

## Article

# Comparative Genomics of *Lotus japonicus* Reveals Insights into Proanthocyanidin Accumulation and Abiotic Stress Response

Zhanmin Sun <sup>1,\*</sup> , Ziyang Liu <sup>1</sup>, Manqing Zhi <sup>1</sup>, Qifan Ran <sup>2</sup>, Wenbo Xue <sup>3</sup>, Yixiong Tang <sup>1</sup> and Yanmin Wu <sup>1</sup>

<sup>1</sup> Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China; liuzy17@lzu.edu.cn (Z.L.); zmq990305@163.com (M.Z.); tangyixiong@caas.cn (Y.T.); wuyanmin@caas.cn (Y.W.)

<sup>2</sup> Chongqing Academy of Animal Sciences, Chongqing 402460, China; ranqifan@outlook.com

<sup>3</sup> BGI Genomics, Shenzhen 518085, China; xuewenbo@genomics.cn

\* Correspondence: sunzhanmin@caas.cn

**Abstract:** *Lotus japonicus*, is an important perennial model legume, has been widely used for studying biological processes such as symbiotic nitrogen fixation, proanthocyanidin (PA) biosynthesis, and abiotic stress response. High-quality *L. japonicus* genomes have been reported recently; however, the genetic basis of genes associated with specific characters including proanthocyanidin distribution in most tissues and tolerance to stress has not been systematically explored yet. Here, based on our previous high-quality *L. japonicus* genome assembly and annotation, we compared the *L. japonicus* MG-20 genome with those of other legume species. We revealed the expansive and specific gene families enriched in secondary metabolite biosynthesis and the detection of external stimuli. We suggested that increased copy numbers and transcription of PA-related genes contribute to PA accumulation in the stem, petiole, flower, pod, and seed coat of *L. japonicus*. Meanwhile, According to shared and unique transcription factors responding to five abiotic stresses, we revealed that MYB and AP2/ERF play more crucial roles in abiotic stresses. Our study provides new insights into the key agricultural traits of *L. japonicus* including PA biosynthesis and response to abiotic stress. This may provide valuable gene resources for legume forage abiotic stress resistance and nutrient improvement.

**Keywords:** *Lotus japonicus*; proanthocyanidins; RNA-seq; abiotic stress; comparative genomics



**Citation:** Sun, Z.; Liu, Z.; Zhi, M.; Ran, Q.; Xue, W.; Tang, Y.; Wu, Y.

Comparative Genomics of *Lotus japonicus* Reveals Insights into Proanthocyanidin Accumulation and Abiotic Stress Response. *Plants* **2024**, *13*, 1151. <https://doi.org/10.3390/plants13081151>

Academic Editor: Changsoo Kim

Received: 26 March 2024

Revised: 16 April 2024

Accepted: 18 April 2024

Published: 20 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Legumes, as the second most crucial plant in the world, contribute a large amount of protein and nutrients as well as provide a vital ecosystem source of reduced nitrogen through symbiotic nitrogen fixation. *Lotus japonicus* is a perennial model legume forage closely related to the cultivated species *L. corniculatus*. Nowadays, because of the availability of high-quality genome assemblies [1–3], easy genetic transformation, and availability of substantial mutant resources, *L. japonicus* has been extensively used for studying molecular mechanisms related to plant–microbial symbiosis, secondary metabolism biosynthesis, particularly proanthocyanidins (PAs), and responses to biotic and abiotic stresses.

Firstly, *L. japonicus* is proposed as an ideal model system for studying regulatory mechanisms of proanthocyanidin (PA) biosynthesis. PAs or condensed tannins (CTs) are polymeric flavonoids contributing quality traits to crucial forage legumes due to preventing pasture bloat and decreasing ruminant methane emissions. Unfortunately, the most crucial pasture legumes such as alfalfa (*Medicago sativa*) and white clover (*Trifolium repens*) cannot accumulate PAs in the leaf and stem tissues, whereas, *L. corniculatus* contains moderate PA levels in the leaves, stems, and other tissues. As a model legume, *L. japonicus*, which possesses PA in most organs excluding leaves, can provide an important source of information for increasing PA levels in alfalfa or white clover leaves and stems through genetic engineering and breeding. Although the genome of *L. japonicus* has been published, comparative genomics and evolutionary analyses of the genes associated with PA biosynthesis are lacking.

Meanwhile, abiotic stresses (including drought, salinity, heat, cold, and heavy metal stresses  $AlCl_3$ ) are considered major limiting factors that qualitatively and quantitatively affect crop production [4]. In fact, numerous *Lotus* species and cultivars exhibit wide and different degrees of tolerance. Almost all *Lotus* species and cultivars are more tolerant of abiotic stress than *Medicago* (lucerne) or *Trifolium* (trefoils) [5].

In this study, based on our previous de novo assembled and annotated *L. japonicus* MG-20 genome [3], we compared the *L. japonicus* genome with those of other legume species, and revealed the expansive and specific gene families enriched in the biosynthesis of secondary metabolites and detection of external stimuli. Expansion of the first key enzyme phenylalanine ammonia lyase (*PAL*) and genes in the late biosynthesis pathway that encode dihydroflavonol 4-reductase (*DFR*) and transparent testa 2 (*TT2*), and transcription expression of genes such as flavonoid 3',5'-hydroxylase (*F3'5'H*), dihydroflavonol 4-reductase (*DFR*), leucoanthocyanidin dioxygenase (*ANS*), and anthocyanidin reductase (*ANR*) in most tissues mainly contribute to PA accumulation in the stems, petioles, flowers, pods, and seed coat of *L. japonicus*. After studying the shared and unique transcription factors (TFs) that respond to five abiotic stresses, we suggested that MYB and AP2/ERF play more prominent roles in abiotic stress responses. This may provide valuable gene resources for legume forage abiotic stress resistance and nutrient improvement.

## 2. Results

### 2.1. Comparison of Genome Changes in Sister Legume Species

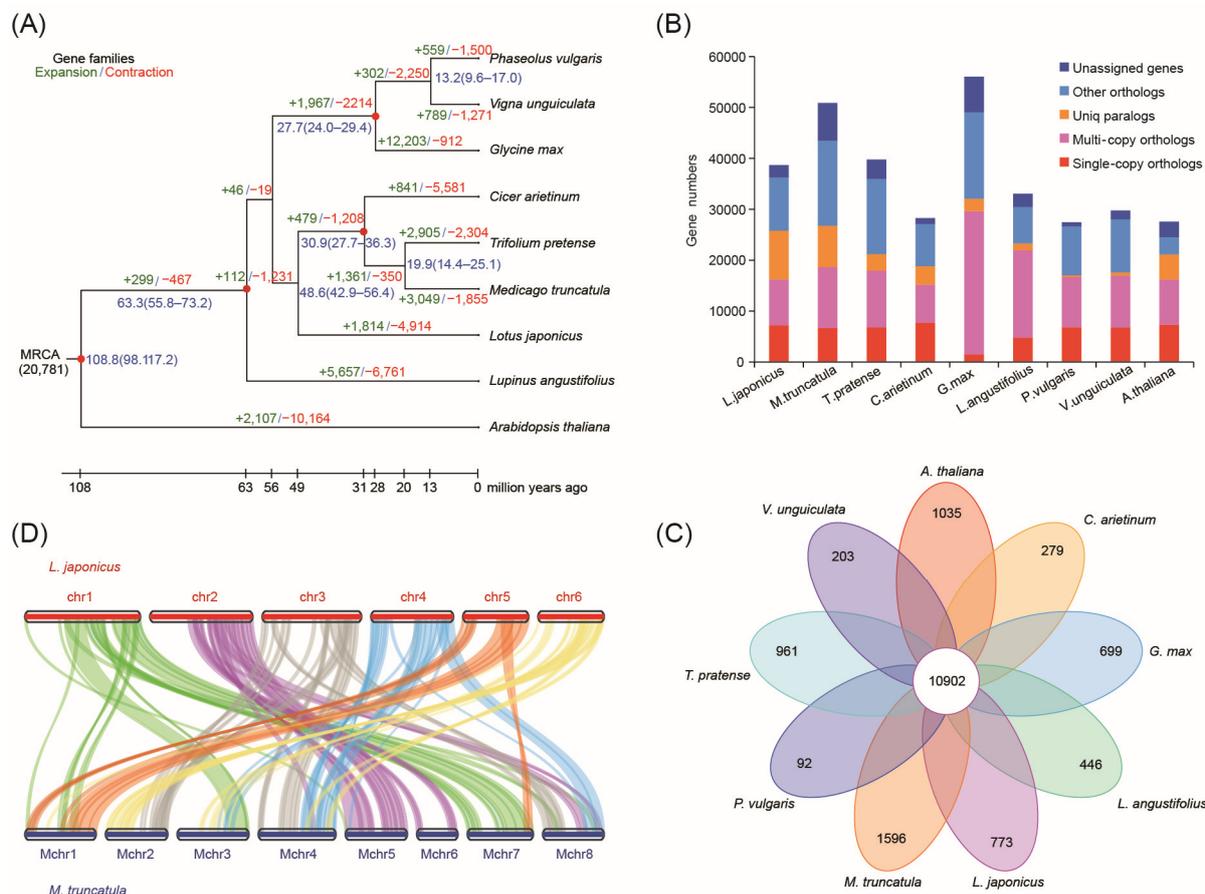
In our previous study, we sequenced, assembled, and annotated the *L. japonicus* MG-20 genome [3]. Based on the aforementioned works, we performed an evolutionary analysis according to the alignment of 592 single-copy gene families shared by a total of eight legume species (*L. japonicus*, *M. truncatula*, *G. max*, *T. pretense*, *P. vulgaris*, *C. arietinum*, *V. unguiculata*, and *Lupinus angustifolius*) and a nonlegume species (*A. thaliana*). The constructed phylogenetic tree not only showed close relationships among *L. japonicus*, *M. truncatula*, *T. pretense*, and *C. arietinum* but also reflected the divergence between *L. japonicus* and the common ancestor of *M. truncatula*, *T. pretense*, and *C. arietinum*, which occurred approximately 48.6 MYA (Figure 1A).

The 38,684 gene sets of *L. japonicus* were compared with 292,810 gene models from the seven sequenced legumes and one nonlegume species (*A. thaliana*), and 36,258 orthogroups and 16,416 families (orthologous groups) were identified using the software OrthoFinder ver.2.3.8 (Table S1). Among them, 7215 orthologs contained only a single *L. japonicus* gene, suggestive of simple orthology (Table S2 and Figure 1B). On the other hand, a total of 10,902 families (orthologous groups) were shared in all nine species, and 773 gene families were composed of only *L. japonicus* proteins (Table S1 and Figure 1C). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of *L. japonicus*-specific genes revealed that these gene families were significantly enriched in genes related to secondary metabolite biosynthesis, detection of external stimuli, circadian rhythm, selenocompound metabolism, and regulation of macromolecule biosynthesis (Figure 2A,B).

Furthermore, we identified the expanded or contracted gene families. Our results revealed that 112 families comprising 1126 genes exhibited significant expansion in the *L. japonicus* genome (Table S3). These genes were functionally annotated. Functional annotation demonstrated that the genes were mainly enriched in functional categories, namely secondary metabolite biosynthesis, plant–pathogen interaction, polysaccharide metabolism, and response to abiotic stimuli (Figure 2C,D and Table S3). Furthermore, 18 families comprising 176 genes exhibited significant contraction in the *L. japonicus* genome (Table S4), and these genes were significantly enriched in functions related to phosphorus metabolism and S-adenosylmethionine biosynthesis/metabolism (Figure 2E).

Synteny block analysis is generally used to study chromosome evolution among related species. Here, we analyzed the aligned protein sequences of *L. japonicus* in comparison with those of *M. truncatula*. The synteny results indicated the presence of shared

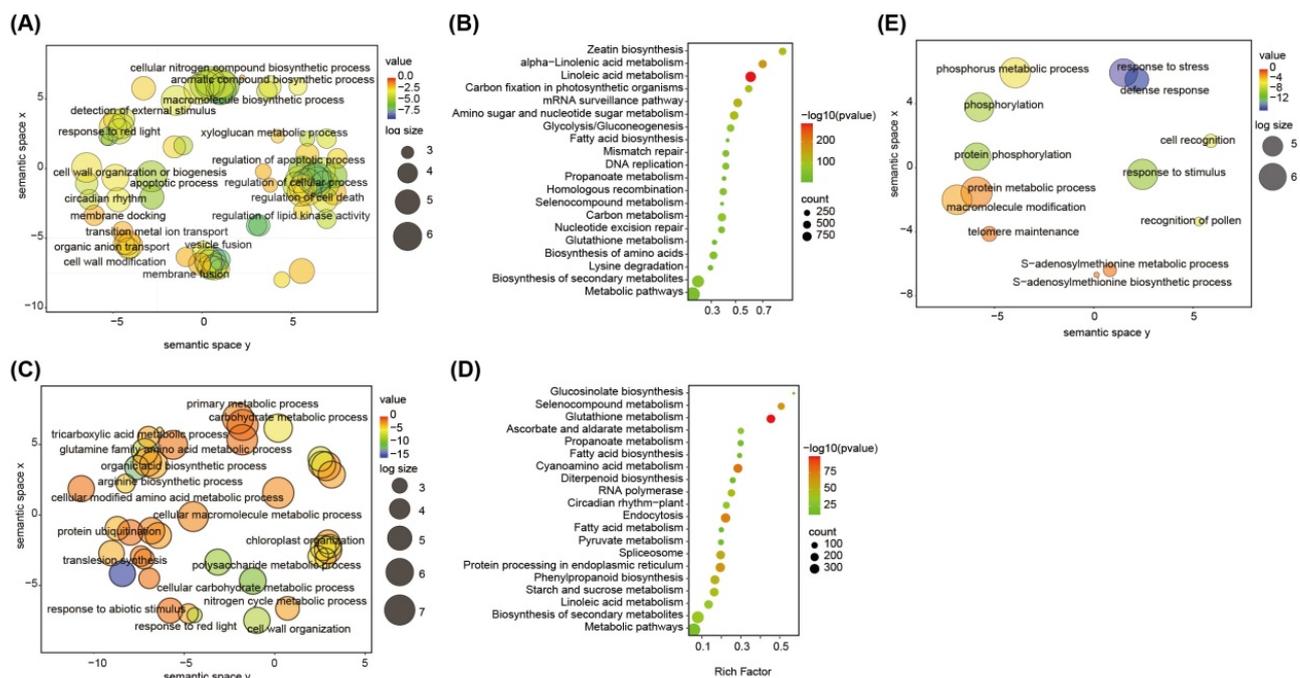
20,024 gene pairs in 833 synteny blocks (Table S5). *L. japonicus* chromosomes 1 and 5 were almost entirely syntenic with *M. truncatula* chromosomes 7 and 1, respectively. The remaining four chromosomes had large synteny blocks each with four or five *M. truncatula* chromosomes (Figure 1D).



**Figure 1.** Evolution and synteny of the *Lotus japonicus* genome. **(A)** The phylogenetic relationship and split-time estimation are based on all single-copy gene families shared by all species used in this analysis. **(B)** An overview of ortholog and paralog genes among *L. japonicus* and eight other plant species (Mt, *Medicago truncatula*, Gm, *Glycine max*, Tp, *Trifolium pretense*, Pv, *Phaseolus vulgaris*, Ca, *Cicer arietinum*, Vu, *Vigna unguiculata*, La, *Lupinus angustifolius*, and Ath, *Arabidopsis thaliana*). **(C)** Venn diagram showing the shared and unique gene families among *L. japonicus* and eight other plant species. **(D)** Synteny analysis between *L. japonicus* and *M. truncatula*.

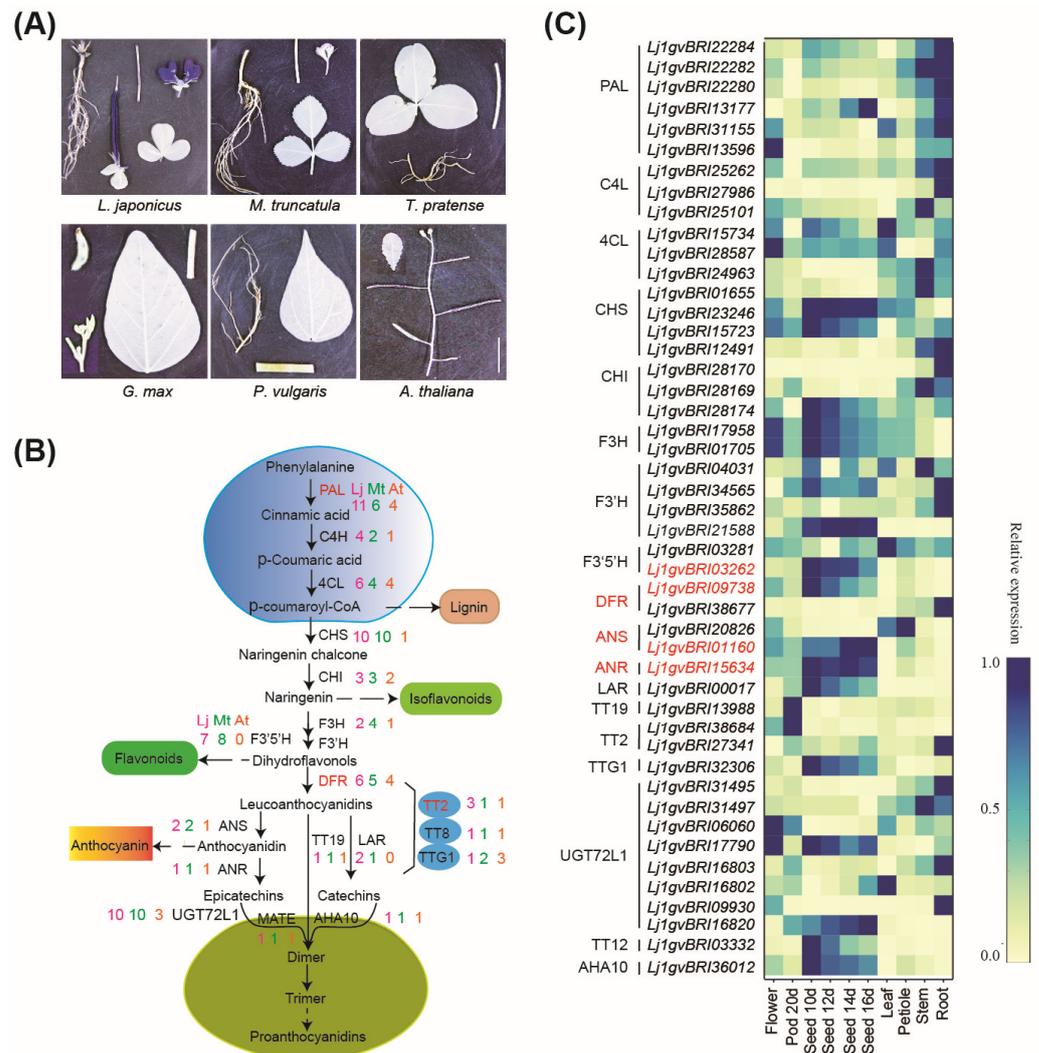
## 2.2. Genes Involved in PA Biosynthesis and Regulation

DMACA staining of organs of five legume species, namely *L. japonicus* MG-20, *M. truncatula*, *T. pretense*, *G. max*, and *P. vulgaris*, was performed, and PAs were found in the seed coat, stem, flower, and pod of *L. japonicus*. However, PAs were only found in the seed coat of *M. truncatula*, *G. max*, and *P. vulgaris* and in the flowers and seed coats of *T. pretense* (Figure 3A) [6].



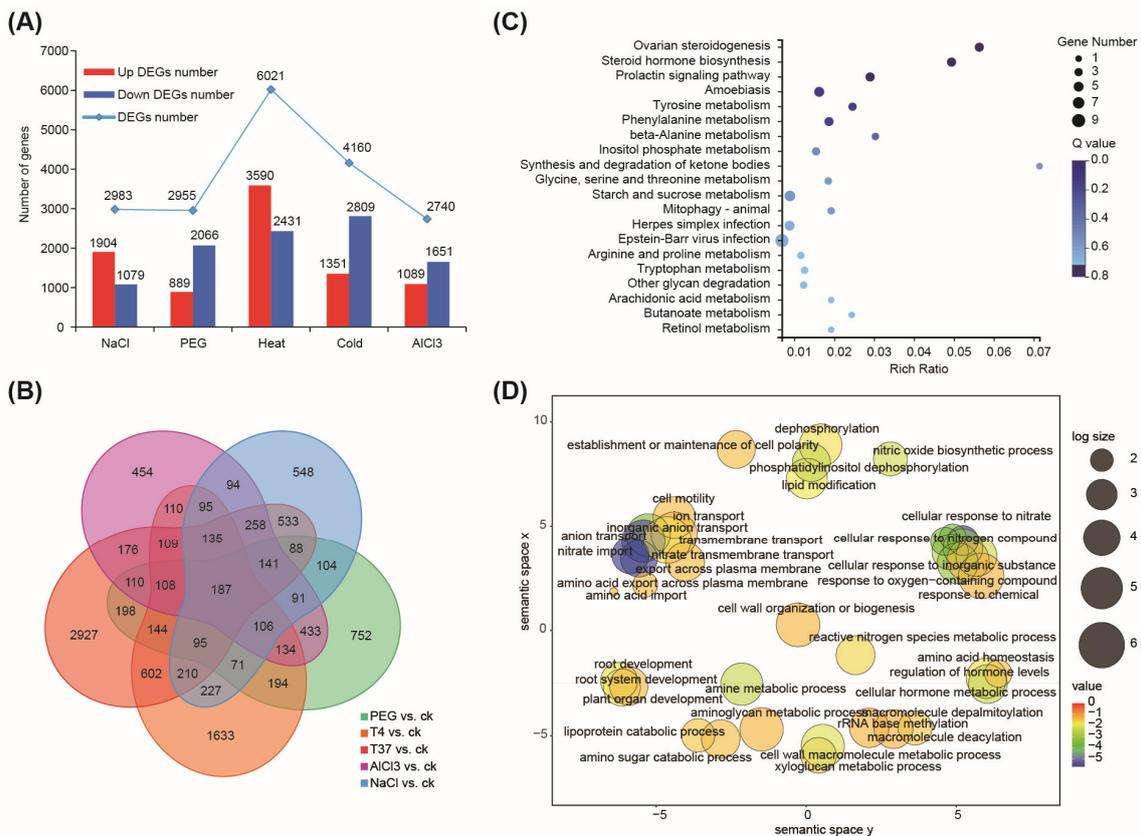
**Figure 2.** Functional enrichment of the *L. japonicus*-specific gene families and expansion or contraction of gene families. (A) GO enrichment analysis of specific gene families. (B) Statistics for the top 20 enriched pathways among the specific gene families. The degree of KEGG enrichment was determined by the enrichment factor,  $q$ -value, and gene number. The sizes and colors of spots represent the number of genes and the  $q$ -value. (C) GO enrichment analysis of contraction gene families. (D) GO enrichment analysis of expansion gene families. (E) Bubble plot of KEGG expansion gene families.

Seventy-two PA biosynthesis-related genes were present in the *L. japonicus* genome, of which 62 were present in *M. truncatula*, 55 in *T. pretense*, 58 in *P. vulgaris*, 72 in *G. max*, and 29 in *Arabidopsis* (Table S6). The copy numbers of the first key enzyme *PAL* and of genes in the late biosynthesis pathway encoding dihydroflavonol 4-reductase (*DFR*) and *TT2* in *L. japonicus* were significantly higher than those in the other species tested (Figure 3B). *PAL* catalyzes the conversion of L-phenylalanine to cinnamic acid, linking primary metabolism with secondary metabolism, a speed-limiting step in phenylpropanoid metabolism [7]. In total, 11 *PAL* genes were present in *L. japonicus* compared with 8 *PAL* genes in *G. max*, 7 in *T. pretense*, and 6 in *M. truncatula*. *DFR* is the first committed enzyme of the flavonoid pathway that leads to the production of common anthocyanins and PAs (Figure 4A). We identified six *DFRs* in the *L. japonicus* genome. A molecular phylogenetic tree of *DFR* from *L. japonicus*, *M. truncatula*, *A. thaliana*, *G. max*, *P. vulgaris*, and *T. pretense* was constructed using MEGA software 5.0 (Figure 4B). *TT2* acts as a key determinant for PA accumulation. We identified three *TT2s* in the *L. japonicus* genome, whereas only one *TT2* was found in the other legumes tested (Figure 4C).



**Figure 3.** Analysis of PA biosynthesis-related genes. **(A)** DMACA staining of different tissues from five legume species, including *L. japonicus* MG-20, *M. truncatula*, *T. pratense*, *G. max*, and *P. vulgaris*, and a nonlegume species *A. thaliana*. **(B)** Diagram of the PA biosynthesis pathway and copy numbers of genes involved in the PA biosynthesis pathway in *L. japonicus*. PAL: Phenylalanine ammonia lyase, C4H: cinnamate 4-hydroxylase, 4CL: 4-coumarate coenzyme A ligase, CHS: chalcone synthase, CHI: chalcone isomerase, F3H: flavonoid 3-hydroxylase, F3'H: flavanone'-hydroxylase, F3'5'H: flavonoid 3,5-hydroxylase, DFR: dihydroflavonol 4-reductase, LAR: leucocyanidin reductase, ANS/LDOX: anthocyanidin synthase/leucoanthocyanidin dioxygenase, ANR: anthocyanidin reductase, GST (TT19): glutathione S-transferase, TT2: transparent testa 2, TT8: transparent testa 8, TTG1: transparent testa glabra 1, MATE (TT12): multidrug and toxic compound extrusion protein, UGT72L1: epicatechin glucosyltransferase, and AHA10: *Arabidopsis* H<sup>+</sup>-ATPase 10. **(C)** Heatmaps showing the expression profiles of genes involved in PA biosynthesis pathways in different organs from *L. japonicus*. The expression values of genes at the row scale were normalized, and the values are indicated by a continuous color scheme. Blue indicates a high expression, and yellow indicates a low expression.





**Figure 5.** Differential gene expression analysis of response to abiotic stressors. **(A)** Number of genes expressed differentially in response to salt, drought, heat, cold, and AlCl<sub>3</sub> stresses. **(B)** Venn diagram showing the shared and unique DEGs in response to abiotic stresses. **(C)** Bubble plot of KEGG of 187 DEGs common to all five stresses. **(D)** GO enrichment analysis of 187 DEGs common to all five stresses.

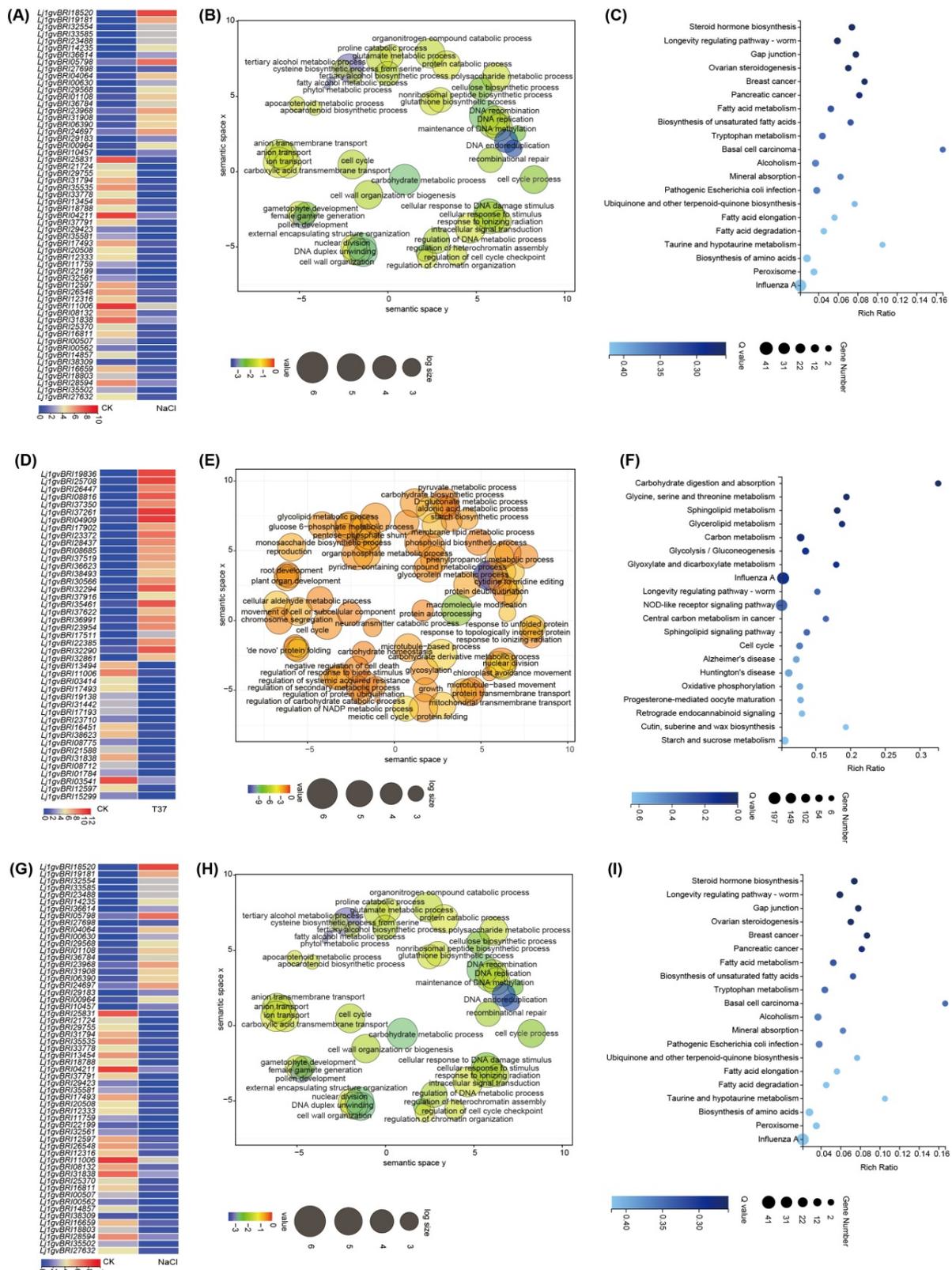
### 2.3.1. Shared Responses under Abiotic Stresses

A total of 187 DEGs were common to all five stresses (Figure 5B and Table S12). They were mainly enriched in steroid hormone biosynthesis, amino acid metabolism, homeostasis and transport, cell wall organization, and starch and sucrose metabolism (Figure 5C,D). *Lj1gvBRI23968.1*, *Lj1gvBRI35026.1*, and *Lj1gvBRI23372.1* were the top upregulated DEGs under the five stress conditions. Only nine TFs were common, including three *ERFs*, two *bZIPs*, one *MYB*, one *WRKY*, one *bHLH*, and one multiprotein bridging factor1c (*MBF*). Among them, the expression of only *MBF* (*Lj1gvBRI17497*) was upregulated and those of the other seven TFs were repressed in response to the five stresses.

Stress-specific transcription patterns are connected to upstream signaling via TFs [8]. We identified 548, 751, 2927, 1623, and 454 DEGs specifically in response to salt, PEG, heat, low temperature/cold, and AlCl<sub>3</sub>, respectively (Figure 5B). Among them, 48, 52, 110, 150, and 30 unique TF families were identified.

### 2.3.2. Cold Stress-Specific Transcription Patterns

MG-20 is the most tolerant to low temperature among the 18 *L. japonicus* ecotypes [9]. We analyzed the specific DEGs under cold stress (Table S13). Twenty-eight genes exhibited >100 fold change under cold stress (Figure 6A). Notably, two genes exhibited >1000 fold change in expression in cold stress, with the expression of *Lj1gvBRI02188* annotated as *CCR4* (a serine/threonine protein kinase-like protein) being increased by 1022 fold. *Lj1gvBRI33915* annotated as unknown was the second top DEG. Twenty-five TF families (150 members) were exclusively expressed during cold stress, such as *AP2/ERF* (35), *C2C2* (18), *MYB* (13), *WRKY* (13), *bHLH* (10), and *GRAS* (9).



**Figure 6.** Cold, heat and salt stress-specific transcription patterns. (A) Heatmaps showing expression profiles of top 38 specific DEGs under cold stress. (B) GO enrichment analysis of 1633 specific DEGs under cold stress. (C) Bubble Plot of KEGG of 1633 specific DEGs under cold stress. (D–F) Heat stress-specific transcription patterns. (G–I) Salt stress-specific transcription patterns.

According to GO analysis results, 1633 cold stress-specific genes were significantly enriched in functions related to secondary metabolism, JA biosynthesis/metabolism process, and chlorophyll catabolic process. Meanwhile, the top 20 KEGG pathways included inflammatory mediator regulation of TRP channels, arginine and proline metabolism, and NOD-like receptor signaling pathway (Figure 6B,C and Table S14).

### 2.3.3. Heat Stress-Specific Transcription Patterns

We analyzed the specific DEGs under heat stress. Seven genes had a fold change of >10 under heat stress. Notably, increase in the expression of *Lj1gvBRI04909*, annotated as unknown, was the highest (by 11.6 folds) (Table S15). *Lj1gvBRI17902* annotated as protein maintenance of meristems-like was the second top DEG (Figure 6D). A total of 15 TF families (110 members) were exclusively expressed during PEG stress, such as *AP2/ERF* (9), *FAR1* (8), *MYB* (7), *NAC* (7), *FHA* (6), and *HSF* (3).

Based on GO analysis results, 2927 heat stress-specific genes were significantly enriched in functions related to the glycolipid metabolic process, response to a topologically incorrect protein, and carbohydrate homeostasis. KEGG analysis indicated that these genes were enriched in the pathways related to cutin, suberine, and wax biosynthesis; starch and sucrose metabolism; and other pathways (Figure 6E,F and Table S16).

### 2.3.4. Salt Stress-Specific Transcription Patterns

We analyzed the expression of specific DEGs under salt stress. The expression of three genes had a fold change of >5 under salt stress (Table S18). Notably, the expression of *Lj1gvBRI14235*, annotated as the MYB TF, was increased 7.6-fold, indicating the highest increase (Figure 6G). In total, 17 TF families (47 members) were exclusively expressed during salt stress, such as *MYB* (9), *LOB* (6), *AP2/ERF* (5), and *bHLH* (5).

Based on the GO analysis results, 548 salt stress-specific genes were significantly enriched in functions related to the proline catabolic process, apocarotenoid metabolic process, anion transmembrane transport, and response to ionizing radiation. Meanwhile, according to KEGG analysis results, these genes were enriched in pathways related to steroid hormone biosynthesis, mineral absorption, and peroxisome. (Figure 6H,I and Table S17).

### 2.3.5. PEG Stress-Specific Transcription Patterns

We analyzed the expression of specific DEGs under drought stress. Three genes had a fold change of >4 under salt stress (Table S19), Notably, expression of *Lj1gvBRI04219*, annotated as the arabinogalactan protein, increased by 6.3-fold, indicating the highest increase (Figure 7A). In total, 15 TF families (41 members) were exclusively expressed during PEG stress, such as *MYB* (8), *bHLH* (7), *NAC* (5), and *WRKY* (4).

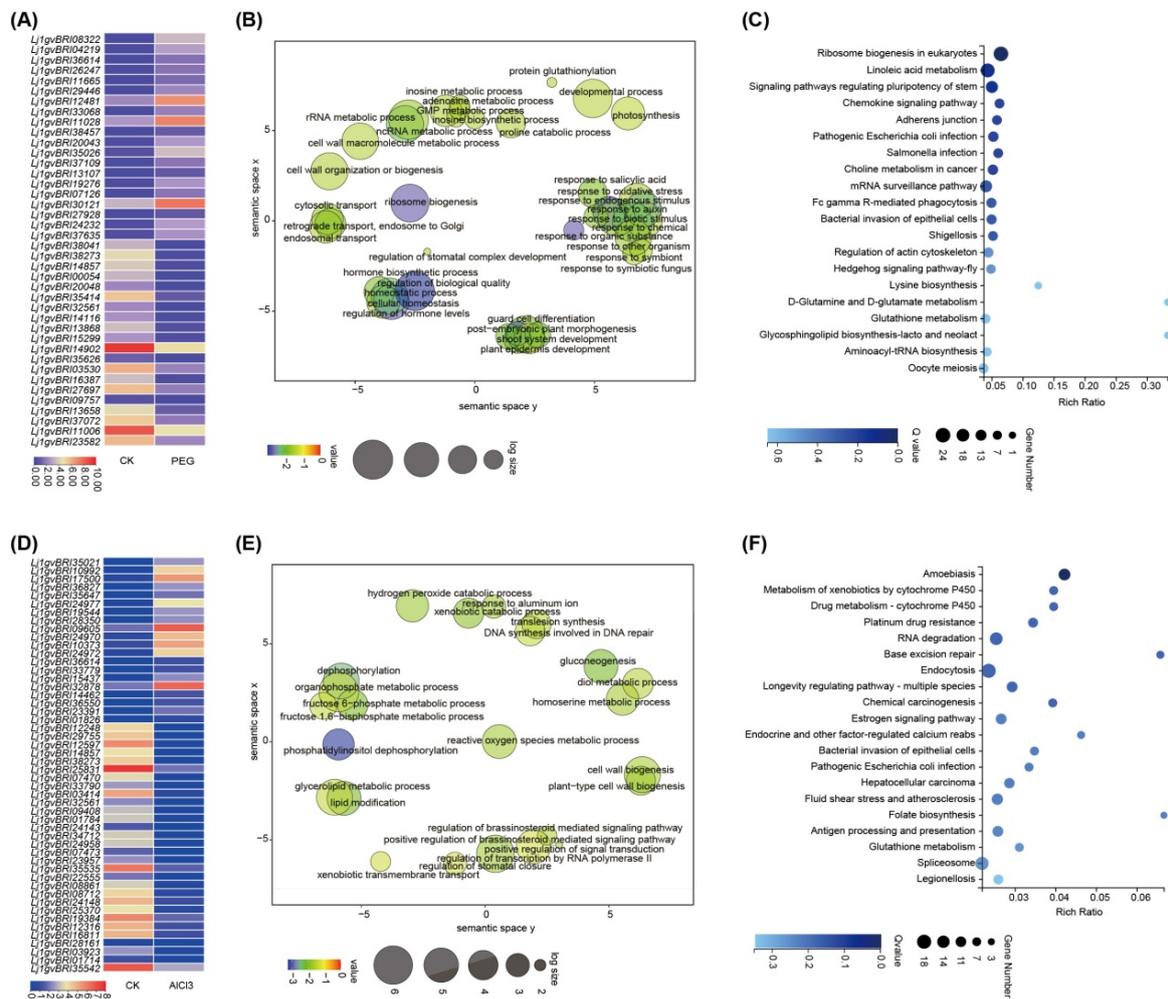
Based on the GO analysis results, 752 drought stress-specific genes were significantly enriched in functions related to cell wall organization or biogenesis, regulation of stomatal complex development, retrograde transport, and response to SA. Meanwhile, the top 20 KEGG pathways were signaling pathways regulating stem cell pluripotency, and glycosphingolipid biosynthesis-lacto and neolact (Figure 7B,C and Table S20).

### 2.3.6. AlCl<sub>3</sub> Stress-Specific Transcription Patterns

We analyzed the expression of specific DEGs under AlCl<sub>3</sub> stress. Six genes had a fold change of >4 under AlCl<sub>3</sub> stress (Table S21). Notably, the expression of *Lj1gvBRI36827*, annotated as VQ domain-containing protein, was the highest by 5.7-fold. *Lj1gvBRI35647* annotated as casein kinase II subunit alpha was the second top DEG (Figure 7D). In total, 15 TF families (30 members) were exclusively expressed during AlCl<sub>3</sub> stress, including *MYB* (6), *WRKY* (5), and *bHLH* (3).

According to GO analysis results, 454 AlCl<sub>3</sub> stress-specific genes were significantly enriched in the fructose 6-phosphate metabolic process, response to aluminum ion, lipid modification, regulation of BR-mediated signaling pathway. Meanwhile, the top 20 KEGG

pathways were related to folate biosynthesis, and drug metabolism—cytochrome P450 (Figure 7E,F and Table S22).



**Figure 7.** Drought and AlCl<sub>3</sub> stress-specific transcription patterns. (A) Heatmaps showing expression profiles of top 40 specific DEGs under salt stress. (B) GO enrichment analysis of 752 specific DEGs under salt stress. (C) Bubble Plot of KEGG of 752 specific DEGs under salt stress. (D–F) AlCl<sub>3</sub> Stress-specific transcription patterns.

### 3. Discussion

Using comparative genomics and evolutionary analysis, this study provides insights into the important agricultural traits of *L. japonicus* including PA biosynthesis and response to abiotic stress. We revealed that the expansive and specific gene families of *L. japonicus* were enriched in secondary metabolite biosynthesis and the detection of external stimuli. Furthermore, according to the copy numbers and transcriptional pattern of PA biosynthesis-related genes, we proposed that the expansion of PAL, DFR, and TT2 in the PA biosynthesis pathway and the transcription expression of F3/5/H, DFR, ANS, and ANR in most tissues mainly contribute to PA accumulation in the stem, petiole, flower, pod, and seed coat of *L. japonicus*. Finally, we reported that MYB and AP2/ERF play crucial roles in response to abiotic stresses.

The genome of *L. corniculatus*, which accumulates PAs in whole plants, remains unknown. The relative model *L. japonicus* may provide clues for PA biosynthesis in most tissues and organs. PA accumulation is a complex quantitative trait, and 14 main enzymes are involved in PA biosynthesis from the aromatic amino acid. Here, 72 PA biosynthesis-related genes were identified in the *L. japonicus* genome, which is higher than those found

in *M. truncatula* (62), *T. pretense* (55), *P. vulgaris* (58), and *Arabidopsis* (29) and the same as those in *G. max* (72). Especially, the expansion of PAL, which is the first speed-limiting key enzyme in phenylpropanoid biosynthesis that links primary metabolism with secondary metabolism [7]; DFR, which is the first committed enzyme of the flavonoid pathway that leads to anthocyanins and PAs [10]; and TT2s, which act as a key determinant for PA accumulation [11–13], have provided fundamental genetic materials for PA accumulation in more tissues. Furthermore, the expression of *F3'5'H*, *DFR*, *ANS* combined *ANR*, which provide the (-)-epicatechin extension, starter, and terminal units of PA polymers [14–17], were most strongly associated with PA accumulation in the stem, petiole, flower, pod, and seed coat of *L. japonicus*. In addition, expression profiles of 62 PA biosynthesis-related genes in *Medicago* showed that *ANS* and *ANR* were highly expressed only in the seed coat where the PAs accumulated exclusively (Figure 3C).

Drought, salinity, heat, cold, and  $AlCl_3$  stresses are considered major abiotic factors that negatively affect crop yield both qualitatively and quantitatively [4]. Identifying common and unique responses under these abiotic stresses is important for understanding the cross-talk mechanism [18]. According to the analysis of common transcriptional responses, the MBF gene *Lj1gvBRI17497.1* was screened. Its expression pattern was consistent with that of its homologue *AtMBF1c* in *Arabidopsis*, and its expression is specifically elevated in response to salinity, drought, heat, hydrogen peroxide, and pathogen infection. Furthermore, constitutive expression enhances the tolerance of transgenic plants to bacterial infection, heat, and osmotic stress [19]. Therefore, we propose that the gene MBF *Lj1gvBRI17497.1* may be an essential candidate gene having a critical role in response to multiple abiotic stresses.

Among the five types of abiotic stress, MG-20 is most tolerant to low temperature among the 18 *L. japonicus* ecotypes [9]. We found some clues based on the expression profile of specific DEGs under cold stress: there were more top DEGs (28 genes with fold change >100), and the expression level of the top DEGs was the highest (two genes with a fold change of >1000). Furthermore, these transcriptomic data and associated analysis results provide a resource for studying the response of *L. japonicus* to abiotic factors. According to shared and unique TF responses under abiotic stress, we proposed that MYB and AP2/ERF play more crucial roles under abiotic stresses.

In summary, based on our previous high-quality *L. japonicus* genome assembly and annotation, this study provides insights into the crucial agricultural traits of *L. japonicus*, including PA biosynthesis and response to abiotic stress by using comparative genomics and evolutionary analyses. We revealed the expansive and specific gene families of *L. japonicus* enriched in the biosynthesis of secondary metabolites and detection of external stimuli. We also proposed that expansion of PAL, DFR, and TT2 in the PA biosynthesis pathway and the transcription of *F3'5'H*, *DFR*, *ANS*, and *ANR* in most tissues mainly contribute to PA accumulation in the stem, petiole, flower, pod, and seed coat of *L. japonicus*. Lastly, we report that MYB and AP2/ERF play more crucial roles in abiotic stresses. This may provide valuable gene resources for legume forage abiotic stress resistance and nutrient improvement.

## 4. Materials and Methods

### 4.1. Phylogenetic Analysis and Gene Synteny Analysis

The genomes of *L. japonicus* (CNA0050696) and eight other plants, namely *M. truncatula*, *Glycine max*, *T. pretense*, *Phaseolus vulgaris*, *C. arietinum*, *Vigna unguiculata*, *Lupinus angustifolius*, and *Arabidopsis thaliana* (<https://phytozome-next.jgi.doe.gov/>, accessed on 22 November 2011), were used for evolutionary analysis. Paralogous and orthologous genes were identified using OrthoFinder [20]. The phylogenetic tree was constructed using RaxML [21]. The divergence times among the nine species were estimated using MCMCTREE within the PAML v4.9 package [22].

The expanded or contracted gene families were defined using CAFÉ v. 4 [23], and the significant genes were used for Kyoto Encyclopedia of Genes and Genomes (KEGG)

pathway enrichment analysis. MCscanX was used to identify synteny blocks, and WGDI was used to determine the whole genome replication time (WGD) of *L. japonicus*, soybean, and alfalfa.

#### 4.2. PA Staining with 4-Dimethylaminocinnamaldehyde

Healthy and fresh organs (roots, stems, leaves, flowers, and pods) of the reproductive stage were selected, decolorized in absolute ethanol containing 30% acetic acid for 12–18 h, stained with a cold 4-dimethylaminocinnamaldehyde (DMACA) reagent (0.3% *w/v* DMACA in 3N HCl/50% *w/v* ethanol), and washed three times using 75% ethanol. The stained organs can be protected in 70% ethanol, following which the color of PA-rich organs or tissues changes to blue [24]. The experiments were performed three times.

*Phaseolus vulgaris* (G19833), *Medicago truncatula* (A17), *Medicago sativa* (XinJiang DaYe), *Trifolium pratense* (Milvus B), *Glycine max* (Williams 82), *Lotus japonicus* (MG-20), *Trifolium repens* (Haifa), and *Arabidopsis thaliana* (Col-0) were used for visualization of PAs by DMACA staining. The plants were cultured at  $25 \pm 2$  °C in a growth chamber with a photoperiod of 16 h/8 h.

#### 4.3. Phylogenetic Tree and Transcriptional Expression Analysis

The complete amino acid sequences of proteins were aligned by ClustalX2 [25], and the phylogenetic tree was constructed by neighbor-joining algorithms of the MEGA 5.0 software (Test of phylogeny: Bootstrap method, No. of bootstrap replications: 1000, Model/Method: p-distance, Gaps/Missing data treatment: Pairwise deletion) [26].

The transcriptional expression data of *Medicago truncatula* were downloaded from the website <https://mtsspdb.zhaolab.org/database/>, accessed on 20 February 2020. We generated the heat maps of the expression patterns using TBtool II software [27].

#### 4.4. Identification of Differentially Expressed Genes under Abiotic Stress

Total RNA was extracted using the RNAprep Pure plant Kit. mRNA was purified using oligo(dT)-linked magnetic beads following the manufacturer's instructions. The purified mRNA was sequenced on the DNBSEQ platform, yielding 91.55 Mb data for each sample. RNA-seq reads were trimmed using SOAPnuke (version 1.5.6) [28] and mapped against *L. japonicus* reference genomes with Bowtie2 [29]. Gene expression values were calculated using the RSEM program [30]. Differential expression analysis was performed using DESeq2 [31]. GO and KEGG pathway enrichment analyses were performed using the phyper function in R package ( $q$ -value  $\leq 0.05$ ). The GO enrichment bubble chart was drawn using REVIGO (<http://revigo.irb.hr/>), then the exported R language pack was run by Rstudio.

#### 4.5. Plant Materials and Treatments

*L. japonicus* seeds were treated with 96% sulfuric acid (H<sub>2</sub>SO<sub>4</sub>) for 10 min and washed five to six times in sterile distilled water. Subsequently, seeds were placed on the Murashige–Skoog (MS) agar medium in a growth chamber with 16 h of light/8 h of dark; 30-day-old seedlings were transferred to the MS liquid medium supplemented with different stress solutions (150 mM NaCl, 15% polyethylene glycol 6000, 500  $\mu$ M AlCl<sub>3</sub> with pH 4.5, 37 °C, 0 °C), MS medium as the control. The seedlings for RNA-seq were harvested 6 h after treatment; untreated seedlings were used as control. All samples were frozen in liquid nitrogen and stored at  $-80$  °C until use.

### 5. Conclusions

In summary, based on our previous high-quality *L. japonicus* genome assembly and annotation, this study provides insights into the crucial agricultural traits of *L. japonicus*, including PA biosynthesis, and response to abiotic stress by using comparative genomics and evolutionary analyses. We revealed the expansive and specific gene families of *L. japonicus* enriched in the biosynthesis of secondary metabolites and detection of external stimuli. We suggested that increased copy numbers and transcription of PA-related genes

contribute to PA accumulation in the stem, petiole, flower, pod, and seed coat of *L. japonicus*. Meanwhile, according to shared and unique transcription factors responding to five abiotic stresses, we revealed that MYB and AP2/ERF play more crucial roles in abiotic stresses. Our study provides new insights into the key agricultural traits of *L. japonicus* including PA biosynthesis and response to abiotic stress. This may provide valuable gene resources for legume forage abiotic stress resistance and nutrient improvement.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants13081151/s1>, Table S1. Gene family analysis of predicted *Lotus japonicus* genes in comparison to eight other dicot genomes; Table S2. Summary of orthologous and paralogous genes in *Lotus japonicus* relative to eight plant (legume); Table S3. GO, KEGG, SwissProt and Nr description of expanded genes in *Lotus japonicus*; Table S4. GO, KEGG, SwissProt and Nr description of contraction genes in *Lotus japonicus*; Table S5. Statics macrosynteny between *Lotus japonicus* and *Medicago truncatula*; Table S6. Copy number of proanthocyanidins-biosynthesis-related genes in legume plants; Table S7. Protein sequence of the genes related proanthocyanidins synthesis in *Arabidopsis*; Table S8. Statistics of transcription reads; Table S9. Transcription reads mapped to genome; Table S10. Transcription reads mapped to gene sets; Table S11. Number of genes differentially expressed in response to abiotic stress; Table S12. Common differentially expressed genes in NaCl, PEG, Heat, Cold and AlCl<sub>3</sub> stress; Table S13. Differentially expressed genes (DEGs) between low temperature (0 °C) stress and control (CK); Table S14. KEGG enrichment of low temperature vs. CK DEGs; Table S15. Differentially expressed genes (DEGs) between high temperature (37 °C) stress and control (CK); Table S16. KEGG enrichment of high temperature vs. CK DEGs; Table S17. Differentially expressed genes (DEGs) between NaCl stress and control (CK); Table S18. KEGG enrichment of NaCl vs. CK DEGs; Table S19. Differentially expressed genes (DEGs) between PEG stress and control (CK); Table S20. KEGG enrichment of PEG vs. CK DEGs; Table S21. Differentially expressed genes (DEGs) between AlCl<sub>3</sub> stress and control (CK); Table S22. KEGG enrichment of AlCl<sub>3</sub> vs. CK DEGs.

**Author Contributions:** Z.S. performed research, analyzed data, and wrote the manuscript; Z.L., M.Z., Q.R., W.X. and Y.T. analyzed the data; Y.W. revised the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by National Natural Science Foundation of China (32001395), Chongqing Natural Science Foundation (stc2020jcyjmsxmX0626), The Agricultural Science and Technology Innovation Program (ASTIP).

**Data Availability Statement:** All data generated or analyzed during this study are included in this published article (and its Supplementary Materials Files). All RNA sequencing data in this study are available on SRA through accession numbers SRP399738 (Cold stress), SRP399738 (Heat stress), SRP399738 (NaCl stress), SRP399738 (PEG stress), SRP399737 (AlCl<sub>3</sub> stress), and SRP399738 (CK).

**Acknowledgments:** We are grateful to researcher Zhe Liang for help, advice, and discussion.

**Conflicts of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Kamal, N.; Mun, T.; Reid, D.; Lin, J.S.; Akyol, T.Y.; Sandal, N.; Asp, T.; Hirakawa, H.; Stougaard, J.; Mayer, K.F.X.; et al. Insights into the evolution of symbiosis gene copy number and distribution from a chromosome-scale *Lotus japonicus* Gifu genome sequence. *DNA Res.* **2020**, *27*, dsaa015. [[CrossRef](#)] [[PubMed](#)]
2. Li, H.; Jiang, F.; Wu, P.; Wang, K.; Cao, Y. A High-Quality Genome Sequence of Model Legume *Lotus japonicus* (MG-20) Provides Insights into the Evolution of Root Nodule Symbiosis. *Genes* **2020**, *11*, 483. [[CrossRef](#)] [[PubMed](#)]
3. Sun, Z.; Jiang, S.; Wang, D.; Li, L.; Liu, B.; Ran, Q.; Hu, L.; Xiong, J.; Tang, Y.; Gu, X.; et al. Single-cell RNA-seq of *Lotus japonicus* provide insights into identification and function of root cell types of legume. *J. Integr. Plant Biol.* **2023**, *65*, 1147–1152. [[CrossRef](#)] [[PubMed](#)]
4. Bashir, K.; Matsui, A.; Rasheed, S.; Seki, M. Recent advances in the characterization of plant transcriptomes in response to drought, salinity, heat, and cold stress. *F1000Research* **2019**, *8*, 658. [[CrossRef](#)] [[PubMed](#)]
5. Márquez, A.J. *Lotus japonicus Handbook*; Springer: Dordrecht, The Netherlands, 2005.
6. Roldan, M.B.; Cousins, G.; Muetzel, S.; Zeller, W.E.; Fraser, K.; Salminen, J.P.; Blanc, A.; Kaur, R.; Richardson, K.; Maher, D.; et al. Condensed Tannins in White Clover (*Trifolium repens*) Foliar Tissues Expressing the Transcription Factor TaMYB14-1 Bind to Forage Protein and Reduce Ammonia and Methane Emissions in vitro. *Front. Plant Sci.* **2021**, *12*, 777354. [[CrossRef](#)] [[PubMed](#)]

7. Li, G.; Wang, H.; Cheng, X.; Su, X.; Zhao, Y.; Jiang, T.; Jin, Q.; Lin, Y.; Cai, Y. Comparative genomic analysis of the *PAL* genes in five Rosaceae species and functional identification of Chinese white pear. *PeerJ* **2019**, *7*, e8064. [[CrossRef](#)]
8. Zhang, H.; Zhu, J.; Gong, Z.; Zhu, J.K. Abiotic stress responses in plants. *Nat. Rev. Genet.* **2022**, *23*, 104–119. [[CrossRef](#)] [[PubMed](#)]
9. Calzadilla, P.I.; Signorelli, S.; Escaray, F.J.; Menendez, A.B.; Monza, J.; Ruiz, O.A.; Maiale, S.J. Photosynthetic responses mediate the adaptation of two *Lotus japonicus* ecotypes to low temperature. *Plant Sci.* **2016**, *250*, 59–68. [[CrossRef](#)] [[PubMed](#)]
10. Chen, X.; Liu, W.; Huang, X.; Fu, H.; Wang, Q.; Wang, Y.; Cao, J. Arg-type dihydroflavonol 4-reductase genes from the fern *Dryopteris erythrosora* play important roles in the biosynthesis of anthocyanins. *PLoS ONE* **2020**, *15*, e0232090. [[CrossRef](#)] [[PubMed](#)]
11. Escaray, F.J.; Passeri, V.; Perea-Garcia, A.; Antonelli, C.J.; Damiani, F.; Ruiz, O.A.; Paolucci, F. The *R2R3-MYB TT2b* and the *bHLH TT8* genes are the major regulators of proanthocyanidin biosynthesis in the leaves of *Lotus* species. *Planta* **2017**, *246*, 243–261. [[CrossRef](#)]
12. Nesi, N.; Jond, C.; Debeaujon, I.; Caboche, M.; Lepiniec, L. The Arabidopsis *TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* **2001**, *13*, 2099–2114. [[CrossRef](#)] [[PubMed](#)]
13. Verdier, J.; Zhao, J.; Torres-Jerez, I.; Ge, S.; Liu, C.; He, X.; Mysore, K.S.; Dixon, R.A.; Udvardi, M.K. MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in *Medicago truncatula*. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1766–1771. [[CrossRef](#)] [[PubMed](#)]
14. Bogs, J.; Downey, M.O.; Harvey, J.S.; Ashton, A.R.; Tanner, G.J.; Robinson, S.P. Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. *Plant Physiol.* **2005**, *139*, 652–663. [[CrossRef](#)] [[PubMed](#)]
15. Liu, C.; Wang, X.; Shulaev, V.; Dixon, R.A. A role for leucoanthocyanidin reductase in the extension of proanthocyanidins. *Nat. Plants* **2016**, *2*, 16182. [[CrossRef](#)] [[PubMed](#)]
16. Abrahams, S.; Lee, E.; Walker, A.R.; Tanner, G.J.; Larkin, P.J.; Ashton, A.R. The Arabidopsis *TDS4* gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. *Plant J.* **2003**, *35*, 624–636. [[CrossRef](#)] [[PubMed](#)]
17. Jun, J.H.; Xiao, X.; Rao, X.; Dixon, R.A. Proanthocyanidin subunit composition determined by functionally diverged dioxygenases. *Nat. Plants* **2018**, *4*, 1034–1043. [[CrossRef](#)] [[PubMed](#)]
18. Sharma, R.; Singh, G.; Bhattacharya, S.; Singh, A. Comparative transcriptome meta-analysis of *Arabidopsis thaliana* under drought and cold stress. *PLoS ONE* **2018**, *13*, e0203266. [[CrossRef](#)] [[PubMed](#)]
19. Suzuki, N.; Rizhsky, L.; Liang, H.; Shuman, J.; Shulaev, V.; Mittler, R. Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c. *Plant Physiol.* **2005**, *139*, 1313–1322. [[CrossRef](#)]
20. Emms, D.M.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157. [[CrossRef](#)]
21. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [[CrossRef](#)]
22. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)] [[PubMed](#)]
23. De Bie, T.; Cristianini, N.; Demuth, J.P.; Hahn, M.W. CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **2006**, *22*, 1269–1271. [[CrossRef](#)] [[PubMed](#)]
24. Abeynayake, S.W.; Panter, S.; Mouradov, A.; Spangenberg, G. A high-resolution method for the localization of proanthocyanidins in plant tissues. *Plant Methods* **2011**, *7*, 13. [[CrossRef](#)] [[PubMed](#)]
25. Larkin, M.A.; Blackshields, G.; Brown, N.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)] [[PubMed](#)]
26. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [[CrossRef](#)] [[PubMed](#)]
27. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **2020**, *13*, 1194–1202. [[CrossRef](#)]
28. Chen, Y.; Chen, Y.; Shi, C.; Huang, Z.; Zhang, Y.; Li, S.; Li, Y.; Ye, J.; Yu, C.; Li, Z.; et al. SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **2018**, *7*, 1–6. [[CrossRef](#)]
29. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)]
30. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)]
31. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.