agriculture

MDPI

Article

Transcriptome Sequencing of *Agave angustifolia* Reveals Conservation and Diversification in the Expression of *Cinnamyl Alcohol Dehydrogenase* Genes in *Agave* Species

Xing Huang [1], Bochao Xu [2], Shibei Tan [1], Yanlei Huang [3], Jingen Xi [1], Xu Qin [4], Tao Chen [4], Helong Chen [1], Xiaohan Yang [5,*] and Kexian Yi [1,6,7,8,*]

[1] Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China; hxalong@gmail.com (X.H.); tanshibei915@163.com (S.T.); xijingen@163.com (J.X.); chenhelong951@126.com (H.C.)
[2] School of Life Sciences, Hainan University, Haikou 570228, China; xbc27960096@163.com
[3] College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China; hyl4532323642022@163.com
[4] Guangxi Subtropical Crops Research Institute, Nanning 530001, China; qinxu847@163.com (X.Q.); 15607718198@wo.cn (T.C.)
[5] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[6] Key Laboratory of Integrated Pest Management on Tropical Crops, Ministry of Agriculture and Rural Affairs, Haikou 571101, China
[7] Key Laboratory of Low Carbon Green Agriculture in Tropical China, Ministry of Agriculture and Rural Affairs, Haikou 571101, China
[8] Hainan Key Laboratory for Monitoring and Control of Tropical Agricultural Pests, Haikou 571101, China
* Correspondence: yangx@ornl.gov (X.Y.); yikexian@126.com (K.Y.)

Citation: Huang, X.; Xu, B.; Tan, S.; Huang, Y.; Xi, J.; Qin, X.; Chen, T.; Chen, H.; Yang, X.; Yi, K. Transcriptome Sequencing of *Agave angustifolia* Reveals Conservation and Diversification in the Expression of *Cinnamyl Alcohol Dehydrogenase* Genes in *Agave* Species. *Agriculture* **2022**, *12*, 1003. https://doi.org/10.3390/agriculture12071003

Academic Editor: Fanchang Zeng

Received: 21 June 2022
Accepted: 9 July 2022
Published: 11 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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/).

Abstract: *Agave angustifolia* is an important crassulacean acid metabolism plant species, with wide applications in beverage and sisal fiber production. In this study, we carried out a transcriptome analysis of *A. angustifolia* leaves, generating a total of 58,482,436 clean reads through Illumina paired-end sequencing. De novo transcriptome assembly generated 67,314 unigenes, with about half of them having homologs in four public databases. In the Nr database, *Asparagus officinalis* was shown to be most closely related to agave, and the early angiosperm *Amborella trichopoda* was selected as an outgroup species. We further identified five, two, seven, seven, seven, six and six *CAD* genes in asparagus, amborella, *A. deserti*, *A. tequilana*, *A. americana*, *A.* hybrid H11648 and *A. angustifolia*, respectively. The maximum likelihood phylogenetic tree revealed the species-specific expansion of *CAD* genes in arabidopsis, rice and agave. The expression analysis indicated the conservatively expressed *CAD1/2/4/6*, providing candidate targets for manipulation to improve lignin traits. The species-specific expression of *CAD3/5/7* indicates the existence of different regulatory mechanisms controlling the expression of these genes in agave species. This study presents the first transcriptome dataset of *A. angustifolia*, facilitating future studies on lignin biosynthesis in agave.

Keywords: transcriptome; *Agave angustifolia*; cinnamyl alcohol dehydrogenase; phylogeny; gene expression

1. Introduction

Agave species are important vegetations in semiarid and arid regions, which are spread from the geographic center of origin in Mexico to many countries or regions in the world [1]. *Agave* species are typical crassulacean acid metabolism plants with large biomass, high tolerance to the stresses of drought, cold and heat temperature and great potential for bioenergy production in arid and semiarid regions [2,3]. Among the 166 *Agave* species, *A. angustifolia* is a local plant species distributed from Mexico to Panama [4]. It is

commonly used for producing the fermented and distilled beverage Bacanora [5]. Additionally, *A. angustifolia* has a high cellulose content (67%) in leaf fiber, which has been applied for breeding the famous agave cultivar *Agave* hybrid H11648 ((*A. amaniensis* Trel. & Nowell × *A. angustifolia* Haw.) × *A. amaniensis*) [6,7]. It has become the main cultivar for sisal fiber production in tropical areas of Brazil, China and African countries, with the potential to the increase yield by improving fiber-related traits [8,9]. However, the molecular basis of fiber biosynthesis in agave is not well understood. To date, there have been several studies related to fiber traits in agave. A transcriptome comparison between wild and cultivated agave species has revealed six candidate genes related to fiber traits [10]. Five *cellulose synthase A* (*CesA*) genes have been identified by transcriptome mining in *A.* hybrid H11648 [11]. The expression patterns of *SMALL AUXIN UP-REGULATED RNA* (*SAUR*) and *NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER FAMILY* (*NPF*) family genes have been well characterized, with several genes from the *SAUR* and *NPF* families identified as candidate regulators of leaf development [12,13]. Recent studies have revealed the expression patterns of cell wall-related genes in *A. sisalana*, *A. fourcroydes*, *A.* hybrid H11648 and *A. tequilana* [9,14]. However, the large genomes and the long life-cycles have significantly restricted genetic research in agave, which impose a big challenge to revealing the molecular mechanism of fiber traits [6]. The next generation sequencing technology has undoubtedly become a useful and efficient tool to generate large transcript sequencing datasets in *A. angustifolia*, facilitating the functional genomics research in this important agave species [15].

Cinnamyl alcohol dehydrogenase (CAD) plays a key enzyme in the metabolic pathway of phenylpropanoid, which catalyzes the reduction of *p*-coumaricaldehyde, coniferylaldehyde and sinapylaldehyde by NADPH [16]. Their alcohol derivatives will be further polymerized into lignin, a main component of fiber [17]. The *CAD* gene family has been functionally characterized in various plants, such as arabidopsis, flax, melon, pear, poplar, sorghum, switchgrass, tobacco and wheat [18,19]. It is also an important target for genetic manipulation to control lignin-related phenotypes [20]. Flax mutants of *CAD* genes show obvious morphological modifications in bast fiber, which provides a typical reference for agave researchers [21]. Thus, we have sequenced and assembled the leaf transcriptome of *A. angustifolia* based on Illumina sequencing to gain a deep understanding of gene expression in relation to fiber traits. Furthermore, we performed a comparative analysis of *CAD* genes in five agave species, including *A. deserti*, *A. tequilana*, *A. americana*, *A.* hybrid H11648 and *A. angustifolia*. The results will provide new insights into the molecular basis of the underlying lignin traits in agave.

2. Materials and Methods

2.1. Plant Materials and RNA Isolation

The plants of *A. deserti*, *A. tequilana*, *A. americana*, *A.* hybrid H11648 and *A. angustifolia* were cultivated in the Wenchang experimental field (19.54° N 110.77° E) of Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences since 2013. The samples of young leaves were collected from three-year-old plants, as previously described [13]. The samples of each species were repeated three times with different individuals as biological replicates and soaked in liquid nitrogen for immediate freezing. The total RNA of each sample was isolated with the Tiangen RNA prep Pure Plant Kit (Tiangen Biomart, Beijing, China) after being ground into powder and then stored in a freezer at −80 °C.

2.2. Illumina Sequencing and Transcriptome Assembly

The total RNA samples of *A. angustifolia* were delivered to Genoseq Technology Co. Ltd. (Wuhan, Hubei, China), where the experiments of library construction and Illumina sequencing were conducted [22]. The RNA samples were examined with the NanoDrop Spectrophotometers (Thermo Fisher Scientific, Waltham, MA, USA) and the 2100

Bioanalyzer Instrument (Agilent Technologies, Santa Clara, CA, USA). The mRNAs were collected from 10 μg of qualified RNA with poly-T oligo-attached magnetic beads and then fragmented with the TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA, USA). The mRNA fragments were used as templates for first strand cDNA synthesis by M-MuLV Reverse Transcriptase (RNase H) together with a random hexamer primer. The second strand cDNA was synthesized with DNA Polymerase I and RNase H. The double-stranded cDNA fragments were modified with single 'A' bases and then added with adapters, which were used to construct a cDNA library after gel purification and PCR amplification. The sequencing was carried out by the Illumina HiSeq platform which generated paired-end raw reads with 150 bp at length.

The raw reads were submitted to the public database Sequence Read Archive (SRA) under the accession PRJNA837721 [23]. Clean reads were obtained after filtering the sequences of the adapter and low quality with Cutadapt and Trimmomatic software, respectively [24,25]. The Trinity software was used for de novo transcriptome assembly, which was annotated according to public databases, including the NCBI non-redundant protein database (Nr), the Gene Ontology (GO), the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Swiss-Prot [26–30].

2.3. Identification and Phylogenetic Analysis of CAD Genes

The CAD proteins of *Arabidopsis thaliana* (9) and *Oryza sativa* (12) were utilized as a query sequence to search homologous genes using tblastn method [31–33]. The previously published transcriptome data of five species in agave genus were selected as the database for the tblastn search, including *A. deserti*, *A. tequilana*, *A. americana*, *A.* hybrid H11648 and *A. angustifolia* [2,11,14,34]. *Asparagus officinalis* (Asparagoideae) was selected as a reference species due to its close phylogenetic position with *Agave* species (Agavoideae) since they were all belonging to Asparagaceae [35]. The early angiosperm *Amborella trichopoda* was selected as an outgroup species [36]. Agave *CAD* transcripts with a full coding region were identified by the ORF-FINDER software [37]. The full-length agave *CAD* genes were selected to calculate the protein length (aa), molecular weight (Da) and theoretical isoelectric point (pI) by using the software ProtParam tool [38]. The subcellular localization of proteins was predicted by using the CELLO software [39]. The full-length CAD proteins of the nine species (amborella, arabidopsis, rice, *A. deserti*, *A. tequilana*, *A. americana*, *A.* hybrid H11648 and *A. angustifolia*) were aligned by the ClustalX method for phylogenetic analysis to construct the maximum likelihood tree with a bootstrap value of 1000 trials using the MEGA 5.0 software [40]. These protein sequences of the nine species were further aligned in DNAMAN7 software to the detection of conserved amino acid residues [41].

2.4. Gene Expression Analysis

The clean reads were selected and mapped to the assembled *A. angustifolia* transcriptome sequences to calculate read counts of unigenes with the RSEM method, which were further normalized into reads per kilobase of transcript per million mapped reads (RPKM) [42,43]. The RPKM values of *CAD* genes in other agave species were calculated by the RSEM method according to previous studies and then normalized with the reference gene *protein phosphatase 2A* (*PP2A*) [2,11,34,42]. A QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) was selected for RT-qPCR validation of the gene expression. The program consisted of the initial stage (94 °C, 30 s), the cycle stage of 40 times (94 °C, 5 s and 60 °C, 30 s) and the dissociation stage. The GoScript Reverse Transcription System (Promega, Madison, WI, USA) was utilized for reverse transcription of RNA samples. The reaction solution of 20 μL contained 10 μL of TransStart Tip Green qPCR Supermix (Transgen Biotech, Beijing, China), 7.6 μL of ddH_2O, 0.5 μL of each gene-specific primer (10 μM), 1 μL of cDNA template and 0.4 μL of Passive Reference Dye (50×) (Transgen Biotech, Beijing, China). A technical replicate was conducted 3 times for each sample. Seven pairs of primers specific for agave *CAD* genes were synthesized after a

primer design by using the Primer 3 software (Table 1) [44]. The ΔΔCt method was used to calculate relative expression levels with the reference gene *protein phosphatase 2A* (*PP2A*) as an endogenous control [12,45].

Table 1. Conserved primers used for qRT-PCR analysis in five agave species.

Gene ID	Forward Primer	Reverse Primer	Product Size (bp)
CAD1	TGAGATCACCGGCATAATCA	TCTGTCAGCAACCAGCATTC	223
CAD2	GACGCCTACTTGGTCAGCTC	GAACTGCATTGGCTGATTGA	171
CAD3	TTGCAGTCAAGTTTGGCAAG	TTGAGCAACTGCAGAATTGG	214
CAD4	TCAGTACAAGCGCATCCAAG	ACCCCACCAACTTTCAACAG	181
CAD5	CAAAAGAGAACGGGAAGCAG	TGCTGTGAAGGTCTGAGTGG	183
CAD6	TGGGATGAAGGTGACAGTGA	AAAAGATCAACGGCACAAGG	182
CAD7	TGTCACCGAAGTAGGCACTG	AATCCATAGGCAAGGTGTCG	248
PP2A	CCTCCTCCTCCTTCGGTTTG	GCCATGAATGTCACCGCAGA	235

3. Results

3.1. De Novo Assembly of A. angustifolia Transcriptome

The leaf sample of *A. angustifolia* was selected for Illumina paired-end sequencing, which generated 72,184,668 raw reads. A total of 58,482,436 clean reads were obtained, with a total of 8,166,023,895 bp in length. The GC content, Q20 value (base quality >20) and Q30 value (base quality > 30) were 49.05%, 98.43% and 95.17%, respectively. The clean data was de novo assembled into 67,314 unigenes with the mean and N50 length of 671 bp and 1169 bp, respectively (Table S1). Among these, the lengths of 44,109 unigenes (65.53%) ranged from 200–500 bp (Figure 1). There were 10,596, 4897, 3335, 3003 and 1374 unigenes with lengths in the range of 500–1000 bp, 1000–1500 bp, 1500–2000 bp, 2000–3000 bp and >3000 bp, respectively.

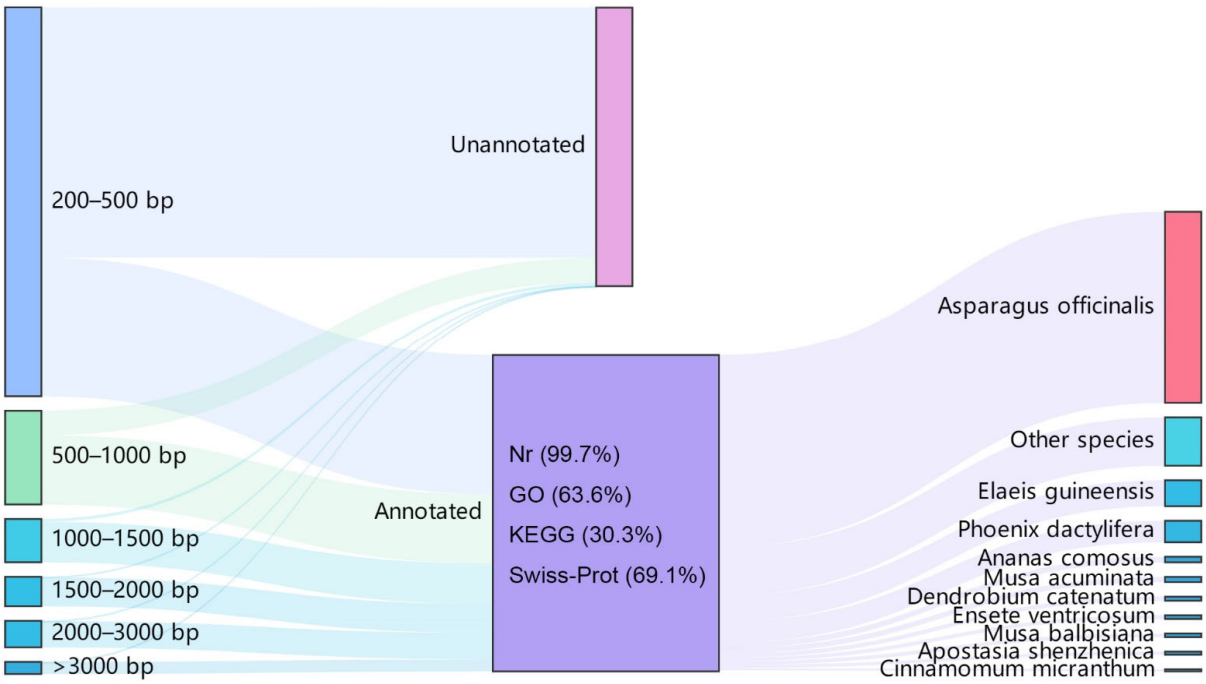


Figure 1. Sequence length distribution and function annotation of *A. angustifolia* unigenes and their closest homologs in the species of the Nr results.

3.2. Functional Annotation

All the unigenes of *A. angustifolia* transcriptome were utilized to search against four public databases for functional annotation (Table S1). A total of 35,798 unigenes were annotated in at least one public database. There were 35,673 and 24,751 unigenes having homologs in the Nr and Swiss-Prot databases, respectively. In the Nr database, the most closely related species is *Asparagus officinalis*, with orthologs to 21,727 (60.9%) *A. angustifolia* unigenes (Figure 1). Based on the GO annotation, 22,765 unigenes were assigned into seven subcategories (Figure 2). Among these, four subcategories were the most abundant, including 'intracellular' (8496 unigenes), 'binding' (6650 unigenes), 'metabolic process' (2014 unigenes) and 'cytoplasm' (1224 unigenes). Based on the KEGG, 10,841 unigenes were categorized into 18 pathways (Figure 2). The top four pathways were 'carbohydrate metabolism', 'translation', 'folding, sorting and degradation' and 'amino acid metabolism' with 3801, 3035, 2797 and 2015 unigenes, respectively.

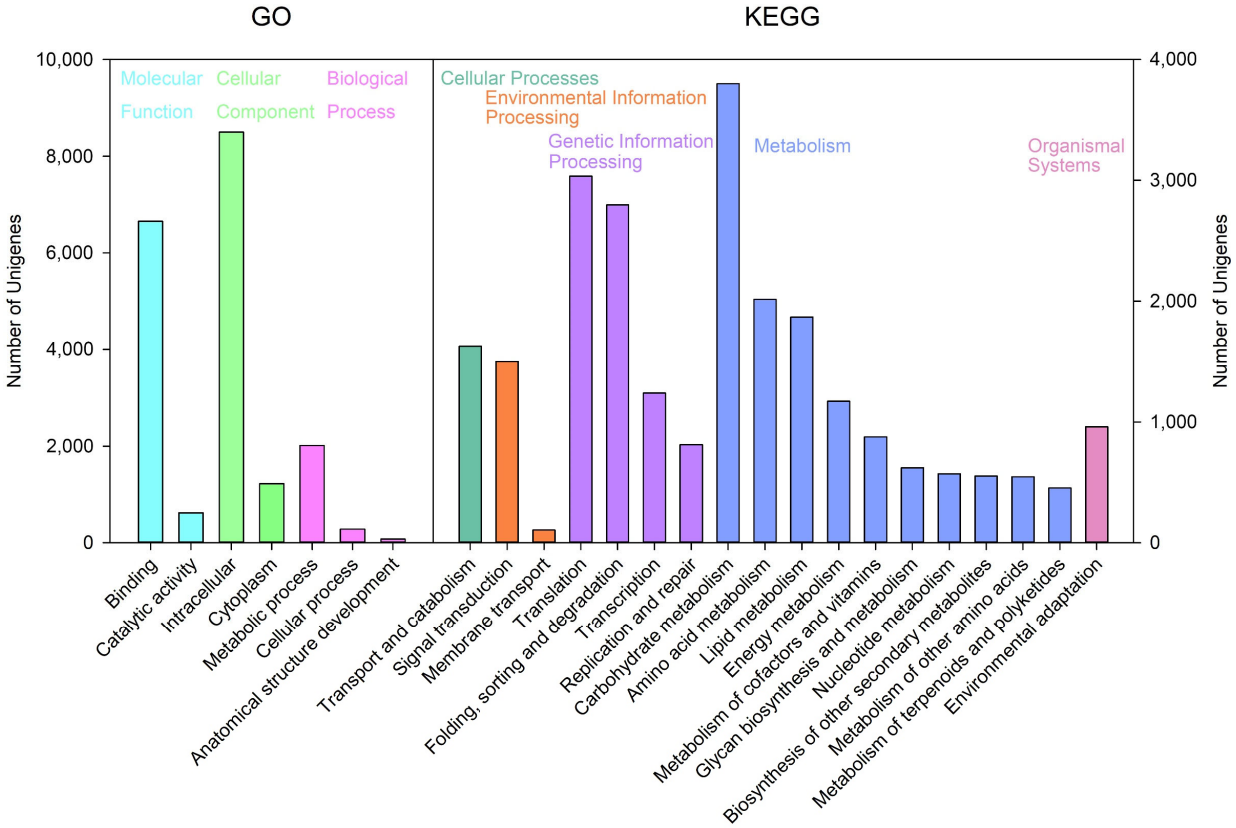


Figure 2. GO and KEGG classifications of the assembled unigenes in *Agave angustifolia*.

3.3. Identification of CAD Genes in Agaves

The CAD proteins from arabidopsis (9) and rice (12) were used as query sequences to search against the asparagus and amborella genomes, which identified five and two *CAD* genes, respectively (Table S2). These 28 CAD proteins were further utilized for sequence retrieval in the transcriptomes of five *Agave* species, including *A. deserti*, *A. tequilana*, *A. americana*, *A.* hybrid H11648 and *A. angustifolia*, with seven, seven, seven, six and six full-length *CAD* genes identified in these five species, respectively (Table S2). We did not find *CAD3* in *A.* hybrid H11648 or *CAD7* in *A. angustifolia*. Their sequence lengths ranged from 1148 to 2548 bp, with the lengths of predicted proteins ranging from 353 aa to 370 aa. The molecular weights and theoretical pI of these proteins were 38,263.10–39,952.18 Da and 5.54–7.15, respectively. Additionally, *CAD1* and *CAD5* genes in *Agave*

species were predicted with their subcellular localizations at the plasma membrane and the chloroplast, respectively. The other agave *CAD* genes, including *CAD2/3/4/6/7*, were predicted to be localized in the cytoplasm.

3.4. Phylogeny of CAD Proteins

The 61 CAD proteins of nine plant species were used for phylogenetic analysis and clustered into three groups according to a previous study (Figure 3) [19]. Group I was further divided into four subgroups. The two amborella proteins were clustered into Group I and II. Agave proteins were distributed in each group or subgroup, except IIIb and IIIc which contain only arabidopsis and rice sequences, respectively. Asparagus proteins spread in group I, II and IIId. In addition, group IIId contained monocot sequences only. All of these proteins and the agave proteins were further aligned to identify conserved domains, revealing 55 and 99 conserved amino acid residues, respectively (Figure S1).

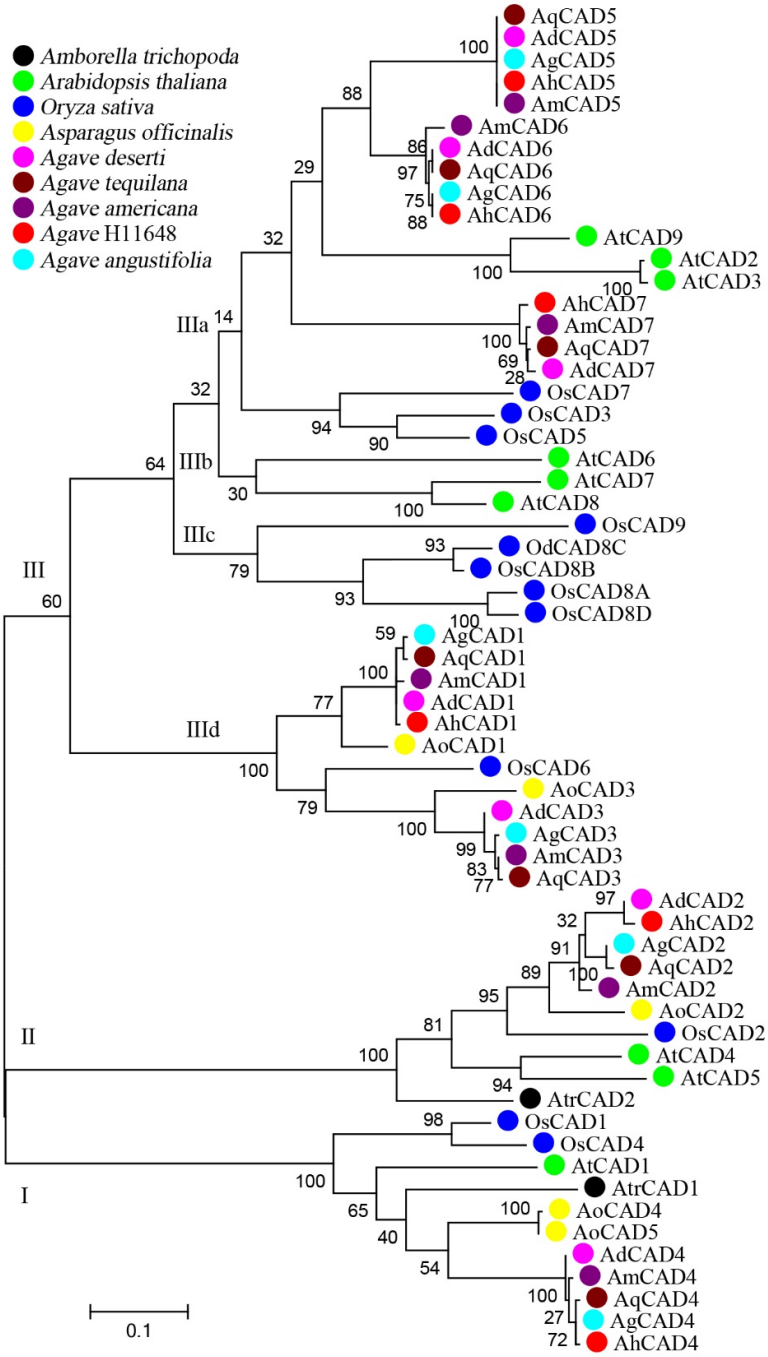


Figure 3. The maximum phylogenetic tree of CAD proteins in amborella (Atr, black), arabidopsis (At, green), rice (Os, blue), asparagus (Ao, yellow), *A. deserti* (Ad, pink), *A. tequilana* (Aq, crimson), *A. americana* (Am, purple), *A.* hybrid H11648 (Ah, red) and *A. angustifolia* (Ag, celeste).

3.5. Expression of Agave CAD Genes

The expression data of agave transcriptomes was selected for in silico expression analysis [2,11,33]. The result indicated the relatively high expression patterns of *CAD1/2/4/6* in all agave species (Figure 4A, Table S1). *CAD3* showed a moderate expression pattern in *A. deserti*, *A. tequilana* and *A. angustifolia* and no expression in *A. americana*

and *A*. H11648. *CAD5* was highly expressed in both *A*. H11648 and *A. angustifolia*, with extremely low or no expression in *A. deserti* and *A. americana*. The high expression of *CAD7* was only detected in *A. tequilana*. We further carried out the qRT-PCR validation (Figure 4B). The result revealed that *CAD1* showed high expression in *A. deserti* and *A.* hybrid H11648, and moderate expression in *A. angustifolia* and *A. americana*. *CAD2* was highly expressed in *A. angustifolia* and moderately expressed *in A. americana*. There were low levels of expression of *CAD3/7* in all of the species. *CAD4* and *CAD6* showed moderate expression in *A. angustifolia* and *A*. hybrid H11648, respectively. *CAD5* had high expression in *A. angustifolia* and moderate expression in *A*. hybrid H11648. We also conducted a correlation analysis between the results of RNA-Seq and qRT-PCR, showing a significant ($p < 0.01$) correlation (Figure 4C).

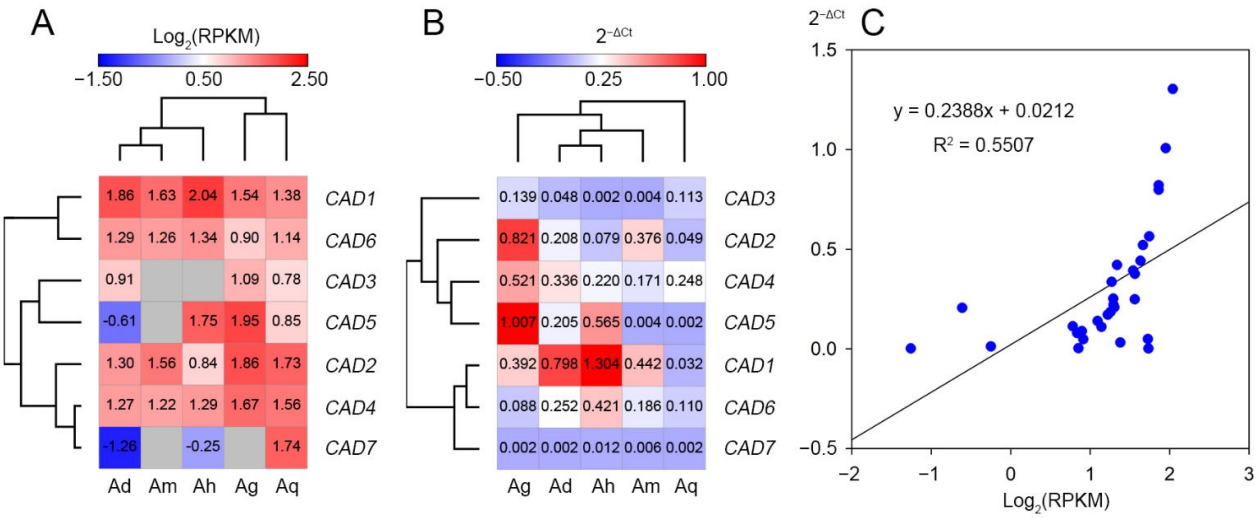


Figure 4. Expression analysis of agave *CAD* genes by RNA-Seq (**A**), qRT-PCR (**B**) and correlation analysis of the two methods (**C**). Grey squares without numbers represent no data obtained by RNA-Seq. The clusters of genes and species were calculated by the hierarchical clustering method in A and B [46].

4. Discussion

4.1. Features of A. angustifolia Transcriptome

The fast development of sequencing technology has provided an efficient tool for studying gene functions in plants without genome sequencing data, especially such as agave species with extremely large genomes [6,15]. So far, few gene sequences of *A. angustifolia* have been submitted to GenBank, leaving a knowledge gap regarding the genetic basis of important traits in this agave species. Hence, we assembled the transcriptome with 67,314 unigenes, which will be a useful resource for studying gene functions in *A. angustifolia*. In this new assembly, the number of unigenes was similar to that in *A. americana* but less than those in *A. deserti*, *A. tequilana* and *A.* hybrid H11648 [2,11,34]. This inconsistency might be caused by the difference in transcriptome assembly method and sample collection [47]. In general, there are about 18% of the unigenes with sequence lengths over 1000 bp (Table S1). In addition, it seems the longer the sequence length, the higher the chance of identifying homologs in public databases (Figure 1). This might be explained by the existence of a large number of non-coding RNAs in *A. angustifolia* transcriptome [48]. Additionally, more than half of the *A. angustifolia* unigenes have homologs in public databases, facilitating the prediction of agave gene functions based on sequence homology [11]. Various genes related to fiber traits have been identified in agave species, including *CesA*, *SAUR* and *NPF* families [11–13]. Moreover, our comparative analysis

revealed the close phylogenetic relationship between *A. angustifolia* and asparagus, both belonging to the Asparagaceae family [49].

4.2. Candidate CAD Genes for Lignin Biosynthesis in Agaves

Lignin is a main component to form the cell wall structure in plants, and also an important secondary metabolite derived from the phenylpropanoid pathway [50]. As the last enzyme in lignin biosynthesis, the manipulation of *CAD* genes has been successfully applied in the improvement of the fiber traits in flax [21]. Thus, we carried out the sequence retrieval of *CAD* genes in asparagus and five agave species. As a result, we identified five, seven, seven, seven, six and six *CAD* genes in asparagus, *A. deserti*, *A. tequilana*, *A. americana*, *A.* hybrid H11648 and *A. angustifolia*, respectively (Table S1). It seems that the *CAD* gene family size in asparagus and agave is smaller than that in arabidopsis and rice. Our phylogenetic analysis revealed diverse evolution patterns of *CAD* genes, including cross-species conservation (group I and II) and lineage-specific expansion in arabidopsis (group IIIb), rice (group IIIc) and monocots (group IIId), respectively (Figure 3). Additionally, our RNA-Seq analysis did not detect the expression of *CAD3* in *A.* hybrid H11648 and *CAD7* in *A. angustifolia*. However, we can assume the existence of *CAD3*/*7* in the genomes of *A.* hybrid H11648 and *A. angustifolia* since the two genes exist in both wild and domesticated agave species [10]. This discrepancy might be caused by their low expressions detected in the leaf samples, which is also validated by the qRT-PCR result (Figure 4). It might be related to the tissue-specific expression of *CAD* genes [51]. Moreover, the extremely low expression of *CAD3* in *A.* hybrid H11648 might also be related to gene expression dominance when compared with one of its parents, *A. angustifolia* [52]. Our expression analysis revealed the conservatively expressed *CAD* genes (*CAD1*/*2*/*4*/*6*) in agave species, which could be considered as candidate targets for manipulation to improve lignin traits [19]. The species-specific expression of *CAD3*/*5*/*7* indicates the existence of a different regulatory mechanism in agave species, including transcription regulations [10,11].

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/article/10.3390/agriculture12071003/s1, Figure S1: Alignment of CAD proteins in the nine species (A) and in agave species (B). Conserved amino acid residues are highlighted in black color.; Table S1: Details and annotation information (in Nr, GO, KEGG and Swiss-Prot databases) of unigenes.; Table S2: The *CAD* gene family in asparagus and agave.

Author Contributions: Conceptualization, X.H., X.Y. and K.Y.; formal analysis, X.H.; investigation, B.X., S.T., Y.H., X.Q. and H.C.; resources, J.X. and T.C.; writing—original draft preparation, X.H.; writing—review and editing, X.Y. and K.Y.; supervision, K.Y.; funding acquisition, X.H. and K.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (32001598), Hainan Provincial Natural Science Foundation of China (322MS112), China Agriculture Research System of MOF and MARA (CARS-16), Central Public-interest Scientific Institution Basal Research Fund (1630042022005) and the innovation platform for Academicians of Hainan Province.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are contained within the article or Supplementary Materials.

Acknowledgments: We would like to thank Bo Wang from Huazhong Agricultural University for his thorough suggestions on experiment design.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Good-Avila, S.V.; Souza, V.; Gaut, B.S.; Eguiarte, L.E. Timing and rate of speciation in *Agave* (Agavaceae). *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9124–9129.
2. Gross, S.M.; Martin, J.A.; Simpson, J.; Abraham-Juarez, M.J.; Wang, Z.; Visel, A. *De novo* transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genom.* **2013**, *14*, 563.

3. Yang, X.; Cushman, J.C.; Borland, A.M.; Edwards, E.; Wullschleger, S.D.; Tuskan, G.A.; Owen, N.A.; Griffiths, H.; Smith, J.A.; De Paoli, H.C.; et al. A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytol*. **2015**, *207*, 491–504.
4. García-Mendoza, A.; Chiang, F. The confusion of *Agave vivipara* L. and *A. angustifolia* Haw., two distinct taxa. *Brittonia* **2003**, *55*, 82–87.
5. Gutiérrez-Coronado, M.; Acedo-Félix, E.; Valenzuela-Quintanar, A. Bacanora industry and its process of production. *Cienc. Tecnol. Aliment. Food* **2007**, *5*, 394–404.
6. Robert, M.L.; Lim, K.Y.; Hanson, L.; Sanchez-Teyer, F.; Bennett, M.D.; Leitch, A.R.; Leitch, I.J. Wild and agronomically important *Agave* species (Asparagaceae) show proportional increases in chromosome number, genome size, and genetic markers with increasing ploidy. *Bot. J. Linn. Soc.* **2008**, *158*, 215–222.
7. Rosli, N.A.; Ahmad, I.; Abdullah, I. Isolation and characterization of cellulose nanocrystals from *Agave angustifolia* fibre. *Bioresources* **2013**, *8*, 1893–1908.
8. Jin, G.; Huang, X.; Chen, T.; Qin, X.; Xi, J.; Yi, K. The complete chloroplast genome of agave hybrid 11648. *Mitochondrial DNA B Resour*. **2020**, *5*, 2345–2346.
9. Raya, F.T.; Marone, M.P.; Carvalho, L.M.; Rabelo, S.C.; Paula, M.S.; Campanari, M.F.Z.; Freschi, L.; Mayer, J.L.S.; Silva, O.R.R.F.; Mieczkowski, P.; et al. Extreme physiology: Biomass and transcriptional profiling of three abandoned *Agave* cultivars. *Ind. Crops Prod.* **2021**, *172*, 114043.
10. Huang, X.; Wang, B.; Xi, J.; Zhang, Y.; He, C.; Zheng, J.; Gao, J.; Chen, H.; Zhang, S.; Wu, W.; et al. Transcriptome comparison reveals distinct selection patterns in domesticated and wild *Agave* species, the important CAM plants. *Int. J. Genom.* **2018**, *2018*, 5716518.
11. Huang, X.; Xiao, M.; Xi, J.; He, C.; Zheng, J.; Chen, H.; Gao, J.; Zhang, S.; Wu, W.; Liang, Y.; et al. De novo transcriptome assembly of *Agave* H11648 by Illumina sequencing and identification of cellulose synthase genes in *Agave* species. *Genes* **2019**, *10*, 103.
12. Deng, G.; Huang, X.; Xie, L.; Tan, S.; Gbokie, T., Jr.; Bao, Y.; Xie, Z.; Yi, K. Identification and Expression of *SAUR* Genes in the CAM Plant Agave. *Genes* **2019**, *10*, 555.
13. Tan, S.; Liang, Y.; Huang, Y.; Xi, J.; Huang, X.; Yang, X.; Yi, K. Phylogeny and Expression Atlas of the *NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER FAMILY* in Agave. *Plants* **2022**, *11*, 1434.
14. Maceda-López, L.F.; Góngora-Castillo, E.B.; Ibarra-Laclette, E.; Morán-Velázquez, D.C.; Girón Ramírez, A.; Bourdon, M.; Villalpando-Aguilar, J.L.; Toomer, G.; Tang, J.Z.; Azadi, P.; et al. Transcriptome Mining Provides Insights into Cell Wall Metabolism and Fiber Lignification in *Agave tequilana* Weber. *Plants* **2022**, *11*, 1496.
15. Schuster, S.C. Next-generation sequencing transforms today's biology. *Nat. Methods* **2008**, *5*, 16–18.
16. Weng, J.K.; Chapple, C. The origin and evolution of lignin biosynthesis. *New Phytol.* **2010**, *187*, 273–285.
17. Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W. Lignin biosynthesis and structure. *Plant Physiol.* **2010**, *153*, 895–905.
18. Qi, K.; Song, X.; Yuan, Y.; Bao, J.; Gong, X.; Huang, X.; Khanizadeh, S.; Zhang, S.; Tao, S. *CAD* Genes: Genome-Wide Identification, Evolution, and Their Contribution to Lignin Biosynthesis in Pear (*Pyrus bretschneideri*). *Plants* **2021**, *10*, 1444.
19. Preisner, M.; Wojtasik, W.; Kostyn, K.; Boba, A.; Czuj, T.; Szopa, J.; Kulma, A. The cinnamyl alcohol dehydrogenase family in flax: Differentiation during plant growth and under stress conditions. *J. Plant Physiol.* **2018**, *221*, 132–143.
20. Vanholme, R.; Morreel, K.; Darrah, C.; Oyarce, P.; Grabber, J.H.; Ralph, J.; Boerjan, W. Metabolic engineering of novel lignin in biomass crops. *New Phytol*. **2012**, *196*, 978–1000.
21. Chantreau, M.; Grec, S.; Gutierrez, L.; Dalmais, M.; Pineau, C.; Demailly, H.; Paysant-Leroux, C.; Tavernier, R.; Trouvé, J.P.; Chatterjee, M.; et al. PT-Flax (phenotyping and TILLinG of flax): Development of a flax (*Linum usitatissimum* L.) mutant population and TILLinG platform for forward and reverse genetics. *BMC Plant Biol*. **2013**, *13*, 159.
22. Huang, X.; Chen, J.; Bao, Y.; Liu, L.; Jiang, H.; An, X.; Dai, L.; Wang, B.; Peng, D. Transcript profiling reveals auxin and cytokinin signaling pathways and transcription regulation during in vitro organogenesis of ramie (*Boehmeria nivea* L. Gaud). *PLoS ONE* **2014**, *9*, e113768.
23. Katz, K.; Shutov, O.; Lapoint, R.; Kimelman, M.; Brister, J.R.; O'Sullivan, C. The Sequence Read Archive: A decade more of explosive growth. *Nucleic Acids Res.* **2022**, *50*, D387–D390.
24. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120.
25. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. **2011**, *17*, 10–12.
26. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. *Nat. Genet*. **2000**, *25*, 25–29.
27. Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* **1999**, *27*, 49–54.
28. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotech.* **2011**, *29*, 644–652.
29. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. **2000**, *28*, 27–30.
30. Pruitt, K.D.; Tatusova, T.; Maglott, D.R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. **2005**, *33*, D501–D504.
31. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.

32. Kim, S.J.; Kim, M.R.; Bedgar, D.L.; Moinuddin, S.G.; Cardenas, C.L.; Davin, L.B.; Kang, C.; Lewis, N.G. Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1455–1460.
33. Park, H.L.; Kim, T.L.; Bhoo, S.H.; Lee, T.H.; Lee, S.W.; Cho, M.H. Biochemical Characterization of the Rice Cinnamyl Alcohol Dehydrogenase Gene Family. *Molecules* **2018**, *23*, 2659.
34. Abraham, P.E.; Yin, H.; Borland, A.M.; Weighill, D.; Lim, S.D.; De Paoli, H.C.; Engle, N.; Jones, P.C.; Agh, R.; Weston, D.J.; et al. Transcript, protein and metabolite temporal dynamics in the CAM plant *Agave*. *Nat. Plants* **2016**, *2*, 16178.
35. Harkess, A.; Zhou, J.; Xu, C.; Bowers, J.E.; Van der Hulst, R.; Ayyampalayam, S.; Mercati, F.; Riccardi, P.; McKain, M.R.; Kakrana, A.; et al. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.* **2017**, *8*, 1279.
36. Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. *Science* **2013**, *342*, 1241089.
37. Rombel, I.T.; Sykes, K.F.; Rayner, S.; Johnston, S.A. ORF-FINDER: A vector for high-throughput gene identification. *Gene* **2002**, *282*, 33–41.
38. ProtParam Tool. Available online: web.expasy.org/protparam/ (accessed on 2 June 2022).
39. Yu, C.S.; Chen, Y.C.; Lu, C.H.; Hwang, J.K. Prediction of protein subcellular localization. *Proteins* **2006**, *64*, 643–651.
40. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol*. **2016**, *33*, 1870–1874.
41. DNAMAN7. Available online: www.lynnon.com (accessed on 2 June 2022).
42. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **2011**, *12*, 323.
43. Mortazavi, A.; Williams, B.A.; Mccue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628.
44. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115.
45. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408.
46. Zolfaghari, F.; Khosravi, H.; Shahriyari, A.; Jabbari, M.; Abolhasani, A. Hierarchical cluster analysis to identify the homogeneous desertification management units. *PLoS ONE* **2019**, *14*, e0226355.
47. Sadamoto, H.; Takahashi, H.; Okada, T.; Kenmoku, H.; Toyota, M.; Asakawa, Y. *De novo* sequencing and transcriptome analysis of the central nervous system of mollusc *Lymnaea stagnalis* by deep RNA sequencing. *PLoS ONE* **2012**, *7*, e42546.
48. Waititu, J.K.; Zhang, C.; Liu, J.; Wang, H. Plant Non-Coding RNAs: Origin, Biogenesis, Mode of Action and Their Roles in Abiotic Stress. *Int. J. Mol. Sci*. **2020**, *21*, 8401.
49. Chase, M.W.; Reveal, J.L.; Fay, M.F. A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. *Bot. J. Linn. Soc.* **2009**, *161*, 132–136.
50. Liu, Q.; Luo, L.; Zheng, L. Lignins: Biosynthesis and Biological Functions in Plants. *Int. J. Mol. Sci.* **2018**, *19*, 335.
51. Shang, H.; Zhang, N.; Xie, Z.; Deng, S.; Yi, L.; Huang, X. Genome-Wide Identification and Expression of the PIN Auxin Efflux Carrier Gene Family in Watermelon (*Citrullus lanatus*). *Agriculture* **2021**, *11*, 447.
52. Bottani, S.; Zabet, N.R.; Wendel, J.F.; Veitia, R.A. Gene Expression Dominance in Allopolyploids: Hypotheses and Models. *Trends Plant Sci*. **2018**, *23*, 393–402.