

Genomic re-annotation and expression analyses of B3 genes reveal lineage-specific evolution and potential roles of REM genes in pepper

Young-Soo Park (✉ pys122836@gmail.com)

University of Seoul

Hye Jeong Cho (✉ chyejeong0725@gmail.com)

University of Seoul

Seungill Kim (✉ ksi2204@uos.ac.kr)

University of Seoul

Research Article

Keywords: B3, transcription factors, Solanaceae, re-annotation, REM, pepper, abiotic stress

DOI: <https://doi.org/>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

1 **Genomic re-annotation and expression**
2 **analyses of *B3* genes reveal lineage-specific**
3 **evolution and potential roles of *REM* genes**
4 **in pepper**

5 Young-Soo Park, Hye Jeong Cho, and Seungill Kim*

6
7 Department of Environmental Horticulture, University of Seoul, Seoul 02504, Republic of
8 Korea

9 *Corresponding author(s): Seungill Kim* (ksi2204@uos.ac.kr)

24

25

Abstract

Background

The *B3* gene family, one of the largest plant-specific transcription factors, plays important roles in plant growth, seed development, and hormones. However, the *B3* gene family, especially the REM subfamily, has not been systematically and functionally studied.

Results

In this study, we performed genome-wide re-annotation of *B3* genes in five Solanaceae plants, *Arabidopsis thaliana*, and *Oryza sativa*, and finally predicted 1,039 *B3* genes, including 231 (22.2%) newly annotated genes. We found a striking abundance of *REM* genes in pepper species (*Capsicum annuum*, *Capsicum baccatum*, and *Capsicum chinense*). Comparative motif analysis revealed that REM and other subfamilies (ABI3/VP1, ARF, RAV, and HSI) consist of different amino acids. We verified that the large number of *REM* genes in pepper were included in the specific subgroup (G8) through the phylogenetic analysis. Chromosome location and evolutionary analyses suggested that the G8 subgroup genes evolved mainly via a pepper-specific recent tandem duplication on chromosomes 1 and 3 after speciation between pepper and other Solanaceae. RNA-seq analyses suggested the potential functions of *REM* genes under salt, heat, cold, and mannitol stress conditions in pepper (*C. annuum*).

Conclusions

51 Our study provides evolutionary and functional insights into the *REM* gene
52 family in pepper.

53

54 **Keywords:** B3, transcription factors, Solanaceae, re-annotation, REM,
55 pepper, abiotic stress

56

Background

The *B3* genes are a superfamily of plant-specific transcription factors. *B3* genes have been characterized as having one or more B3 domains consisting of approximately 110 amino acids, two α -helices, and seven β -sheets [1]. This domain was named because it was first discovered in the third basic domain of the maize *Viviparous-1* (*Vp1*) gene [2]. Based on domain architectures and motifs, *B3* genes are classified into five major subfamilies: ABI3/VP1 [3], ARF (Auxin Response Factor) [4], RAV (Related to ABI3/VP1) [5], REM (Reproductive Meristem) [6], and HSI (High-level expression of sugar-inducible gene) [7, 8]. ABI3/VP1, ARF, RAV, and HSI have been reported to play important roles such as seed development, auxin signaling pathway, flowering time, and maturation [9]. Recently, ABI3 in Arabidopsis has been reported to control several downstream genes to resist dehydration stress [10]. *ARF* gene known to be associated with resistance to *Bradyrhizobium* infection [11]. In addition, it was known that ARFs have potential roles in adaptation by regulating soluble sugar content, maintaining chlorophyll content, and promoting root development under salt and drought stresses [12]. RAVs play an important role in plant disease resistance such as cassava bacterial blight [13]. It is known that MED13, which are subunits of the CDK8 module, depend on *HSI* gene to suppress the seed maturation [14]. Although it was recently reported that downregulated *REM34* and *REM35* lead to early arrest of gametophytic development in both male and female Arabidopsis [15], REM still has been recognized as one of the subfamilies and not yet studied in major crops.

Recent genomic studies for the *B3* gene family have revealed that a significant presence of *REM* genes within the B3 subfamilies, but the primary focus of these studies has been to investigate the structural characteristics and phylogenetic relationships within the REM subfamily for classification [16, 17].

The Solanaceae family, which belongs to the asteroid phylogeny of eudicots, includes economically important major crops such as tomato (*Solanum lycopersicum*), pepper (*C. annuum*), and potato (*Solanum tuberosum*). In addition, advances in sequencing technologies have enabled the construction of high-quality genome resources for these species, which are well deposited in public databases [18-21]. Using these resources, the genomic structures and molecular functions of the B3 subfamilies in Solanaceae were also well analyzed. Specifically, the function of *ARF* in tomato genome was known to be a defense response through the regulation of the auxin pathway [22]. It was also known that *CaRAV1* has a role as a transcriptional activator that induces resistance to bacterial infection in pepper [23]. However, the evolutionary process and potential roles of *REM* genes in Solanaceae remain unclear.

In this study, we conducted a re-annotation of the *B3* genes in seven plants: *A. thaliana*, *O. sativa*, and five Solanaceae species. The 1,039 *B3* genes were identified, including 231 (22.2%) genes that were omitted in the previous annotation. We found that *REM* genes were mostly abundant in pepper genomes. Through comparative and evolutionary analyses, we classified *B3* genes in the seven genomes into 12 subgroups (G1-12) and

identified that a large number of *REM* genes were clustered in the pepper-specific subgroup, G8. The microsynteny, chromosome location, and duplication analyses suggested that the pepper *REM* genes in G8 were recently expanded by lineage-specific tandem gene duplications after the divergence between pepper and other Solanaceae. In addition, expression analyses suggested that pepper *REM* genes are associated with functions under abiotic stresses such as cold, heat, mannitol, and salt. Our results with updated *B3* gene annotations will serve as a fundamental resource for genomic, functional, and breeding research, especially in pepper.

Results and Discussion

Annotation update and genomic characteristics of *B3* genes

We performed re-annotation of the *B3* genes for seven species, including *A. thaliana*, *O. sativa*, and five species of Solanaceae. A total of 1,039 *B3* genes were annotated containing 231 (22.2%) newly identified genes (Table 1). Specifically, we found 75.3% of newly identified *B3* genes in three pepper genomes (*Capsicum* species), ranging from 53 to 67 in individual species. We investigated the domain architectures for *B3* genes and classified them into five known subfamilies: RAV, HSI, ABI3/VP1, ARF, and REM. Specifically, *ABI3/VP1* and *REM* genes had only B3 domain(s), whereas *RAV*, *HSI*, and *ARF* genes had additional AP2, CW-type zinc finger, and auxin response factor domains, respectively, with B3 domain(s) (Fig. 1A). When we examined the ratio and number of *B3* genes from the five subfamilies, *REM* genes were mostly abundant compared to those of other

families, especially in three pepper species (Fig. 1B). Specifically, *REM* genes accounted for 78.2% of the newly identified genes in pepper genomes (Table S3). These results indicate that updated *B3* genes from the re-annotation process provide accurate gene repertoires of *B3* genes, especially for *REM* genes in pepper genomes.

To explore the sequence differences of the B3 domain among the five subfamilies, we examined the amino acid sequence of the B3 domain using the updated *B3* genes (Fig. 1C). The B3 domain consisted of approximately 110 amino acids, including two α -helices and seven β -sheets. The position of the first deduced α -helix from the B3 domain was located between the second and third β -sheets, otherwise, the second deduced α -helix was located between the fifth and sixth β -sheets. Given the secondary structure, we subdivided sections of the B3 domain and found that most of the sections were not conserved. Because *REM* genes were particularly variable, we separated *B3* genes in the five subfamilies into two types (*REM* and other families) and observed high conservation among genes in other families (Fig. S1). This indicates that the *REM* genes have contributed to increased sequence diversity among the B3 domains. We also examined the motif structure of the B3 domain and verified representative motifs abundant in *REM* or other subfamilies. Motifs #3, #9, #14, and #17 were enriched in other families, whereas motifs #5, #6, #19, and #20 were abundant in *REM* (Fig. 1D). This implies that *REM* and other families contain distinct B3 domains and thus may have different functions.

Gene Ontology (GO) analysis was performed to understand the potential functions of *B3* genes in seven species (Fig. 1E). Many *B3* genes were predicted to be associated with a binding in molecular function and nucleus function in the cellular component. Because transcription factors are generally known to control the transcription of specific genes [24], the *B3* genes were also predicted to play a role as a transcription factor. To examine the putative function of *B3* genes in subfamilies, we separated the results of GO analysis by families and found different functions in specific subfamilies. Specifically, *ARF* genes were thought to function in biological processes related to the auxin-activated signaling pathway, suggesting that these genes are involved in the auxin response factor and AUX/IAA proteins (Fig. S2). Taken together, our data generated from the updated *B3* genes provide accurate subfamily repertoires, genomic structures, and potential function of *B3* genes in Solanaceae species with *A. thaliana* and *O. sativa*.

Motif compositions of REM and other subfamilies of *B3* genes

To compare the motif composition of REM and other families, we analyzed the motif structure and verified 50 conserved motifs in the updated *B3* genes. A total of 36 motifs were mapped at 19 positions, except for 14 motifs that were observed at multiple locations as repetitive motif sequences (Fig. 2). Specifically, two regions of B3 domains were presented in the *REM* genes, at positions 4th to 9th and 15th to 18th, respectively (Fig. 2A). However, we found that *B3* genes from other families have a B3

domain located at positions 4th to 10th (Fig. 2B). We also analyzed the non-B3 domain regions and found significant differences between REM and other families mainly due to additional domains in RAV, HSI, and ARF families. These results illustrate distinct amino-acid sequence repertoires between REM and other subfamilies. When we examined motif positions in REM or other subfamilies, enriched motifs of *REM* genes were mainly observed in B3 domain regions, whereas the majority of specifically abundant motifs in other subfamilies were positioned in the flanking region as well as within the B3 domain (Fig. 2). These represent that the *REM* genes consisted of different sequences compared to the *B3* genes in other families, and thus may have undergone an independent evolutionary process. Among the B3 domains in the *REM* genes, we found that the first and remaining B3 domain regions have similar motif configurations (Fig. 2A). This suggests that *REM* genes may evolve by gaining repetitive domains.

Copy number expansion of specific *REM* genes in pepper

To elucidate the evolutionary relationship of *B3* genes, we constructed a phylogenetic tree with 231 updated *B3* genes in the seven plant genomes. Based on the motif compositions of genes and tree branches, we classified 1,010 genes into 12 subgroups (G1-12) (Fig. 3A). Specifically, *REM* genes were constructed into a large lineage that was grouped into G5-11. The motifs in the front and back of the B3 domain were examined to identify the characteristics of each subgroup (Fig. 3B). Specifically, we found that

motif #11 was observed within the ARF family, but motifs #13 and #24 were unique to G3 and G4, respectively. This indicates that these motifs contributed to the generation of the genomic diversity of *ARF* genes among those subgroups. Conversely, we observed that dominant motifs among the REM subgroups were conserved overall and shared with motifs #27 and/or #4. However, motif #4 of G7 and G8 had the characteristic of being accompanied by motifs #28 and #41, respectively. The G9 and G10 also showed different characteristics for each subgroup, such as a difference in the number of B3 domains. Based on these conserved structures, this result implies that the duplication of *REM* genes occurred rapidly, resulting in the conservation of genomic sequences of genes in REM subgroups. We then examined the copy number of *B3* genes in each subgroup by species and verified that the overall number of genes in each subgroup of Solanaceae was similar, except for the number of genes in G8 (Fig. 3C). The largest number of genes were congregated in G8, followed by G4. In the G8 subgroup, a large number of *REM* genes (90.64%) were clustered in three pepper genomes, whereas genes in potato and tomato were rarely observed. This suggests that the specific *REM* genes in pepper species have been expanded by lineage-specific evolution, resulting in the conservation of a large pool of *REM* genes in pepper species.

Next, we conducted the chromosome location of the genes for pepper (*C. annuum*), potato, and tomato (Fig. S3). We found that most of the genes were unevenly distributed across 12 chromosomes, suggesting different repertoires of *B3* genes among Solanaceae. In particular, the

tandem array of genes in G8, located on chromosomes 1 and 3 of the pepper genome, was observed. To compare the genomic regions of the G8 genes on chromosomes 1 and 3 in the pepper genome with the corresponding regions in the Solanaceae genomes, we performed synteny analyses for chromosomes 1 and 3 of pepper and chromosomes 8 and 9 of potato and tomato, respectively (Fig. 3D). Of the 17 and 14 pepper *B3* genes of G8 on chromosomes 1 and 3, respectively, we found only 12.5% and none of the orthologous genes in their corresponding regions in potato and tomato. Because we observed only a few orthologous relationships between the pepper *B3* genes in G8 and the other two species, we assumed that the copy number expansion of the pepper *B3* genes in G8 had recently occurred mainly after speciation. To verify this, synonymous substitution rate (Ks) values were calculated between duplication pairs of G8 in pepper, potato, and tomato, respectively, to estimate the emergence time of genes from G8 in the three species (Fig. 3E). The average MYA of pepper, potato, and tomato were about 14.9 (0.21 Ks), 30.8 (0.43 Ks), and 50.1 (0.7 Ks), respectively. Because 80% of the G8 genes in pepper were smaller than 21.6 MYA (0.3 Ks), we constructed dendrogram based on duplication time for the pepper G8 genes that located on the chromosomes 1 and 3 to identify how tandem duplications were occurred (Fig. S4). Considering that the divergence point between *Capsicum* and *Solanum* species was around 0.3 Ks [19], our result suggests that the pepper *REM* genes in G8 have been rapidly duplicated and specifically expanded after speciation by tandem duplication especially in chromosome 1 and 3. Taken together, our

data indicates that pepper-specific copy number expansion of *REM* genes via recent tandem duplication events probably has played a crucial role in the construction of distinct *B3* gene repertoires in pepper compared to other Solanaceae species.

Expression and potential roles of pepper *B3* genes under abiotic stress conditions

To investigate the potential role of the *B3* genes in pepper (*C. annuum*) under abiotic stress conditions, we performed RNA-seq analyses and identified differentially expressed genes (DEGs) under cold, heat, mannitol, and salt stresses. Because previous studies reported the speculation of gene functions by detecting genes that had similar expression patterns with DEGs under certain conditions [25, 26], we also compared the expression of *B3* genes and DEGs and grouped them into three clusters (C1-3) for each stress condition given similar expression patterns (Fig. 4A). The expressed *B3* genes were most abundant in mannitol stress with 85 genes, followed by 84 genes, 79 genes, and 69 genes for heat, salt, and cold stress, respectively. This suggests that these *B3* genes, which were similarly expressed with DEGs, may have roles related to each stress. Among the subgroups, we observed many *REM* genes of G8 in clusters, containing 15 genes under cold, 24 genes under heat, 22 genes under mannitol, and 22 genes under salt (Fig. 4B). Specifically, these genes in G8 were abundant in specific clusters. For example, 60% (cold C3), 42% (heat C3), 59% (mannitol C3), and 59% (salt C2) of the *B3* genes were in

G8. These suggest that the potential role of pepper-specific expanded *REM* genes is relevant to various abiotic stresses. When we examined the GO terms of whole genes in two clusters by each abiotic stress condition, including abundant *REM* genes in G8, we observed various abiotic stress-related functions, such as response to auxin (GO:0009733) in cold and heat, small molecule biosynthetic process (GO:0044283) in heat and mannitol, and signaling receptor activity (GO:0038023) in salt (Fig. 4C). In addition, many genes in specific clusters, such as cold C1, heat C3, mannitol C3, and salt C3, were associated with an RNA modification (GO:0009451) encoding a pentatricopeptide repeat (PPR) domain. Our results suggest the potential role of pepper *B3*, especially *REM* genes in G8 under abiotic stress conditions via association with a variety of other genes.

Co-expression network and functional association of pepper *REM* genes under abiotic stress conditions

We detected *B3* DEGs in expression clusters and verified an overall similar distribution across four stresses: 33 (cold), 26 (heat), 26 (mannitol), and 26 (salt). Of these, the *REM* DEGs, especially in the G8, were mostly abundant regardless of the stresses (Fig. 5A). Furthermore, we identified 25 stress-specific DEGs, and 16 (8) of them belonged to REM (G8) (Fig. 5B). Co-expression networks of target genes suggest their potential roles given the repertoires of linked genes in the expression network [27]. We constructed co-expression networks of stress-specific *REM* DEGs with other DEGs in the same expression clusters to understand the specific

roles of pepper *REM* genes under abiotic stress conditions. Our analyses revealed that pepper *REM* genes in G8 could be involved in cold and mannitol stress-induced functions of various genes (Fig. 5C). In particular, we detected that *CaREM210* and *CaREM205* were co-expressed with three *PPRs* and a variety of genes under cold and mannitol conditions, respectively. Previous studies have reported the functions of *PPRs* under cold and mannitol conditions. The repressed expression of *TCD10* (*LOC_Os10g28600*) and *OsV4* (*LOC_Os04g39970*) genes in rice caused abnormal chloroplast development at low temperatures [28, 29]. *SOAR1* (*At5g11310*) in Arabidopsis was the positive response to particularly cold and osmotic stress through the regulation of ABA signaling [30]. In addition to the *PPR* genes, *OsRH42* (*Os08g0159900*), which encodes a DEAD-box helicase in rice, is important for pre-mRNA splicing under cold stress [31]. The *Rboh* gene (*LOC107862088*, *LOC107875997*), encoding several domains such as EF-hand, NADPH oxidase, and FAD-binding in pepper, was known to be activated by the binding of Ca cations to the EF-hand, causing the accumulation of malondialdehyde in response to mannitol stress [32]. Our data suggest that pepper *REM* genes in G8 could be involved in the regulation of cold and mannitol stress-related traits with *PPR* and other various genes via examination of co-expressed genes with pepper *REM* genes in G8 under those stress conditions.

Furthermore, we observed diverse genes that co-expressed with pepper *REM* genes in G8 and G9 under salt and heat stress (Fig. 5D). In salt stress, we also found genes involved in the regulation of ABA [30, 33],

the regulation of ROS production [34], and the early events of the signal transduction pathway [35] such as CCT, PPR, and protein kinase domain, as described in previous studies. For example, overexpression of *AtCOL4* (*At5g24930*), which encodes the CCT domain in Arabidopsis, is known to regulate ABA synthesis and stress-related genes under salt stress [33]. These results suggest that the pepper *REM* genes in G8 may play a role in the salt stress condition through co-expression with other salt stress-related genes. In heat stress, *CaREM365* in G9 was co-expressed with nine genes having various domains such as protein kinase, α -amylase, and lactate/malate dehydrogenase, suggesting the putative role of the pepper *REM* gene in G9 with these genes. Taken together, our data comprehensively suggest the underlying roles of pepper *REM* genes, especially those belonging to the pepper-specific expanded G8 via association with a variety of genes involved in abiotic stress responses based on the investigation of DEGs and co-expression networks.

Conclusions

Construction of gene annotations without omission of genes that existed in genome assemblies is a crucial step for gene family studies [36-40]. In this study, we performed an annotation update of the *B3* genes and identified 231 new gene models in five Solanaceae, *A. thaliana*, and *O. sativa* genomes. Notably, newly annotated genes in pepper accounted for 78.2% of novel *REM* genes. The motif analyses of the updated *B3* genes showed the different amino acid composition of B3 domains between *REM*

and other *B3* genes, indicating that they may have undergone independent evolutionary processes. Based on the phylogenetic relationship, we divided the *B3* genes into 12 subgroups and found a lineage-specific burst of pepper *REM* genes in G8. These pepper *REM* genes in G8 comprised 50.6 % of the 174 newly annotated pepper *B3* genes, indicating the significantly improved *B3* gene annotation of the pepper genome through the identification of pepper-enriched *REM* genes in this study. These pepper *REM* genes in G8 were mainly clustered on chromosomes 1 and 3. These pepper *REM* genes were overall non-syntenic with *REM* genes in corresponding regions of the tomato and potato genomes. These also indicated pepper-specific evolution of those *REM* genes via recent tandem duplication after speciation between pepper and other Solanaceae species.

Because our analyses suggested that pepper *REM* genes, especially in G8, could be pepper-specific genes and thus related to pepper-specific traits, we focused on RNA-seq analyses mainly for *REM* genes in pepper (*C. annuum*) under cold, heat, mannitol, and salt stress conditions. We found that the DEGs in expression clusters, including abundant *REM* genes, were overall involved in RNA modification in response to abiotic stresses. Furthermore, the co-expression regulatory networks of stress-specific *REM* genes suggested that specific *REM* genes in G8, in particular, were co-expressed with previously known functional genes such as several PPRs under cold and mannitol stresses [28-30]. Consequently, our results with the updated *B3* genes provide insights into the genomic structural features,

evolutionary history, and potential roles of pepper *B3* genes, including pepper-specific evolved *REM* genes.

Materials and Methods

Re-annotation of the *B3* gene family in seven plant genomes

Data for genome sequences of *A. thaliana* [41], *O. sativa* [42], *C. annuum* [19], *C. baccatum* [20], *C. chinense* [20], *S. tuberosum* [21], and *S. lycopersicum* [18] were downloaded for re-annotation of *B3* genes. Re-annotation of *B3* genes was conducted using TGFam-Finder v1.20 [43], which tool was developed to focus on identification of missing genes in annotations based on protein mapping, transcriptome data, and ab initio prediction. As described in the previous study [43] with parameters 'EXTENSION_LENGTH' = 200,000, 'MAX_INTRON_LENGTH'=200,000, 'HMM_CUTOFF'=1e-4 (Table S1). The TSV file was created from InterProScan 5 (-f tsv -appl Pfam) [44] using "TSV_FOR_DOMAIN_IDENTIFICATION", and the target ID was set to "PF02362 (B3)" by the PFAM (<http://pfam.xfam.org/>) and Hidden Markov Model (HMM) databases. We used several Pfam IDs such as PF06507 (Auxin response factor domain), PF02309 (AUX/IAA family), PF00847 (AP2 domain), and PF07496 (CW-type zinc finger domain) to separate the *B3* subfamilies. Finally, we assigned new names to the updated *B3* genes instead of using the published annotation names (Table S2).

Conserved amino acids in the B3 domain

The acid sequence of the B3 domains was extracted for the seven plant genomes. To align the B3 domain sequences, we utilized MAFFT v7.470 (--reorder --maxiterate 1000) [45], and then TrimAL v1.4 [46] with the gt 0.2 parameter to trim the alignment. We visualized the amino acid sequence composition of all B3 domains using WebLogo v2.8.22 [47]. To elucidate the consensus sequence of the B3 domain according to the subfamily, the EMBOSS Cons program [48] was used with the plurality 0.1 option. Multiple alignments were performed using Jpred (default parameters) [49] to predict the secondary protein structure of the B3 domain. We divided the B3 domain into 19 subdivisions according to the protein structure. Average consensus scores for each division were calculated by MACSIMS from Jalview programs [50].

Statistical test for enrichment

To identify which motif is specifically enriched in genes belonging to REM or others (ABI3/VP1, ARF, RAV, and HSI), we used Fisher's test and Chi-square test using an in-house Perl script executed with the Statistics::R module R. P-values were computed by the Monte Carlo test (B=10,000).

Gene ontology term of *B3* genes

Functional annotation of *B3* genes was conducted using OmicsBox v2.2.4 [51]. To align B3 protein sequences to the NCBI database for non-redundant protein database (nr v5), BLASTP was used with a 1e-3 e-value cut-off. Blast2GO mapping and annotation were performed using the Blast

results that were merged with the InterProScan results [44]. Next, we divided the GO terms of each *B3* gene into three categories (biological process, molecular function, and cellular component). From direct GO, the top five GO terms for each category were displayed.

Motif compositions of *B3* genes

To determine the conserved motifs of the *B3* genes, MEME v5.1.1 (-protein -mod zoops -nmotifs 50 -minw 10 -maxw 50 -objfun se -markov_order 0) [52] was conducted. A total of 50 motifs were matched to protein sequences by MAST v5.1.1 [53]. Conserved motif positions were determined manually based on sequence alignments and motif compositions, except for recurring motifs at various positions.

Phylogenetic analysis of *B3* genes

The 1,039 re-annotated *B3* genes were aligned using MAFFT v7.470 [45]. Ambiguous alignments were removed with TrimAL v1.4 with trim option gt 0.1 [46]. To infer phylogenetic relationships, the maximum likelihood tree was constructed using IQ-TREE v2.0.6 [54] with bootstrap replicates of 1,000 ultrafast parameters. Interactive Tree of Life (iTOL) v6 was utilized to visualize the tree. The final tree of *B3* genes was organized into 12 subgroups (G1-12) based on domains and motifs.

Chromosome distribution and microsynteny of *B3* genes in pepper, potato, and tomato

The physical locations of the *B3* genes, excluding unanchored scaffolds, were determined from the GFF files generated by TGFam-Finder v1.20 [43]. We used Mapchart to show the distribution of *B3* genes on the chromosomes [55]. Subgroups of all genes were represented in different colors according to the phylogenetic tree.

To show the orthologous relationships of the G8 genes, microsynteny analysis was performed. All-by-all comparison of BLASTP [56] and GFF files obtained by TGFam-Finder v1.20 [43] results were identified using the Multiple Collinearity Scan toolkit (MCScanX) [57] with parameters such as a match score 50, and a match size 3. We used RIdeogram in R packages to represent the genomic locations for each gene pair [58].

Duplication history of G8 genes in pepper, potato, and tomato

Duplicated pairs of *B3* genes were identified using DupGen_Finder [59], and the duplication time of those pairs was calculated. Multiple alignments were conducted with PRANK (-codon) using the coding sequences of each gene pair. We estimated the synonymous substitution rates (Ks) of each duplicated *B3* gene using KaKs_Calculator 2.0 (-m MYN) [60]. To calculate million years ago (MYA), we used the formula $T = Ks/2\lambda$. Each of λ was assumed to be 6.96×10^{-9} [61-63]. We displayed the duplication time with Beeswarm in R packages.

Transcriptome analyses of pepper and tomato *B3* genes under abiotic stresses

463 We first downloaded RNA sequencing data of pepper (SRR8692625,
464 SRR8692629, SRR8692600, SRR8692599, SRR8692602, SRR8692601,
465 SRR8692596, SRR8692595, SRR8692598, SRR8692597, SRR8692624,
466 SRR8692623, SRR8692622, SRR8692621, SRR8692620, SRR8692619,
467 SRR8692618, SRR8692580, SRR8692592, SRR8692575, SRR8692576,
468 SRR8692577, SRR8692610, SRR8692609, SRR8692608, SRR8692591,
469 SRR8692568, SRR8692569, SRR8692570, SRR8692571, SRR8692572,
470 SRR8692573, SRR8692574, SRR8692626, SRR8692630, SRR8692615,
471 SRR8692581, SRR8692639, SRR8692583, SRR8692628, SRR8692582,
472 SRR8692584, SRR8692586, SRR8692612, SRR8692643, SRR8692640,
473 SRR8692585, SRR8692578, SRR8692607, SRR8692631, SRR8692632,
474 SRR8692637, SRR8692638, SRR8692635, SRR8692636, SRR8692606,
475 SRR8692605, SRR8692604, SRR8692603, SRR8692614, SRR8692613,
476 SRR8692633, SRR8692634, SRR8692641, SRR8692593, SRR8692617,
477 SRR8692616, SRR8692627, SRR8692645, SRR8692611, SRR8692642,
478 SRR8692588, SRR8692587, SRR8692590, SRR8692589, SRR8692579,
479 SRR8692644, SRR8692594) [64] under abiotic stresses for cold, heat,
480 mannitol, and salt at different times:3h, 6h, 12h, 24h, and 72h. To
481 eliminate low-quality RNA sequencing data, trimming was performed by
482 CLC Assembly Cell (CLC Bio, Aarhus, Denmark) using fastq raw files. Next,
483 we conducted HISAT2 (-dta -x) [65] and StringTie (-e -B -G) [66] to map
484 the *C. annuum* reference genome and calculate the fragment per kilobase
485 transcript per million mapped reads (FPKM) values of the updated *B3*
486 genes. To convert FPKM values to read counts, we used Python scripts

(prepDE.py). We examined differentially expressed genes (DEGs) with DESeq2 in R software with $|\log_2\text{FoldChange}| > 1$ and adjusted p-value < 0.05 [67]. Clustering analyses were completed on *B3* genes and DEGs from Mfuzz [68] programs in R packages with $\log_2(\text{FPKM} + 1)$ under abiotic stress. Three clusters of each stress were determined according to the k-means algorithm. GO annotation was then retrieved with OmicsBox v2.2.4 [51] for clusters containing abundant *REM* genes in G8. In addition, we carried out an enrichment test of GO terms from Fisher's exact test (false discovery rate corrected p-value ≤ 0.01) and showed more specific GO terms with the reduction to the most specific option. Co-expression network analysis based on expression clusters was performed using WGCNA [69] in R packages with optimal β (soft thresholding power) values selected by SFT.R.sq over 0.8 for all stress conditions and minModuleSize of 30. We visualized networks using the Cytoscape v3.9.1 program.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets provided in this study can be detected online, and the accession numbers are written in the article or additional files.

Competing interests

None declared.

Funding

This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2022R1C1C1004918) to S.K., and by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPET) through the Digital Breeding Transformation Technology Development Program funded by Ministry of Agriculture, Food, and Rural Affairs (MAFRA) (322075-3) to S.K.

Author's contributions

S.K. designed the study. Y.-S.P., H.J.C., and S.K. carried out the re-annotation and comparative analyses. Y.-S.P. wrote the first manuscript draft. All authors contributed to the editing and review of the final version.

Acknowledgments

Not applicable.

Abbreviations

- 534 ARF, Auxin Response Factor
- 535 RAV, Related to ABI3/VP1
- 536 REM, Reproductive Meristem
- 537 HSI, High-level expression of sugar-inducible gene
- 538 GO, Gene Ontology
- 539 DEG, Differentially expressed gene
- 540 FPKM, Fragment Per Kilobase of transcript per Million mapped reads

541 **Table 1.** The number of re-annotated *B3* genes in seven species.

Species	Previously annotated genes	Newly annotated genes	Total
<i>O. sativa</i>	96	5	101
<i>A. thaliana</i>	116	5	121
<i>C. annuum</i>	120	54	174
<i>C. baccatum</i>	119	53	172
<i>C. chinense</i>	118	67	185
<i>S. tuberosum</i>	125	15	140
<i>S. lycopersicum</i>	114	32	146
Total	808	231	1,039

542

Figure legends

Figure 1. Genomic features of the *B3* genes in seven plant genomes.

(A) The number of B3 subfamilies and their domain architectures in seven plant genomes. Different symbols located on the left represent the corresponding domains. The color of the bars indicates subfamilies. The dots next to B3-REM express one or more of the B3 domains. (B) The ratio and number of subfamilies per species. The bar colors present each subfamily. (C) Multiple sequence alignments and secondary structures of B3 domain sequences in seven plant genomes. The secondary structure positions are shown above the amino acid logo. The height of the logos within the stack shows the relative frequency at each position. In the sequence alignment, it is capitalized if more than half of the genes have a particular amino acid at that position, otherwise, it is lowercase. Gaps are marked with dashes. The calculated sequence conservation is presented as a bar diagram. (D) Enriched motifs in the B3 domain of REM and other subfamilies. The numbers in brackets manifest the motif sequences shown in Table S4. The size of the numbers shows the frequency of motifs from genes. Colors display the result of the enrichment test for REM and other families ($p < 0.0001$). (E) The gene ontology (GO) term for updated B3 genes. The GO categories are listed below the heatmap.

Figure 2. Motif compositions of *B3* genes in seven species.

(A) The motif compositions of the *REM* genes. If one or more B3 domains are present, only the first B3 is depicted in the REM first B3 domain box, and

the remaining B3 domains are illustrated in the additional B3 box. (B) The motif compositions of the non-*REM* genes. The integrated domains are shown in the gray box. (A-B) The number in the bar indicates the motif sequence in Table S4. The result of the enrichment test is displayed as REM-enriched, Others-enriched, or neutral in different colored boxes ($p < 0.0001$).

Figure 3. Lineage-specific copy number expansion of *REM* genes in pepper. (A) The phylogenetic relationship of the *B3* genes among seven species. The outer and middle rings represent subfamilies and subgroups, respectively. Colored dots at the end of the branch indicate species. (B) The most abundant motifs near the B3 domain are shown. Motifs are portrayed as numbers in the box. The boxes are colored to show the type of enriched result of the motifs ($p < 0.0001$). (C) The number of subgroups indicated by species on the heat map. (D) Microsynteny relationship of G8 genes in pepper (*C. annuum*), potato, and tomato. The line implies an orthologous relationship between genes. Black and gray blocks mark the chromosome locations of orthologous and non-orthologous *B3* genes. White blocks mean orthologous genes of non-*B3* genes. (E) Duplication history of G8 genes in pepper, potato, and tomato. The Ks values of duplication pairs of genes are shown in a dot plot.

Figure 4. Expression analyses of pepper *B3* genes under abiotic stress. (A) Expression patterns of whole DEGs with *B3* genes. (B) The

number of *B3* genes in expression clusters. The heatmap represents the number of *B3* genes by subgroup. (C) Abundant GO descriptions of genes in representative clusters. Dot plots display the top three GO enrichment results for each cluster under abiotic stress. The shape and size of symbols indicate the cluster number and frequency of GO descriptions, respectively, depicted on the right side of the dot plots.

Figure 5. Co-expression network analysis of *REM* genes in pepper under abiotic stress. (A) The number of *B3* DEGs under each abiotic stress. Colors and shapes next to the diagram symbolize the subfamilies and subgroups. (B) DEGs of the *B3*, *REM*, and G8 genes. (C) The networks illustrate associations with G8 genes and other genes under cold and mannitol conditions. (D) Co-expression network of stress response-related *REM* genes under salt and heat (C-D) The number written in the circle provides domain information, which is located on the right side of the network. The empty circle represents an uncharacterized gene.

Additional files

Additional file 1: Fig. S1. Comparison of amino acid sequence alignment for REM and other (ABI3/VP1, ARF, RAV, and HSI) families. The α -helices and β -sheets are presented above the amino acid logo. Each logo consists of stacks of symbols indicating the most conserved amino acid for each position. The height of the logos within the stack shows the relative frequency at each position. In the sequence, it is uppercase if more than

half the genes have that amino acid, otherwise, it is lowercase. Gaps are marked with dashes. The calculated sequence conservation is presented as a bar diagram based on the secondary structure positions. **Fig. S2.** The proportion of top five gene ontology (GO) terms for each category among subfamilies. The GO descriptions are located on the left side of the heat map. **Fig. S3.** Chromosomal locations of *B3* genes in pepper (*C. annuum*), potato, and tomato. (A-C) (A), (B), and (C) show the distribution of *B3* genes for pepper, potato, and tomato in this order. The genes are denoted by different colors based on the subgroups. The tandem array genes are connected to the black lines to the left of the gene names. **Fig. S4.** Predicted the time and order of tandem duplication on chromosomes 1 and 3 for pepper G8 genes via hierarchical clustering. Gray blocks mark the chromosome locations of pepper G8 genes. The duplication history of G8 genes is shown in the dendrogram plot

Additional file 2: Table S1. List of genomic resources for five Solanaceae species, *Arabidopsis thaliana*, and *Oryza sativa*. **Table S2.** Data for re-annotated *B3* genes. **Table S3.** The number of updated *B3* genes for five subfamilies in seven species. **Table S4.** Motif sequences of *B3* genes in seven plant genomes. **Table S5.** log₂(Fold change) values of pepper *B3* genes in abiotic stress. **Table S6.** Domain information of co-expressed genes with core genes.

Figure 1

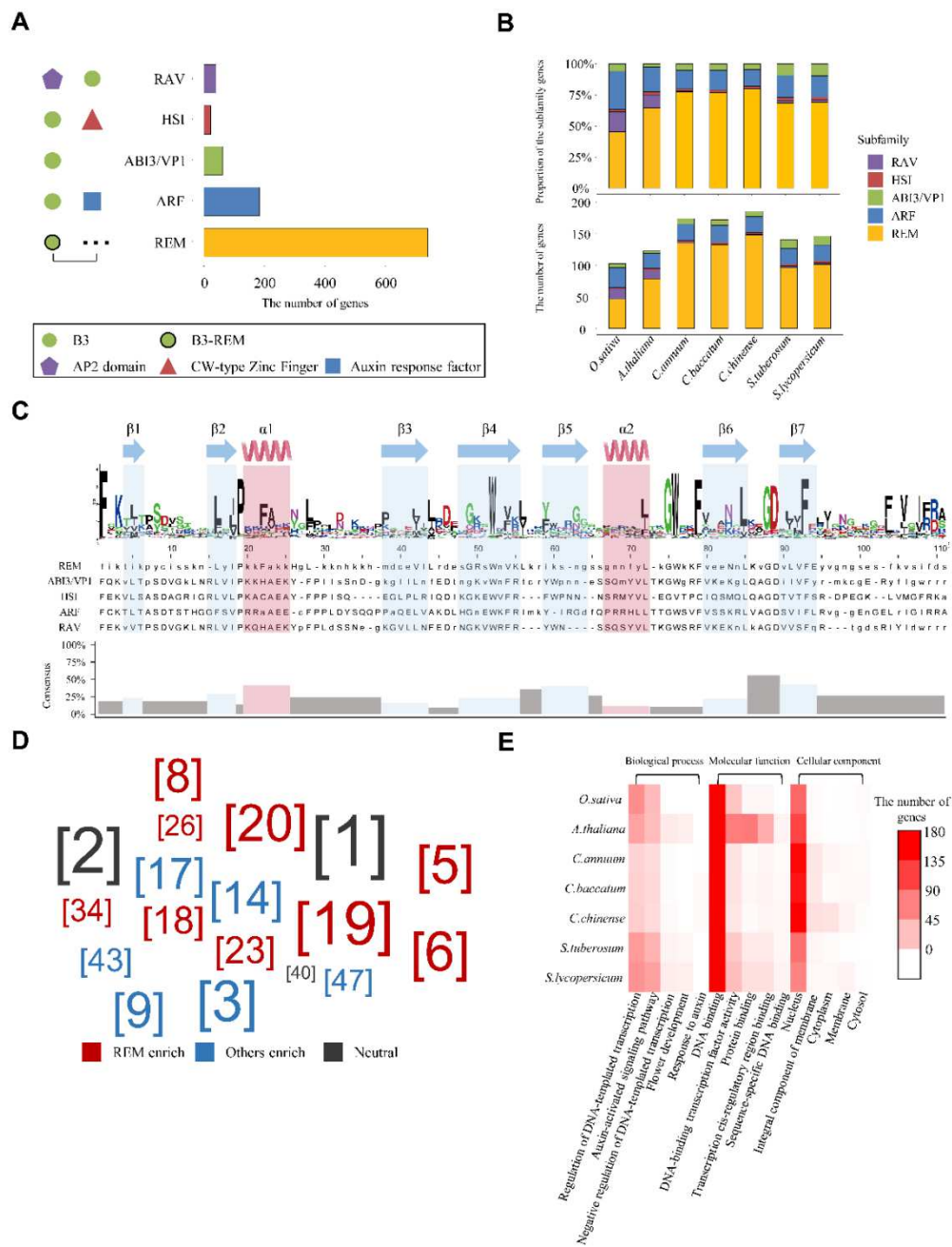
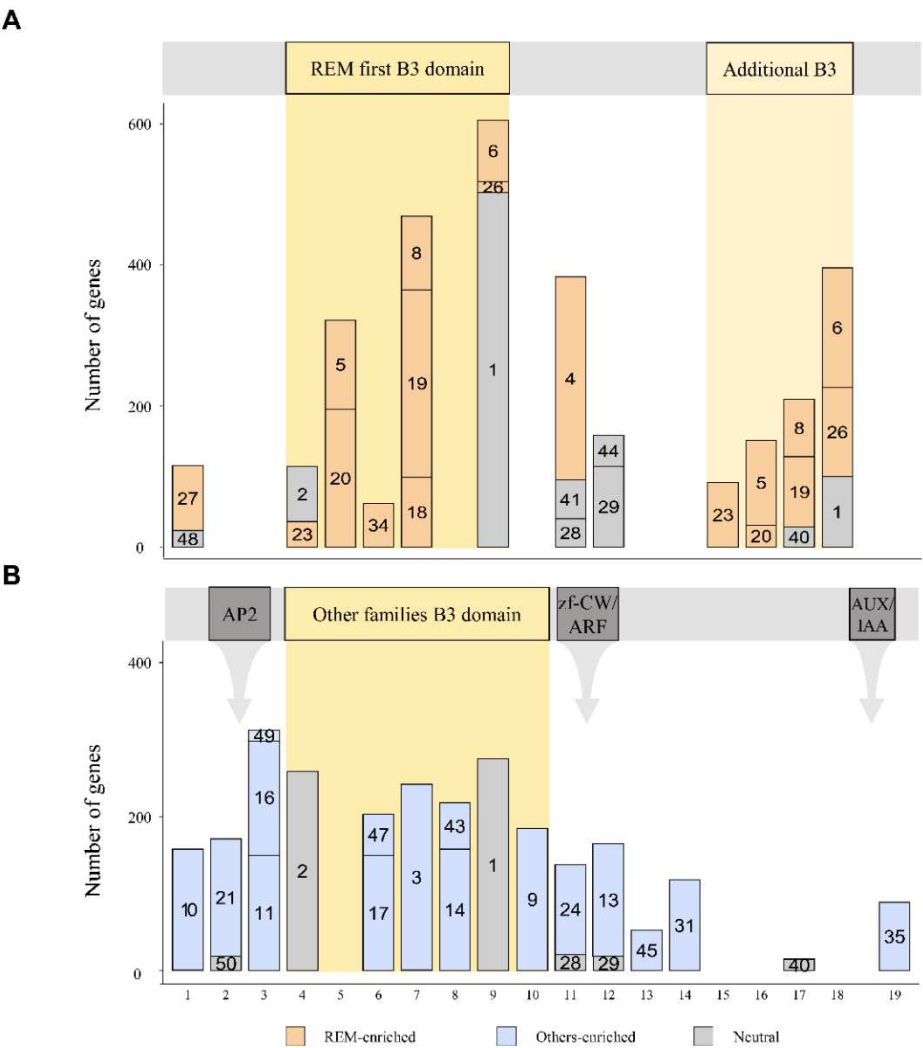


Figure 2



640

641

Figure 3

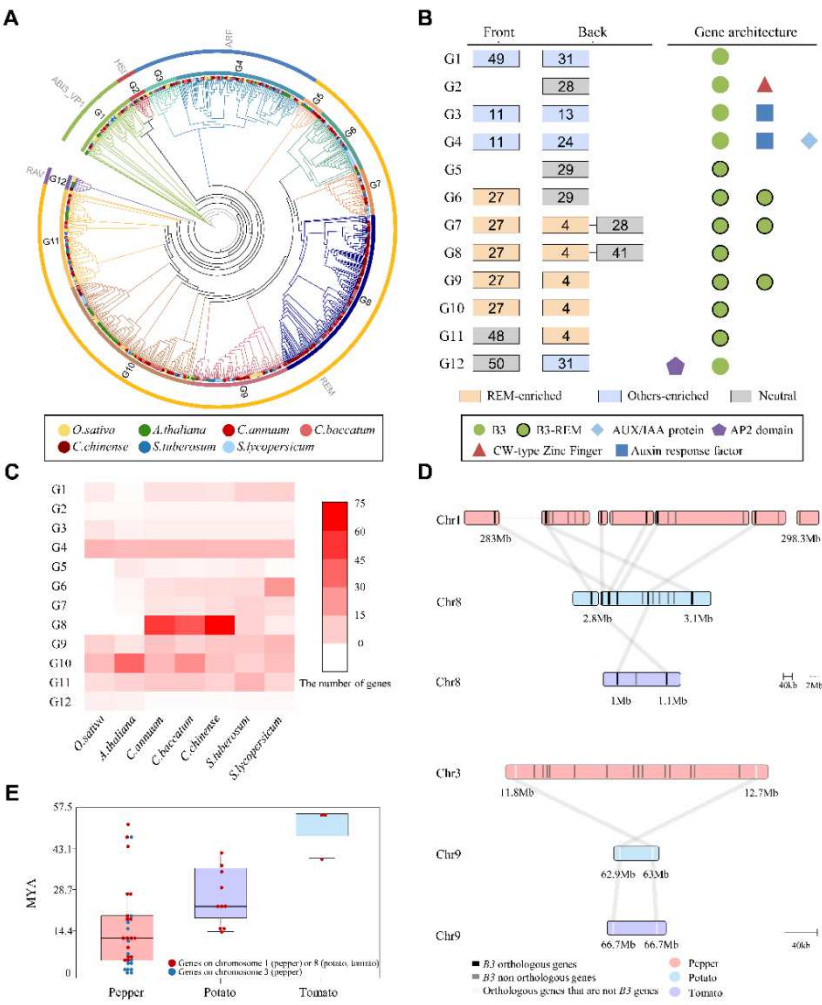


Figure 4

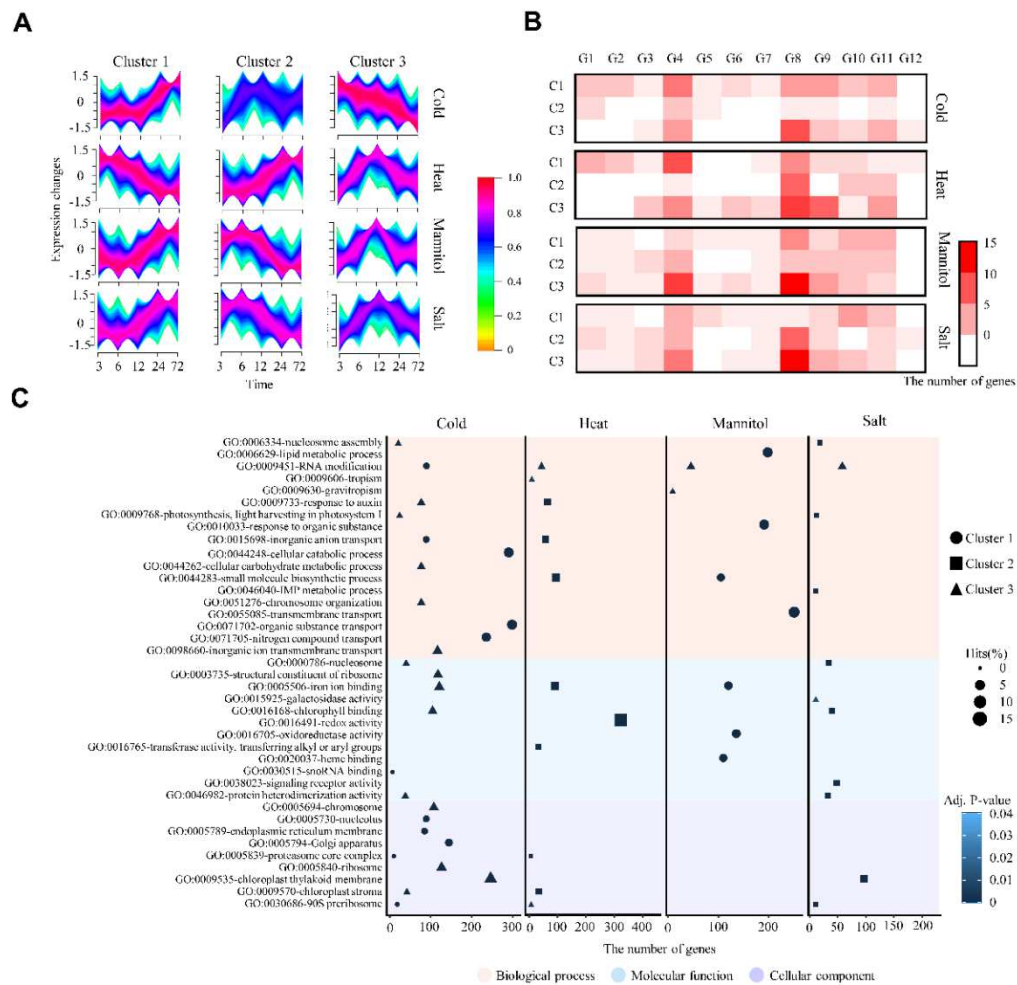
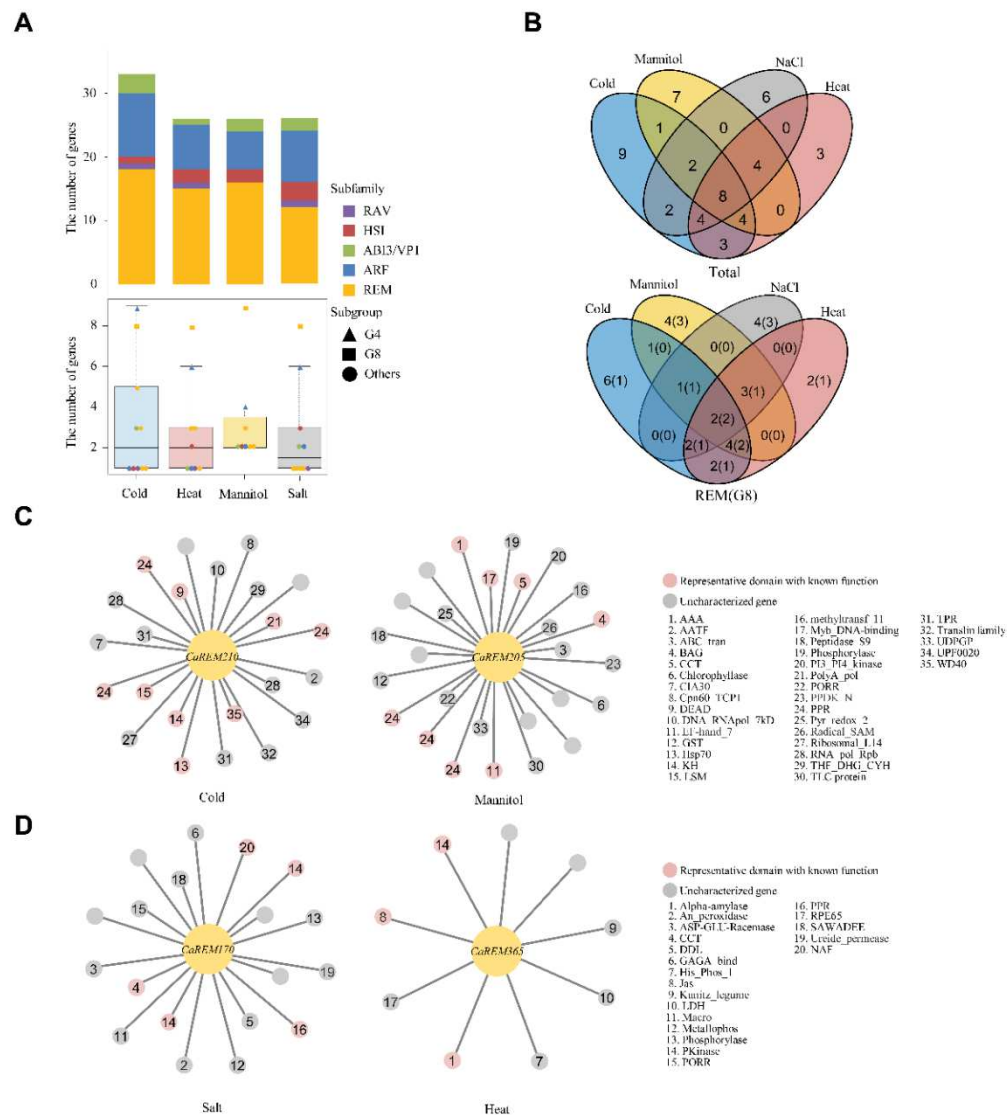


Figure 5



References

1. Swaminathan K, Peterson K, Jack T. The plant B3 superfamily. Trends Plant Sci. 2008; 13(12):647-655.
2. McCarty DR, Hattori T, Carson CB, Vasil V, Lazar M, Vasil IK. The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. Cell. 1991; 66(5):895-905.
3. Suzuki M, Kao CY, McCarty DR. The conserved B3 domain of *VIVIPAROUS1* has a cooperative DNA binding activity. Plant Cell. 1997; 9(5):799-807.
4. Ulmasov T, Murfett J, Hagen G, Guilfoyle TJ. Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. Plant Cell. 1997; 9(11):1963-1971.
5. Kagaya Y, Ohmiya K, Hattori T. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. Nucleic Acids Res. 1999; 27(2):470-478.
6. Franco-Zorrilla JM, Cubas P, Jarillo JA, Fernández-Calvin B, Salinas J, Martínez-Zapater JM. *AtREM1*, a member of a new family of B3 domain-containing genes, is preferentially expressed in reproductive meristems. Plant Physiol. 2002; 128(2):418-427.
7. Tsukagoshi H, Saijo T, Shibata D, Morikami A, Nakamura K. Analysis of a sugar response mutant of Arabidopsis identified a novel B3 domain protein that functions as an active transcriptional repressor. Plant Physiol. 2005; 138(2):675-685.

- 670 8. Suzuki M, Wang HH, McCarty DR. Repression of the *LEAFY*
671 *COTYLEDON 1/B3* regulatory network in plant embryo development by
672 *VP1/ABSCISIC ACID INSENSITIVE 3*-LIKE B3 genes. *Plant Physiol.* 2007;
673 143(2):902-911.
- 674 9. Romanel EA, Schrago CG, Couñago RM, Russo CA, Alves-Ferreira
675 M. Evolution of the B3 DNA binding superfamily: new insights into REM
676 family gene diversification. *PLoS One.* 2009; 4(6):e5791.
- 677 10. Bedi S, Sengupta S, Ray A, Chaudhuri RN. ABI3 mediates
678 dehydration stress recovery response in *Arabidopsis thaliana* by regulating
679 expression of downstream genes. *Plant Science.* 2016; 250:125-140.
- 680 11. Tang Y, Du G, Xiang J, Hu C, Li X, Wang W, Zhu H, Qiao L, Zhao C,
681 Wang J. Genome-wide identification of auxin response factor (*ARF*) gene
682 family and the miR160-ARF18-mediated response to salt stress in peanut
683 (*Arachis hypogaea* L.). *Genomics.* 2022; 114(1):171-184.
- 684 12. Verma S, Negi NP, Pareek S, Mudgal G, Kumar D. Auxin response
685 factors in plant adaptation to drought and salinity stress. *Physiol Plant.*
686 2022; 174(3):e13714.
- 687 13. Wei Y, Chang Y, Zeng H, Liu G, He C, Shi H. RAV transcription
688 factors are essential for disease resistance against cassava bacterial blight
689 via activation of melatonin biosynthesis genes. *J Pineal Res.* 2018; 64(1).
- 690 14. Chhun T, Chong SY, Park BS, Wong EC, Yin JL, Kim M, Chua NH.
691 HSI2 Repressor Recruits MED13 and HDA6 to Down-Regulate Seed
692 Maturation Gene Expression Directly During Arabidopsis Early Seedling
693 Growth. *Plant Cell Physiol.* 2016; 57(8):1689-1706.

- 694 15. Caselli F, Beretta VM, Mantegazza O, Petrella R, Leo G, Guazzotti
695 A, Herrera-Ubaldo H, de Folter S, Mendes MA, Kater MM et al. REM34
696 and REM35 Control Female and Male Gametophyte Development in
697 *Arabidopsis thaliana*. Front Plant Sci. 2019; 10:1351.
- 698 16. Ruan CC, Chen Z, Hu FC, Fan W, Wang XH, Guo LJ, Fan HY, Luo
699 ZW, Zhang ZL. Genome-wide characterization and expression profiling of
700 B3 superfamily during ethylene-induced flowering in pineapple (*Ananas*
701 *comosus* L.). BMC Genom. 2021; 22(1):561.
- 702 17. Wang W-B, Ao T, Zhang Y-Y, Wu D, Xu W, Han B, Liu A-Z. Genome-
703 wide analysis of the B3 transcription factors reveals that RcABI3/VP1
704 subfamily plays important roles in seed development and oil storage in
705 castor bean (*Ricinus communis*). Plant Divers. 2022; 44(2):201-212.
- 706 18. Zhou Y, Zhang Z, Bao Z, Li H, Lyu Y, Zan Y, Wu Y, Cheng L, Fang Y,
707 Wu K et al. Graph pangenome captures missing heritability and empowers
708 tomato breeding. Nature. 2022; 606(7914):527-534.
- 709 19. Kim S, Park M, Yeom S-I, Kim Y-M, Lee JM, Lee H-A, Seo E, Choi J,
710 Cheong K, Kim K-T. Genome sequence of the hot pepper provides insights
711 into the evolution of pungency in Capsicum species. Nat Genet. 2014;
712 46(3):270-278.
- 713 20. Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM,
714 Cheong K, Shin HS et al. New reference genome sequences of hot pepper
715 reveal the massive evolution of plant disease-resistance genes by
716 retroduplication. Genome Biol. 2017; 18(1):210.

- 717 21. Pham GM, Hamilton JP, Wood JC, Burke JT, Zhao H, Vaillancourt B,
718 Ou S, Jiang J, Buell CR. Construction of a chromosome-scale long-read
719 reference genome assembly for potato. *Gigascience*. 2020; 9(9).
- 720 22. Bouzroud S, Gouiaa S, Hu N, Bernadac A, Mila I, Bendaou N,
721 Smouni A, Bouzayen M, Zouine M. Auxin Response Factors (ARFs) are
722 potential mediators of auxin action in tomato response to biotic and abiotic
723 stress (*Solanum lycopersicum*). *PLoS One*. 2018; 13(2):e0193517.
- 724 23. Sohn KH, Lee SC, Jung HW, Hong JK, Hwang BK. Expression and
725 functional roles of the pepper pathogen-induced transcription factor RAV1
726 in bacterial disease resistance, and drought and salt stress tolerance. *Plant*
727 *Mol Biol*. 2006; 61(6):897-915.
- 728 24. Lehti-Shiu MD, Panchy N, Wang P, Uygun S, Shiu SH. Diversity,
729 expansion, and evolutionary novelty of plant DNA-binding transcription
730 factor families. *Biochim Biophys Acta*. 2017; 1860(1):3-20.
- 731 25. Feng MQ, Lu MD, Long JM, Yin ZP, Jiang N, Wang PB, Liu Y, Guo
732 WW, Wu XM. miR156 regulates somatic embryogenesis by modulating
733 starch accumulation in citrus. *J Exp Bot*. 2022; 73(18):6170-6185.
- 734 26. Shen C, Wei C, Li J, Zhang X, Wu Y. Integrated single-molecule long-
735 read sequencing and Illumina sequencing reveal the resistance
736 mechanism of *Psathyrostachys huashanica* in response to barley yellow
737 dwarf virus-GAV. *Phytopathology Res*. 2020; 2:1-15.
- 738 27. Wang H, Gu L, Zhang X, Liu M, Jiang H, Cai R, Zhao Y, Cheng B.
739 Global transcriptome and weighted gene co-expression network analyses

740 reveal hybrid-specific modules and candidate genes related to plant height
 741 development in maize. *Plant Mol Biol.* 2018; 98(3):187-203.

742 28. Wu L, Wu J, Liu Y, Gong X, Xu J, Lin D, Dong Y. The rice
 743 pentatricopeptide repeat gene *TCD10* is needed for chloroplast
 744 development under cold stress. *Rice.* 2016; 9(1):1-13.

745 29. Gong X, Su Q, Lin D, Jiang Q, Xu J, Zhang J, Teng S, Dong Y. The
 746 rice *OsV4* encoding a novel pentatricopeptide repeat protein is required
 747 for chloroplast development during the early leaf stage under cold stress.
 748 *J Integr Plant Biol.* 2014; 56(4):400-410.

749 30. Jiang SC, Mei C, Liang S, Yu YT, Lu K, Wu Z, Wang XF, Zhang DP.
 750 Crucial roles of the pentatricopeptide repeat protein SOAR1 in Arabidopsis
 751 response to drought, salt and cold stresses. *Plant Mol Biol.* 2015; 88(4-
 752 5):369-385.

753 31. Lu CA, Huang CK, Huang WS, Huang TS, Liu HY, Chen YF. DEAD-
 754 Box RNA Helicase 42 Plays a Critical Role in Pre-mRNA Splicing under
 755 Cold Stress. *Plant Physiol.* 2020; 182(1):255-271.

756 32. Zhang J, Xie Y, Ali B, Ahmed W, Tang Y, Li H. Genome-wide
 757 identification, classification, evolutionary expansion and expression of
 758 RbOH family genes in pepper (*Capsicum annuum* L.). *Trop Plant Biol.* 2021;
 759 14:251-266.

760 33. Min JH, Chung JS, Lee KH, Kim CS. The CONSTANS-like 4
 761 transcription factor, AtCOL4, positively regulates abiotic stress tolerance
 762 through an abscisic acid-dependent manner in Arabidopsis. *J Integr Plant*
 763 *Biol.* 2015; 57(3):313-324.

- 764 34. Zsigmond L, Szepesi A, Tari I, Rigo G, Kiraly A, Szabados L.
765 Overexpression of the mitochondrial *PPR40* gene improves salt tolerance
766 in *Arabidopsis*. *Plant Sci.* 2012; 182:87-93.
- 767 35. Chung E, Park JM, Oh SK, Joung YH, Lee S, Choi D. Molecular and
768 biochemical characterization of the *Capsicum annuum calcium-dependent*
769 *protein kinase 3 (CaCDPK3)* gene induced by abiotic and biotic stresses.
770 *Planta.* 2004; 220(2):286-295.
- 771 36. Young Chae G, Hong WJ, Jeong Jang M, Jung KH, Kim S. Recurrent
772 mutations promote widespread structural and functional divergence of
773 MULE-derived genes in plants. *Nucleic Acids Res.* 2021; 49(20):11765-
774 11777.
- 775 37. Jang MJ, Hong WJ, Park YS, Jung KH, Kim S. Genomic basis of
776 multiphase evolution driving divergent selection of zinc-finger
777 homeodomain genes. *Nucleic Acids Res.* 2023; 51(14):7424-7437.
- 778 38. Choi JW, Choi HH, Park YS, Jang MJ, Kim S. Comparative and
779 expression analyses of AP2/ERF genes reveal copy number expansion and
780 potential functions of ERF genes in Solanaceae. *BMC Plant Biol.* 2023;
781 23(1):48.
- 782 39. Iliopoulos I, Tsoka S, Andrade MA, Enright AJ, Carroll M, Pouillet P,
783 Promponas V, Liakopoulos T, Palaios G, Pasquier C et al. Evaluation of
784 annotation strategies using an entire genome sequence. *Bioinform.* 2003;
785 19(6):717-726.
- 786 40. Xing Z, Geng W, Li C, Sun Y, Wang Y. Comparative genomics of
787 *Lactobacillus kefiranofaciens* ZW3 and related members of *Lactobacillus*.

788 spp reveal adaptations to dairy and gut environments. Sci Rep. 2017;
789 7(1):12827.

790 41. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan
791 R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M et al. The
792 Arabidopsis Information Resource (TAIR): improved gene annotation and
793 new tools. Nucleic Acids Res. 2012; 40(Database issue):D1202-1210.

794 42. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie
795 WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S et al. Improvement
796 of the *Oryza sativa* Nipponbare reference genome using next generation
797 sequence and optical map data. Rice 2013; 6(1):4.

798 43. Kim S, Cheong K, Park J, Kim MS, Kim J, Seo MK, Chae GY, Jang MJ,
799 Mang H, Kwon SH et al. TGFam-Finder: a novel solution for target-gene
800 family annotation in plants. New Phytol. 2020; 227(5):1568-1581.

801 44. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam
802 H, Maslen J, Mitchell A, Nuka G et al. InterProScan 5: genome-scale
803 protein function classification. Bioinform. 2014; 30(9):1236-1240.

804 45. Katoh K, Standley DM. MAFFT multiple sequence alignment
805 software version 7: improvements in performance and usability. Mol Biol
806 Evol. 2013; 30(4):772-780.

807 46. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. TrimAl: a tool
808 for automated alignment trimming in large-scale phylogenetic analyses.
809 Bioinform. 2009; 25(15):1972-1973.

810 47. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a
811 sequence logo generator. Genome Res. 2004; 14(6):1188-1190.

812 48. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O,
813 Madhusoodanan N, Kolesnikov A, Lopez R. Search and sequence analysis
814 tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 2022;
815 50(W1):W276-W279.

816 49. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein
817 secondary structure prediction server. *Nucleic Acids Res.* 2015;
818 43(W1):W389-394.

819 50. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ.
820 Jalview Version 2--a multiple sequence alignment editor and analysis
821 workbench. *Bioinform.* 2009; 25(9):1189-1191.

822 51. Bioinformatics B, Valencia S. OmicsBox-Bioinformatics made easy.
823 *March.* 2019; 3:2019.

824 52. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and
825 analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 2006;
826 34(Web Server issue):W369-373.

827 53. Bailey TL, Gribskov M. Combining evidence using p-values:
828 application to sequence homology searches. *Bioinformatics.* 1998;
829 14(1):48-54.

830 54. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD,
831 von Haeseler A, Lanfear R. IQ-TREE 2: New Models and Efficient Methods
832 for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol.* 2020;
833 37(5):1530-1534.

834 55. Voorrips RE. MapChart: software for the graphical presentation of
835 linkage maps and QTLs. *J Hered.* 2002; 93(1):77-78.

836 56. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W,
837 Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein
838 database search programs. *Nucleic Acids Res.* 1997; 25(17):3389-3402.

839 57. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H,
840 Marler B, Guo H et al. MCScanX: a toolkit for detection and evolutionary
841 analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;
842 40(7):e49.

843 58. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. RIdeogram:
844 drawing SVG graphics to visualize and map genome-wide data on the
845 idiograms. *PeerJ Comput Sci.* 2020; 6:e251.

846 59. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. Gene
847 duplication and evolution in recurring polyploidization-diploidization
848 cycles in plants. *Genome Biol.* 2019; 20(1):38.

849 60. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. KaKs_Calculator:
850 calculating Ka and Ks through model selection and model averaging.
851 *Genom Proteom Bioinform.* 2006; 4(4):259-263.

852 61. Lopez-Ortiz C, Pena-Garcia Y, Natarajan P, Bhandari M, Abburi V,
853 Dutta SK, Yadav L, Stommel J, Nimmakayala P, Reddy UK. The ankyrin
854 repeat gene family in *Capsicum* spp: Genome-wide survey,
855 characterization and gene expression profile. *Sci Rep.* 2020; 10(1):4044.

856 62. Vatansever R, Koc I, Ozyigit, II, Sen U, Uras ME, Anjum NA, Pereira
857 E, Filiz E. Genome-wide identification and expression analysis of sulfate
858 transporter (SULTR) genes in potato (*Solanum tuberosum* L.). *Planta.*
859 2016; 244(6):1167-1183.

860 63. Cai X, Zhang Y, Zhang C, Zhang T, Hu T, Ye J, Zhang J, Wang T, Li
861 H, Ye Z. Genome-wide analysis of plant-specific Dof transcription factor
862 family in tomato. *J Integr Plant Biol.* 2013; 55(6):552-566.

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

866 65. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based
867 genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat*
868 *Biotechnol.* 2019; 37(8):907-915.

869 66. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT,
870 Salzberg SL. StringTie enables improved reconstruction of a
871 transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015; 33(3):290-295.

872 67. Love MI, Huber W, Anders S. Moderated estimation of fold change
873 and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;
874 15(12):550.

875 68. Kumar L, M EF. Mfuzz: a software package for soft clustering of
876 microarray data. *Bioinform.* 2007; 2(1):5-7.

877 69. Langfelder P, Horvath S. WGCNA: an R package for weighted
878 correlation network analysis. *BMC Bioinform.* 2008; 9:559.

879

Figures



Figure 1



Figure 2



Figure 3



Figure 4



Figure 5

Supplementary Files

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

- [20231110B3BMCsupplementaryfigures.pdf](#)
- [20231110B3BMCsupplementarytables.xlsx](#)
- [20231110B3BMCsupplementaryfigures.pdf](#)
- [20231110B3BMCsupplementarytables.xlsx](#)