

# Genome-Wide Identification, Evolutionary And Expression Analyses of LEA Gene Family In Peanut (*Arachis Hypogaea* L.)

RuoLan Huang

Guangxi University

Dong Xiao (✉ [xiaodong@gxu.edu.cn](mailto:xiaodong@gxu.edu.cn))

Guangxi University

Xin Wang

Guangxi University

Yi Shen

Guangxi University

Jie Zhan

Guangxi University

AiQing Wang

Guangxi University

LongFei He

Guangxi University

---

## Research Article

**Keywords:** *Arachis hypogaea* L., Late embryogenesis abundant, Expression profiles, Abiotic stress

**Posted Date:** August 31st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-775523/v1>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Plant Biology on March 30th, 2022. See the published version at <https://doi.org/10.1186/s12870-022-03462-7>.

# Abstract

**Background:** Late embryogenesis abundant (LEA) proteins are a group of highly hydrophilic glycine-rich proteins, which accumulate in the late stage of seed maturation and are associated with many abiotic stresses. However, few peanut *LEA* genes had been reported, and the research on the number, location, structure, molecular phylogeny and expression of *AhLEAs* was very limited.

**Results:** In this study, 126 *LEA* genes were identified in the peanut genome through genome-wide analysis and were further divided into eight groups. Sequence analysis showed that most of the *AhLEAs* (85.7 %) had no or only one intron. *LEA* genes were randomly distributed on 20 chromosomes. Compared with tandem duplication, segmental duplication played a more critical role in *AhLEAs* amplification, and 93 segmental duplication *AhLEAs* and 5 pairs of tandem duplication genes were identified. Synteny analysis showed that some *AhLEAs* genes come from a common ancestor, and genome rearrangement and translocation occurred among these genomes. Almost all promoters of *LEAs* contain ABRE, MYB recognition sites, MYC recognition sites, and ERE cis-acting elements, suggesting that the *LEA* genes were involved in stress response. Gene expression analyses revealed that most of the *LEAs* were expressed in the late stages of peanut embryonic development. *LEA3* (AH16G06810.1, AH06G03960.1), and *Dehydrin* (AH07G18700.1, AH17G19710.1) were highly expressed in roots, stems, leaves and flowers. Moreover, 100 *AhLEAs* were involved in response to drought, low-temperature, or AI stresses. Some *LEAs* that were regulated by different abiotic stresses were also regulated by hormones including ABA, brassinolide, ethylene and salicylic acid. Interestingly, *AhLEAs* that were up-regulated by ethylene and salicylic acid showed obvious subfamily preferences.

**Conclusions:** *AhLEAs* are involved in abiotic stress response, and segmental duplication plays an important role in the evolution and amplification of *AhLEAs*. The genome-wide identification, classification, evolutionary and expression analyses of the *AhLEA* gene family provide a foundation for further exploring the *LEA* genes' function in response to abiotic stress in peanuts.

## Background

Plant in nature often encounters various abiotic stresses including drought, cold, high temperature, and salinity, which affect growth and development, reduce its yield and survival rate. Plants have evolved many mechanisms to cope with various environmental stresses. It is known that the late embryogenesis abundant (LEA) proteins play important roles in protecting cells under abiotic stresses, and many *LEAs* are induced by cold, drought, salinity, abscisic acid (ABA), and ethylene (Kentaro et al., 2014; Zegzouti et al., 1997; Park et al., 2003). Moreover, it has been confirmed that *AdDHN1*, a member of the Dehydrin family, can improve the drought resistance of transgenic *Arabidopsis*, but it is more sensitive to nematodes (Mota et al., 2018), which indicated that some of the *LEAs* may respond to abiotic stress as well as biotic stress.

LEA proteins are highly hydrophilic glycine-rich proteins, which accumulate largely in the later stage of seed maturation and fade away following germination (Battaglia and Covarrubias, 2013; Lan et al., 2013). As water-binding molecules, the role of LEA proteins is enhancing the stability of protein and membrane. Subcellular localization analysis has indicated that LEA proteins are mainly located in nuclear regions and the cytoplasm (Adrien et al., 2014). LEA proteins have been observed in the roots, leaves, buds, and seedlings, although they mainly appear in seeds of plants (Shao, 2005; Hundertmark et al., 2008).

LEA protein families were identified in many plant species by genome-wide identification and analysis, such as *Arabidopsis thaliana* (Bies-Ethève et al., 2008), *Populus trichocarpa* (Lan et al., 2013), *Camellia sinensis* (Jin et al., 2019), *Brassica napus* (Liang et al., 2016), and *Triticum aestivum* (Liu et al., 2019). During the growth and development of plants, LEA proteins are considered to play important roles. It was reported that *Medicago falcate* *LEA3* conferred multiple abiotic stress tolerance by involving the protection of catalase activity (Shi et al., 2020). A heterologous expression of a barley *LEA3* protein gene, *HVA1*, improved tolerance to water stress in rice and wheat (Sivamani et al., 2000; Xu et al., 1996). *AtLEA5* protects yeast cells against oxidative stress (Mowla et al., 2010). *Escherichia coli* can grow in high salt and extreme temperature conditions due to the over-expression of soybean PM2 protein (*LEA3*) (Yun and Zheng, 2005; Yun et al., 2010). ABA can regulate the expression of many LEA proteins, and it was proved that the expression of *LEA4* subfamily members was upregulated by exogenous ABA (Zamora-Briseño et al., 2016).

Peanut is one of the main oils and cash crops all over the world. Peanut is a rainfed crop, but it is sensitive to water deficit stress in the flowering and pegging stages, which would impact the yield of peanuts (Bhogireddy et al., 2020). Also, Al stress inhibition of growth reduces peanut yield in acid soil (Qiu et al., 2019). To date, the function of the *LEA* gene family in peanuts has little been reported. In this study, we identified the *LEAs* in peanuts and analyzed the structure, evolution, and chromosome location of peanut *LEAs*. Our findings provide a foundation for the evolutionary and functional characterization of *LEA* gene families in peanut and other plant species.

## Results

### 2.1. Identification and characteristics of *AhLEA* gene in peanut

By using the publicly available peanut genome sequence data, the genome-wide identification of *LEAs* in peanuts was based on sequence homology with 51 *Arabidopsis* *LEAs* (Hundertmark et al., 2008) (Table 1 and Additional file 1: Table S1). Proteins that contained a conserved LEA domain were screened by the NCBI-BLAST online tool. Eventually, 126 *AhLEAs* were identified. All of these genes were grouped with 51 *AtLEAs* by phylogenetic analyses. The *AhLEAs* were classified into eight subfamilies including *LEA1*, *LEA2*, *LEA3*, *LEA4*, *LEA5*, *PvLEA18*, *SMP*, and *Dehydrin* (Fig. 1). The *LEA2* family was the largest, with 78 members. The *LEA3s* and *LEA5s* had 14 and 10 members, respectively. The *LEA1s* had 8 members, *SMP*

had 6 members and *PvLEA18* had 4 members. The *LEA4* and *Dehydrin* families had 3 members. The species-specific group (*AtM*) of *Arabidopsis* was absent in the peanut.

Table 1 The classification of LEA proteins in *Arachis hypogaea* is based on *Arabidopsis*.

| In this study   | IPR ID              | Pfam ID | Hundertmark et al. (2008) | <i>Arabidopsis</i> | <i>A. hypogaea</i> |
|-----------------|---------------------|---------|---------------------------|--------------------|--------------------|
| <i>LEA1</i>     | IPR005513           | PF03760 | <i>LEA 1</i>              | 7                  | 8                  |
| <i>LEA2</i>     | IPR004864/IPR013990 | PF03168 | <i>LEA 2</i>              | 3                  | 78                 |
| <i>LEA3</i>     | IPR004926           | PF03242 | <i>LEA 3</i>              | 4                  | 14                 |
| <i>LEA4</i>     | IPR004238           | PF02987 | <i>LEA 4</i>              | 18                 | 3                  |
| <i>LEA5</i>     | IPR000389           | PF00477 | <i>LEA 5</i>              | 2                  | 10                 |
| <i>PvLEA18</i>  | IPR018930           | PF10714 | <i>PvLEA18</i>            | 3                  | 4                  |
| <i>SMP</i>      | IPR007011           | PF04927 | <i>SMP</i>                | 6                  | 6                  |
| <i>Dehydrin</i> | IPR000167           | PF00257 | <i>Dehydrin</i>           | 10                 | 3                  |
|                 |                     |         | <i>AtM</i>                | 2                  | 0                  |

## 2.2. Chromosomal locations, gene duplication and synteny analysis of the *AhLEAs*

The identified 126 *AhLEAs* were further located on the 20 chromosomes (Fig. 2). The largest number of *AhLEAs* was found on chromosome 12, fourteen genes, followed by chromosome 14 (eleven genes). The lowest loci density was observed on chromosome 8, with only two genes. Eight genes were located on chromosomes 13 and 15, seven genes were located on chromosome 6. Five chromosomes (chr1, chr2, chr3, chr4, and chr16) carried six *AhLEAs* and four chromosomes (chr5, chr7, chr19, and chr20) carried five *AhLEAs*. Chromosomes 9, 10, and 11 contained four *AhLEAs*, and chromosomes 17, 18 contained three *AhLEAs*. The *AhLEAs* were distributed unevenly among the 20 chromosomes in peanut.

Different *LEA* subfamilies were located on different chromosomes. All chromosomes contained the *LEA2s*, and all of the *LEAs* on chromosomes 3, 8, 10, 13, 18, and 20 belonged to the *LEA2s*. The *LEA1s* were distributed on chromosomes 1, 6, 9, 11, 16, and 19. The *LEA3s* were mainly distributed on chromosomes 1, 2, 4, 6, 11, 12, 14, and 16. The *LEA4s* were mainly found on chromosomes 6, 12, and 16. The *LEA5s* were distributed on chromosomes 2, 12, 14, and 16. The *PvLEA18s* were distributed on chromosomes 1, 5, 11, and 15. The *SMPs* were distributed on chromosomes 2, 6, 9, 12, and 19. The *Dehydrins* were distributed on chromosomes 7, 15, and 17. Chromosomal location analysis of *AhLEAs* indicated that eight subfamilies were distributed unevenly in the genome.

The generation and evolution of gene families may be caused by tandem duplication and segmental duplication (Mehan et al., 2004; Kent et al., 2003). To investigate the evolutionary relationships of the

*AhLEA* gene family, we analyzed the duplication events of *AhLEAs* (Fig. 3). In this study, five pairs of tandem duplication and 93 pairs of segmental duplication were identified (Fig. 3, Additional file 1: Table S2). Five tandem duplication pairs (AH12G24920.1: AH12G24910.1, AH12G24930.1: AH12G24910.1, AH15G25110.1: AH15G24880.1, AH15G25120.1: AH15G24880.1, AH15G25120.1: AH15G25110.1) belong to the *LEA5s* and *LEA2s*, and located on chromosomes 12 and 15. The segmental duplication genes were mainly distributed on chromosome 12. All members of the *LEA1s*, *SMPs*, and *PvLEA18s* were segmental duplication genes, followed by *LEA2s* (79.5 %) and *LEA3s* (71.4 %). The Ka/Ks values of all the tandem duplication gene pairs were less than 1. Except for four segmental duplication gene pairs whose Ka/Ks values could not be calculated, the Ka/Ks values of the most segmental duplication gene pairs were less than 1, and only two pairs (2.2 %) were more than 1. (Fig. 4, Additional file 1: Table S2). The divergence time of tandem duplication events was mainly 0–10 million years ago (MYA), and 49.5 % (46/93) of segmental duplication events occurred between 0–5 MYA (Fig. 5, Additional file 1: Table S2).

To explore the evolutionary process of the peanut *LEA* genes, we performed synteny analysis among peanut, Arabidopsis, and soybean. *AhLEAs* showed a more syntenic to soybean than Arabidopsis (Fig. 6, Additional file 1: Table S3). Thirteen orthologous pairs exhibited single gene correspondences between peanut and Arabidopsis, and five orthologous pairs exhibited single gene correspondences peanut and soybean. Five *AhLEAs* were associated with multiple *AtLEAs*, and fourteen *AhLEAs* were associated with *GmLEAs*. Additionally, there were nine cases that peanut segmental duplications that corresponded to a single Arabidopsis gene, and eleven cases that *AhLEAs* corresponded to a single soybean gene. Finally, some genes showed more-to-more correspondence, for example, AH02G02040.1/ AH12G02210.1/ AH04G26920.1/ AH14G31640.1-AT1G01470.1/ AT2G46140.1 and AH01G11560.1/ AH05G04840.1/ AH11G11350.1/ AH15G00880.1-Glyma05g22030.1/ Glyma17g17860.1.

## 2.3. Analysis of gene structure and protein motifs of *LEAs* in peanut

To examine the structural characteristics of *AhLEAs*, an unrooted phylogenetic tree that combines the UTR-CDS structures and motifs were constructed based on the full lengths of the 126 peanut *LEA* gene sequences by using the Maximum-Likelihood method (Fig. 7). The majority of the *AhLEAs* contained zero or one intron, with 55 and 53, respectively. Sixteen genes had two introns. One gene, AH19G03360.1, contained three introns, and one gene, AH12G35940.1, contained seven introns. All the *LEA1s* and *Dehydrins* contained only one intron, and the main members of the *LEA3* and *LEA5* subfamilies had one intron. The majority of the *LEA2s* had no intron. To identify the conserved protein motifs, the MEME (<http://meme-suite.org/tools/meme>) online software was used to predict putative motifs of these proteins, with a maximum number of the different motifs at 20. Motif analysis indicated that members of each subfamily had the group-specific conserved domain, and *AhLEAs* with closer evolutionary relationships had more similar motif numbers. MEME analysis revealed that most *AhLEAs* contained motif 3 and all the *LEA4s* and *LEA1s* had motif 13. The *LEA2s* had the greatest number of motifs, which were approximately 7, while other subfamily members had 1 to 4 motifs.

## 2.4. Analysis of cis-acting elements in promoters of *AhLEAs*

To investigate the cis-acting elements of *AhLEAs*, 2 kb upstream of the translation initiation sites of all the *LEA* genes were obtained from the peanut genome database. Many cis-acting regulatory elements that may be involved in the plant's response to environmental stresses, including ABRE, WRE3, ERE, MYB recognition sites, MYC recognition sites, TC-rich repeats, STRE, and MRE, were detected (Fig. 8). The promoter of subfamily *LEA2* contained the most cis-acting elements, followed by subfamily *LEA3*, *LEA5*, and *LEA1*. The promoter of subfamily *LEA4*, *SMP*, *PvLEA18*, and *Dehydrin* contained the least elements. Among the identified cis-acting elements, ABRE (22.2 %), ERE (55.6 %), MYB recognition sites (65.9 %), and MYC recognition sites (70.6 %) cis-acting elements were over-represented.

## 2.5. Expression profiles of *AhLEAs* at different stages of embryo development

To investigate the expression profiles of *AhLEAs* across different stages of embryo development and different tissues, the transcriptomic data of a cultivated variety (*A. hypogaea* L.) in gene bank were further scrutinized (<http://peanutgr.fafu.edu.cn/Transcriptome.php>) (Fig. 9, Additional file 1: Table S4). Not all *AhLEAs* were expressed at the four embryo development stages. Meanwhile, twenty-seven genes were not detected at any tested stages. Sixty-eight *LEAs* had different expression levels among the four stages. In the early embryo development stages, most *LEA3s* were up-regulated. Among them, three *LEA3s* (AH01G27080.1, AH01G27080.2, and AH11G30560.1) exhibited very high expression levels in the early stages, which showed up to 10-fold higher than those in the late stages. Nevertheless, *AhLEA1s*, *AhLEA4s*, and *AhLEA5s* were up-regulated mainly in the late stages. Four genes including two *LEA5s* (AH12G24910.1 and AH12G24920.1) and two *LEA1s* (AH06G01030.1 and AH16G03650.1) exhibited very high expression levels in the late stages. Two genes of the *Dehydrins* expressed at a high level in stages I, and II, while another *Dehydrin* (AH17G19580.1) expressed at a high level in stages III, and IV. The expression of most *AhLEA2s* was not changed as embryo development, while the expression level one *LEA2* (AH12G34850.1) in the early stages showed up to 26-fold higher than those in the later stages.

As shown in Fig. 10, the expression profiles of eight subfamilies, including *LEA1s*, *LEA2s*, *LEA3s*, *LEA4s*, *LEA5s*, *SMPs*, *PvLEA18s*, and *Dehydrins*, were similar in roots, stems, leaves, and flowers. Among them, the members of *LEA2s*, *LEA3s*, and *Dehydrins* were expressed at a high level in all four tissues. Twenty-four *LEAs* were highly expressed in roots, 21 in stems, 15 in leaves, and 20 in flowers. Two *Dehydrins* (AH07G18700.1 and AH17G19710.1) and two *LEA3s* (AH16G06810.1 and AH06G03960.1) had the highest expression levels in the stem (Additional file 1: Table S5).

## 2.6. Expression profiles of *AhLEAs* in response to drought and low-temperature stresses.

To investigate the transcriptional changes of the *AhLEAs* under cold and drought stresses, the expression profiles of these genes were examined by using transcriptomic data (Fig. 11). Under drought treatment, 28.6 % (36 out of 126) of the *AhLEAs* were more than 2-fold up-regulated compared with the control, while the expression levels of 21.4 % (27 out of 126) genes were down-regulated more than 2-fold. Among the 27 genes that down-regulated more than 2-fold, 24 genes belonged to the *LEA2* subfamily.

Two *LEA3s* (AH01G27080.1, and AH01G27080.2) showed the highest expression levels under drought stress (Additional file 1: Table S6).

Under low-temperature treatment, 28.6 % (36 out of 126) of the *AhLEAs* were more than 2-fold up-regulated compared with the control, while the expression levels of 14.3 % (18 out of 126) genes were down-regulated more than 2-fold. It was found that 21 genes of *LEA2s* were up-regulated and 11 genes were down-regulated. It is noteworthy that all *Dehydrins* were up-regulated under drought and low-temperature stresses. Interestingly, the genes expressed the highest under low-temperature stress were also two *LEA3* subfamily genes (AH16G06810.1, AH06G03960.1). Besides, 60 % of *LEA5s* genes were not detected under drought and low-temperature (Fig. 11: Additional file 1: Table S6).

## 2.7. Expression profiles of *AhLEA* genes in response to hormone

To understand the expression changes of the *AhLEAs* under drought and low-temperature stresses, the responses of 126 *AhLEAs* to four stress-related hormones (abscisic acid, brassinolide, ethylene, and salicylic acid) were investigated (Fig. 12). The expression profiles of these genes were examined by using transcriptomic data. After ABA treatment, 8 *LEAs* were induced more than 2-fold, while 19 *LEAs* were down-regulated more than 2-fold. After brassinolide treatment, 5 genes were up-regulated more than 2-fold, while and 31 genes were down-regulated more than 2-fold. The expression of 13 *AhLEAs* was up-regulated more than 2-fold after ethylene treatment, while 28 genes were down-regulated more than 2-fold. The expression of 10 *AhLEAs* was up-regulated more than 2-fold after salicylic acid treatment, while 16 genes were down-regulated more than 2-fold. Although the main *AhLEAs* were down-regulated by these four hormones, half of the *LEA3s* (7 out of 14) were up-regulated more than 2-fold after ethylene treatment, all members of *LEA4s* were induced by salicylic acid. Moreover, the expression of five *AhLEAs* was up-/down-regulated more than 2-fold by all four tested hormones. These genes included four *LEA2s* (AH06G19190.1, AH16G23780.1, AH20G34490.1, and AH16G06810.1) which were down-regulated after hormone treatment and a *PvLEA18* (AH11G11350.1) that was up-regulated (Additional file 1: Table S7).

## 2.8. Expression pattern of *AhLEAs* under Al stress

To gain a broader understanding of the putative functions of peanut *LEAs* in response to Al stress, the expression profiles of these genes were examined by using the RNA-Seq data which was generated from the root tips of two peanut cultivars that exhibited different Al sensitivity and had already been deposited in NCBI (Xiao et al., 2021). ZH2 is known as an Al sensitive peanut cultivar and 99-1507 is proved as an Al tolerant peanut cultivar (Zhan et al., 2008). Here, a total of 50 *AhLEAs* were found to be aluminum stress-responsive genes (Fig. 13, Additional file 1: Table S8). *LEA2s* which included twenty-three DEGs had the most aluminum stress-responsive genes. All of the members in *LEA4s* and *Dehydrins* were aluminum stress-responsive genes, and both of these two subfamilies were composed of three genes. The aluminum stress-responsive genes accounted for 75 % (3 out of 4), 60 % (6 out of 10), 50 % (3 out of 6, and 4 out of 8), and 35.7 % (5 out of 14) of the members in *PvLEA18s*, *LEA5s*, *SMPs* (and *LEA1s*), and *LEA3s*, respectively.

Five *AhLEAs* were significantly up-regulated after 8h of Al treatment in ZH2. Seventeen *AhLEAs* were significantly up-regulated after 8h of Al treatment in 99-1507, including 66 % of *LEA4s* and *Dehydrins* (2 out of 3) and 50 % of *SMPs* (3 out of 6). Nineteen *AhLEAs* were significantly up-regulated after 24 h of Al treatment in ZH2, including all members of *LEA5s* and *PvLEA18s* and half of *LEA1s* (4 out of 8) and *SMPs* (3 out of 6). Nine *AhLEAs* were significantly up-regulated after 24 h of Al treatment in 99-1507. Twelve *AhLEAs* were down-regulated after 8h of Al treatment in ZH2, and all *PvLEA18s* were down-regulated. Two *AhLEAs* were down-regulated after 8h of Al treatment in 99-1507. Seven *AhLEAs* were down-regulated after 24h of Al treatment in ZH2. Eight *AhLEAs* were down-regulated after 24h of Al treatment in 99-1507.

Furthermore, we compared the DEGs in Al stress and the genes that were up-/down-regulated more than 2-fold under drought and low-temperature stresses. As shown in Fig S1, a total of 100 *AhLEAs* were regulated under drought, low temperature, and Al stress. Among these genes, 35 common *AhLEAs* were involved in the responses to low-temperature and drought stresses, 29 common *AhLEAs* that were involved in the responses to drought and Al stresses, and 22 common *AhLEAs* that were involved in the responses to low-temperature and Al stresses. Sixteen *AhLEAs* were overlaps among the three abiotic stresses (Additional file 2: Fig S1).

## Discussion

### 3.1. Identification of the *LEA* gene family

In this study, 126 *LEA* genes were identified from whole peanut genome sequences. Based on the phylogenetic relationship with *Arabidopsis*, these 126 *AhLEAs* were distributed to eight groups. The number of peanut *LEAs* was twice that in *Arabidopsis*. According to the studies of the *LEA* family in other species, the number of *LEAs* may be related to the polyploidy of plants (Ibrahime et al., 2019). For example, many *LEA* genes were found in polyploids of *upland cotton* (Fang and Magwanga, 2018), *Triticum aestivum* (Shi et al., 2020), and *Brassica napus* (Liu et al., 2019). The *LEA2s* had more members than other subfamilies in the majority of species so far studied (Ibrahime et al., 2019). The *LEA2s* were the largest *LEA* subfamily in *Citrus sinensis*, *Oryza sativa*, *Populus trichocarpa* (Pedrosa et al., 2015), and *upland cotton* (Fang and Magwanga, 2018). Similarly, the *AhLEAs* mainly belong to the *LEA2s*, which accounted for 61.9 % of the *LEA* genes. However, the *LEA2s* were not found as a large subfamily in the previous works in *Arabidopsis* (Hundertmark and Hinch, 2008), *Oryza sativa* (Wang et al., 2007), and *Populus trichocarpa* (Lan et al., 2013). This result can be partly explained by the fact that improved annotation of higher plant genomes can be found on phytochrome (v10.2), and *LEA2* is an unusual component of "a typical" *LEA* proteins because they are more hydrophobic. In addition, there were three pairs of tandem duplication in *AhLEAs*, which belong to the *LEA2s*. This result supported the view that tandem duplications have contributed significantly to the expansion and diversity of the *LEA2s* in most species (Artur et al., 2019).



## 3.2. Analysis of conserved domains and introns revealed that *LEAs* might be stress-response genes

Motif analysis of the *AhLEAs* showed that members of each *LEA* group contained specific conserved motifs. Most members of the same subfamily have similar motifs, indicating an important role of these conserved motifs in the evolution of the *LEA* gene family. Big differences were found in the structure of different clades. For example, *LEA1s* contained motifs 13 and 17, whereas *LEA5s* contained motif 2 and 9, which indicated the complexity and group-specific of *LEA* protein function. The conserved motifs observed in each *LEA* group suggested that genes with the same motif might be amplified from genes within the same evolutionary clade or group. It has been reported that stress-responding genes usually contain fewer introns (Fang and Magwanga, 2018; Xie et al., 2015). Introns have harmful effects on gene expression by delaying transcription products (Lane et al., 2010). Moreover, introns can extend the length of new transcripts, resulting in additional energy consumption for transcription (Jeffares et al., 2008). Up to 85.7 % of the *AhLEAs* had zero or only one intron, which further suggested that *AhLEAs* were stress-related genes.

## 3.3. Segmental duplication plays an important role in the evolution and amplification of *AhLEAs*

Gene duplication plays an important role in the evolution and amplification of gene families (Flagel and Wendel, 2009). In this study, 93 pairs of segmental duplication and 5 pairs of tandem duplication were identified, it could be inferred that segmental duplication and tandem duplication contribute to the common expansion of the *AhLEAs* family, but the former played a predominant role. This finding was similar to previous studies on *Brassica napus* and upland cotton (*Gossypium hirsutum*) (Liu et al., 2019; Fang and Magwanga, 2018) and consistent with our previous study on receptor-like protein kinase (RLK) in peanut (Wang et al., 2021). According to Ka/Ks estimation, 94.9 % of the duplication gene pairs of *AhLEAs* were less than 1, indicating the results of the purification selection. The Ka/Ks ratios of two gene pairs (AH01G27080.1 and AH11G30560.1, AH05G16640.1 and AH15G06250.1) were more than 1, which indicated that these genes were in a state of positive selection in peanuts. We calculated the divergence time, and the results showed that many duplication events appeared to have occurred during relatively recent key periods. For example, all tandem duplication events occurred at 0–10 MYA, and 49.5 % of segmental duplication occurred at 0–5 MYA. These results indicated that many *AhLEAs* were produced by the recent gene duplication events in peanuts. This may be related to the origin of cultivated peanut, through a single and recent polyploidization event, and then continuous selection in breeding work, forming a highly conserved genome (Cunha et al., 2008).

The closer the species are, the greater the genome coverage of synteny fragments and the more genes they contain (Ye et al., 2020). Syntenic analysis showed that more homologous gene pairs were found between peanut and soybean. There were 13 single peanut-to-Arabidopsis *LEA* gene correspondences and 5 single peanut-to-soybean *LEA* gene correspondences. These results indicated that these genes come from a common ancestor. Among these genes, two soybean genes (Glyma11g02290.1 and

Glyma09g30400.1) did not belong to the *LEA* family. The rest genes showed one-to-more, more-to-one and more-to-more correspondence, and most of the genes included in these cases appeared more than once. However, 15 of the 23 orthologs of *AhLEAs* in soybean (Glyma09g38980.1, Glyma19g37350.1, Glyma08g22050.1, Glyma12g09590.1, Glyma12g32090.1, Glyma13g38380.1, Glyma03g34670.1, Glyma10g07360.1, Glyma13g21240.1, Glyma19g37340.1, Glyma06g01170.1, Glyma07g06960.1, Glyma13g43610.1, Glyma09g30400.1, and Glyma20g35880.1) were not *LEA* genes, which implied that their genomes underwent multiple rounds of chromosomal rearrangement and fusions. Allotetraploid-cultivated peanut composed of A and B genomes and was generated from diploid *A. duranesis* (AA) and *A. ipaensis* (BB) (Zhuang et al., 2019). Taking into account the divergence time of the duplications, we inferred that the divergence of many *AhLEAs* duplications occurred after the divergence of peanut and Arabidopsis/soybean from their last common ancestor. Combined with the results of phylogenetic tree analysis, there were nine orthologs including nineteen peanut *LEA* genes (AH12G35940.1-AT2G36640.1, AH17G19580.1-AT2G21490.1, AH12G35940.1-AT3G22500.1, AH02G22690.1/ AH12G24910.1-AT3G51810.1, AH05G04840.1/ AH15G00880.1-AT2G23110.1, AH04G10170.1/ AH14G12410.1-AT4G15910.1, AH06G03960.1/ AH12G32330.1/ AH12G37270.1/ AH16G06810.1-AT1G02820.1/ AT4G02380.3, AH02G06810.1/ AH12G08270.2-AT1G03120.1/ AT3G22490.1, AH02G02040.1/ AH04G26920.1/ AH12G02210.1/ AH14G31640.1-AT2G46140.1/ AT1G01470.1) that could be clustered together in the phylogenetic tree and were also contained in the syntenic map. We speculate that the functions of these *AhLEAs* are more similar to their Arabidopsis homologs than the other *AhLEAs* in the phylogenetic tree and syntenic map.

### 3.4. MYB and MYC recognition sites may be involved in the response of *AhLEAs* to abiotic stress

Many studies have shown that *LEAs* play an important role in abiotic stress. In this study, many cis-acting elements related to abiotic stress and plant hormones were identified, such as ABRE, ERE, MYB recognition sites, MYC recognition sites, and STRE. We found that the MYB and MYC recognition sites were presented in the most promoters of the *AhLEAs*. It is reported that MYBs and MYCs are transcription factors that participate in ABA-dependent signaling pathways to cope with abiotic stresses such as drought, salt, and low-temperature (Li et al., 2015; Boter, 2014). Consistently, the *LEAs* that contain MYB recognition sites and MYC recognition sites, including AH16G06810.1, and AH06G03960.1, were induced under ABA, salicylic acid, drought, and low-temperature stress. Besides, most of the *LEA1s*, *LEA5s*, *SMPs*, and *Dehydrins* were highly expressed under aluminum stress, and these genes contained a large number of MYB and MYC recognition sites. Therefore, we speculated that the up-regulation of *LEAs* expression under aluminum stress might be regulated by MYB and MYC transcription factors. This provides a theoretical basis for further exploring the response regulation mechanism of *LEAs* containing cis-acting elements of MYB and MYC recognition sites under stress. In this study, many *LEAs* responses to abiotic stresses were found to be regulated by ABA. However, only five *AhLEAs* (AH03G11350.1, AH03G03840.1, AH02G02050.1, AH01G27080.1, AH01G27080.2) contained ABRE cis-acting elements. The promoter

region of many *AhLEAs* had recognition sites for MYB and MYC. These results suggested that *AhLEAs* responded to abiotic stresses such as drought, low-temperature, and AI stress might be direct or indirect activated ABA-dependent signaling pathways.

### 3.5. Expression analysis revealed *AhLEAs* respond to different abiotic stresses

It can obtain clues from gene expression patterns to explore the function of genes (Chen et al., 2019). We investigated the expression level of *AhLEAs* in different tissues, at different embryo development stages, under different abiotic stresses (drought, low-temperature, and AI treatment), and after different hormone treatments. In four different embryo development stages, there were sixty-eight differentially expressed genes. Consistent with previous studies (Liang et al., 2016) that *LEAs* were up-regulated as the embryo developed, most of the *AhLEAs* were expressed at a high level at stages Ⅱ, and Ⅲ. However, the majority of *LEA3s* were highly expressed at an early stage, suggesting the potential roles of *LEA3s* in the early embryo development stage. As shown in Fig. 9, subfamily *LEA2* was the biggest subfamily, but the expression levels of most *LEA2s* at four embryo development stages were stable, suggesting that *LEA2s* might not play important roles during embryo development.

The expression level of most *AhLEAs* in the root, stem, leaf, and flower tissues was similar. The expression level of many *AhLEAs* was low, while there were still several genes of subfamily *LEA2*, *LEA3*, and *Dehydrin* that exhibited a high expression level in the four tissues. Two *LEA3s* (AH16G06810.1, AH06G03960.1) were very highly expressed in different peanut tissues (Fig. 10). It was reported that the *LEA3s* play an important role in plant growth, development, and response to abiotic stresses (Yu et al., 2019; Koubaa and Brini 2020), and these two genes might be suitable candidates to understand the role of *LEA3s* in peanut.

Under drought stress, 50 % of the *AhLEAs* were up-/down-regulated for more than 2-fold compared with control. Among them, *LEA2s* contributed most genes, containing 10 up-regulated genes and 24 down-regulated genes. This is consistent with the fact that *LEA2s* were the largest subfamily in peanuts. Among the genes that were down-regulated for more than 2-fold, most of them were *LEA2s*. Additionally, four *AhLEA1s* and three *AhLEA3s* were induced more than 60-fold by drought stress, implying their potential roles in enhancing drought stress tolerance in peanuts.

Under low-temperature stress, 36 *AhLEAs* were up-regulated more than 2-fold, while 18 genes were down-regulated more than 2-fold. *LEA2s* also contributed to most genes. Twenty-one *AhLEA2s* were up-regulated and eleven genes were down-regulated. Interestingly, the *LEA2s* that down-regulated under drought stress was also down-regulated under low-temperature stress, which suggested that there was a common mechanism to regulate *LEA2s* expression.

Seventeen genes were up-regulated after 8h of AI treatment in 99-1507, and two of their (AH16G20700.1 and AH06G16990.1) were also up-regulated after 24h of AI treatment. In ZH2, only five *AhLEAs* were up-regulated after 8h of AI treatment, while sixteen *AhLEAs* were up-regulated after 24h of AI treatment.

Interestingly, three *SMPs* (AH12G08270.1, AH12G08270.2, and AH02G06810.1) were up-regulated after 8h of Al treatment in both cultivars, suggesting that these genes might play important roles in Al tolerance in peanuts. Together, the Al-tolerant cultivar 99-1507 exhibited a rapid response to Al treatment, and the *LEAs* that induced rapidly should be studied in future work.

As shown in Fig. S1, the majority of the 126 *LEAs* were induced under at least one stress condition. Sixty-three (50 %), fifty-four (42.9 %), and fifty (39.7 %) the *AhLEAs* were induced by drought, low-temperature, and Al stresses, respectively. Among these genes, sixteen were induced only under drought stress, fourteen were induced only under low-temperature, and sixteen were induced only under Al stress (Fig S1). These results implied that these genes play distinct roles in response to different abiotic stresses in peanuts.

Some *AhLEAs* were regulated by different stress conditions. Three genes including two *LEA5s* (AH12G24910.1 and AH12G24920.1) and one *LEA1* (AH19G11740.1) were up-regulated greatly under both drought and Al stresses (Additional file 1: Table S6, Table S 8). The expression of *LEA1* (AH19G11740.1) was induced more than 2-fold by ABA treatment. Two *LEA3s* (AH01G27080.1 and AH11G30560.1) and one *LEA4* (AH12G35940.1) were down-regulated under drought, low-temperature, and Al stresses. The expression of that two *LEA3s* was significantly induced by ethylene, while *LEA4* (AH12G35940.1) was down-regulated by ABA treatment. Two *LEA2s* (AH02G02040.1 and AH12G02210.1) were up-regulated under drought, low-temperature, and Al stresses, and they were also up-regulated by ABA.

Many *AhLEAs* that were regulated more than 2-fold by hormones such as abscisic acid, brassinolide, ethylene, and salicylic acid were found to be down-regulated. As revealed by table S7, these down-regulated genes showed no obvious subfamily preference. However, *AhLEAs* that were up-regulated more than 2-fold by ethylene and salicylic acid showed obvious subfamily preference. Seven *AhLEA3s* were induced by ethylene. Five *AhLEA3s* induced by ethylene were also involved in response to drought and low-temperature stresses. The expression level of AH12G37280.1 was increased up to 8.45-fold under low-temperature stress. AH12G32330.1 was up-regulated 3.5-fold under drought stress. Moreover, three *AhLEA3s* (AH01G27080.1, AH01G27080.2, AH11G30560.1) were up-regulated greatly under both drought and low-temperature stresses. These results revealed the important roles of the *AhLEA3* subfamily in the ethylene-mediated response under drought and low-temperature stresses. Additionally, all *AhLEA4s* were induced by salicylic acid, and all *AhLEA4s* were also regulated greatly under drought and low-temperature stresses. Among them, two genes (AH06G16990.1 and AH12G35940.1) were induced more than 6-fold under drought and low-temperature stresses, and one gene (AH16G20700.1) was down-regulated 3.5-fold under low-temperature stress, which implied that subfamily *AhLEA4* played important roles in SA-mediated response under drought and low-temperature stresses in peanut.

Taken together, these results suggested that common mechanisms might be initiated in peanuts to cope with different abiotic stresses. Hormones were involved in regulating *LEA's* expression under abiotic

stresses. The role of hormones in regulating gene expression had a preference among *AhLEA* gene families.

## Conclusions

In this study, 126 *LEA* genes in *Arachis hypogaea* were identified. They were divided into eight groups according to homologous in *Arabidopsis thaliana*. *AhLEAs* are randomly distributed on the chromosome, and most of them may be segmental duplication. The exon-intron and motif structures indicated that the *LEAs*' family functions are highly conserved. Some cis-elements of abiotic stress response were also found in the upstream sequences of most *AhLEAs*. The comprehensive analysis of *AhLEAs* gene expression profiles showed that the *LEA3s*, *LEA4s*, and *SMPs* played an important role in abiotic stress response, and also showed the functional differences among other subfamilies. This study provided a reference for further exploring the mechanism of *LEAs* in response to abiotic stress in peanuts.

## Materials And Methods

### 5.1. Identification of *LEAs* in peanut

To identify the *AhLEAs*, we used 51 *LEA* genes (Hundertmark et al., 2008) in *Arabidopsis thaliana* acquire Pfam ID (PF03760, PF03168, PF03242, PF02987, PF00477, PF10714, PF04927, PF00257) and InterPro ID (IPR005513, IPR004864, /IPR013990, IPR004926, IPR004238, IPR000389, IPR018930, IPR007011, IPR000167) from Peanut Base (<https://www.peanutbase.org/>). By acquiring *LEA* peanut protein sequences based on InterPro ID search of Peanut Genome Resource (PGR) (<http://peanutgr.fafu.edu.cn/>). NCBI's Conserved Domains Database (<https://www.ncbi.nlm.nih.gov/cdd>) and PFAM (<http://pfam.xfam.org/>) database were used to verify the presence of the *LEA* domains and finally obtained 126 *AhLEAs*.

### 5.2. Phylogenetic relationships, gene structures, conserved motifs, and chromosomal locations of the *AhLEAs*

The phylogenetic tree was constructed by the maximum-likelihood method with 1000 bootstrap replicates in MEGA 7.0 software (Sudhir et al., 2016). Multiple Expectation Maximization for Motif Elicitation (MEME) (<http://meme-suite.org/tools/meme>) (Bailey et al., 2009) was used to identify the conserved protein motifs, with a maximum number of the different motif at 20. The exon-intron structures were identified using the TBtools software (Chen et al., 2020). The physical location of each *AhLEA* is determined by identifying the starting position of all genes on each chromosome, searching the local database of Peanut Genome Resources by BLAST. Using TBtools of Gene location visualize from GFF/GFF3 to draw chromosome mapping and tandem duplication pairs.

### 5.3. Promoter cis-element analysis

Genomic data were obtained from Peanut Genome Resource (PGR) (<http://peanutgr.fafu.edu.cn/>), and TBtools software was used to extract all *LEA* upstream 2kd promoter sequences. Transcriptional response elements of *LEA* gene promoters were predicted using the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Higo et al., 1999).

## 5.4. Gene duplication and evolutionary analysis

We used Virtual Machine to construct the tandem and segmental of the putative duplication of the *AhLEAs* and calculate the ratio of the nonsynonymous substitution rate ( $K_a$ ) to the synonymous substitution rate ( $K_s$ ) by the Simple  $K_a/K_s$  calculator (NG) of TBtools (Chen et al., 2020). *LEAs* clustered together within 100 kb, length of the alignable sequence covers > 75 % of longer gene and similarity of aligned regions > 75 % were regarded as tandem duplicated genes. The relationship between  $K_a/K_s$  ratio and value 1,  $K_a$  larger than  $K_s$  (or  $K_a/K_s > 1$ ),  $K_a$  equals  $K_s$  ( $K_a/K_s = 1$ ), and  $K_a$  less than  $K_s$  (or  $K_a/K_s < 1$ ), which represent positive (or diversifying) selection, neutral evolution and negative (or purifying) selection, respectively. Divergence time was calculated with the formula  $T = K_s/2r$ , where  $r$  is  $1.5 \times 10^{-8}$  synonymous substitutions per site per year and it is the rate of divergence for nuclear genes from plants (Koch et al., 2000). We used Multiple Synteny Plot software (Chen et al., 2020) to explore the collinear relationship between the *AhLEA* and *LEA* genes from *Arabidopsis thaliana* and *Glycine max*. All the soybean *LEA* domain-containing protein sequences were downloaded from the Soybase Glyma.Wm82.a2.v1 (<http://www.soybase.org/>). The NCBI's Conserved Domains Database (<https://www.ncbi.nlm.nih.gov/cdd>) and PFAM (<http://pfam.xfam.org/>) database were used to verify the presence of the *LEA* domains. The *GmLEAs* that were identified in the previous study were also screened (Li et al., 2011). After eliminating the invalid sequence, a total of 132 *GmLEAs* were identified.

## 5.5. Expression analysis of *AhLEAs*

The blast was performed in the transcriptome of the PGR database using the protein sequences of 126 *AhLEAs*. RNA-Seq data were downloaded from PGR and used to generate the expression patterns of *AhLEAs* in different tissues (root, stem, leaf, and flower), different embryo development stages, and various abiotic stresses (cold, and drought), and different hormones treatment on leaves. Transcriptome data that were generated from peanut root tips under Al stress were used to generate the expression patterns of *AhLEAs* under Al stress. The data had been deposited in the database of the National Center for Biotechnology Information (NCBI) under accession number PRJNA525247 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA525247>). TBtools were used to generate heat maps and combine phylogenetic tree, gene, and protein structure (Chen et al., 2020).

## Abbreviations

Al: Aluminum

*At*: *Arabidopsis thaliana*

*Ah*: *Arachis hypogaea*. L

ABRE: ABA-responsive element

ERE: Ethylene response element

WRE3: Water response element

MYB: Transcription factor

MYC: Transcription factor

TC-rich repeats: Cis-acting element involved in defense and stress responsiveness

MRE: Metal responsive element

STRE: Stress response element

DEGs: Differentially expressed genes

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

**Competing interests:** All authors declare no conflicting interest.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant No. 31701356, 32060419) and College Students' Innovative Entrepreneurial Training Plan Program (201910593083). Apart from providing financial support, funding bodies were not involved in the study design, data analyses, and interpretation of results or manuscript preparation. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Authors' contribution:** RLH: Writing Original Draft, Data analysis, Conceptualization; XW and YS: Data analysis; LFH, AQW, and JZ: manuscript review; DX: Conceptualization, supervision, Writing—Review & Editing. All authors read and approved the final manuscript.

**Acknowledgments:** We thank Mr Yun Xiong Zhao and Miss Xu Fang for providing the software needed for drawing and data collection.

**Availability of data and materials:** The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. The following are all databases in this study and are open.

Peanut Base (<https://www.peanutbase.org/>)

Peanut Genome Resource (PGR) (<http://peanutgr.fafu.edu.cn/>)

Transcriptome of Peanut Genome Resource (PGR) (<http://peanutgr.fafu.edu.cn/Transcriptome.php>)

NCBI's Conserved Domains Database (<https://www.ncbi.nlm.nih.gov/cdd>)

PFAM (<http://pfam.xfam.org/>)

Expectation Maximization for Motif Elicitation (MEME) (<http://meme-suite.org/tools/meme>)

PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)

Soybase Glyma.Wm82.a2.v1 (<http://www.soybase.org/>)

PRJNA525247 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA525247>)

## References

1. Adrien C., Gal P., Martine N. et al. The Ubiquitous Distribution of Late Embryogenesis Abundant Proteins across Cell Compartments in Arabidopsis Offers Tailored Protection against Abiotic Stress. [J]. *The Plant Cell*, 2014(7): 7. DOI: 10.1105/tpc.114.127316
2. Artur, M. A. S., Zhao T., Ligterink W. et al. Dissecting the Genomic Diversification of Late Embryogenesis Abundant (LEA) Protein Gene Families in Plants). [J]. *Genome Biology and Evolution*, 2019. 11(2): 459–471. DOI: 10.1093/gbe/evy248
3. Anne R., Auer P. L., Marc L. et al. The fate of duplicated genes in a polyploid plant genome. [J]. *Plant Journal*, 2013, 73(1): 143–153. DOI: 10.1111/tpj.12026
4. Bailey T.L., Boden M., Buske F. A. et al. MEME SUITE: tools for motif discovery and searching. [J]. *Nucleic Acids Research*, 2009. 37(Web Server): W202-W208. DOI: 10.1093/nar/gkp335
5. Battaglia M., and Covarrubias A. A. Late Embryogenesis Abundant (LEA) proteins in legumes. [J]. *Frontiers in Plant Science*, 2013, 4(190): 190. DOI: 10.3389/fpls.2013.00190
6. Bhogireddy S., Xavier A., Garg V. et al. Genome-wide transcriptome and physiological analyses provide new insights into peanut drought response mechanisms. [J]. *Scientific Reports*, 2020, 10(1). DOI: 10.1038/s41598-020-60187-z
7. Bies-Ethève N., Gaubier-Comella P., Debures A. et al. Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. [J]. *Plant Molecular Biology*, 2008, 67(1–2): 107–124. DOI: 10.1007/s11103-008-9304-x
8. Blanc G., Wolfe K. H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. [J]. *The Plant Cell Online*, 2004, 16(7): 1667–1678. DOI: 10.1016/j.livsci.2009.01.009



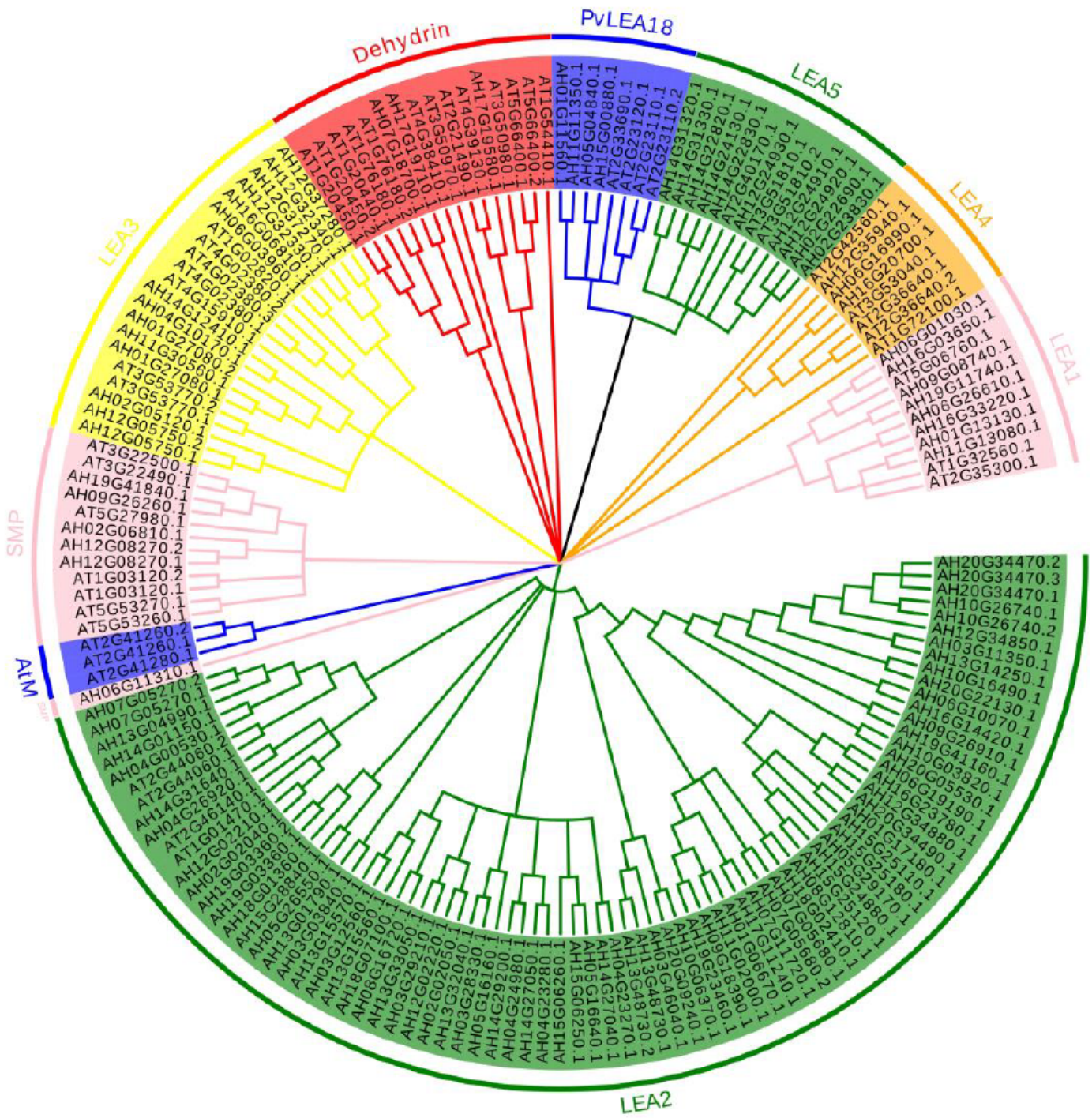
9. Boter, M. Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and *Arabidopsis*. [J]. *Genes and Development*, 2004, 18(13): 1577–1591. DOI: 10.1101/gad.297704
10. Cao J., and Li X. Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*). [J]. *Planta*, 2015, 241(3): 757–772. DOI: 10.1007/s00425-014-2215-y
11. Chen C. J., Chen H., Zhang Y. et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. [J]. *Molecular Plant*, 2020, 13(8). DOI: 10.1016/j.molp.2020.06.009
12. Chen Q., Chen Q. J., Sun G. Q. et al. Genome-Wide Identification of Cyclophilin Gene Family in Cotton and Expression Analysis of the Fibre Development in *Gossypium barbadense*. [J]. *International Journal of Molecular Sciences*, 2019, 20(2): 349. DOI: 10.3390/ijms20020349
13. Cunha F., Nobile P. M., Hoshino A. A. et al. Genetic relationships among *Arachis hypogaea* L. (AABB) and diploid *Arachis* species with AA and BB genomes. [J]. *Genetic Resources and Crop Evolution*, 2008, 55(1): 15–20. DOI: 10.1007/s00425-014-2215-y
14. Du D., Zhang Q., Cheng T. et al. Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*. [J]. *Molecular Biology Reports*, 2013, 40(2): 1937–1946. DOI: 10.1007/s11033-012-2250-3
15. Fang L., Magwanga R. O. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in *upland cotton*. [J]. *BMC Genetics*, 2018. DOI: 10.1186/s12863-017-0596-1
16. Flagel L. E., and Wendel J. F. Gene duplication and evolutionary novelty in plants. [J]. *New Phytologist*, 2009, 183(3): 557–564. DOI: 10.1111/j.1469-8137.2009.02923.x
17. Feng K. W., Cui L. C., Wang L. et al. The improved assembly of 7DL chromosome provides insight into the structure and evolution of bread wheat. [J]. *Plant Biotechnology Journal*, 2020, 18(3). DOI: 10.1111/pbi.13240
18. Hernandez-Garcia C. M., and Finer J. J. A novel cis-acting element in the GmERF3 promoter contributes to inducible gene expression in soybean and tobacco after wounding. [J]. *Plant Cell Reports*, 2016, 35(2): 303–316. DOI: 10.1007/s00299-015-1885-7
19. Higo K., Ugawa Y., Iwamoto M. et al. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. [J]. *Nucleic Acids Research*, 1999, 27(1): 297–300. DOI: 10.1093/nar/27.1.297
20. Hundertmark M., and Hinch D. K. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. [J]. *BMC Genomics*, 2008, 9(1): 118. DOI: 10.1186/1471-2164-9-118
21. Ibrahime M., Kibar U., Kazan K., et al. Genome-wide identification of the LEA protein gene family in grapevine (*Vitis vinifera* L.). [J]. *Tree Genetics and Genomes*, 2019, 15(4): 55.1-55.14. DOI: 10.1007/s11295-019-1364-3
22. Jeffares, D. C., Penkett C.J., Bahler J. Rapidly regulated genes are intron poor. [J]. *Trends in Genetics*, 2008, 24(8): 375–378. DOI: 10.1016/j.tig.2008.05.006
23. Jin X. F., Cao D., Wang Z. J. 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. [J]. *Scientific Reports*, 2019, 9(1): 14123–14123. DOI: 10.1038/s41598-019-50645-8

23. Kent W. J., Baertsch R., Hinrichs A. et al. Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. [J]. *Proceedings of the National Academy of Sciences*, 2003, 100(20): 11484–11489. DOI: 10.1073/pnas.1932072100
24. Kentaro S., Kirilov C. N., Sakae T. et al. Identification of a novel LEA protein involved in freezing tolerance in wheat. [J]. *Plant and cell physiology*, 2014, 55(1) 136–147. DOI: 10.1093/pcp/pct164
25. Koch M. A., Bernhard H., Thomas M. O. Comparative Evolutionary Analysis of Chalcone Synthase and Alcohol Dehydrogenase Loci in *Arabidopsis*, *Arabis*, and Related Genera (Brassicaceae). [J]. *Molecular Biology and Evolution*, 2000(10): 1483-98. DOI: 10.1161/01.STR.0000221702.75002.66
26. Koubaa, S., and Brini F. Functional analysis of a wheat group 3 late embryogenesis abundant protein (TdLEA3) in *Arabidopsis thaliana* under abiotic and biotic stresses. [J]. *Plant Physiology and Biochemistry*, 2020, 156: 396–406. DOI: 10.1016/j.plaphy.2020.09.028
27. Lan, T., Jie, G., Zeng Q. Y. Genome-wide analysis of the lea (late embryogenesis abundant) protein gene family in *Populus trichocarpa*. [J]. *Tree Genetics & Genomes*, 2013, 9(1), 253–264. DOI: 10.1007/s11295-019-1364-3
28. Lane N., and Martin W. The energetics of genome complexity. [J]. *Nature*, 2010, 467(7318): 929–34. DOI: 10.1038/nature09486
29. Li L., Xu H. L., Yang X L. et al. Genome-Wide Identification, Classification and Expression Analysis of LEA Gene Family in Soybean. [J]. *Scientia Agricultura Sinica*, 2011. DOI: 10.3864/j.issn.0578-1752
30. Li C., Ng K. Y., Fan L. M. MYB transcription factors, active players in abiotic stress signaling. [J]. *Environmental and Experimental Botany*, 2015, 114: 80–91. DOI: 10.1016/j.envexpbot.2014.06.014
31. Liang Y., Xiong Z., Zheng J. et al. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*. [J]. *Scientific Reports*, 2016, 6: 24265. DOI: 10.1038/srep24265
32. Liu H., Xing M., Yang W. et al. Genome-wide identification of and functional insights into the late embryogenesis abundant (LEA) gene family in bread wheat (*Triticum aestivum*). [J]. *Scientific Reports*, 2019, 9(1): 1–11. DOI: 10.1038/s41598-019-49759-w
33. Mehan M. R., Freimer N. B., Ophoff R. A. A genome-wide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture. [J]. *Human Genomics*, 1, 5(2004-08-01), 2004, 1(5): 335–344. DOI: 10.1186/1479-7364-1-5-335
34. Mota A., Oliveira T. N., Vinson C. C. et al. Contrasting Effects of Wild *Arachis* Dehydrin Under Abiotic and Biotic Stresses. [J]. *Frontiers in Plant Science*, 2019, 10. DOI: 10.3389/fpls.2019.00497
35. Mowla S. B., Cuyper A., Driscoll S. P. et al. Yeast complementation reveals a role for an *Arabidopsis thaliana* late embryogenesis abundant (LEA)-like protein in oxidative stress tolerance. [J]. *Plant Journal*, 2010, 48(5): 743–756. DOI: 10.1111/j.1365-313X.2006.02911.x
36. Park J. A., Cho S. K., Kim J. E. et al. Isolation of cDNAs differentially expressed in response to drought stress and characterization of the Ca-LEAL1 gene encoding a new family of atypical LEA-like protein homologue in hot pepper (*Capsicum annuum L. cv. Pukang*). [J]. *Plant Science*, 2003, 165(3): 471–481. DOI: 10.1016/S0168-9452(03)00165-1

37. Pedrosa A. M., Martins C., Gonçalves L. P. et al. Late Embryogenesis Abundant (LEA) Constitutes a Large and Diverse Family of Proteins Involved in Development and Abiotic Stress Responses in Sweet Orange (*Citrus sinensis* L. Osb.). [J]. Plos One, 2015, 10(12): e0145785. DOI: 10.1371/journal.pone.0145785
38. Qiu W., Wang N., Dai J. et al. AhFRDL1-mediated citrate secretion contributes to adaptation to iron deficiency and aluminum stress in peanuts. [J]. Journal of Experimental Botany, 2019. 70(10): 2873–2886. DOI: 10.1093/jxb/erz089
39. Shao H. B., Liang Z. S., Shao M. A. LEA proteins in higher plants: structure, function, gene expression and regulation. [J]. Colloids and Surfaces B Biointerfaces, 2005, 45(3–4): 131–135. DOI: 10.1016/j.colsurfb.2005.07.017
40. Shi H. F., He X. Y., Zhao Y. J. et al. Constitutive expression of a group 3 LEA protein from *Medicago falcata* (MfLEA3) increases cold and drought tolerance in transgenic tobacco. [J]. Plant Cell Reports, 2020, 39(7): 851–860. DOI: 10.1007/s00299-020-02534-y
41. Sivamani E., Bahieldin A., Wraith J. M. et al. Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley HVA1 gene. [J]. Plant Science, 2000, 155(1): 1–9. DOI: 10.1016/S0168-9452(99)00247-2
42. Sudhir K., Glen S., Koichiro T. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. [J]. Molecular Biology and Evolution, 2016, 33(7): 1870–1874. DOI: 10.1093/molbev/msw054
43. Wang X. S., Zhu H. B., Jin G. L. et al. Genome-scale identification and analysis of LEA genes in rice (*Oryza sativa* L.). [J]. Plant Science, 2007, 172(2): 414–420. DOI: 10.1016/j.plantsci.2006.10.004
44. Wang X., Wu M. H., Xiao D. et al. Genome-wide identification and evolutionary analysis of RLKs involved in the response to aluminium stress in peanut. [J]. BMC Plant Biology, 2021, 21(1). DOI: 10.1186/s12870-021-03031-4
45. Xiao D, Li X, Zhou Y. Y. et al. Transcriptome analysis reveals significant difference in gene expression and pathways between two peanut cultivars under Al stress. [J]. Gene, 2021, 781(145535). DOI: 10.1016/J.GENE.2021.145535
46. Xie D. W., Wang X. N., Fu L. S. et al. Identification of the trehalose-6-phosphate synthase gene family in winter wheat and expression analysis under conditions of freezing stress. [J]. Journal of Genetics, 2015, 94(1): 55–65. DOI: 10.1007/s12041-015-0495-z
47. Xu D., Duan X., Wang B. et al. Expression of a Late Embryogenesis Abundant Protein Gene, HVA1, from Barley Confers Tolerance to Water Deficit and Salt Stress in Transgenic Rice. [J]. Plant Physiology, 1996, 110(1): 249–257. DOI: 10.1104/pp.110.1.249
48. Yamaguchi-Shinozaki K., and Shinozaki K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. [J]. Trends in Plant Science, 2005, 10(2): 88–94. DOI: 10.1016/j.tplants.2004.12.012
49. Yu L., Kai K., Lu G. et al. Drought-responsive genes, late embryogenesis abundant group 3 (LEA3) and vicinal oxygen chelate (VOC), function in lipid accumulation in *Brassica napus* and Arabidopsis

- mainly via enhancing photosynthetic efficiency and reducing ROS. [J]. *Plant Biotechnology Journal*, 2019, 17(11): 2123–2142. DOI: 10.1111/pbi.13127
50. Yun L., and Zheng Y. PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in *Escherichia coli*. [J]. *Biochemical and Biophysical Research Communications*, 2005, 331(1): 325–332. DOI: 10.1016/j.bbrc.2005.03.165
51. Yun L., Zheng Y., Zhang Y. et al. Soybean PM2 Protein (LEA3) Confers the Tolerance of *Escherichia*. [J]. 2010, 60(5): 373–378. DOI: 10.1007/s00284-009-9552-2
52. Ye J., Yang X., Hu G. et al. Genome-Wide Investigation of Heat Shock Transcription Factor Family in Wheat (*Triticum aestivum* L.) and Possible Roles in Anther Development. [J]. *International Journal of Molecular Sciences*, 2020, 21(2). DOI: 10.3390/ijms21020608
53. Zamora-Briseño, J. A. and de Jiménez E. S. A LEA 4 protein up-regulated by ABA is involved in drought response in maize roots. [J]. *Molecular Biology Reports*, 2016. 43(4): 221–228. DOI: 10.1007/s11033-016-3963-5
54. Zegzouti H., Jones B., Marty C. et al. Er5, a tomato cDNA encoding an ethylene-responsive LEA-like protein: characterization and expression in response to drought, ABA and wounding. [J]. *Plant Molecular Biology*, 1997, 35(6): 847–854. DOI: 10.1023/A:1005860302313
55. Zhan J., Kou R. J., He L. F. et al. Effects of aluminum on morphological structure of peanut root tips. [J]. *Chinese Journal of Oil Crop Sciences*, 2008(01): 79–83. DOI: 10.3724/SP.J.1011.2008.00534
56. Zhuang W. J., Chen H., Yang M. et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution, and crop domestication. [J]. *Nature Genetics*, 2019, 51(5): 865–876. DOI: 10.1038/s41588-019-0402-2

## Figures

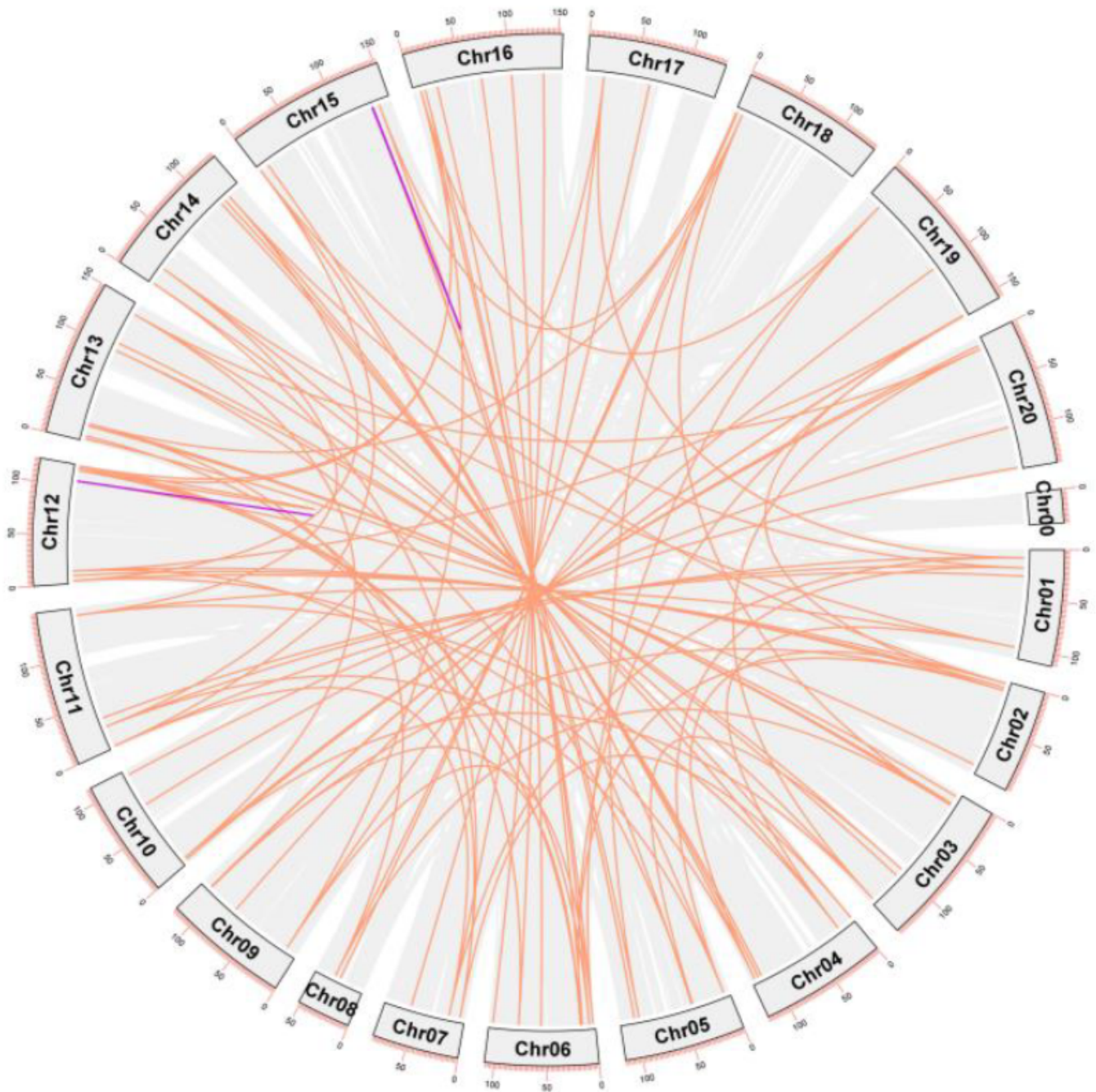


**Figure 1**

Phylogenetic relationships of the AhLEAs and AtLEAs. The Maximum Likelihood (ML) tree was generated using MEGA7 with 1000 bootstrap replicates. LEA gene families are distinguished by different colors.

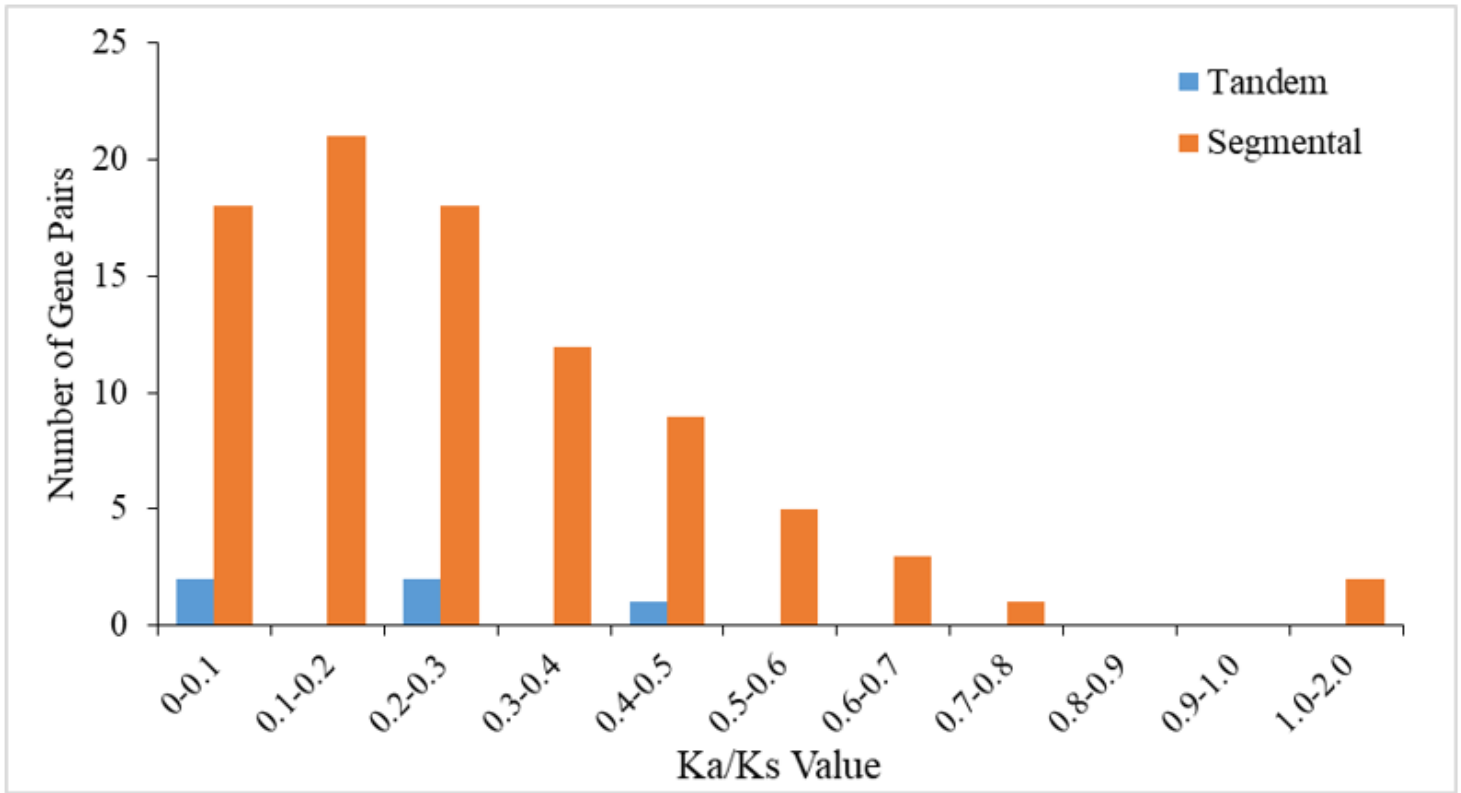






**Figure 3**

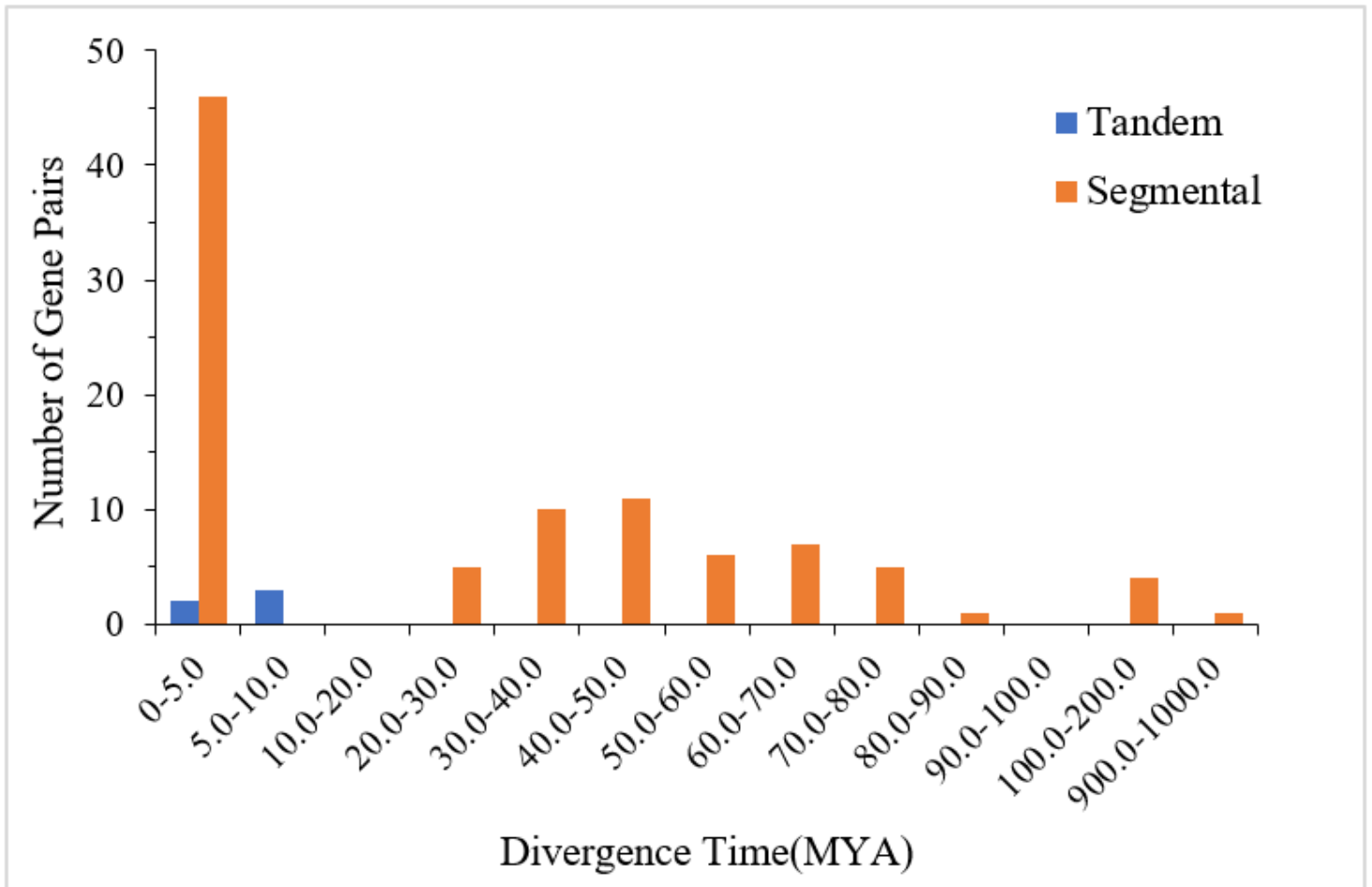
Duplication analysis of 126 AhLEAs. The rectangle on the outer ring represents peanut chromosome 00-20. The purple line on chromosomes 12, 15 represents tandem duplication gene pairs, and light orange lines on chromosomes represent segmental duplication gene pairs.



**Figure 4**

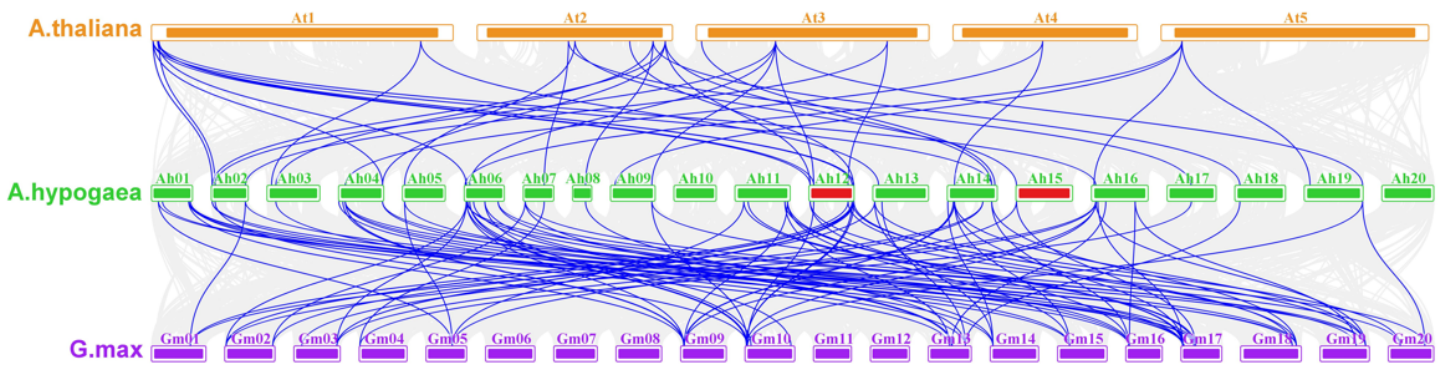
The distribution of Ka/Ks values in all tandem and segmental duplicated AhLEAs.





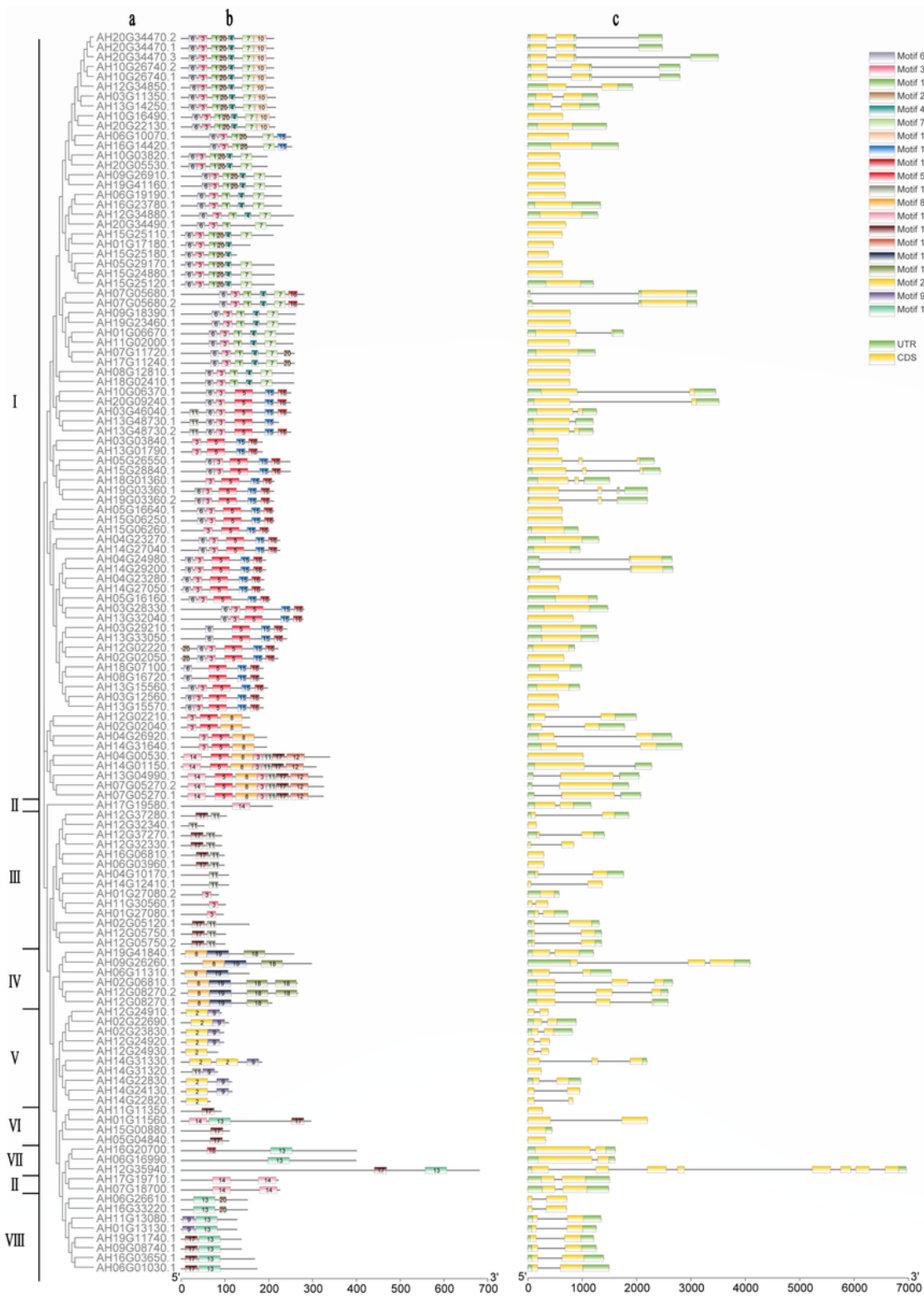
**Figure 5**

The distribution of divergence time (MYA) in all tandem and segmental duplicated AhLEAs.



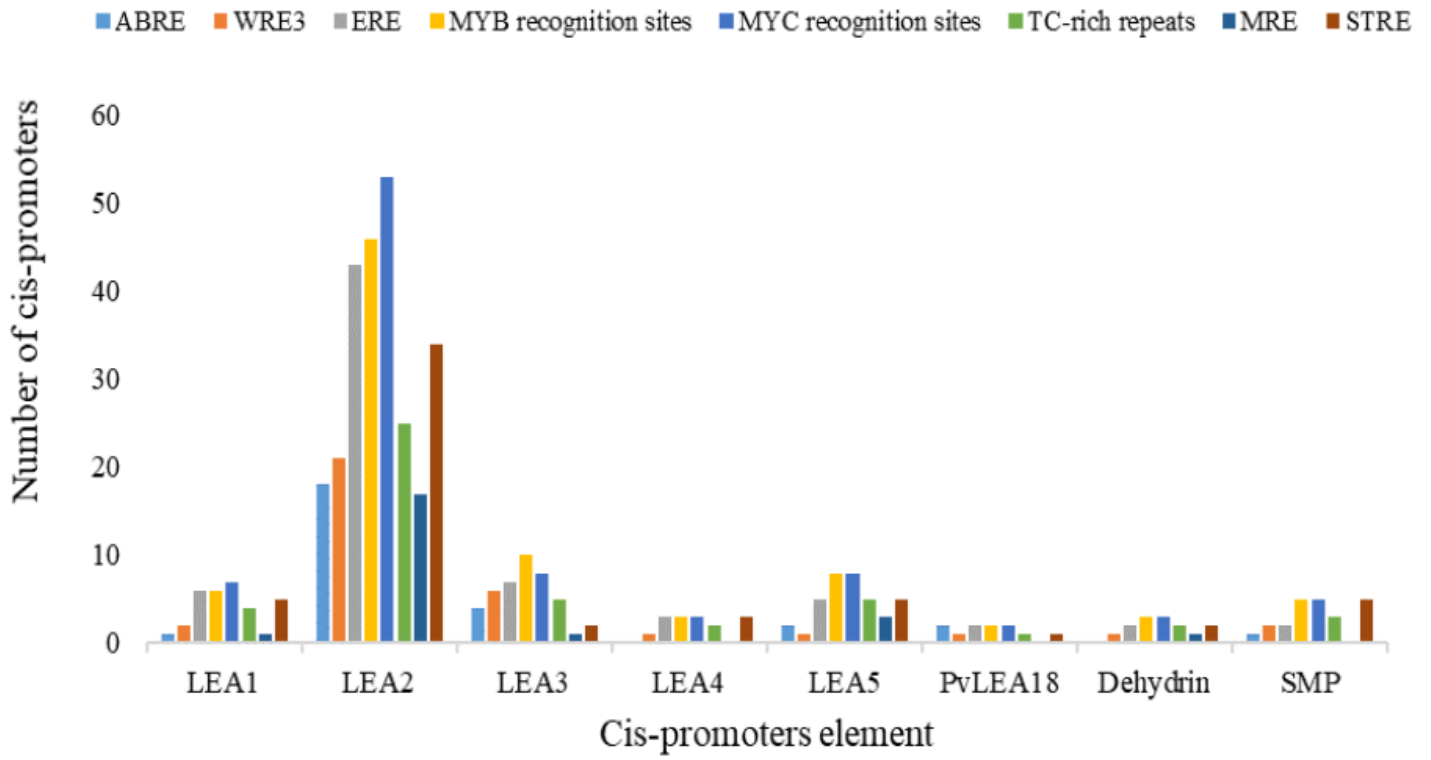
**Figure 6**

Synteny analyses of AhLEAs to Arabidopsis and G. max. Gray lines in the background indicate collinear blocks within peanut and Arabidopsis, soybean genomes, while blue lines highlight syntenic LEA gene pairs, Red chromosome blocks represent tandem duplicated genes.



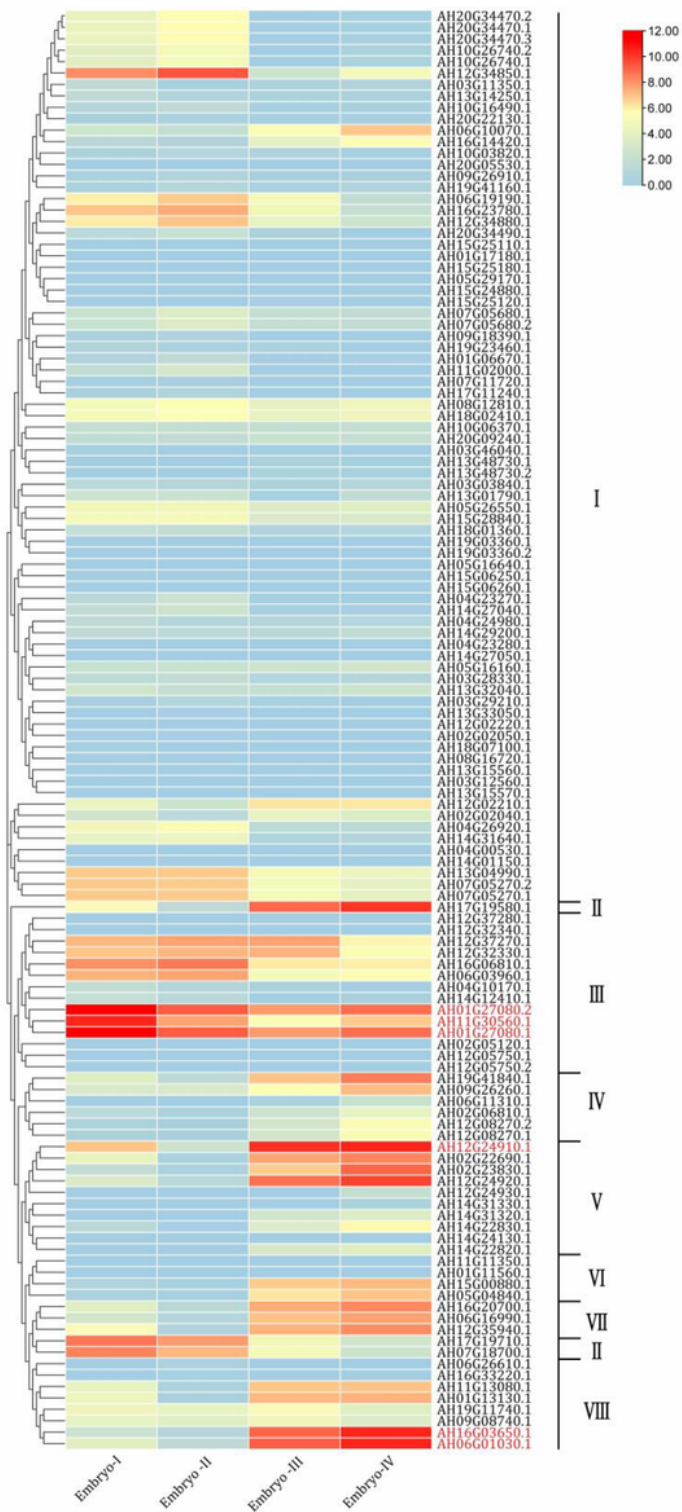
**Figure 7**

Phylogenetic relationships, gene structures, and compositions of the conserved protein motifs of the AhLEAs. ☒: LEA2; ☒: Dehydrin; ☒: LEA3; ☒: SMP; ☒: LEA5; ☒: PvLEA18; ☒: LEA4; ☒: LEA1; a: Phylogenetic relationships, b: conserved motif, c: UTR–CDS organization, black lines represent intron.



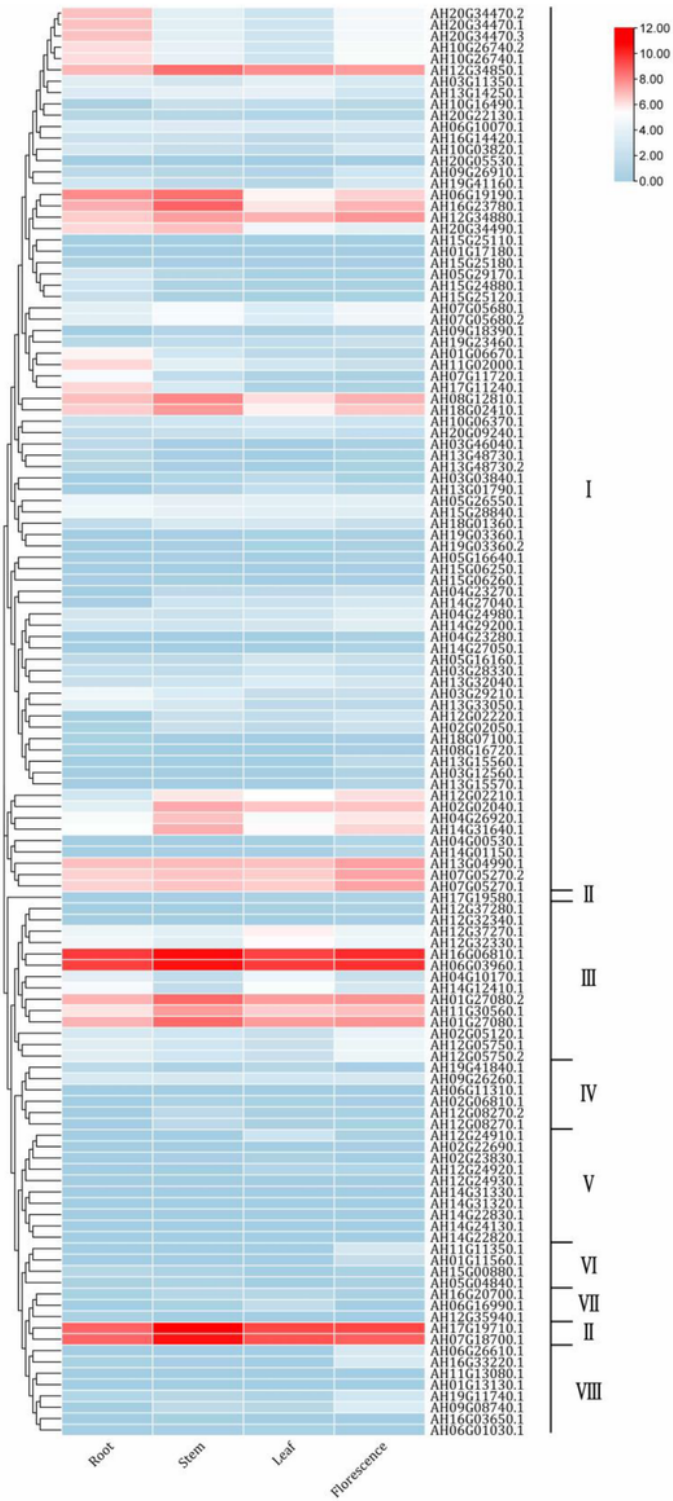
**Figure 8**

Distribution of major abiotic stress-responsive cis-elements in the promoter sequences of the 126 AhLEAs.



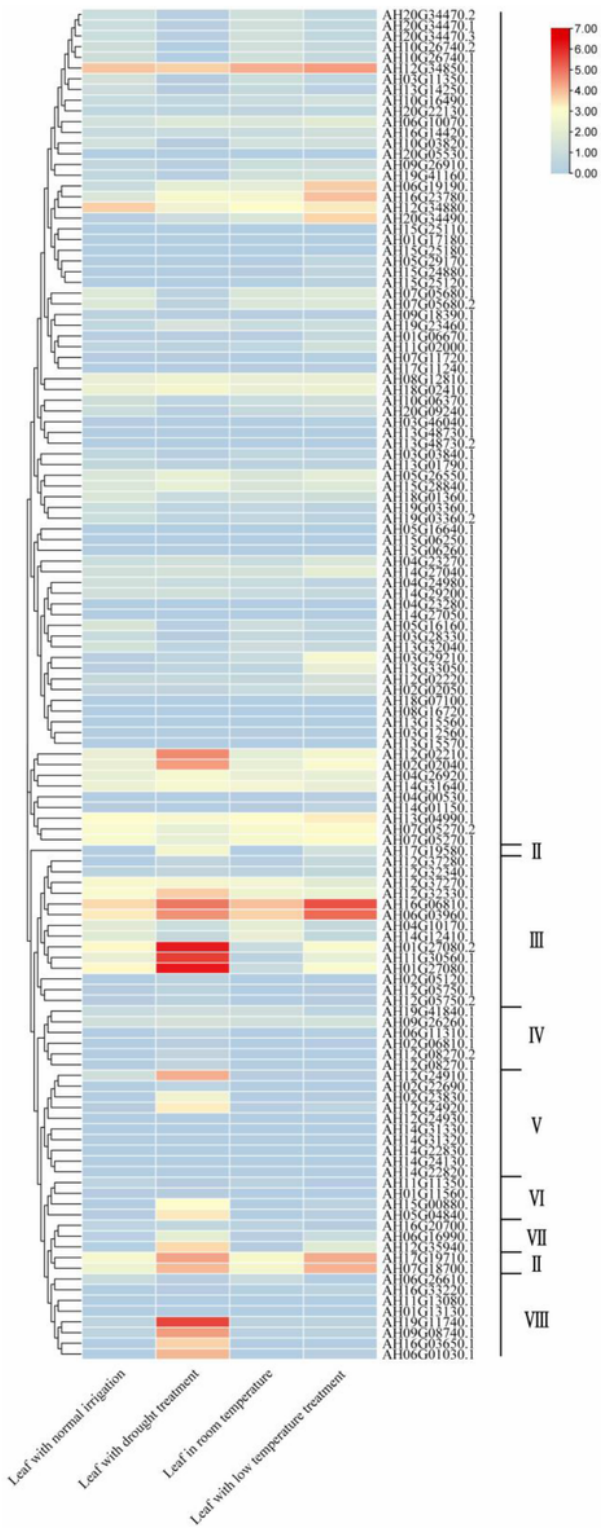
**Figure 9**

A heatmap showing the hierarchical clustering of the expression levels of the 126 AhLEAs in the four embryo periods in peanut. ☐: LEA2; ☐: Dehydrin; ☐: LEA3; ☐: SMP; ☐: LEA5; ☐: PvLEA18; ☐: LEA4; ☐: LEA1.



**Figure 10**

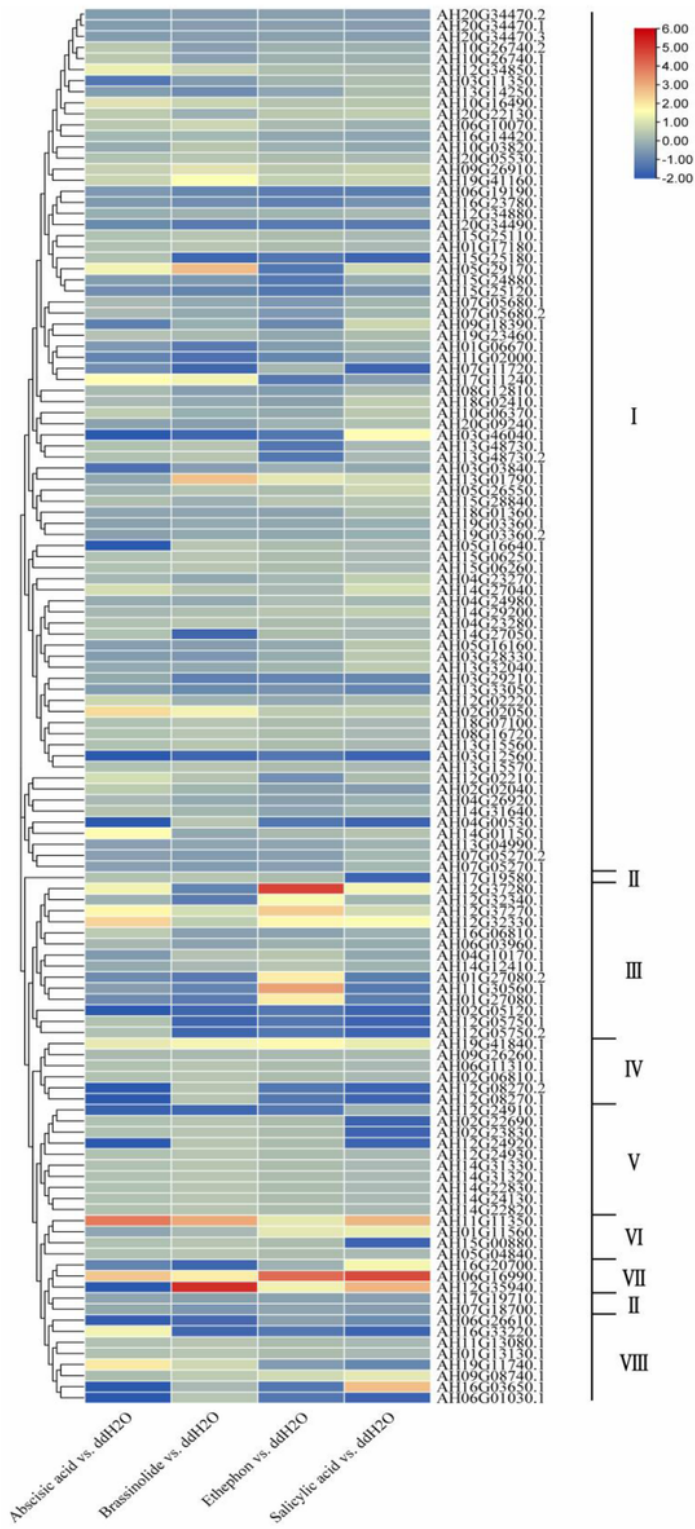
A heatmap showing the hierarchical clustering of the expression levels of the 126 AhLEAs in the roots, stems, leaves, and flowers of peanut. ☐: LEA2; ☐: Dehydrin; ☐: LEA3; ☐: SMP; ☐: LEA5; ☐: PvLEA18; ☐: LEA4; ☐: LEA1.



**Figure 11**

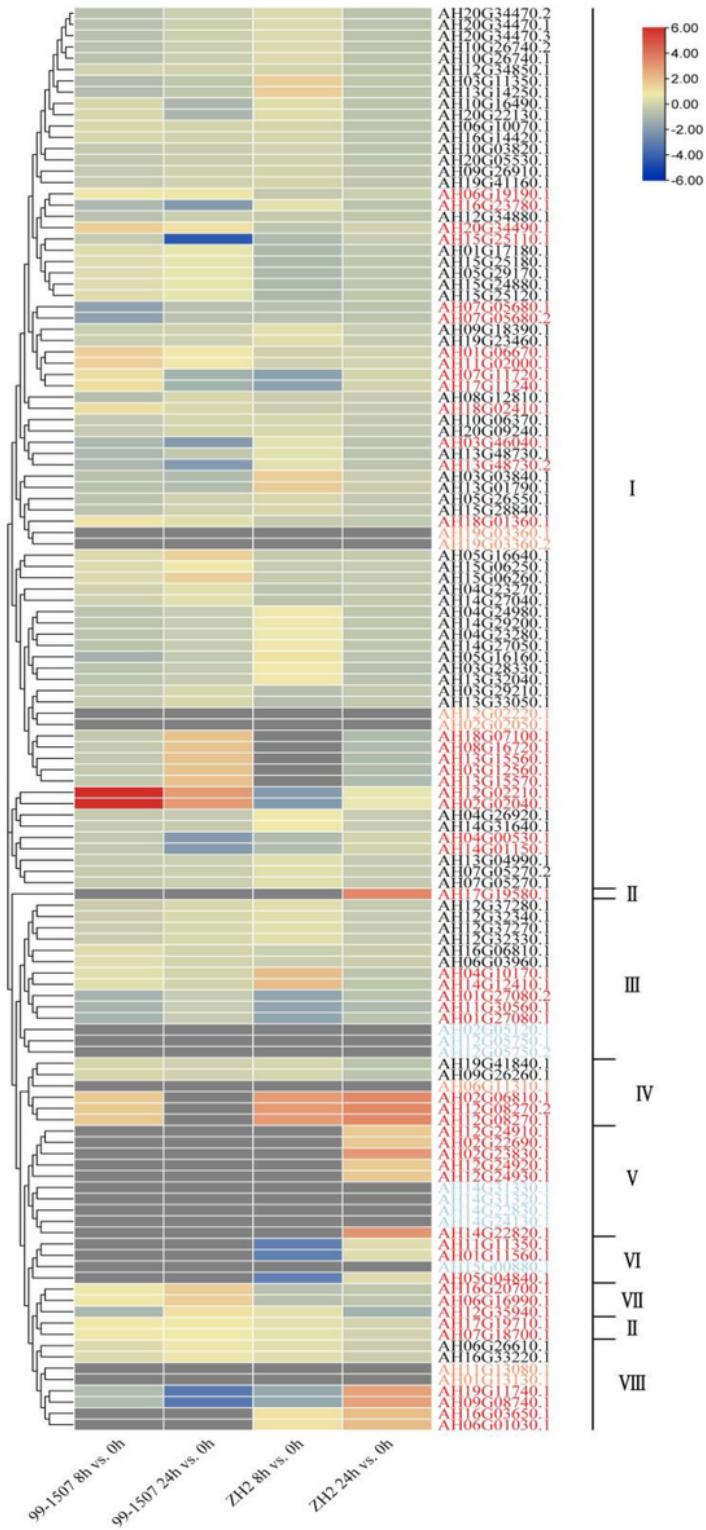
Expression profiles of the AhLEAs in peanut. Dynamic expression profiles of AhLEAs drought and low-temperature treatments using heatmap of hierarchical clustering. ☒: LEA2; ☒: Dehydrin; ☒: LEA3; ☒: SMP; ☒: LEA5; ☒: PvLEA18; ☒: LEA4; ☒: LEA1.





**Figure 12**

A heatmap showing the hierarchical clustering of the expression levels of the 126 AhLEAs under different hormone treatments in peanuts. ☐: LEA2; ☐: Dehydrin; ☐: LEA3; ☐: SMP; ☐: LEA5; ☐: PvLEA18; ☐: LEA4; ☐: LEA1.



**Figure 13**

Expression profiles of Al-responsive AhLEAs in two varieties. The blue words represent not detected in the RNA-Seq dataset, and the orange words represent LEA genes were not expressed, the red words represent 50 differentially expressed genes. ☐: LEA2; ☐: Dehydrin; ☐: LEA3; ☐: SMP; ☐: LEA5; ☐: PvLEA18; ☐: LEA4; ☐: LEA1.



## Supplementary Files

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

- [Additionalfile1.xlsx](#)
- [Additionalfile2.docx](#)