

RESEARCH

Open Access



The LEA gene family in tomato and its wild relatives: genome-wide identification, structural characterization, expression profiling, and role of *SILEA6* in drought stress

Chunping Jia^{1,2†}, Bin Guo^{1,3†}, Baike Wang¹, Xin Li^{1,3}, Tao Yang¹, Ning Li¹, Juan Wang^{1*} and Qinghui Yu^{1,2*}

Abstract

Background Late embryogenesis abundant (LEA) proteins are widely distributed in higher plants and play crucial roles in regulating plant growth and development processes and resisting abiotic stress. Cultivated tomato (*Solanum lycopersicum*) is an important vegetable crop worldwide; however, its growth, development, yield, and quality are currently severely constrained by abiotic stressors. In contrast, wild tomato species are more tolerant to abiotic stress and can grow normally in extreme environments. The main objective of this study was to identify, characterize, and perform gene expression analysis of LEA protein families from cultivated and wild tomato species to mine candidate genes and determine their potential role in abiotic stress tolerance in tomatoes.

Results Total 60, 69, 65, and 60 *LEA* genes were identified in *S. lycopersicum*, *Solanum pimpinellifolium*, *Solanum pennellii*, and *Solanum lycopersicoides*, respectively. Characterization results showed that these genes could be divided into eight clusters, with the LEA_2 cluster having the most members. Most *LEA* genes had few introns and were non-randomly distributed on chromosomes; the promoter regions contained numerous *cis*-acting regulatory elements related to abiotic stress tolerance and phytohormone responses. Evolutionary analysis showed that *LEA* genes were highly conserved and that the segmental duplication event played an important role in evolution of the *LEA* gene family. Transcription and expression pattern analyses revealed different regulatory patterns of *LEA* genes between cultivated and wild tomato species under normal conditions. Certain *S. lycopersicum* *LEA* (*SILEA*) genes showed similar expression patterns and played specific roles under different abiotic stress and phytohormone treatments. Gene ontology and protein interaction analyses showed that most *LEA* genes acted in response to abiotic stimuli and water deficit. Five *SILEA* proteins were found to interact with 11 *S. lycopersicum* *WRKY* proteins involved in development or resistance to stress. Virus-induced gene silencing of *SILEA6* affected the antioxidant and reactive oxygen species defense systems, increased the degree of cellular damage, and reduced drought resistance in *S. lycopersicum*.

Conclusion These findings provide comprehensive information on *LEA* proteins in cultivated and wild tomato species and their possible functions under different abiotic and phytohormone stresses. The study systematically

[†]Chunping Jia and Bin Guo, contributed equally to this work.

*Correspondence:

Juan Wang

drjuanwang@126.com

Qinghui Yu

yuqinghui98@sina.com

Full list of author information is available at the end of the article



broadens our current understanding of LEA proteins and candidate genes and provides a theoretical basis for future functional studies aimed at improving stress resistance in tomato.

Keywords Cultivated tomato, Wild tomato, LEA, Genome-wide identification, Expression patterns, Evolutionary analyses, Abiotic stressors, Phytohormone treatments

Background

Abiotic stressors such as drought, salinity, and extreme weather affect normal crop growth and development, causing massive yield losses and quality degradation [1, 2]. Through continuous evolution, a series of mechanistic changes have occurred in plants at the molecular, physiological, and biochemical levels to maximize resistance to the adverse effects of abiotic stressors [3]. For example, transcription factors (TFs) and protein kinases can regulate the downstream signaling pathways that activate and regulate specific stress-related genes, ultimately leading to physiological responses to stress such as accumulation of osmoregulatory substances [4, 5]. Specific functional proteins related to stress tolerance in plants, for example, late embryogenesis abundant (LEA) proteins that are highly enriched in late embryogenesis and nutritional tissues in response to water deficit, can eliminate reactive oxygen species (ROS) from cells to protect macromolecules and mitigate the damage caused by abiotic stressors [6, 7].

LEA proteins is the collective term for a large group of glycine-rich, hydrophilic, and structurally intrinsically disordered proteins that are widely distributed in the plant kingdom [8]. A study in 1981 first identified a LEA protein in *Gossypium hirsutum* and showed that it accumulates in large quantities during seed dehydration and maturation to protect the seed from damage [9]. Subsequently, with the development of whole-genome sequencing technology, additional LEA proteins were identified in model crops (*Arabidopsis thaliana* and *Oryza sativa*) [10, 11], food crops (*Zea mays* and *Triticum aestivum*) [12, 13], horticultural crops (*Cucumis sativus* and *Vitis vinifera*) [14–16], and other species (microbes and invertebrates) [17]. Most LEA proteins, except for a small proportion in the endoplasmic reticulum of plant cells, are highly diverse and are distributed in the nucleus, cytoplasm, and mitochondria [18]. The classification of LEA proteins differs among species. Typically, LEA proteins are classified into eight different clusters based on sequence similarity, specific motifs, and conserved structural domains [19, 20].

As LEA proteins have unique biochemical properties, such as high hydrophilicity [21], presence of a large number of charged amino acid (AA) residues [22], glycine, or other tiny AAs [23], small number of cysteine residues [24], and a lack of tryptophan residues [7], they

are tolerant to heat and acid and participate in numerous physiological processes associated with plant development and responses to adversity, thereby contributing to various stress tolerance responses [7, 20, 25].

It is widely known that an extremely rapid increase in the cytosolic ion concentration can cause irreversible cell damage in plants subjected to adversity [26]. LEA proteins with bound divalent cations can redirect water molecules in cells, bind salt ions, and activate abiotic stress-induced oxidative scavengers in cells [27]. LEA proteins effectively slow down the inactivation of malate dehydrogenase, lactate dehydrogenase, catalase, citrate synthase, and some mitochondrial enzymes in response to water deficiency and exhibit good protective properties against enzymatic activity [28–30]. A few LEA proteins act by stabilizing the membranes; sugar-bound membrane maintenance and ion segregation help in preventing the collapse of cellular structures [6, 31, 32]. Furthermore, LEA proteins can repair misassembled proteins to recover their biological activity by utilizing molecular chaperones that bind to misfolded proteins [33, 34].

The imperative role of LEA proteins in plant reactions to abiotic stressors has been reported in many studies. Overexpression of *AtLEA14* overactivates the transcripts of salt stress response marker genes in transgenic *A. thaliana*, which exhibits high salt tolerance [35]. Overexpression of *OsLEA3-2* increases salt and drought tolerance in transgenic *A. thaliana* and *O. sativa* [36]. During the nutritional growth period, *OsEm1* responds to the hormone abscisic acid (ABA) and abiotic stressors, and overexpression of *OsEm1* increases the survival rate and expression of other LEA genes in transgenic *O. sativa*, thereby improving its sensitivity to ABA and enhancing tolerance to osmotic stressors [37]. Both abiotic stress and signaling molecules induce *ZmLEA3* expression in leaves, and overexpression of *ZmLEA3* improves the tolerance of transgenic *Nicotiana tabacum* to osmotic and oxidative stress [38]. In addition, the *CaLEA1* gene in *Capsicum annuum* [39], *TaLEA3* gene in *Leymus chinensis* [40], *IpLEA* gene in *Ipomoea pes-caprae* [41], *SmLEA* gene in *Salvia miltiorrhiza* [42], and *SiLEA14* gene in *Setaria italica* [43] have also been reported to play roles in stress tolerance.

S. lycopersicum is an important vegetable crop and a model plant belonging to the Solanaceae family.

However, its growth and development, yield, and quality are severely constrained by abiotic stressors [44]. Conversely, wild tomato species have become tolerant to abiotic stressors and have been able to grow normally in extreme environments after differential expression of different genes activated certain molecular and physiological mechanisms that allowed them to adapt effectively [45, 46]. Previous genome sequencing-based research on *S. lycopersicum*, *Solanum pimpinellifolium*, *Solanum pennellii*, and *Solanum lycopersicoides* provides an excellent opportunity to identify and analyze the *LEA* gene family in the context of abiotic stress [47, 48].

In the present study, we aimed to perform genome-wide identification of *LEA* genes in the above-mentioned tomato species and systematically analyze the phylogeny, gene structure, conserved structural domains, motifs, chromosome distribution, purification pressure, duplication date, tandem duplication, and segmental duplication events. We also conducted homology and synteny analyses among different species, analyzed *cis*-acting regulatory elements, transcriptional expression profiles, expression patterns under different abiotic stressors and phytohormone treatments, gene ontology (GO) annotation, protein interaction networks, and performed virus-induced gene silencing (VIGS) to provide new information regarding the *LEA* gene family and a reference for future molecular and biofunctional studies aimed at developing genetic improvement strategies in tomato.

Results

LEA identification and characterization

We used the protein families (Pfam) database identity documents (IDs) (PF03760, PF03168, PF03242, PF02987, PF0477, PF10714, PF04927, and PF00257) from *S. lycopersicum*, *S. pimpinellifolium*, *S. pennellii*, and *S. lycopersicoides*, which were identified as having 60, 69, 65, and 60 *LEA* genes, respectively. The identified *LEA* genes were divided into eight clusters—*LEA_1* (18 genes), *LEA_2* (160 genes), *LEA_3* (16 genes), *LEA_4* (12 genes), *LEA_5* (7 genes), *LEA_6* (2 genes), dehydrin (25 genes), and seed maturation protein (SMP) (14 genes). The largest and smallest numbers of *LEA* genes were observed in the *LEA_2* and *LEA_6* clusters, respectively. All *LEA* proteins were predicted to have different physicochemical properties; the peptide length ranged from 78–737 AA, with a mean of 228.44 AA. The molecular weight (MW) of these 254 *LEA* proteins ranged from 8.87–78.60 kDa. The isoelectric point (pI) ranged from 4.42–10.42, with 25.20% being acidic proteins (pI < 7) and 74.80% being basic proteins (pI > 7). The instability coefficients ranged from 6.65–59.1, with 53.94% of *LEA* proteins having low instability index (< 40), indicating that > 50% of the

LEA proteins were stable at the theoretical level. The grand average of the hydropathy index (GRAVY) suggested that > 83.46% of *LEA* proteins were hydrophilic (GRAVY < 0). We observed similar physicochemical parameters for most *LEA* genes in the same family, which was in agreement with the main characteristics of *LEA* genes (i.e., low hydrophobicity and high net charge, which can help plants to resist desiccation effects). The predicted possible subcellular localization of all *LEA* genes revealed that they had a broad distribution at the cellular level, with the majority targeting the nucleus (39.37%) or plasma membrane (26.77%). Contrastingly, some other members were localized in cellular structures such as the cytoplasm, extracellular matrix, mitochondria, and chloroplasts, thus establishing protective mechanisms and enhancing intermembrane interactions to effectively respond to various external stressors (Additional file 1: Table S1).

Phylogeny, gene structure, conserved domains, and motif analysis of *LEA* genes

To investigate the homology and similarity, unrooted phylogenetic trees were created based on the comparison of *LEA* protein sequences of cultivated and wild tomato species (Fig. 1). Phylogenetic analysis indicated that they were separated into eight clusters belonging to two main branches; generally, the *LEA_2* cluster was in one branch and the other clusters were in the other branch. Furthermore, the phylogeny of *LEA* genes exhibited strong correlations with gene structure, conserved domain, and motif. Genetic structure analysis revealed that the introns and exons of *LEA* genes had different structural features and that the most closely related *LEA* genes in the same group had similar gene structures in terms of intron number and exon length. The majority of genes had intron numbers ranging from 0–3. Ninety-six of the genes had no introns, which implied that the *LEA* genes may have lost numerous introns during their evolution. Conserved domain analysis showed that except for SMP, the other *LEA* protein subfamily members had a simple structure, mostly containing a *LEA* structural domain; most of these were near the C-terminus of the protein, with a few at the N-terminal or intermediate position. The same *LEA* subfamily contained the same protein structure domain, suggesting that their main functions may be the same. Our further conserved motif analysis showed that among the 20 different motifs, motifs 1, 2, 3, and 10 were common in the *LEA_2* subfamily of cultivated and wild tomato species. Except for the *LEA_2* subfamily, the other *LEA* subfamily members had a relatively simple motif composition, mostly comprising 1–3 motifs. Each *LEA* subfamily member contained several identical motifs specific to the group; this taxon specificity implied

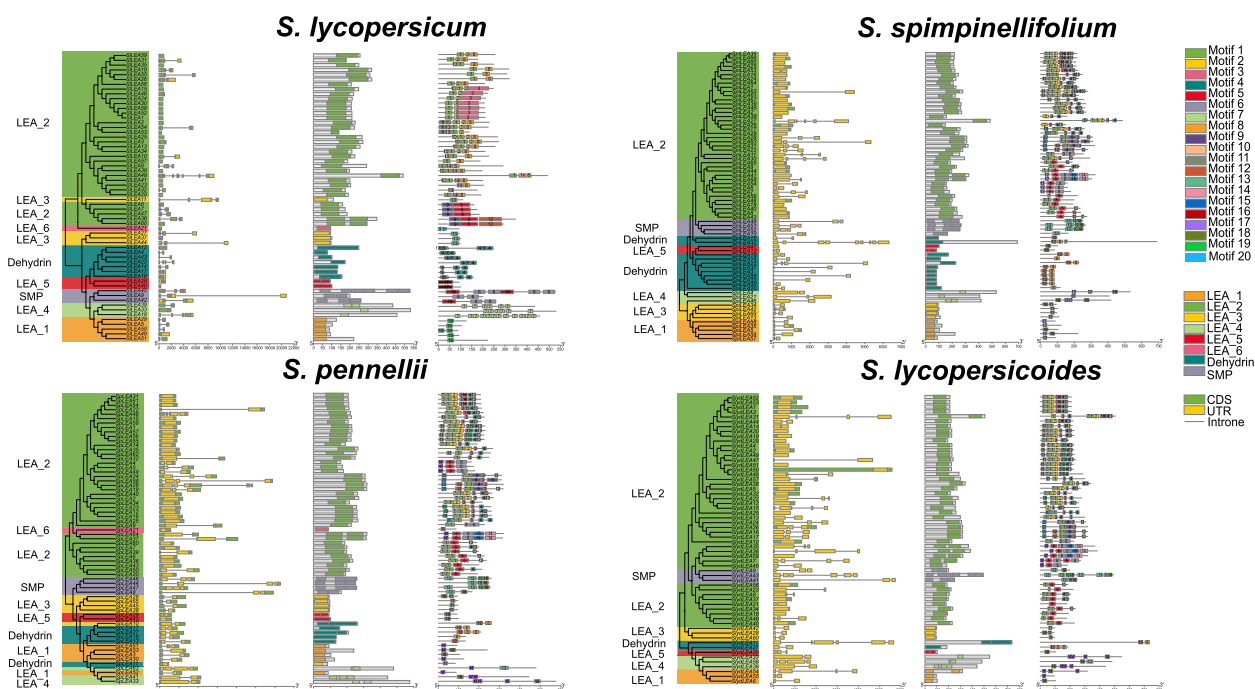


Fig. 1 Phylogeny, gene structure, conserved domains, and motifs of *LEA* genes in cultivated and wild tomato species. Each diagram represents, in order from left to right, the phylogenetic tree of *LEA* proteins; structure of exons, introns, and untranslated regions (UTR) in *LEA* genes; distribution of conserved domains of *LEA* proteins; different motifs present in *LEA* proteins

that the members belonging to the same taxon have similar biological functions. The formation of motif patterns suggested that *LEA* proteins are positively engaged in a multitude of biological processes (BP). The diverse nature of the conserved motifs suggested that *LEA* proteins evolved from the amplification of genes within their particular gene families.

Phylogenetic tree analysis of *LEA* proteins

Considering the large differences in length and different structural domains of the *LEA* protein family members, we first performed a phylogenetic tree analysis of all 254 *LEA* proteins between cultivated and wild tomato species (Fig. 2a). The results showed that they were mainly divided into eight evolutionary branches, with the *LEA_2* subfamily at one end of the main branch and the remaining subfamilies at the other end of the main branch. Except for the dehydrin subfamily (which was divided into five sub-branches), the other branches included the corresponding *LEA* subfamily members. Subsequently, we performed a phylogenetic analysis of all *LEA* proteins with the 97 *LEA* proteins present in the *A. thaliana* and *O. sativa* genomes (Fig. 2b), which we re-predicted and identified. Previous studies predicted that the *A. thaliana* [10] and *O. sativa* [11] genomes contain 51 and 35

LEA proteins, respectively; because of the large differences, in this study they were renamed according to the order of their distribution from top to bottom in each chromosome (further information is provided in Additional file 2: Table S2). This enabled comprehension of the evolutionary relationships of *LEA* families among different model plants. The results of phylogenetic tree analysis showed that the main evolutionary branches corresponded to different *LEA* subfamilies and that each subfamily could be divided into two subfamilies. In general, the *LEA_2* subfamily contained the most members, the *LEA_6* subfamily contained the least members, and the members in the same subfamily had similar origins and evolutionary relationships. The different subfamilies contained *LEA* members from *O. sativa*, *A. thaliana*, and cultivated and wild tomato species, and their origins presumably appeared before the divergence of monocotyledons and dicotyledons.

Chromosome distribution, duplication events, purification pressure, and synteny analysis

The genomic distribution of *LEA* genes varied depending on the chromosome (Fig. 3). They were distributed on all chromosomes of *S. pimpinellifolium* and all chromosomes of *S. lycopersicum*, *S. pennellii*, and *S.*

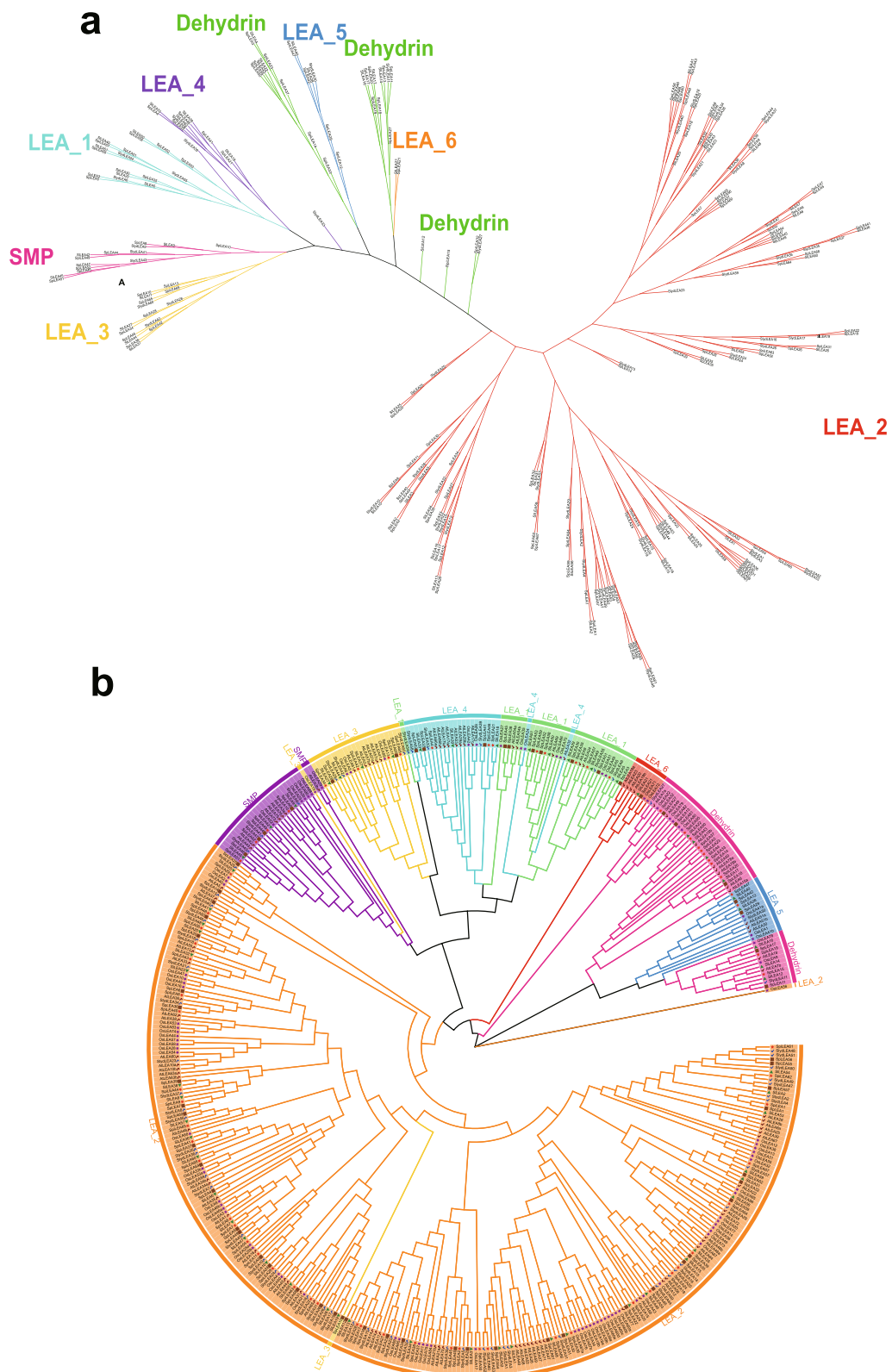


Fig. 2 Phylogenetic tree analysis of LEA proteins. **a** Phylogenetic tree of LEA proteins between cultivated and wild tomato species. **b** Phylogenetic tree of LEA proteins between tomato and other species. The sequences of all LEA proteins were compared with multiple sequences, and a phylogenetic tree was constructed using the maximum likelihood method and 1000 bootstrap resampling in MEGA 11 software. Each color represents a cluster

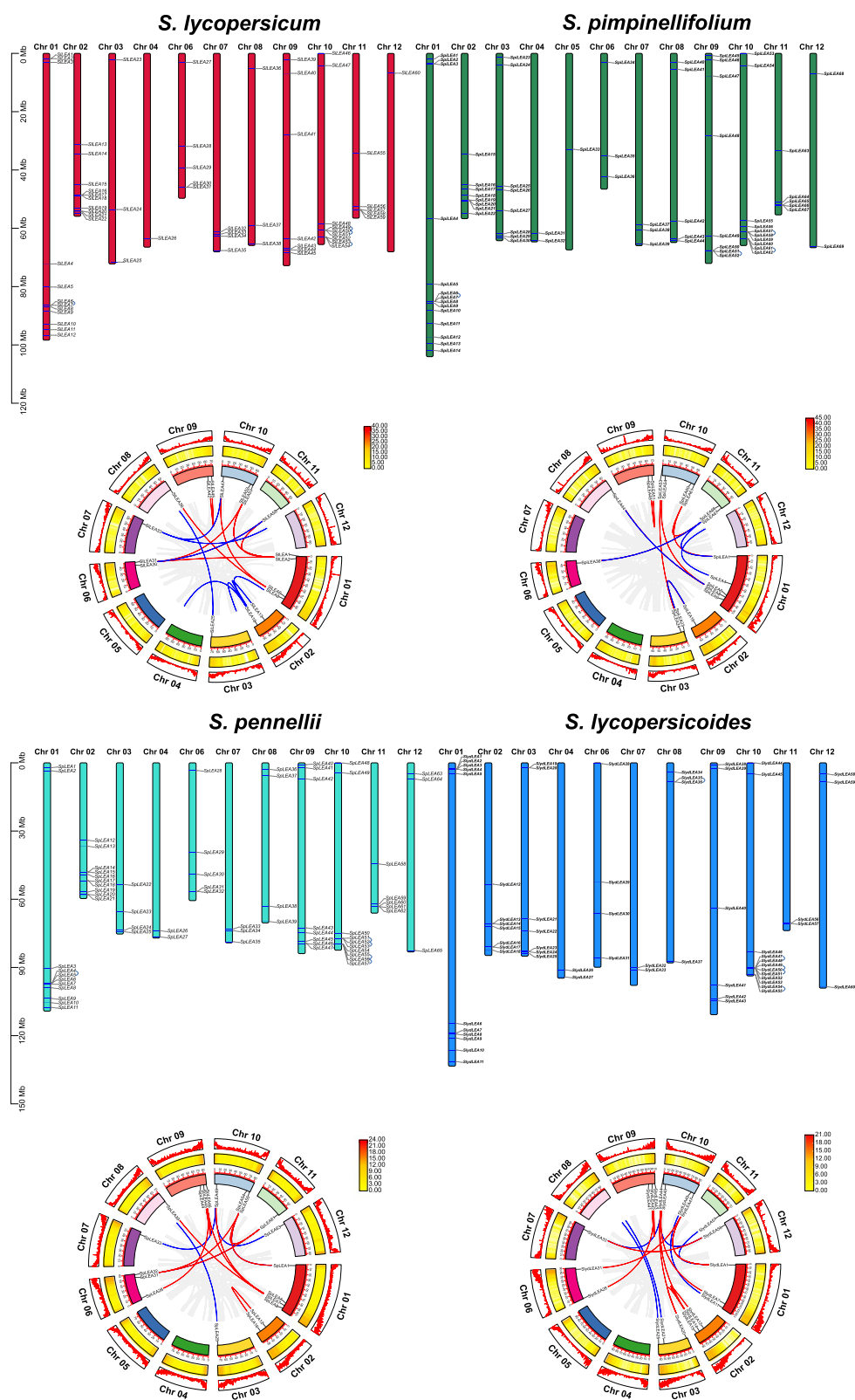


Fig. 3 Chromosomal localization and gene duplication events of *LEA* genes in cultivated and wild tomato species. *LEA* distribution on the chromosomes of cultivated and wild tomato species is represented by different colored bars; each blue arc represents a pair of tandem duplicated genes. The different colored lines in the corresponding circle plots indicate synthetic pairs of *LEA* genes, and the gray lines indicate synthetic pairs of all genes

lycopersicoides except for chromosome 5. Among *S. lycopersicum*, *S. pimpinellifolium*, *S. pennellii*, and *S. lycopersicoides*, all *LEA* genes were most densely distributed on chromosome 1, with 12, 14, 11, and 11 members, respectively. In contrast, they had a sparse distribution on chromosomes 4 and 12, on which only 1–3 members were distributed. Most *LEA* genes were distributed at both ends of the chromosomes. An analysis of intra-genomic duplication events in all *LEA* genes showed that *S. lycopersicoides* (located on chromosomes 8 and 10) produced five pairs of tandem duplication events, and *S. lycopersicum*, *S. pimpinellifolium*, and *S. pennellii* (located on chromosomes 1, 9, and 10) produced four, four, and six pairs of tandem duplication events, respectively (Fig. 3, Additional file 3: Table S3).

Furthermore, we found that *S. lycopersicum*, *S. pimpinellifolium*, *S. pennellii*, and *S. lycopersicoides* exhibited seven, seven, nine, and seven pairs of segmental duplication events distributed on the corresponding chromosomes, respectively (Fig. 3, Additional file 3: Table S3). Segmental duplication events were more pronounced than tandem duplication events, which may imply that segmental duplication events were a major contributor to the evolutionary process.

We first assessed species selection and evolution using the non-synonymous substitution (*Ka*)/synonymous substitution (*Ks*) ratio. The ratios of tandem duplication events in *LEA* genes ranged from 0.08 to 2.59, with a mean of 0.83. The ratios of segmental duplication events ranged from 0.18 to 2.20, with a mean of 0.98 (Additional file 3: Table S3). The mean values of the ratios of all segmental duplication and tandem duplication gene pairs were < 1, suggesting that *LEA* genes were subjected to purifying selection during evolution. We then used the *Ks* values to predict the dispersion time of *LEA* duplication events. Tandem duplication of *LEA* genes occurred between 3.17 and 92.64 million years ago (MYA), and segmental duplication occurred between 31.18 and 149.38 MYA (Additional file 3: Table S3).

We also discovered that seven pairs of segmental duplication events occurred between 40 and 50 MYA. Furthermore, we predicted that the tandem duplication events of the *SpLEA55/SpLEA56*, *SILEA53/SILEA54*, and *SlydLEA50/SlydLEA51* gene pairs occurred at approximately 3.17, 3.33, and 3.94 MYA, i.e., after the tomato–*Solanum tuberosum* division [49].

By constructing divergence time trees and synteny profiles between cultivated and wild tomato species, we gained further insights into the divergence time and synteny relationships of *LEA* genes (Fig. 4a). The results showed that they were mainly divided into two evolutionary branches at around 5.9 MYA, with *S. lycopersicum*

and *S. pimpinellifolium* having similar divergence times at around 0.68 MYA; contrastingly, their divergence times with respect to other species were farther apart. Similar synteny relationships among *LEA* genes existed between cultivated and wild tomato species. Most of them formed approximately three synteny gene pairs, which suggested a conserved evolutionary role of the *LEA* family in cultivated and wild tomato species.

We further elucidated the sequence similarity of *S. lycopersicum* *LEA* (*SILEA*) genes with related *LEA* genes and homologous genes of other species; to this end, we constructed synthetic maps of *SILEA* genes with the *LEA* genes of two monocots (*Z. mays* and *O. sativa*) and two dicotyledons (*A. thaliana* and *S. tuberosum*) (Fig. 4b). The outcome revealed that the synteny profiles of *SILEA* genes were largely consistent with those of the remaining four species. We identified 14, 27, 47, and 57 repeat events (shown as red lines) in *Z. mays*, *O. sativa*, *A. thaliana*, and *S. tuberosum*, respectively. We found no synteny between *S. lycopersicum* and *Z. mays* and *O. sativa* on chromosomes 4, 5, and 12; synteny modules were mainly distributed on the remaining chromosomes. *S. lycopersicum* had synteny modules with *S. tuberosum* except for chromosome 5, whereas *S. lycopersicum* had synteny modules with *A. thaliana* on all chromosomes. *SILEA* genes had no homologous genes in *Z. mays*. *SILEA38* on *S. lycopersicum* chromosome 8 was homologous to *Os01t0234700-01* on *O. sativa* chromosome 1 (shown as green lines), and *OsLEA4* on *O. sativa* chromosome 1 was homologous to *Solyc03g120720.3.1* on *S. lycopersicum* chromosome 3 (shown as blue lines). There were multiple pairs of homologous genes between *S. lycopersicum* and *A. thaliana* and between *S. lycopersicum* and *S. tuberosum*, with *SILEA18* and *SILEA48* being homologous in both *A. thaliana* and *S. tuberosum*. Similarly, *SILEA38* had homologous genes in both *O. sativa* and *A. thaliana*.

Analyses of divergence times and syntenic relationships of *LEA* genes among the other 10 species showed that they mainly divided into two evolutionary branches around 108 MYA, with tomato and *S. tuberosum* having similar divergence times to rice and maize at approximately 32 MYA (Fig. 4c). More synteny relationships existed among *LEA* genes in monocotyledons (*O. sativa*, *T. aestivum*, and *Z. mays*). Except for *S. tuberosum* and *V. vinifera*, synthetic relationships existed between the *LEA* genes of adjacent species. Some *LEA* genes formed 2–5 synteny gene pairs, which may have played an important role in the evolution of the *LEA* gene family, and indirectly suggested that the structure and function of *LEA* genes are variable and diverse.

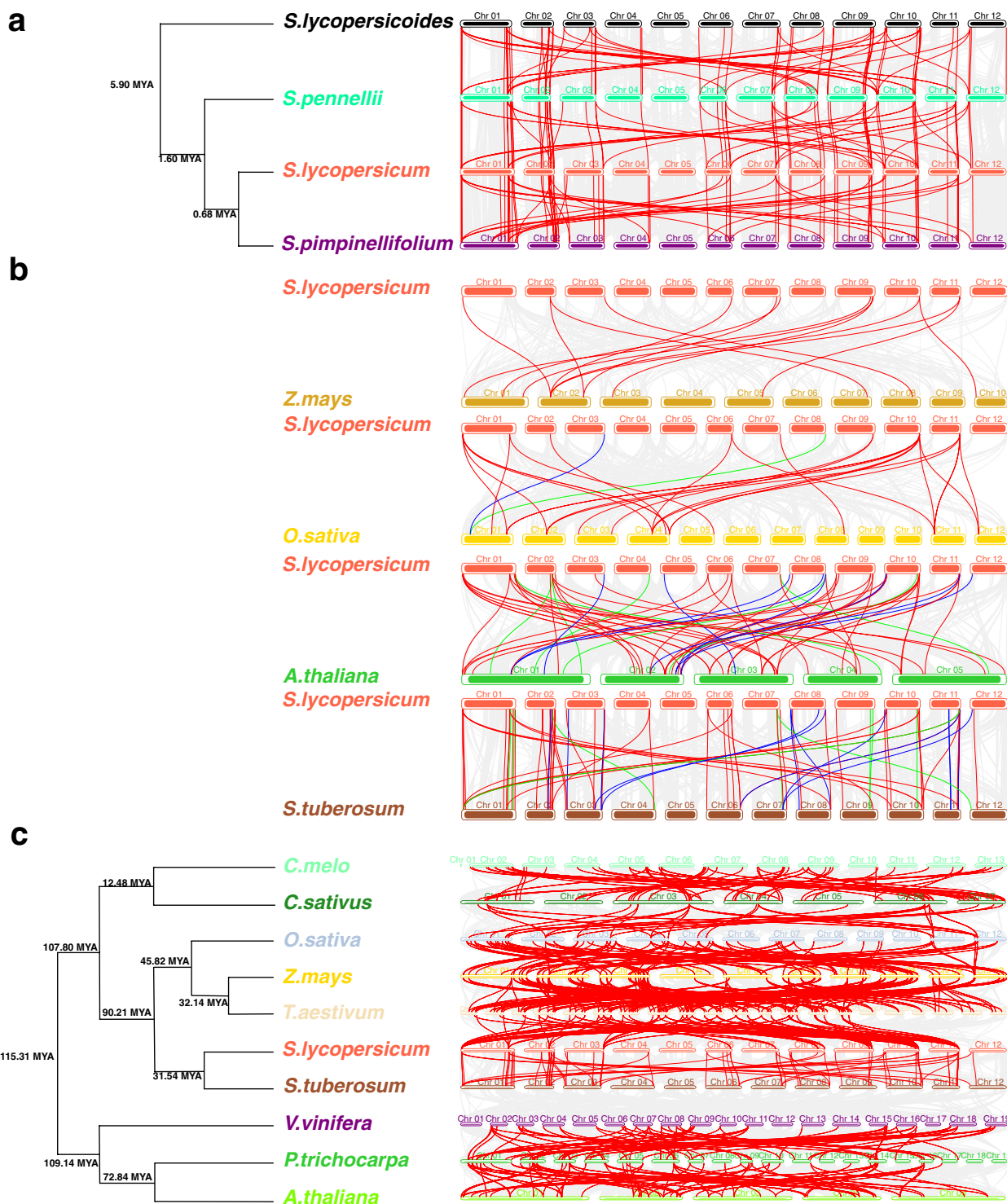


Fig. 4 Synteny plot of *LEA* genes. **a** The red line indicates the synteny of *LEA* genes in cultivated and wild tomato species. **b** Synteny plot of *SILEA* genes with those of the other four species. Gray lines indicate synteny between the *S. lycopersicum* genome and those of other species. Red lines indicate synteny between *SILEA* genes and the *LEA* genes of other species. Blue lines indicate synteny between *SILEA* genes and the non-*LEA* genes of other species. Green lines indicate synteny between *LEA* genes of other species and non-*LEA* genes of *S. lycopersicum*. **c** Red lines indicate synteny of *LEA* genes among the 10 species

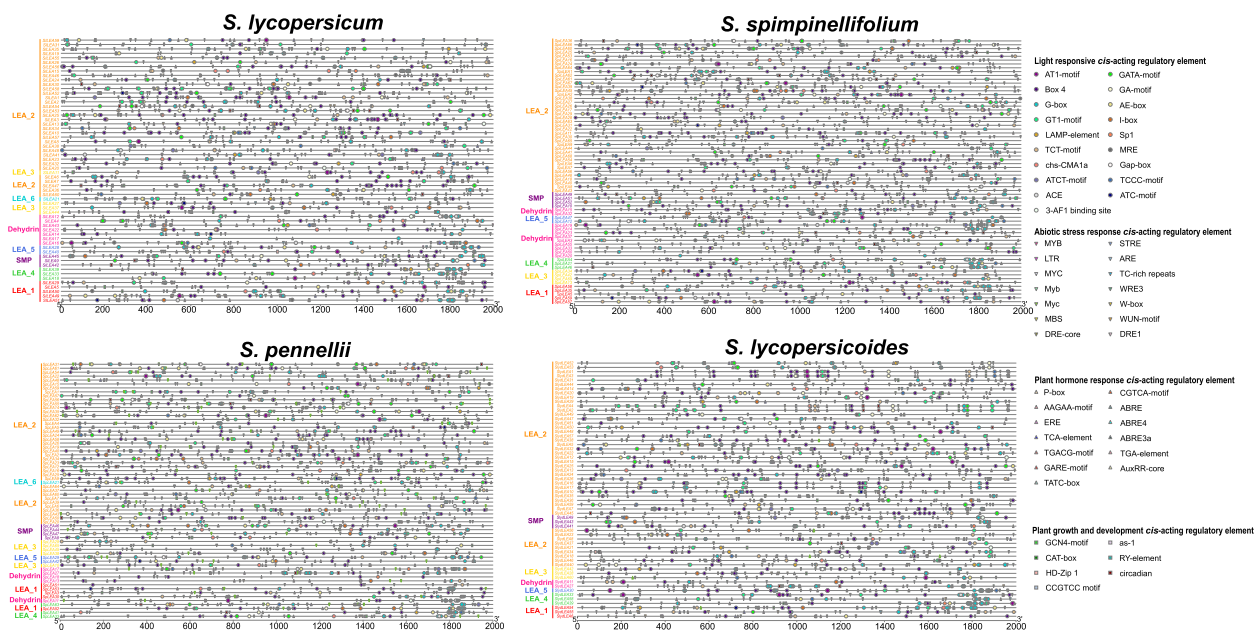


Fig. 5 Predicted *cis*-acting regulatory elements in the promoter regions of *LEA* genes in cultivated and wild tomato species. Gene names are arranged according to the order of the phylogenetic tree. Different colors and shapes of patterns represent different classes of *cis*-acting regulatory elements

Analysis of *cis*-acting regulatory elements of *LEA*

Our prediction analysis of *cis*-acting regulatory elements at the 2000 bp nucleotide sequence upstream (from ATG) of each *LEA* in cultivated and wild tomato species showed that except for the common *cis*-acting regulatory elements, i.e., TATA-box and CAAT-box, in the promoter region, the other *cis*-acting regulatory elements present could be classified into four categories (Fig. 5). The first category was light-responsive *cis*-acting regulatory elements; all *LEA* genes in cultivated and wild tomato species contained 19 such *cis*-acting regulatory elements (including AT1-motif, GATA-motif, GA-motif, Box4, and G-box), and almost all *LEA* genes comprised at least one light-responsive *cis*-acting regulatory element in their promoter regions. The second category was *cis*-acting regulatory elements related to responses to abiotic stress environments or external stressors. Thirteen such *cis*-acting regulatory elements (including MYB, STRE, LTR, ARE, and MYC) suggested that *SILEA* genes may play key roles in abiotic stress defense and stress, drought response, and anaerobic induction. In addition, some *LEA* genes contained W-box motifs, which were combining locus for *WRKY* TFs. The third category was phytohormone-responsive *cis*-acting regulatory elements. Thirteen such *cis*-acting regulatory elements, including P-box, CGTCA-motif, AAGAA-motif, ABRE, and ERE, were related to the response to various hormones such as ABA, methyl

jasmonate (MeJA), salicylic acid, gibberellins (GA), and auxin (IAA), respectively. The fourth category of *cis*-acting regulatory elements was involved in plant growth and development (such as GCN4-motif, as-1, and CAT-box) and included seven such *cis*-acting regulatory elements, which were associated with growth and developmental processes such as tomato meristem expression, chloroplast differentiation, and endosperm expression.

Analysis of the expression pattern and transcriptional profiles of *LEA* genes

We first analyzed the expression profiles of *LEA* genes in *S. lycopersicum*, *S. pimpinellifolium*, *S. habrochaites*, and *S. pennellii* leaves in an attempt to reveal the differential regulation patterns between cultivated and wild tomato species under normal conditions (Fig. 6a). The clustered heat map results showed that 16 *LEA* genes, including *LEA1/6/17*, were significantly overexpressed in *S. pennellii*, *LEA30/36/38/39* were significantly overexpressed in *S. habrochaites*, *LEA12/19/31/35* were significantly highly expressed in *S. pimpinellifolium*, and *LEA3/15/16/26/58* were significantly highly expressed in *S. lycopersicum*. Different patterns of differential regulation were observed between cultivated and wild tomato species under normal conditions; this up and downregulation relationship of *LEA* genes among different cultivated and wild tomato species may also have been closely

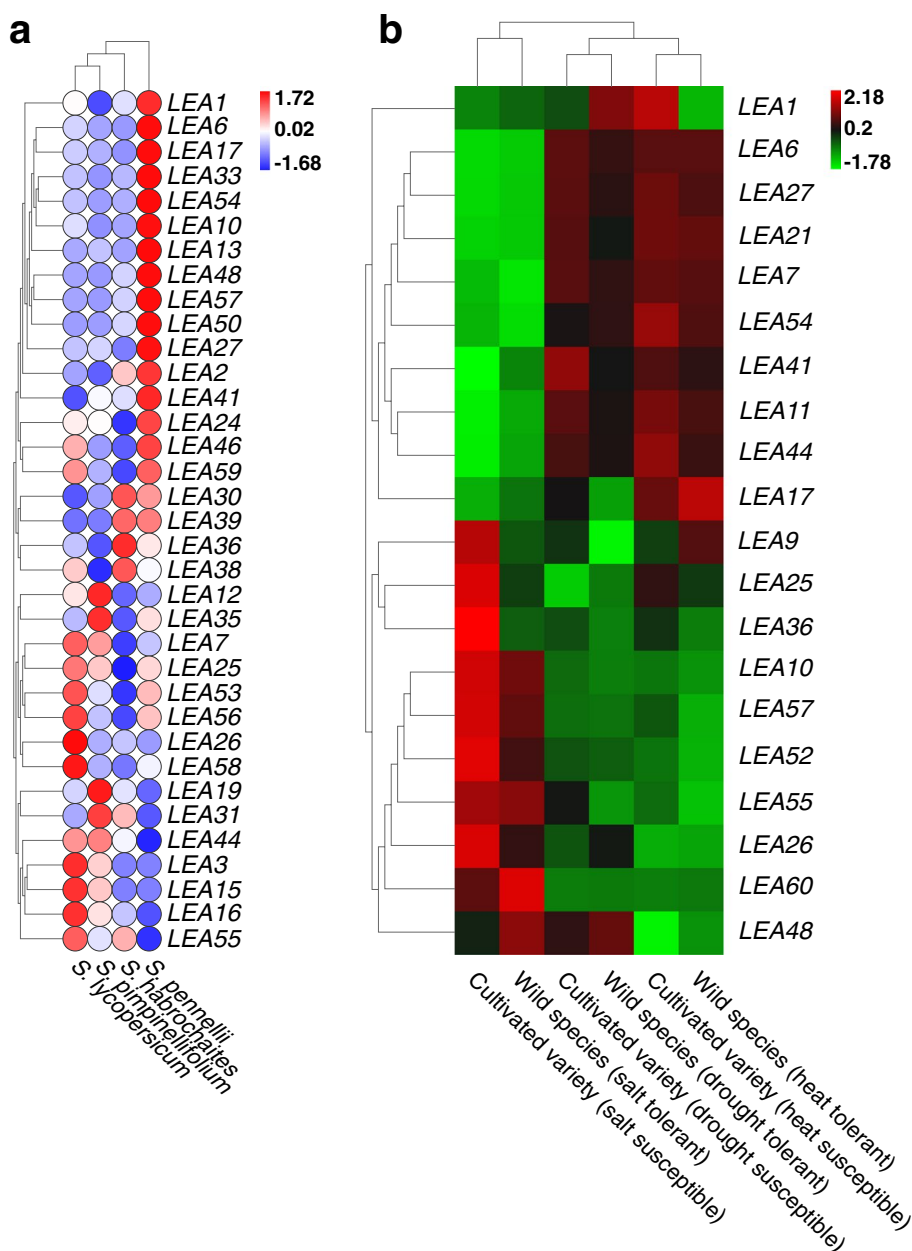


Fig. 6 Clustering heat map of *LEA* genes from leaves of cultivated and wild tomato species under (a) normal and (b) abiotic stress conditions. Salt stress: *S. lycopersicum* (salt-susceptible cultivar) versus *S. pimpinellifolium* (salt-tolerant wild tomato species) (GEO accession number GSE16401). Plants were treated with 200 mM NaCl for 5 h. Drought stress: *S. lycopersicum* (drought-susceptible cultivar) versus *S. habrochaites* (drought-tolerant wild tomato species) (GEO accession number GSE22304). Plants were subjected to water stress by holding water for 7 d. Heat stress: *S. lycopersicum* (heat-susceptible cultivar) versus *S. habrochaites* (heat-tolerant wild tomato species) (GEO accession no. GSE22304). Plants were treated at a high temperature of 40 °C for 1 h

related to their abiotic stress tolerance-related expression patterns.

To obtain more information on the role of *LEA* genes under abiotic stress conditions, we analyzed their transcriptional expression profiles under salt, drought,

and high-temperature stress and compared the results to those of the wild species (Fig. 6b). Under salt stress, *LEA9/25/26/36/52* were downregulated in wild species, whereas *LEA48/60* were upregulated in wild species compared to those in cultivars. Under drought stress,

LEA11/21/41/44 were downregulated in the wild species, whereas *LEA1/48/54* were upregulated in the wild species compared to those in cultivars. Under high-temperature stress, *LEA1/25/44/54* were downregulated in the wild species, whereas *LEA9/17* were upregulated in the wild species compared to those in cultivars. *LEA9* was downregulated under salt stress and upregulated under high-temperature stress compared to that in the cultivar. *LEA1/54* were upregulated under drought stress and down-regulated under high-temperature stress compared to those in the cultivar. *LEA25* was downregulated under both salt and high-temperature stress compared to that in the cultivar, *LEA44* was downregulated under both drought and high-temperature stress compared to that in the cultivar, and *LEA48* was upregulated under both salt and drought stress compared to that in the cultivar. The differential up or downregulation of *LEA* genes under different stress conditions indicated their different patterns of regulation in cultivated and wild tomato species, which are complex and diverse.

To further explore the potential features of *SILEA* genes, the transcript profile expression values of each *SILEA* on the TomExpress online platform were systematically analyzed. The expression patterns of *SILEA* genes in nutritional and reproductive tissues were presented in the form of an additional clustered heat map (Fig. 7). *SILEA* genes were expressed in several examined tissues, and the expression levels varied considerably among tissues and periods. *SILEA* genes were generally underexpressed in meristematic tissues, leaves, and flowers and generally overexpressed in seeds and fruits. Among them, certain genes exhibited unique expression profiles. For example, *SILEA20* showed the highest expression during mid-seed development (10 days post anthesis, DPA), *SILEA35* in roots, *SILEA8* during mid-meristem tissue development (10 DPA), *SILEA22* during mid-leaf development (11 d), and *SILEA34* during early flower development (8 d). *SILEA2/5/15/24/31/38/55/58* were consistently expressed at high levels during leaf and phloem tissue growth and development. *SILEA18/47/49* showed the highest expression in flower buds (3 mm). *SILEA26/30/57* were consistently highly expressed during the initial periods of fruit development (0, 1, 2, and 5 DPA). More than half of the *SILEA* genes were persistently highly expressed during late seed development (38, 41, and 44 DPA), which suggested a potential role of these genes in regulating seed maturation. *SILEA1/3/7/13/36/53/54/59* showed high expression levels during the middle and late periods of fruit development, which suggested that they may be engaged in the growth and development of *S. lycopersicum* fruits (from immature to mature). The above-mentioned

analysis of *SILEA* expression pattern provides additional insights into their possible roles.

Analysis of *SILEA* expression patterns in response to abiotic stressors and phytohormone treatments

We identified the *cis*-acting regulatory elements of all *LEAs* that may be responsible for directing their differential activation or repression concerning an abiotic stress environment or external stress and phytohormone responses and analyzed the expression patterns and transcriptional profiles of *SILEA* genes. Following this, we randomly selected 24 *SILEA* genes in an attempt to identify the genes contributing to responses to various abiotic stressors and phytohormone treatments. General expression characteristics of *SILEA* genes in response to different abiotic stressors and phytohormone treatments were examined to better understand the possible roles of various *SILEA* family members.

In the present analysis, we only considered genes with a >fivefold changes in transcript abundance and *P*-values < 0.05. Using this criterion, we observed that none of the *SILEA* genes showed a significant change in expression at any period under the control treatment in normal conditions (Additional file 4: Figure S1). However, under simulated drought stress, *SILEA4/6/12/14/17/27/50* were significantly upregulated at the early stage (1 or 3 h) and maintained a high expression profile until the middle stage (12 h). Among them, *SILEA6* and *SILEA17* were rapidly upregulated by approximately 50-fold at the initial (1 h) and middle (12 h) stages, respectively (Additional file 5: Figure S2). Under simulated high salt stress, *SILEA4/6/12/14/16/17/18/27/33/50* were significantly upregulated at the early stage (1 or 3 h) and maintained high expression until the late stage (24 h). Among them, *SILEA6/50/17* were strongly upregulated by more than 100-fold in the initial (1 h), middle (12 h), and late (24 h) stages, respectively (Additional file 6: Figure S3). Under simulated high-temperature stress, *SILEA27/37/50/60* responded rapidly and were significantly upregulated at the early stage (1 or 3 h); *SILEA60* maintained high expression from the initial (1 h) to the late (24 h) stage, and *SILEA37* was rapidly upregulated by more than 60-fold at the initial stage (1 h) (Additional file 7: Figure S4). Under simulated low-temperature stress, *SILEA6/14/27* were significantly upregulated at the early stage (1 or 3 h) and maintained high expression until the late stage (24 h); *SILEA50* was significantly upregulated at the middle and late stages (12 h and 24 h), and *SILEA6* was rapidly upregulated by more than 60-fold at the middle stage (12 h) (Additional file 8: Figure S5).

SILEA genes showed no significant changes in expression at any period under the control treatment of ddH₂O spraying (Additional file 9: Figure S6). Contrastingly,

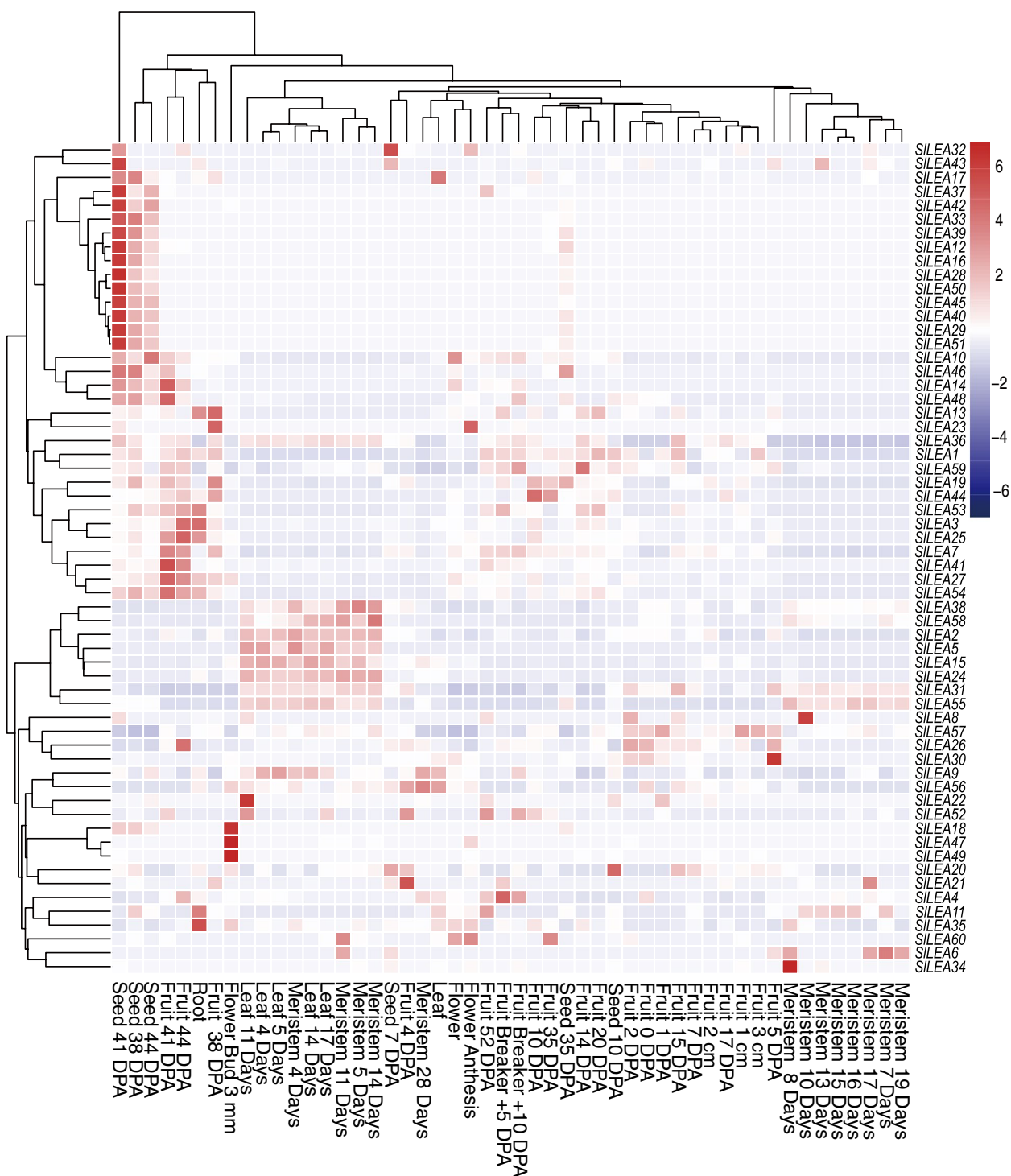


Fig. 7 Clustering heat map of *SILEA* expression patterns in various tissues. Horizontal coordinates indicate the seeds, roots, meristems, leaves, flowers, and fruits of cultivated tomatoes at different growth periods. Developmental stages include the seed (7, 10, 35, 38, 41, and 44 DPA), root, meristem (4, 5, 7, 8, 10, 11, 13, 14, 15, 16, 17, 19, and 28 d), leaf (4, 5, 11, 14, and 17 d), flower (anthesis and bud 3 mm), and fruit (0, 1, 2, 4, 5, 7, 10, 14, 15, 17, 20, 35, 38, 41, 44, and 52 DPA; 1, 2, and 3 cm; breaker +5 and +10). Vertical coordinates indicate *SILEA* expression pattern. The color gradient from blue to red indicates low to high expression

under the phytohormone ABA treatment, *SILEA6* and *SILEA27* were rapidly induced and their expression was highly significantly increased at the initial stage (1 h). *SILEA6* was rapidly upregulated by approximately 150-fold, and *SILEA14* expression was highly significantly increased at the middle stage (12 h) (Additional file 10: Figure S7). Under the phytohormone MeJA treatment, *SILEA6* was rapidly induced and highly significantly upregulated at the initial (1 h) and middle (6 h) stages, *SILEA4* and *SILEA17* were highly significantly upregulated at the late stage (24 h), and *SILEA27* maintained high expression from the early (1 h) to the late (24 h) stage (Additional file 11: Figure S8). Under the strigolactones (*rac*-GR24) treatment, *SILEA6* was rapidly induced at the early stage (1 h), was highly significantly up-regulated, and maintained high expression until the middle stages (6 h). *SILEA27* responded rapidly at the early stage (1 h), was highly significantly upregulated, and maintained high expression until the middle stage (12 h) (Additional file 12: Figure S9). *SILEA7* was rapidly induced and highly significantly up-regulated at the initial stage (1 h) upon treatment with the phytohormone γ -aminobutyric acid (GABA). *SILEA6/16/27* responded rapidly and were significantly upregulated at the early stage (1 h and 3 h), with *SILEA6* being strongly upregulated by more than 80-fold at the initial stage (1 h) (Additional file 13: Figure S10). The different expression patterns of these *SILEA* genes suggested that they performed particular functions under different stressor and phytohormone treatments. Overall, *SILEA* expression was more sensitive to high salt and drought stressors than temperature. Some *SILEA* genes showed similar expression profiles in response to various abiotic stressors; *SILEA4/12/17* responded to both high salt and drought stressors; *SILEA6/14/50* responded to high salt, drought, and low-temperature stressors; *SILEA27* simultaneously responded to all four of these stress conditions. Similarly, *SILEA6* and *SILEA27* exhibited comparable expression patterns in response to different phytohormone treatments and simultaneously responded to the four phytohormone treatments. *SILEA6* simultaneously responded to all of the conditions mentioned above except high-temperature stress, and *SILEA27* simultaneously responded to all of the conditions mentioned above. These findings indicate that *SILEA6* and *SILEA27* may be the key genes in the *SILEA* family that participate in defense and stress responses to external stressors.

GO annotation and enrichment analyses

We first investigated the BP, molecular functions (MF), and cellular components (CC) of *LEA* genes in cultivated and wild tomato species. The GO analysis results indicated that 254 *LEA* genes may be involved in a range

of BP. Among them, most *LEA* genes were predicted to play roles in the responses to abiotic stimuli and water deficit. The prediction of MF indicated that most *LEA* genes were involved in single-stranded DNA binding and transferase activity. Genes involved in condensed nuclear chromosome and egg chorion were higher in the predictions for CC (Additional file 14: Figure S11).

We further performed GO annotation and enrichment analyses of 60 *SILEA* genes to gain a more in-depth understanding of the functions of these genes. The GO secondary classification statistical graph results indicated that only two major categories (i.e., BP and CC) were observed in GO terms, with no MF (Fig. 8a). The GO-directed acyclic graph confirmed that these *SILEA* genes were involved in a range of BP; most *SILEA* genes were predicted to play a role in the responses to desiccation and water deficit and to respond to abiotic stimuli and stressful environments (Fig. 8b). The GO significance bubble plots indicated that the CC of *SILEA* genes was predicted to be the most functionally enriched, involving the membrane, cell, and cytoplasmic fractions (Fig. 8c). The innermost layer of the enrichment circle plot was particularly visually focused to show the rich factor values (i.e., the number of foreground genes divided by the number of background genes) for each taxon, including response to desiccation (GO: 0,009,269) (6/8), cold acclimation (GO: 0,009,631) (5/11), and response to water (GO: 0,009,415) (14/78) (Fig. 8d), further suggesting an essential role of *SILEA* genes in the regulation of these BP.

Protein interaction network analysis

Based on String 11.5, we constructed the interaction network of *SILEA* proteins with *S. lycopersicum* WRKY (SIWRKY) proteins in an attempt to improve our current understanding of the functional proteome-wide regulatory network (Fig. 9a). We found a total of five *SILEA* proteins interacting with 11 SIWRKY proteins, including *SILEA16* with SIWRKY30, *SILEA23* with SIWRKY26/70, *SILEA13* with SIWRKY14/56/66, *SILEA27* with SIWRKY14/48/56/66, and *SILEA17* with SIWRKY23/47/65/77. They each covered different SIWRKY subclasses [50], of which Classes I and III each had one (i.e., SIWRKY14 and SIWRKY70, respectively) and Classes IIc and IId + IIe each had four (SIWRKY30/47/48/56) and five (SIWRKY23/26/65/66/77) SIWRKY members, respectively. Additionally, we constructed the interaction network of *AtLEA* proteins with *AtWRKY* proteins to further explore this network (Fig. 9b). We observed that a total of seven *AtLEA* proteins interacted with 14 *AtWRKY* proteins, including *AtLEA11* with *AtWRKY19/62/72*, *AtLEA14* with *AtWRKY19/58/62/72*, *AtLEA74* with *AtWRKY72*, *AtLEA28* with *AtWRKY6/16/27/32/47/58*, *AtLEA57* with *AtWRKY6/20/27*, *AtLEA67* with *AtWRKY16/27/28/48/54*,

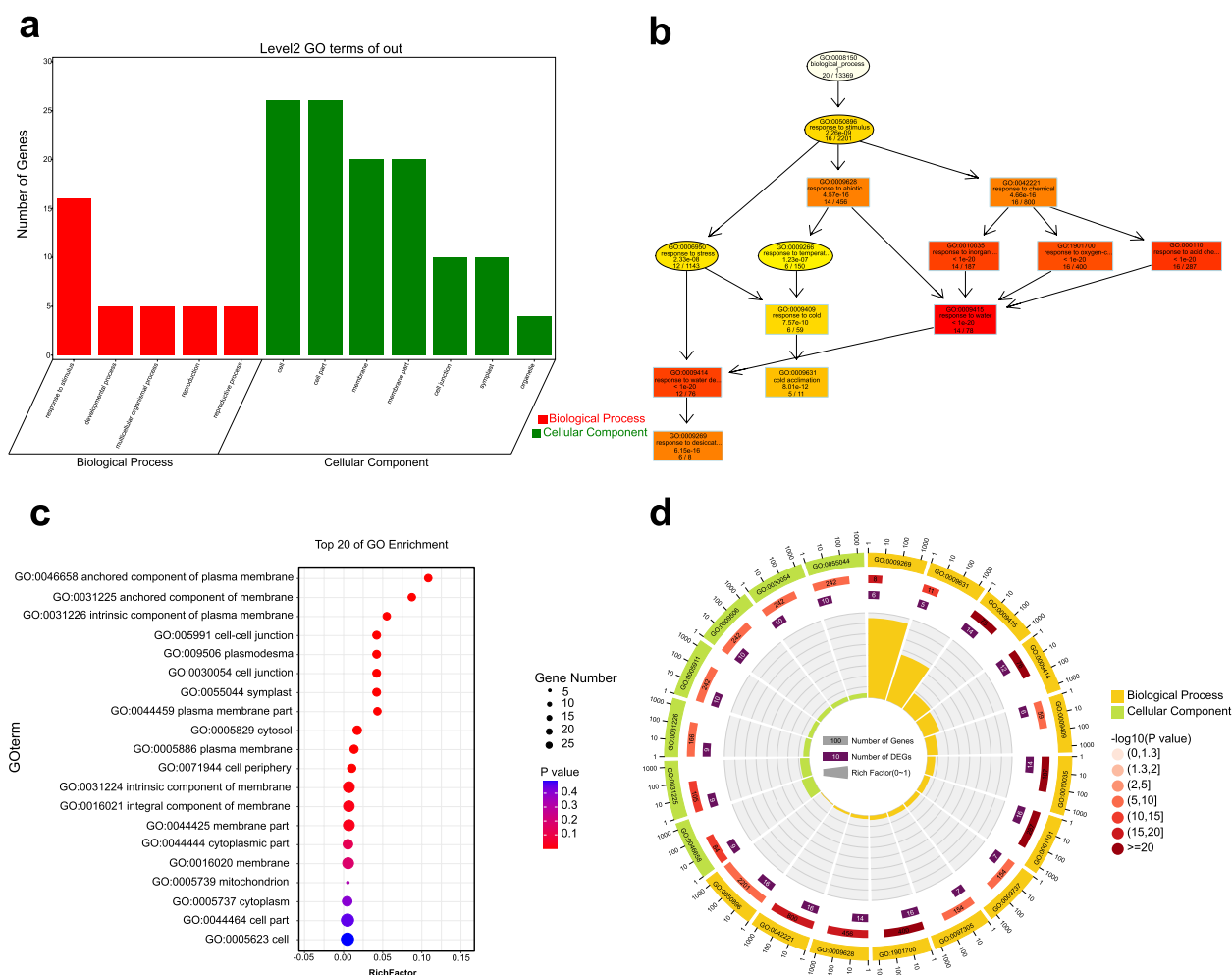


Fig. 8 GO annotation and enrichment analyses of *SILEA* genes. **a** Statistical graph of the secondary classification of GO. Horizontal coordinates represent the secondary classification of GO ontology, and vertical coordinates represent the number of genes present in each classification entry, thus indicating the number and enrichment of target genes in each GO term. **b** Directed acyclic graph of GO. The GO term closer to the root node is more general, and the GO term branching down is annotated to a finer level. The first 10 GO terms with the highest salience are rectangular, and the other GO terms are circular. The darker the color, the more significant the GO term, from light to dark: no color, light yellow, dark yellow, and red, respectively. **c** Significant bubble plot. The top 20 significantly enriched GO terms are displayed by default. The color represents the *P*-value, the bubble size represents the number of genes enriched to this GO term, and the rich factor in the horizontal coordinate of the bubble plot represents the ratio of the number of differentially expressed genes located in this GO term to the total number of genes located in this GO term among all annotated genes (i.e., the larger the ratio, the greater the enrichment). **d** Enrichment circle map. The first circle is the enriched classification, and the outside of the circle is a coordinate scale for the number of genes, with different colors representing different classifications. The second circle shows the number of that classification in the background genes and the *P*-value, whereby the more genes the longer the bar and the smaller the value the redder the color. The third circle shows the total number of foreground genes. The fourth circle shows the rich factor value for each classification (i.e., the number of foreground genes in that classification divided by the number of background genes), with each small cell of the background auxiliary line indicating 0.1

and AtLEA21 with AtWRKY16/54/69 (Fig. 9c). They also covered different AtWRKY subclasses [51], with three (AtWRKY20/32/58) and two (AtWRKY28/48) for Classes Ia and Ib, respectively; three (AtWRKY6/47/72) and two (AtWRKY27/69) for Classes IIa and IIb, respectively; and one (AtWRKY16) and two (AtWRKY54/62) for Classes IIc

and III, respectively. In particular, AtWRKY19 is a member of the A1 subgroup of the family encoding MAPK/ERK kinases. In addition, we found that the core binding domains of SiWRKY and AtWRKY were highly conserved, suggesting that they may perform similar roles, such as interacting with LEA proteins (Fig. 9d).

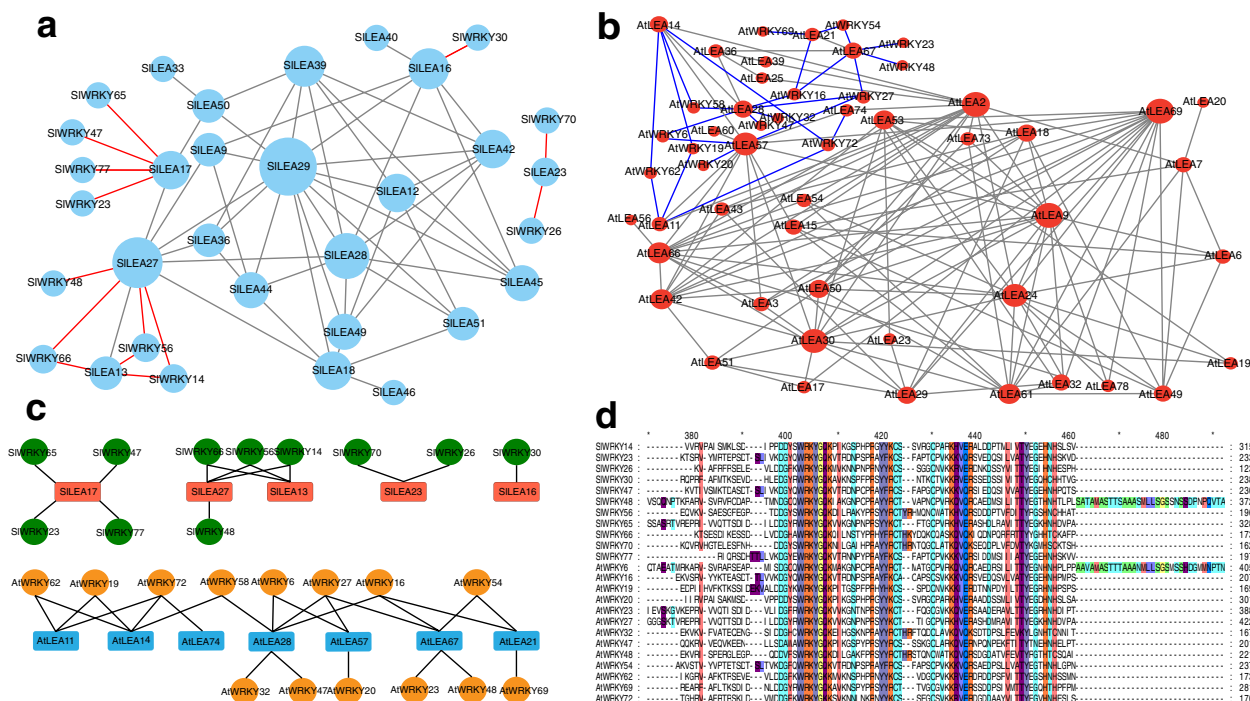


Fig. 9 Protein interaction network diagrams of LEA and WRKY proteins in *S. lycopersicum* and *A. thaliana*. **a** Protein interaction network of SILEA and SIWRKY proteins in *S. lycopersicum*. Network construction was performed using String 11.5. The basic settings were as follows: active interaction sources, experiments, databases, and gene fusion. The maximum number of interactors was 100, as calculated using the show simplification function to hide disconnected nodes in the network; the threshold for the confidence parameter was set to 0.40. Gray lines indicate interactions between different SILEA proteins, and red lines indicate interactions between SILEA and SIWRKY proteins. **b** Protein interaction network of AtLEA and AtWRKY proteins in *A. thaliana*. Gray lines indicate interactions among different AtLEA proteins, and blue lines indicate interactions between AtLEA and AtWRKY proteins. **c** Redrawing of a simpler and more intuitive protein interaction network diagram. The top panel shows the protein interaction network diagram of SILEA and SIWRKY proteins. The bottom panel shows the protein interaction network diagram of AtLEA and AtWRKY proteins. **d** Comparative analysis of WRKY structural sequences interacting with LEA proteins in *S. lycopersicum* and *A. thaliana*

Changes in phenotypic and physiological indices in the SILEA6-silenced lines under stress conditions

The VIGS technique is one of the most effective reverse genetics techniques used to characterize gene function. Our results of silencing and functional analyses of *SILEA6* using the VIGS method showed that gene silencing of *SILEA6* occurred at the posttranscriptional level in tomato leaves 2 weeks after injection inoculation, with a highly significant decrease in expression by approximately 60% compared to that in the negative control (Fig. 10a). After 1 h of salt, high, and low-temperature stress treatment respectively (note: *SILEA6* expression peaked at this point), none of the plants

showed any degree of phenotypic changes compared to those at 0 h (Additional file 15: Figure S12). Interestingly, after 1 h of drought stress, the *pTRV1-pTRV2-SILEA6* plants exhibited more severe leaf curling and wilting than the negative and positive control plants compared to those at 0 h (Fig. 10a).

We further collected the top leaves of plants after drought stress for DAB and NBT staining and measured the enzymatic activities and physiological index contents of their biological antioxidant and ROS scavenging systems to systematically determine the effect of *SILEA6* silencing on the drought tolerance of tomato plants (Fig. 10b). The results showed that after 1 h of

(See figure on next page.)

Fig. 10 Phenotypic analysis and determination of physiological indices under drought stress conditions after *SILEA6* silencing. **a** Phenotypic changes in plants after 1 h of drought stress and *SILEA6* expression after being silenced. **b** DAB and NBT staining of leaves after 1 h of drought stress and determination of various physiological indices. Three independent biological replicates were included to calculate the mean. Error bars show the standard deviations (SD) of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett’s multiple comparisons test (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and **** $P < 0.0001$)

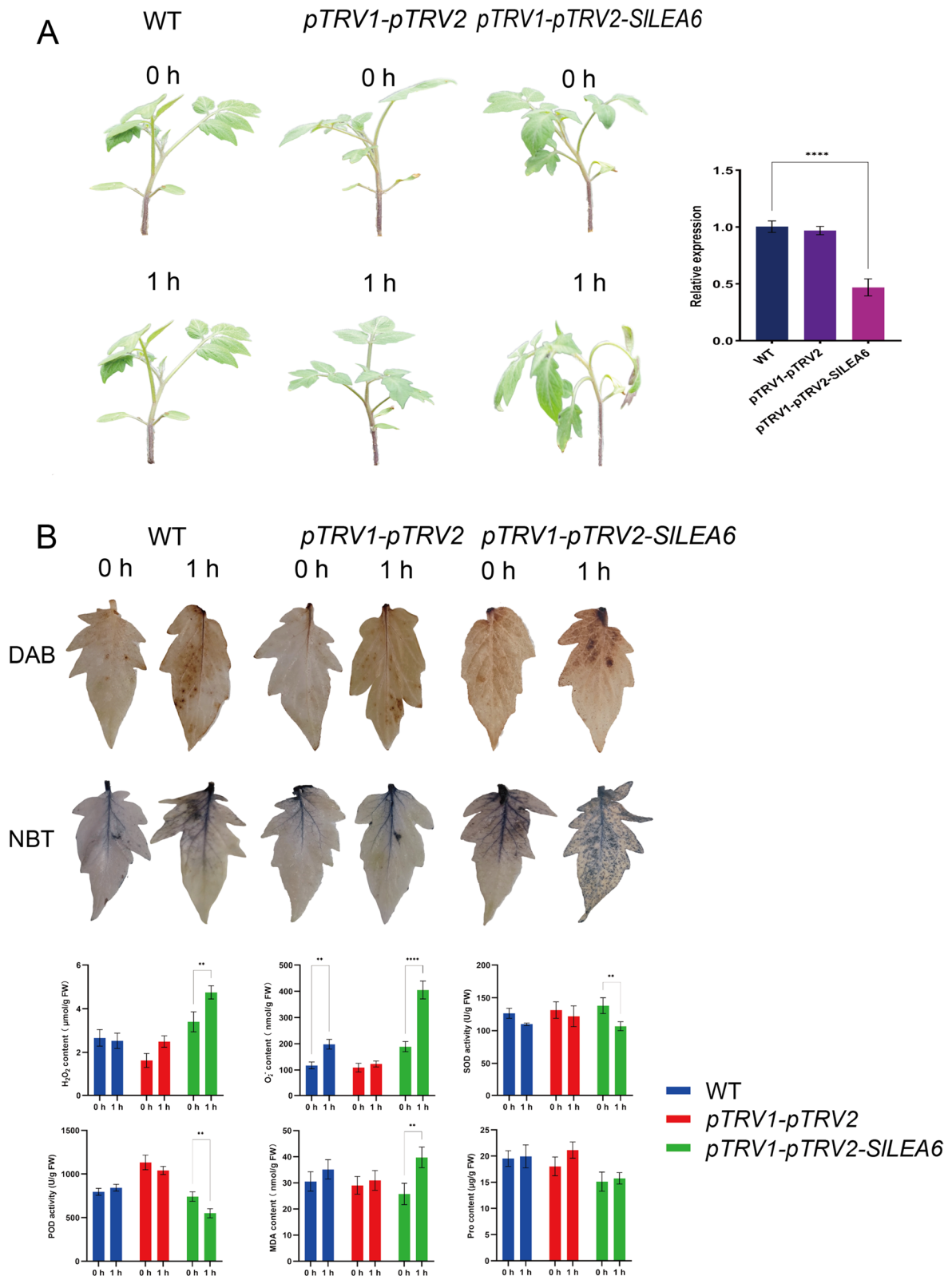


Fig. 10 (See legend on previous page.)

drought stress, the superoxide anion (O_2^-), hydrogen peroxide (H_2O_2), malondialdehyde (MDA), and proline (Pro) contents were increased in the leaves of all plants compared to those at 0 h. However, none of the controls showed significant differences, except for the O_2^- content of the negative control, indicating that the control treatment resulted in relatively less cell membrane and biomolecular damage. In contrast, the leaves of *pTRV1-pTRV2-SILEA6* plants showed highly significant differences in the accumulation of O_2^- , H_2O_2 , and MDA (but not of Pro), the contents of which were significantly higher than those in the negative and positive controls. Furthermore, superoxide dismutase (SOD) and peroxidase (POD) activities were highly significantly reduced, suggesting that the antioxidant and ROS defense systems in *pTRV1-pTRV2-SILEA6* plants had a greater impact, which resulted in their inability to promptly catalyze degradation and increase oxidation, causing more severe cell damage. The above-mentioned phenotypic, physiological, and biochemical measurements suggested that silencing *SILEA6* reduced the tolerance of tomato plants to drought stress and that it may be a candidate gene involved in drought stress response.

Expression analysis of drought-induced *SIWRKY* genes in *SILEA6*-silenced lines under drought stress

Given that *WRKY* TFs played a key role in drought stress tolerance, we further analyzed the expression pattern of drought-induced *SIWRKYs* under drought stress after constructing an interaction network between *SILEA* and *SIWRKY* proteins to obtain information on *SILEA6*-silenced strains involved in drought stress response. This analysis was conducted to explore their interaction and verify the potential regulatory functions. The selected drought-induced *SIWRKYs* were reported from previous studies [50] and consisted of eight genes, which were divided into two groups (upregulated and downregulated group), each containing the four genes with the strongest upregulation (*SIWRKY3/30/58/72*, corresponding to 13.27, 14.42, 125.37, and 36.76-fold changes, respectively) and downregulation (*SIWRKY33/39/45/46*, corresponding to -3.73, -3.25, -3.61, and -3.27-fold changes, respectively).

The expression patterns of drought-induced *SIWRKYs* were different in *SILEA6*-silenced lines under drought stress (Additional file 16: Figure S13). Compared with that at 0 h, *SIWRKY3* in the upregulated group was not upregulated, but rather was downregulated, and there was no significant difference between the silenced lines and WT. *SIWRKY30* and *SIWRKY72* showed significant upregulation trends; however, there was no significant difference between the silenced line and WT. *SIWRKY58* showed a significant upregulation trend and there was a

significant difference between the silenced lines and WT. Upregulation of *SIWRKY58* was significantly lower in the *SILEA6*-silenced lines than in WT. Overexpression of *GhWRKY41* (i.e., the nearest homolog of *SIWRKY58*) in *Nicotiana benthamiana* has been shown to result in tolerance to drought and salt stress by enhancing stomatal closure and regulating ROS scavenging [52]. The antioxidant and ROS defense systems of *SILEA6*-silenced lines were more negatively affected by ROS burst, implying that it positively regulated the expression of downstream drought-responsive genes to enhance drought tolerance in tomato plants during drought stress.

Compared with that at 0 h, *SIWRKY39* in the originally downregulated group was not downregulated, but rather was significantly upregulated; *SIWRKY33/46* were downregulated (but not significantly); and *SIWRKY45* was significantly downregulated. Their expression was not significantly different between the *SILEA6*-silenced lines and WT. Under drought stress, only *SIWRKY45* expression was significantly reduced, suggesting that it may be a negative regulator of the response to drought stress in tomato plants; however, this requires further validation through functional analysis. Except for *SIWRKY3/39*, for which the expression patterns were the opposite to those originally reported [50], the expression patterns of the other *SIWRKYs* were largely consistent with the trends in the previously divided groups, indicating the reliability of the quantitative reverse transcription-polymerase chain reaction (qRT-PCR) results.

Discussion

The *LEA* family is widespread in plants and plays important roles, particularly in protecting cells from abiotic stressors and regulating normal growth and developmental processes [7, 28]. In the present study, we identified 60, 69, 65, and 60 *LEA* genes in the genomes of *S. lycopersicum*, *S. pimpinellifolium*, *S. pennellii*, and *S. lycopersicon*, respectively. A total of 254 *LEA* genes were identified, which could be classified into eight clusters. We identified more genes and clusters than a previous study in which only 27 *LEA* genes were identified in the *S. lycopersicum* genome [14]. Continuous refinement of plant genome annotation and finer refined gene identification and classification are the most likely reasons for such differential results [19]. We compiled and summarized the main findings and analyzed them in comparison with previous results [14] to facilitate further clarification of the dynamics of the *LEA* family in tomatoes (Additional file 17: Table S4).

The phylogeny of *LEA* proteins in cultivated and wild tomato species was strongly correlated with gene structure, conserved structure domains, and motifs. These results not only support the reliability of phylogenetic

tree analysis but also further support the evolutionarily conserved features of the LEA family. Similar to previous findings, the LEA_2 cluster was the largest subfamily among all eight clusters. This phenomenon was also corroborated by studies on *Sorghum bicolor* [53], *Camellia sinensis* [54], *Populus trichocarpa* [19], and *G. hirsutum* [25]. The LEA_6 cluster was the subfamily containing the smallest number of family members, which was also consistent with studies on *T. aestivum* [55], *V. vinifera* [16], and *G. hirsutum* [25]; i.e., it represents the smallest subfamily among most other plant species [56]. As expected, our results differed from those of studies on *O. sativa* [11] and *Brassica napus* [57] wherein the subfamilies with the largest number of LEA family members were dehydrin and LEA_4, respectively. This difference is because the composition of LEA subfamilies varies considerably among species.

LEA proteins of cultivated and wild tomato species exhibited a broad cytoarchitectural hierarchical distribution pattern and were mainly located in the nucleus (39.37%) and the protoplasmic membrane (26.77%). The findings of the present study were largely consistent with those of previous studies on *A. thaliana* [10] and *G. hirsutum* [25], which showed that LEA proteins are primarily situated in subcellular areas, with a positive correlation between subcellular positioning and the effect of LEA proteins in the cell.

Typically, genes that responded to stress included fewer introns than other genes, and members of the LEA family in cultivated and wild tomato species had relatively few introns, with 37.80% members belonging to the no-intron family. These findings are similar to those of previous studies into LEA families in other species [13, 19]. It may be that the LEA genes of cultivated and wild tomato species experienced a significant loss of introns during their evolution. Furthermore, the absence of introns reduces the deleterious effects on gene expression, i.e., it reduces the delay in transcript production and additional energy consumption, thus promoting rapid expression of LEA genes in cultivated and wild tomato species and improving adaptation to adversity [58–61].

Most LEA proteins in the same cluster shared one or more conserved structural domains and motifs, which indicated that they may play an important role in the primary and specific functions of the clade. Conserved structural domains are closely related to the physiological functions of proteins and are the structural basis of protein physiological functions. However, the structure varied greatly between the different branches, indicating that LEA proteins of cultivated and wild tomato species had functional complexity. Overall, it can be inferred with high probability from the distribution of conserved

motifs that genes containing the same motifs are generated by the evolution of gene extensions within the same clade or taxon. Because different clusters consisted of different conserved motifs, it can also be preliminarily determined that they evolved from different ancestors with different conserved motifs [14].

In the present study, a phylogenetic tree analysis of LEA proteins among different model species was performed to elucidate the evolutionary relationships among the LEA families of cultivated and wild tomato species, *A. thaliana*, and *O. sativa*. The clustering of members into the same subfamily among different species implied that they have similar origins and evolutionary relationships. The different subfamilies contained LEA members from *O. sativa*, *A. thaliana*, and cultivated and wild tomato species, and their origins presumably appeared before the divergence of monocotyledons and dicotyledons [10, 11].

All LEA genes of cultivated and wild tomato species were distributed in a non-random pattern on their respective genomic chromosomes; they were more densely distributed on chromosomes 1 and 2, accounting for 32.68% of the total. In contrast, only *S. pimpinellifolium* had a single LEA gene distributed on chromosome 5. This layout makes it easier to associate that the genomes of cultivated and wild tomato species may have undergone multiple genome-wide replication events of chromosomal rearrangements and fusions during evolution, which allowed the expansion of numerous gene families [62, 63].

Universally, source drivers of gene family amplification mainly include segmental duplication, tandem duplication, and transposition events [64]. Our results showed that tandem duplication and segmental duplication events coexist in the LEA family and are the main mechanisms responsible for the increase in the number of LEA. Tandem duplication was much less frequent than segmental duplication, suggesting that the latter plays a major role in promoting LEA amplification. This mechanism of LEA amplification, wherein segmental duplication predominates, was similar to findings in *G. hirsutum* [25] and *P. trichocarpa* [19] and different from those in *A. thaliana* [10, 65] and *T. aestivum* [55]. In general, after a duplication event, a portion of duplicated genes is removed from the genome as a result of random mutations and loss of functions, and another portion is retained through functional differentiation [10]. This rapid expansion of the LEA family has adaptive implications for establishing desiccation tolerance [66]. The emergence of new genes is inseparable from the generation of replication events, and the diversity of gene functions requires the continuous replenishment of new genes, which can further help plants to resist adversity stress in the process of continuous evolution [59].

Usually, the AAs in proteins are considered to be under different selective pressures and are often measured by different Ka/Ks ratios, which can provide an important theoretical reference for species selection and evolution [67]. The Ka/Ks ratios of >1 , 1 , and <1 represent positive selection, neutral evolution, and purifying selection, respectively [68]. Our results indicated that the mean ratios of all segmental duplication and tandem duplication gene pairs were <1 , indicating that the *LEA* genes of cultivated and wild tomato species have experienced purifying selection through evolution, which may play an important role in sustaining the consistent stability of the biological structure of *LEA* genes over time. Because the Ks value reflects the timing of homologous gene production to some extent, we used this value to make predictions regarding the dispersion time of *LEA* duplication events [69]. Based on the dispersion rate of 1.50×10^{-8} synonymous substitution rate loci per year, it can be deduced that segmental duplication of *LEA* genes occurred between 31.18 and 149.38 MYA and tandem duplication of *LEA* genes occurred between 3.17 and 92.64 MYA. It has been postulated that a minimum of two large-scale genome duplication events occurred during the evolution of the plant kingdom [69, 70]. This may imply that the expansion of *LEA* family members in cultivated and wild tomato species also contributed to these processes. Furthermore, we predicted that the tandem replication events of the *SpLEA55/SpLEA56*, *SILEA53/SILEA54*, and *SlydLEA50/SlydLEA51* gene pairs occurred at approximately 3.17, 3.33, and 3.94 MYA, respectively, i.e., after the tomato–*S. tuberosum* division [49]. Furthermore, similar synteny relationships existed between cultivated and wild tomato species for *LEA* genes. The majority of *LEA* genes formed approximately three synteny gene pairs, suggesting a conserved evolutionary role of the *LEA* family in cultivated and wild tomato species. Interestingly, more synteny relationships existed among *LEA* genes in monocotyledons than in dicotyledons. Some *LEA* genes in 10 species formed 2–5 synteny gene pairs, which may have played an important role in the *LEA* family evolution.

Subsequently, by constructing synteny maps of *SILEA* genes with the *LEA* genes of two monocotyledons (*Z. mays* and *O. sativa*) and two dicotyledons (*A. thaliana* and *S. tuberosum*), we further elucidated the sequence similarity of *SILEA* genes with related *LEA* genes and homologous genes of other species. The synteny profiles of *SILEA* genes were largely consistent with those of the remaining four species. We identified 14, 27, 47, and 57 repeat events in *Z. mays*, *O. sativa*, *A. thaliana*, and *S. tuberosum*, respectively. *SILEA* genes had more synteny gene pairs with dicotyledons than with monocotyledons.

In addition, *SILEA* genes had no homologous genes in *Z. mays* and was the least homologous with *Z. mays*; had homologous genes in *O. sativa*, *A. thaliana*, and *S. tuberosum*; and was the most homologous with *S. tuberosum*. This was most likely because they belong to the same family (Solanaceae).

Cis-acting regulatory elements are required for gene regulation and expression processes. In particular, rapid and accurate identification of genes associated with specific functions in processes related to plant resistance to adversity and growth and development cannot be achieved without the aid of *cis*-acting regulatory element studies [71]. The *cis*-acting regulatory elements of *LEA* genes that were screened in the cultivated and wild tomato species were classified into four major categories—light response; abiotic stress environment or external stress response; phytohormone response; and *cis*-acting regulatory elements that participate in plant growth and development. In addition, some *LEA* genes also contained W-box motifs, which are combining locus for *WRKY* TFs. Among them, the first three major classes of *cis*-acting regulatory elements are widely available, and the promoter regions of all *LEA* genes contain several *cis*-acting regulatory elements from these three classes. This evidence suggested that members of the *LEA* family in cultivated and wild tomato species may be associated with multiple phytohormone responses and play key roles in abiotic stress defense as well as stress, wound response, and anaerobic induction. These inferences are consistent with the findings of previous studies on *cis*-acting regulatory elements of *LEA* family members in *P. trichocarpa* [19], *T. aestivum* [55], and *G. hirsutum* [25]. Of the *LEA* promoters in *A. thaliana*, 82% and 69% were reported to contain ABRE or LTRE *cis*-acting regulatory elements, respectively, in previous studies; among them, approximately 76.2% and 34.3% of *LEA* genes were significantly induced by ABA and low temperature, respectively [10]. In *Prunus mume*, 83.3% of the *LEA* promoters contained MYB *cis*-acting regulatory elements, of which approximately 63.3% of the *LEA* genes were induced by ABA [72]. These studies provide strong support for our inference. It is evident that the *LEA* genes of cultivated and wild tomato species play a critical role in stress resistance; these findings warrant further in-depth study.

Since gene function was visualized as changes in gene expression patterns, we first used normalized count-per-million (CPM) values to observe changes in the regulatory patterns of *LEA* genes in cultivated and wild tomato species under normal conditions. A differential regulation pattern was observed between them, and this up and downregulation may have been closely related to their abiotic stress tolerance-related expression patterns. Meanwhile, we analyzed the transcriptional expression profiles

of *LEA* genes under salt, drought, and high-temperature stressors, which indicated the complexity and diversity of different gene regulation patterns in *LEA* genes in cultivated and wild tomato species. Thereafter, we examined the spatiotemporal expression differences among *SILEA* genes in various tissues and during different developmental periods using transcriptional profiling data. *SILEA* genes were expressed in multiple tissues, and expression levels varied greatly between tissues and periods. Overall, the expression of *SILEA* genes was generally low in meristematic tissues, leaves, and flowers and generally high in seeds and fruits; this represented a tissue-specific expression profile that implies their involvement in specific tissue and developmental processes. Our results are supported by studies into the role of *LEA* genes in protecting *Citrus unshiu* [73], *S. tuberosum* [74], *Cucumis melo*, and *Citrullus lanatus* [75] against abiotic stressors while also functioning in normal cellular functions and fruit developmental stages. We also observed that certain *SILEA* genes exhibited unique expression profiles and that their persistent high expression in specific organs may play a key role in the growth and development of tomato plants, in particular in regulating seed maturation and fruit maturation from immaturity to ripening. This finding was similar to the results of previous findings that showed that *LEA* proteins accumulate at high levels during late seed maturation and are absent after germination [55, 76]

Moreover, we also investigated the spatiotemporal expression modes of *SILEA* genes in response to various stressors and phytohormone treatments using qRT-PCR in an attempt to elucidate the possible effects of various *SILEA* family members. *SILEA* expression patterns were affected by a wide range of abiotic stressors and phytohormone treatments, and certain *SILEA* genes were rapidly induced and persistently highly expressed during specific periods, which demonstrated that they performed particular functions under different stressors and hormone treatments. Overall, *SILEA* expression was more sensitive to high salt and drought stressors than extreme temperature. We also observed that some *SILEA* genes showed similar expression patterns in response to various abiotic stressors and phytohormone treatments. *SILEA6/27* may be the key genes in the *SILEA* family involved in defense and stress response to external stressors. *TaLEA3-2/6-3* were previously reported to be strongly responsive to drought and salt stress in *T. aestivum* [55]. *ZmLEA31* gene in *Z. mays* was upregulated under drought, high-salt, low-temperature, and IAA treatments [12], which was similar to our findings. Furthermore, overexpression of *LEA* genes resulted in enhanced salt tolerance, drought tolerance, sensitivity to ABA, osmotic stress tolerance, and oxidative stress in transgenic *A. thaliana*

[35, 36], *O. sativa* [37], and *N. tabacum* [38]. In the present study, we found that silencing *SILEA6* significantly increased the H_2O_2 , O_2^- , and MDA contents in leaves while significantly reducing SOD and POD activities. This caused a greater impact on biological antioxidant and ROS defense systems, increased oxidation (thereby increasing the degree of cellular damage), and led to an increase in leaf wilting, reducing the tolerance of tomato plants to drought stress. The *SILEA6* identified using VIGS technology can be used as a candidate gene for subsequent studies to improve the current understanding of drought tolerance in tomato plants.

We performed GO annotation analysis of 254 *LEA* genes in cultivated and wild tomatoes and 60 *SILEA* genes to obtain the BP, MF, and CC of all *LEA* genes. A series of BP related to *SILEA* genes were clarified and were mostly predicted to function in response to desiccation and water deficit as well as to abiotic stimuli and stressful environments. The enrichment results also revealed that the most abundant functions in their CC involved the membrane, cellular, and cytoplasmic fractions. This was similar to the analysis of the *LEA* family in *G. hirsutum* and is presumably consistent with their function in securing membranes and enzymes to maintain cellular activity under drought stress [25, 32].

WRKY TFs are one of the most populated families of transcriptional regulators that regulate plant growth and development and can interact with all kinds of proteins or TFs to perform their corresponding biological functions and regulate plant defense responses. *WRKY* proteins can bind to the W-box in the promoter region of target genes to activate or inhibit the expression of downstream genes, thereby regulating their stress response [77]. However, the functional mechanisms underlying the interaction between *SILEA* and *SIWRKY* proteins remain poorly understood. The protein interactions regulatory network constructed in this study showed that *S. lycopersicum* and *A. thaliana* *LEA* proteins interacted with multiple *WRKY* proteins to regulate a variety of physiological processes. Our analysis of the expression of drought-induced *SIWRKY* genes in *SILEA6*-silenced lines under drought stress showed that the expression patterns of the eight selected *SIWRKYs* differed. Only *SIWRKY45* expression was significantly reduced under drought stress, indicating that it may be a negative regulator of tomato plant response to drought stress; however, this requires further verification via functional analysis. In addition, the increased expression of *SIWRKY58* was significantly lowered in *SILEA6*-silenced lines compared to that in WT; overexpression of *GhWRKY41* (the nearest homolog of *SIWRKY58*) in *N. benthamiana* has been shown to confer drought and salt-stress tolerance by enhancing stomatal closure and regulating ROS clearance [52]. Contrastingly,

the *SILEA6*-silenced lines did have a serious negative impact on antioxidant and ROS defense systems, which implied that they positively regulate the expression of downstream drought-responsive genes under drought stress to enhance drought resistance in tomato plants.

To date, AtWRKY19 [78], AtWRKY22/29 [79], AtWRKY38/48/62 [80], and AtWRKY72 [81] in *A. thaliana* have been reported to be involved in biotic stress defense; AtWRKY25/26/33 [82] and AtWRKY39 [83] respond to high-temperature stress; AtWRKY34 [84] responds to low-temperature stress; AtWRKY25/33 [82], AtWRKY46 [85], and AtWRKY6/8/22/30/39/48/53/75 [86] respond to salt, drought, and oxidative stress; AtWRKY53/54/70 [87] and AtWRKY57 [88] are involved in regulating leaf senescence mechanisms, and AtWRKY27 [89] plays a key role in pollen growth and development. In *S. lycopersicum*, SIWRKY1/11 [90], SIWRKY39 [91], and SIWRKY70 [92] are involved in biotic stress defense; SIWRKY12/13/23/50/51 [93] respond to low-temperature stress; SIWRKY39 [91], SIWRKY23/31/32/33/41/57/74 [94], SIWRKY58 [50], and SIWRKY81 [95] respond to salt and drought stress; and SIWRKY23 [96] and SIWRKY16/17/53/54 [97] play key roles in growth and fruit ripening. Meanwhile, SIWRKY39/40/43/45/46 are closely homologous to AtWRKY18/40/60 in Class IIa and have been shown to respond to both abiotic and biotic stressors [98]. These previous findings on AtWRKY and SIWRKY proteins imply that some of the AtLEA and SILEA proteins identified in the present study that interact with WRKY proteins may also have corresponding biological functions. Nonetheless, it should be emphasized that slight variations in the DNA binding domain may have critical implications for binding specificity and that sequence homologs may be very similar but have various functions [99].

Conclusions

In summary, the present study provides a comprehensive and systematic study of the LEA family in cultivated and wild tomato species. We identified 254 LEA genes divided into eight clusters with different physicochemical properties. They were non-randomly distributed on chromosomes, and promoter regions contained more *cis*-acting regulatory elements linked to abiotic stress tolerance and phytohormone response. Evolutionary analysis revealed replication events, purifying selection, divergence time, intraspecific and interspecific homology, and synteny of LEA genes. Transcriptional profiling and gene expression pattern analyses indicated specific roles of certain SILEA genes during growth and development or exposure to various abiotic stressors and phytohormone treatments. Gene functional annotations were then developed and protein-based interaction networks were constructed to

further elucidate the gene and functional proteomic regulatory networks. Finally, *SILEA6* was shown to be a candidate gene involved in response to drought stress. Our results provide strong support for an improved appreciation of the evolutionary history and function of LEA genes and lay the foundation for subsequent in-depth studies of the molecular and biological functions of this family.

Materials and methods

LEA identification and characterization

Genome-wide data were obtained from the Sol Genomics Network (SGN) database (<https://solgenomics.net/>), and based on a hidden Markov model (HMM), the conserved domains of LEA proteins were obtained. The potential HMM profile was utilized as an inquiry separately against the genomes of cultivated (*S. lycopersicum*) and wild tomato species (*S. pimpinellifolium*, *S. pennellii*, and *S. lycopersicoides*) via the HMMER search (<https://www.ebi.ac.uk/Tools/hmmer/>) [100]. Subsequently, the clustering comparison was first performed and all redundant sequences were discarded. The Simple Modular Architecture Research Tool (SMART) database (<http://smart.embl-heidelberg.de/>) [101] and Pfam database (<http://pfam.xfam.org/>) [102] were then used to confirm the existence of LEA domains. Physicochemical parameters such as MW and pI of LEA proteins were anticipated using the ExPASy server tool (<http://www.web.expasy.org/protparam/>). The subCELLular LOCALization predictor (CELLO) website (<http://cello.life.nctu.edu.tw/>) was used for analyzing the subcellular location of all LEA proteins.

Phylogeny, gene structure, conserved domains, and motif analyses of LEA proteins

LEA protein sequences were clustered to construct an evolutionary tree, and the coding regions and genomic sequences were compared to analyze the gene structure. The conserved structural domains and motifs of proteins were evaluated using the HMMER search and Multiple Expectation Maximization for Motif Elicitation (MEME) website. (<https://meme-suite.org/meme/tools/meme>) Finally, all of these were visualized using TBtools [103].

Phylogenetic tree analysis of LEA proteins

To investigate the phylogenetic relationships between the model crops (*A. thaliana* and *O. sativa*) and the LEA family, we first obtained the protein sequences of AtLEA and OsLEA after screening and identification using the HMMER search. Then, a phylogenetic tree analysis was performed using MEGA 11 software [104] and the

maximum likelihood (ML) method with bootstrap analysis for 1000 repetitions.

Chromosome distribution, replication events, purification pressure, and synteny analysis

The DNA sequences of each member of the *LEA* family were localized to the genomes of cultivated and wild tomato species, and the gene distribution on chromosomes was calculated and mapped using TBtools. Gene duplication events for *LEA* genes were calculated using the MCSanX toolkit [105]. We used the latest KaKs_Calculator 3.0 software to calculate the K_a and K_s ratios for the duplicate gene pairs of *LEA* [106]. Then, the replication date of each gene pair was calculated using the K_s value by applying the formula $T = K_s / 2\lambda$, where λ was assumed to be 1.5×10^{-8} synonymous/substitution site/year [69]. Single-copy homologous genes were obtained using diamond comparison in the OrthoFinder 2.5.2 software [107], evolutionary divergence trees of different species were constructed using the RAxML 8.0 tool [108], and fossil times were added to the evolutionary trees using the TimeTree database (<http://www.timetree.org/>). Then, the mcmctree command in PAML 4.9 software [109] was used to estimate the divergence times of different species. Moreover, the synteny of *LEA* genes between different species was calculated, analyzed, and visualized using the dual synteny plotter module in TBtools.

Analysis of the cis-acting regulatory elements of *LEA* genes

Sequences in the range of 2000 base pairs (bp) upstream of the start codon (ATG) in *LEA* genes were downloaded from the genomes of cultivated and wild tomato species and analyzed using the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>) for promoter cis-acting regulatory element screening.

Analysis of *LEA* expression pattern

The CPM values for *LEA* genes of *S. lycopersicum* and its wild relatives (*S. pimpinellifolium*, *S. habrochaites*, and *S. pennellii*) were obtained from the Maloof Lab database (<http://malooflab.phytonetworks.org/apps/tomato-expression/>). Transcriptomic data of cultivated and wild tomato species under stress conditions were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Salt stress corresponds to the GEO accession GSE16401 [110], and drought and high-temperature stressors correspond to the GEO accession GSE22304 [111]. *SILEA* expression data in various tissues (including seeds, roots, meristem, leaves, flowers, and fruits) are available on the TomExpress database, i.e.,

a unified tomato RNA-Seq platform (<http://tomexpress.toulouse.inra.fr/>). Cluster heat map analysis was conducted using the OmicShare tool, i.e., a free online data analysis platform (<https://www.omicshare.com/tools>).

Plant material and treatments

S. lycopersicum var. M82 was used as experimental material. Seeds were cultivated in a 3:1 (w/w) mixture of nutrient soil and vermiculite, and plants were grown in an artificial greenhouse at a controlled temperature of 25 ± 1 °C, relative humidity of $50 \pm 10\%$, and a 16 h/8 h photoperiod (daytime: 08:00–00:00). After 4 weeks, seedlings were sprayed with 100 μ M ABA, 100 μ M MeJA, 50 μ M *rac*-GR24, and 100 μ M GABA for analysis of the response to phytohormone treatment. Seedlings were sprayed with double distilled water (ddH₂O) in the control treatment. Each solution was sprayed evenly on each leaf, and spraying was stopped before the solution dripped from the leaf. Similarly, for 4-week-old seedlings, abiotic stress responses were analyzed after treatment with 200 mM NaCl and 300 mM mannitol. The corresponding solution was slowly poured from the top of the seedlings into the planting trays with holes at the bottom, and watering was ceased when the solution began to leak from the holes, indicating saturation. Seedlings were placed in 4 °C and 35 °C light-temperature incubators for low and high temperature treatments, respectively, and the control group followed the standard light and temperature as described previously. The control and treated seedlings were collected at 0, 1, 3, 6, 12, and 24 h after processing, whereby the top first fully expanded leaves of three uniformly grown seedlings of each treatment were mixed as biological replicates and replicated three times. Samples were frozen in liquid nitrogen and stockpiled in an ultra-low temperature refrigerator at -80 °C until use.

RNA extraction, qRT-PCR, and statistical analysis

Total RNA was extracted from the harvested samples using the Polysaccharide Polyphenol Plant Total RNA Extraction Kit (DP441, Tiangen, Beijing, China) according to the manufacturer's instructions. First-strand cDNA synthesis was completed using the 5× All-In-One RT MasterMix (with AccuRT Genomic DNA Removal Kit) (G492, ABM, Vancouver, Canada) according to the manufacturer's instructions. Primers specific for *SILEA* and *SIWRKY* genes were designed using the NCBI primer design tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome), and *Actin* (*Solyc03g078400.2*) was used as an internal reference control. qRT-PCR was performed on a LightCycler® real-time fluorescent quantitative PCR system (Roche, Basel, Switzerland) using ChamQ Universal SYBR qPCR Master Mix (Q711, Vazyme, Nanjing, China) and the

primers listed in Additional file 18: Table S5. The amplification process consisted of one cycle of 95 °C pre-denaturation for 30 s, followed by 40 cycles of 95 °C denaturation for 5 s, and 60 °C annealing and extension for 30 s. Finally, one cycle of 95 °C for 15 s, 60 °C for 60 s, and 95 °C for 15 s was performed for melting curve acquisition. For each sample, three technical replicates were performed to calculate the mean cycle threshold (C_t) value. Relative expression was calculated using the $2^{-\Delta\Delta C_t}$ method [112]. All data were subjected to variance analysis; the mean of values from all treatments is presented. Three independent biological replicates were used for each determination. A one-way analysis of variance (ANOVA) was performed using GraphPad Prism for Windows (version 9.0.0, GraphPad Software, San Diego, CA, USA), followed by Dunnett's multiple comparisons test.

GO and protein interaction network analyses

GO analysis of *LEA* genes in cultivated and wild tomato species was performed using the OmicShare tool. The ensemble IDs of *LEA* genes were loaded into the program and used to complete the three steps of functional analysis—(i) mapping to each term of the GO database (<http://www.geneontology.org/>) and calculating the number of genes per term; (ii) applying hypergeometric tests to recognize GO entries that were remarkably enriched in genes contrasted to the total genomic context; and (iii) using a corrected- P value ≤ 0.05 as the threshold. GO terms that satisfied this requirement were recognized as being significantly enriched in genes. Thus, the MF, CC, and BP involved in the genes were described separately.

The protein sequence of SIWRKY was acquired from the *S. lycopersicum* genome database and localized to the AtWRKY protein on The *Arabidopsis* Information Resource (TAIR) website (<https://www.arabidopsis.org/>) using the BLASTP tool. Subsequently, the interactions of SILEA and SIWRKY proteins and of AtLEA and AtWRKY proteins were predicted according to String 11.5, and their interaction networks were plotted. Sequence alignment of SIWRKY and AtWRKY proteins was performed using MUSCLE 5.1 software.

VIGS vector construction, gene silencing, and determination of physiological indices

VIGS vectors for the target genes were designed and constructed using the SGN VIGS tool (<https://vigs.solgenomics.net/>), following the procedure reported in the literature [113]. Briefly, *pTRV1*, *pTRV2* empty, and *pTRV2* recombinant vectors were cloned into *Agrobacterium tumefaciens* GV3101 (pSoup-p19), cultured overnight in a shaker, centrifuged to collect the organisms, and adjusted with a resuspension solution to a final concentration corresponding to $OD_{600} = 1.0$. The *pTRV1-pTRV2* empty vector and *pTRV1-pTRV2*

recombinant vector resuspension were mixed at a ratio of 1:1 (v/v) and incubated at 28 °C for 3 h in darkness and then infiltrated into the abaxial surface of the leaves of 4-week-old tomato plants using a 1-mL syringe without the needle. After 12 h of dark treatment, a normal photoperiod was provided. Untreated plants served as the negative controls and plants inoculated with *pTRV1-pTRV2* empty served as the positive controls. Abiotic stress conditions were applied 2 weeks later as previously described. Top leaves were collected from the same parts at 0 h and 1 h post-stress and stained using an NBT solution (0.1%, pH 7.8) (SL18061, Coolaber, Beijing, China) and a DAB solution (1 mg/mL, pH 3.8) (SL1805, Coolaber, Beijing, China), respectively. Six test kits (SA-1-G, H₂O₂-1-Y, MDA-1-Y, PRO-1-Y, SOD-1-Y, and POD-1-Y, Comin, Suzhou, China) were used sequentially to detect O₂⁻, H₂O₂, MDA, Pro, SOD, and POD activity/content according to the manufacturer's instructions; three replicates were included.

Abbreviations

LEA	Late embryogenesis abundant
Pfam	The protein families
IDs	Identity documents
ROS	Reactive oxygen species
TFs	Transcription factors
AA	Amino acids
ABA	Abcisic acid
GO	Gene ontology
VIGS	Virus-induced gene silencing
SMP	Seed maturation protein
MW	Molecular weight
pI	Isoelectric point
GRAVY	Grand average of hydropathy index
UTR	Untranslated regions
Ka	Non-synonymous substitution
Ks	Synonymous substitution
MYA	Million years ago
MeJA	Methyl jasmonate
GA	Gibberellins
IAA	Auxin
GABA	γ -Aminobutyric acid
BP	Biological processes
MF	Molecular functions
CC	Cellular components
MEKK	MAPK/ERK kinases
O ₂ ⁻	Superoxide anion
H ₂ O ₂	Hydrogen peroxide
MDA	Malondialdehyde
Pro	Proline
SOD	Superoxide dismutase
POD	Peroxidase
CPM	Count-per-million
SGN	Sol Genomics Network
HMM	Hidden Markov model
SMART	Simple modular architecture research tool
MEME	Maximization for motif elicitation
ML	Maximum likelihood
GEO	Gene expression omnibus
ddH ₂ O	Distilled water
qRT-PCR	Quantitative real-time PCR
C_t	Cycle threshold
TAIR	The <i>Arabidopsis</i> information resource

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-022-03953-7>.

Additional file 1: Table S1. Identification and characterization of LEAs in cultivated and wild tomato species.

Additional file 2: Table S2. Basic information on the AtLEAs and OsLEAs re-predicted, identified, and renamed in this study.

Additional file 3: Table S3. The list of replication events, *Ka/Ks* ratios, and their replication time for LEAs.

Additional file 4: Figure S1. Expression patterns of *SILEAs* under normal conditions (control). Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 5: Figure S2. Expression patterns of *SILEAs* under simulated drought stress (300 mM mannitol). Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 6: Figure S3. Expression patterns of *SILEAs* under simulated high salt stress (200 mM NaCl). Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 7: Figure S4. Expression patterns of *SILEAs* under simulated high-temperature stress (35°C). Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 8: Figure S5. Expression patterns of *SILEAs* under simulated low-temperature stress (4°C). Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 9: Figure S6. Expression patterns of *SILEAs* in response to ddH₂O control treatment. Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 10: Figure S7. Expression patterns of *SILEAs* in response to 100 μ M ABA treatment. Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 11: Figure S8. Expression patterns of *SILEAs* in response to 100 μ M MeJA treatment. Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 12: Figure S9. Expression patterns of *SILEAs* in response to 50 μ M *rac*-GR24 treatment. Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 13: Figure S10. Expression patterns of *SILEAs* in response to 100 μ M GABA treatment. Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 14: Figure S11. GO analysis of 254 LEAs in this study.

Additional file 15: Figure S12. Phenotypic analysis under abiotic stress conditions after silencing of *SILEA6* (**a**: salt stress; **b**: high temperature stress; **c**: low temperature stress).

Additional file 16: Figure S13. Expression patterns of eight drought-induced *SIWRKYs* in *SILEA6* silenced lines under drought stress. Three independent biological replicates were included to calculate the mean. Error bars show the SD of the three biological replicates. Values represent mean \pm SD. Statistical significance of the differences was confirmed using Dunnett's multiple comparisons test ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$).

Additional file 17: Table S4. Comparison and summary of the main study results.

Additional file 18: Table S5. The primer sequences used in this study for qRT-PCR.

Acknowledgements

We thank the useful technical assistance provided by our colleagues in the laboratory. We also thank the rigorous evaluation of the manuscript made by the editors and reviewers.

Authors' contributions

QY and JW designed the experiment and methodology. CJ and BG carried out most experimental work. BW and XL conducted data analysis. TY and NL finished plant material treatment. QY, JW, CJ, and BG wrote the manuscript. All authors contributed to experimental design and data analysis, commented on the manuscript, and gave final approval for publication.

Funding

This work was supported by the Key Programs for Crop Important Traits Formation and Cutting-edge Technologies in Biological Breeding (jxjny-wdzc-2022001–8), the Special Incubation Project of Science & Technology Renovation of Xinjiang Academy of Agricultural Sciences (xjkcpy-2021001), and the Key Research and Development Task Special Project of Xinjiang Uygur Autonomous Region (2022B02002).

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information file. Transcriptome sequencing data used during the study were obtained from Maloof Lab (<http://maloooflab.phytonetworks.org/apps/tomato-expression/>), TomExpress (<http://tomexpres.toulouse.inra.fr/>), and GEO databases (GEO accession GSE16401 and GSE22304), respectively.

Declarations

Ethics approval and consent to participate

The plant material used in this study was cultivated tomato variety M82, grown in the laboratory of the Institute of Horticultural Crops, Xinjiang Academy of Agricultural Sciences, Urumqi, China, and no permits were required to collect plant samples. In the present study, all methods were carried out following relevant guidelines and regulations. Ethical approval or consent was not required for this study because no endangered or protected species were involved.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflicts of interest.

Author details

¹Institute of Horticulture Crops, Xinjiang Academy of Agricultural Sciences (Key Laboratory of Genome Research and Genetic Improvement of Xinjiang Characteristic Fruits and Vegetables), Urumqi, China. ²College of Life Science and Technology, Xinjiang University, Urumqi, China. ³College of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi, China.

Received: 1 July 2022 Accepted: 21 November 2022

Published online: 19 December 2022

References

- Ahuja I, de Vos RC, Bones AM, Hall RD. Plant molecular stress responses face climate change. *Trends Plant Sci.* 2010;15(12):664–74. <https://doi.org/10.1016/j.tplants.2010.08.002>.
- Gechev T, Petrov V. Reactive oxygen species and abiotic stress in plants. *Int J Mol Sci.* 2020;21(20):7433. <https://doi.org/10.3390/ijms21207433>.
- Zhu JK. Abiotic stress signaling and responses in plants. *Cell.* 2016;167(2):313–24. <https://doi.org/10.1016/j.cell.2016.08.029>.
- Campo S, Baldrich P, Messegue J, Lalanne E, Coca M, San SB. Overexpression of a calcium-dependent protein kinase confers salt and drought tolerance in rice by preventing membrane lipid peroxidation. *Plant Physiol.* 2014;165(2):688–704. <https://doi.org/10.1104/pp.113.230268>.
- Shah WH, Rasool A, Saleem S, Mushtaq NU, Tahir I, Hakeem KR, et al. Understanding the integrated pathways and mechanisms of transporters, protein kinases, and transcription factors in plants under salt stress. *Int J Genomics.* 2021;2021:5578727. <https://doi.org/10.1155/2021/5578727>.
- Candat A, Paszkiewicz G, Neveu M, Gautier R, Logan DC, Avelange-Macherel MH, et al. The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in *Arabidopsis* offers tailored protection against abiotic stress. *Plant Cell.* 2014;26(7):3148–66. <https://doi.org/10.1105/tpc.114.127316>.
- Olvera-Carrillo Y, Luis Reyes J, Covarrubias AA. Late embryogenesis abundant proteins: versatile players in the plant adaptation to water limiting environments. *Plant Signal Behav.* 2011;6(4):586–9. <https://doi.org/10.4161/psb.6.4.15042>.
- Saucedo AL, Hernandez-Dominguez EE, de Luna-Valdez LA, Guevara-Garcia AA, Escobedo-Moratilla A, Bojorquez-Velazquez E, et al. Insights on structure and function of a late embryogenesis abundant protein from *Amaranthus cruentus*: an intrinsically disordered protein involved in protection against desiccation, oxidant conditions, and osmotic stress. *Front Plant Sci.* 2017;8:497. <https://doi.org/10.3389/fpls.2017.00497>.
- Dure L, Greenway SC, Galau GA. Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. *Biochemistry.* 1981;20(14):4162–8. <https://doi.org/10.1021/bi00517a033>.
- Hundertmark M, Hinch DK. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics.* 2008;9:118. <https://doi.org/10.1186/1471-2164-9-118>.
- Wang XS, Zhu HB, Jin GL, Liu HL, Wu WR, Zhu J. Genome-scale identification and analysis of LEA genes in rice (*Oryza sativa* L.). *Plant Sci.* 2007;172(2):414–20. <https://doi.org/10.1016/j.plantsci.2006.10.004>.
- Li X, Cao J. Late embryogenesis abundant (LEA) gene family in maize: identification, evolution, and expression profiles. *Plant Mol Biol Rep.* 2015;34(1):15–28. <https://doi.org/10.1007/s11105-015-0901-y>.
- Liu D, Sun J, Zhu D, Lyu G, Zhang C, Liu J, et al. Genome-wide identification and expression profiles of late embryogenesis-abundant (LEA) genes during grain maturation in wheat (*Triticum aestivum* L.). *Genes.* 2019;10(9):696. <https://doi.org/10.3390/genes10090696>.
- Cao J, Li X. Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*). *Planta.* 2015;241(3):757–72. <https://doi.org/10.1007/s00425-014-2215-y>.
- Celik Altunoglu Y, Baloglu P, Yer EN, Pekol S, Baloglu MC. Identification and expression analysis of LEA gene family members in cucumber genome. *Plant Growth Regul.* 2016;80(2):225–41. <https://doi.org/10.1007/s10725-016-0160-4>.
- Ibrahim M, Kibar U, Kazan K, Yüksel Özmen C, Mutaf F, Demirel Aşçı S, et al. Genome-wide identification of the LEA protein gene family in grapevine (*Vitis vinifera* L.). *Tree Genet Genomes.* 2019;15(4). <https://doi.org/10.1007/s11295-019-1364-3>.
- Hand SC, Menze MA, Toner M, Boswell L, Moore D. LEA proteins during water stress: not just for plants anymore. *Annu Rev Physiol.* 2011;73:115–34. <https://doi.org/10.1146/annurev-physiol-012110-142203>.
- Tunnacliffe A, Wise MJ. The continuing conundrum of the LEA proteins. *Sci Nat-heidelberg.* 2007;94(10):791–812. <https://doi.org/10.1007/s00114-007-0254-y>.
- Cheng Z, Zhang X, Yao W, Zhao K, Liu L, Fan G, et al. Genome-wide search and structural and functional analyses for late embryogenesis-abundant (LEA) gene family in poplar. *BMC Plant Biol.* 2021;21(1):110. <https://doi.org/10.1186/s12870-021-02872-3>.
- Hunault G, Jaspard E. LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics.* 2010;11:221. <https://doi.org/10.1186/1471-2164-11-221>.
- Singh KK, Graether SP. The in vitro structure and functions of the disordered late embryogenesis abundant three proteins. *Protein Sci.* 2021;30(3):678–92. <https://doi.org/10.1002/pro.4028>.
- Battaglia M, Olvera-Carrillo Y, Garcarrubio A, Campos F, Covarrubias AA. The enigmatic LEA proteins and other hydrophilins. *Plant Physiol.* 2008;148(1):6–24. <https://doi.org/10.1104/pp.108.120725>.
- Garay-Arroyo A, Colmenero-Flores JM, Garcarrubio A, Covarrubias AA. Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J Biol Chem.* 2000;275(8):5668–74. <https://doi.org/10.1074/jbc.275.8.5668>.
- Cuevas-Velazquez CL, Reyes JL, Covarrubias AA. Group 4 late embryogenesis abundant proteins as a model to study intrinsically disordered proteins in plants. *Plant Signal Behav.* 2017;12(7): e1343777. <https://doi.org/10.1080/15592324.2017.1343777>.
- Magwanga RO, Lu P, Kirungu JN, Lu H, Wang X, Cai X, et al. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. *BMC Genet.* 2018;19(1):6. <https://doi.org/10.1186/s12863-017-0596-1>.
- Nadarajah KK. ROS homeostasis in abiotic stress tolerance in plants. *Int J Mol Sci.* 2020;21(15):5208. <https://doi.org/10.3390/ijms21155208>.
- Liu G, Xu H, Zhang L, Zheng Y. Fe binding properties of two soybean (*Glycine max* L.) LEA4 proteins associated with antioxidant activity. *Plant Cell Physiol.* 2011;52(6):994–1002. <https://doi.org/10.1093/pcp/pcr052>.
- Goyal K, Walton LJ, Tunnacliffe A. LEA proteins prevent protein aggregation due to water stress. *Biochem J.* 2005;388:151–7. <https://doi.org/10.1042/BJ20041931>.
- Grelet J, Benamar A, Teyssier E, Avelange-Macherel MH, Grunwald D, Macherel D. Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. *Plant Physiol.* 2005;137(1):157–67. <https://doi.org/10.1104/pp.104.052480>.
- Furuki T, Sakurai M. Group 3 LEA protein model peptides protect enzymes against desiccation stress. *Biochim Biophys Acta.* 2016;1864(9):1237–43. <https://doi.org/10.1016/j.bbapap.2016.04.012>.
- Hinch DK, Zuther E, Popova AV. Stabilization of dry sucrose glasses by four LEA_4 proteins from *Arabidopsis thaliana*. *Biomolecules.* 2021;11(5):615. <https://doi.org/10.3390/biom11050615>.
- Tolleter D, Hinch DK, Macherel D. A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state. *Biochim Biophys Acta.* 2010;1798(10):1926–33. <https://doi.org/10.1016/j.bbamem.2010.06.029>.
- Kovacs D, Agoston B, Tompa P. Disordered plant LEA proteins as molecular chaperones. *Plant Signal Behav.* 2008;3(9):710–3. <https://doi.org/10.4161/psb.3.9.6434>.
- Mertens J, Aliyu H, Cowan DA. LEA proteins and the evolution of the Why domain. *Appl Environ Microbiol.* 2018;84(15):15. <https://doi.org/10.1128/AEM.00539-18>.
- Jia F, Qi S, Li H, Liu P, Li P, Wu C, et al. Overexpression of *Late Embryogenesis Abundant 14* enhances *Arabidopsis* salt stress tolerance. *Biochem Biophys Res Commun.* 2014;454(4):505–11. <https://doi.org/10.1016/j.bbrc.2014.10.136>.

36. Duan J, Cai W. *OsLEA3-2*, an abiotic stress induced gene of rice plays a key role in salt and drought tolerance. *PLoS ONE*. 2012;7(9): e45117. <https://doi.org/10.1371/journal.pone.0045117>.
37. Yu J, Lai Y, Wu X, Wu G, Guo C. Overexpression of *OsEm1* encoding a group I LEA protein confers enhanced drought tolerance in rice. *Biochem Biophys Res Commun*. 2016;478(2):703–9. <https://doi.org/10.1016/j.bbrc.2016.08.010>.
38. Liu Y, Wang L, Xing X, Sun L, Pan J, Kong X, et al. ZmLEA3, a multifunctional group 3 LEA protein from maize (*Zea mays* L.), is involved in biotic and abiotic stresses. *Plant Cell Physiol*. 2013;54(6):944–59. <https://doi.org/10.1093/pcp/pct047>.
39. Lim CW, Lim S, Baek W, Lee SC. The pepper late embryogenesis abundant protein CaLEA1 acts in regulating abscisic acid signaling, drought and salt stress response. *Physiol Plantarum*. 2015;154(4):526–42. <https://doi.org/10.1111/pp1.12298>.
40. Wang L, Li X, Chen S, Liu G. Enhanced drought tolerance in transgenic *Leymus chinensis* plants with constitutively expressed wheat *TaLEA3*. *Biotechnol Lett*. 2009;31(2):313–9. <https://doi.org/10.1007/s10529-008-9864-5>.
41. Zheng J, Su H, Lin R, Zhang H, Xia K, Jian S, et al. Isolation and characterization of an atypical LEA gene (*IpLEA*) from *Ipomoea pes-caprae* conferring salt/drought and oxidative stress tolerance. *Sci Rep-UK*. 2019;9(1):14838. <https://doi.org/10.1038/s41598-019-50813-w>.
42. Wu Y, Liu C, Kuang J, Ge Q, Zhang Y, Wang Z. Overexpression of *SmlEA* enhances salt and drought tolerance in *Escherichia coli* and *Salvia miltiorrhiza*. *Protoplasma*. 2014;251(5):1191–9. <https://doi.org/10.1007/s00709-014-0626-z>.
43. Wang MZ, Li P, Li C, Pan YL, Jiang XY, Zhu DY, et al. SiLEA14, a novel atypical LEA protein, confers abiotic stress resistance in foxtail millet. *BMC Plant Biol*. 2014;14:290. <https://doi.org/10.1186/s12870-014-0290-7>.
44. Solankey SS, Singh RK, Baranwal DK, Singh DK. Genetic Expression of Tomato for Heat and Drought Stress Tolerance: An Overview. *International Journal of Vegetable Science*. 2014;21(5):496–515. <https://doi.org/10.1080/19315260.2014.902414>.
45. Loukehaich R, Wang T, Ouyang B, Ziaf K, Li H, Zhang J, et al. *SpUSP*, an annexin-interacting universal stress protein, enhances drought tolerance in tomato. *J Exp Bot*. 2012;63(15):593–606. <https://doi.org/10.1093/jxb/ers220>.
46. Loyola J, Verdugo I, Gonzalez E, Casaretto JA, Ruiz-Lara S. Plastidic isoprenoid biosynthesis in tomato: physiological and molecular analysis in genotypes resistant and sensitive to drought stress. *Plant Biol*. 2012;14(1):149–56. <https://doi.org/10.1111/j.1438-8677.2011.00465.x>.
47. Tomato GC. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*. 2012;485(7400):635–41. <https://doi.org/10.1038/nature11119>.
48. Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, et al. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat Genet*. 2014;46(9):1034–8. <https://doi.org/10.1038/ng.3046>.
49. Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, Siepel A, et al. Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. *Genetics*. 2008;180(1):391–408. <https://doi.org/10.1534/genetics.108.087981>.
50. Karkute SG, Gujjar RS, Rai A, Akhtar M, Singh M, Singh B. Genome wide expression analysis of WRKY genes in tomato (*Solanum lycopersicum*) under drought stress. *Plant Gene*. 2018;13:8–17. <https://doi.org/10.1016/j.plgene.2017.11.002>.
51. Wang Q, Wang M, Zhang X, Hao B, Kaushik SK, Pan Y. WRKY gene family evolution in *Arabidopsis thaliana*. *Genetica*. 2011;139(8):973–83. <https://doi.org/10.1007/s10709-011-9599-4>.
52. Chu X, Wang C, Chen X, Lu W, Li H, Wang X, et al. The cotton WRKY gene *GhWRKY41* positively regulates salt and drought stress tolerance in transgenic *Nicotiana benthamiana*. *PLoS ONE*. 2015;10(11): e0143022. <https://doi.org/10.1371/journal.pone.0143022>.
53. Prasad M, Nagaraju M, Kumar SA, Reddy PS, Kumar A, Rao DM, et al. Genome-scale identification, classification, and tissue specific expression analysis of late embryogenesis abundant (LEA) genes under abiotic stress conditions in *Sorghum bicolor* L. *PLoS ONE*. 2019;14(1): e0209980. <https://doi.org/10.1371/journal.pone.0209980>.
54. Jin X, Cao D, Wang Z, Ma L, Tian K, Liu Y, et al. Genome-wide identification and expression analyses of the LEA protein gene family in tea plant reveal their involvement in seed development and abiotic stress responses. *Sci Rep*. 2019;9(1):14123. <https://doi.org/10.1038/s41598-019-50645-8>.
55. Zan T, Li L, Li J, Zhang L, Li X. Genome-wide identification and characterization of late embryogenesis abundant protein-encoding gene family in wheat: Evolution and expression profiles during development and stress. *Gene*. 2020;736: 144422. <https://doi.org/10.1016/j.gene.2020.144422>.
56. Artur MAS, Zhao T, Ligterink W, Schranz E, Hilhorst HWM. Dissecting the genomic diversification of late embryogenesis abundant (LEA) protein gene families in plants. *Genome Biol Evol*. 2019;11(2):459–71. <https://doi.org/10.1093/gbe/evy248>.
57. Liang Y, Xiong Z, Zheng J, Xu D, Zhu Z, Xiang J, et al. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*. *Sci Rep*. 2016;6:24265. <https://doi.org/10.1038/srep24265>.
58. Jeffares DC, Penkett CJ, Bähler J. Rapidly regulated genes are intron poor. *Trends Genet*. 2008;24(8):375–8. <https://doi.org/10.1016/j.tig.2008.05.006>.
59. Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. *New Phytol*. 2009;183(3):557–64. <https://doi.org/10.1111/j.1469-8137.2009.02923.x>.
60. Lane N, Martin W. The energetics of genome complexity. *Nature*. 2010;467(7318):929–34. <https://doi.org/10.1038/nature09486>.
61. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*. 2009;136(4):701–18. <https://doi.org/10.1016/j.cell.2009.02.009>.
62. Raskina O, Barber JC, Nevo E, Belyayev A. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenet Genome Res*. 2008;120(3–4):351–7. <https://doi.org/10.1159/000121084>.
63. Waminal NE, Pellerin RJ, Kang SH, Kim HH. Chromosomal mapping of tandem repeats revealed massive chromosomal rearrangements and insights into *Senna tora* dysploidy. *Front Plant Sci*. 2021;12: 629898. <https://doi.org/10.3389/fpls.2021.629898>.
64. Kong H, Landherr LL, Frohlich MW, Leebens-Mack J, Ma H, dePamphilis CW. Patterns of gene duplication in the plant *SKP1* gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *Plant J*. 2007;50(5):873–85. <https://doi.org/10.1111/j.1365-313X.2007.03097.x>.
65. Bies-Etheve N, Gaubier-Comella P, Debures A, Lasserre E, Jobet E, Raynal M, et al. Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Mol Biol*. 2008;67(1–2):107–24. <https://doi.org/10.1007/s11103-008-9304-x>.
66. Lan T, Gao J, Zeng QY. Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in *Populus trichocarpa*. *Tree Genet Genomes*. 2013;9(1):253–64. <https://doi.org/10.1007/s11295-012-0551-2>.
67. Yang Z, Nielsen R, Goldman N, Pedersen AMK. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*. 2000;155(1):431–49. <https://doi.org/10.1093/genetics/155.1.431>.
68. Hurst LD. The *Ka/Ks* ratio: diagnosing the form of sequence evolution. *Trends Genet*. 2002;18:486–7. [https://doi.org/10.1016/S0168-9525\(02\)02722-1](https://doi.org/10.1016/S0168-9525(02)02722-1).
69. Blanc G, Wolfe KH. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*. 2004;16(7):1667–78. <https://doi.org/10.1105/tpc.021345>.
70. Ku HM, Vision T, Liu J, Tanksley SD. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proc Natl Acad Sci USA*. 2000;97(16):9121–6. <https://doi.org/10.1073/pnas.160271297>.
71. Mehrotra R, Sethi S, Zutshi I, Bhalothia P, Mehrotra S. Patterns and evolution of ACGT repeat *cis*-element landscape across four plant genomes. *BMC Genomics*. 2013;14:203. <https://doi.org/10.1186/1471-2164-14-203>.
72. Du D, Zhang Q, Cheng T, Pan H, Yang W, Sun L. Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in

- Prunus mume*. Mol Biol Rep. 2013;40(2):1937–46. <https://doi.org/10.1007/s11033-012-2250-3>.
73. Jeon OS, Kim CS, Lee SP, Kang SK, Kim CM, Kang BG, et al. Fruit ripening-related expression of a gene encoding group 5 late embryogenesis-sabundant protein in *Citrus*. J Plant Biol. 2006;49(5):403–8. <https://doi.org/10.1007/BF03178819>.
 74. Charfeddine S, Saidi MN, Charfeddine M, Gargouri-Bouزيد R. Genome-wide identification and expression profiling of the late embryogenesis abundant genes in potato with emphasis on dehydrins. Mol Biol Rep. 2015;42(7):1163–74. <https://doi.org/10.1007/s11033-015-3853-2>.
 75. Celik Altunoglu Y, Baloglu MC, Baloglu P, Yer EN, Kara S. Genome-wide identification and comparative expression analysis of LEA genes in watermelon and melon genomes. Physiol Mol Biol Plants. 2017;23(1):5–21. <https://doi.org/10.1007/s12298-016-0405-8>.
 76. Galau GA, Hughes DW, Dure L III. Abscisic acid induction of cloned cotton late embryogenesis-abundant (*Lea*) mRNAs. Plant Mol Biol. 1986;7:155–70. <https://doi.org/10.1007/BF00021327>.
 77. Jiang J, Ma S, Ye N, Jiang M, Cao J, Zhang J. WRKY transcription factors in plant responses to stresses. J Integr Plant Biol. 2017;59(2):86–101. <https://doi.org/10.1111/jipb.12513>.
 78. Warmerdam S, Sterken MG, Sukarta OCA, van Schaik CC, Oortwijn MEP, Lozano-Torres JL, et al. The TIR-NB-LRR pair *DSC1* and *WRKY19* contributes to basal immunity of Arabidopsis to the root-knot nematode *Meloidogyne incognita*. BMC Plant Biol. 2020;20(1):73. <https://doi.org/10.1186/s12870-020-2285-x>.
 79. Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, Gomez-Gomez L, et al. MAP kinase signalling cascade in *Arabidopsis* innate immunity. Nature. 2002;28:977–83. <https://doi.org/10.1038/415977a>.
 80. Xing DH, Lai ZB, Zheng ZY, Vinod KM, Fan BF, Chen ZX. Stress- and pathogen-induced *Arabidopsis* WRKY48 is a transcriptional activator that represses plant basal defense. Mol Plant. 2008;1(3):459–70. <https://doi.org/10.1093/mp/ssn020>.
 81. Bhattarai KK, Atamian HS, Kaloshian I, Eulgem T. WRKY72-type transcription factors contribute to basal immunity in tomato and Arabidopsis as well as gene-for-gene resistance mediated by the tomato *R* gene *Mi-1*. Plant J. 2010;63(2):229–40. <https://doi.org/10.1111/j.1365-3113.2010.04232.x>.
 82. Li S, Fu Q, Chen L, Huang W, Yu D. *Arabidopsis thaliana* WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermo-tolerance. Planta. 2011;233(6):1237–52. <https://doi.org/10.1007/s00425-011-1375-2>.
 83. Li S, Zhou X, Chen L, Huang W, Yu D. Functional characterization of *Arabidopsis thaliana* WRKY39 in heat stress. Mol Cells. 2010;29(5):475–83. <https://doi.org/10.1007/s10059-010-0059-2>.
 84. Zou C, Jiang W, Yu D. Male gametophyte-specific WRKY34 transcription factor mediates cold sensitivity of mature pollen in *Arabidopsis*. J Exp Bot. 2010;61(14):3901–14. <https://doi.org/10.1093/jxb/erq204>.
 85. Ding ZJ, Yan JY, Xu XY, Yu DQ, Li GX, Zhang SQ, et al. Transcription factor WRKY46 regulates osmotic stress responses and stomatal movement independently in *Arabidopsis*. Plant J. 2014;79(1):13–27. <https://doi.org/10.1111/tpj.12538>.
 86. Davletova S, Rizhsky L, Liang H, Shengqiang Z, Oliver DJ, Coutu J, et al. Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of *Arabidopsis*. Plant Cell. 2005;17(1):268–81. <https://doi.org/10.1105/tpc.104.026971>.
 87. Besseau S, Li J, Palva ET. WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in *Arabidopsis thaliana*. J Exp Bot. 2012;63(7):2667–79. <https://doi.org/10.1093/jxb/err450>.
 88. Jiang Y, Liang G, Yang S, Yu D. *Arabidopsis* WRKY57 functions as a node of convergence for jasmonic acid- and auxin-mediated signaling in jasmonic acid-induced leaf senescence. Plant Cell. 2014;26(1):230–45. <https://doi.org/10.1105/tpc.113.117838>.
 89. Mukhtar MS, Liu X, Somssich IE. Elucidating the role of WRKY27 in male sterility in *Arabidopsis*. Plant Signal Behav. 2017;12(9):e1363945. <https://doi.org/10.1080/15592324.2017.1363945>.
 90. Moghaddam GA, Rezaatmand Z, Nasr Esfahani M, Khozaei M. Genetic defense analysis of tomatoes in response to early blight disease. *Alternaria alternata* Plant Physiol Biochem. 2019;142:500–9. <https://doi.org/10.1016/j.plaphy.2019.08.011>.
 91. Sun XC, Gao YF, Li HR, Yang SZ, Liu YS. Over-expression of *SlWRKY39* leads to enhanced resistance to multiple stress factors in tomato. J Plant Biol. 2015;58(1):52–60. <https://doi.org/10.1007/s12374-014-0407-4>.
 92. Atamian HS, Eulgem T, Kaloshian I. *SlWRKY70* is required for *Mi-1*-mediated resistance to aphids and nematodes in tomato. Planta. 2012;235(2):299–309. <https://doi.org/10.1007/s00425-011-1509-6>.
 93. Chen L, Yang Y, Liu C, Zheng Y, Xu M, Wu N, et al. Characterization of WRKY transcription factors in *Solanum lycopersicum* reveals collinearity and their expression patterns under cold treatment. Biochem Biophys Res Commun. 2015;464(3):962–8. <https://doi.org/10.1016/j.bbrc.2015.07.085>.
 94. Huang S, Gao Y, Liu J, Peng X, Niu X, Fei Z, et al. Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*. Mol Genet Genomics. 2012;287(6):495–513. <https://doi.org/10.1007/s00438-012-0696-6>.
 95. Ahammed GJ, Li X, Yang Y, Liu C, Zhou G, Wan H, et al. Tomato *WRKY81* acts as a negative regulator for drought tolerance by modulating guard cell H₂O₂-mediated stomatal closure. Environ Exp Bot. 2020;171:103960. <https://doi.org/10.1016/j.envexpbot.2019.103960>.
 96. Singh D, Debnath P, Roohi, Sane AP, Sane VA. Expression of the tomato WRKY gene, *SlWRKY23*, alters root sensitivity to ethylene, auxin and JA and affects aerial architecture in transgenic *Arabidopsis*. Physiol Mol Biol Plants. 2020;26(6):1187–99. <https://doi.org/10.1007/s12298-020-00820-3>.
 97. Wang L, Zhang XL, Wang L, Tian Y, Jia N, Chen S, et al. Regulation of ethylene-responsive *SlWRKYs* involved in color change during tomato fruit ripening. Sci Rep. 2017;7(1):16674. <https://doi.org/10.1038/s41598-017-16851-y>.
 98. Bai Y, Sunarti S, Kissoudis C, Visser RGF, van der Linden CG. The role of tomato WRKY genes in plant responses to combined abiotic and biotic stresses. Front Plant Sci. 2018;9:801. <https://doi.org/10.3389/fpls.2018.00801>.
 99. Tao Z, Liu H, Qiu D, Zhou Y, Li X, Xu C, et al. A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. Plant Physiol. 2009;151(2):936–48. <https://doi.org/10.1104/pp.109.145623>.
 100. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res. 2018;46(W1):W200–4. <https://doi.org/10.1093/nar/gky448>.
 101. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2021;49(D1):D458–60. <https://doi.org/10.1093/nar/gkaa937>.
 102. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412–9. <https://doi.org/10.1093/nar/gkaa913>.
 103. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202. <https://doi.org/10.1016/j.molp.2020.06.009>.
 104. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7. <https://doi.org/10.1093/molbev/msab120>.
 105. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49. <https://doi.org/10.1093/nar/gkr1293>.
 106. Zhang Z. KaKs_calculator 3.0: Calculating selective pressure on coding and non-coding sequences. Genom Proteom Bioinf. 2022. <https://doi.org/10.1016/j.gpb.2021.12.002>.
 107. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238. <https://doi.org/10.1186/s13059-019-1832-y>.
 108. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
 109. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91. <https://doi.org/10.1093/molbev/msm088>.
 110. Sun W, Xu X, Zhu H, Liu A, Liu L, Li J, et al. Comparative transcriptomic profiling of a salt-tolerant wild tomato species and a salt-sensitive tomato cultivar. Plant Cell Physiol. 2010;51(6):997–1006. <https://doi.org/10.1093/pcp/pcq056>.
 111. Mishra U, Rai A, Kumar R, Singh M, Pandey HP. Gene expression analysis of *Solanum lycopersicum* and *Solanum habrochaites* under drought conditions. Genom Data. 2016;9:40–1. <https://doi.org/10.1016/j.gdata.2016.04.001>.

112. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*. 2001;25(4):402–8. <https://doi.org/10.1006/meth.2001.1262>.
113. Nam H, Gupta A, Nam H, Lee S, Cho HS, Park C, et al. JULGI-mediated increment in phloem transport capacity relates to fruit yield in tomato. *Plant Biotechnol J*. 2022:1–13. <https://doi.org/10.1111/pbi.13831>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

