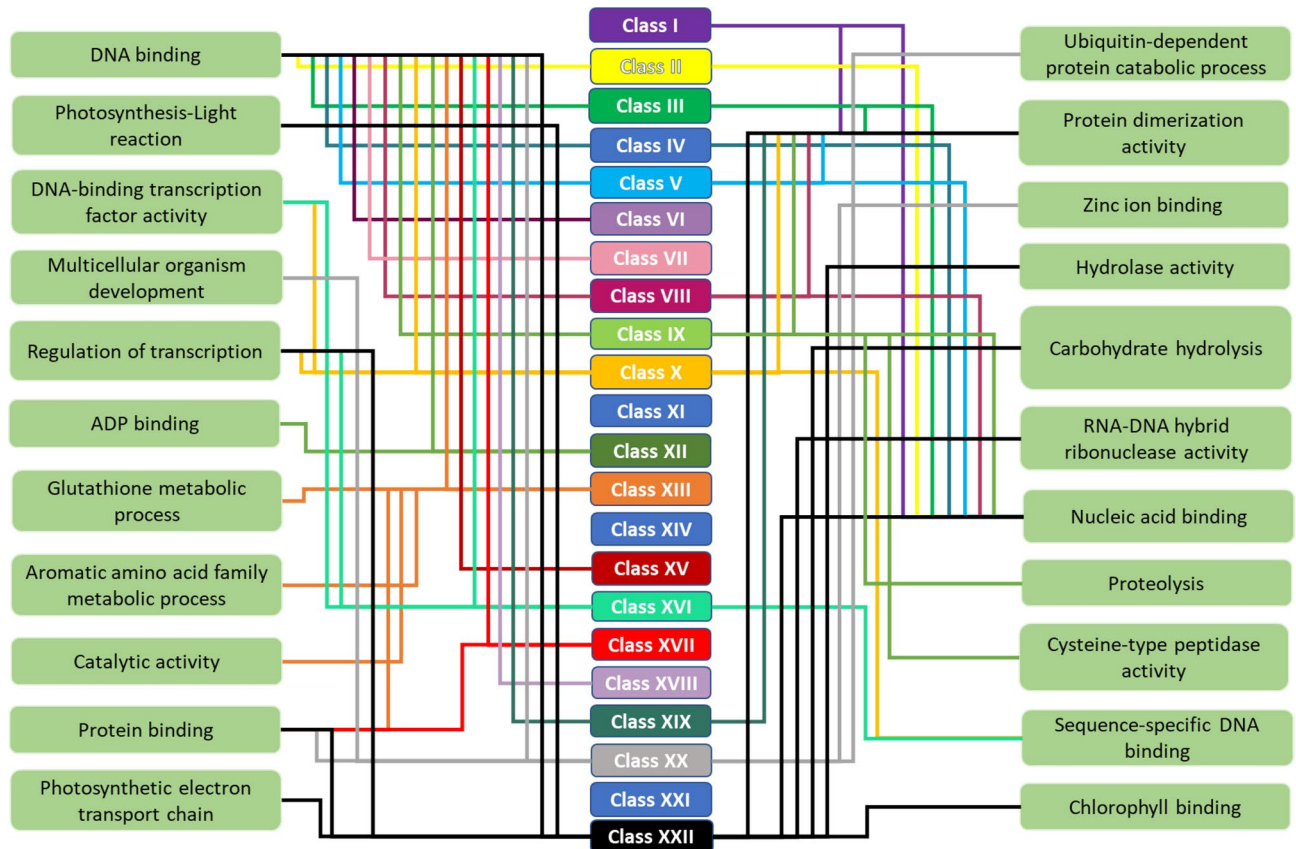
**Figure 3.** Evolutionary summary; species-tree with duplication events at the terminal and internal nodes and duplication at orthogroup level.

study provided clear differences in the structure, active sides, and ligands. For instance, class VI, VII, XX, and XXI proteins showed their association with zinc ion ligands in addition to some additional small molecules. Whereas, some classes e.g. X, XVI, and XVIII did not show any ligands in their structures. Furthermore, some classes (XV, XII, XIV, and XIX) were associated with DNA molecules. Similarly, class XIII was associated with glutathione (GSH) and NAG molecules (Fig. 5).

**Evolution of *Zf-BED* gene family in *Gossypium* sp..** To provide a detailed and comprehensive analysis of the *Zf-BED* genes in *Gossypium* species, we also presented the evolutionary relationship among five *Gossypium* sp., (*G. hirsutum*, *G. barbadense*, *G. arboreum*, *G. herbaceum* and *G. raimondii*) with *T. cacao*. The comparative genomics summarized that all *Zf-BED*s from 6 species were divided into 15 orthogroups encompassing 179 genes (98.9% of genes in orthogroups). Of these orthogroups, only four orthogroups were common in all species, while two orthogroups were species-specific. Similarly, only four genes were in species-specific orthogroups. Overall, the orthogroup mean and medians were computed as 11.9 and 11 genes, respectively.

The species-wise orthogroups distribution represented that most of the identified genes belonged to orthogroups. For instance, 98.0% of genes were in orthogroups with all genes of *Tca*, *Ghe*, *Gra*, and *Gba* and the minimum genes in orthogroups were present in *Ghi* and *Gar*. The orthogroup sharing analysis demonstrated that the highest number of orthogroups were shared between *Ghe* and *Ghi* with *Gba*. Similarly, the orthologs multiplicity also depicted significant similarity and uniqueness among six species. The species-based phylogenetic tree showed that *Tca* is the common ancestor for all *Gossypium* species, and *Ghe* has the most conserved *Zf-BED* encoding genes with *Tca*, followed by the other four cotton species. The duplication events at each terminal node demonstrated that *Tca* has had the high duplication rate, followed by *Ghe*, *Ghi*, *Gar*, and *Gba*, while *Gra* has experienced only a single duplication event (Fig. 6).

**Gene features and physio-chemistry of *Zf-BED* protein sequences.** The analyzed gene features included physio-chemistry and gene-specific attributes. The physio-chemistry included molecular weight, protein charge, isoelectric point, and grand average of hydropathy. Similarly, gene features comprised transcript length, CDS length, GC contents in CDS sequence, number of exons and introns, lengths of exons and introns, etc. We provided a detailed study of the five species. In the case of *G. arboreum* *Zf-BED*, the transcript length ranged between 0.25 and 5.5 kb, the coding sequence length ranged from 0.25 to 5.5 kb, the percentage of GC content in the coding sequence was 34% to 46%, exon number of genes of *G. arboreum* ranged from a minimum of 1 to a maximum of 17 exons, while the mean exon length ranged between 1 and 6 kbp, mean intron length of genes of *G. arboreum* ranged from 0 to 7 kb, longest protein consisted of 1713 amino acids and the shortest



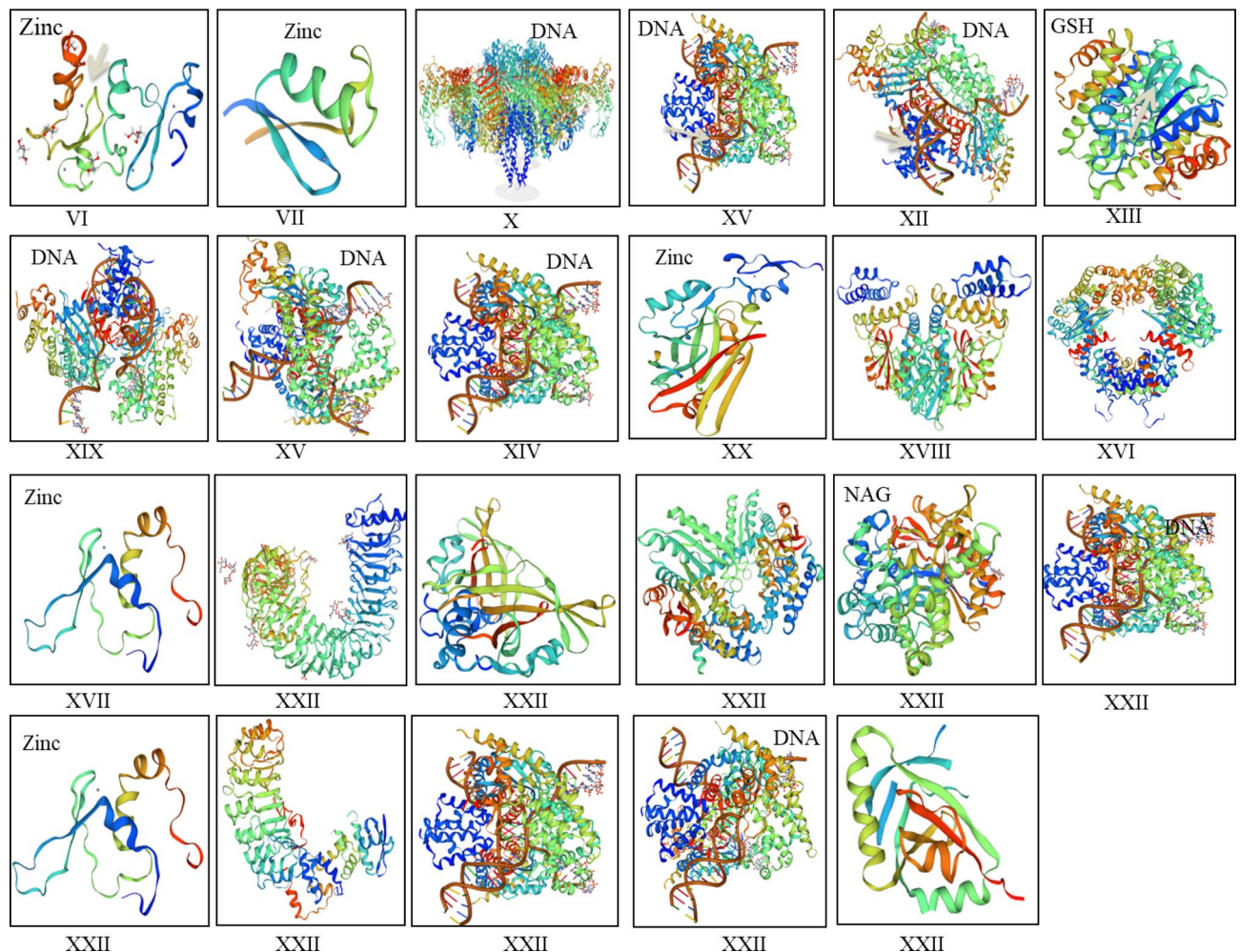
**Figure 4.** Zf-BEDs classes with their gene ontology-predicted molecular functions.

protein had 95 amino acid residues, the molecular weight was observed between 10.58 to 197.52 kDa, protein charge range was  $-20$  to  $+47$ , the isoelectric point range was 5.15 to 10.28 and the grand average of hydropathy range was  $-0.75$  to  $-0.21$ . Similar results were also observed in other *Gossypium* species (Figures S7–S8).

**Chromosomal mapping of Zf-BED genes.** To understand the effect of polyploidization on Zf-BED genes between diploid and tetraploid species, a genome-wide comparative distribution of Zf-BED genes on chromosomes was determined in all *Gossypium* species. Diverse species had a varied number of genes on different chromosomes. In A-like genomes (*Gar* and *Ghe*), all genes were relatively evenly distributed on all chromosomes except Chr#02, while in the D-like genome (*Gra*), all genes were localized to only six chromosomes. In the case of tetraploid species (*Ghi* and *Gba*), 20 chromosomes and 18 chromosomes possessed Zf-BED genes in *Ghi* and *Gba*, respectively. However, some chromosomes showed the highest number of genes in all species. For instance, D02 has the highest number of Zf-BED genes in *G. barbadense* and *G. hirsutum*, followed by Chr#07 in all species except *G. arboreum*. Similarly, some chromosomes did not contain any Zf-BED genes, such as Ga2, Ga5, Ga7, Ga10 did not bear any Zf-BED genes in *G. arboreum*. A similar pattern was also observed in *G. raimondii* (Gr2, Gr3, Gr5, Gr8, Gr10, Gr11, and Gr12), *G. barbadense* (GbA01, GbA04, GbD01, and GbD10), and in *G. hirsutum* (GhA01, GhA04, GhA09, GhA10, GhD07, and GhD10). In addition to *Gossypium* sp., the chromosomal locations of the *AtZf-BED* and *ZmZf-BED* genes were also mapped on their respective chromosomes. We observed that all *AtZf-BED* genes were localized only on three chromosomes (AT1, AT3, and AT4) while *ZmZf-BED* genes were distributed on all chromosomes except chromosome no. 6 (Zm6) (Figure S9).

**Cis-acting regulatory elements prediction of Zf-BED genes.** The identified cis-acting regulatory elements were classified into different categories including response to hormones, response to stresses, growth, and during development. We observed that most of the gene promoters contained gibberellin, abscisic acid, salicylic acid, jasmonic acid, and auxin hormone-responsive elements. Similarly, in stress-responsive elements, we observed low temperature, light, wound, and drought, anoxic and anaerobic responsive factors. Furthermore, the growth and developmental cis-acting elements included zein metabolism regulation, seed-specific regulation, meristem expression, endosperm expression, circadian control, and cell cycle regulation. Some additional elements including MYBHv1 binding site, ATBP-1, a MYB binding site involved in flavonoid biosynthetic genes regulation, and CMA3 were also observed in some gene regulatory regions. The comparison of cis-acting elements among *G. arboreum*, *G. hirsutum*, and *G. barbadense* demonstrated common regulatory elements including gibberellin-responsive elements, low-temperature responsiveness, as well as wound-responsive elements, light-responsive, and salicylic acid responsive elements (Tables S24–S25).



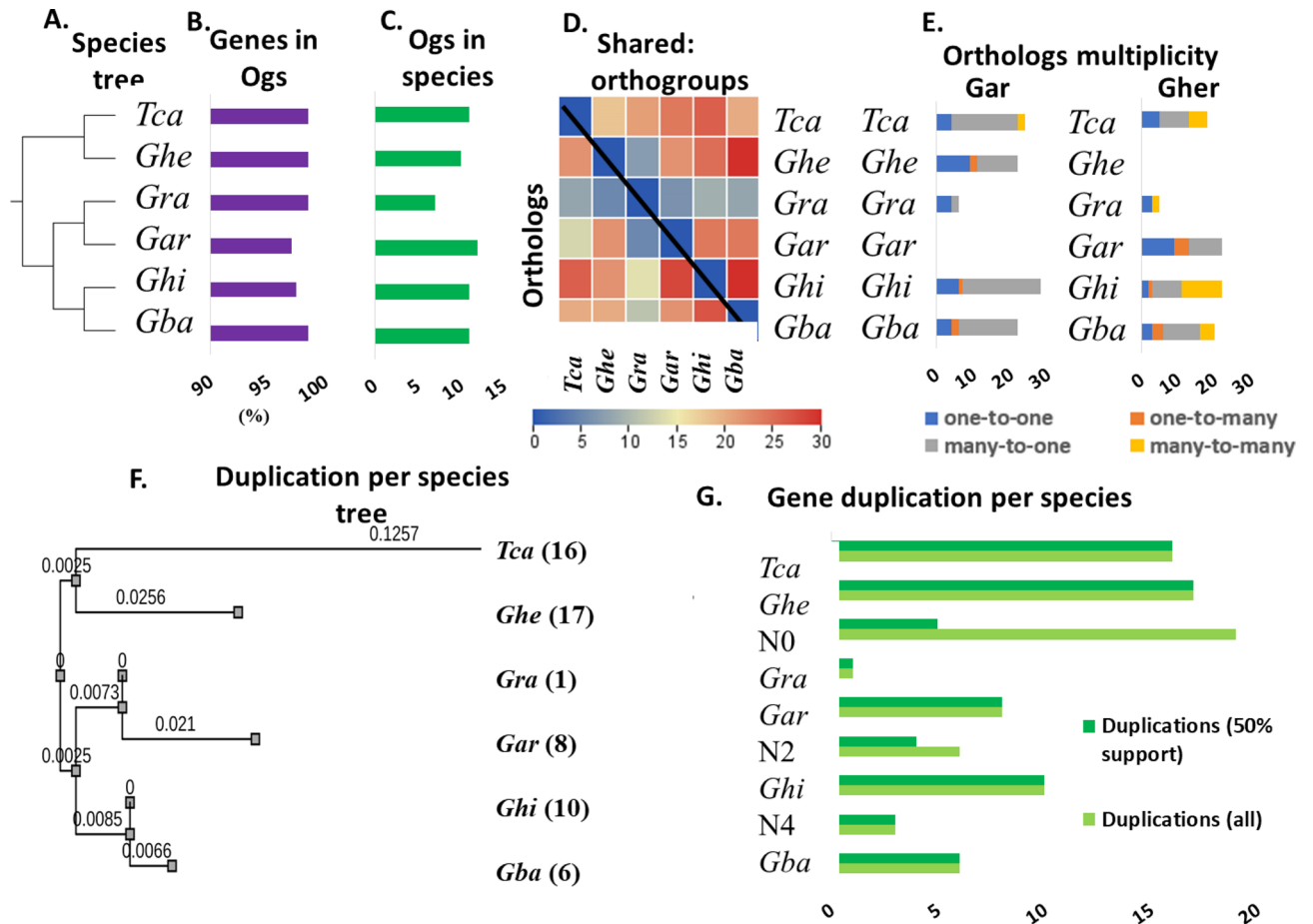


**Figure 5.** Protein 3D models for different classes of Zf-BED with their ligands; Zinc, DNA, GSH.

**Expression profiling of Zf-BED genes in cotton, maize, and *Arabidopsis*.** The expression profiling of Zf-BED genes showed variation with respect to tissues and stresses. In this study, we have included expression profiling of Zf-BED genes in *Gossypium* sp., maize, and *Arabidopsis* in different tissues and under various stresses. The expression profiling and cluster correlation of Zf-BED in *G. arboreum* demonstrated multiple clusters in different tissues including fiber, leaf, ovule, flower, stem, and root. For instance, a cluster containing GaZf-BED20, GaZf-BED17, and GaZf-BED31 showed relatively high expression in leaf, flower, stem, and root. Similarly, another cluster (GaZf-BED11, GaZf-BED33, GaZf-BED30, GaZf-BED09, GaZf-BED16) had high transcript abundance in the ovule, flower, stem, and root. Some genes showed high expression levels in stem and root tissues, while other genes displayed little expression in ovule, flower, stem, and root. Interestingly, however, none of the genes displayed any expression in leaf and fiber. In summary, only a few GaZf-BED genes presumably play roles in tissue development (Figure S10, Table S27).

The expression profiling of Zf-BED genes in *G. hirsutum* in diverse tissues (like fiber, flower, ovule, and leaf at different time intervals) showed different expression patterns of genes. The gene clustering cladogram demonstrated that some genes have putative roles in all tissues. For instance, gene Gh\_A08G102900.1 had increased transcript levels in all four tissues. On the other hand, a set of genes exhibited tissue-specific expression such as Gh\_D08G097100.1, Gh\_D07G157800.1, Gh\_A08G102900.1, and Gh\_D07G157800.1. Gh\_D06G202500.1 had the highest expression in the leaf (Table S28). The stress-specific expression of Zf-BED demonstrated their putative functions under different abiotic stresses including cold, drought, heat, and salt at different time intervals. A cluster of genes such as Gh\_A08G102900.1, Gh\_D08G097100.1, Gh\_D07G157800.1, Gh\_D06G202500.1, and Gh\_D09G000500.1 showed high expression in all stresses including cold, drought, heat, and salt (Table S28). However, several genes exhibited tissue-specific expression. Moreover, we observed a cluster of genes that did not show any expression during the applied stresses.

The expression profiling of *Z. mays* Zf-BED genes also demonstrated tissue-specific and stress-specific. For instance, ZmZf-BED43 (Zm00001d033361) gene is highly expressed only in anther and endosperm, while a cluster of genes including ZmZf-BED43 (Zm00001d013336), ZmZf-BED26 (Zm00001d017846), ZmZf-BED34 (Zm00001d022534), ZmZf-BED38 (Zm00001d026358), and ZmZf-BED02 (Zm00001d003128) share a co-expression pattern in all tissues (Figure S11). In addition, the abiotic stress response of ZmZf-BED genes also followed



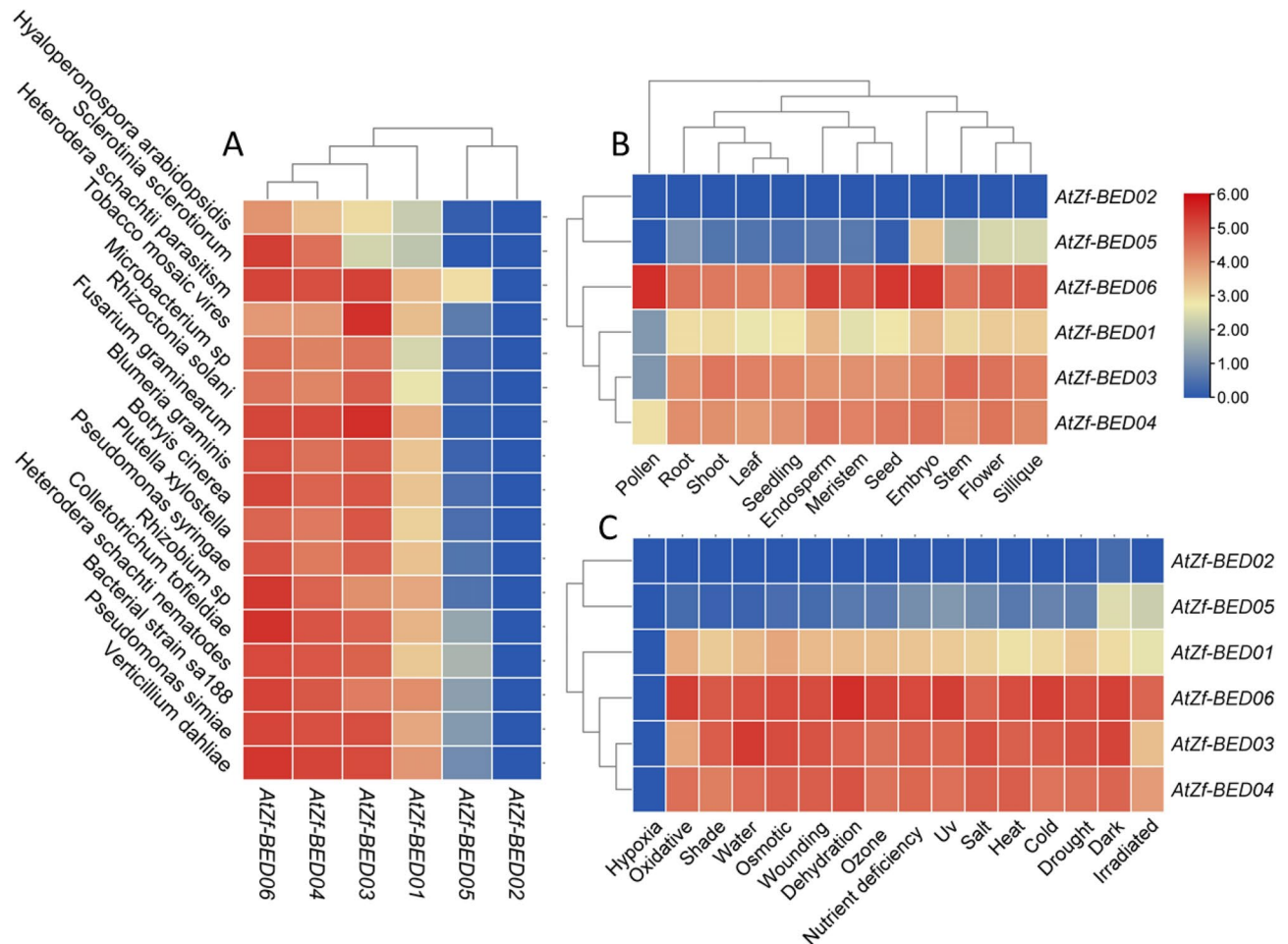
**Figure 6.** Summary of evolutionary analysis of *Zf-BEDs* in *Gossypium* sp. (A) Species-based phylogenetic tree, (B) percentage of genes in orthogroups, (C) number of species-specific orthogroups, (D) heatmap of orthologs statistics one-to-one, (E) Heatmap of orthologs statistics one-to-many.

the same pattern. A cluster of genes (*ZmZf-BED34*, *ZmZf-BED45*, *ZmZf-BED17*, and *ZmZf-BED26*) co-expressed under different abiotic stresses with a variable range of FPKM values (Figure S12). In contrast to biotic and tissues specific stresses, the response of *ZmZf-BED* was highly variable for biotic stresses. For instance, a pair of genes (*ZmZf-BED26* and *ZmZf-BED17*) displayed high expression under all biotic stresses indicating their key roles in biotic stress regulation in *Z. mays*. Whereas, other *ZmZf-BED* genes were highly treatment-specific e.g. *ZmZf-BED34* had high transcript levels under pathogen stresses including Maize Mosaic Virus, *Fusarium virguliforme*, etc. (Figure S13, Table S29).

The tissue-specific expression profiling of *AtZf-BED* in *Arabidopsis* demonstrated that all genes had high relative expression except *AtZf-BED05* (AT3G48770; class-II; *Zf-BED*) and *AT1G36095* (class-IV; *Zf-BED-DUF-domain*) forming a common cluster on the dendrogram. Another cluster, which included *AtZf-BED06* (AT4G15020), *AtZf-BED01* (AT1G18560), *AtZf-BED03* (AT1G79740), and *AtZf-BED04* (AT3G17450), had high transcripts abundance in all tissues (pollen, root, shoot, leaf, seedling, endosperm, meristem, seed, embryo, stem, flower and silique). However, the level of transcripts varied among the tested tissues. In contrast to tissue-specific expression, the biotic stress-specific expression was highly variable depending on stress sources. For instance, *AtZf-BED01* (AT1G18560), *AtZf-BED03* (AT1G79740), *AtZf-BED04* (AT3G17450), and *AtZf-BED06* (AT4G15020) had increased expression when challenged with *Verticillium dahliae*, followed by *Fusarium graminearum*, Tobacco Mosaic Virus, and bacterial disease. Similarly, under abiotic stresses, *AtZf-BED06* (AT4G15020), *AtZf-BED03* (AT1G79740), and *AtZf-BED04* (AT3G17450) showed heightened transcript accumulation under all stresses including cold, dark, drought, salt, wounding and water deficit (Fig. 7, Table S30).

### Experimental verification of *Zf-BED* genes expression in *Arabidopsis* and Micro-Tom tomato plants.

To decipher the relationship of *Zf-BED* family members in *Arabidopsis* and Micro-Tom in response to biotic and abiotic stress, three *Zf-BED* genes in *Arabidopsis* and four orthologous genes in Micro-Tom were selected to examine the transcript level changes under various stresses using qRT-PCR. In *Arabidopsis*, *AtZf-BED01* (AT1G18560) and *AtZf-BED03* (AT1G79740) were downregulated upon DC3000 infection compared to HrcC<sup>-</sup>, while *AT4G15020* showed the inverse trend (Fig. 8). Under heat treatment, *AtZf-BED01* (AT1G18560) transcripts were elevated whereas other *Zf-BED* genes didn't display significant changes. Furthermore, *AtZf-BED01* (AT1G18560) and *AtZf-BED03* (AT1G79740) were upregulated significantly in response to drought



**Figure 7.** Expression profiling of *AtZF-BEDs*. (A) under various biotic stresses, (B) in different tissues, (C) under different abiotic stresses.

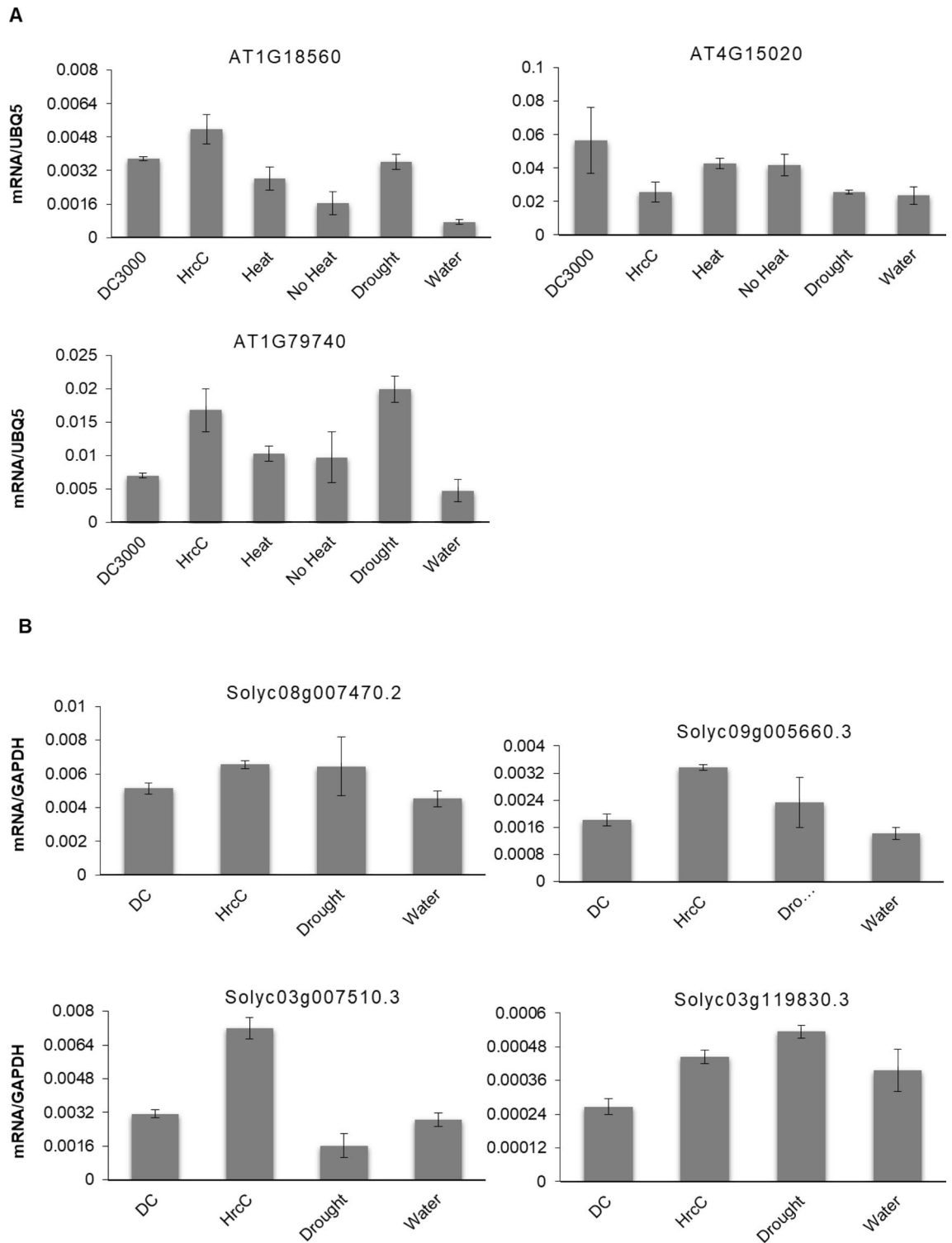
stress. In Micro-Tom, all *Zf-BED* genes showed lower expression levels under DC3000 treatment compared to HrcC-. As for abiotic stress, *Solyc08g007470.2*, *Soly09g005660.3*, and *Solyc03g119830.1* expression was upregulated upon drought, while *Solyc03g007510.3* was downregulated (Fig. 8). Overall, the representative *Zf-BED* genes in *Arabidopsis* and Micro-Tom were shown to be involved in biotic and abiotic stresses.

## Discussion

The study of plant evolution provides evidence to explain the diversity of plants across the geologic time frame. Plant evolution has been shaped by key morphological and developmental novelties that improved plant adaptation, and studying morphological traits is a prime feature of paleobiology and leads to fundamental insights<sup>50</sup>. The gene family expansion and domain rearrangements in land plants hamper the identification of orthologous proteins in different plant species<sup>50</sup>. However, the comparative study on a genome-wide scale enables researchers to discover duplications and other events known to be major drivers of plant speciation<sup>51</sup>.

Owing to the evolution of diverse transcription factors, plants have acquired well-developed defense mechanisms and developed resilience to environmental stresses<sup>52</sup>. Zinc fingers execute their function via binding to nucleic acids and participate in transcriptional regulation processes. The *Zf-BED* genes are one of the important contributing factors in the plant defense system and have been functionally studied in many plant species including *Arabidopsis*, wheat, maize, *Medicago*, sorghum, etc.<sup>15</sup>.

The *Zf-BED* domains are widely spread in the eukaryotic genomes and actively participate in DNA binding and protein–protein interactions<sup>53–56</sup>. Recent studies demonstrated their integration as decoy domains with other resistance-related domains and their role in the regulation of diverse host functions. The *Zf-BED* domain with NLR conferred resistance to yellow rust and blast in rice<sup>57,58</sup>. With the identification of *Zf-BED* as a resistance domain, several experimental studies were carried out and validated their putative role in plant development and adaptation<sup>58–61</sup>. In contrast to previous gene-based studies, here we used genome-wide approaches for large-scale, unbiased identification and characterization of plant *Zf-BEDs*. In the current study, we have conducted an evolutionary study of the *Zf-BED* genes in 35 land plants including mosses, bryophytes, gymnosperms, and angiosperms. A total of 750 *Zf-BED* domain encoding genes were identified, representing 32 plant species. The lower plants such as *Physcomitrella patens* (mosses) and *Selaginella moellendorffii* (lycophyte) did not



**Figure 8.** RT-qPCR analysis of *Zf-BED* mRNAs accumulation in Col-0 and Micro-Tom treated by biotic and abiotic stresses. **(A)** Transcript levels of *Zf-BED* genes in 4-week-old Col-0 genotype upon treatment with *Pseudomonas syringae* pv. *tomato* DC3000 (DC3000) and effector less mutant strain Pto DC3000 HrcC- (HrcC-) at 48 h, 37 °C (Heat) and room temperature (No heat) at 1 h, as well as no irrigation (Drought) and normal irrigation (Water) at 7 days. Gene expression was assessed using reference gene *UBQ5* in *AT1G18560*, *AT4G15020* and *AT1G79740*. **(B)** Transcript levels of *Zf-BED* genes in 3-week-old Micro-Tom genotype upon treatment with DC3000 and HrcC- at 48 h, as well as no irrigation (Drought) and normal irrigation (Water) at 7 days. Gene expression was assessed using reference gene *GAPDH* in *Solyc08g007470.2*, *Solyc09g005660.3*, *Solyc03g007510.3* and *Solyc03g119830.3*.

possess any Zf-BED encoding genes. Gene copy numbers varied from species to species, showing evolutionary diversity among land plants with higher plants showing numbers larger than those in lower plants. All identified genes were classified into 22 major classes, presenting the most comprehensive Zf-BED proteins analysis to date. In these classes, the presence of “Dimer\_Tnp\_hAT” as a decoy domain was common in all plants, pointing toward their essential roles in plant development<sup>60</sup>. Despite their important functionalities, very few classes of Zf-BED proteins have been reported in the literature to date; for instance, zFBED-NB-ARC(NLR) (class XII; ZfBED-NBS and class XIV; ZfBED-NBS-LRR) domain architecture was reported as conferring disease resistance in plants<sup>19,57,61</sup>. These classes were only found in *Brachypodium distachyon* (*Bradi5g22179.1*), *Oryza sativa* (*OsR498G0409301900.01.T01*), *Medicago truncatula* (*XP\_013447741.2*), and *Populus trichocarpa* (*Potri.001G404800*, *Potri.001G405100*, *Potri.019G002700*, *Potri.T013632*, and *Potri.T107066*). However, they were only studied in wheat<sup>57</sup> and rice<sup>58</sup>, while other identified ZfBED-NLRs may also have great potential in the plant immune system. Other than ZfBED-hAT and ZfBED-NLR, the remaining classes also have important decoy domains like DUF26, WRKY<sup>62,63</sup>, GRAS (transcription factors that regulate plant development)<sup>64,65</sup>, Sina, Glyco\_hydro\_1, and many others (Table S3). The comparative study of these 22 major classes among land plants demonstrated tandem duplications, deletions, and insertions of the domain during genome evolution. Similar evolutionary mechanisms were also observed in other protein families<sup>66–68</sup>. The molecular functions of 22 classes revealed their role in multiple biological processes and the functional diversity of classes appears to be due to the presence of decoy domains with the Zf-BED domain. For example, the hAT decoy domain has a potential role in plant development<sup>60</sup>. Similarly, the presence of the NLR domain provides an additional link to plant immunity<sup>58</sup>.

The sequence alignment and motif discovery identified important conserved signatures in land plants. We observed some highly conserved residues like W<sup>7</sup>, H<sup>9</sup>, and C<sup>20</sup>-C<sup>23</sup> in all species. These conserved residues and signatures show their important roles in the structure and function of the Zf-BED domain<sup>9</sup>. Based on these signature variations, the Zf-BED proteins were divided into nine groups<sup>11–15</sup>. The expression profiling of *Zf-BED* genes in *Arabidopsis*, *Z. mays*, *G. arboreum*, and *G. hirsutum* also demonstrated their putative role in different tissues, as well as under biotic and abiotic stresses. Different orthogroups (OG1, OG2, and OG3) of Zf-BED responded differently in various tissues under diverse stresses. The OG1 (*AT4G15020.1*, *Solyc09g005660.2*, *Ghir\_D05G011710.1*, *Ghir\_D13G002100.1*, *Ghir\_D13G002100.2*, *Zm00001d000412*, *Zm00001d004256*, *Zm00001d010895*, *Zm00001d016617*) was mainly involved in biotic and abiotic stresses as compared to other two orthogroups (Table S31). Several prior studies presented the responses of *Zf-BED* genes under biotic and abiotic stresses<sup>59</sup>. The overexpression of Zf-BED in rice lines increased drought tolerance<sup>59</sup>. The prediction of *cis*-regulatory elements also provides insight into the response of genes to different stresses. We also observed and identified several additional stress-responsive elements in the regulatory region of *Gossypium* sp. *Zf-BED* genes.

## Conclusions

Zf-BED encoding genes play important roles in plant development and adaptation. We identified a total of 750 Zf-BED encoding genes in 35 land plants and classified them into 22 major classes. The comparative study of land plants highlighted several duplications, deletions, and insertions in the genome during the evolutionary processes. A handful of genes were common in all plant species, while the shared and unique genes were different at different classification levels of kingdom Planta. The addition of the decoy domain with the Zf-BED domain provided an additional structural and functional role to the *Zf-BED* genes. The expression profiling also demonstrated that closely related genes have conserved stress-responsive functions in different plants. This is the first report of a genome-wide identification, characterization, and evolutionary study of Zf-BED encoding genes in land plants, which provides primary data for further functional studies that will help guide research efforts on plant adaptation under biotic and abiotic stresses.

## Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Received: 29 January 2022; Accepted: 12 July 2022

Published online: 19 July 2022

## References

- Feurtado, J. A. *et al.* The Arabidopsis C2H2 zinc finger INDETERMINATE DOMAIN1/ENHYDROUS promotes the transition to germination by regulating light and hormonal signaling during seed maturation. *Plant Cell* **23**, 1772–1794 (2011).
- Giri, J., Vij, S., Dansana, P. K. & Tyagi, A. K. Rice A20/AN1 zinc-finger containing stress-associated proteins (SAP1/11) and a receptor-like cytoplasmic kinase (OsRLCK253) interact via A20 zinc-finger and confer abiotic stress tolerance in transgenic Arabidopsis plants. *New Phytol.* **191**, 721–732 (2011).
- Zhou, H., Hirata, M., Osawa, R., Fujino, K. & Kishima, Y. Detainment of Tam3 transposase at plasma membrane by its BED-zinc finger domain. *Plant Physiol.* **173**, 1492–1501 (2017).
- Park, S. *et al.* Structural basis for interaction of the tandem zinc finger domains of human muscleblind with cognate RNA from human cardiac troponin T. *Biochemistry* **56**, 4154–4168 (2017).
- Babu, M. M., Iyer, L. M., Balaji, S. & Aravind, L. The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* **34**, 6505–6520 (2006).
- AbuZayed, R. *et al.* Correlates of zinc finger BED domain-containing protein 3 and ghrelin in metabolic syndrome patients with and without prediabetes. *Hormone Mol. Biol. Clin. Investig.* **37** (2019).
- Tue, N. T. *et al.* DREF plays multiple roles during Drosophila development. *Gene Regulat. Mech.* **1860**, 705–712 (2017).
- Michel, K. & Atkinson, P. Nuclear localization of the Hermes transposase depends on basic amino acid residues at the N-terminus of the protein. *J. Cell. Biochem.* **89**, 778–790 (2003).
- Aravind, L. The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.* **25**, 421–423 (2000).

10. Berg, J. M. & Shi, Y. The galvanization of biology: A growing appreciation for the roles of zinc. *Science* **271**, 1081–1085 (1996).
11. Jenkins, T. H., Li, J., Scutt, C. P. & Gilmartin, P. M. Analysis of members of the *Silene latifolia* Cys 2/His 2 zinc-finger transcription factor family during dioecious flower development and in a novel stamen-defective mutant *ssf1*. *Planta* **220**, 559–571 (2005).
12. Ori, N. *et al.* The I2C family from the wilt disease resistance locus I2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. *Plant Cell* **9**, 521–532 (1997).
13. Schumann, U. *et al.* Requirement of the C3HC4 zinc RING finger of the Arabidopsis PEX10 for photorespiration and leaf peroxisome contact with chloroplasts. *Proc. Natl. Acad. Sci.* **104**, 1069–1074 (2007).
14. Siddiqui, K., Mansoor, S., Briddon, R. W. & Amin, I. Diversity of alphasatellites associated with cotton leaf curl disease in Pakistan. *Virology* **6**, 41–52 (2016).
15. Gupta, S. K., Rai, A. K., Kanwar, S. S. & Sharma, T. R. Comparative analysis of zinc finger proteins involved in plant disease resistance. *PLoS ONE* **7**, e42578 (2012).
16. Takatsuji, H. Zinc-finger proteins: The classical zinc finger emerges in contemporary plant science. *Plant Mol. Biol.* **39**, 1073–1078 (1999).
17. Chen, C. *et al.* BED domain-containing NLR from wild barley confers resistance to leaf rust. *Plant Biotechnol. J.* **19**, 1206–1215. <https://doi.org/10.1111/pbi.13542> (2021).
18. Hussain, A. *et al.* Genome-wide identification and classification of resistance genes predicted several decoy domains in *Gossypium* sp. *Plant gene* **24**, 100250 (2020).
19. Marchal, C., Haberer, G., Spannagl, M. & Uauy, C. Comparative genomics and functional studies of wheat BED-NLR loci. *Genes* **11**, 1406 (2020).
20. Chen, C. *et al.* Ancient BED-domain-containing immune receptor from wild barley confers widely effective resistance to leaf rust. *bioRxiv* (2020).
21. Li, X., Kapos, P. & Zhang, Y. NLRs in plants. *Curr. Opin. Immunol.* **32**, 114–121 (2015).
22. Mishra, B., Kumar, N. & Mukhtar, M. S. systems biology and machine learning in plant–pathogen interactions. *Mol. Plant Microbe Interact.* **32**, 45–55. <https://doi.org/10.1094/MPMI-08-18-0221-FI> (2019).
23. Mukhtar, M. S. Engineering NLR immune receptors for broad-spectrum disease resistance. *Trends Plant Sci.* **18**, 469–472. <https://doi.org/10.1016/j.tplants.2013.08.005> (2013).
24. Mishra, B., Kumar, N. & Mukhtar, M. S. Network biology to uncover functional and structural properties of the plant immune system. *Curr. Opin. Plant Biol.* **62**, 102057. <https://doi.org/10.1016/j.pbi.2021.102057> (2021).
25. McHale, L., Tan, X., Koehl, P. & Michelmore, R. Plant NBS-LRR proteins: Adaptable guards. *Genome Biol.* **7**, 1–11 (2006).
26. Mishra, B., Sun, Y., Howton, T. C., Kumar, N. & Mukhtar, M. S. Dynamic modeling of transcriptional gene regulatory network uncovers distinct pathways during the onset of Arabidopsis leaf senescence. *NPJ Syst. Biol. Appl.* **4**, 35. <https://doi.org/10.1038/s41540-018-0071-2> (2018).
27. Bernacki, M. J., Rusaczek, A., Czarnocka, W. & Karpinski, S. Salicylic acid accumulation controlled by LSD1 is essential in triggering cell death in response to abiotic stress. *Cells* <https://doi.org/10.3390/cells10040962> (2021).
28. Naqvi, R. Z. *et al.* Transcriptomics reveals multiple resistance mechanisms against cotton leaf curl disease in a naturally immune cotton species, *Gossypium arboreum*. *Sci. Rep.* **7**, 15880. <https://doi.org/10.1038/s41598-017-15963-9> (2017).
29. Zaidi, S. S. *et al.* Molecular insight into cotton leaf curl geminivirus disease resistance in cultivated cotton (*Gossypium hirsutum*). *Plant Biotechnol. J.* **18**, 691–706. <https://doi.org/10.1111/pbi.13236> (2020).
30. Naqvi, R. Z. *et al.* Transcriptomic analysis of cultivated cotton *Gossypium hirsutum* provides insights into host responses upon whitefly-mediated transmission of cotton leaf curl disease. *PLoS ONE* **14**, e0210011. <https://doi.org/10.1371/journal.pone.0210011> (2019).
31. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
32. Kelly, S. & Maini, P. K. DendroBLAST: Approximate phylogenetic trees in the absence of multiple sequence alignments. *PLoS ONE* **8**, e58537 (2013).
33. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
34. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
35. Consortium & G. O., The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
36. DeLano, W.L., Pymol: An open-source molecular graphics tool. *Protein Crystallogr.* p. 82–92 (2002).
37. Bailey, T. L. *et al.* MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
38. Yu, J. *et al.* CottonGen: A genomics, genetics and breeding database for cotton research. *Nucleic Acids Res.* **42**, D1229–D1236 (2014).
39. Poole, R. L. *The TAIR database*, in *Plant bioinformatics*. p. 179–212 (Springer, 2005).
40. Lawrence, C. J., Dong, Q., Polacco, M. L., Seigfried, T. E. & Brendel, V. MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* **32**, D393–D397 (2004).
41. Zhu, T. *et al.* CottonFGD: An integrated functional genomics database for cotton. *BMC Plant Biol.* **17**, 1–9 (2017).
42. Rombauts, S., Déhais, P., Van Montagu, M. & Rouzé, P. PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Res.* **27**, 295–296 (1999).
43. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*, 571–607 (2005).
44. Hu, Y. *et al.* *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* **51**, 739–748 (2019).
45. Wendel, J. F. & Grover, C. E. Taxonomy and evolution of the cotton genus, *Gossypium*. *J. Cotton* **57**, 25–44 (2015).
46. Zhang, H. *et al.* A comprehensive online database for exploring ~ 20,000 public Arabidopsis RNA-seq libraries. *Mol. Plant* **13**, 1231–1233 (2020).
47. Liu, X. *et al.* Bacterial leaf infiltration assay for fine characterization of plant defense responses using the *Arabidopsis thaliana*-*Pseudomonas syringae* pathosystem. *JoVE Immunol. Infect.* <https://doi.org/10.3791/53364> (2015).
48. Liu, X. *et al.* Arabidopsis thaliana atGCN2 kinase is involved in disease resistance against pathogens with diverse life styles. *Int. J. Phytopathol.* **4**, 93–104 (2015).
49. Hetherington, A. J., Berry, C. M. & Dolan, L. Networks of highly branched stigmatic rootlets developed on the first giant trees. *Proc. Natl. Acad. Sci.* **113**, 6695–6700 (2016).
50. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732. <https://doi.org/10.1038/nrg2600> (2009).
51. Li, J.-T. *et al.* The fate of recently duplicated genes following a fourth-round whole genome duplication in a tetraploid fish, common carp (*Cyprinus carpio*). *Sci. Rep.* **5**, 1–9 (2015).
52. Davis, J. E., Kolozsvary, M. B., Pajeroska-Mukhtar, K. M. & Zhang, B. Toward a universal theoretical framework to understand robustness and resilience: From Cells to systems. *Front. Ecol. Evol.* <https://doi.org/10.3389/fevo.2020.579098> (2021).
53. Boocock, G., Marit, M. & Rommens, J. Phylogeny, sequence conservation, and functional complementation of the SBDS protein family. *Genomics* **87**, 758–771 (2006).

54. Gourcilleau, D. *et al.* Phylogenetic study of plant Q-type C2H2 zinc finger proteins and expression analysis of poplar genes in response to osmotic, cold and mechanical stresses. *DNA Res.* **18**, 77–92 (2011).
55. Emerson, R. O. & Thomas, J. H. Adaptive evolution in zinc finger transcription factors. *PLoS Genet.* **5**, e1000325 (2009).
56. Mackay, J. P. & Crossley, M. Zinc fingers are sticking together. *Trends Biochem. Sci.* **23**, 1–4 (1998).
57. Marchal, C. *et al.* BED-domain-containing immune receptors confer diverse resistance spectra to yellow rust. *Nature Plants* **4**, 662–668 (2018).
58. Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X. & Morel, J. B. Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytol.* **210**, 618–626 (2016).
59. Zuluaga, A. P. *et al.* The rice DNA-binding protein ZBED controls stress regulators and maintains disease resistance after a mild drought. *Front. Plant Sci.* **11**, 1265 (2020).
60. Bundock, P. & Hooykaas, P. An Arabidopsis hAT-like transposase is essential for plant development. *Nature* **436**, 282–284 (2005).
61. Read, A. C. *et al.* Genome assembly and characterization of a complex zfBED-NLR gene-containing disease resistance locus in Carolina Gold Select rice with Nanopore sequencing. *PLoS Genet.* **16**, e1008571 (2020).
62. Li, W., Pang, S., Lu, Z. & Jin, B. Function and mechanism of WRKY transcription factors in abiotic stress responses of plants. *Plants* **9**, 1515 (2020).
63. Chen, X., Li, C., Wang, H. & Guo, Z. WRKY transcription factors: evolution, binding, and action. *Phytopathol. Res.* **1**, 1–15 (2019).
64. Hirsch, S. & Oldroyd, G. E. GRAS-domain transcription factors that regulate plant development. *Plant Signal. Behav.* **4**, 698–700 (2009).
65. Hofmann, N. R. A structure for plant-specific transcription factors: the GRAS domain revealed. *Plant Cell* **28**, 993–994 (2016).
66. Vaattovaara, A. *et al.* Mechanistic insights into the evolution of DUF26-containing proteins in land plants. *Commun. Biol.* **2**, 1–18 (2019).
67. Hussain, A. *et al.* Comparative phylogenetic analysis of aquaporins provides insight into the gene family expansion and evolution in plants and their role in drought tolerant and susceptible chickpea cultivars. *Genomics* **112**, 263–275 (2020).
68. Zhang, S.-D., Ling, L.-Z. & Yi, T.-S. Evolution and divergence of SBP-box genes in land plants. *BMC Genomics* **16**, 1–10 (2015).

### Author contributions

M.S.M. and K.M.P.M. conceived the study and supervised the research. A.H., A.B., Z.N., R.B., A.A., and M.Z. performed computational data analyses. J.L. and B.M. performed qRT-PCR studies on the Zf-BED genes. A.H., M.S.M. and K.M.P.M. wrote the manuscript.

### Funding

This research study was funded by International Foundation for Science (IFS), Sweden, under project C-6501-1 to Athar Hussain. This work was supported by NSF awards IOS-1557796 and IOS-2038872 to MSM, and IOS-1350244 and IOS-2038872 to KMPM.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-16602-8>.

**Correspondence** and requests for materials should be addressed to M.S.M. or K.M.P.-M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022