Genome-Wide Survey and Identification of AP2/ERF Genes Involved in Shoot and Leaf Development in Liriodendron Chinense

Yaxian Zong
Nanjing Forestry University https://orcid.org/0000-0001-6263-8278
Ziyuan Hao
Nanjing Forestry University
Zhonghua Tu
Nanjing Forestry University
Yufang Shen
Nanjing Forestry University
Chengge Zhang
Nanjing Forestry University
Shaoying Wen
Nanjing Forestry University
Lichun Yang
Nanjing Forestry University
Jikai Ma
Nanjing Forestry University
Huogen Li (hgli@njfu.edu.cn)
Nanjing Forestry University

Research article

Keywords: Liriodendron chinense, AP2/ERF family, transcription factors, shoot-specific, leaf development

Posted Date: August 12th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-48902/v1

Version of Record: A version of this preprint was published at BMC Genomics on November 8th, 2021.
See the published version at https://doi.org/10.1186/s12864-021-08119-7.

Abstract

Background: *Liriodendron chinense* is a distinctive ornamental tree species due to its unique-shaped leaves and tulip-like flowers. The discovery of genes involved in leaf development and morphogenesis is critical for uncovering the underlying genetic basis of these traits. Genes in the AP2/ERF family were recognized as plant-specific transcription factors contributing to plant growth, hormone-induced development, ethylene response factors and stress responses.

Results: In this study, we identified 104 putative AP2/ERF genes in the recently released *Liriodendron* genome database and RNA-seq dataset. Accordingly, all 104 genes were grouped into four subfamilies, including the AP2, ERF, RAV and Soloist subfamilies. This classification was further supported by the results of gene structure and conserved motif analyses. Moreover, based on the expression profiles of various tissues and organs, we discovered three LcAP2 genes as well as two VIII group genes that were significantly enriched in a shoot-specific manner by applying expression pattern and category enrichment analyses. Of these five genes, the relative expression levels of LcERF 94, LcERF 96 and LcERF 98 in the RT-qPCR assay were highly consistent with the RNA-seq results. Furthermore, we illustrated by dissection of *Liriodendron* shoots and subsequent qPCR assays that LcERF 94, LcERF 96 and LcERF 98 were expressed extensively in the early stage of leaf primordium development but rarely in tender leaves. In addition, these three genes displayed nuclear subcellular localizations through transient transformation of tobacco epidermal cells.

Conclusions: Taken together, we identified all the AP2/ERF family members at the genome-wide level and provided candidates that might participate in the development and morphogenesis of the leaf primordium in *L. chinense*.

Background

Plant morphogenesis is mainly related to the shoot as well as the activity of the shoot apical meristem (SAM) [1, 2], which further gives rise to stems, tender leaves or other tissues and organs. Plant endogenous hormones are recognized to play a crucial role in regulating leaf development and morphogenesis [3, 4]. Moreover, functional dissections of plant transcription factors (TFs), such as NAC [5], KNOX [6], and AP2/ERF [7, 8], which have been largely described as involved in shoot and leaf development, have been revealed from studies on *Arabidopsis*. Hence, it is extremely essential to discover the genes and to unpack the genetic basis of shoot and leaf development.

The APETALA2/ethylene responsive element-binding proteins (AP2/EREBPs) are a well-known family of TFs that have been reported to be involved in ethylene response, biotic or abiotic stress resistance, cell differentiation, cell expansion and stress signaling pathways in plants [9–11]. Accordingly, the AP2/EREBP family is generally regarded as one of the largest families of plant TFs and is composed of four subfamilies of over 100 members in various taxa[12–14]. A feature of the AP2/ERF-type DNA binding domain, which consists of 60 ~ 70 residues, is universally exposed in this family [15]. Moreover,

on the basis of the types and quantities of conserved domains, the AP2/ERF family can be divided into the AP2, RAV, ERF and Soloist subfamilies. Proteins of the AP2 subfamily contain two repetitive AP2 domains; however, they have only one of the ERF proteins. In the RAV subfamily, tandem repetition of an AP2 domain and a B3 domain was found in the primary protein sequences [16–18]. Finally, the Soloist subfamily was historically regarded as an ERF member and is currently regrouped in a novel subfamily due to its single AP2 domain and strong sequence divergence [16, 19]. In addition, based on the sequence similarity of the AP2/ERF motifs, the ERF family is further divided into the ERF and DREB subfamilies [18, 20].

Although four clades have been maintained in the AP2/ERF family, their functions largely depend on the interaction of their motifs with specific regulatory elements [16]. In general, the ERF and DREB subfamilies primarily function in resistance to biotic or abiotic stresses [21], and the “WLG” motif is considered a typical feature of these subfamilies [18]. Additionally, the DREB subfamily is mainly involved in the resistance of plants via the interaction of a core motif of A/GCCGAC with downstream dehydration responsive elements (DREs) [22]. The ERF subfamily participates in defense by integrating the cis-acting element AGCCGCC with its GCC box [23, 24]. However, this is not always the case, and this resistance can be interrupted by the VIII groups of these families [25]. DRN and DRNL genes in Arabidopsis hierarchically interact in auxin signaling and patterning of the apical embryo. In addition, LEAFY PETIOLE (LEP) acts as a positive regulator of gibberellic acid-induced germination and is involved in the formation of petioles [26]. In addition, the AP2 subfamily usually regulates the development of shoots as well as the stem cell niche during embryonic pattern formation [27, 28]. In summary, these inferences from previous studies provide direction for comprehensively understanding AP2/ERFs and the discovery of novel genes involved in leaf and shoot development.

Liriodendron chinense (Hemsl.) Sarg. is a distinctive ornamental tree species native to southern China. Owing to its unique leaf shape and yellowish green flower, it is widely used as an ornamental in landscapes and gardens. As a valuable foliage plant, it is essential to understand the development and morphogenesis of *Liriodendron* leaves. However, limited by a lack of genomic sequencing and expression profile analyses reported for this tree species, the underlying genetic mechanism of leaf development and morphogenesis is poorly understood. Along with the recent release of the *Liriodendron* genome and massive transcriptome information [29, 30], it is now possible for us to illustrate which genes are involved in this process as well as the underlying mechanisms.

In this study, we identified 104 LcAP2/ERF genes through genome-wide scanning and complementation of the transcriptome from various tissues and periods. Meanwhile, we described conserved motif, gene structure and phylogenetic analyses and divided the LcAP2/ERF genes into four subfamilies of approximately 14 groups. Through expression profile analysis and functional enrichment, we discovered the genes involved in shoot and leaf development and further illustrated the expression patterns of these candidates in the SAM and spires at different developmental stages. This work will lay a foundation for the comprehensive understanding of the LcAP2/ERF family and will also be helpful for determining candidate genes involved in leaf development in *L. chinense*.

Results

Identification of AP2/ERF TFs in Liriodendron chinense

Based on the HMM profiles (PF00847) and homology searches, a total of 104 putative AP2/ERF genes that have been designated from LcERF1 to LcERF104 were identified in *L. chinense*. All these candidates contained one or more AP2/ERF domains according to conservative domain analysis. Then, we described the characteristics of their proteins, including coding sequence (CDS) length, protein length, molecular weight (MW), isoelectric point (PI) and predicted subcellular localization (see Additional file 1: Table S1). Accordingly, the protein lengths of these 104 AP2/ERFs ranged from 100 aa (*LcERF29*) to 758 aa (*LcERF42*), with an average of approximately 317 aa (Table 1). Moreover, the molecular weight of the proteins varied from 11.48 kDa (*LcERF29*) to 84.42 KDa (*LcERF42*). In addition, the isoelectric point values of these proteins ranged from 4.72 (*LcERF27*) to 10.22 (*LcERF67*). The predicted subcellular localization results showed that 83 LcERF proteins were located in the nuclear region, 13 LcERF proteins were located in the chloroplast region, and the remaining genes were distributed in the cytoplasm, mitochondria, plasma membrane and other areas.

Table 1
List of the 104 *AP2/ERF* genes identified in *Liriodendron Chinense*

Gene name	Gene ID	Location	Protein length(aa)	Introns	Family group
LcERF1	Unigene40981_All	Scaffold211	261	0	I
LcERF2	Lchi03057	Scaffold506	328	1	I
LcERF3	Lchi07965	Scaffold708	325	1	I
LcERF4	Lchi07966	Scaffold708	561	2	I
LcERF5	Lchi22931	Scaffold1519	432	1	I
LcERF6	Lchi09796	Scaffold2048	316	1	I
LcERF7	Lchi16995	Scaffold3097	316	0	I
LcERF8	Lchi23250	Scaffold142	152	0	II
LcERF9	Lchi16170	Scaffold408	193	1	II
LcERF10	Unigene12650_All	Scaffold416	188	0	II
LcERF11	Lchi16911	Scaffold480	199	1	II
LcERF12	Unigene40401_All	Scaffold525	185	0	II
LcERF13	Unigene20830_All	Scaffold836	186	0	II
LcERF14	Lchi11957	Scaffold345	224	1	III
LcERF15	Lchi04946	Scaffold530	235	1	III
LcERF16	Lchi04947	Scaffold530	295	1	III
LcERF17	CL2522.Contig2_All	Scaffold530	223	0	III
LcERF18	CL10877.Contig3_All	Scaffold530	223	0	III
LcERF19	Unigene6126_All	Scaffold530	211	0	III
LcERF20	Lchi33109	Scaffold1203	229	1	III
LcERF21	Lchi33111	Scaffold1203	227	1	III
LcERF22	Lchi34895	Scaffold1374	513	4	III
LcERF23	Lchi29925	Scaffold1675	425	3	III
LcERF24	Lchi08587	Scaffold39	420	1	III
LcERF25	CL5589.Contig2_All	Scaffold345	203	0	III
LcERF26	Unigene11386_All	Scaffold432	211	0	III
LcERF27	CL5589.Contig1_All	Scaffold530	246	0	III

Gene name	Gene ID	Location	Protein length(aa)	Introns	Family group
LcERF28	Lchi00950	Scaffold723	217	0	III
LcERF29	Lchi01616	Scaffold1191	100	1	III
LcERF30	Unigene5530_All	Scaffold1191	252	0	III
LcERF31	Lchi32377	Scaffold1289	210	0	III
LcERF32	Lchi26370	Scaffold1364	193	1	III
LcERF33	Lchi08922	Scaffold3419	244	0	III
LcERF34	Lchi28169	Scaffold654	418	1	IV
LcERF35	Lchi23878	Scaffold1043	141	0	IV
LcERF36	Lchi22387	Scaffold1263	229	1	IV
LcERF37	Lchi13652	Scaffold1315	429	1	IV
LcERF38	Lchi30363	Scaffold2365	475	3	IV
LcERF39	Lchi30365	Scaffold2365	354	1	IV
LcERF40	Lchi31374	Scaffold3032	404	1	IV
LcERF41	Lchi34724	Scaffold3708	355	1	IV
LcERF42	Lchi10868	Scaffold159	758	2	V
LcERF43	Lchi11945	Scaffold345	421	7	V
LcERF44	Lchi25937	Scaffold1371	183	1	V
LcERF45	Lchi16637	Scaffold2432	226	1	V
LcERF46	Lchi34468	Scaffold2926	206	1	V
LcERF47	Lchi05084	Scaffold3476	136	1	V
LcERF48	Lchi07311	Scaffold172	268	0	VI
LcERF49	Lchi22103	Scaffold920	369	2	VI
LcERF50	Lchi17039	Scaffold3097	359	1	VI
LcERF51	Lchi02638	Scaffold416	310	1	VII
LcERF52	Lchi02639	Scaffold416	362	1	VII
LcERF53	Lchi11452	Scaffold525	383	1	VII
LcERF54	Lchi04620	Scaffold775	290	1	VII
LcERF55	Lchi04621	Scaffold775	289	1	VII

Gene name	Gene ID	Location	Protein length(aa)	Introns	Family group
LcERF56	Lchi04623	Scaffold775	236	1	VII
LcERF57	Lchi07083	Scaffold135	273	1	VIII
LcERF58	Lchi07084	Scaffold135	203	0	VIII
LcERF59	Lchi13371	Scaffold1075	207	1	VIII
LcERF60	Lchi11824	Scaffold1130	375	3	VIII
LcERF61	Lchi13392	Scaffold1763	207	1	VIII
LcERF62	Unigene7795_All	Scaffold1763	205	0	VIII
LcERF63	Lchi31572	Scaffold1784	180	1	VIII
LcERF64	Lchi08484	Scaffold39	204	1	IX
LcERF65	Lchi09908	Scaffold79	316	1	IX
LcERF66	Unigene24905_All	Scaffold79	170	0	IX
LcERF67	Lchi01406	Scaffold432	211	1	IX
LcERF68	Unigene35921_All	Scaffold432	301	0	IX
LcERF69	Lchi08172	Scaffold580	324	0	IX
LcERF70	Lchi07909	Scaffold708	174	1	IX
LcERF71	Lchi31530	Scaffold803	105	2	IX
LcERF72	CL9762.Contig1_All	Scaffold1024	307	0	IX
LcERF73	Lchi05992	Scaffold1024	250	1	IX
LcERF74	Lchi05993	Scaffold1024	350	2	IX
LcERF75	Lchi26525	Scaffold1934	272	1	IX
LcERF76	Lchi26532	Scaffold1934	361	3	IX
LcERF77	Lchi28702	Scaffold54	309	2	X
LcERF78	Unigene10666_All	Scaffold100	235	1	X
LcERF79	Lchi02215	Scaffold682	229	1	X
LcERF80	Lchi02216	Scaffold682	196	1	X
LcERF81	Lchi18461	Scaffold943	403	1	X
LcERF82	Lchi11856	Scaffold1130	128	1	X
LcERF83	Lchi01932	Scaffold1191	292	1	X

Gene name	Gene ID	Location	Protein length(aa)	Introns	Family group
LcERF84	Lchi20453	Scaffold1167	328	2	VI-L
LcERF85	Lchi14855	Scaffold41	394	6	AP2
LcERF86	Lchi23120	Scaffold192	405	7	AP2
LcERF87	Lchi16948	Scaffold480	550	9	AP2
LcERF88	Lchi08043	Scaffold502	680	6	AP2
LcERF89	Lchi11241	Scaffold503	375	8	AP2
LcERF90	Lchi28881	Scaffold509	474	7	AP2
LcERF91	Lchi06162	Scaffold527	524	7	AP2
LcERF92	Lchi03252	Scaffold764	535	6	AP2
LcERF93	CL7987.Contig2_All	Scaffold805	572	12	AP2
LcERF94	Unigene5404_All	Scaffold2118	490	6	AP2
LcERF95	Lchi33401	Scaffold2225	563	6	AP2
LcERF96	Lchi13837	Scaffold2467	662	7	AP2
LcERF97	CL6967.Contig2_All	Scaffold2956	327	6	AP2
LcERF98	Unigene39546_All	Scaffold3476	468	7	AP2
LcERF99	Lchi08779	Scaffold67	376	1	RAV
LcERF100	Lchi02516	Scaffold100	607	4	RAV
LcERF101	Lchi02519	Scaffold100	428	2	RAV
LcERF102	Lchi15640	Scaffold1242	354	2	RAV
LcERF103	Lchi23744	Scaffold1330	361	1	RAV
LcERF104	Lchi32356	Scaffold3563	235	5	Soloist

Phylogenetic analysis and classification of LcERF genes

On the basis of conservative domain analysis and multiple alignments of LcAP2/ERF protein sequences, consistent with the classification results in Arabidopsis, the 104 LcERF proteins were categorized into four subfamilies, including ERF, AP2, RAV, and Soloist subfamilies. All 84 ERF genes contained a single AP2/ERF domain, and based on the characteristics of the amino acid sequences and domains which they encode, these genes were further divided into two subfamilies, which were named the DREB and ERF subfamilies and covered 41 and 43 members, respectively. However, among the remaining genes, 14

genes were identified as members of the AP2 family owing to their tandemly repeated double AP2/ERF domain. In addition, 5 genes that did not possess a single AP2/ERF domain but displayed a B3 domain were classified in the RAV subfamily. The last one, *LcERF104*, is homologous with the Arabidopsis Soloist gene (At4g13040) and was classified in the Soloist subfamily. According to the description of Nakano's study [19], the DREB subfamily is comprised of four parts, named I, II, III, and IV, which contain 7, 6, 20, and 8 members, respectively. The ERF subfamily genes can be divided into seven groups based on phylogenetic analysis and belong to the V, VI, VII, VIII, IX, X, and VI-L parts, with 6, 3, 6, 7, 13, 7, and 1 members, respectively. The sequence alignment of LcERF genes showed that the WLG element was highly conserved in the ERF, DREB and RAV subfamilies but less conserved in the AP2 subfamily. However, the RAYD, AA, and other elements were conserved in the AP2 subfamily (Fig. 1).

The evolutionary relationships of all the candidate genes were further illustrated by phylogenetic analysis. According to the unrooted tree profile, AP2, RAV and Soloist were clustered to a separate branch within the subfamily. However, ERF genes were divided into 2 large branches, the ERF branch and the DREB branch, and the ERF and DREB branches were divided into 7 and 4 groups, respectively (Fig. 2). Moreover, these findings coincided with the grouping of the ERF subfamily described in part 3.1 based on the conserved motifs (Fig. 1). In addition, this result showed the same clustering pattern as that obtained by the classification method based on alignment with Arabidopsis (Table 1). As a result, we propose that these 104 putative genes are indeed AP2/EFR family genes in *L. chinense*.

Gene structure and conserved motif analysis of LcERF genes

To further understand the structural composition of LcERF genes, we analyzed the genomic DNA sequence using the online Gene Structure Display Server, with the locations of exons and introns provided by the Liriodendron genomic resource. According to the structural characteristics of LcAP2/ERF genes, the number of introns varied among the distinct subfamilies (Fig. 3A). Except for a few members carrying more than one intron, most of the DREB and ERF subfamily genes have only one intron or even no introns in their genomic DNA. In the AP2 subfamily, all the genes possess numerous introns, with intron numbers ranging from 6 to 12. Furthermore, *LcERF93* is considered to have the most introns with 12, even though most *AP2* genes contain 6 or 7 introns. Moreover, four of the five RAV members possessed one or two introns, and the single Soloist member contained five introns. In addition, the position of introns also presented interesting differences in different subfamilies. As far as the sequences with an intron are concerned, the position of their intron was mostly near the N-terminal or C-terminal, rarely in the middle of the sequence, because these sequences usually consist of a long exon and a very short exon. In general, the members with close evolutionary relationships and those from the same subfamily had similar exon and intron structures in terms of intron number and position and exon length.

Conserved motifs of 104 LcERF genes were identified using the MEME (Multiple Em for Motif Elicitation) tool. A total of 15 conserved motifs were displayed in the 104 LcERF proteins. The amino acid length of

the 15 motifs ranged from 15 to 50. As AP2 DNA-binding motifs, motif 1 and motif 2 joined together and appeared in both DREB and ERF subfamilies, expect for special cases of motif 1, which also existed independently in the RAV and Soloist subfamilies. Even though most of the ERF subfamily members shared the two conserved motifs of motif 1 and motif 2, the other motifs were varied in the different proteins (Fig. 3B). In the DREB subfamily, proteins contained relatively more conserved motifs than in the ERF subfamily, especially in group III and group IV. Motifs 5, 7, 8 and 15 were detected in some group III proteins, and motifs 9, 10, 11 and 12 were found in most group IV members. In the AP2 subfamily, all 14 proteins carried motif 3 and motif 6, and some of them also had motif 4 and motif 12. In addition, motifs 13 and 14 only existed in the B3 domain and were also considered specific to the RAV subfamily.

Expression profiles of LcERF genes in different tissues

We investigated the expression profiles of LcAP2/ERF genes in various tissues by Illumina RNA-Seq data[31] and constructed a heatmap, revealing that 86 LcERF genes were detected in the various tissues, including 34 genes in the DREB subfamily, 35 genes in the ERF subfamily, 12 genes in the AP2 subfamily, 4 genes in the RAV subfamily, and one Soloist gene. To explore the differential expression of these genes in different tissues, the FPKM values were standardized by row with TBtools software. Then, the standardized results were clustered by row and column (Figure 4A). The results showed that several genes were expressed in all tissues and clustered in a large group. In addition, the column cluster divided the other genes based on their different expression patterns, including pistil-specific, stamen-specific, leaf-specific, shoot-specific and other patterns.

Expression patterns of LcERF genes and discovery of shoot-specific genes

To reveal genes involved in shoot and leaf development, we intentionally focused on genes that were expressed specifically in the shoot tissue. All the LcAP2/ERF family genes were divided into ten clusters in accordance with the K-means method in the STEM program. Accordingly, cluster IV and cluster V showed tissue-specific expression in leaves and shoots, respectively. Cluster V contained eight genes, while cluster IV contained only one (Figure 5A). In addition, based on enrichment analysis, the expressed genes were categorized into different LcAP2/ERF groups with an adjusted p-value. Interestingly, six of eight genes showed significant enrichment in cluster V, while the single member of cluster IV failed to pass the significance test (Figure 5B). Among these six genes, three genes are part of the LcERF VIII group (*LcERF57*, *LcERF58* and *LcERF63*), and another three genes belong to the LcAP2 subfamily (*LcERF94*, *LcERF96* and *LcERF98*).

We then annotated the functions of these six genes by submitting sequences to The Arabidopsis Information Resource (TAIR) database. Through alignment and annotation, all three genes from the LcAP2 subfamily were mapped to the *AINTEGUMENTA* (*ANT*) or *AINTEGUMENTA*-like (*AIL*) gene, which is

also considered to be involved in the maintenance of the shoot apical meristem (GO:0010492), the auxin-mediated signaling pathway involved in phyllotactic patterning (GO:0060774), plant organ morphogenesis (GO:1905392), cell division (GO:0051301) and cell growth (GO:0016049). However, *LcERF57*, *LcERF58* and *LcERF62* from the LcERF VIII group showed extensive functions, such as negative regulation of the ethylene-activated signaling pathway (GO:0010105, GO:0009873) and glucosinolate metabolic process (GO:0019760).

Potentiality of shoot-specific gene involvement in shoot and leaf development

The expression of the six candidate genes showing shoot-specific patterns was further verified using RT-qPCR. We determined the expression of these six genes in seven tissues, including leaf, shoot, sepal, petal, stamen, pistil and stem tissues (Figure 4B). Consistent with the RNA-seq results, the patterns of *LcERF94*, *LcERF96* and *LcERF98*, which are from the AP2 subfamily, were relatively shoot specific. *LcERF58* and *LcERF63*, which are from Group VIII in the ERF subfamily, were primarily expressed in shoot as well as flower. However, *LcERF57* was not amplified from cDNA or gDNA after repetitive optimization of primer design and the amplification conditions; as a result, we did not describe its function in the following assay (details are not discussed in this article). Considering the potential functions annotated in the NCBI GenBank and Gene Ontology (GO) databases, dual roles may exist for these three genes in Group VIII, as we inferred. Moreover, this conjecture has been clarified in previous studies, which have proven that ERF VIII subgroup genes play an important role in in vitro shoot regeneration and development [32].

In addition to these expression patterns, we further illustrated the molecular functions by separating shoots into multiple layers of tender leaves and then detecting the expression of the candidate genes in different leaf development stages (Figure 6A). RT-qPCR was performed to detect the expression of target genes from P1 to P6 as well as the SAM (Figure 6B). The results revealed that different expression patterns were present among different subfamily genes. Specifically, expression of the *LcERF58* genes was gradual, rising from P1 to P6, yet the exact opposite was observed for *LcERF94*, *LcERF96* and *LcERF98* from the *LcAP2* subfamily as well as *LcERF63* from the ERF group VIII. Opposite expression patterns in leaf primordia (P1 ~ P2) and tender leaves (P3 ~ P6) suggest that the functions of these genes were fairly different in regulating shoot and leaf development. Compared to the expression of *LcERF58/63*, *LcERF94/96/98* were specifically expressed in shoots, and their expression was almost 100 times that of other tissues. These results also indicated that *LcERF94/96/98* genes may play an essential as well as unique role both in the development and morphogenesis of shoots and leaves.

Subcellular localization of LcERF genes

To investigate the potential function of AP2 genes in transcriptional regulation, we detected the subcellular localization of *LcERF94/96/98* using young tobacco leaves. Confocal microscopy was used to observe and photograph the transient transformed lower epidermal cells of tobacco leaves, and visible, GFP fluorescence, chlorophyll fluorescence and merge field images were obtained (Fig. 7). *35S::GFP*, as a control sample, showed GFP fluorescence in the whole cell. The GFP fluorescence of pBI121-*35S::GFP-LcERF94/96/98* were observed only in the nucleus, which is consistent with the characteristics of TFs, and these histological observations demonstrated the alleged role of *LcERF94/96/98* as TFs localized in the nucleus.

Discussion

The AP2/ERF TFs is ubiquitous in all plant species and plays an important role in various biological processes. It has been widely studied in many species, including *A. thaliana* [19], rice [19], grapes [33], *Populus trichocarpa* [34], tartary buckwheat [12], *Chinese jujube* [35], pear (*Pyrus*) [36], *Medicago truncatula* [37], and *Dimocarpus longan* Lour [38]. As a plant-specific TFs family, global surveys and investigations of the AP2/ERF family help us understand the molecular mechanisms of these genes in plants and, in particular, increase awareness of the roles and functions of these genes in the process of plant morphogenesis. In the present study, we identified 104 AP2/ERF genes in *Liriodendron chinense*, including 84 ERF genes, 14 AP2 genes, 5 RAV genes and only one soloist gene. Based on the statistics of other species [39], the total number of AP2/ERF family members in most plants is over 100, and this number ranges widely. Regardless of whether the number of AP2/ERF genes varied, the proportion of each subfamily in the AP2/ERF family of *L. chinense* was similar to that in other species, and the ERF subfamily was the most abundant group.

Compared with other subfamilies, the ERF subfamily engaged with the majority of genes in the AP2/ERF TF family; depending on the domain features and divergent functions, ERF subfamily genes were divided into several phylogenetic groups in previous studies. In 2002, Sakuma *et al.* [18] classified 121 *Arabidopsis thaliana* ERF genes into 2 main groups, DREB (group A) and ERF (group B), based on the sequence identities of the DNA-binding domains, and each group was further divided into 6 small subgroups. Nakano further refined Sakuma's classification in 2006 and subdivided ERF genes into 12 groups according to the genome annotation and phylogenetic analysis of *A. thaliana* and rice [19]. Similarly, in this study, we also combined the results of blasting all proteins with the *Arabidopsis* database and identified the proteins by phylogenetic analysis. Finally, we classified and named these proteins, consisting of the DREB (group I to IV) and ERF (group V to X and VI-L) subfamilies, based on Nakano's classification.

Since domains and motifs are related to transcriptional activity and protein interactions, generally, the function and characteristics of TFs can be determined by domain and motif analyses [40]. The results of conservative domain analysis showed that motifs 1 and 2 are related to the AP2/ERF domain and exist in all ERF subfamily members. Multiple alignment analysis of the AP2/ERF domain revealed that the residues Trp-28, Leu-29 and Gly-30, also known as "WLG", were completely conserved among the DREB

and ERF subfamilies, and its conservation has also been proven in other species, such as Arabidopsis [19], rice [21] and sorghum [41]. While “WLG” was converted into “YLG” elements in all AP2 subfamily members, the same substitution was found in tartary buckwheat (*Fagopyrum tataricum*) [12]. In addition, different residues are conserved in positions 14 and 19 of the AP2/ERF domain between DREB and ERF; V14 (valine) and E19 (glutamic) are conserved in DREB, while A14 (alanine) and D19 (aspartic) exist in ERF [18]. Most of the genes of the ERF subfamily have alanine at position 14, and valine occupies the corresponding position of the DREB proteins in this study; position 19 is not perfectly conserved. It is noteworthy that these amino acids are both located in the β-sheet of the ERB/AP2 domain, and this is also important for binding to the target elements. The Ala38 and Ala39 of DREB and ERF amino acids were previously reported on the α-helix and might also play a crucial role in binding with the DRE and GCC box or stabilizing the AP2/ERF domain [21, 42]. In addition, motifs 3, 4 and 6 constituted another AP2 domain and existed specifically in the AP2 subfamily. This further indicated that the amino acid structure of the AP2 domain shows low similarity between the AP2 and ERF subfamilies, which leads to their different functions in the plant development process [16].

The AP2/ERF TF family has been widely studied in plants, and there is considerable evidence that it plays an important role in regulating plant growth and development and resistance to abiotic stress [12]. Furthermore, the AP2/ERF family was also thought to be essential for the response to plant development and morphogenesis, especially in hormone-mediated morphological development of leaf and flower organs [43–45]. Particularly, the AP2 subfamily genes were reported to be associated with plant organ-specific regulation of higher plants [46]. Previous studies have demonstrated that the *ANT* and *AINTEGUMENTA-LIKE* (*AIL*) TFs of the AP2 subfamily are expressed in all dividing tissues, and they have central roles in developmental processes such as embryogenesis[47], root development[48], organ initiation and growth[49]. The *ANT* and *AIL6* genes are essential for SAM maintenance[50].

Generally, the function of genes can be preliminarily predicted by analysis of their expression patterns. Discovery of genes involved in tissue-specific expression helps us understand organogenesis and tissue formation and then uncover the process of morphogenesis. According to previous reports, the AP2 gene tends to exhibit organ specificity and undoubtedly controls organ development [46, 51]. It was clear from the expression profiles of various tissues that most LcERF genes were not tissue-specific. This is mainly because of the extensive roles of the genes in this family in stress adaptability in plants. The significant enrichment of *LcERF57/58/62/94/96/98* in cluster V with expression trend analysis indicated shoot specificity. In addition, according to the alignment and annotation results, *LcERF94*, *LcERF96* and *LcERF98*, which originated from the AP2 subfamily, were mapped to the *AINTEGUMENTA* (*ANT*) or *AINTEGUMENTA-Like* (*AIL*) gene, which stimulated our interest in further research. Interestingly, the same situation was obtained with RT-qPCR; *LcERF94/96/98* were nearly specific to shoots, particularly in the SAM and leaf primordia. Hence, we assumed that these three genes were probably involved in the early stage of leaf formation and development. This conjecture was similar to those made in studies on the *AINTEGUMENTA* and *ANT-LIKE* genes in other plants [47, 52]. Additionally, the gene expression profiles of the *LcERF58* and *LcERF63* genes were also shoot specific; however, this was not always the case according to the qPCR assay, and *LcERF58* and *LcERF63* were primarily expressed in shoots and petals.

One of the reasons for this finding was that group VIII genes have dual or multiple roles in controlling the morphogenesis of vegetative organs as well as floral organs [25]. In addition, different sampling periods were also a main consideration. Shoot organs enter the stage of flower-bud differentiation in a short time, followed by the end of flowering in the current season for many flowering plants. As a result, we cannot deny that the two genes may be involved in leaf development, and attention should also be paid to this in follow-up studies on shoot and leaf development.

One of the most important pieces of information required to define the function of a protein is its subcellular location, and GFP tagging is widely used to detect the functional sites of proteins [53, 54]. TFs usually interact with specific promoter cis-elements to initiate the transcription process and activate or repress the expression of a particular gene. Moreover, pairs of nuclear localization signals exist in several TFs, and basic amino acid residues play essential roles in their function [40]. Kzizek *et al.* [55] found that ANT proteins contain specific amino acids for nuclear localization and full transcriptional activation and proved that ANT is localized in the nucleus and can act as a transcriptional activator in plants. Subsequently, however, Aaron M. Rashotte *et al.* [43] confirmed that ERF proteins not only display constitutive nuclear localization but also present nuclear localization after hormone induction. In addition, other studies of the AP2/ERF TF family showed that most of the AP2/ERF proteins are detected in the nucleus [56, 57]. In this study, *LcERF94/96/98* were all localized in the nucleus only, confirming the prediction results, and this is consistent with the locations of most TFs. This result further indicated that these TFs transcribe and perform their functions in the nucleus.

Conclusions

In summary, a genome-wide survey of the *L. chinense* AP2/ERF family and the expression profiling of the AP2/ERF genes provide comprehensive information for the identification of potential target genes and allow conjecture regarding the roles of these genes. Furthermore, the expression and subcellular location analysis of *LcERF94/96/98* indicated that these genes might have an important function in shoot and leaf development. In any case, this study provides a new perspective for research on the function of LcERFs in regulating plant growth and development and facilitates further research into how their regulatory networks could control shoot and leaf development.

Methods

Plant materials

Plant materials were collected from a forest farm attached to Nanjing Forestry University, Jiangsu Province, China (119°13'20"E, 32°7'8"N). The provenance of sample trees in Lushan, Jiangxi Province (116°0'E, 29°32'N) (Specimen No. 20010020016, deposited in the specimen room of Nanjing Forestry University). In the middle of spring 2018, various tissues, including shoot, leaf, sepal, petal, stamen, pistil and stem tissues, were removed from a mature *L. chinense* tree. In addition, young leaves in distinct developmental periods (dissected from shoots) were also collected from the same tree in summer 2019.

Accordingly, the developmental stages of young leaves were defined based on their developmental morphology as well as the position of the stalks on the shoots according to previous reports [30]. Materials of various developmental stages, *i.e.,* P0 (SAM), P_1, P_2, P_3, P_4, P_5 and P_6, are shown in Figure 6A (P1 ~ P2 were the leaf primordia, while P3 ~ P6 were identified as young leaves). All plant samples were removed and immediately frozen in liquid nitrogen and stored at -80°C.

Identification and characteristics of the AP2/ERF genes of *L. chinense*

The *Liriodendron* genome, which was published in 2019, was downloaded from the NCBI genome resource database [29]. Arabidopsis AP2/ERF proteins were obtained from the Plant Transcription Factor database (http://plntfdb.bio.uni-potsdam.de/v3.0/) and utilized for BLASTP searches against genome sequences of *L. chinense* with expected values less than 10^{-3}. Moreover, the AP2/ERF domain (Pfam accession is PF00847) was downloaded from the Pfam entrance database (http://pfam.xfam.org/) and then used to retrieve AP2/ERF-domain amino acid sequences from all annotated genes of the *Liriodendron* genome by the HMMER program (v3.0). In the process of retrieval, blast protein entries and the hits of the HMMER program were all regarded as candidate AP2/ERF-domain sequences. However, because of the incomplete annotation of gene structures, de novo assembled RNA-seq data were used to correct the mismatched transcripts, and sequences of orthologous AP2/ERF genes of Arabidopsis (Tair 11) and poplar (JGI) were adopted as references simultaneously.

All candidate AP2/ERF-domain amino acid sequences were assessed based on the presence of the conserved AP2/ERF domain with Pfam search (http://pfam.xfam.org/search#tabview=tab1) and CDD search (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) procedures. In addition, sequences that incorrectly occupied or even did not carry a complete domain were removed from the list of putative genes. Conserved motifs are essential characteristics of a gene family that perform various functions. Thus, all the putative AP2/ERF-domain amino acid sequences were divided into distinct subfamilies according to their motif characteristics.

Phylogenetic analysis

To determine the evolutionary relationship of all AP2/ERF sequences, subfamily classification of all AP2/ERF sequences was further confirmed by constructing an unrooted phylogenetic tree with the neighbor-joining (NJ) method. The conserved domain extracted from the whole length of the AP2/ERF sequences was used for multiple sequence alignment through Clustal W with the default parameters in the MEGA X software package [58]. The phylogenetic tree was then constructed using the NJ method with 1000 bootstrap replicates in MEGA X. Finally, the network profile of the phylogenetic tree was visualized by Evoview (https://www.evolgenius.info/evolview/#login).

Conserved motif and gene structure analyses

To investigate the gene structure of the AP2/ERF family, the annotation profile of the *L. chinense* genome was retrieved from the China National GeneBank (CNGB). Information on all exon and intron loci was extracted and later visualized with GSDS (v2.0) (http://gsds.cbi.pku.edu.cn/). In parallel with gene structure surveys, the conserved motifs of the *L. chinense* AP2/ERF family were predicted utilizing MEME (v5.1.1) (http://meme-suite.org/tools/meme) based on the full-length protein sequences with the following parameters: the maximum number of motifs was set to 15, and the motif sites were distributed among sequences with zero or one per sequence model. Then, the results of the motif analysis were visualized by TBtools software. [59]

Expression pattern and trend analysis of LcAP2/ERF genes in various tissues using Illumina RNA-seq profiles

The Illumina RNA-Seq data of various tissues (including bract, sepal, petal, stamen, pistil leaf, shoot and a mixture of all the floral organs named BSPSP) were adopted to demonstrate the expression profiles of LcAP2/ERF genes (RNA-Seq data are under review in other manuscripts). Expression values of all annotated LcAP2/ERF genes in distinct tissues were reflected as fragments per kilobase per million mapped reads (FPKM) and later standardized by Z-score normalization. Then, we generated heat maps of all genes using TBtools according to the instructions [59].

To discover the genes involved in shoot development, all the LcAP2/ERF family genes were clustered in the STEM program using the K-means clustering method [60]. Moreover, according to the verification results, we treated genes that were significantly enriched in the shoot-specific cluster as candidate genes and then performed functional annotation. Additionally, the results of the enrichment analysis are displayed in the bubble plots.

RNA extraction and RT-qPCR analysis of AP2/ERF genes

Total RNA was extracted from samples with the RNAprep pure kit (Tiangen, Beijing, China) according to the manufacturer's instructions. Then, cDNA was synthesized from 500 ng of total RNA using PrimeScirptTM RT Master Mix (TaKaRa, Dalian, China) in a 10 μL reaction volume according to the instructions. Before polymerase chain reaction, the cDNA was diluted 1/10 with mother liquor made with deionized water to reduce systematic error. Then, the specific primers of LcERF genes were designed with Oligo 7.0 software following the instructions strictly (see Additional file 2: Table S2). RT-qPCR was performed on a StepOnePlus^{TM} System (Applied Biosystems) with 10 μL reaction mixtures containing 5 μL of 2× SYBR Premix Ex Taq, 0.2 μL of 50× ROX Reference Dye (TaKaRa, Dalian, China), primers and cDNA. The relative expression levels of the LcERF genes were calculated by using the $\Delta\Delta C_T$ method. The Actin97 primers were used as reference genes in this process.

Subcellular localization assay

We further constructed recombinant proteins that fused the eGFP marker with the LcERF proteins in the C-terminus. The pBI121-eGFP vector (GUS was replaced by GFP in the original PBI121 vector) was restricted by the XbarI and BamHI enzymes along with insertion of the recombinant proteins into the digested vector. Subsequently, the recombinant plasmid sequences were verified by the Sanger sequencing platform and then transferred into *Agrobacterium tumefaciens* (GV3101). After incubation overnight at 28 ℃, the OD600 value of the bacterial solution reached 0.6 ~ 0.8. We collected recombinant bacteria by centrifugation at 4000 rpm and resuspended the bacteria in infection buffer (10 mM $MgCl_2$, 10 mM MES, 150 μM HO-AS, with a final pH=5.6). Moreover, the helper vector P19 was subjected to the same treatment and mixed with ERF equally. The resuspended mixtures were injected into tender tobacco leaves. After 1 ~ 2 days of coculture, we observed and recorded the GFP fluorescence signal under a laser confocal microscope.

Abbreviations

TFs: Transcription factors

AP2/EREBPs: APETALA2/ethylene responsive element-binding proteins

ERF: Ethylene responsive rlement binding factor

DREB: Dehydration responsive element binding factor

RAV: Related to ABI3/VP

ANT: AINTEGUMENTA

AIL: AINTEGUMENTA-Like

SAM: Shoot Apical Meristem

DRN: DORNROESCHEN

DRNL: DRN-LIKE

FPKM: Fragments Per Kilobase Million

GFP: Green fluorescent protein

Declarations

Ethics approval and consent to participate

Experimental materials provided by Nanjing Forestry University, Nanjing, Jiangsu, China. The authors confirm that the material collection presented here were conducted in accordance with the wild plant care regulations and natural reserves regulations set forth by the Decree of the state council of the People's Republic of China.

Consent for publication

Not applicable.

Availability of data and materials

All data presented in this study are provided either in the manuscript or additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

The design and experiments of this study were supported by the National Natural Science Foundation of China (31770718 and 31470660) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Authors' Contributions

H. L. designed the research. Y. Z. and Z. H. analyzed the data and prepared figures. Y. Z., Z. H., Z. T., Y. S., C. Z., S. W., and L. C. performed the experiments. Y. Z. and Z. H. did final editing of the manuscript. Y. Z. and Z. H. have contributed equally to this work. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors would like to thank Dr Wenjie Ding for advice on data analysis; Dr Huanhuan Liu for assistance in experiments. In addition, the authors also like to thank other colleagues in the laboratory for their help.

References

1. Ma Y, Wang F, Guo J, Zhang XS: **Rice OsAS2 Gene, a Member of LOB Domain Family, Functions in the Regulation of Shoot Differentiation and Leaf Development**. *Journal of Plant Biology* 2009, **52**(5):374-381.
2. Barkoulas M, Hay A, Kougioumoutzi E, Tsiantis M: **A developmental framework for dissected leaf formation in the Arabidopsis relative Cardamine hirsuta**. *Nature Genetics* 2008, **40**(9):1136-1141.
3. Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ: **The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs**. *Development* 2005, **133**(2):251-261.
4. Gutierrez C: **The Arabidopsis Cell Division Cycle**. *The Arabidopsis Book* 2009, **7**(7).
5. Takada S, Hibara K, Ishida T, Tasaka M: **The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation**. *Development* 2001, **128**(7):1127-1135.
6. Barton MK, Poethig RS: **Formation of the shoot apical meristem in Arabidopsis thaliana: an analysis of development in the wild-type and in the SHOOT MERISTEMLESS mutant**. *Development* 1993, **119**(3):823-831.
7. Wurschum T, Groshardt R, Laux T: **APETALA2 Regulates the Stem Cell Niche in the Arabidopsis Shoot Meristem**. *The Plant Cell* 2006, **18**(2):295-307.
8. Ohto M, Floyd SK, Fischer RL, Goldberg RB, Harada JJ: **Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in Arabidopsis**. *Sexual Plant Reproduction* 2009, **22**(4):277-289.
9. McGrath KC, Dombrecht B, Manners JM, Schenk PM, Edgar CI, Maclean DJ, Scheible WR, Udvardi MK, Kazan K: **Repressor- and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of Arabidopsis transcription factor gene expression**. *Plant Physiology* 2005, **139**(2):949-959.
10. Yamaguchishinozaki K, Shinozaki K: **Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses**. *Plant Biology* 2006, **57**(1):781-803.
11. Marschmartinez N, Greco R, Becker JD, Dixit S, Bergervoet JHW, Karaba A, De Folter S, Pereira A: **BOLITA, an Arabidopsis AP2/ERF-like transcription factor that affects cell expansion and proliferation/differentiation pathways**. *Plant Molecular Biology* 2006, **62**(6):825-843.
12. Liu M, Sun W, Ma Z, Zheng T, Huang L, Wu Q, Zhao G, Tang Z, Bu T, Li C: **Genome-wide investigation of the AP2/ERF gene family in tartary buckwheat (Fagopyum Tataricum)**. *BMC Plant Biology* 2019, **19**(1):1-19.
13. Agarwal G, Garg V, Kudapa H, Doddamani D, Pazhamala LT, Khan AW, Thudi M, Lee S, Varshney RK: **Genome-wide dissection of AP2/ERF and HSP90 gene families in five legumes and expression profiles in chickpea and pigeonpea**. *Plant Biotechnology Journal* 2016, **14**(7):1563-1577.
14. Du D, Hao R, Cheng T, Pan H, Yang W, Wang J, Zhang Q: **Genome-Wide Analysis of the AP2/ERF Gene Family in Prunus mume**. *Plant Molecular Biology Reporter* 2013, **31**(3):741-750.
15. Jofuku KD, Den Boer BGW, Van Montagu M, Okamuro JK: **Control of Arabidopsis flower and seed development by the homeotic gene APETALA2**. *The Plant Cell* 1994, **6**(9):1211-1225.

16. Licausi F, Ohmetakagi M, Perata P: **APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs**. *New Phytologist* 2013, **199**(3):639-649.
17. Riechmann JL, Meyerowitz EM: **The AP2/EREBP Family of Plant Transcription Factors**. *Biological Chemistry* 1998, **379**(6):633-646.
18. Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchishinozaki K: **DNA-Binding Specificity of the ERF/AP2 Domain of Arabidopsis DREBs, Transcription Factors Involved in Dehydration- and Cold-Inducible Gene Expression**. *Biochemical and Biophysical Research Communications* 2002, **290**(3):998-1009.
19. Nakano T, Suzuki K, Fujimura T, Shinshi H: **Genome-Wide Analysis of the ERF Gene Family in Arabidopsis and Rice**. *Plant Physiology* 2006, **140**(2):411-432.
20. Sakuma Y, Maruyama K, Osakabe Y, Qin F, Seki M, Shinozaki K, Yamaguchishinozaki K: **Functional Analysis of an Arabidopsis Transcription Factor, DREB2A, Involved in Drought-Responsive Gene Expression**. *The Plant Cell* 2006, **18**(5):1292-1309.
21. Sharoni AM, Nuruzzaman M, Satoh K, Shimizu T, Kondoh H, Sasaya T, Choi I, Omura T, Kikuchi S: **Gene Structures, Classification, and Expression Models of the AP2/EREBP Transcription Factor Family in Rice**. *Plant and Cell Physiology* 2011, **52**(2):344-360.
22. Liu Q: **Two Transcription Factors, DREB1 and DREB2, with an EREBP/AP2 DNA Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought- and Low-Temperature-Responsive Gene Expression, Respectively, in Arabidopsis.** *The Plant Cell* 1998, **10**(8):1391–1406.
23. Hao D, Ohmetakagi M, Sarai A: **Unique Mode of GCC Box Recognition by the DNA-binding Domain of Ethylene-responsive Element-binding Factor (ERF Domain) in Plant**. *Journal of Biological Chemistry* 1998, **273**(41):26857-26861.
24. Welsch R, Maass D, Voegel T, Dellapenna D, Beyer P: **Transcription Factor RAP2.2 and Its Interacting Partner SINAT2: Stable Elements in the Carotenogenesis of Arabidopsis Leaves**. *Plant Physiology* 2007, **145**(3):1073-1085.
25. Chandler JW: **Class VIIIb APETALA2 Ethylene Response Factors in Plant Development**. *Trends in Plant Science* 2017, **23**(2):151-162.
26. Der Graaff EV, Den Dulkras A, Hooykaas PJJ, Keller B: **Activation tagging of the LEAFY PETIOLE gene affects leaf petiole development in Arabidopsis thaliana**. *Development* 2000, **127**(22):4971-4980.
27. Banno H, Ikeda Y, Niu Q, Chua N: **Overexpression of Arabidopsis ESR1 Induces Initiation of Shoot Regeneration**. *The Plant Cell* 2001, **13**(12):2609-2618.
28. Aida M, Beis D, Heidstra R, Willemsen V, Blilou I, Galinha C, Nussaume L, Noh Y, Amasino RM, Scheres B: **The PLETHORA Genes Mediate Patterning of the Arabidopsis Root Stem Cell Niche**. *Cell* 2004, **119**(1):109-120.
29. Chen J, Hao Z, Guang X, Zhao C, Wang P, Xue L, Zhu Q, Yang L, Sheng Y, Zhou Y: **Liriodendron genome sheds light on angiosperm phylogeny and species–pair differentiation**. *Nature plants* 2019, **5**(1):18-25.

30. Ma J, Wei L, Li J, Li H: **The Analysis of Genes and Phytohormone Metabolic Pathways Associated with Leaf Shape Development in Liriodendron chinense via De Novo Transcriptome Sequencing**. *Genes* 2018, **9**(12):577.
31. Tu z, Shen Y, Wen S, Liu H, Wei L, Li H: **Gene co-expression analysis and tissue-specific gene identification in Liriodendron chinense via hybrid sequencing**. *PREPRINT (Version 1) available at Research Square* 2020.
32. Matsuo N, Makino M, Banno H: **Arabidopsis ENHANCER OF SHOOT REGENERATION (ESR)1 and ESR2 regulate in vitro shoot regeneration and their expressions are differentially regulated**. *Plant Science* 2011, **181**(1):39-46.
33. Zhu Y, Li Y, Zhang S, Zhang X, Yao J, Luo Q, Sun F, Wang X: **Genome-wide identification and expression analysis reveal the potential function of ethylene responsive factor gene family in response to Botrytis cinerea infection and ovule development in grapes (Vitis vinifera L.)**. *Plant Biology* 2019, **21**(4):571-584.
34. Zhuang J, Cai B, Peng R, Zhu B, Jin X, Xue Y, Gao F, Fu X, Tian Y, Zhao W: **Genome-wide analysis of the AP2/ERF gene family in Populus trichocarpa**. *Biochemical and Biophysical Research Communications* 2008, **371**(3):468-474.
35. Zhang Z, Li X: **Genome-wide identification of AP2/ERF superfamily genes and their expression during fruit ripening of Chinese jujube**. *Scientific Reports* 2018, **8**(1):15612.
36. Li X, Tao S, Wei S, Ming M, Huang X, Zhang S, Wu J: **The mining and evolutionary investigation of AP2/ERF genes in pear (Pyrus)**. *BMC Plant Biology* 2018, **18**(1):46.
37. Shu Y, Liu Y, Zhang J, Song L, Guo C: **Genome-Wide Analysis of the AP2/ERF Superfamily Genes and their Responses to Abiotic Stress in Medicago truncatula**. *Frontiers in Plant Science* 2015, **6**:1247-1247.
38. Zhang S, Zhu C, Lyu Y, Chen Y, Zhang Z, Lai Z, Lin Y: **Genome-wide identification, molecular evolution, and expression analysis provide new insights into the APETALA2/ethylene responsive factor (AP2/ERF) superfamily in Dimocarpus longan Lour**. *BMC Genomics* 2020, **21**(1):62.
39. Srivastava R, Kumar R: **The expanding roles of APETALA2/Ethylene Responsive Factors and their potential applications in crop improvement**. *Brief Funct Genomics* 2018, **18**(4):240-254.
40. Liu L, White MJ, Macrae TH: **Transcription factors and their genes in higher plants functional domains, evolution and regulation**. *FEBS Journal* 1999, **262**(2):247-257.
41. Mathur S, Priyadarshini SS, Singh V, Vashisht I, Jung K-H, Sharma R, Sharma MK: **Comprehensive phylogenomic analysis of ERF genes in sorghum provides clues to the evolution of gene functions and redundancy among gene family members**. *3 Biotech* 2020, **10**(3).
42. Liu Y, Zhao T, Liu J, Liu W, Liu Q, Yan Y, Zhou H: **The conserved Ala37 in the ERF/AP2 domain is essential for binding with the DRE element and the GCC box**. *FEBS Letters* 2006, **580**(5):1303-1308.
43. Rashotte AM, Mason MG, Hutchison CE, Ferreira F, Schaller GE, Kieber JJ: **A subset of Arabidopsis AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway**.

Proceedings of the National Academy of Sciences of the United States of America 2006, **103**(29):11081-11085.

44. Horstman A, Willemsen V, Boutilier K, Heidstra R: **AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks**. *Trends in Plant Science* 2014, **19**(3):146-157.
45. Elliott RC, Betzner AS, Huttner E, Oakes MP, Tucker WQJ, Gerentes D, Perez P, Smyth DR: **AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth**. *The Plant Cell* 1996, **8**(2):155-168.
46. Wuddineh WA, Mazarei M, Turner GB, Sykes RW, Decker SR, Davis MF, Stewart CN, Jr.: **Identification and Molecular Characterization of the Switchgrass AP2/ERF Transcription Factor Superfamily, and Overexpression of PvERF001 for Improvement of Biomass Characteristics for Biofuel**. *Front Bioeng Biotechnol* 2015, **3**:101.
47. Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu C, Van Lammeren AAM, Miki B: **Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth**. *The Plant Cell* 2002, **14**(8):1737-1749.
48. Galinha C, Hofhuis H, Luijten M, Willemsen V, Blilou I, Heidstra R, Scheres B: **PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development**. *Nature* 2007, **449**(7165):1053-1057.
49. Mizukami Y, Fischer RL: **Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis**. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(2):942-947.
50. Mudunkothge JS, Krizek BA: **Three Arabidopsis AIL/PLT genes act in combination to regulate shoot apical meristem function**. *Plant Journal* 2012, **71**(1):108-121.
51. Zhang JY, Lee YC, Torres-Jerez I, Wang M, Yin Y, Chou WC, He J, Shen H, Srivastava AC, Pennacchio C *et al*: **Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (Panicum virgatum L.)**. *Plant J* 2013, **74**(1):160-173.
52. Nolewilson S, Tranby TL, Krizek BA: **AINTEGUMENTA-like (AIL) genes are expressed in young tissues and may specify meristematic or division-competent states**. *Plant Molecular Biology* 2005, **57**(5):613-628.
53. Millar AH, Carrie C, Pogson B, Whelan J: **Exploring the function-location nexus: using multiple lines of evidence in defining the subcellular location of plant proteins**. *Plant Cell* 2009, **21**(6):1625-1631.
54. Narsai R, Devenish J, Castleden I, Narsai K, Xu L, Shou H, Whelan J: **Rice DB: an Oryza Information Portal linking annotation, subcellular location, function, expression, regulation, and evolutionary information for rice and Arabidopsis**. *Plant J* 2013, **76**(6):1057-1073.
55. Krizek BA, Sulli C: **Mapping sequences required for nuclear localization and the transcriptional activation function of the Arabidopsis protein AINTEGUMENTA**. *Planta* 2006, **224**(3):612-621.
56. Zhou L, Li J, He Y, Liu Y, Chen H: **Functional characterization of SmCBF genes involved in abiotic stress response in eggplant (Solanum melongena)**. *Scientia Horticulturae* 2018, **233**:14-21.

57. Djemal R, Mila I, Bouzayen M, Pirrello J, Khoudi H: **Molecular cloning and characterization of novel WIN1/SHN1 ethylene responsive transcription factor HvSHN1 in barley (Hordeum vulgare L.)**. *Journal of Plant Physiology* 2018, **228**:39-46.
58. Kumar S, Stecher G, Li M, Knyaz C, Tamura K: **MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms**. *Molecular Biology and Evolution* 2018, **35**(6):1547-1549.
59. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R: **TBtools - an integrative toolkit developed for interactive analyses of big biological data**. *bioRxiv* 2020.
60. Ernst J, Nau GJ, Barjoseph Z: **Clustering short time series gene expression data**. *Bioinformatics* 2005, **21**(1):159-168.

Figures

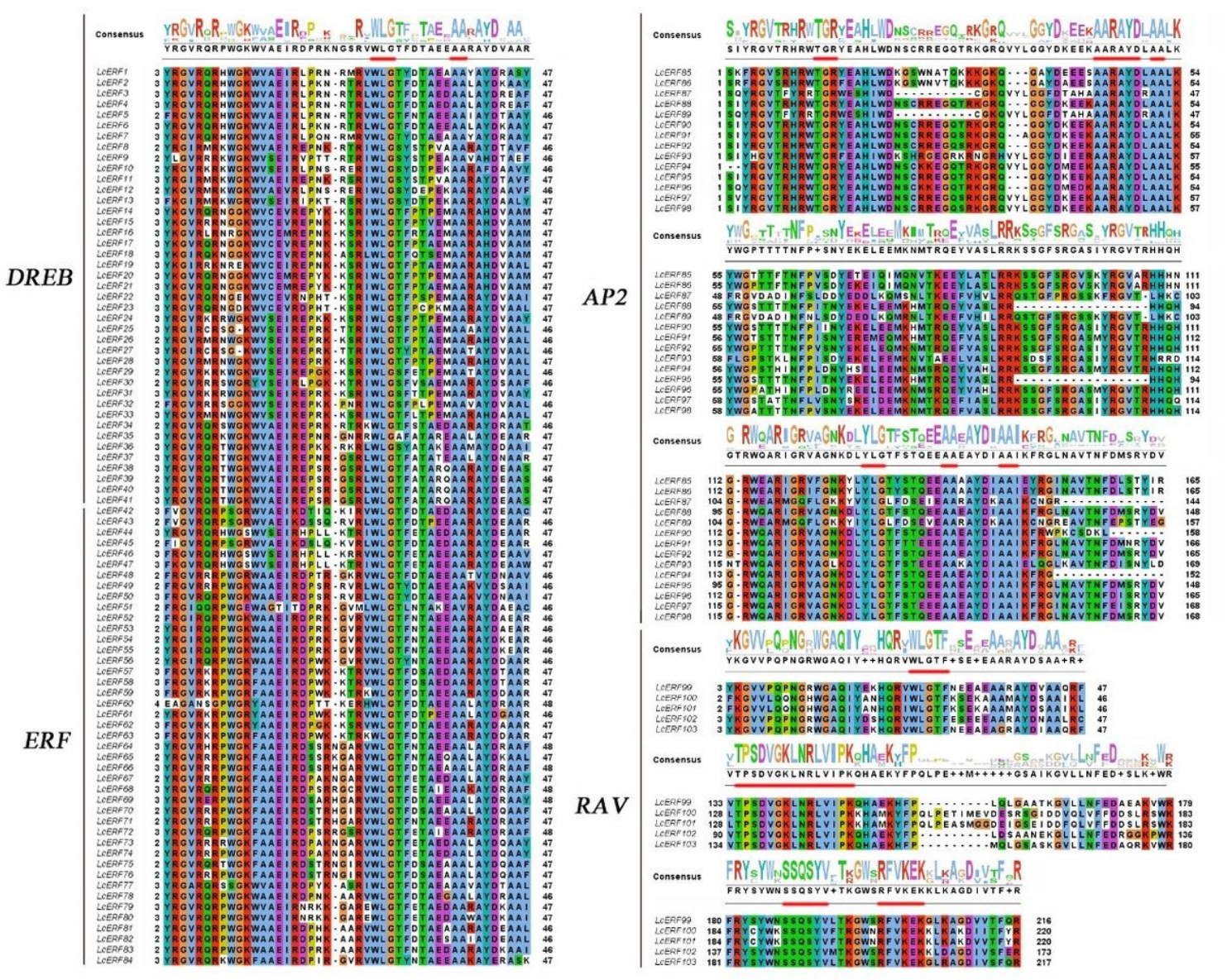

Figure 1

Multiple sequence alignment of the AP2/ERF DNA-binding domains of LcERF proteins. Multiple sequence alignment of different subfamilies using Clustal W and visualized by Jalview.

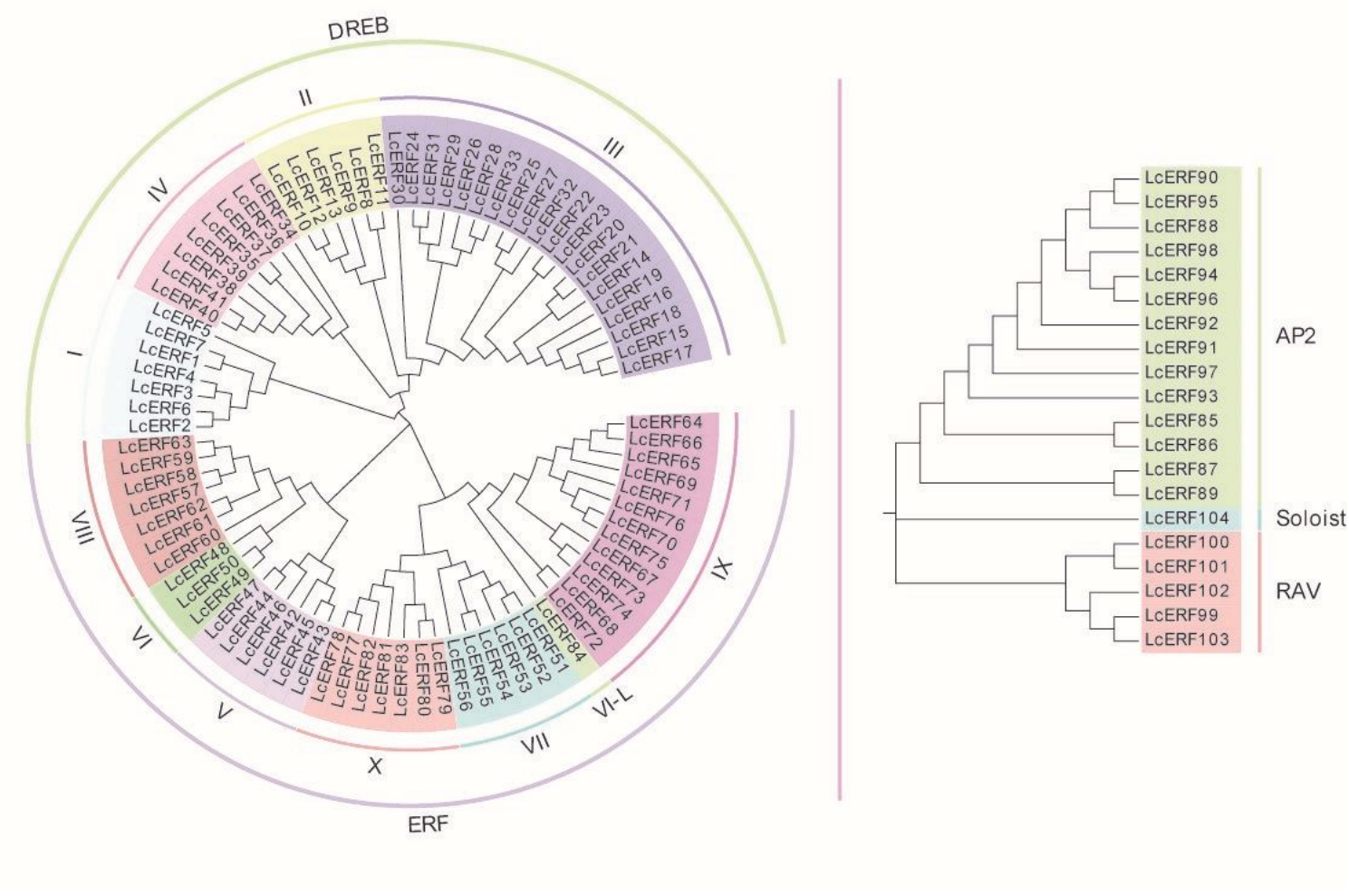

Figure 2

Phylogenetic tree of 104 AP2/ERF proteins from Liriodendron chinense. The unrooted phylogenetic tree was constructed in MEGA X using the neighbor - joining (NJ) method.

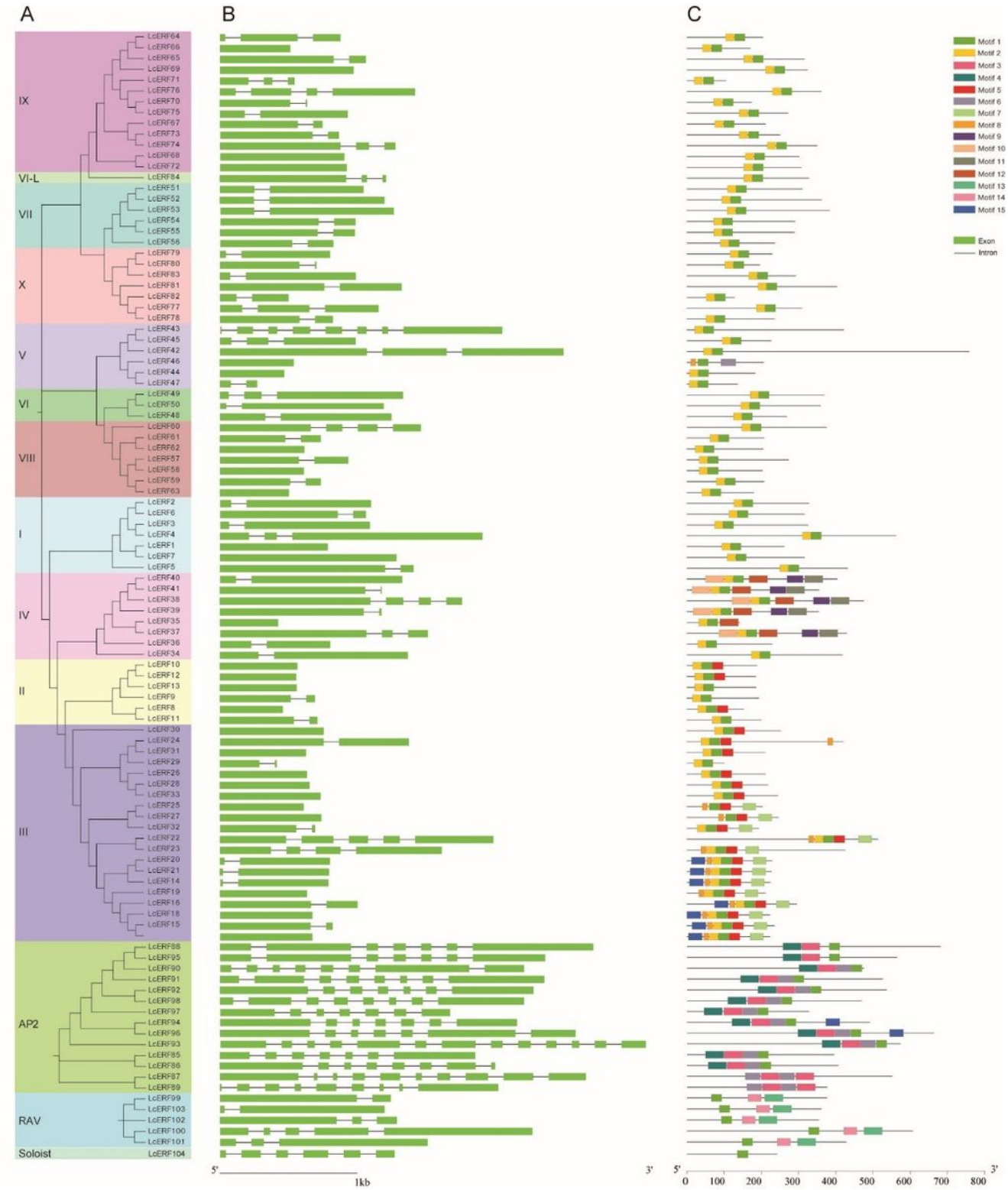

Figure 3

Gene structure and conserved motifs analysis of 104 LcERF genes. (A) Phylogenetic tree of 104 LcERF proteins. (B) Exon-intron structures of LcERF genes. Green box and black lines represent exons and introns, respectively. (C) Motif composition of LcERF genes. Boxes in different color represent different conserved motifs.

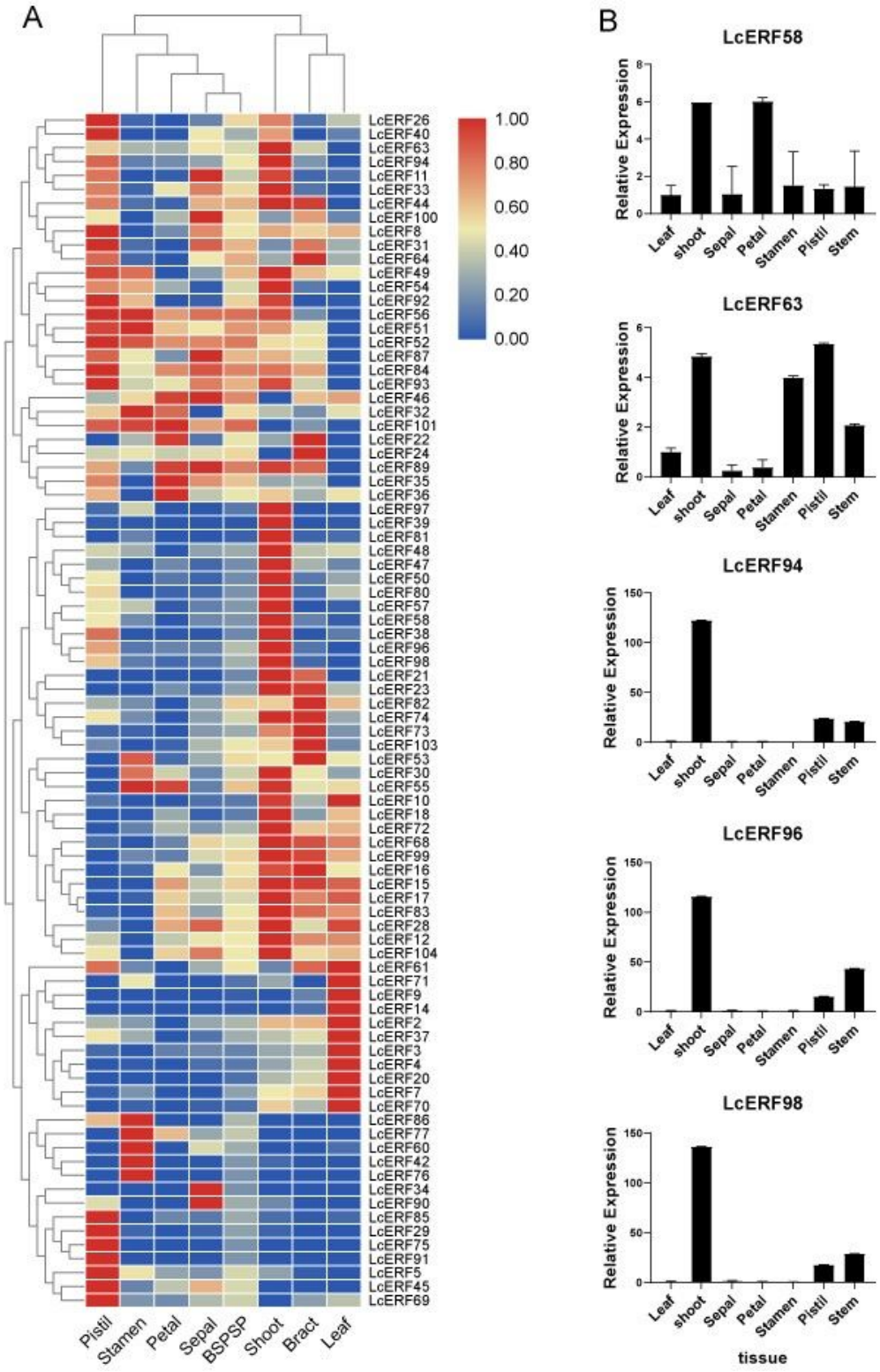

Figure 4

Expression analysis based on RNA-seq data and relative expression levels of LcERF genes in various tissues. (A) Expression analysis of 86 LcERF genes based on RNA-seq data, FPKM value were standardized by the Z-score Normalization and the heatmap constructed by TBtools software. (B) relative expression levels of five potential genes in various tissues.

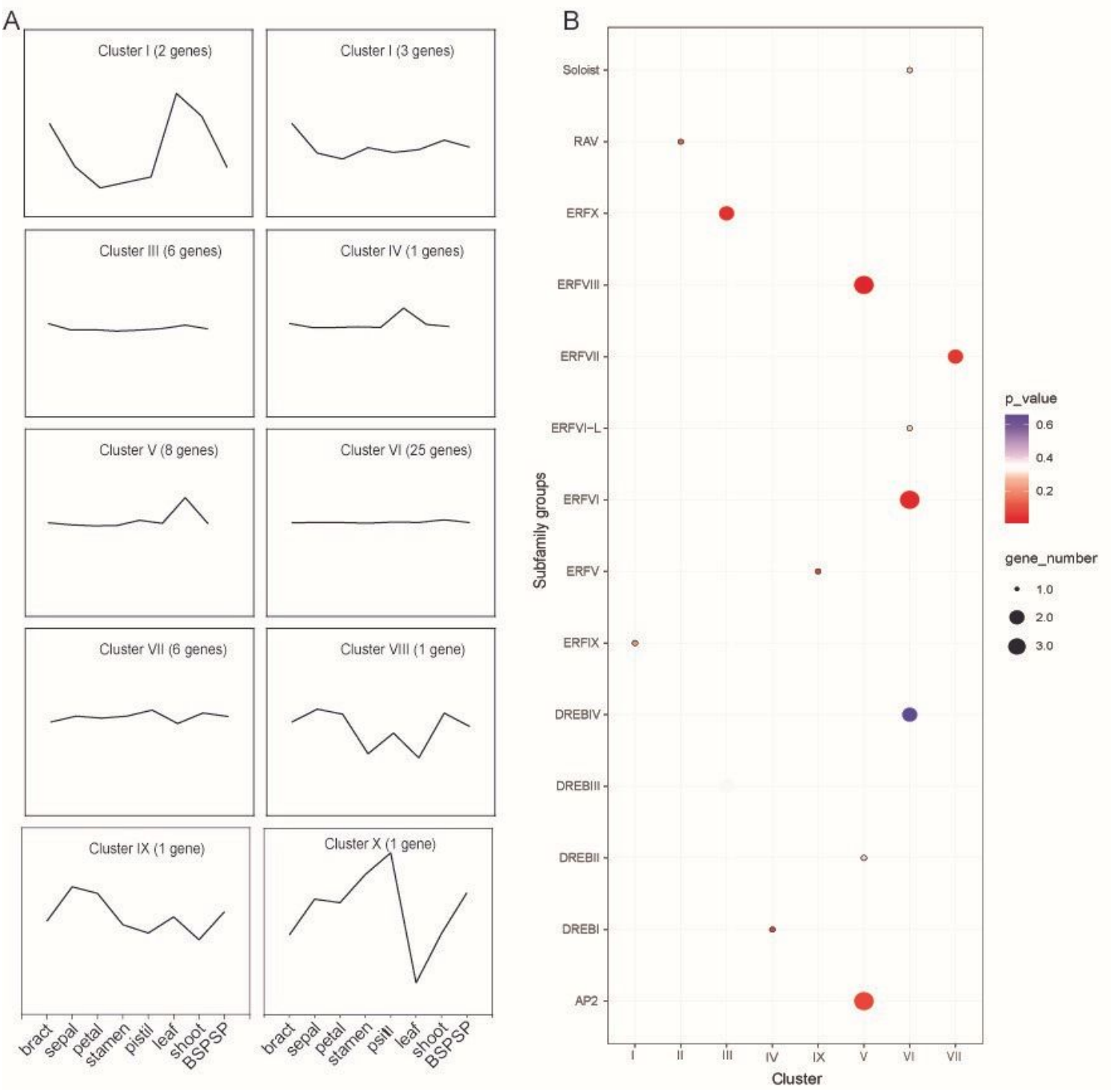

Figure 5

The expression trend analysis of LcERF genes. (A) Ten clusters with the K-means method in STEM program. (B) Category enrichment of DEGs.

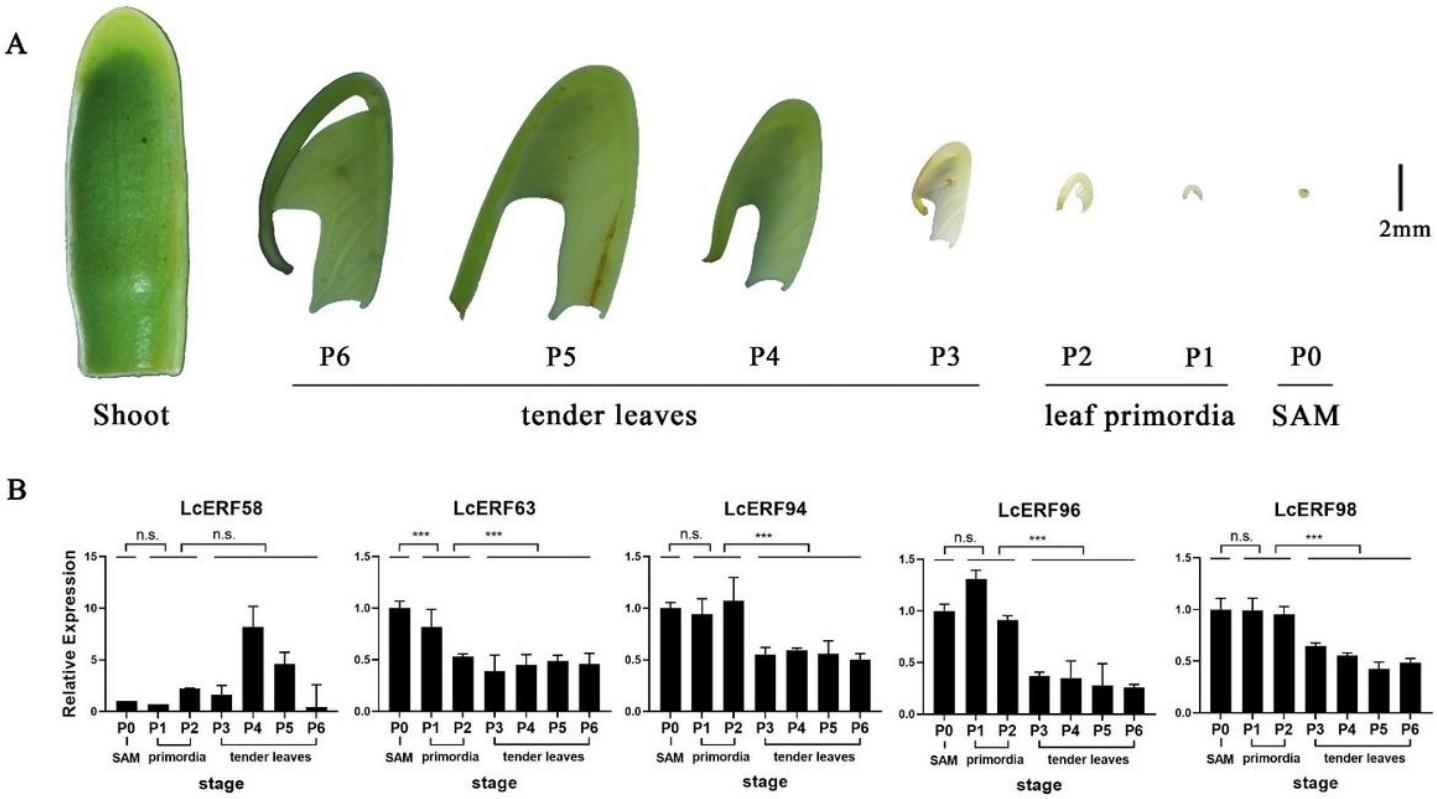

Figure 6

The dynamic development and relative expression of Liriodendron chinense leaves. (A) Different development stage leaves in the stereomicroscope. (B) Relative expression of five potential genes in different leaves development stage. Date and error bars represent the mean values the SD from three replications, respectively. The significant differences were analyzed using T-test: *** indicates P-value < 0.001; n.s. indicates no significant.

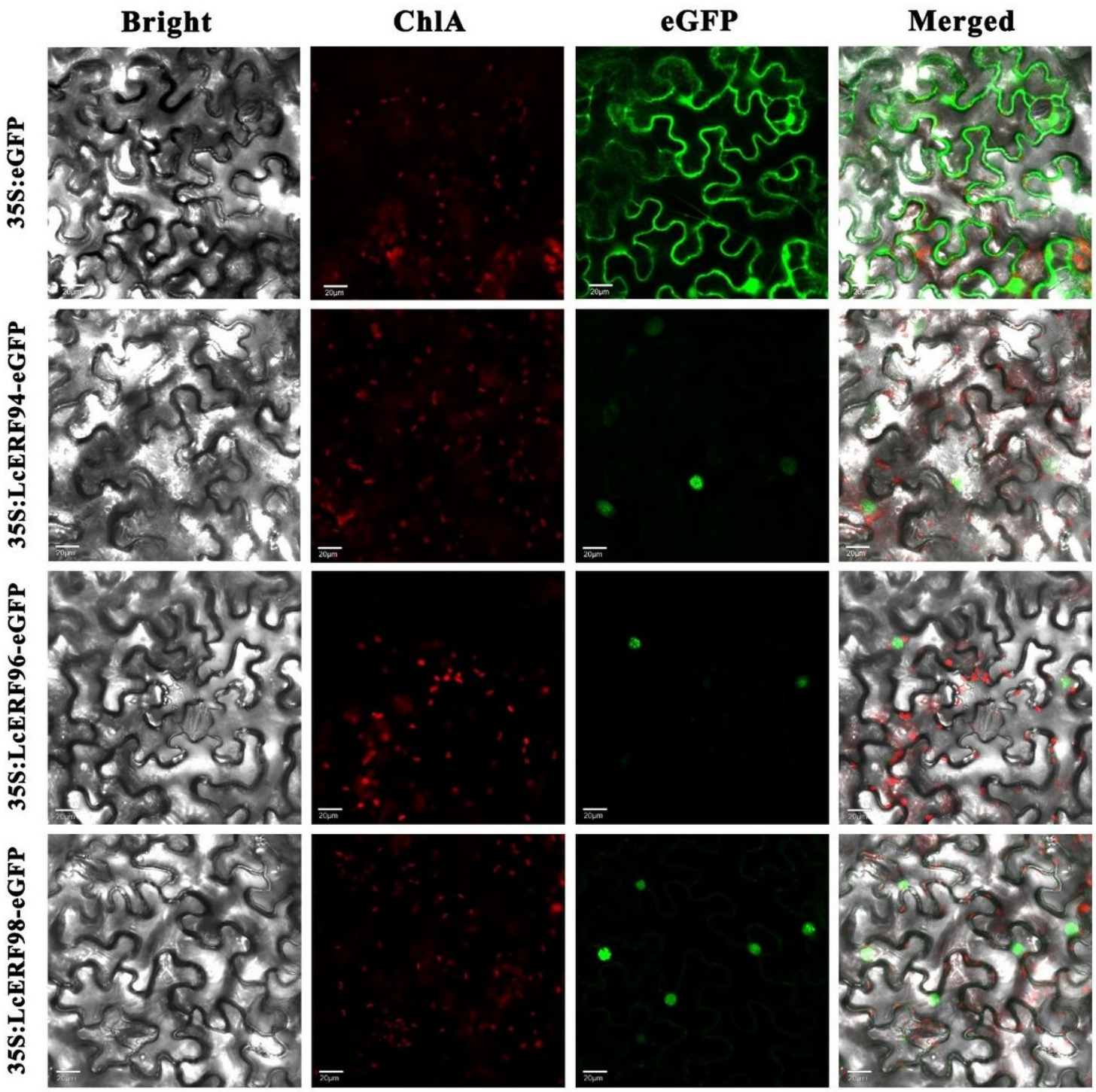

Figure 7

Subcellular localization of LcERF94/96/98 proteins. 35S: eGFP are presented as a control sample. The signal mode was observed at GFP fluorescence (green), chlorophyll fluorescence (red).

Supplementary Files

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

- Additionalfile2.xlsx
- Additionalfile1.xlsx