Comparative and expression analyses of AP2/ERF genes reveal copy number expansion and potential functions of ERF genes in Solanaceae

Jin-Wook Choi, Hyeon Ho Choi, Young-Soo Park, Min-Jeong Jang and Seungill Kim*

Abstract

Background The AP2/ERF gene family is a superfamily of transcription factors that are important in the response of plants to abiotic stress and development. However, comprehensive research of the AP2/ERF genes in the Solanaceae family is lacking.

Results Here, we updated the annotation of AP2/ERF genes in the genomes of eight Solanaceae species, as well as *Arabidopsis thaliana* and *Oryza sativa*. We identified 2,195 AP2/ERF genes, of which 368 (17%) were newly identified. Based on phylogenetic analyses, we observed expansion of the copy number of these genes, especially those belonging to specific Ethylene-Responsive Factor (ERF) subgroups of the Solanaceae. From the results of chromosomal location and synteny analyses, we identified that the AP2/ERF genes of the pepper (*Capsicum annuum*), the tomato (*Solanum lycopersicum*), and the potato (*Solanum tuberosum*) belonging to ERF subgroups form a tandem array and most of them are species-specific without orthologs in other species, which has led to differentiation of AP2/ERF gene repertory among Solanaceae. We suggest that these genes mainly emerged through recent gene duplication after the divergence of these species. Transcriptome analyses showed that the genes have a putative function in the response of the pepper and tomato to abiotic stress, especially those in ERF subgroups.

Conclusions Our findings will provide comprehensive information on AP2/ERF genes and insights into the structural, evolutionary, and functional understanding of the role of these genes in the Solanaceae.

Keywords AP2/ERF, Transcription factors, Re-annotation, Solanaceae, species-specific duplication, Abiotic stress

Background

Plants are exposed to a broad range of biotic and abiotic stresses. They have evolved defense responses to allow them to cope with and adapt to stress [1, 2]. Transcription factors (TFs) are often involved in plant responses to stress and typically bind to specific cis-elements, thereby regulating the expression of the target gene downstream [3, 4]. The APETALA2 (AP2)/ Ethylene-Responsive Factor (ERF) superfamily is one of the largest families of TFs in plants. Members of this family have an essential role in plant responses to diverse stresses as well as in development [5, 6]. The TFs in the AP2/ERF superfamily contain one or more AP2/ERF domains, which have 60–70 conserved amino acids [7]. The family is generally divided into six subfamilies based on domain architectures and specific amino acid residues. These subfamilies are APETELA2 (AP2), AINTEGUMENTA (ANT),

*Correspondence:
Seungill Kim
ksi2204@uos.ac.kr
Department of Environmental Horticulture, University of Seoul, Seoul 02504, Republic of Korea

Choi *et al. BMC Plant Biology*   (2023) 23:48

Page 2 of 15

ERF, Dehydration-Responsive Element-Binding protein (DREB), ABI3/VP1 (RAV), and Soloist [7, 8]. AP2/ERF genes in the AP2 and ANT subfamilies have two conserved AP2 domains, while members with just one AP2 domain are assigned to ERF or DREB subfamilies. Members with the AP2 domain and the B3 binding domain belong to the RAV subfamily [9], and those with a sequence in the AP2 domain that is distinct from the other subfamilies are in Soloist [7]. For genes in which the domain architecture is identical in different subfamilies, the specific amino acid sequences in the AP2 domain are considered for the classification. Specifically, the TFs that belong to the AP2 subfamily contain 10 amino acid insertions in the AP2 domain region, while those in the ANT subfamily do not have these insertions [10]. The ERF and DREB subfamilies have distinct amino acid residues in the 14th and 19th positions. Proteins in the ERF subfamily have alanine (Ala) and aspartic acid (Asp), and proteins in the DREB subfamily have valine (Val) and glutamic acid (Glu) in the 14th and 19th positions, respectively [8].

The genes in the AP2/ERF superfamily have been studied in several plant species, including *Arabidopsis thaliana, Oryza sativa, Zea mays, Glycine max,* and *Populus trichocarpa* [7, 11–13]. These genes have also been identified in the Solanaceae family, which includes many economically important crops, notably the tomato (*Solanum lycopersicum*), the potato (*Solanum tuberosum*), and the pepper (*Capsicum annuum*) [14–16]. In these studies, the AP2/ERF genes were first identified and, next, evolution and expression analyses, based on previously published annotations, were performed. However, many important genes were not included in the published annotations and there has been little or no interactive research with updated annotations of AP2/ERF genes in the Solanaceae family [17–21]. For these reasons, we regard comprehensive comparative and functional analyses as indispensable to understanding the AP2/ERF gene families in the Solanaceae.

We have performed annotation updates on Solanaceae AP2/ERF genes and comparative analyses of these AP2/ERF genes with those from *A. thaliana* and *O. sativa*. We identified 2,195 AP2/ERF genes; of these, 368 (17%) were newly annotated. From the results of structural and phylogenetic analyses, we verified that the AP2/ERF genes were grouped into 12 subgroups according to discrete domain architectures; these groups were: A1-A4, B1-B4, AP2, ANT, RAV, and Soloist. Synteny analysis between pepper, tomato, and potato revealed repertory changes of AP2/ERF genes by copy number expansion in specific ERF subgroups of specific species. From our analysis, we showed that these expanded genes in the three Solanaceae species mainly emerged via recent gene

duplications in individual species after the completion of divergence of these species. Furthermore, the expression profiles and gene ontology (GO) enrichment test in pepper and tomato revealed diverse putative functions in ERF subgroups under various abiotic stress conditions. We believe that our study can be used to illustrate the evolutionary and expressional characteristics of the AP2/ERF genes in the Solanaceae, and we hope the information will serve as a fundamental genomic resource for further functional and breeding studies on Solanaceae crops.

Results and discussion

Re-annotation and characterization of the AP2/ERF gene family in Solanaceae

We performed a re-annotation of AP2/ERF genes in 10 genomes, specifically eight Solanaceae species, *A. thaliana,* and *O. sativa*. This was done to construct advanced AP2/ERF gene models of these genomes. We identified 2,195 AP2/ERF genes containing 368 (17%) newly annotated genes in these genomes. Of them, 48% of genes (176) were newly annotated based on RNA-seq data or protein evidence, suggesting that these genes are annotated based on high-confident evidence, whereas the other genes could be inactive. The numbers of genes per species ranged from 142 in *A. thaliana* to 355 in *Nicotiana benthamiana* (Table 1). *A. thaliana* had the fewest newly annotated genes (1) and *Capsicum baccatum* had the most (74). We investigated the structures of the updated AP2/ERF genes to characterize the general features of AP2/ERF genes in Solanaceae. Following the rule of dividing subfamilies described previously [7], we divided the AP2/ERF genes into six subfamilies according

Table 1 Numbers of re-annotated AP2/ERF genes in the eight Solanaceae plants studied here, and in *Arabidopsis thaliana* and *Oryza sativa*

Species	Previously annotated genes	Newly annotated genes	Total
Oryza sativa	163	14	177
Arabidopsis thaliana	141	1	142
Nicotiana benthamiana	323	32	355
Petunia axillaris	187	32	219
Capsicum annuum	181	67	248
Capsicum baccatum	173	74	247
Solanum tuberosum	212	42	254
Solanum melongena	153	23	176
Solanum pimpinellifolium	127	59	186
Solanum lycopersicum	167	24	191
Total	1827	368	2195

Choi *et al. BMC Plant Biology*   (2023) 23:48

Page 3 of 15

to the number of AP2/ERF domains and bases coding for specific amino acid residues: ERF, DREB, ANT, AP2, RAV, and Soloist (Fig. 1A). Specifically, 1,263 genes were classified into the ERF subfamily, which comprises one AP2/ERF domain and codes for Alanine and Aspartate in the 15th and 20th positions, respectively. The proportion of genes in the subfamilies indicates that most were in the ERF subfamily in the Solanaceae (Fig. 1B). As in *Vitis vinifera* and *G. max,* in which the ERF subfamily is a dominant group, more than half of AP2/ERF genes in the Solanaceae family were classified into the ERF subfamily [13, 22].

We also explored structural diversity within subfamilies and examined the coding sequence for specific amino acids in the AP2/ERF domain. As discussed, subfamilies of AP2/ERF genes code for specific amino acid residues in specific positions, and the proteins encoded for by the AP2/ERF domain in the Solanaceae family also had amino acid differences according to subfamily in positions 15, 20, and 22–29 (Fig. 1C). Specifically, the Soloist subfamily included distinct amino acid sequences of the AP2/ERF domain compared to the other subfamilies [23]. We also investigated the secondary structure encoded by the AP2/ERF domain related to the essential role of binding to the GCC box. Consistent with other findings, the AP2/ERF domain coded for three β-sheets and one α-helix: positions 1–7, β1; 14–20, β2; 31–33, β3; 48–57, α (Fig. 1C) [24]. Based on the region of the four secondary structures (three β-sheets and one α-helix), we divided the AP2/ERF domain into eight divisions and calculated the consensus score of each division to compare the conservation degree among subfamilies. We found overall conservation of most of the regions, including the four secondary structures, with the exception that we observed low conservation of the fourth division (14%) among subfamilies. These results suggest that the structural diversity between subfamilies occurred due to the divergence in specific residues and divisions, which may contribute to the functional diversity of AP2/ERF genes in Solanaceae.

Functional annotation through GO analysis was performed to shed light on the putative functions of the AP2/ERF genes in the Solanaceae family. The 2,166 (99%) GO terms for AP2/ERF genes were classified into three categories based on biological processes, molecular functions, and cellular components (Fig. 1D). The predominant GO descriptions in three categories were "regulation of cellular process" (99%; 2,139 genes), "organic cyclic compound binding" (97%; 2,093 genes), and "intracellular anatomical structure" (73%; 1,576 genes). These results were consistent with previous studies, in which it was concluded that functions of AP2/ERF genes were related to regulating target genes [25]. Taken together,

our analyses display the overall characteristics of updated AP2/ERF genes in the Solanaceae family by investigating repertoires of the subfamily, amino acid composition encoded within the AP2/ERF domain, and functional prediction of AP2/ERF genes.

Phylogenetic relationship and copy number expansion of ERF genes in Solanaceae

We constructed a phylogenetic tree with the updated AP2/ERF genes of the 10 species to explore the evolutionary relationships of AP2/ERF genes in Solanaceae. We divided AP2/ERF genes into 12 subgroups, based on their phylogeny and the previously described rule for dividing subfamilies (Fig. 2A) [7–10]. The updated AP2/ERF genes within the same subfamily were distinctly clustered into subgroups (Fig. 2A). For example, genes from the ERF and DREB subfamilies were clearly clustered into B (B1-B4) and A (A1-A4) subgroups, respectively. We then compared copy numbers of AP2/ERF genes among subgroups from each species (Fig. 2B). Overall, the copy numbers of genes in subgroups among species were similar. However, genes in subgroups B2 and B4 were specifically abundant in Solanaceae species. This suggests that genes belonging to these subgroups from the ERF subfamily in Solanaceae were expanded in a lineage-specific manner.

Different amino acid residues coded for by the AP2/ERF genes within the ERF subfamily have apparently led to divergent DNA-binding specificities [26]. We extended these findings and investigated the amino acid residues of the AP2/ERF domain of genes in B1-4 subgroups to explore structural diversity within the ERF subfamily. The B1-4 subgroups had significant differences in specific positions of amino acids in the protein. Position 49 had phenylalanine, leucine, phenylalanine, and methionine, in B1-4 subgroups, respectively, and position 56 had leucine, threonine, threonine, and leucine, respectively. These results concur with those of Shoji et al. [26], and indicate that these specific amino acids within ERF subgroups B1-4 contributed to structural diversification, resulting in the construction of distinct lineages of ERF genes [7, 26]. Taken together, our findings provide an overview of the phylogenetic relationship and lineage-specific structural features of AP2/ERF genes that will serve as a fundamental genomic resource for future genetic and functional research.

Copy number expansion of ERF genes through recent gene duplication among Solanaceae

We next verified the physical position of the AP2/ERF genes in the pepper, tomato, and potato. Most of the AP2/ERF genes in these species were evenly distributed
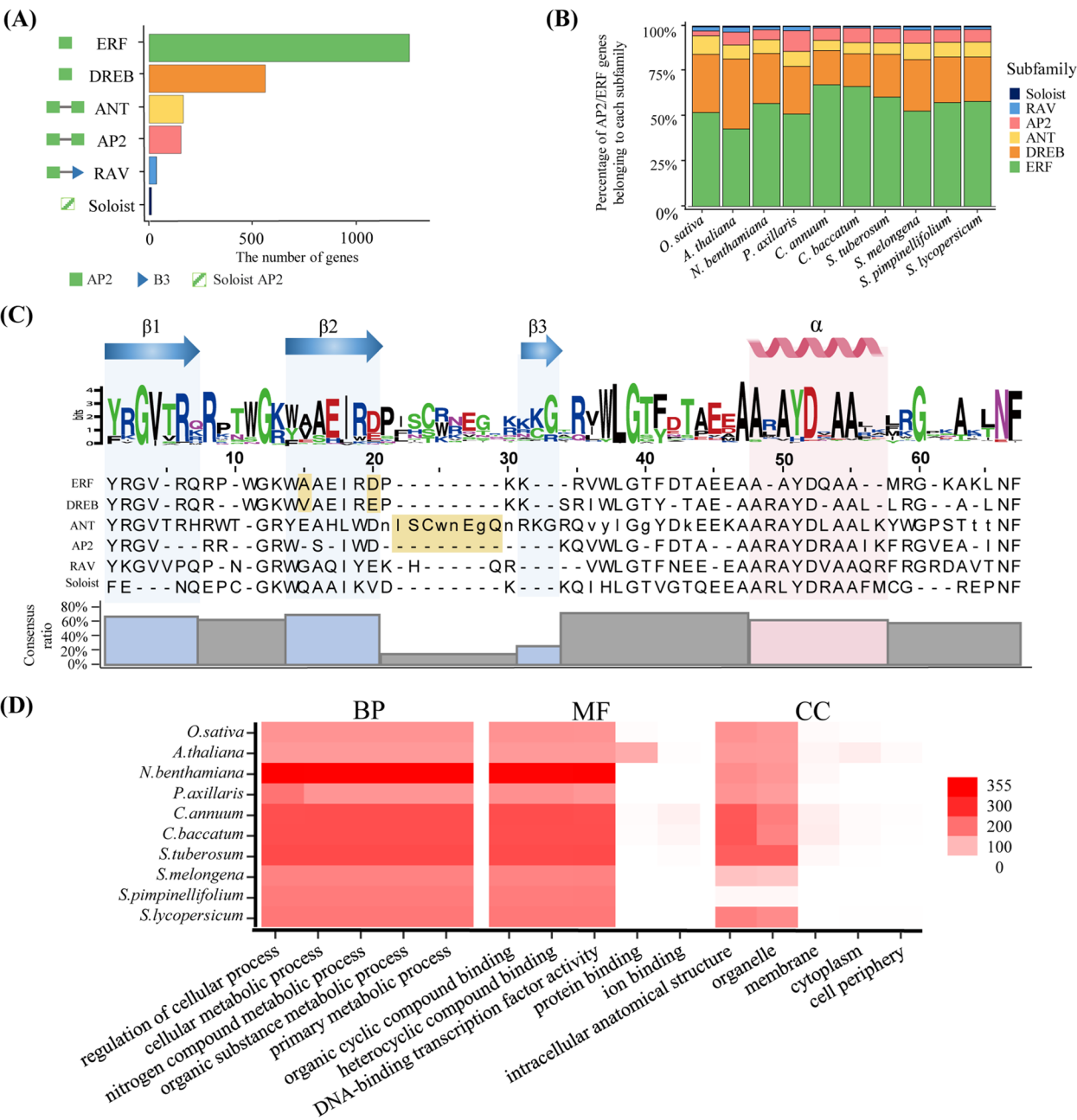
Fig. 1 Characteristics of AP2/ERF genes in Solanaceae. **A** The number of AP2/ERF genes from six subfamilies in 10 plant genomes. The domain repertoires of six subfamilies are represented by symbols positioned on the left side of the bar graph. The color of each bar indicates the subfamily. **B** The proportion of genes from subfamilies in the 10 species. The color of each bar indicates the subfamily. **C** Representation of the amino acid sequence and secondary structures in the AP2 domain from 10 species. The x- and y-axes in the diagram of the amino acid logo indicate the position of amino acids in the AP2 domain and the relative frequency of amino acids at each position, respectively. The location of the secondary structure is depicted on the upper side of the amino acid logo. The alignment of amino acid sequences positioned below the logo shows the most conserved amino acids in each position. For the sequence in uppercase, more than half of the genes contain specific amino acids in that position; if this is not the case, the sequence is lowercase. The specific amino acid residues that are critical in distinguishing the subfamily are highlighted in yellow. The bar graph below the alignment shows the consensus ratio of eight divisions considering the secondary structure of the AP2 domain. The colors of the bar display different divisions: pink bar indicates division in α-helix, blue bars indicate divisions in β-sheets, and grey bars indicate the interval regions. **D** Heat map of the number of the top five gene ontology (GO) terms of AP2/ERF genes. Three GO categories are listed over the heat map. The number of GO terms is indicated by the color of the boxes next to the heatmap
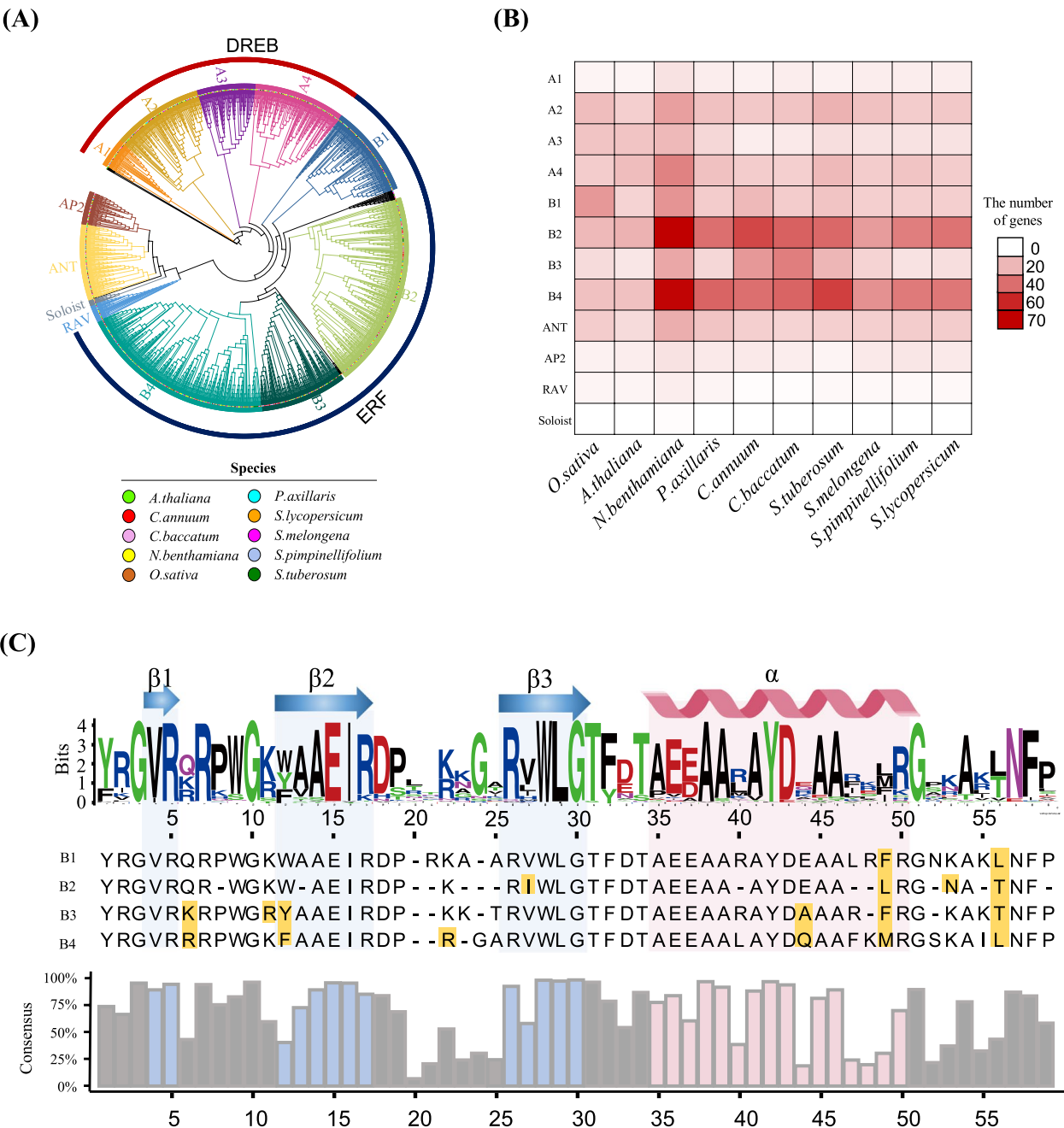
Choi *et al. BMC Plant Biology*   (2023) 23:48

Page 5 of 15

(A)

(B)

(C)

Fig. 2 Phylogenetic relationship and subgroup classification of AP2/ERF genes in 10 species. **A** Phylogenetic tree, in which species are represented by different colors of dots on the end of the branch. The color of each branch and the strap outside of colored dots display different subgroups. The outer ring identifies ERF subgroups (B1-B4) and DREB subgroups (A1-A4). **B** Heat map showing the number of AP2/ERF genes in subgroups from each species. **C** The conserved amino acid sequence of AP2 domain in ERF subgroups. The stacked logo depicts multiple amino acids in a specific position. The height of the logo indicates the frequency of amino acids. The letter under the logo indicates the most conserved amino acid in each position. The different conserved residues between ERF subgroups are highlighted in yellow. The position of α-helices and β-sheets are displayed on the upper side of the logos. The calculated consensus ratio of each position is shown as the bar with the color of the secondary structure of the AP2 domain

throughout the ends of chromosomes with a few exceptions (Fig. S1). These exceptions were: the genes belonging to the ERF subgroups, abundant in Solanaceae, formed a tandem array on specific chromosomes; genes in the B2 subgroup were clustered on the short arm of chromosome 1 in all three species; the genes in the B3

Choi *et al. BMC Plant Biology* (2023) 23:48

Page 6 of 15

subgroup formed a tandem array on the short arm of chromosome 4 in pepper.

We further explored possible mechanisms for the evolution of the AP2/ERF genes in Solanaceae by examining the micro-syntenic region for the subgroups of these species (Fig. 3A and 3B). The copy numbers of genes belonging to the B2 and B3 subgroups were quite variable in the syntenic region of the three genomes, mainly due to the abundantly clustered genes in the B2 and B3 subgroups of pepper. Most of the genes were not in orthologous relationships within the three genomes. These suggest that the AP2/ERF genes in pepper that are in the B2 and B3 subgroups expanded in chromosomes 1 and 4.

When we performed synteny analyses for all AP2/ERF genes, 70%, 33%, and 50% of AP2/ERF genes were not in an orthologous relationship with genes of the pepper, tomato, and potato, respectively. This indicates that species-specific genes in the Solanaceae may have generated differences in the AP2/ERF gene repertories (Fig. S2 and Fig. 3C) [27]. Of these species-specific genes, 69% of the pepper genes, 68% of the tomato genes, and 70% of the potato genes were in ERF subgroups. In particular, 33 (65%) of the pepper genes in subgroup B2 and 24 (83%) of the genes in subgroup B3 had no ortholog in other species. These imply that many genes in specific ERF subgroups may have undergone species-specific gene duplication in specific chromosomes.

To verify whether the AP2/ERF genes without orthologous genes mainly emerged after the divergence of Solanaceae species, we estimated the duplication period of AP2/ERF genes in the pepper, tomato, and potato genomes through calculation of the synonymous substitution rates between duplicated AP2/ERF gene pairs (Fig. 3D). We observed 62 AP2/ERF genes in the pepper with Ks values < 0.3; 84% of these genes had no orthologous relationship with either the tomato or potato. Likewise, eight genes in the tomato and 34 genes in the potato had Ks values < 0.1, and 25% and 65% of these genes, respectively, had no orthologous relationship with the other two species. Given the divergence time between the pepper and *Solanum* species (0.3 Ks) and between the tomato and potato (0.1 Ks) [28], it may be that recent species-specific duplication of AP2/ERF genes occurred mainly in the pepper and potato. We investigated the Ks distribution of AP2/ERF genes for each subgroup in the three species (Fig. 3E). We observed that 58 (93%) of pepper AP2/ERF genes with Ks values < 0.3 belonged to ERF subgroups (B1-4). Most of these genes (84%) did not have an orthologous relationship with the other two species. In tomato and potato, respectively, 75% and 82% of AP2/ERF genes with Ks values between 0 and 0.1 belonged to ERF subgroups. Most of these genes in the potato (64%) did not have an orthologous relationship with the other

two species. These suggest that the repertory of AP2/ERF genes among the three genomes differentiated mainly due to the emergence of species-specific AP2/ERF genes, especially those in ERF subgroups. This occurred through recent gene duplications after the divergence of the three Solanaceae species.

Expression and functional investigation of AP2/ERF genes in pepper

AP2/ERF genes regulate plant response to abiotic stress [29]. We conducted expression analyses for all pepper genes, including the newly identified AP2/ERF genes, to learn more about their roles in abiotic stress. The expression profiles of pepper AP2/ERF genes were acquired from RNA-seq data in a time course (data were taken at 3, 6, 12, 24, and 72 h after the imposition of stress) to four stresses (cold, heat, mannitol, or salt). By comparing the data from plants under stress with the unstressed control, we identified the following numbers of differentially expressed genes (DEGs): 10,720 under cold stress, 9,992 under heat stress, 3,542 under mannitol stress, and 3,194 under salt stress. Of these, 70, 48, 31, and 46 were AP2/ERF genes that were differentially expressed during cold, heat, mannitol, and salt stress, respectively (Fig. S3 and Table S3). Most of the AP2/ERF DEGs (56%) were in the ERF subgroups: 33 were in B1, 30 were in B2, 14 were in B3, and 34 were in B4. This result suggests that many AP2/ERF genes in ERF subgroups B1-4 participate in the stress response (Fig. S3) [14, 30–32]. In agreement with our suggestion, it has been reported that *CaPF1* (caAP2_140) was up-regulated under cold stress [33].

We then performed an expression clustering analysis to investigate expression patterns of pepper DEGs including both AP2/ERF and others under abiotic stress conditions. The DEGs were divided into four different clusters (C1-C4) for each stress condition (Fig. 4A). We detected abundant AP2/ERF DEGs in C1 and C4 under cold stress, in C4 under heat stress, in C2 under mannitol stress, and in C1 and C3 under salt stress (Fig. 4B). This represents that the AP2/ERF genes and other pepper genes in these clusters may interact with each other for specific functions under these conditions.

We next conducted GO-enrichment analysis for the genes in the six of 16 clusters mentioned above, which contained abundant AP2/ERF DEGs (Fig. 4C). This was done to gain some insight into the functions of AP2/ERF genes in pepper. We surveyed enriched GO terms in the clusters. We detected many GO terms in the clusters related to regulation, and it may be that these genes have regulatory roles during abiotic stress. For example, GO terms such as "regulation of
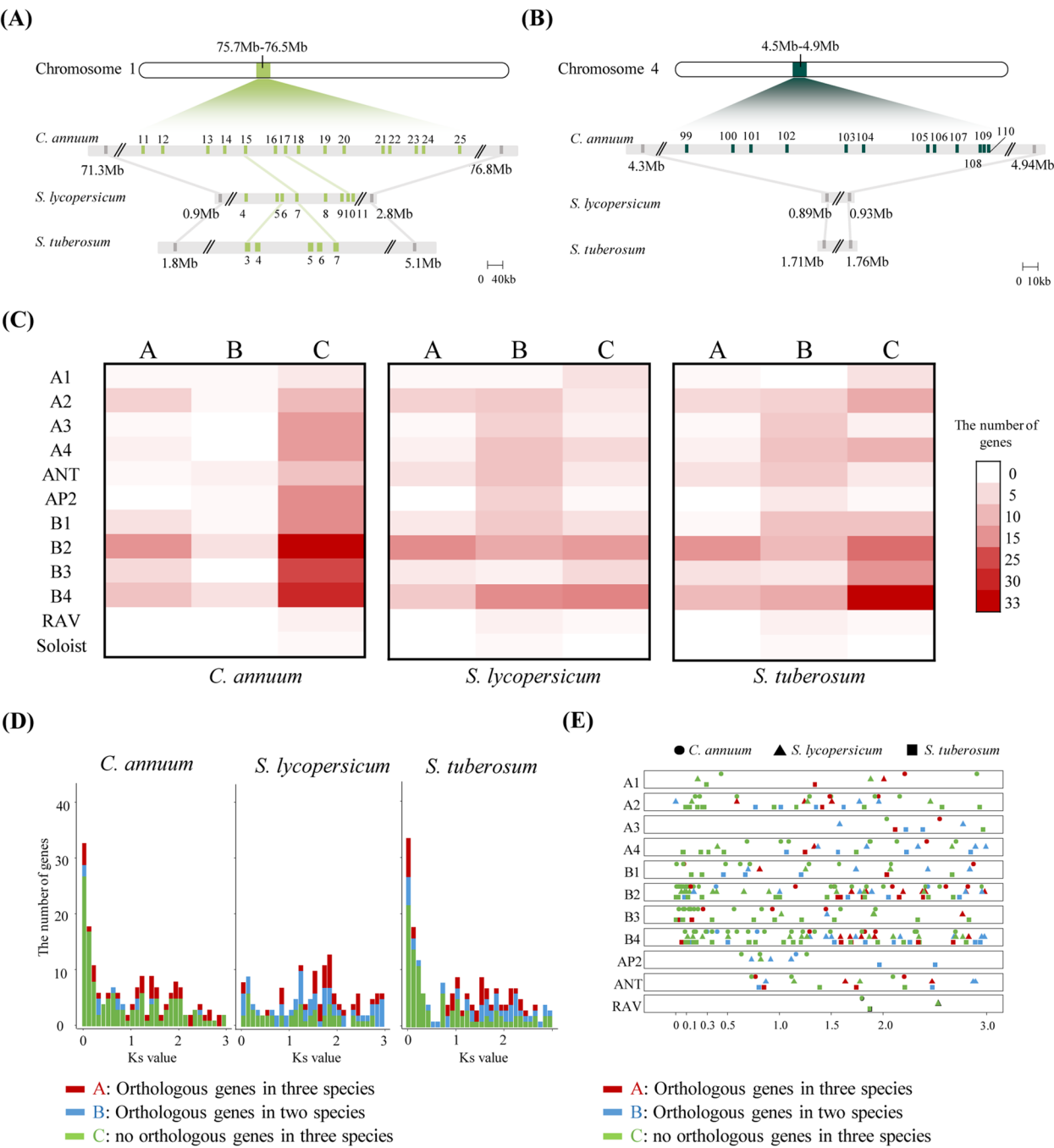
Fig. 3 Comparison of the syntenic region and duplication history of the AP2/ERF pepper, tomato, and potato genes. **A**-**B** Micro-synteny analysis of AP2/ERF genes in the subgroup B2 **A** and B3 **B** on chromosomes 1 and 4. The gene IDs in the syntenic region are represented and the orthologous gene is linked with the line. The genes colored with grey indicate orthologous genes that are not AP2/ERF genes. **C** Heat map of the number of genes with orthologous relationships. The number of genes having orthologs in three, and two species are displayed on A and B, respectively. The number of genes with no orthologs is represented on the **C**. **D** Histogram showing the duplication history of AP2/ERF genes. The color of the bar indicates an ortholog relationship between the three species. **E** The distribution of duplication history of AP2/ERF. The shapes and colors indicate species and ortholog relationships between the three species, respectively

Choi *et al. BMC Plant Biology*   (2023) 23:48

Page 8 of 15

cellular process (GO:0050794)", "regulation of biological process (GO:0050789)", and "biological regulation (GO:0065007)" were enriched in cold cluster 4 and in salt cluster 1, and many of them were included in the ERF subgroups (B1-4). This is consistent with previous studies that AP2/ERF genes in the ERF subgroup are important in the regulation of plant resistance to abiotic stress [30, 33, 34]. Specifically, it is known that *CaPF1* of pepper in the ERF subgroup containing "biological regulation" (GO:0009873) promoted cold tolerance by binding to both sequences of GCC and DRE/CRT (dehydration-responsive element/C-repeat) in PR and COR genes [33]. At the time this work was done because only a few functional AP2/ERF genes had been reported, the pepper genes that had an orthologous relationship with functional tomato genes were analyzed. The purpose was to better understand the function of genes in the ERF subgroups in pepper under salt stress. The CA.PGAv.1.6.scaffold790.71, an orthologous tomato gene of *TSRF1*, known for improving salt resistance, had the child GO description of biological regulation (GO:0009873) [35]. Thus, it seems likely that the pepper AP2/ERF genes in cold cluster 4 and salt cluster 1 are involved in regulatory mechanisms associated with abiotic stress. Taken together, our analyses provide overall expressional and functional features of pepper AP2/ERF genes, that will build a broad understanding of these genes and enhance future studies.

Transcriptome analyses of AP2/ERF tomato genes

We next examined the expression profile of tomato genes under cold, drought, heat, and salt stress. We examined the whole genes, including the newly-annotated AP2/ERF genes, using the RNA-seq data. The following DEGs were identified in different stress conditions: 8,975 under cold stress, 2,698 under drought stress, 4,643 under heat stress, and 2,651 under salt stress. Of these DEGs, 60, 19, 23, and 11 AP2/ERF were identified in tomato under cold, drought, heat, and salt stress, respectively (Fig. 5A and Fig. S4). We detected that many AP2/ERF DEGs were included in ERF subgroups as pepper DEGs (B1:4, B2:16, B3:6, and B4:37). To compare the expression level of AP2/ERF with other genes in the tomato under abiotic stress, we classified DEGs into two groups: up-regulated and down-regulated (Fig. 5B). We verified up-regulation of the expression of *SlDREB2* (slAP2_161) under salt stress, as has been reported previously [36]. This also validated our analyses of the expression of tomato AP2/ERF genes (Fig. S4).

We performed a GO-enrichment test for all genes including AP2/ERF DEGs in the eight groups (Up- and down-regulated groups in four abiotic stresses) (Fig. 5C).

Diverse functional descriptions appeared in specific groups. For example, metabolism-related GO terms such as "metabolic process (GO:0008152)", "cellular metabolic process (GO:0044237)", and "nitrogen compound metabolic process (GO:0006087)" were present in six groups (down- and up-regulated under cold stress, down- and up-regulated under drought, down-regulated under heat stress, down-regulated under salt stress). Binding-related GO terms such as "binding" (GO:0005488), "ion binding" (GO:0043167), "metal ion binding" (GO:0046872), and "cation binding" (GO:0043169) were enriched in six groups (down-and up-regulated genes under cold stress, down- and up-regulated genes under drought, down-regulated genes under heat stress, down-regulated genes under salt stress). This suggests that tomato genes, including AP2/ERF genes, are involved in diverse functions under abiotic stresses. The response-related GO term "response to stimulus" (GO:0050896) was enriched in seven groups (down-and up-regulated genes under cold stress, down- and up-regulated genes under drought, down-regulated genes under heat stress, down-and up-regulated genes under salt stress). This result is consistent with previous studies that AP2/ERF genes are involved in stress response mechanisms. Specifically, AP2/ERF genes in ERF subgroups (B1-4) are key regulators for several stresses, such as jasmonate (JA), ethylene, salt, drought, and cold treatments [37]. *JERF3* in the tomato, which belongs to ERF subgroups, had the child term of response to stimulus (GO:0009873) and binds to the ethylene/JA responsive GCC box, thereby enhancing salt tolerance [37]. These results suggest that tomato AP2/ERF genes belong to ERF subgroups (B1-4) and associate with stress response genes.

Conclusions

In the past, annotations have been broadly performed. However, researchers have highlighted the need for analyses of gene families with updated annotation and suggested that such analyses will provide new insight into the evolutionary and functional characteristics of genes [18–21]. In this study, we performed a re-annotation of AP2/ERF genes in eight Solanaceae genomes as well as the genomes of *A. thaliana* and *O. sativa*, and enable to conduct comparative, evolutionary, and functional analyses of the AP2/ERF genes, providing insights into comprehensive structural and functional characteristics of Solanaceae AP2/ERF genes. In particular, our data reveals that there has been species-specific copy number expansion of AP2/ERF genes in ERF subgroups B2-4. We found that genes in ERF subgroups form a tandem array in pepper, tomato, and potato chromosomes. Most of these genes do not have orthologs among the three species: most of the genes without orthologs in ERF subgroups were specific
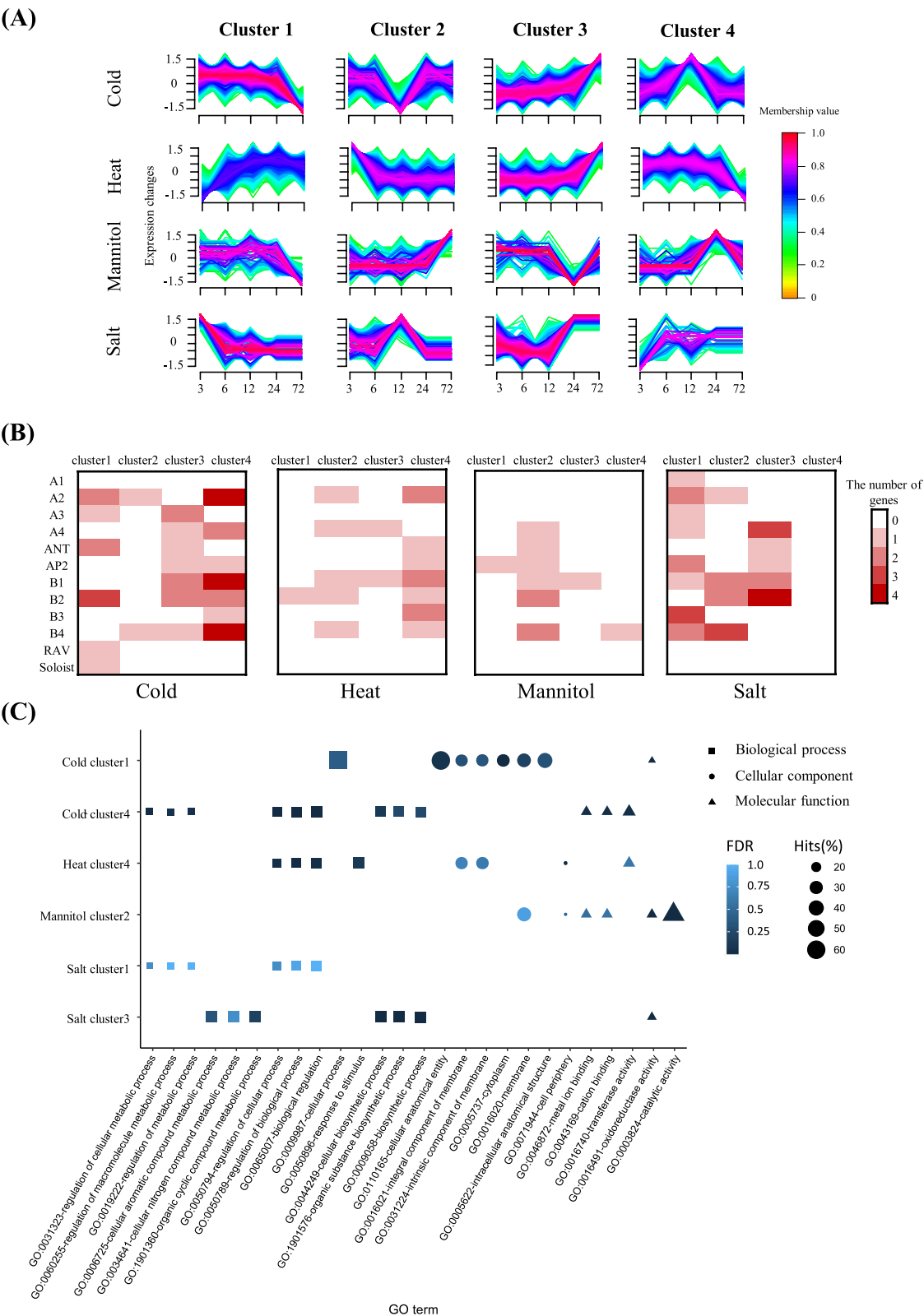
Fig. 4 Transcriptome analyses with gene ontology (GO) enrichment test in differentially expressed genes (DEGs) of pepper under abiotic stress. **A** For each stress, expression patterns of whole DEGs, including AP2/ERFs are represented in four clusters. **B** Heat map showing the distribution of AP2/ERF genes in each subgroup. **C** Dot plot showing the top five GO descriptions abundance in specific clusters. The shape and size of the symbols illustrate the type and frequency of the GO description, respectively
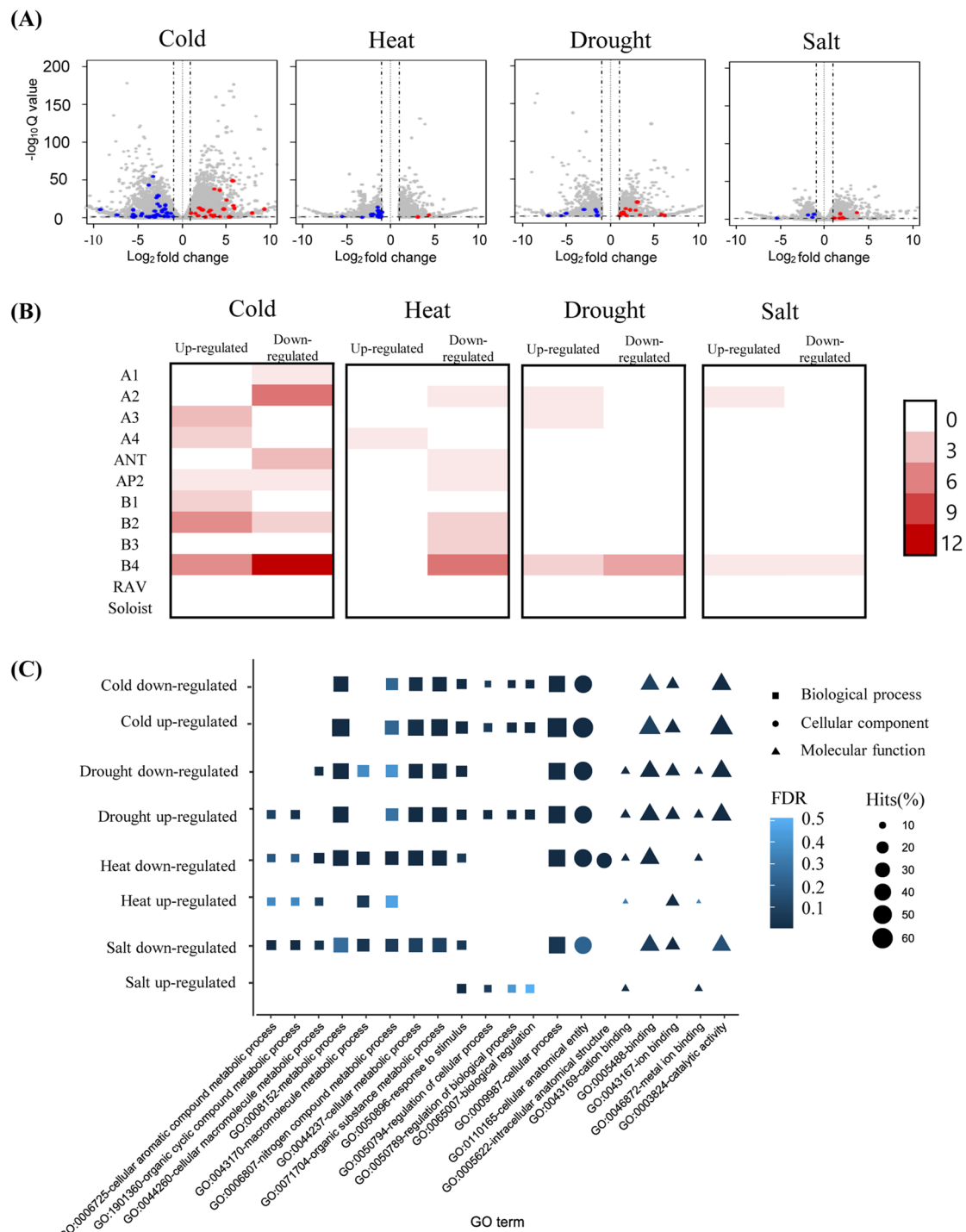
Fig. 5 Expression and putative functions of AP2/ERF genes in tomato exposed to abiotic stress. **A** Differentially expressed genes under four abiotic stresses are shown as a volcano plot. The color of scattered points represents the expression difference of AP2/ERF genes (grey: whole genes; red: up-regulated AP2/ERF DEGs; blue: down-regulated AP2/ERF DEGs). **B** Heatmap showing the number of AP2/ERF DEGs in each subgroup. The scale bar on the right side of the heatmap indicates the heatmap value. **C** Depiction of results of the GO-enrichment test. The top five enriched GO terms in each group are listed under the plot. The different shapes and sizes of symbols indicate different categories and the frequency of GO descriptions (DEGs; differentially expressed genes; GO, gene ontology)

to their species and emerged through recent gene duplication after the divergence of the species. This could explain the different AP2/ERF gene repertoires in the Solanaceae family. Based on transcriptome analyses and investigation of functional genes and their orthologs, our findings suggest that genes in ERF subgroups are involved in response mechanisms and interact in the response to abiotic stress. Furthermore, we investigated the similarity of expression patterns between orthologous genes in pepper and tomato (Fig. S5). We found that the expression of orthologous genes had a low correlation under cold, heat, and salt stresses, probably due to different experimental conditions in RNA-seq data, such as differences in the developmental level of plant materials or growth conditions. Consequently, our study provides a comprehensive understanding of structural characteristics, evolutionary history, and potential functions of AP2/ERF genes in the Solanaceae family.

Materials and methods

Identification and re-annotation of AP2/ERF genes in 10 species

We downloaded publicly available genomic resources and RNA-seq data of *A. thaliana* [38], *O. sativa* [39], and eight other Solanaceae crops: *N. benthamiana* [40], *Petunia axillaris* [41], *C. annuum* [42], *C. baccatum* [28], *S. tuberosum* [43], *S. melongena* [44], *S. pimpinellifolium* [45], and *S. lycopersicum* [46] (Table S1). For re-annotation of AP2 genes, we operated TGFam-Finder v1.20 [47] with "260,000" and "110,000" for "EXTENSION_LENGTH" and "MAX_INTRON_LENGTH" considering maximum gene and intron lengths of AP2/ERF genes in published annotations, respectively. The tab-separated value (TSV) files of AP2/ERF genes were generated from InterproScan 5 (-f tsv, -appl Pfam) [48], which were used for "TSV_FOR_DOMAIN_IDENTIFICATION". The AP2 domain, which was indicated as "PF00847" from the Pfam database was utilized for "TARGET_DOMAIN_ID".

All re-annotated genes were given new names instead of using locus tag names. In this, we considered species names and the physical position in chromosomes such as "atAP2_1" (Table S2). In cases of the AP2/ERF genes in *A. thaliana*, *O. sativa*, *S. lycopersicum*, and *S. tuberosum*, we matched previously assigned gene names to new names.

Structure of AP2/ERF genes

The tab-separated value (TSV) files generated by InterproScan 5 (-f tsv, -appl Pfam) [48] according to the Pfam database were used to identify the domain structures of the updated AP2/ERF genes. The domain region of AP2/ERF genes was aligned by the MAFFT program (with default

parameter) [49]. We trimmed inconclusive sequences using trimAL v1.4 (gt: 0.05) [50]. To obtain more accurate information for further analyses, we excluded domains with high e-values (>1e-5) or that overlapped with the AP2/ERF domain.

Amino acid sequence composition within AP2/ERF domain

We wished to represent graphically the amino acid composition of all AP2/ERF domains in 10 species with a multiple sequence alignment. For this, we utilized a WebLogo program [51]. We used EMBOSS Cons programs (Plurality 0.1, default option) [52] to estimate the consensus sequence of the AP2/ERF domain according to subfamily. The default option, Jpred, [53] predicted the secondary protein structure of the AP2/ERF domain by multiple alignments (JNETPSSM model). From the protein structure, we divided the AP2/ERF domain into eight divisions. We calculated the average consensus scores of each division by multiple alignments of the complete sequences information management system (MACSIMS) in Jalview programs [54].

Gene Ontology (GO) analysis

We used the OmicsBox tool (v 1.4) [55] to conduct a functional annotation of APE/ERF genes in Solanaceae. With an e-value cut off of 1e-3 as the default setting in the OmicsBox tool, BLASTP [56] was used to match APE/ERF protein sequences to the National Center for Biological Information (NCBI) non-redundant proteins database (nr v5). We combined the result of InterProScan [48] with BLAST results for prediction. We used default parameters of Blast2GO Mapping and Blast2GO Annotation [57] to match GO terms and subdivided the results of GO analysis into three categories: biological process, molecular function, and cellular component. We visualized the top five GO terms in level 3 of each category with a heatmap.

Classification and phylogenetic analysis of AP2/ERF genes

The 1937 AP2/ERF genes with intact AP2 domain(s) in 10 species were aligned by MAFFT v7.470 [49] and the alignments were trimmed by trimAL v1.4 (-gt: 0.5) [50]. We constructed a maximum likelihood tree with 500 rapid bootstraps in random parsimony (-m PROTGAMMAJTT -p 12345-× 12345 -#500) via RAxML v8.2.12 [58], which predicted the PROTGAMMAJTT model as the best model (-m PROTGAMMAAUTO -p 12345). The constructed mid-point rooted tree was represented by the Interactive Tree of Life (iToL v5). The AP2/ERF genes were clustered into 12 subgroups. These subgroups were distinguished by the domain architecture and specific residues (A1-A4, B1-B4, AP2, ANT, RAV, and Soloist).

Chromosomal location and synteny of the AP2/ERF genes

The chromosome distribution of newly annotated genes was obtained from GFF files generated by TGFam-Finder v1.20 [47]. The MapChart [59] was used to visualize the diagram illustrating the chromosomal location of AP2/ERF genes except for unanchored scaffolds. We represented the subgroups of all genes in the phylogenetic tree with different colors.

Syntenic analyses were conducted on pepper (*C. annuum*), tomato (*S. lycopersicum*), and potato (*S. tuberosum*) AP2/ERF genes. We used BLASTP [56] for an all-by-all comparison to detect putative orthologous gene pairs. To find orthologous syntenic chains including gene location information that was generated by TGFam-Finder v1.20 [47], we utilized MCScanX programs [60]. The genomic positions of each gene pair were visualized with RIdeogram packages in R software [61].

Duplication and synonymous substitution rates (Ks) value analysis

We utilized the DupGen_Finder pipeline [62] to identify duplicated AP2/ERF gene pairs as described in previous studies. The coding sequences of gene pairs were probabilistic multiple aligned with PRANK (-codon) [63]. We calculated the gene pairs of the synonymous substitution rates (Ks) with the KaKs_calculator v2.0 (-m MYN) [64]. We visualized the gene pairs (Ks value < 3) with ggplot2 [65] in the R package.

Expression analyses of pepper and tomato AP2/ERF genes

We acquired RNA sequencing (RNA-seq) data from the leaves of pepper [66] and tomato (SRR7652567, SRR7652566, SRR7652565, SRR7652564, SRR7652571, SRR7652570, SRR7652569, SRR7652568, SRR7652563, SRR15410554, SRR15410555, SRR15410556, SRR15410551, SRR15410552, SRR15410553, SRR15607561, SRR15607560, SRR15607558, SRR15607557, SRR15607556, and SRR15607555) under diverse abiotic conditions to evaluate the expression patterns of the AP2/ERF genes in these species. The RNA-seq data from the pepper was acquired under cold, heat, salt, and mannitol stress at various times (3, 6, 12, 24, and 72 h) after the imposition of stress. Tomato RNA-seq data was collected from cold, heat, drought, and salt treatments without specific time points. Three biological replicates were performed.

The raw FASTQ files by CLC Assembly Cell (CLC Bio, Aarhus, Denmark) were trimmed to remove low-quality RNA-seq results. The files generated were mapped to each reference genome of *C. annuum* and *S. lycopersicum* by HISAT2 (-dta -x) [67]. The fragment per kilobase of transcript per million mapped reads (FPKM) values of whole genes was calculated by the StringTie (-e, -B -G) [68], including the newly annotated pepper and tomato AP2/ERF. Python scripts (prepDE.py) were used to convert FPKM values to read counts. We used DESeq2 in R software [69] to identify DEGs with the following criteria: $\log_2$ FoldChange > 1 or < -1, and adjusted p-value < 0.05.

Clustering analysis was performed on all pepper genes by Mfuzz in the R package [70] to inspect AP2/ERF expression arrangement. Based on the k-means algorithm, four clusters with up-regulated genes and down-regulated genes were identified. We also arranged all DEGs in the tomato into two groups: up- or down-regulated groups, in each stress treatment. We performed GO annotation and Fisher's exact test (false discovery rate [FDR] p-value ≤ 0.01) for each cluster or group with OmicsBox v1.4 [55] to test GO-enrichment.

As we considered the number of functional genes in the pepper to be deficient, we surveyed orthologs of functional genes. We used Exonerate v2.2.0 (-model protein2genome, -showtargetgff yes, -showquerygff yes, -querytype protein, -targettype DNA) to detect the candidate regions of the orthologs [71].

Investigation of similarity of expression pattern

We examined FPKM values of orthologous genes between pepper and tomato under cold, heat, and salt stress (Fig. S5). As we can not check the time points of treatment in tomato RNA-seq data, we calculated the average FPKM values between five time series of pepper FPKM values. We used pearson method to test correlation between pepper and tomato expression patterns.

Abbreviations

AP2	APETALA2
ERF	Ethylene-Responsive Factor
ANT	AINTEGUMENTA
DREB	Dehydration-Responsive Element-Binding protein
GO	Gene ontology
DEG	Differentially expressed gene
FPKM	Fragment Per Kilobase of transcript per Million mapped reads
FDR	False discovery rate

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-022-04017-6.

Additional file 1: Table S1. Genomic resources of 10 plant species. **Table S2.** Data for re-annotated AP2/ERF genes. **Table S3.** $\log_2$ (Fold change) values of Capsicum annuum in stress condition. **Table S4.** $\log_2$ (Fold change) values of Solanum lycopersicum in stress condition.

Additional file 2: Fig. S1. Chromosome allocation of pepper, tomato and potato AP2/ERF genes. White rectangular boxes indicate chromosomes. **Fig. S2.** Syntenic analyses of all pepper, tomato and potato chromosomes. Colored rectangles represent the chromosomes (chr) of each species. The genes with orthologous relationships among the three species are linked with lines. The line color indicates details of the orthologous relationships, as indicated. **Fig. S3.** Expression profiles of pepper AP2/ERF genes under various abiotic stresses. **A** Heat map showing the relative differences of

expression of AP2/ERF genes under abiotic stress. The heat map values (log$_2$ fold-change) are calculated by FPKM (abiotic stress) /FPKM (control). The names of the AP2/ERF genes are displayed next to the heat map. The normalized values of the heat map are represented in the scale bars, positioned on the right side of the heat map; green implies a low level of expression, whereas red represents a high level of expression. **B** The distribution of AP2/ERF DEGs in the subgroups is illustrated in the heat map (inset). FPKM, fragments per kilo base of exon per million mapped fragments. **Fig. S4.** The expression value of tomato AP2/ERF genes under various abiotic stresses. **A** The heat map represents the value of expression of AP2/ERF genes under abiotic stress. The values are calculated by log$_2$ fold-change of tomato AP2/ERF genes and fold-change values are calculated by dividing FPKM from each stress to control (C: Cold, D: Drought, H: Heat, S: Salt). The colored bar positioned on the right represents the following range of expression levels:-3 (green) to +3 (red). **B** The number of tomato AP2/ERF DEGs in are displayed in heat map (inset). FPKM, fragments per kilo base of exon per million mapped fragments. **Fig. S5.** The expression patterns of orthologous AP2/ERF genes between pepper and tomato under cold, heat, and salt stresses. **A** The heatmap shows expression profiles of orthologous genes. The scale bar on the bottom represents FPKM values. **B** Pearson's correlation analysis between orthologous genes in pepper and tomato.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Verma V, Ravindran P, Kumar PP. Plant hormone-mediated regulation of stress responses. Bmc Plant Biol. 2016;16:86.
2. Ahanger MA, Akram NA, Ashraf M, Alyemeni MN, Wijaya L, Ahmad P. Plant responses to environmental stresses-from gene to biotechnology. Aob Plants. 2017;9:plx025.
3. Schmitz RJ, Grotewold E, Stam M. Cis-regulatory sequences in plants: their importance, discovery, and future challenges. Plant Cell. 2022;34(2):718–41.
4. Bilas R, Szafran K, Hnatuszko-Konka K, Kononowicz AK. Cis-regulatory elements used to control gene expression in plants. Plant Cell Tiss Org. 2016;127(2):269–87.
5. Xie ZL, Nolan ROR, Jiang H, Tang BY, Zhang MC, Li ZH, Yin YH. The AP2/ERF Transcription factor TINY modulates brassinosteroid-regulated plant growth and drought responses in Arabidopsis. Plant Cell. 2019;31(8):1788–806.
6. Licausi F, Ohme-Takagi M, Perata P. APETALA 2/Ethylene Responsive Factor (AP 2/ERF) transcription factors: Mediators of stress responses and developmental programs. New Phytol. 2013;199(3):639–49.
7. Nakano T, Suzuki K, Fujimura T, Shinshi H. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiol. 2006;140(2):411–32.
8. Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K. DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration-and cold-inducible gene expression. Biochem Bioph Res Co. 2002;290(3):998–1009.
9. Romanel EAC, Schrago CG, Counago RM, Russo CAM, Alves-Ferreira M. Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. PLoS ONE. 2009;4(6):e5791.
10. Elliott RC, Betzner AS, Huttner E, Oakes MP, Tucker W, Gerentes D, Perez P, Smyth DR. AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. Plant Cell. 1996;8(2):155–68.
11. Zhou ML, Tang YX, Wu YM. Genome-wide analysis of AP2/ERF transcription factor family in Zea Mays. Curr Bioinform. 2012;7(3):324–32.
12. Zhuang J, Cai B, Peng RH, Zhu B, Jin XF, Xue Y, Gao F, Fu XY, Tian YS, Zhao W, et al. Genome-wide analysis of the AP2/ERF gene family in Populus trichocarpa. Biochem Bioph Res Co. 2008;371(3):468–74.
13. Zhang GY, Chen M, Chen XP, Xu ZS, Guan S, Li LC, Li AL, Guo JM, Mao L, Ma YZ. Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (Glycine max L.). J Exp Bot. 2008;59(15):4095–107.
14. Hsieh TH, Lee JT, Yang PT, Chiu LH, Charng YY, Wang YC, Chan MT. Heterology expression of the Arabidopsis C-repeat/dehydration response element binding factor 1 gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato. Plant Physiol. 2002;129(3):1086–94.
15. Charfeddine M, Saidi MN, Charfeddine S, Hammami A, Bouzid RG. Genome-wide analysis and expression profiling of the ERF transcription factor family in potato (Solanum tuberosum L.). Mol Biotechnol. 2015;57(4):348–58.
16. Jin JH, Wang M, Zhang HX, Khan A, Wei AM, Luo DX, Gong ZH. Genome-wide identification of the AP2/ERF transcription factor family in pepper (Capsicum annuum L.). Genome. 2018;61(9):663–74.
17. van den Berg BHJ, McCarthy FM, Lamont SJ, Burgess SC. Re-annotation is an essential step in systems biology modeling of functional genomics data. PLoS ONE. 2010;5(5):e10642.
18. Bayer PE, Edwards D, Batley J. Bias in resistance gene prediction due to repeat masking. Nat Plants. 2018;4(10):762–5.
19. Guk JY, Jang MJ, Kim S. Identification of novel PHD-finger genes in pepper by genomic re-annotation and comparative analyses. Bmc Plant Biol. 2022;22(1):206.
20. Choi JW, Kim H, Kim S. Two different domain architectures generate structural and functional diversity among bZIP genes in the Solanaceae family. Front Plant Sci. 2022;13:967546.
21. Chae GY, Hong WJ, Jang MJ, Jung KH, Kim S. Recurrent mutations promote widespread structural and functional divergence of MULE-derived genes in plants. Nucleic Acids Res. 2021;49(20):11765–77.
22. Licausi F, Giorgi FM, Zenoni S, Osti F, Pezzotti M, Perata P. Genomic and transcriptomic analysis of the AP2/ERF superfamily in Vitis vinifera. BMC Genomics. 2010;11:719.
23. Giri MK, Swain S, Gautam JK, Singh S, Singh N, Bhattacharjee L, Nandi AK. The Arabidopsis thaliana At4g13040 gene, a unique member of the AP2/EREBP family, is a positive regulator for salicylic acid accumulation and basal defense against bacterial pathogens. J Plant Physiol. 2014;171(10):860–7.

24. Sun XL, Malhis N, Zhao B, Xue B, Gsponer J, Rikkerink EHA. Computational disorder analysis in ethylene response factors uncovers binding motifs critical to their diverse functions. Int J Mol Sci. 2020;21(1):74.

25. Hsieh EJ, Cheng MC, Lin TP. Functional characterization of an abiotic stress-inducible transcription factor AtERF53 in Arabidopsis thaliana. Plant Mol Biol. 2013;82(3):223–37.

26. Shoji T, Mishima M, Hashimoto T. Divergent DNA-binding specificities of a group of ethylene response factor transcription factors involved in plant defense. Plant Physiol. 2013;162(2):977–90.

27. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. J Genet. 2013;92(1):155–61.

28. Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS, et al. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biol. 2017;18:21.

29. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AP2/ERF family transcription factors in plant abiotic stress responses. Bba-Gene Regul Mech. 2012;1819(2):86–96.

30. Sharma MK, Kumar R, Solanke AU, Sharma R, Tyagi AK, Sharma AK. Identification, phylogeny, and transcript profiling of ERF family genes during development and abiotic stress treatments in tomato. Mol Genet Genomics. 2010;284(6):455–75.

31. Gilmour SJ, Sebolt AM, Salazar MP, Everard JD, Thomashow MF. Overexpression of the Arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. Plant Physiol. 2000;124(4):1854–65.

32. Ohmetakagi M, Shinshi H. Structure and expression of a tobacco beta-1,3-glucanase gene. Plant Mol Biol. 1990;15(6):941–6.

33. Yi SY, Kim JH, Joung YH, Lee S, Kim WT, Yu SH, Choi D. The pepper transcription factor CaPF1 confers pathogen and freezing tolerance in Arabidopsis. Plant Physiol. 2004;136(1):2862–74.

34. Jin XY, Yin XF, Ndayambaza B, Zhang ZS, Min XY, Lin XS, Wang YR, Liu WX. Genome-wide identification and expression profiling of the ERF gene family in medicago sativa L. Under various abiotic stresses. DNA Cell Biol. 2019;38(10):1056–68.

35. Wang MZ, Liu C, Li SX, Zhu DY, Zhao Q, Yu JJ. Improved nutritive quality and salt resistance in transgenic maize by simultaneously overexpression of a natural lysine-rich protein gene, SBgLR, and an ERF transcription factor gene, TSRF1. Int J Mol Sci. 2013;14(5):9459–74.

36. Guo J, Wang MH. Expression profiling of the DREB2 type gene from tomato (Solanum lycopersicum L.) under various abiotic stresses. Hortic Environ Biote. 2011;52(1):105–11.

37. Wang H, Huang ZJ, Chen Q, Zhang ZJ, Zhang HB, Wu YM, Huang DF, Huang RF. Ectopic overexpression of tomato JERF3 in tobacco activates downstream gene expression and enhances salt tolerance. Plant Mol Biol. 2004;55(2):183–92.

38. Lamesch P, Berardini TZ, Li DH, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012;40(D1):D1202–10.

39. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu JZ, Zhou SG, et al. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice. 2013;6:4.

40. Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB. A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research. Mol Plant Microbe In. 2012;25(12):1523–30.

41. Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, Bliek M, Boersma MR, Borghi L, Bruggmann R, et al. Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. Nat Plants. 2016;2(6):16074.

42. Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat Genet. 2014;46(3):270.

43. Sharma SK, Bolser D, de Boer J, Sonderkaer M, Amoros W, Carboni MF, DAmbrosio JM, delaCruz G, DiGenova A, Douches DS, et al. Construction of reference chromosome scale Pseudomolecules for potato integrating the potato genome with genetic and physical maps. Genes G3 Genom Genet. 2013;3(11):2031–47.

44. Barchi L, Pietrella M, Venturini L, Minio A, Toppino L, Acquadro A, Andolfo G, Aprea G, Avanzato C, Bassolino L, et al. A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. Sci Rep-Uk. 2019;9:11769.

45. Wang X, Gao L, Jiao C, Stravoravdis S, Hosmani PS, Saha S, Zhang J, Mainiero S, Strickler SR, Catala C, et al. Genome of Solanum pimpinellifolium provides insights into structural variants during tomato breeding. Nat Commun. 2020;11(1):5817.

46. Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, et al. The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. Nucleic Acids Res. 2015;43(D1):D1036–41.

47. Kim S, Cheong K, Park J, Kim MS, Kim J, Seo MK, Chae GY, Jang MJ, Mang H, Kwon SH, et al. TGFam-Finder: a novel solution for target-gene family annotation in plants. New Phytol. 2020;227(5):1568–81.

48. Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.

49. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772–80.

50. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.

51. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188–90.

52. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022;50(W1):W276–9.

53. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 2015;43(W1):W389–94.

54. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.

55. BioBam Bioinformatics. OmicsBox - Bioinformatics Made Easy. 2019. https://www.biobam.com/omicsbox.

56. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

57. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36(10):3420–35.

58. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.

59. Voorrips RE. MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered. 2002;93(1):77–8.

60. Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin HZ, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):49.

61. Hao ZD, Lv DK, Ge Y, Shi JS, Weijers D, Yu GC, Chen JH. Rideogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. Peerj Comput Sci. 2020;6:e251.

62. Qiao X, Li QH, Yin H, Qi KJ, Li LT, Wang RZ, Zhang SL, Paterson AH. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome Biol. 2019;20:38.

63. Loytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. 2014;1079:155–70.

64. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics. 2006;4(4):259–63.

65. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer; 2016.

66. Kang WH, Sim YM, Koo N, Nam JY, Lee J, Kim N, Jang H, Kim YM, Yeom SI. Transcriptome profiling of abiotic responses to heat, cold, salt, and osmotic stress of Capsicum annuum L. Sci Data. 2020;7(1):17.

67. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907.

68. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290.

Choi *et al. BMC Plant Biology*   (2023) 23:48

Page 15 of 15

69. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
70. Kumar L. M EF: Mfuzz: a software package for soft clustering of microarray data. Bioinformation. 2007;2(1):5–7.
71. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. Bmc Bioinformatics. 2005;6:31.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.