

# Genome-wide identification of the plant U-box (PUB) gene family and their global expression analysis in tomato (*Solanum lycopersicum*)

Gaofeng Liu<sup>2#</sup>, Qian Hu<sup>1,3#</sup>, Jing Zhang<sup>1</sup>, Entong Li<sup>3</sup>, Xiao Yang<sup>2</sup> and Zixin Zhang<sup>1\*</sup>

<sup>1</sup> Key Laboratory of Bio-resource and Eco-environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, China

<sup>2</sup> Institute of Urban Agriculture, Chinese Academy of Agricultural Sciences, Chengdu National Agricultural Science & Technology Center, Chengdu 610213, China

<sup>3</sup> State Key Laboratory of Crop Genetics and Germplasm Enhancement, the Key laboratory of Landscaping, Ministry of Agriculture, College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China

# These authors contributed equally: Gaofeng Liu, Qian Hu

\* Corresponding author, E-mail: [zixinzhang@scu.edu.cn](mailto:zixinzhang@scu.edu.cn)

## Abstract

Plant ubiquitination plays an important role in protein post-translational changes that occur in a wide variety of eukaryotic organisms and regulate a broad range of biological processes. However, little is known about the evolutionary relationships, gene duplication, and functional evolution of *PUB* genes in tomato. Herein, we explored 48 *PUB* genes in the tomato genome which were further classified into seven major groups based on their sequences and relative genetic similarities with *Arabidopsis*. Gene structure analysis suggested that most of the *PUB*s carry similar features and with slight differentiation in their sources of functions during the process of evolution. Collinear analysis showed a high degree of conservation among *SIPUB*s, with a total of ten pairs identified as collinear with potential interactions with each other and another family. Based on their duplication forms, a total of 23 pairs were chosen for *Ka/Ks* calculation: 13 pairs of dispersed, one tandem, and nine pairs of WGD or segmental. The majority of the gene pairs from the three forms of duplication had a *Ka/Ks* ratio of less than 1.00, indicating purifying selection and reduced divergence after duplication. Only three pairs i.e., *SIPUB16-SIPUB22*, *SIPUB30-SIPUB32*, and *SIPUB36-SIPUB37* were perceived by higher than 1.00 values, signifying positive selection. In addition, three major RNA-seq datasets analyses from the cultivated and its relatives, as well as Micro-Tom (MT), revealed their highly tissue-specific role, with the expression correlation analysis of duplicated pairs indicating highly putative neo-functionalization as compared to functional conservation or sub-functionalization after duplication. *SIPUB10* was screened out and preliminary verified as a key gene in salt and cold stress response with validation of transgenic strains. These results are indicative of their response during the gene duplication and evolution process, while further functional validation step is required to determine their specific role.

**Citation:** Liu G, Hu Q, Zhang J, Li E, Yang X, et al. 2023. Genome-wide identification of the plant U-box (*PUB*) gene family and their global expression analysis in tomato (*Solanum lycopersicum*). *Vegetable Research* 3:16 <https://doi.org/10.48130/VR-2023-0016>

## Introduction

Ubiquitination is an important process of post-translational modification of proteins that occurs in eukaryotic organisms, modulating a broad range of biological processes. There are three types of catalytic activities involved in this process: ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin ligase (E3)<sup>[1]</sup>. Initially, the E1 enzyme initiates the ubiquitin molecule in an ATP-dependent manner, and then these molecules are further progressed with the aid of E2. Finally, E3 covalently attaches E2 to the target protein. Target proteins are then tagged by multiple ubiquitin molecules and can be degraded by the 26S proteasome. E3 ubiquitin ligase is the most effective of these three enzymes in evaluating the specificity of the ubiquitination mechanism by selecting appropriate candidate proteins<sup>[2]</sup>. E3 is the most important factor in ubiquitination pathways, as it is present in a wide range of substrates. In *Arabidopsis*, more than 1,400 genes encode the functional components of the ubiquitination pathway, of which about 90% of genes are related to E3<sup>[3,4]</sup>. Based on the presence of domains and functional mechanism, E3 ubiquitin ligases are classified into single-subunit type, (HECT, RING/U-box ligases)<sup>[4]</sup> and multi-subunit type, such as

SCF (skp1-cullin-F-box), APC (anaphase-promoting complex)<sup>[5]</sup> VBC (VHL-Elongin B-Elongin C)<sup>[6]</sup>, etc. U-box proteins are categorized by the existence of their domain and that is mainly equal to up to 70 amino acids<sup>[7,8]</sup>. Both U-box and Ring domain proteins shared almost identical tertiary structures with slight differences such as the formation of salt-bridges for the stabilization of its structure<sup>[8]</sup>. So far, the presence of U-box proteins has been reported in a variety of organisms like yeast, plants, and animals<sup>[9–11]</sup>. The abundance of *U-box* genes varies from organism to organism. For example, in yeast genome two *U-box* genes have been reported, while 21 has been annotated in the human genome<sup>[8,12,13]</sup>.

To date, the main abundance of U-box protein has been reported in plants<sup>[14]</sup>, and the presence of U-box protein in plants is known as *PUB*s. *PUB* proteins are members of the E3 ubiquitin ligase family<sup>[9]</sup>. *PUB* genes have been reported in a variety of plant species, including, 64 in *Arabidopsis thaliana*, 70 *Oryza sativa*, 33 *Chlamydomonas reinhardtii*, 101 *Brassica rapa*, and 125 *Glycine mix*<sup>[15–19]</sup>, suggesting that *PUB* genes are widely dispersed in plants. Yet, a comprehensive genome-wide analysis on their identification and expression profiling under multiple organs-tissues along with abiotic stresses still lacking in tomato.

The presence of a U-box domain in PUB proteins is typically used to classify them. As a result, they're divided into subgroups<sup>[9,16,20–22]</sup>. PUB proteins with ARM domains and PUB proteins with Kinase domains are the two most common types found in plants. However, several other species-specific groups have also been reported particularly in *Arabidopsis* and rice, such as the occurrence of two PUB-MIF4G proteins and one PUB-TPR-Kinase<sup>[9,15,16]</sup>, respectively.

Increasing evidence proposes that PUB proteins play critical roles in various plant developmental and physiological processes, such as abiotic stress responses. *CaPUB1*, a U-box E3 protein found in water-stressed hot pepper plants, and its overexpression was found to be more susceptible to water stress and mild salinity<sup>[23]</sup>. In *Arabidopsis*, *AtPUB22* and *AtPUB23* proteins act as negative regulators against drought tolerance by RPN12a ubiquitination<sup>[24]</sup>. U-box E3 protein *AtPUB19* is induced by drought, salt, cold, and ABA in *Arabidopsis*. Down-regulating the *AtPUB19* expression resulted in hypersensitivity to ABA, enhancing ABA-induced stomata closing, and drought tolerance, while overexpression of *AtPUB19* exhibited the reverse phenotypes<sup>[25]</sup>. Furthermore, *Arabidopsis AtPUB44/SAUL1* E3 ubiquitin ligase inhibits premature leaf senescence by decreasing the ABA level, and *AtPUB43* mutants displayed an increased seed germination rate compared with the wild type<sup>[24]</sup> at inhibitory concentrations of ABA<sup>[26,27]</sup>. In rice, the overexpression of *OsPUB15* significantly regulated the drought and salt resistance positively, and further reduced the intracellular oxidative stress<sup>[28]</sup>.

Tomato (*Solanum lycopersicum*) is one of the vital economic crops and ranks fourth in global vegetable production worldwide. It is also a model plant species for fleshy fruit development and a rich source of nutrients<sup>[29]</sup>. The scale of production of tomato is however mainly constrained due to a series of various biotic and abiotic factors<sup>[30]</sup>. Hence, the identification of the gene family is a useful strategy for improving tomato development and crop production improvement. To the best of our knowledge, there are no reports for genome-wide investigation and wide-ranging transcriptional profiling of *PUB* genes in tomato to date. While the availability of the tomato whole-genome sequence<sup>[29]</sup> provides valuable resources for an in-depth understanding of gene functioning and evolution. In the present study, we systematically studied the *PUB* gene family in tomatoes such as basic properties, chromosomal localization, collinear analysis, synonymous-non synonymous mutation, gene duplications, organ-tissue expression profiles, abiotic stress, and functional evolution. These results will facilitate future research for understanding the *PUB* gene in functional genomic studies, as well as their utilization under abiotic stress in this important crop.

## Results

### Basic information of the PUB gene family in tomato

In this study, a total of 48 *PUB* genes were explored from the tomato genome and were renamed as *SIPUB1-SIPUB48* based on their chromosomal positions from Chromosome 1–12 (Chr1–12). For these genes, we explored various gene properties including, chromosomal distribution, coding sequence (CDS) length (bp), and five different types of protein properties which include, protein length<sup>[26]</sup>, molecular weight (MW) kDa, isoelectric point (PIs), grand average of hydrophobicity (GRAVY),

and subcellular prediction for each of *PUBs* protein (Dataset 1). The *PUB* genes in tomato were randomly distributed on the 1–12 chromosomes of the tomato genome, except none of the genes were detected on Chr8 and Chr10. The CDS length of *SIPUB* ranged from 831–3,141 bp (*SIPUB32-SIPUB27*), the protein length ranged from 276–1,046 aa (*SIPUB32-SIPUB27*). Similarly, the MW ranged from 31.77–118.442 kDa (*SIPUB32-SIPUB14*), the PIs 5.2–9.12 (*SIPUB15-SIPUB2*), and GRAVY –0.528 (*SIPUB32*) to 0.137 (*SIPUB22*). The observed variability among *SIPUB* genes specifically for GRAVY analysis revealed most of them were hydrophilic and only four genes i.e., *SIPUB11*, *SIPUB7*, *SIPUB25*, and *SIPUB22* were hydrophobic having positive values. Also, the subcellular localization analysis predicted that the maximum *PUB* genes were involved in the nucleus, cytoplasm, chloroplast, plasma membrane, vacuole, and endoplasmic reticulum.

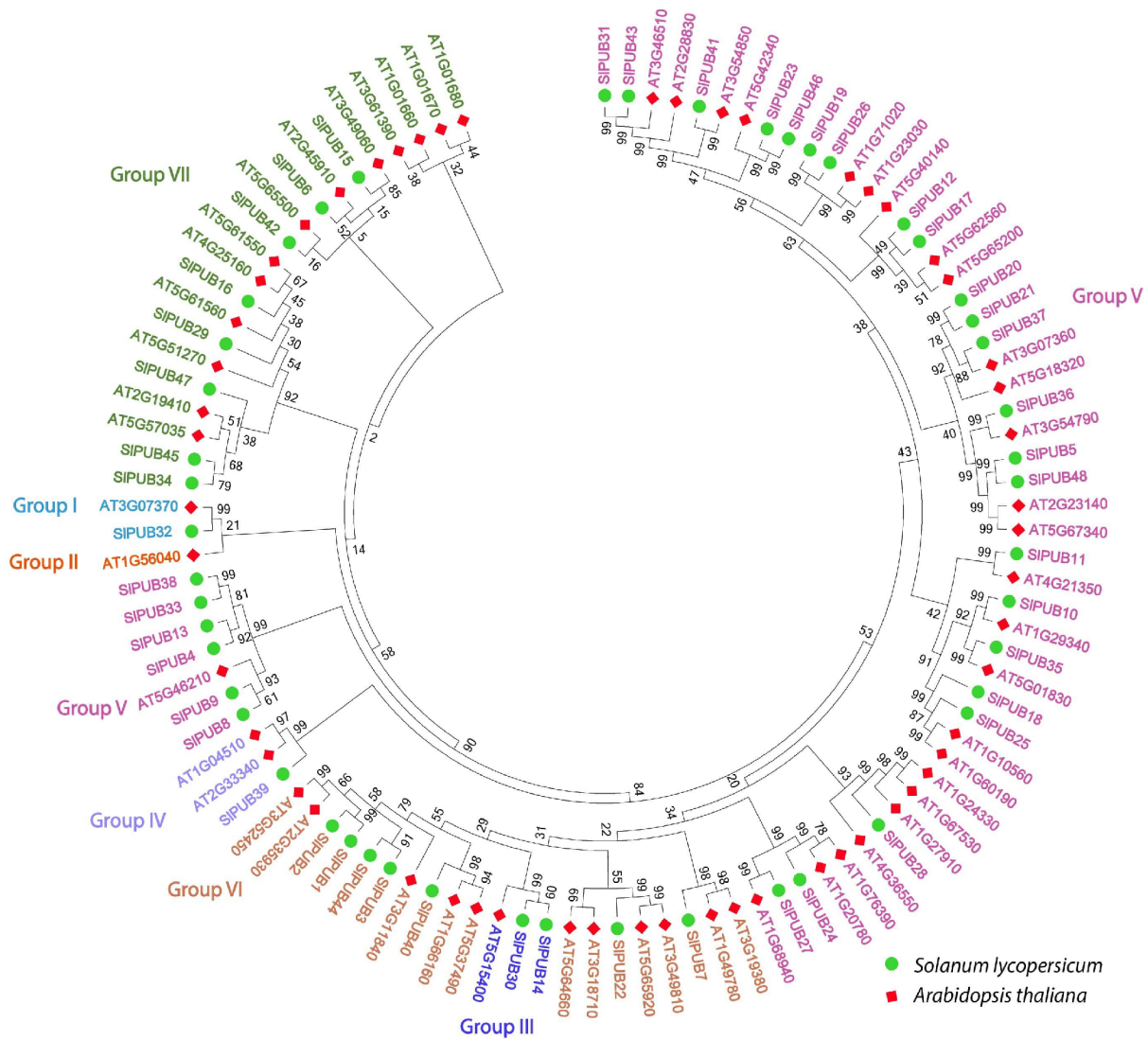
### Phylogenetic characterization, motif and gene structure analysis of PUB genes in tomato

To explore the evolutionary relationships, we constructed a neighbor-joining phylogenetic tree by utilizing the protein sequences i.e., 48 from tomato and 64 from *Arabidopsis*. The results revealed that *PUB* genes were clustered into seven subgroups (Fig. 1) and show the uneven distribution of *SIPUB* when compared with *AtPUB*. Notably, group V was observed with the most number of genes (29), followed by group VII and VI with eight and six genes respectively. Based on the group-wise comparison of *SIPUB*, group V and VI were scattered in the phylogenetic tree and showed relatively close genetic relationships with *Arabidopsis*.

To better understand the conservation and diversification among *PUBs*, we analyzed the conserved motifs and gene structure. The motifs analysis was explored by the MEME program, and motifs ranged from 1–10 for *SIPUBs*. The results showed that several motifs were widely dispersed, however, motifs 1 and 4 were highly conserved amongst almost all the members of *PUBs*, suggesting their common evolution and the rest of *PUB* members indicated slightly structural diversification among the *SIPUB* gene family (Fig. 2a). Again, the MEME server was explored for the *PUB* protein LOGOS of these motifs. Motif6 was found with the highest number (100) of consensus sequences, while motif4 with a fewest number (50) of sequences (Supplemental Fig. S1). The gene structure analysis showed similar patterns to the motif composition, most of the *PUBs* generally exhibited similar gene structure in terms of their lengths (Fig. 2b). However, two of them showed contrasting results such as *SIPUB8* and *SIPUB9*. These results suggested that most of the *PUBs* carry similar features and slight differentiation in their sources of functions during the process of evolution.

### Gene collinearity, duplication, and Ka/Ks analysis of PUB

The chromosomal mapping of 48 *PUB* genes in tomato from Chr1-Chr12 was drawn by using TBtools software. All the *PUB* genes were distributed across the tomato genome except Chr8 and Chr10 as shown in Fig. 3. The distribution of the number of genes varies from chromosome such as the maximum amount of genes (7) was identified on Chr1 and Chr2, followed by Chr4 possessing six genes, while five of each gene were located on Chr5, Chr9 and Chr11. Thus, these results unveiled their uneven patterns among *PUB* members in the tomato genome. Collinearity relationships between *PUB* genes were also



**Fig. 1** Phylogenetic relationships of the PUB genes between tomato and *Arabidopsis*. The seven different subgroups of PUBs are indicated in different colors.

demonstrated by drawing circos using the TBtools software. Among SIPUB, high conservation of nucleotide was detected and a total of 11 pairs were detected as collinear with possible interaction with each other and another family. These collinear pairs include: *SIPUB6-SIPUB15*, *Solyc01g099290-SIPUB38*, *SIPUB8-SIPUB9*, *SIPUB12-SIPUB17*, *SIPUB18-SIPUB25*, *SIPUB19-SIPUB26*, *SIPUB23-SIPUB46*, *Solyc06g008710-SIPUB38*, *SIPUB31-SIPUB43*, *SIPUB34-SIPUB45*, and *SIPUB41-SIPUB43*. The two unknown genes (*Solyc01g099290* and *Solyc06g008710*) were identified as cullins genes based on their domain analysis, and these genes are recognized for their vital role in the process of ubiquitination in combination with diverse cellular processes<sup>[31]</sup>. Additionally, the gene duplication investigation displayed that the maximum of the genes (26) were dispersed type followed by whole-genome duplication (WGD) or segmental (19), and tandem (3) as shown in Fig. 3.

The synonymous with non-synonymous mutation analysis was also performed among the different types of gene duplications in tomato i.e., dispersed, tandem, and WGD or segmental. In the process of evolution, genes are typically

associated with diverse kinds of selection pressure that include, purifying selection ( $Ka/Ks < 1$ ), positive selection ( $Ka/Ks > 1$ ), and neutral selection ( $Ka/Ks = 1$ )<sup>[32]</sup>. A possible combination of 23 pairs was selected for  $Ka/Ks$  calculation based on their duplication types i.e., 13 pairs of dispersed, one tandem, and nine pairs of WGD or segmental as described in Table 1. The results have shown that the maximum of gene pairs having a  $Ka/Ks$  ratio of less than 1.00, from three different types of duplication. These analyses signifying that purifying selection and reducing divergence occurred among them. Only three pairs of dispersed genes that include, *SIPUB16-SIPUB22*, *SIPUB30-SIPUB32*, and *SIPUB36-SIPUB37* were perceived with greater than 1.00 values, suggesting positive selection.

**Transcriptional analysis of PUB genes in multiple tissues and developmental stages in tomato**

Three major Illumina RNA-seq datasets of various multiple tissues were utilized from cultivated tomato (*Solanum lycopersicum* cv. Heinz) and the wild relative *Solanum pimpinellifolium*<sup>[33,34]</sup>, and the Micro-Tom (MT) (unpublished data). The ten



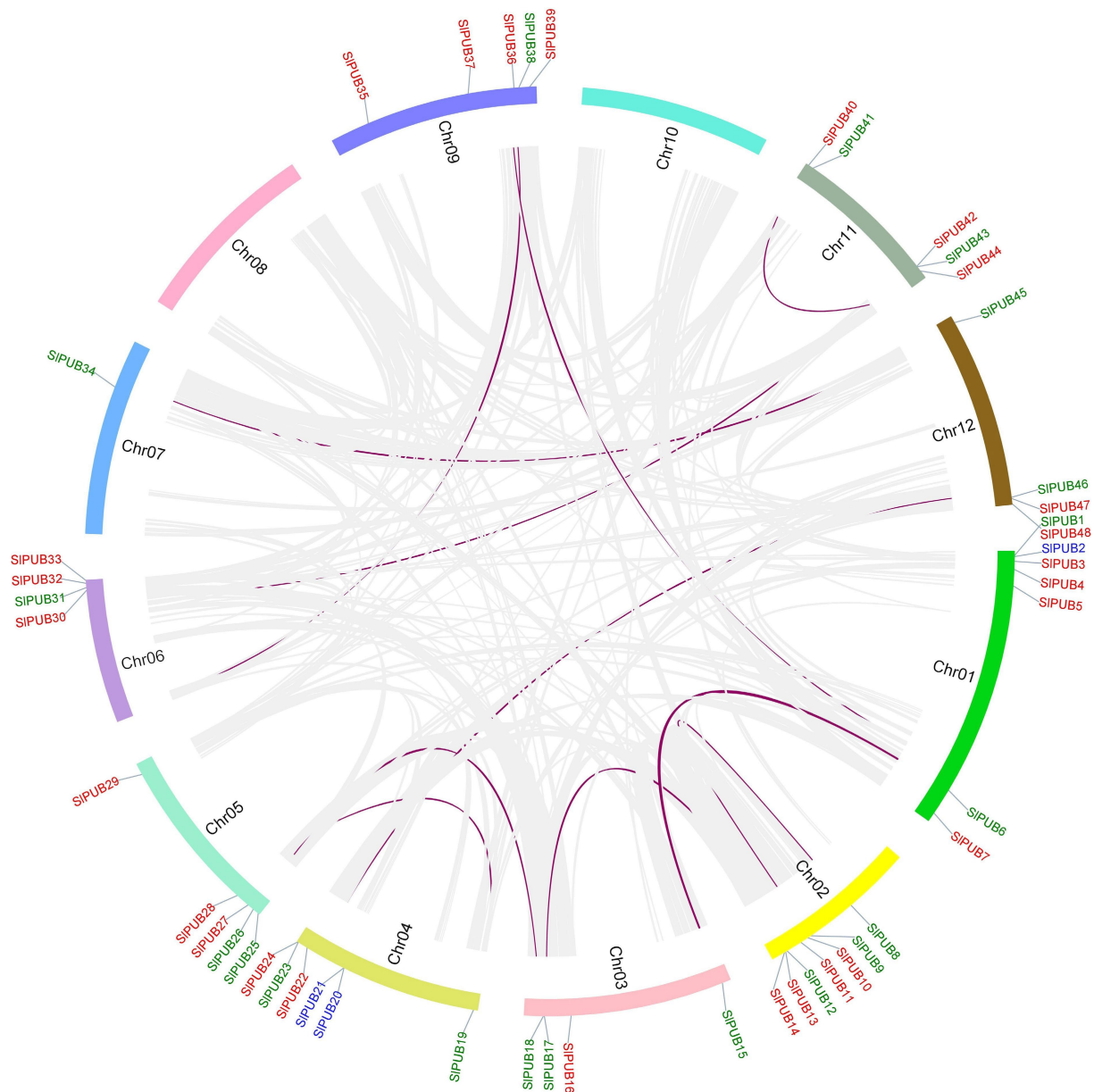
**Fig. 2** (a) Motif structures and (b) gene structures of PUBs in tomato. The different motifs (Motif 1 – Motif 10) are displayed in different colors. The gene structures of PUBs are based on the coding sequences (CDS) and untranslated region (UTR) which are shown in yellow and green.

tissues from the cultivated tomato include bud, flower, leaf, roots, 1 cm, 2 cm, 3 cm, mature green, breaker, and breaker+ 10 fruits. From the wild relative, eight tissues and organs, which included anthesis flower (ODPA), 10 d post-anthesis (DPA) fruit, 20 DPA fruit, ripening fruit (33 DPA), mature leaves (ML), whole root (WR), young flower buds (YFB), and young leaves<sup>[30]</sup>, were selected for analysis. The data from our unpublished RNA-seq data of the model variety Micro-Tom (MT) was also utilized for four different tissues at various stages such as root 30, 45, and 85 days post germination (DPG), stem 30, 45, and 85 DPG, leaf 30, 45, and 85 DPG, and flower 30 and 45 DPG. Several developmental stages like, 10 DPA (55 DPG), 20 DPA (65 DPG),

immature green (IMG) at 75 DPG, mature green (MG) at 80 DPG, Br (85 DPG), Br3 (88 DPG), Br7 (92 DPG), Br10 (95 DPG) and Br15 (100 DPG).

To represent the tissue-specific expression of 48 *SIPUB* genes, a heatmap was generated on RPKM values. Results of Heinz and MT data showed an almost identical expression pattern and slight inconsistency with its wild relative such as *SIPUB4-SIPUB19*, *SIPUB22*, *SIPUB24*, *SIPUB26*, *SIPUB28*, *SIPUB30-SIPUB39*, *SIPUB41-43*, *SIPUB47*, and *SIPUB48* significantly expressed in both of them (Figs 4a, 5a & 5b) and Dataset 2 and 3. While, the wild relative showed that *SIPUB4-SIPUB17*, *SIPUB19*, *SIPUB22*, *SIPUB24*, *SIPUB26*, *SIPUB28*, *SIPUB30-SIPUB34*, *SIPUB36-SIPUB39*,





**Fig. 3** Chromosomal locations of the PUB genes in the tomato genome from Chr1-Chr12. The collinear genes are presented inside the circle in purple. The different types of duplication such as dispersed, segmental, and tandem are marked in red, green, and blue.

*SIPUB41-43*, *SIPUB47* and *SIPUB48* were significantly expressed entirely in the tissues, indicating their foremost contribution in tissue-specific response and development (Fig. 4b) and Dataset 4. The remaining *PUB* genes showed either very low, moderate, and even a few of them were not expressed in or several selected tissues, suggesting their slightly partial response in tomato.

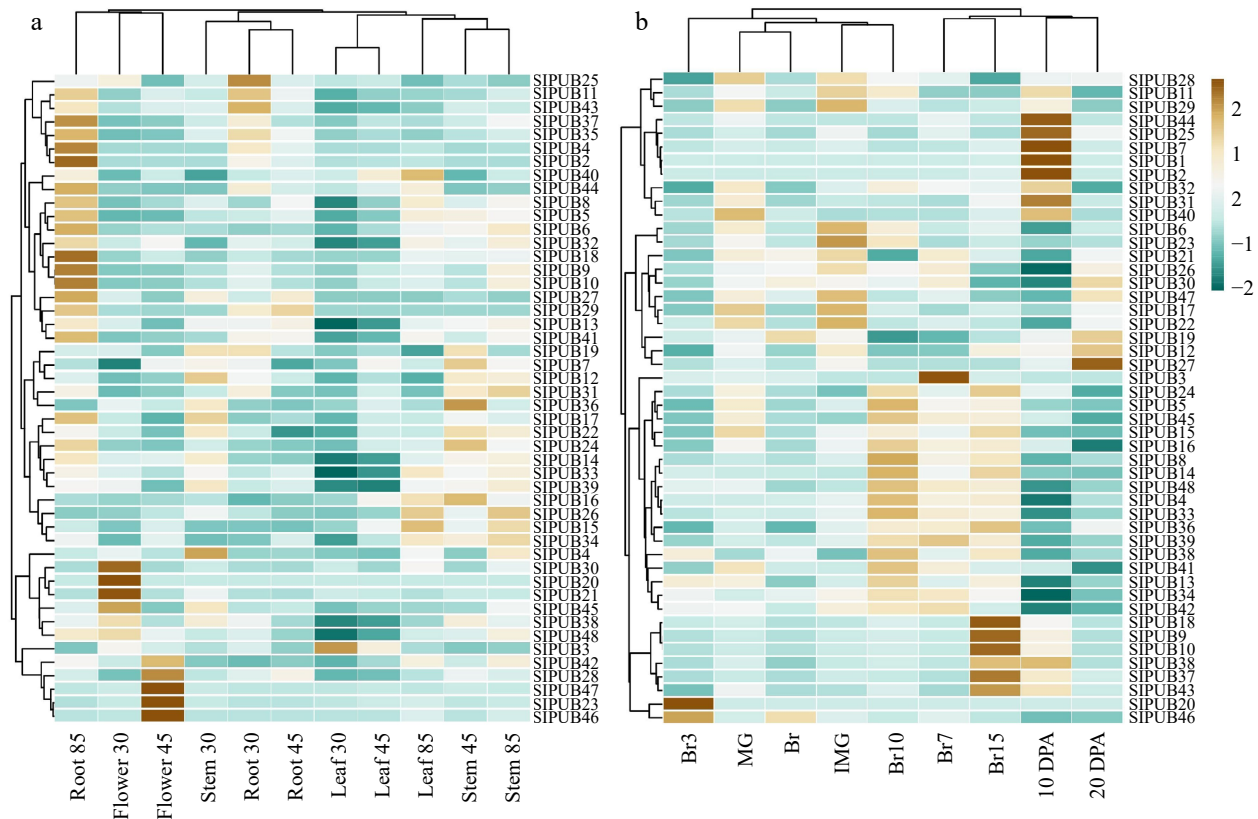
#### qRT-PCR analysis of *PUB* genes in different tissues, and their response to salt and cold stress

The expression profiling of 30 *PUB* genes was randomly selected and then quantified by qRT-PCR in five different tissues including root, stem, leaf, flower bud, and flower (Fig. 6a). As shown in Fig. 6c, the same genes were tested in response to salt and cold stress, as well as control (CK). In most tissues, the majority of genes were highly expressed, and only a

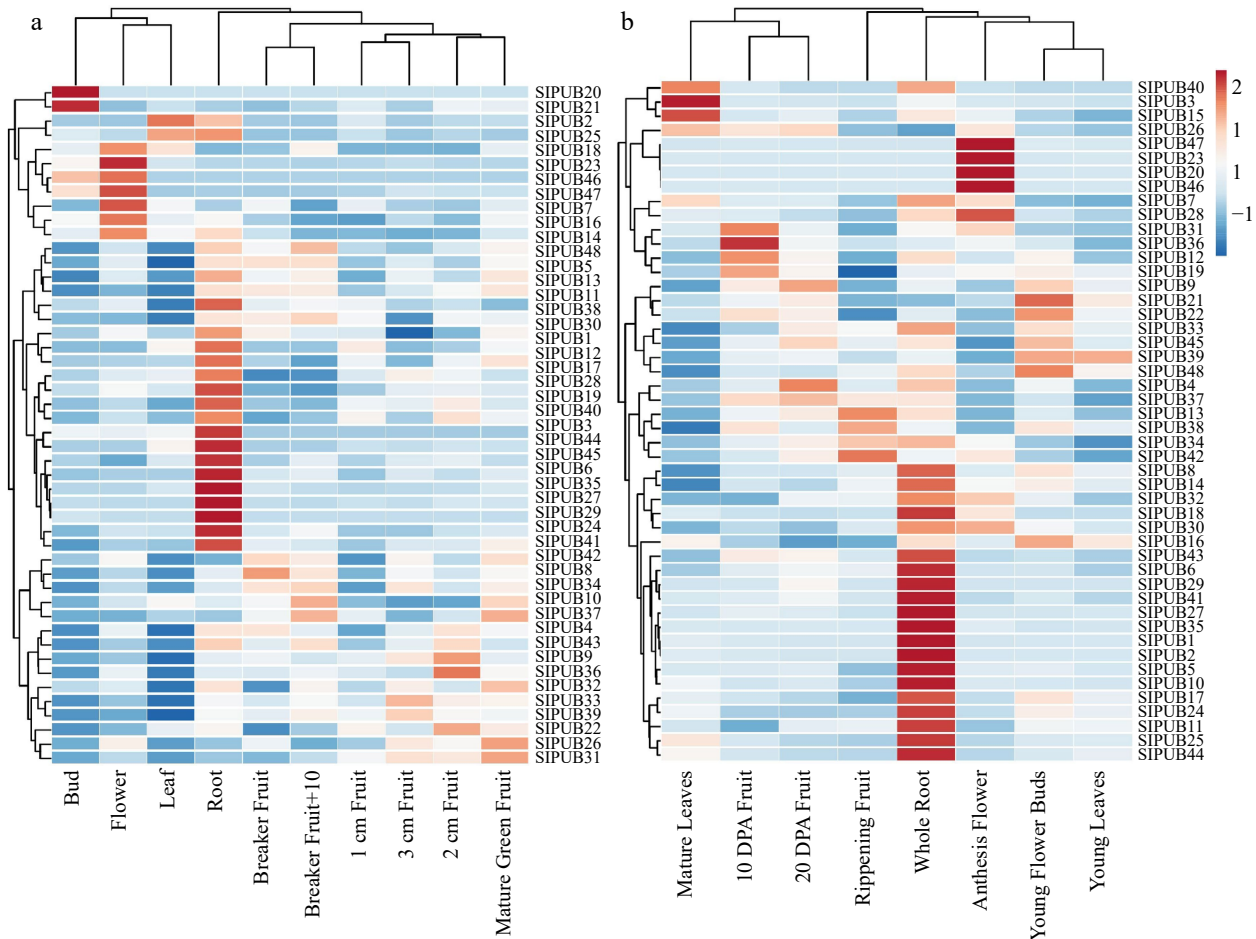
few genes were down-regulated such as *SIPUB3*, *SIPUB23*, *SIPUB25*, and *SIPUB47*. Similarly, findings showed that most *PUB* genes expressed distinctly in response to both stresses, implying that these genes can play a variable role in tomato plant resistance. When it came to salt stress, for example, most of the genes were significantly up-regulated, especially *SIPUB10*, *SIPUB33*, *SIPUB30*, *SIPUB43*, *SIPUB5*, *SIPUB13* and *SIPUB31*. When compared to low temperature, *SIPUB10*, *SIPUB43*, *SIPUB13* and *SIPUB31* were significantly up-regulated. Several genes, including *SIPUB3*, *SIPUB16*, and *SIPUB34*, were down-regulated in both stress conditions. In response to both treatments, the rest of the genes showed either poor or moderate expression. Additionally, we also performed principal component analysis (PCA) to gain deeper insights into their involvement in different tissues and against abiotic stress. For different tissues, PCA analysis of *PUBs* transcripts showed

**Table 1.** Gene duplications of *PUB* genes in tomato with outlier *Ka/Ks* values.

Gene 1	Gene 2	<i>Ks</i>	<i>Ka</i>	<i>Ka/Ks</i>	Selection pressure	Gene duplications
<i>SIPUB3</i>	<i>SIPUB4</i>	0.74	0.59	0.80	Dispersed	Purifying
<i>SIPUB5</i>	<i>SIPUB7</i>	0.68	0.62	0.92	Dispersed	Purifying
<i>SIPUB10</i>	<i>SIPUB11</i>	0.65	0.59	0.90	Dispersed	Purifying
<i>SIPUB13</i>	<i>SIPUB14</i>	0.80	0.59	0.73	Dispersed	Purifying
<i>SIPUB16</i>	<i>SIPUB22</i>	0.58	0.61	1.06	Dispersed	Positive
<i>SIPUB24</i>	<i>SIPUB27</i>	0.70	0.68	0.98	Dispersed	Purifying
<i>SIPUB28</i>	<i>SIPUB29</i>	0.88	0.63	0.71	Dispersed	Purifying
<i>SIPUB30</i>	<i>SIPUB32</i>	0.44	0.45	1.03	Dispersed	Positive
<i>SIPUB33</i>	<i>SIPUB35</i>	0.88	0.62	0.70	Dispersed	Purifying
<i>SIPUB36</i>	<i>SIPUB37</i>	0.46	0.67	1.45	Dispersed	Positive
<i>SIPUB39</i>	<i>SIPUB40</i>	0.78	0.65	0.83	Dispersed	Purifying
<i>SIPUB42</i>	<i>SIPUB44</i>	0.72	0.53	0.73	Dispersed	Purifying
<i>SIPUB47</i>	<i>SIPUB48</i>	0.75	0.61	0.82	Dispersed	Purifying
<i>SIPUB20</i>	<i>SIPUB21</i>	0.22	0.11	0.52	Tandem	Purifying
<i>SIPUB1</i>	<i>SIPUB6</i>	0.63	0.46	0.74	WGD or Segmental	Purifying
<i>SIPUB8</i>	<i>SIPUB9</i>	0.65	0.05	0.07	WGD or Segmental	Purifying
<i>SIPUB12</i>	<i>SIPUB15</i>	0.74	0.61	0.82	WGD or Segmental	Purifying
<i>SIPUB17</i>	<i>SIPUB18</i>	0.76	0.65	0.86	WGD or Segmental	Purifying
<i>SIPUB19</i>	<i>SIPUB23</i>	1.08	0.55	0.51	WGD or Segmental	Purifying
<i>SIPUB25</i>	<i>SIPUB26</i>	0.76	0.66	0.87	WGD or Segmental	Purifying
<i>SIPUB31</i>	<i>SIPUB34</i>	0.77	0.62	0.80	WGD or Segmental	Purifying
<i>SIPUB38</i>	<i>SIPUB41</i>	0.79	0.61	0.77	WGD or Segmental	Purifying
<i>SIPUB43</i>	<i>SIPUB45</i>	0.80	0.64	0.80	WGD or Segmental	Purifying



**Fig. 4** Expression profiling of the 48 differentially expressed genes in multiple tissues based on RPKM values, including (a) *Solanum lycopersicum* cv. Heinz and (b) the wild relative *Solanum pimpinellifolium*. The brown bars represent up-regulated genes and dark blue bars represent down regulation.



**Fig. 5** Expression profiling of the 48 differentially expressed PUB genes from MT data based on RPKM values, including (a) different developmental stages and (b) different organs. Red bars represent up-regulated genes and blue bars represent down-regulation.

39.37% variation in PC1 and 31.06% variation in PC2, with 67.03% in PC1 accounting for 32.07% of total variation in PC2 for stress conditions (Fig. 6b).

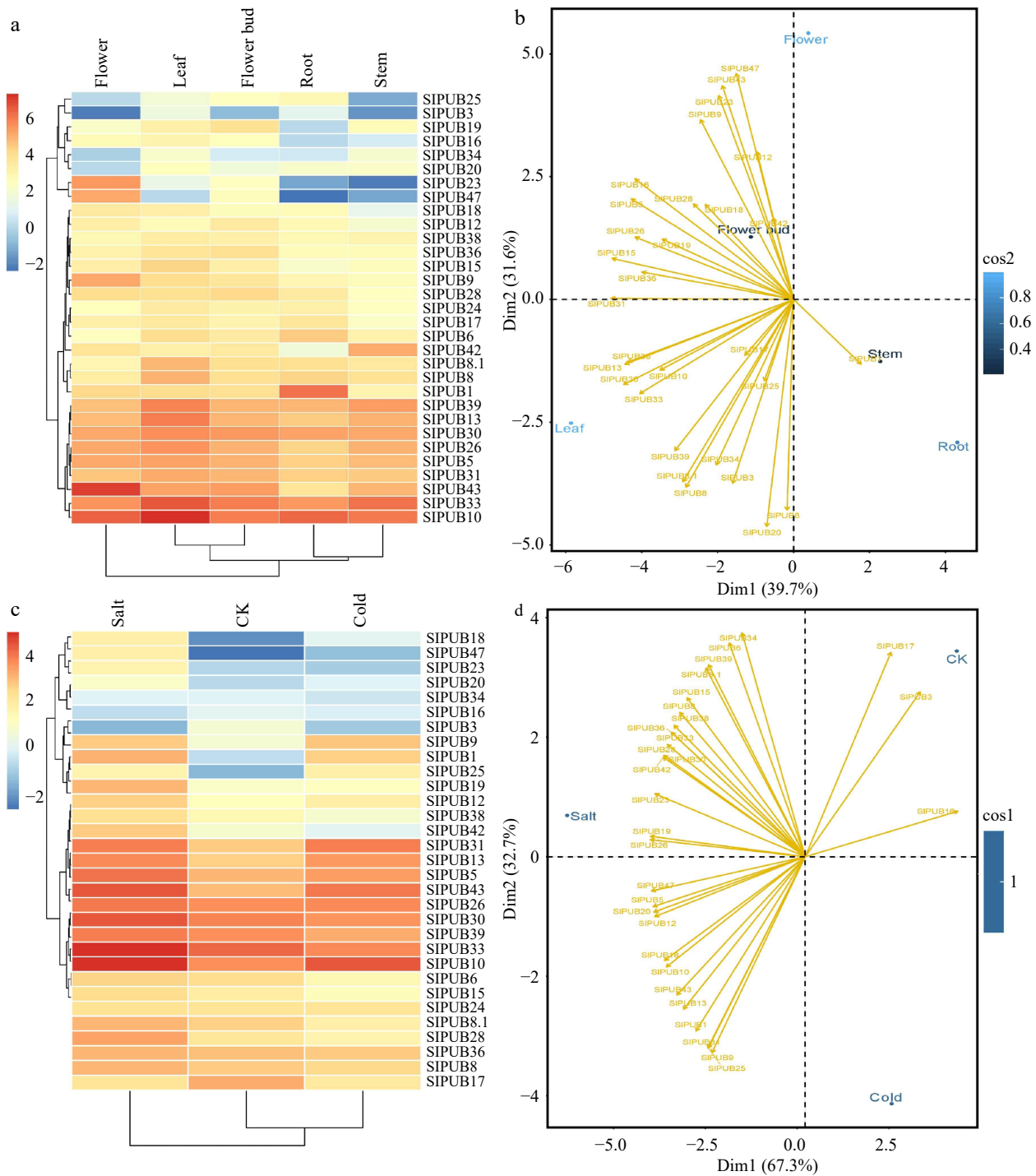
### **SIPUB10 can effectively regulate tomato's tolerance to low temperature and salt stress**

Due to the relatively high expression of *SIPUB10* in roots and leaves, and the significant increase in expression after low temperature and salt treatment, we conducted further research on this gene. The localization of protoplasts shown that *SIPUB10* were expressed in the nucleus and cytoplasm (Fig. 7a). In the leaves of mock-treated plants, there was no genotypic effect on the activity of any of reactive oxygen species (SOD) or Peroxidase (POD) (Fig. 7b). This was not the case when the plants were challenged by salt and cold stress: here, the genotypic ranking for SOD and POD was OX>WT>RNAi when treated by salt or low temperature. In the salinity treatment, the activity of the two enzymes were higher than CK. The determination was consistent with the relative expression of *SIPUB10* in transgenic lines and wild-type<sup>[24]</sup> (Fig. 7c). After salt or cold treatment, the overexpression lines of *SIPUB10* did not show any difference from those before treatment, while RNAi lines showed more severe wilting than the WT (Fig. 7d), which suggested that *SIPUB10* can enhance the salt tolerance and cold resistance of plants.

## **Discussion**

*PUB* genes have been found in a variety of plants, including 64 in *Arabidopsis thaliana*, 70 *Oryza sativa*, 33 *Chlamydomonas reinhardtii*, 101 *Brassica rapa*, 125 *Glycine mix*, and 93, 96, 185, 208, from four different cotton sequenced species including, *G. raimondii*, *G. arboreum*, *G. hirsutum*, and *G. barbadense*<sup>[15–19,35]</sup>. However, a comprehensive genome-wide analysis of their identification, expression profiling, gene duplication, and evolutionary analysis is still lacking in tomato. As a result, we used multiple bioinformatics methods to investigate the *PUB* gene for their physicochemical properties, phylogenetic relationships, gene structure analysis, chromosomal localization, gene duplications, gene collinearity analysis, expression profiling of multiple tissues, stress conditions, and developmental stages. Additionally, we explored the mode and tempo of *SIPUB* duplicated gene pairs based on their transcriptional profiling to determine their functional evolution.

The comparative phylogenetic analysis between tomato (*SIPUB*) and *Arabidopsis* (*AtPUB*), suggested an almost identical and relatively close genetic relationship. Following the previously reported studies<sup>[18,35]</sup> on *PUB* with slight modifications, the *PUB* genes were divided into seven different groups according to the phylogenetic analysis. *PUB* genes were found to have high conservation and diversification, according to gene structure analysis based on coding and protein



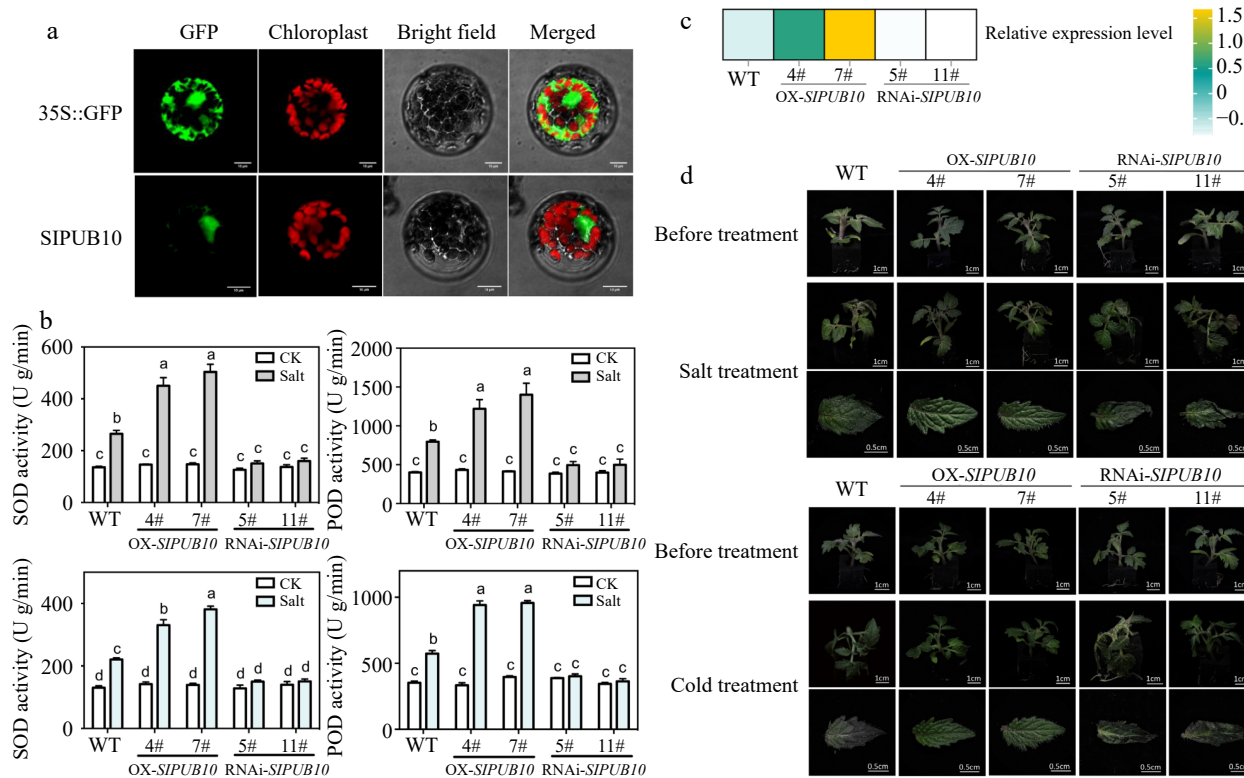
**Fig. 6** Relative expressions of SIPUBs (a) and (c) in five different tissues including root, stem, leaf, flower, and flower bud, and two stress conditions i.e., salt (200 mM) and cold stress (4 °C). The principal component analysis for (b) different tissues and (d) stress conditions.

sequences. For instance, several motifs were widely scattered, but motifs 1 and 4 were strongly conserved almost across all PUB members, implying their common evolution, whereas the rest of the PUB members suggested slight structural diversification among SIPUBs. While the gene structure analysis revealed nearly identical patterns with motif composition, two members, *SIPUB8* and *SIPUB9*, showed contrasting results. These findings indicated that most PUBs had similar characteristics and only small variations in their origins of functions as they evolved. During the process of evolution, structural

diversification is vital for their important role in the multi-gene family<sup>[36]</sup>.

Due to gene duplications, genomic evolution and functional importance are likely to be the most popular ingredients of variations<sup>[37]</sup>. The observed special genomic functions and evolutionary innovation are also due to gene duplications<sup>[38,39]</sup>. In this study, we discovered three forms of *PUB* gene duplication in the tomato genome: 26 dispersed, 19 WGD, and three tandem. Two types of duplication events, segmental and tandem, are the most important contributors to gene family





**Fig. 7** Gene function verification of SIPUB10. (a) The proteins' subcellular localization in Arabidopsis protoplasts. (b) The activity of SOD and POD in leaves of wild-type (WT) and transgenic tomato lines after salt (200 mM NaCl) and cold treatment (4 °C) for 4 h. The P values indicate the results from pairwise comparisons of one-way ANOVA tests. Different letters represent a significant difference at  $P < 0.05$ . (c) The relative expression level of SIPUB10 in transgenic lines. (d) Phenotypes of WT and transgenic lines after 48 h of salt (200 mM NaCl) and cold treatment (4 °C).

expansion during evolution<sup>[40,41]</sup>. Our findings showed, however, that both dispersed and WGD greatly enhanced and extended the *PUB* gene family in tomato. We also measured the rate of synonymous to non-synonymous mutation among the three forms of gene duplications, namely dispersed, tandem, and WGD, in light of the significance of gene duplication. It is a well-known fact that genes are typically associated with various types of selection pressure including purifying selection ( $Ka/Ks < 1$ ), positive selection ( $Ka/Ks > 1$ ), and neutral selection ( $Ka/Ks = 1$ )<sup>[32]</sup>. In our study, we make 23 pairs of possible combinations for  $Ka/Ks$  calculation based on their duplication types i.e., 13 pairs of dispersed, one tandem, and nine pairs of WGD. We discovered that 20 pairs of duplications from various types had a  $Ka/Ks$  ratio of less than 1.00, indicating purifying selection and reduced divergence after gene duplications. Just three scattered gene pairs, *SIPUB16-SIPUB22*, *SIPUB30-SIPUB32*, and *SIPUB36-SIPUB37*, had  $Ka/Ks$  standards greater than 1.00, indicating positive selection.

The angiosperm evolution is generally governed by several events such as polyploidization through the process of whole-genome duplication, followed by diploidization, which is mostly associated with considerable gene loss<sup>[42]</sup>. Several polyploidy events in flowering plants have been identified so far in different flowering plant lineages. After the event of polyploidy, the gene duplication mainly experiences several fates over the process of evolutionary time. For example, most likely the gene loss among the pairs of paralogous or become pseudogene<sup>[43]</sup>. However, the gene duplication pairs can survive for many years through the process of natural selection

and numerous functional importance<sup>[44,45]</sup>. A category of gene products that include several factors like, transcription, ribosomal protein, or protein kinases are specifically reserved<sup>[46,47]</sup>. For example, the interaction of genes with one of the others in the group tends to be highly retained due to gene dosage hypothesis<sup>[48]</sup>. In summary, the gene duplication tends to undergo several evolutionary fates i.e., functional conservation, sub-functionalization, neo-functionalization, or pseudogenization<sup>[43]</sup>. To determine the functional evolution of SIPUB duplicated gene pairs, we explored its different modes based on gene transcriptional analysis that are generally used for predicting the diverse fates. Thus, we took a major benefit from the two RNA-seq datasets by calculating the Pearson correlation coefficient ( $r$ ) analysis of the same duplicated gene pairs which were used for  $Ka/Ks$  ratio, across various tissues. Thus, based on predicted expression analysis of the *PUB* genes in tomato after the process of gene duplication might be likely involved in functional or sub-functionalization conservation (Dataset 5). In short, the possible predicted functions of *PUB* genes in tomato were largely enhanced and expanded through the experience of gene duplications. However, further functional validation is required to determine the fates of gene pairs either through the consequences of structural variation or epigenetic modification.

Gene expression profiling is useful to the approach for predicting its gene functions. In this work, we have utilized the three major datasets of Illumina RNA-seq dataset of various multiple tissues from cultivated tomato (*Solanum lycopersicum* cv. Heinz), the wild relative *Solanum pimpinellifolium*<sup>[33,34]</sup> and

MT. Both Heinz and our MT data showed an almost identical expression pattern and slight inconsistency with it wild relative such as, *SIPUB4-SIPUB19*, *SIPUB22*, *SIPUB24*, *SIPUB26*, *SIPUB28*, *SIPUB30-SIPUB39*, *SIPUB41-43*, *SIPUB47*, and *SIPUB48* significantly expressed in both. While, the wild relative showed that *SIPUB4-SIPUB17*, *SIPUB19*, *SIPUB22*, *SIPUB24*, *SIPUB26*, *SIPUB28*, *SIPUB30-SIPUB34*, *SIPUB36-SIPUB39*, *SIPUB41-43*, *SIPUB47*, and *SIPUB48* displayed high expression among all tissues, involving their foremost contribution in tomato growth and developmental based on their high tissue-specific response. The remaining *PUB* genes displayed a very modest or weak transcriptional response among the different tissues and organs used in our study, which further suggested their slightly restricted role in tomato development. Further, the RT-PCR analysis of 30 *PUB* genes in five different tissues and their response to salt and low temperature was performed. These results showed most of the genes were significantly up-regulated in the majority of the tissues, and response to salt stress when compared to low temperature. Thus, these analyses supporting the hypothesis that *PUB* genes play a vital role in tissue-specific patterns and various environmental stimuli. While, together with the multiple tissue-specific patterns and abiotic stresses, these analyses provided basic resources for the examination of tomato development and stress resistance.

In summary, 48 *PUB* genes were comprehensively discovered and presented in the tomato genome. Based on evolutionary phylogenetic relationships and high sequence similarities with *Arabidopsis*, the *PUB* genes in tomato were classified into seven groups. Gene duplication analysis suggested that the *PUB* gene family in tomato were mainly expanded as a consequence of dispersed and segmental duplications. The multiple transcriptional profiling of RNA-seq and RT-PCR analyses provided a valuable clue to their tissue-specific abundance, developmental stages and response to salt and low temperature. The expression correlation analysis further suggested that putative neo-functionalization largely was observed among the SIPUB of at least one of the duplicated pairs compared to functional conservation or sub-functionalization. These results will greatly facilitate the critical role of *PUB* genes in tomato and other crop species. Moreover, *SIPUB10* was screened out and verified as a key gene in salt and cold stress response.

## Materials and methods

### Genome-wide investigation of *PUB* genes families in tomato

The *PUB* genes were identified using the 64 reference sequences of *Arabidopsis* (TAIR10) against tomato (version 3) genome by BLASTP search at 1E-100 with the help of BioEdit tools. Both of their genomes sequences were retrieved from online sources such as phytozome (version 12.1) (<https://phytozome.jgi.doe.gov/pz/portal.html>)<sup>[49]</sup> and the *Arabidopsis* information resource TAIR ([www.arabidopsis.org](http://www.arabidopsis.org))<sup>[50]</sup>. Domain analysis for each of SIPUB was further examined by two major servers such as the NCBI-Conserved Domain (version 3.18) ([www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi](http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi))<sup>[35]</sup> and SMART (version 8.0) (<http://smart.embl-heidelberg.de/>)<sup>[51]</sup>. Those *PUB* sequences with error either in length (< 100 aa) or absence of their domain were removed before any further analysis.

### Phylogenetic analysis of the *PUB* gene family

Phylogenetic analysis was performed with the help of Molecular Evolutionary Genetics Analysis (MEGA: version 7.0)<sup>[52]</sup>, firstly the multiple sequences alignment (MSA) of *PUB* genes for both tomato and *Arabidopsis* genome was implemented by MUSCLE<sup>[53]</sup> in MEGA. Several models were tested before proceeding with the neighbor-joining tree with the help of Jones, Taylor, and Thornton amino acid substitution model (JTT model), and keeping 1,000 replication as bootstrap. These 1,000 replications were tested only to satisfy the consistency of the phylogenetic analysis, and other parameters in MEGA were set as the default option.

### Synonymous (Ks) and non-synonymous (Ka) for duplicated pairs of *PUB* genes

Synonymous with non-synonymous (*Ka/Ks*) mutation analysis was performed for different types of duplicated *PUB* gene pairs identified in tomato with the help of MEGA<sup>[52]</sup>. These ratios from synonymous with non-synonymous were also calculated by using the MEGA and the Nei-Gojobori method (Juke-Cantor) model was tested by performing 1,000 bootstrap replications.

### Gene structure, motif composition, and physicochemical analysis of *PUB* protein

A generic file format (GFF3) was downloaded from the tomato genome, and then gene structure analysis was performed by using TBtools software<sup>[54]</sup>. Similarly, the motif composition of *PUB* proteins was carried out by using the MEME suite: version 5.1.1. The following parameter was set-up: the number of motifs 10 and width ranged from 50–100 while keeping the other parameter as the default option<sup>[55]</sup>. Both the GFF3 and XML file were utilized for gene structure visualization with the help of TBtools software<sup>[54]</sup>, and the XML was downloaded from the MEME server.

Further, the ExPASy PROTPARAM server (<http://web.expasy.org/protparam/>) was utilized aimed at different sequence properties of each *PUB* gene in tomato including, molecular weight (MW), isoelectric points (PIs), and GRAVY). Subcellular prediction analysis was performed with the help of the WOLF PSORT online resource (<https://wolfsort.hgc.jp/>).

### Gene collinearity and duplication analysis of *PUB* genes in tomato

Tomato genomic resources were downloaded from Phytozome as discussed above, both gene collinearity and duplication analysis were carried out by utilizing the tomato genomic sequences, protein sequences, and GFF3. These genomic resources were utilized with the help of TBtools software<sup>[54]</sup>, and the collinearity analysis was demonstrated with the help of Circos.

### Tissue-specific transcriptional profiling of the *PUB* gene family in tomato

The tissue-specific transcriptional profiling of *SIPUB* genes in numerous tissues was utilized by two major publically available RNA-seq datasets, these resources were utilized from the Tomato Functional Genomics Database (TFGD, <http://ted.bti.cornell.edu/cgi-bin/TFGD/digital/home.cgi>). The two Illumina RNA-seq datasets based on several multiple tissues were used from cultivated tomato (*Solanum lycopersicum* cv. Heinz) and the wild relative *Solanum pimpinellifolium*<sup>[33,34]</sup>. The data from the RNA-seq data of the model variety Micro-Tom (MT) was also

## SIPUBs gene family analysis in tomato

utilized for four different tissues at various stages such as root 30, 45, and 85 d post germination (DPG), stem 30, 45, and 85 DPG, leaf 30, 45, and 85 DPG, and flower 30 and 45 DPG. 10 DPA (55 DPG), 20 DPA (65 DPG), immature green (IMG) at 75 DPG, mature green (MG) at 80 DPG, Br (85 DPG), Br3 (88 DPG), Br7 (92 DPG), Br10 (95 DPG) and Br15 (100 DPG)<sup>[56]</sup>. The ten tissues from the cultivated tomato include bud, flower, leaf, roots, 1 cm, 2 cm, 3 cm, mature green, breaker, and breaker + 10 fruits. While the data from the wild relative, eight different tissues and organs were used, which included anthesis flower (0 DPA), 10 d post-anthesis (DPA) fruit, 20 DPA fruit, ripening fruit (33 DPA), mature leaves (ML), whole root (WR), young flower buds (YFB), and young leaves<sup>[30]</sup>, were selected for analysis. Heat maps were generated based on RPKM values by using the following method: bidirectional cluster analysis, and maximum distance with complete linkage method using the ClustVis<sup>[57]</sup>. Pearson correlation coefficient (*r*) analysis was performed on three different RNA-seq datasets of various tissues among the duplicated pairs of *PUB* genes using Rstudio (Version 3.5.2).

**Plant material and RT-PCR analysis**

The experiments were carried out on the Micro-Tom cultivar, which was grown in a 1:1 mixture of garden soil and vermiculite under a 16 h photoperiod (day/night temperature regime of 23 °C/18 °C, relative humidity of 70%). The plants were irrigated with 3 mL/cm<sup>3</sup> soil of either 200 mM NaCl or placed in an incubator at 4 °C once they had grown seven to eight completely expanded leaves. Different tissues were sampled and used for RT-PCR quantification, including roots, stems, leaves, flower buds, and flowers in full bloom.

Total RNA was extracted, using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) from five different tissues and the leaves with the same pitch position from plants exposed to salinity stress for 6 h, or low temperature treatment for 2 h were used. A 1 µg aliquot of RNA was used for the synthesis of the cDNA first strand, using a PrimeScript<sup>TM</sup>RT reagent kit containing gDNA Eraser (Takara, Shiga, Japan). The cDNA was used as the template for qRT-PCRs based on Fast SYBR Green Master Mix (www.bimake.com). The reference sequence was the *SUB18* gene, and relative transcript abundances were calculated using the 2<sup>-ΔΔCT</sup> method<sup>[58]</sup> and heatmap of fold changes was plotted on a log<sub>2</sub> scale. Rstudio (Version 3.5.2) was used to construct a heatmap and Pearson component analysis (PCA) analysis and the set of RT-PCR primer sequences used in this study is listed in [Dataset 6](#).

**Subcellular localization**

The full-length coding region without the termination codon was fused with 35S::GFP. It was then transformed into protoplasts of *Arabidopsis* after incubation for 14 h at 28 °C. A confocal laser scanning microscope was used for GFP fluorescence detecting.

**Determination of antioxidant enzyme activity**

Antioxidation enzyme activities were measured in the fifth or sixth fully expanded leaf (counting from the apex). A ~1 g sample of leaf was equilibrated in 10 mL 10 mM phosphate buffer (pH 7.8) for 40 min. SOD activity was measured as follows: the assay comprised a 100 µL aliquot of the enzymatic extract, to which was added 30 µM EDTA-Na<sub>2</sub>, 60 µM riboflavin and 2.25 mM nitroblue tetrazolium; the reaction was exposed to 4,000 lx fluorescent light for 20 min and then its absorbance was measured at 560 nm. For the estimation of POD activity,

the reaction mixture comprised a 100 µL aliquot of the enzymatic extract, containing 25 mM phosphate buffer (pH 7.8), 0.05% w/v guaiacol and 10 mM H<sub>2</sub>O<sub>2</sub>. The absorbance of the reaction was measured at 470 nm.

**Conclusions**

Herein, we identified 48 *PUB* genes in the tomato genome that were further characterized into seven different groups based on their phylogenetic analysis. We carried out a thorough genome-wide investigation search of the tomato genome for *PUB* genes identification by several bioinformatics tools, such as the physicochemical properties, gene structure, and motif composition, gene duplication, gene collinearity analysis, evolutionary rates, expression profiling, and functional evolution. Gene duplication analysis proposed mainly three different types among *PUB* genes i.e., segmental, tandem, and dispersed. Among these duplications notably, the dispersed genes contributed most to the expansion of *PUB* in tomato. The transcriptional analysis demonstrated significant changes specifically their likely diversification and fates of duplicated gene pairs during the process of evolution. The common fates among these genes were predicted to be involved during their functional or subfunctional conservation and neo-functionalization. These analyses, based on the *PUB* gene family, lay a solid basis for future investigation and their functional characterization in tomato.

**Acknowledgments**

This study was financially supported by grants from the Agricultural Science and Technology Innovation Program (34-IUA-03), the Central Public-interest Scientific Institution Basal Research Fund (S2022001), the National Natural Science Foundation of China (32200260), China Postdoctoral Science Foundation Funded Project (2020M673207), and the Fundamental Research Funds for the Central Universities (2020SCU12061).

**Conflict of interest**

The authors declare that they have no conflict of interest.

**Supplementary Information** accompanies this paper at (<https://www.maxapress.com/article/doi/10.48130/VR-2023-0016>)

**Dates**

Received 28 November 2022; Accepted 13 March 2023; Published online 16 May 2023

**References**

1. Dye BT, Schulman BA. 2007. Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins. *Annual Review of Biophysics and Biomolecular Structure* 36:131–50
2. Ciechanover A. 1998. The ubiquitin-proteasome pathway: on protein death and cell life. *The EMBO journal* 17:7151–60
3. Moon J, Parry G, Estelle M. 2004. The ubiquitin-proteasome pathway and plant development. *The Plant Cell* 16:3181–95
4. Smalle J, Vierstra RD. 2004. The ubiquitin 26S proteasome



- proteolytic pathway. *Annual Review of Plant Biology* 55:555–90
5. Harper JW, Burton JL, Solomon MJ. 2002. The anaphase-promoting complex: it's not just for mitosis any more. *Genes & Development* 16:2179–206
  6. Jackson PK, Eldridge AG, Freed E, Furstenthal L, Hsu JY, et al. 2000. The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. *Trends in Cell Biology* 10:429–39
  7. Aravind L, Koonin EV. 2000. The U box is a modified RING finger - a common domain in ubiquitination. *Current Biology* 10:PR132–R134
  8. Ohi MD, Vander kooi CW, Rosenberg JA, Chazin WJ, Gould KL. 2003. Structural insights into the U-box, a domain associated with multi-ubiquitination. *Nature Structural & Molecular Biology* 10:250–55
  9. Azevedo C, Santos-Rosa MJ, Shirasu K. 2001. The U-box protein family in plants. *Trends in Plant Science* 6:354–58
  10. Meacham GC, Patterson C, Zhang W, Younger JM, Cyr DM. 2001. The Hsc70 co-chaperone CHIP targets immature CFTR for proteasomal degradation. *Nature Cell Biology* 3:100–5
  11. Stone SL, Anderson EM, Mullen RT, Goring DR. 2003. ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible *Brassica* pollen. *The Plant Cell* 15:885–98
  12. Koegl M, Hoppe T, Schlenker S, Ulrich HD, Mayer TU, et al. 1999. A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* 96:635–44
  13. Hatakeyama S, Yada M, Matsumoto M, Ishida N, Nakayama KI. 2001. U box proteins as a new family of ubiquitin-protein ligases. *Journal of Biological Chemistry* 276:33111–20
  14. Patterson C. 2002. A new gun in town: The U box is a ubiquitin ligase domain. *Science's STKE* 2002:pe4
  15. Wiborg J, O'shea C, Skriver K. 2008. Biochemical function of typical and variant *Arabidopsis thaliana* U-box E3 ubiquitin-protein ligases. *Biochemical Journal* 413:447–57
  16. Zeng LR, Park CH, Venu RC, Gough J, Wang GL. 2008. Classification, expression pattern, and E3 ligase activity assay of rice U-Box-containing proteins. *Molecular Plant* 1:800–15
  17. Luo Q, Li Y, Wang W, Fei X, Deng X. 2015. Genome-wide survey and expression analysis of *Chlamydomonas reinhardtii* U-box E3 ubiquitin ligases (CrPUBs) reveal a functional lipid metabolism module. *PLoS One* 10:e0122600
  18. Wang C, Duan W, Riquicho AR, Jing Z, Liu T, et al. 2015. Genome-wide survey and expression analysis of the PUB family in Chinese cabbage (*Brassica rapa* ssp *pekinensis*). *Molecular Genetics and Genomics* 290:2241–60
  19. Wang N, Liu Y, Cong Y, Wang T, Zhong X. 2016. Genome-wide identification of soybean U-box E3 ubiquitin ligases and roles of GmPUB8 in negative regulation of drought stress response in *Arabidopsis*. *Plant and Cell Physiology* 57:1189–209
  20. Andersen P, Kragelund BB, Olsen AN, Larsen FH, Chua NH, et al. 2004. Structure and biochemical function of a prototypical *Arabidopsis* U-box domain. *Journal of Biological Chemistry* 279:40053–61
  21. Mudgil Y, Shiu SH, Stone SL, Salt JN, Goring DR. 2004. A large complement of the predicted *Arabidopsis* ARM repeat proteins are members of the U-box E3 ubiquitin ligase family. *Plant Physiology* 134:59–66
  22. Yee D, Goring DR. 2009. The diversity of plant U-box E3 ubiquitin ligases: from upstream activators to downstream target substrates. *Journal of Experimental Botany* 60:1109–21
  23. Cho SK, Chung HS, Ryu MY, Park MJ, Lee MM, et al. 2006. Heterologous expression and molecular and cellular characterization of *CaPUB1* encoding a hot pepper U-box E3 ubiquitin ligase homolog. *Plant Physiology* 142:1664–82
  24. Cho SK, Ryu MY, Song C, Kwak JM, Kim WT. 2008. *Arabidopsis* PUB22 and PUB23 are homologous U-box E3 ubiquitin ligases that play combinatory roles in response to drought stress. *The Plant Cell* 20:1899–914
  25. Liu Y, Wu Y, Huang X, Sun J, Xie Q. 2011. *AtPUB19*, a U-box E3 ubiquitin ligase, negatively regulates abscisic acid and drought responses in *Arabidopsis thaliana*. *Molecular Plant* 4:938–46
  26. Raab S, Drechsel G, Zarepour M, Hartung W, Koshiba T, et al. 2009. Identification of a novel E3 ubiquitin ligase that is required for suppression of premature senescence in *Arabidopsis*. *The Plant Journal* 59:39–51
  27. Salt JN, Yoshioka K, Moeder W, Goring DR. 2011. Altered germination and subcellular localization patterns for PUB44/SAUL1 in response to stress and phytohormone treatments. *PLoS One* 6:e21321
  28. Park JJ, Yi J, Yoon J, Cho LH, Ping J, et al. 2011. *OsPUB15*, an E3 ubiquitin ligase, functions to reduce cellular oxidative stress during seedling establishment. *The Plant Journal* 65:194–205
  29. Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, et al. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635–41
  30. Bai YL, Kissoudis C, Yan Z, Visser RGF, van der Linden G. 2018. Plant behaviour under combined stress: tomato responses to combined salinity and pathogen stress. *The Plant Journal* 93:781–93
  31. Moin M, Bakshi A, Madhav MS, Kirti PB. 2019. Comprehensive expression profiling reveals the possible involvement of Cullins in developmental and stress regulation in rice. *Environmental and Experimental Botany* 160:101–11
  32. Khan N, Fatima F, Haider MS, Shazadee H, Liu Z, et al. 2019. Genome-wide identification and expression profiling of the polygalacturonase (PG) and pectin methylesterase (PME) genes in grapevine (*Vitis vinifera* L.). *International Journal of Molecular Sciences* 20:3180
  33. Dai Q, Geng L, Lu M, Jin W, Nan X, et al. 2017. Comparative transcriptome analysis of the different tissues between the cultivated and wild tomato. *PLoS One* 12:e0172411
  34. Pattison RJ, Csukasi F, Zheng Y, Fei Z, van der Knaap E, et al. 2015. Comprehensive tissue-specific transcriptome analysis reveals distinct regulatory programs during early tomato fruit development. *Plant Physiology* 168:1684–701
  35. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, et al. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research* 48:D265–D268
  36. Han Y, Ding T, Su B, Jiang H. 2016. Genome-wide identification, characterization and expression analysis of the chalcone synthase family in maize. *International Journal of Molecular Sciences* 17:161
  37. Moore RC, Purugganan MD. 2003. The early stages of duplicate gene evolution. *Proceedings of the National Academy of Sciences of the United States of America* 100:15682–87
  38. Gilbert W, de Souza SJ, Long MY. 1997. Origin of genes. *Proceedings of the National Academy of Sciences* 94:7698–703
  39. Long M, Langley CH. 1993. Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* 260:91–95
  40. Cao Y, Han Y, Meng D, Li D, Jin Q, et al. 2016. Structural, evolutionary, and functional analysis of the Class III peroxidase gene family in Chinese pear (*Pyrus bretschneideri*). *Frontier in Plant Science* 7:1874
  41. Khan N, Ke H, Hu CM, Naseri E, Haider MS. 2019. Genome-wide identification, evolution, and transcriptional profiling of PP2C gene family in *Brassica rapa*. *BioMed Research International* 2019:2965035
  42. Stebbins GL. 1950. *Variation and evolution in plants*. pp 644. London: Oxford University Press. pp 643–44.
  43. Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–55
  44. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–83
  45. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. 2009. The B73



- maize genome: complexity, diversity, and dynamics. *Science* 326:1112–15
46. Aderemi F, Wuraola A. 2006. Effects of dietary replacement of maize with malted or unmalted sorghum on the performance of weaner rabbits. *African Journal of Food, Agriculture, Nutrition and Development* 10:4032–46
  47. Maere S, de Bodt S, Raes J, Casneuf T, van Montagu M, et al. 2005. Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 102:5454–59
  48. Birchler JA, Veitia RA. 2007. The gene balance hypothesis: from classical genetics to modern genomics. *The Plant Cell* 19:395–402
  49. Goodstein DM, Shu SQ, Howson R, Neupane R, Hayes RD, et al. 2011. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* 40:D1178–D1186
  50. Lamesch P, Berardini TZ, Li DH, Swarbreck D, Wilks C, et al. 2012. The *Arabidopsis* information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40:D1202–D1210
  51. Letunic L, Bork P. 2018. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research* 46:D493–D496
  52. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870–74
  53. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–97
  54. Cheng CJ, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Molecular Plant* 13(8):1194–202
  55. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. 2009. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* 37:W202–W208
  56. Li Y, Chen Y, Zhou L, You S, Deng H, et al. 2020. MicroTom metabolic network: Rewiring tomato metabolic regulatory network throughout the growth cycle. *Molecular Plant* 13:1203–18
  57. Metsalu T, Vilo J. 2015. ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Research* 43:W566–W570
  58. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  Method. *Methods* 25:402–8



Copyright: © 2023 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.