Article

Transcriptome Mining Provides Insights into Cell Wall Metabolism and Fiber Lignification in *Agave tequilana* Weber

Luis F. Maceda-López [1], Elsa B. Góngora-Castillo [2], Enrique Ibarra-Laclette [3], Dalia C. Morán-Velázquez [1], Amaranta Girón Ramírez [4], Matthieu Bourdon [5], José L. Villalpando-Aguilar [1], Gabriela Toomer [6], John Z. Tang [7], Parastoo Azadi [7], Jorge M. Santamaría [4], Itzel López-Rosas [8], Mercedes G. López [9], June Simpson [10] and Fulgencio Alatorre-Cobos [8,*]

1 Colegio de Postgraduados, Campus Campeche, Carretera Haltunchén-Edzná km 17.5, Sihochac, Campeche 24450, Mexico; maceda.luis@outlook.com (L.F.M.-L.); carolina.moran21@hotmail.com (D.C.M.-V.); jose.va@china.tecnm.mx (J.L.V.-A.)
2 CONACYT-Centro de Investigación Científica de Yucatán, Unidad de Biotecnología, Calle 43 No. 130 × 32 y 34, Chuburná de Hidalgo, Mérida 97205, Mexico; elsa.gongora@cicy.mx
3 Red de Estudios Moleculares Avanzados, Instituto de Ecología A. C. Carretera Antigua a Coatepec 351, El Haya, Xalapa 91070, Mexico; enrique.ibarra@inecol.mx
4 Centro de Investigación Científica de Yucatán, Unidad de Biotecnología, Calle 43 No. 130 × 32 y 34, Chuburná de Hidalgo, Mérida 97205, Mexico; amaranta.giron@cicy.mx (A.G.R.); jorgesm@cicy.mx (J.M.S.)
5 Sainsbury Laboratory, University of Cambridge, Cambridge CB2 1LR, UK; matthieu.bourdon@slcu.cam.ac.uk
6 Division of Microbiology and Molecular Biology, IIT Research Institute, Chicago, IL 60616, USA; gtoomer@iitri.org
7 Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA; johnzhoutang@gmail.com (J.Z.T.); azadi@ccrc.uga.edu (P.A.)
8 CONACYT-Colegio de Postgraduados Campus Campeche, Carretera Haltunchén-Edzná km 17.5, Sihochac, Campeche 24450, Mexico; itzel.rosas@colpos.mx
9 Departamento de Biotecnología y Bioquímica, Centro de Investigación y Estudios Avanzados del IPN, Irapuato 36824, Mexico; mercedes.lopez@cinvestav.mx
10 Departamento de Ingeniería Genetica, Centro de Investigación y Estudios Avanzados del IPN, Irapuato 36824, Mexico; june.simpson@cinvestav.mx
* Correspondence: fulgencio@colpos.mx

Abstract: Resilience of growing in arid and semiarid regions and a high capacity of accumulating sugar-rich biomass with low lignin percentages have placed Agave species as an emerging bioenergy crop. Although transcriptome sequencing of fiber-producing agave species has been explored, molecular bases that control wall cell biogenesis and metabolism in agave species are still poorly understood. Here, through RNAseq data mining, we reconstructed the cellulose biosynthesis pathway and the phenylpropanoid route producing lignin monomers in *A. tequilana*, and evaluated their expression patterns in silico and experimentally. Most of the orthologs retrieved showed differential expression levels when they were analyzed in different tissues with contrasting cellulose and lignin accumulation. Phylogenetic and structural motif analyses of putative CESA and CAD proteins allowed to identify those potentially involved with secondary cell wall formation. RT-qPCR assays revealed enhanced expression levels of AtqCAD5 and AtqCESA7 in parenchyma cells associated with extraxylary fibers, suggesting a mechanism of formation of sclerenchyma fibers in Agave similar to that reported for xylem cells in model eudicots. Overall, our results provide a framework for understanding molecular bases underlying cell wall biogenesis in Agave species studying mechanisms involving in leaf fiber development in monocots.

Keywords: agave; cell walls; lignocellulose; CAD protein; CESA protein; sclerenchyma

1. Introduction

The cell wall is the boundary between the plant cell and the environment. Its composition affects cell growth, expansion, proliferation, adhesion, and responses to external

stimuli [1]. Because the cell wall skeleton is primarily based on fermentable polysaccharides, cell walls are now highly valorized as a sugar source for bioethanol production and other bio-products obtained by biorefinery [2–4]. Cellulose, the most abundant polymer on earth, is the major structural polysaccharide of plant primary and secondary cell walls (PCW and SCW, respectively). This compound consists de linear chains of several hundred to many thousands of β (1, 4) linked D-glucose units, which can be defragmented and used for saccharification. Cellulose is up 40–50% and close to 90% in SCW in specialized cells as mature fibers, respectively [5]. However, here, cellulose is found tightly impregnated by lignin, a complex polyphenolic biopolymer that increases the stiffness and strength of SCW but complicates access to cellulose [6]. Understanding lignin and cellulose biosynthesis pathways during SCW formation is critical to dealing with an increasing demand of low-lignin biomass more suitable for producing of second-generation bioethanol from plant biomass.

In Agave species used for spirits production, only pines are used for fermentation, while leaves that account for 40% of total wet weight are considered waste. Agave species are also emerging as biofuel crops because of their drought resilience, adaptation to poor soil, high biomass yield, and notably high cellulose/lignin ratio [7–10]. Recently, biofuel potential has been evaluated in *A. americana* and *A. deserti*, two species adapted to marginal soils. The results showed biomass with low K-lignin content and enriched with soluble and structural sugars [11], which could be helpful for the bioethanol industry.

Reference genomes have not been reported so far for any Agaves species. However, high-throughput RNA-sequencing has revealed transcriptomic landscapes for several Agave species during responses to environmental stimuli and development. These catalogs of expressed genes include transcriptional profiles of *A. tequilana* Weber, *A. deserti* Engelm, *A. victoriae-reginae* T. Moore, *A. striata* Zucc., *A. americana* L., *A. salmiana* Otto, *A. sisalana* Perr. ex. Engelm, *A. fourcroydes* Lem., and the hybrid H11648 ((*A. amaniensis* Trel and Nowell x *A. angustifolia* Haw) x *A. amaniensis*) [9,12–19]. Because of the economic importance tequila production represents, analysis of RNA-seq data mainly has focused on *A. tequilana* to uncover and characterize sets of genes involved in the fructan biosynthesis and transition to flowering [12,15,16]. Moreover, two Agave species also traditionally used in Mexico since the Precolombin ages, *A. americana* and *A. salmiana*, have been used as models to provide new insights into the molecular mechanisms underlying the CAM metabolism and regeneration in plants, respectively [13,14].

In contrast to eudicots, biosynthetic genes and regulatory network hubs that control the metabolism and deposition of cell wall components have been poorly characterized in monocots. Using comparative genomics and expression profiling analysis, large sets of genes putatively involved in cell wall biogenesis, especially for the components cellulose and lignin, have been reported for maize and rice, but the functional significance of most of those is missing still [20–25]. Recently, encouraged by a world economy that is moving back to natural or bio-based fibers, leaf fibers from Agave species have drawn the attention again of industry and science. Transcriptomes of *A. sisalana*, *A. fourcroydes*, and the hybrid H11648, widely cultivated in Brazil, Africa, and China for fiber production, have been explored to know fiber-related genes subjected to positive selection during Agave domestication, those involved with cellulose biosynthesis, and responses to drought [15,17–19]. Here, we carried out mining of publicly available RNA-seq data from *A. tequilana* to predict the closest orthologous for the sets of genes involved with cellulose and lignin biosynthesis pathways. Most of the retrieved orthologous genes showed differential expression levels when analyzed in eight transcriptomes from different tissues of *A. tequilana*. Gene expression patterns found in enriched lignocellulosic tissues are in line with the predicted functions for key genes in metabolic pathways of cellulose and lignin. Phylogenetic and structural motif analysis for putative CESA and CAD proteins allowed identifying those potentially involved with SCW formation. RT-qPCR assays revealed enhanced expression levels of AtqCAD5 and AtqCESA7 quantified in parenchymatous cells associated with extraxylary fibers, supporting the accuracy of our orthology analyses and suggesting a

mechanism of formation of sclerenchyma fibers in Agave similar to that reported for xylem cells in model eudicots.

2. Results

2.1. Identification of Orthologous Biosynthetic Genes for Cellulose and Lignin in A. tequilana

Using the orthologous groups obtained by Markov clustering algorithm and a similarity search with Arabidopsis proteins as queries, the closest orthologous genes encoding for all the enzymes involved in the cellulose and lignin pathways were identified in the *A. tequilana* transcriptomes mined (File S1). Gene nomenclature designated for each Agave orthologous was based on gene prefixes previously used (Atq) [12,16,26,27]. For cellulose biosynthesis, all the nine gene families previously annotated and characterized in *A. thaliana*, the best plant model to study cell wall-related genes [28,29], were found expressed in the Agave transcriptomes mined (Table 1). Cellulose Synthase (CESA) and Fructose Kinase (FK) families had the highest number of orthologous genes identified. Ten CESA genes and seven FK genes have been reported in the Arabidopsis genome, while here, we retrieved six and five genes expressed, respectively. Although this study just reflects transcriptomic landscapes, only one Agave orthologous gene was found for *Sucrose Phosphate Synthase* (*SPS*), contrary to those four reported in Arabidopsis. Size, calculated molecular weight, and theoretical isoelectric points were relatively similar within all the gene families analyzed (Table 1), except those for *Fructokinase* (*FK*) and *UDP-Glucose Pyrophosphorylase* (*UGP*) genes. FK1 is 191 aa smaller compared to FK4, while UGP1 has 380 aa fewer with respect to UGP3 (Table 1). For these genes, incomplete open reading frames (ORF) cannot be ruled out. Subcellular localization prediction for most enzymes was consistent with previously reported for carbohydrate metabolism: cytosolic, plastids, or associated to the membrane or cell wall (Table 1) [29–32]. For example, as expected for plant CESA proteins, all Agave CESA isoforms found in this study were predicted to be targeted to the plasma membrane (Table 1).

Table 1. Detailed sequence information of genes involved in cellulose biosynthesis in *Agave tequilana*.

Gene	ORF (nt)	Protein (aa)	Calculated Mw (Da)	Theoretical pI	Predicted Subcellular Location and Score *
AtqSPS1	3168	1055	118,083.33	5.77	nucl: 5, cyto: 4, chlo: 3
AtqSPS2	3252	1083	120,102.55	6.11	nucl: 5, cyto: 5, chlo: 2
AtqSPP1	1275	424	47,733.35	5.54	cyto: 5, nucl: 4, chlo: 2
AtqSPP2	1275	424	48,024.71	6.00	chlo: 7, nucl: 3, mito: 3
AtqFK1	651	216	23,897.62	5.62	chlo: 6, cyto: 5, cysk_nucl: 2
AtqFK2	969	322	34,489.54	5.20	cyto: 8, chlo: 3, pero: 2
AtqFK3	996	331	35,476.84	5.03	pero: 11, chlo: 1, nucl: 1
AtqFK4	1224	407	43,435.34	6.13	chlo: 14
AtqFK5	1107	368	39,509.00	5.50	cysk: 9, cyto: 3, chlo: 1
AtqGPI1	1521	506	55,292.25	4.97	cyto: 9, chlo: 3, mito: 1
AtqPGM1	1872	623	68,121.76	6.82	chlo: 14
AtqPGM2	1749	582	63,450.95	6.02	cyto: 11, chlo: 2, nucl: 1
AtqSUS1	2532	843	96,280.19	8.07	cyto: 12, chlo: 1, plas: 1
AtqSUS2	2448	815	92,525.91	6.03	cyto: 9, chlo: 2, nucl: 1
AtqSUS3	2424	807	92,436.65	5.83	cyto: 5, mito: 4, chlo: 3
AtqHXK1	1488	495	53,179.08	6.07	chlo: 13, mito: 1
AtqHXK2	1497	498	54,454.50	5.87	chlo: 11, nucl: 1, extr: 1
AtqUGP1	1422	473	51,920.96	6.09	cyto: 11, chlo: 2, E.R.: 1
AtqUGP2	1575	524	58,194.63	7.74	chlo: 7, mito: 6, nucl: 1
AtqUGP3	2262	853	93,824.45	6.52	chlo: 10, mito: 2, nucl: 1.5
AtqCESA1	3264	1087	122,300.01	6.46	plas: 13, E.R.: 1
AtqCESA3	3180	1059	119,098.30	8.14	plas: 10, nucl: 2, mito: 1

Table 1. *Cont.*

Gene	ORF (nt)	Protein (aa)	Calculated Mw (Da)	Theoretical pI	Predicted Subcellular Location and Score *
AtqCESA4	3174	1057	120,445.31	6.90	plas: 11, nucl: 2, E.R.: 1
AtqCESA5	3273	1090	122,477.03	6.85	plas: 13, E.R.: 1
AtqCESA7	3153	1058	118,824.62	6.34	plas: 12, nucl: 1, E.R.: 1
AtqCESA8	3063	1020	114,790.86	6.57	plas: 12, nucl: 1, vacu: 1

Sucrose Phosphate Phosphatase (SPP), Sucrose Phosphate Synthase (SPS), Fructokinase (FK), Glucose-6-PhosphateIisomerase (GPI), Phosphoglucomutase (PGM), Sucrose Synthase (SUS), Hexokinase (HXK), UDP-Glucose Pyrophosphorylase (UGP), and Cellulose Synthase (CESA). Molecular weight (Mw), pI: Isoelectric point (pI). Chloroplast or thylakoid lumen (clo), cytoskeleton (cysk), cytosol (cyto), endoplasmic reticulum (E.R.), extracellular or cell wall (extra), mitochondrion (mito), nucleus (nucl), peroxisome (pero), plasma membrane (plas), vacuolar membrane (vacu). Localization names joining two abbreviations indicate dual localization. * The score means the number of nearest neighbors with the same subcellular localization indicated. Only the three highest scores are shown.

Similar to cellulose-related genes, orthologous Agave genes were found for all eleven genes families encoding for enzymes that participate in the biosynthesis of lignin subunits (Table 2). The gene family with more members expressed in the transcriptomes mined was *Cinnamyl Alcohol Dehydrogenase* (*CAD*). Nine *CAD* genes have been functionally characterized in Arabidopsis. Here, seven orthologous *CAD* genes were retrieved in the leaves' assembled transcriptomes. For the rest of the lignin biosynthetic genes, gene family size was quite similar to that reported in Arabidopsis, except for *Shikimate Hydroxycinnamoyl Transferase* (*HCT*), *Cinnamoyl CoA Reductase* (*CRR*), and *Caffeic Acid 3-O-Methyltransferase* (*COMT*). Interestingly, in *A. tequilana*, more isoforms expressed were found for *HCT*, *CCR*, and *COMT*, respecting those reported in Arabidopsis (Table 2). Predicted protein size and physic-chemical properties were similar among the homologous genes involved with monolignol biosynthesis (Table 2).

2.2. Agave Cellulose and Lignin-Related Predicted Genes Are Differentially Expressed in Different Agave Tissues

Analysis of mRNA abundance for all the biosynthetic orthologous predicted for lignin and cellulose was conducted in Agave transcriptomes previously reported [16]. These RNA-seq data are from several Agave tissues and organs in different developmental stages, and thus, represent a valuable dataset to assess the potential biological roles of the genes identified in this study. For cellulose biosynthesis, all the genes' isoforms showed differential expression patterns in the eight transcriptomes analyzed that were obtained from tissues highly differentiated as the root, pine, and leaves, or undifferentiated cells of the shoot apical meristem (SAM), for example. Heatmap clustering analysis showed that all the genes encoding enzymes that participate in getting sucrose (Suc) have differential expression levels in the most transcriptomes sequenced (Figure 1, Supplementary Table S3). *Sucrose Phosphate Synthase2* (*AtqSPS2*) and *Sucrose Phosphate Phosphatase1* (*AtqSPP1*) showed medium to high expression in pine and reproductive tissues, while *Fructokinase3* (*AtqFK3*) had similar levels for root and leaves. *Phosphoglucomutase2* (*AtqPGM2*), with a cytosolic localization predicted, showed a broad expression in all the transcriptomes analyzed (Figure 1; Table S1), which is in line with the biological role reported for cytosolic PGM. Plant cytosolic PGM isoforms are involved in Suc catabolism to provide intermediates used in glycolysis and as substrates for synthesizing a variety of cellular constituents [33,34]; thus, it is expected PGMs are ubiquitous in all plant tissues. Among the *three Sucrose Synthase* (*SUS*) isoforms retrieved in our orthology analysis, *AtqSUS2* was the most expressed in all the Agave transcriptomes mined. It was moderately expressed in roots, pines, and leaves and interestingly, highly expressed in reproductive tissues (Figure 1; Table S3). Likewise, *Pgm1, UDP-Glucose Pyrophosphorylase1* (*UGP1*), and four *Cellulose Synthase* (*CESA*) isoforms, *AtqCESA1*, *AtqCESA3*, *AtqCESA4*, and *AtqCESA8*, showed a medium expression, fairly uniform, in all the tissues tested (Figure 1; Table S3). Contrary to most of the *CESAs*

retrieved, *AtqCESA5* and *AtqCESA7* were found to be expressed only in roots and leaves, which are samples enriched with lignocellulosic cell walls [10,35], suggesting that these two *CESA* orthologous may have a functional role during secondary cell wall biosynthesis in Agave plants.

Table 2. Detailed sequence information of genes involved in lignin biosynthesis in *Agave tequilana*.

Gene	ORF (nt)	Protein Size (aa)	Calculated Mw (Da)	Theoretical pI	Predicted Subcellular Location and Score *
AtqPAL1	2103	700	75,908.60	5.60	chlo: 4, plas: 3, cyto: 2
AtqPAL2	2112	703	75,997.81	5.92	chlo: 4, plas: 3, cyto: 2
AtqC4H	1518	505	57,804.27	9.19	plas: 10, E.R.: 3, vacu: 1
Atq4CL	1689	562	59,781.88	5.52	E.R.: 5.5, E.R._plas: 3.5, cyto: 3
AtqCSE	990	329	36,767.97	6.09	nucl: 4, cyto: 4, mito: 2,
AtqHCT1	1326	441	48,196.19	6.22	cyto: 7, chlo: 4, nucl: 2
AtqHCT2	1302	433	47,684.51	6.02	cyto: 10, chlo: 3, mito: 1
AtqC3H	1539	512	58,118.34	7.23	chlo: 13, mito: 1
AtqCCoAOMT	771	256	28,666.75	5.30	cyto: 7, chlo: 3, cysk: 2
AtqCCR1	981	326	35,680.92	6.18	chlo: 6, plas: 3, vacu: 2
AtqCCR2	1017	338	36,883.26	5.73	chlo: 6, plas: 4, vacu: 2
AtqF5H	1545	514	57,197.12	6.81	chlo: 5, plas: 4, nucl: 2.5
AtqCOMT1	1092	363	39,707.87	5.58	cyto: 9, chlo: 2, cysk: 2
AtqCOMT2	1095	364	39,372.20	4.76	cyto: 10, cysk: 2, chlo: 1
AtqCAD1	1062	353	38,482.99	7.05	cysk: 8, extr: 3, chlo: 1
AtqCAD2	1107	368	39,485.53	6.75	cysk: 11, cyto: 2, pero: 1
AtqCAD3	1071	356	38,497.48	6.16	cyto: 8, cysk: 4, plas: 1
AtqCAD4	1065	354	37,854.72	6.86	cyto: 11, cysk: 2, plas: 1
AtqCAD5	1074	357	38,286.33	5.82	cyto: 10, mito: 1, plas: 1
AtqCAD6	1113	370	39,936.18	6.60	cyto: 4, chlo: 3, nucl: 2
AtqCAD7	1065	354	38,779.55	7.55	cysk: 9, cyto: 2, extr: 2

L-Phenylalanine Ammonia-Lyase (PAL), Cinnamic Acid 4-Hydroxylase (C4H), 4-Hydroxycinnamate CoA Ligase (4CL), Caffeoyl Shikimate Esterase (CSE), Hydroxycinnamoyl CoA: Shikimate Hydroxycinnamoyl Transferase (HCT), Coumarate 3-Hydroxylase (C3H), Caffeoyl CoA 3-O-Methyltransferase (CCoAOMT), Cinnamoyl CoA Reductase (CRR), Ferulic Acid/Coniferaldehyde 5-Hydroxylase (F5H), Caffeic Acid 3-O-Methyltransferase (COMT), and Cinnamyl Alcohol Dehydrogenase (CAD). Molecular weight (Mw), pI: Isoelectric point (pI). Chloroplast or thylakoid lumen (clo), cytoskeleton (cysk), cytosol (cyto), endoplasmic reticulum (E.R.), extracellular or cell wall (extra), mitochondrion (mito), nucleus (nucl), peroxisome (pero), plasma membrane (plas), vacuolar membrane (vacu). Localization names joining two abbreviations indicate dual localization. * The score means the number of nearest neighbors with the same subcellular localization indicated. Only the three highest scores are shown.

Unlike the expression patterns of genes for cellulose formation, a transcriptomic landscape that is more diverse and contrasting was observed for lignin biosynthesis. All the gene isoforms encoding for L-Phenylalanine Ammonia-Lyase2 (PAL2), Cinnamic Acid 4-Hydroxylase (C4H), 4-Hydroxycinnamate CoA Ligase (4CL), Hydroxycinnamoyl CoA: Shikimate Hydroxycinnamoyl Transferase1 (HCT1), and Coumarate 3-Hydroxylase (C3H), enzymes that catalyze shikimate ester intermediates, were found to be from medium to highly expressed in roots and reproductive tissues (Supplementary Figure S2 and Table S3). Expression of *Cinnamoyl CoA Reductase* (*CCR*), a critical gene that drives the first specific step synthesizing the lignin monomers [36], was restricted to two lignified tissues. *AtqCCR1* had a basal expression in roots, while *AtqCCR2* was found to have enhanced expression in roots and pine. Similar to *AtqCCR2*, *Ferulic Acid/Coniferaldehyde 5-Hydroxylase* (*AtqF5H*) was highly expressed in roots, pine, and anthers. Both isoforms for *Caffeic Acid 3-O-Methyltransferase1* and *2* (*AtqCOMT1*, *2*) were widely expressed in almost all the Agave tissues and organs evaluated (Figure 2; Table S3). It is worth mentioning that *AtqCOMT2* was the lignin-related gene with the highest expression level that was found in Agave leaves. With some exceptions, all the *Cinnamyl Alcohol Dehydrogenase* (*CAD*) identified had a discreet expression pattern; *AtqCAD1* had a medium expression level, expressed throughout all the samples analyzed, and *AtqCAD3* had a high expression in tepals (Figure 2; Table S2).

AtqCAD5 was only found in samples with highly lignified tissues, roots, pine, and leaves, suggesting an essential role in monolignol biosynthesis.
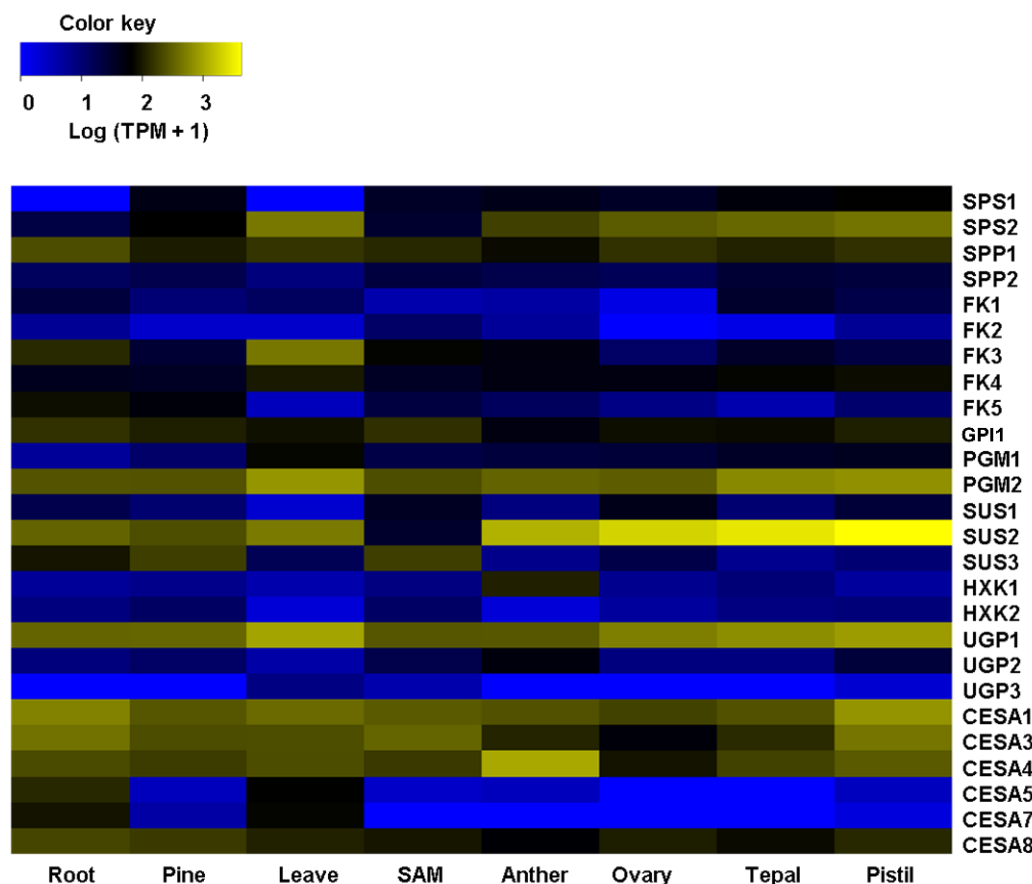

Figure 1. Heatmap depicting expression profiles of orthologous genes involved with cellulose biosynthesis across several tissues/organs of *A. tequilana*. Genes enlisted on the left side are plant biosynthetic genes for cellulose according to the literature. *Sucrose PhosphateSsynthase* (*SPS*), *Sucrose Phosphate Phosphatase* (*SPP*), *Fructokinase* (*FK*), *Phosphoglucomutase* (*PGM*), *Sucrose Synthase* (*SUS*), *Hexokinase* (*HXK*), *UDP-Glucose Ppyrophosphorylase* (*UGP*), and *Cellulose Synthase* (*CESA*). At the bottom, the last row indicates the RNA-sequencing samples previously reported [16]. SAM: Shoot apical meristem. Hierarchical clustering was conducted on the log (TPM+1). In the color key, yellow denotes high expression, black refers to medium, and blue denotes low expression.

2.3. Structural and Phylogenetic Characteristics of AtqCESA and AtqCAD Proteins

To analyze in deep of the sets of orthologous genes found, we focused on *CESA* and *CAD*, two genes encoding for key enzymes in determining cellulose and lignin content in plants, respectively, and that were found upregulated in transcriptomes coming from lignocellulose-enriched samples (roots, pine, and leaves). Multiple sequence alignments between all the predicted AtqCESA proteins and *A. thaliana* revealed high sequence conservation, with percentages of identity and similarity ranging from 54–84% and 66–89%, respectively. When these analyses were carried out with *Asparagus officinalis*, the closest phylogenetically monocot species to Agave with publicly available CESA sequences, percentages increased to 44–93% for identity and 54–96% for similarity (Table S3). For CAD proteins, the identity comparison between *A. tequilana* and *A. thaliana* showed 41–75% and 58–81% for similarity. CAD proteins have not been reported for *A. officinalis*; thus, analyses were performed with *O. sativa*. AtqCAD proteins showed 39–80% identity and 52–88% similarity with respect to rice (Table S4).
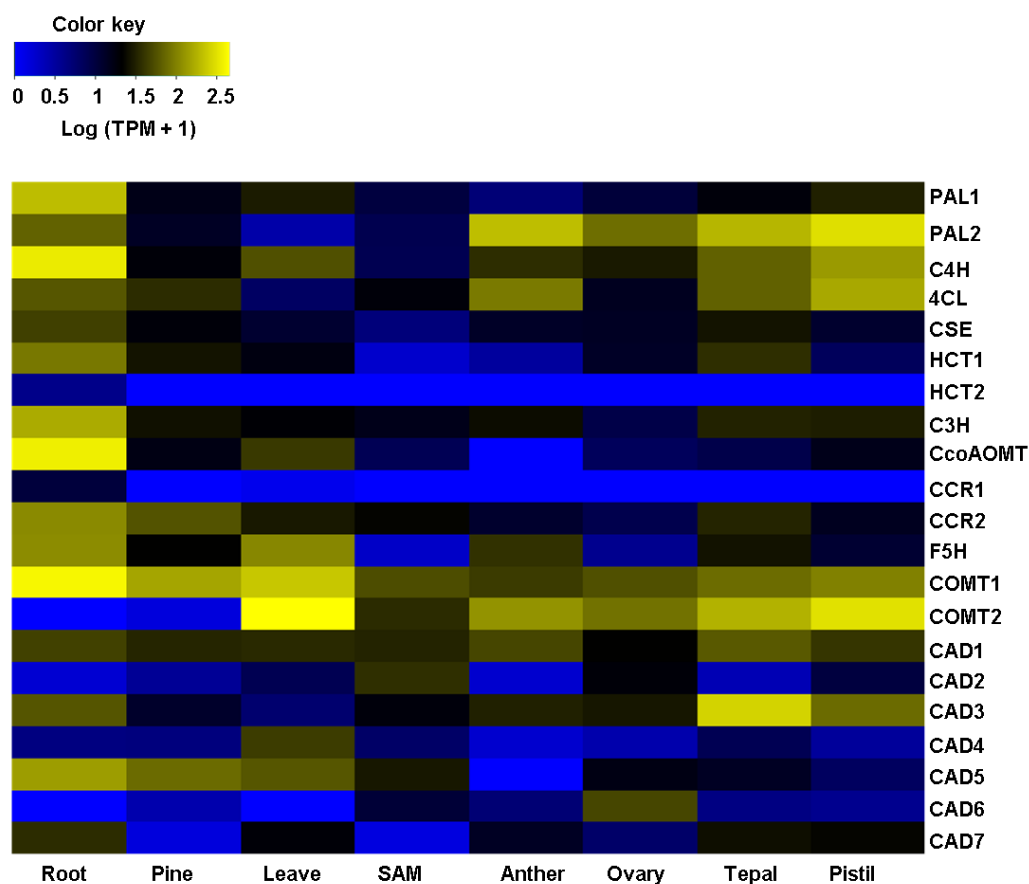
Figure 2. Heatmap of expression profiles of orthologous genes involved with lignin biosynthesis across several tissues/organs of *A. tequilana*. Genes enlisted on the left side are plant biosynthetic genes for lignin according to the literature. *L-Phenylalanine Ammonia-lyase* (*PAL*), *Cinnamic Acid 4-Hydroxylase* (*C4H*), *4-Hydroxycinnamate CoA Ligase* (*4CL*), *Caffeoyl Shikimate Esterase* (*CSE*), *Hydroxycinnamoyl CoA: Shikimate Hydroxycinnamoyl Transferase* (*HCT*), *Coumarate 3-Hydroxylase* (*C3H*), *Caffeoyl CoA 3-O-Methyltransferase* (*CCOAOMT*), *Cinnamoyl CoA Reductase* (*CRR*), *Ferulic Acid/coniferaldehyde 5-Hydroxylase* (*F5H*), *Caffeic Acid 3-O-Methyltransferase* (*COMT*), and *Cinnamyl Alcohol Dehydrogenase* (*CAD*). At the bottom, the last row indicates the RNA-sequencing samples previously reported [16]. SAM: Shoot apical meristem. Hierarchical clustering was conducted on the log (TPM+1). In the color key, yellow denotes high expression, black refers to medium expression, and blue denotes low expression.

Putative AtqCESA proteins are 1020–1090 amino acids (aa) in length, similar in size to those previously reported for other monocots [37]. When multiple aa alignments with CESAs from Arabidopsis, rice, maize, and garden asparagus were carried out, all the specific structural elements reported previously for plant CESA proteins [37–39] were found to be highly conserved in the six deduced AtqCESAs (Figure 3). The zinc-finger domain (ZnF), the two high variable class-specific regions (C-SR-I and C-SR-II), and the two plant-conserved regions (P-CR) were fully conserved with respect to Arabidopsis and other monocots included in this analysis (Figure 3). The motif CXXC, located within the ZnF domain and a signature for genuine CESA proteins and useful to differentiate them from CESA-like proteins (CS-L) [37], was also present in all the AtqCESA proteins; in *A. tequilana*, CXXC had two residues C that were invariable in all the sequences analyzed, while the middle residues showed changes (Figure 3). CQ(I/L)C was found in AtqCESA1, AtqCESA3, AtqCESA5, and AtqCESA7, and across all most Arabidopsis and monocot CESAs compared. C(R/K)(V/A)C was found in AtqCESA4 and AoCESA4 (Figure 3), and their corresponding orthologous AtCESA4, ZmCESA10, and TaCESA4. C(N/S)SC was observed only in AtqCESA8 and AoCESA2 (Figure 3). All these variations in the motif

CXXC are conservative replacements indeed. For CQ(I/L)C, isoleucine (I) and leucine (L) are isomers, while in C(R/K)(V/A)C and CN/SSC, the replacements occur with aa with equivalent charges in the lateral chain. With respect to the transmembrane helices (TME) that anchor CESA proteins to the plasma membrane, the two amino terminus-located TMEs, TME1 and TME2, and the six TMEs close to the carboxyl terminus, TME3–6, showed slight variation in the AtqCESA sequences analyzed (Figure S3). The full-conserved motif QXXRW, a catalytic site of CESA proteins [40–43], was identified between the second and third TME. Moreover, the residues D and DCD within P-CR-I and D within P-CR-II, all suggested to be involved in catalyzing long-chain polysaccharides [42], were also found to be full conservatives in all the CESA sequences analyzed (Figure 3).
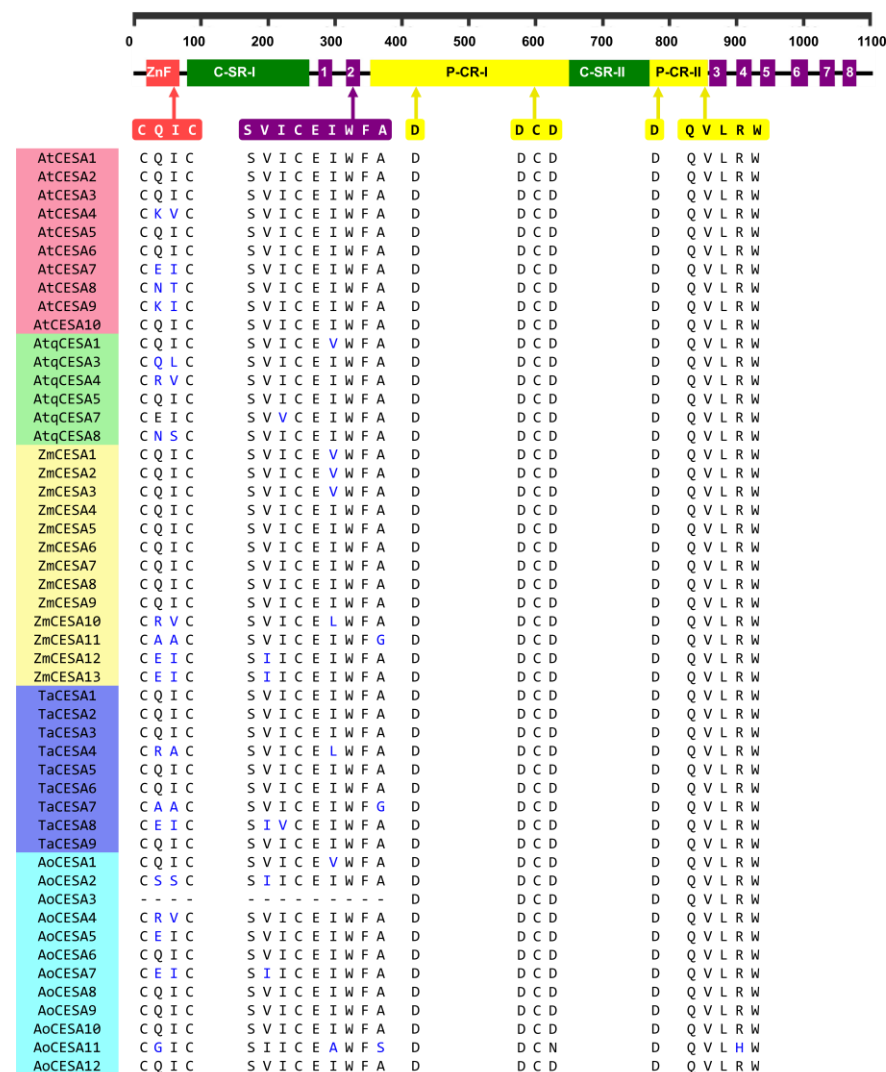

Figure 3. Motif alignment of AtqCESAs with CESAs from other species. *Arabidopsis thaliana* (AtCESA), *Agave tequilana* (AtqCESA), *Zea mays* (ZmCESA), *Triticum aestivum* (TaCESA), and *Asparus officinalis* (AoCESA). Non-conserved residues are highlighted in blue.

The seven *AtqCAD* orthologous predicted for *A. tequilana* encodes for putative proteins with a molecular size ranging from 353 aa to 370 aa. Structural features analysis allowed us to identify the two conserved Zn binding motifs, GHEX2GX5GX2V and GDX9CX2CX2CX7C, previously named Zn1-BD and Zn2-BD, respectively (Figure 4) [44]. The motif Zn1-BD is considered the Zn-catalytic center in bona fide CAD proteins, while Zn2-BD is a structural domain [44–47]. In AtqCAD proteins, the domains CX4H and the glycine-rich repeat GLGGVG, which participate in binding NADP and NADPH, respec-

tively [5,45], were also identified and were highly conserved compared to all other AtCAD and monocot CADs analyzed (Figure 4). The twelve residues that compose the substrate-binding pocket of CAD proteins, previously reported in rice [48], were conserved or substituted by aa with similar properties (Figure S4). Among all the AtqCADs analyzed, the key phenylalanine (F) residue of bona fide CAD proteins [49,50] was only present in AtqCAD5 at position 299, similar to the well-characterized AtCAD5. Taking all these data together, it is suggested that the deduced AtqCAD proteins belong to the zinc-dependent alcohol dehydrogenases with a potential functional role in the lignin biosynthesis in *A. tequilana*.
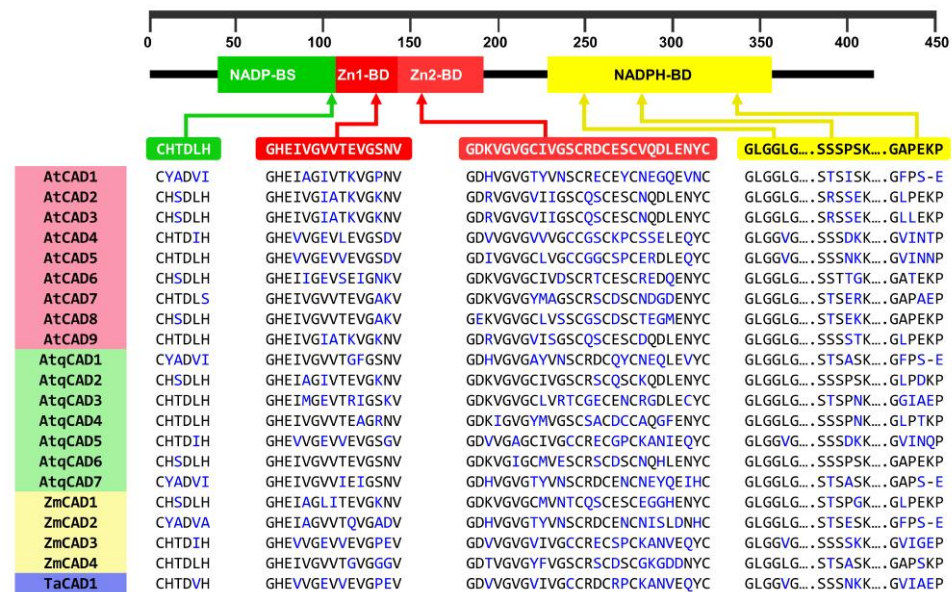

Figure 4. Motif alignment of AtqCADs with CADs from other species. *Arabidopsis thaliana,* (AtCAD) *Agave tequilana* (AtqCAD), *Zea mays* (ZmCAD), and *Triticum aestivum* (TaCAD). Non-conserved residues are highlighted in blue.

To gain insight into the evolutionary relationships of putative *CESA* and *CAD* genes of *A. tequilana*, we carried out phylogenetic analysis with all orthologous found in this study and *CESAs* and *CADs* from seven dicots and five monocots, including those well-studied models, such as *A. thaliana*, *A. sativa*, and *Z. mays*. An unrooted tree showed a distribution of all the paralogues *AtqCESAs* in six different clades that grouped closely related orthologous from monocots species, excepting *AtqCESA7*, which was grouped with *AtCESA7* first (Figure 5). This clustering, supported by relatively high bootstrap values, was consistent with the identity and similarity analyses and orthology prediction. Interestingly, three *AtqCESA* genes, *AtqCESA4*, *AtqCESA7*, and *AtqCESA8*, shared clades (red-colored clades, Figure 5) with the well-characterized CESA genes from *A. thaliana* for SCW, *AtCESA4*, *AtCESA7*, and *AtCESA8*, suggesting a similar biological role during cell wall formation in *A. tequilana*.

According to previous phylogenetic studies for monocot *CADs* [44], our analyses for Agave CAD proteins and 32 CADs from other plant species revealed three major classes clearly divided, named Class I, II, and III. CAD Class I can be subdivided into two subgroups (Figure 6). Clade Class I clusters monocot CADs and another for eudicots; this group contains AtqCAD5 and another five members previously considered as bona fide CAD proteins (i.e., those whose products catalyze genuine enzymatic reactions) [44,45,50]. OsCAD2, the closest in silico match with Agave CAD proteins (88% similarity concerning AtqCAD5) (Table S4) was the only rice CAD protein included in this class. Clade CAD class II clusters most of the AtqCADs predicted, AtqCAD2, AtqCAD3, AtqCAD4, and AtqCAD6, and shows no a clear separation for the monocots and eudicots species analyzed. Clade

Class III included AtqCAD1 and AtqCAD7 and was the smallest group in the phylogenetic tree obtained (Figure 6).
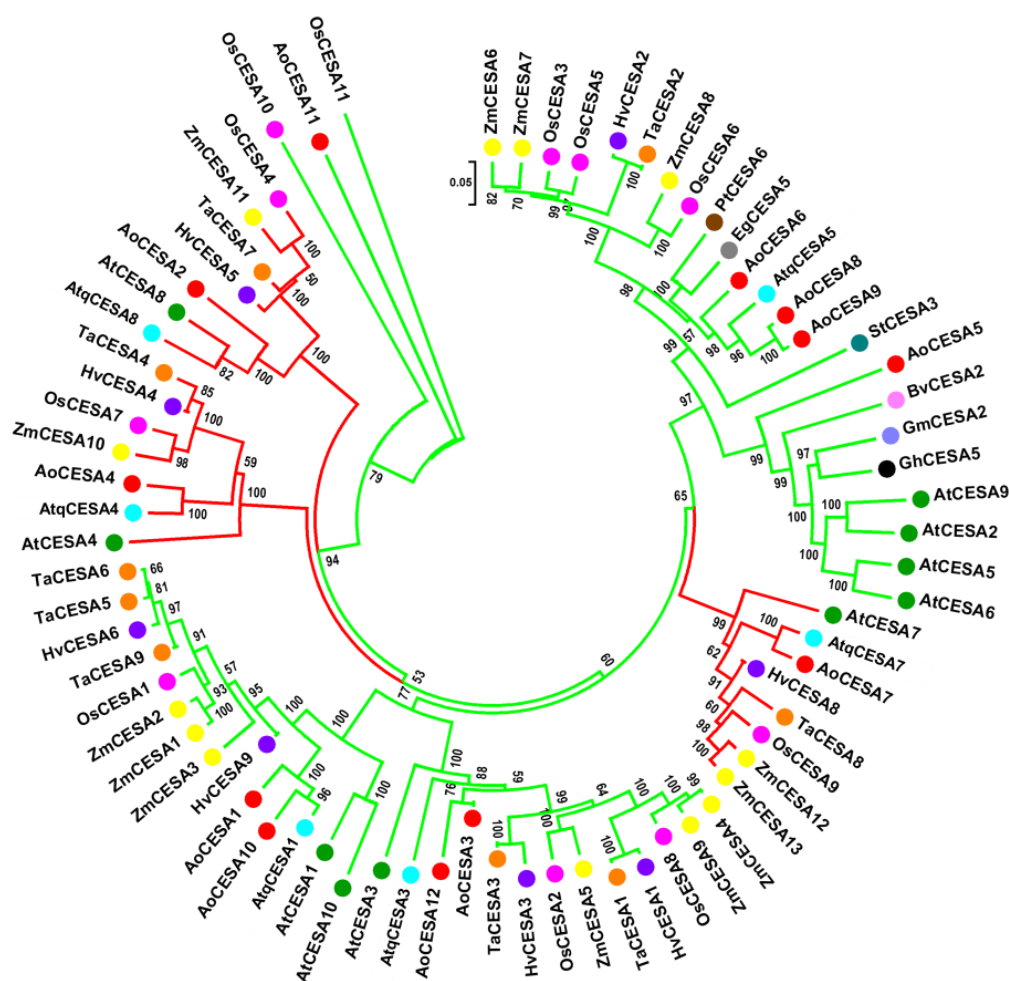

Figure 5. Unrooted phylogenetic tree of the CESAs of *Agave tequilana* (AtqCESA), *Arabidopsis thaliana* (AtCESA), *Beta vulgaris* (BvCESA), *Eucalyptus grandis* (EgCESA), *Glycine max* (GmCESA), *Gossypium hirsutum* (GhCESA), *Hordeum vulgare* (HvCESA), *Oryza sativa* (OsCESA), *Populus trichocarpa* (PtCESA), *Solanum tuberosum* (StCESA), *Asparagus officinalis* (AoCESA), *Triticum aestivum* (TaCESA), and *Zea mays* (ZmCESA). The phylogenetic tree was inferred using a maximum likelihood method.

2.4. Retrieved Agave Orthologous Genes for Cellulose and Lignin Biosynthesis Are Preferentially Expressed in Cells Surrounding Leaf Sclerenchyma Fibers

Agave leaves are SCW-enriched organs since their high content of sclerenchyma fibers [51]. Thus, it was expected that SCW-related genes were highly expressed in the leaf transcriptome mined here, although leaf transcriptomic heatmaps showed null, low, and medium expression levels of key genes for cellulose and lignin metabolism. We suppose these expression levels could be underrated because of sample-processing issues during total RNA isolation. During Agave sample processing, fibers are eliminated usually when leaf samples are grounded in liquid nitrogen, throwing out cells associated with fibers or developing fibers that potentially express biosynthetic genes for SCW components. In this last part of our study, we tested this hypothesis by quantifying the expression levels of some Agave genes using qPCR, predicted to be working in biosynthetic pathways for cellulose or lignin during SCWs formation, according to our orthology inference, transcriptomic data clustering, structural motif, and phylogeny analysis.
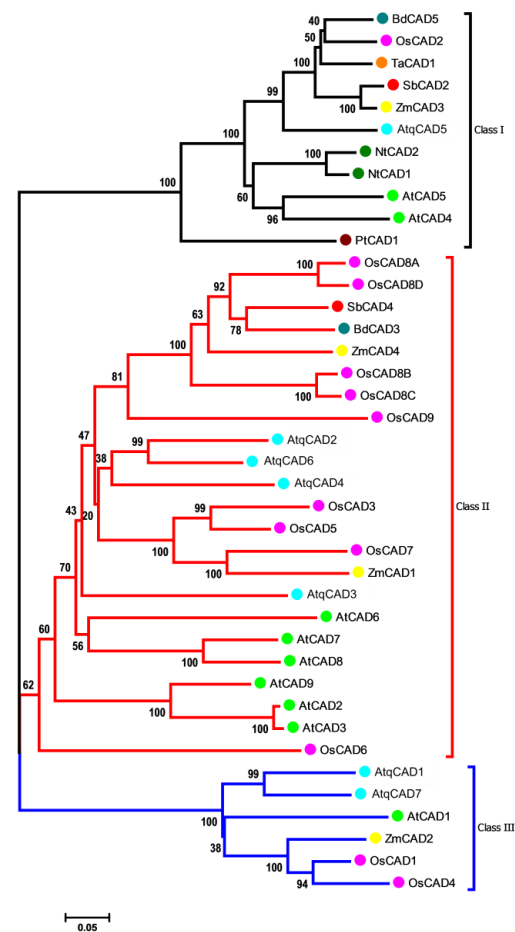
Figure 6. Unrooted phylogenetic tree of the CADs from *Agave tequilana* (AtqCAD), *Zea mays* (ZmCAD), *Oryza sativa* (OsCAD), *Sorghum bicolor* (SbCAD), *Nicotiana tabacum* (NtCAD), *Pinus taeda* (PtCAD), *Brachypodium distachyon* (BdCAD), *Triticum aestivum* (TaCAD), and *Arabidopsis thaliana* (AtCAD).

Histology analysis of cross sections showed typical sclerenchyma fibers surrounding vascular bundles in Agave leaves, which according to their position with respect to xylem can be classified as extraxylary fibers [51] (Figure 7A, left). An enhanced cellulose and lignin deposition was detected in the cell walls of these extraxylary fibers compared to those of surrounding parenchyma tissue (Figure 7A, middle and right). These observations were confirmed by monomers quantification of cellulose, hemicellulose, and lignin (Figure 7B). Although no statistical significance was found between cellulose levels fibers and whole leaf, determined by quantifying their glucose subunits, a tendency of higher content was observed in fibers. When hemicellulose components were measured, parenchyma cell walls had a higher content for all the monosaccharides compared with fibers, except gluconic acid (GA) (Figure 7B, middle). On the other hand, lignin analysis revealed that total lignin content is about 1.6-fold higher in fibers compared to that of surrounding parenchyma, mostly explained by higher levels of syringyl (H) monolignols in the sclerenchymatous tissue (Figure 7B, left), in line with that known for fibers [52]. To circumvent the potential bias described above in the SCW-involved gene expression analysis, we contrasted the gene expression patterns *AtqSUS2*, *AtqCESA1*, *AtqCESA7*, *AtqCCR2*, and *AtqCAD5* between cells associated to leaf fibers (Figure 7C, left) and whole leaf. All the evaluated genes were more expressed in the fibers-associated cells than the in whole leaf, except for *AtqCCR2*. Among the three genes analyzed for cellulose biosynthesis, *AtqCESA7*, which we predicted to be the orthologous for *AtCESA7* that is widely known for its vital role during CSW biogenesis, was the highest expressed in cells surrounding fibers, with a 4.5-fold change respect to the levels found in whole leaf (Figure 7C, right). Similarly, *AtqCAD5*, the Agave CAD isoform

clustered in our phylogenetic analysis with bona fide CAD proteins, had an expression level almost 2-fold higher in the fiber-associated cells than in the whole leaf (Figure 7C, right). Overall, this upregulation of predicted genes involved with SCW biogenesis in cells associated with extraxylary fibers is consistent with the chemical composition of such tissues.
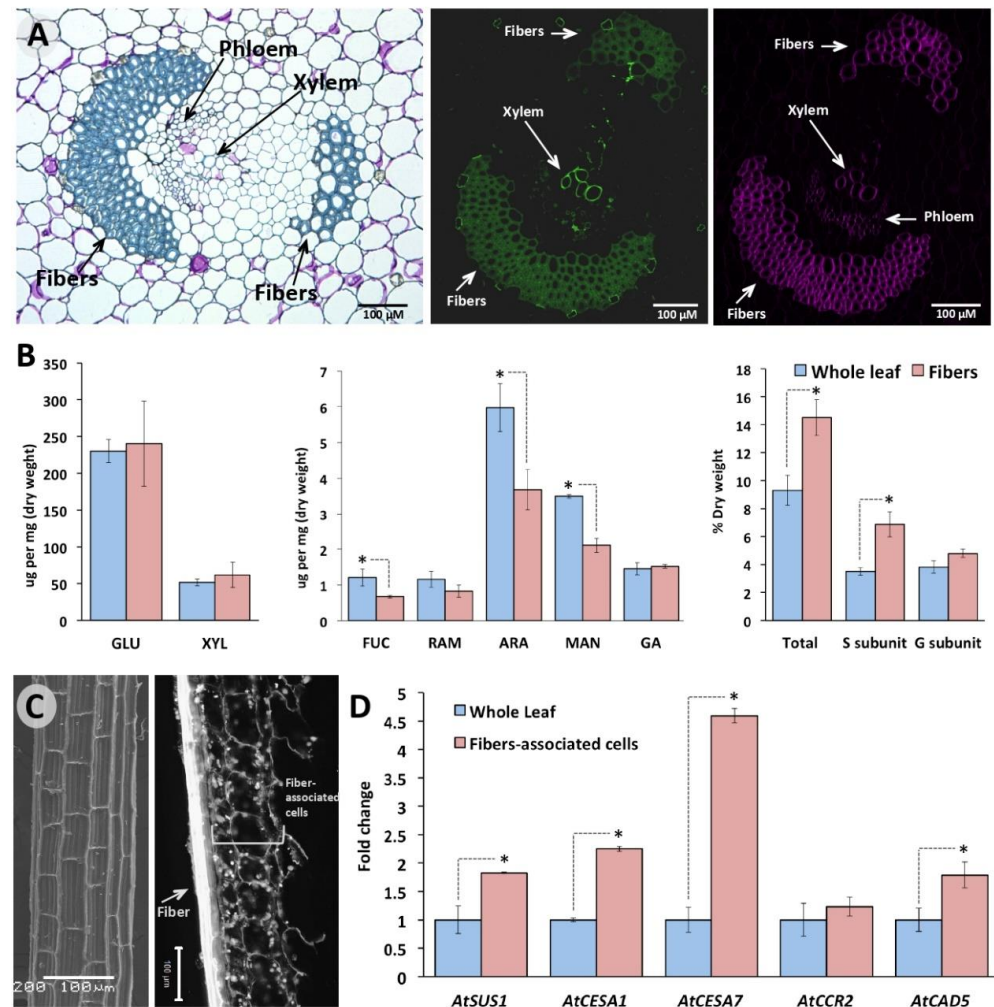

Figure 7. Key genes for cellulose and lignin have a high expression in cells surrounding fibers of A. tequilana leaves. (**A**) Sclerenchyma fibers surrounding vascular bundles. Heavily thickened secondary cells forming fiber caps walls were visualized in cross leaf sections after Toluidine Blue O (TBO) staining (left); lignification of fibers and xylem cells were visualized by lignin autofluorescence imaged by confocal laser scanning microscopy (LCSM) (middle). Cellulose was detected by Direct Red 23 staining and visualized by LCSM (right). (**B**) Monosaccharides analysis (left and middle) by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) and lignin quantification (left) by pyrolysis gas chromatography/mass spectrometry (py-MBMS) of sclerenchyma fibers and cell walls of whole leaf. FUC: Fuccose, RAM: ramose, ARA: arabinose, MAN: mannose, GA: galacturonic acid, GLU: glucose; XYL: xylose. (**C**) Scanning Electronic Microscopy (SEM) of fiber (left) and fibers-associated cells stained with propidium iodide and imaged by LSCM (right). (**D**) Expression levels of key genes for cellulose and lignin biosynthesis quantified by qPCR assays. Total RNAs isolated from whole leaf and fibers-associated cells were used for RT-qPCR assays (right). Glycerol-3-Phosphate Dehydrogenase (GPDH) was used as a load gene. Values are means ± SD. Asterisks indicate significant statistical differences between fibers/fibers-associated cells and whole leaf determined by Tukey's HSD test ($p \leq 0.05$).

3. Discussion

Unlike in dicots, biosynthetic and regulatory genes involved with cell wall metabolism are less known in monocots. Although significant advances have been achieved for maize, wheat, and rice, for emerging bioenergy crops such as Agave, our knowledge about cell wall biogenesis is still very limited. For monocot species, in the absence of reference nuclear genomes, transcriptome resources have provided insights into the biosynthesis and regulatory mechanisms for cellulose and lignin biosynthesis, two key components for cell walls.

Although the biosynthetic pathways for cellulose and lignin have been explored in *A. sisalana*, *A. fourcroydes*, and the hybrid H11648 [17,19], our study dissects both routes in *A. tequilana* for the first time, and provides their transcriptomic landscapes in contrasting tissues and organs for gaining more insights into the biological significance of the predicted orthologous. Variation in the numbers of isoforms found in both metabolic routes between our results and earlier analyses may reflect species or tissue-specific expression patterns, expressed according to the physiological status, development, and environment tested, as reported previously in Agave species [12,19,53]. For example, in silico expression profiling and qPCR assays of genes of the glycoside hydrolase family 32 revealed new vacuolar invertase isoforms only for *A. tequilana* and *A. deserti* but not for *A. victoriae-reginae*, and *A. striata* [12]. Immunolocalization of mayahuelin, a type I ribosome-inactivating protein (RIP), showed differential protein abundance, including absence, depending on the Agave species or cultivar screened [53]. Several studies support the idea that *CESA* genes belong large families in monocots. In wheat, maize, and rice, 14, 10, and 11 CESAs members have identified, respectively [37,54–56]. In grasses, in *Panicum virgatum* 14 *CESA* genes have been reported [56], while an RNA-seq analysis revealed 27 *CESA* genes expressed during stem development of *Pennisetum purpureum* [57]. Here, we retrieved only six *CESA* orthologous in *A. tequilana*, which is quite comparable with that found in the hybrid H11648 (five CESA genes) [17] but contrasts with that reported recently in *A. fourcroydes*, *A. sisalana*, and H11648 (14 CESA genes) [19].

For COMT genes, which drive the first specific step conducting lignin syringyl unit biosynthesis, Agave species seem to harbor small families. We found only two putative *AtqCOMT* genes in *A. tequilana*, while three genes have been identified in *A. fourcroydes*, *A. sisalana*, and the hybrid H11648 [19], findings similar to that found in maize [58] but different from the 18 genuine COMT and COMT-like sequences found in sugarcane transcriptomes [59]. For *CAD* genes, the existence of multiple orthologous found in *A. tequilana* is also consistent is a high number *CAD* genes reported in fiber-producing Agave species [19], and other monocots, such as rice (12 genes) and *Setaria viridens* (11 genes) [60,61]. For the rest of cellulose and phenylpropanoid pathway genes, they were grouped in small families or even classified as unigenes, comparable with that reported in *A. fourcroydes*, *A. sisalana*, and the hybrid H11648 [17,19]. This fact, and considering that these two Agave species show that different ploidy levels ranging from 2n to 5n [62–64] support the previous suggestion of null existence of expanded gene families in Agaves [9], at least for lignin metabolism. Among the genes identified in our study, it is worth mentioning the finding of *CSE*, that encodes an enzyme that converts caffeoyl CoA to caffeic acid [65], previously also reported in transcriptomes of fiber-producing Agaves [19]. The existence of *CSE* genes expressed in diverse Agavaceae members allows us to speculate this gene has a functional role in this taxonomic family; however, its real participation in lignin biosynthesis in plants is still controversial as *CSE* is absent in many monocots genomes, including *B. distachyon*, *S. italica*, *S. bicolor*, *H. vulgare*, *Oriza sativa*, and *Z. mays* [66].

Interestingly, our analysis also revealed that orthologous predicted for genes located upstream of lignin-specific steps but still into the general phenypropanoid pathway have isoforms highly expressed in reproductive tissues of *A. tequilana*. It is known that CCR and CAD-upstream enzymes provide intermediates not only for lignin biosynthesis but also for hydroxycinnamoyl esters, isoflavones, and coumarin derivatives [67,68]. Recently, it was shown that genes of the phenylpropanoid pathway, including lignin-specific biosynthetic

genes, are expressed in the tapetum in Arabidopsis [69]. In line with this, in *A. tequilana*, all these genes are also highly expressed in anther, ovary, tepals, and pistil. We found enhanced expression for some isoforms of *AtqPAL*, *AtqC4H*, *Atq4CL*, *AtqCOMT*, and *Atq-CAD*, which may yield phenylpropanoid derivaties; indeed, a high content of kaempferol glycosides has been quantified in pollen and flowers of *A. americana* and *A. durangneisis*, respectively [70,71]. Slight transcriptional changes for *AtqHCT1* and *AtqC3H* observed in *A. tequilana tepal* and pistil could explain the previously reported chlorogenin levels of Agave flowers [70]. In Agave plants, all these phenolic compounds may prove UV protection to reproductive tissues, as assumed in Arabidopsis [69], and potential protection against pathogens.

Our detailed analyses of key enzymes for cellulose and lignin, *CESA* and *CAD*, respectively, gave some clues about in their participation in cell wall biosynthesis. Protein sequences analysis was consistent with phylogenetic clustering obtained for both types of enzymes, suggesting a high evolutionary conservation within monocots and with respect to dicots. Identification of all the structural and functional motifs of typical plant CESA proteins in all the *AtqCESAs* predicted here indicate we retrieved full-length *CESA* sequences, encoding genuine CESA proteins, and potentially involved in UDP-glucose monomers polymerization. Although we were unable to distinguish a clear conservation of the new motifs proposed to classify monocot CESA into PCW CESAs or SCW CESAs [37], the phylogenetic analysis and transcriptional profiling of these *AtqCESAs* in tissues and organs with contrasting development level suggest *AtqCESA1*, *AtqCESA3*, and *AtqCESA4* are putative PCW *AtqCESAs*, while *AtqCESA5* and *AtqCESA7* are SCW *AtqCESAs*. *AtqCESA8*, in spite of its orthologous in *A. thaliana* (*AtCesA8/IRX1*), is an SCW CESA [72] that was found to be expressed relatively uniform across all the transcriptomes mined, giving a unclear idea about their role in cell metabolism. Among the *AtqCADs* identified, *AtqCAD5* is the phylogenetic closest to bona fide CAD proteins previously reported in monocots. This finding suggests that *AtqCAD5* is a bona fide CAD protein in *A. tequilana*. This is also supported by another evidence line: high expression in parenchyma cells associated to extraxylar fibers.

Molecular mechanisms involved in cell wall formation of sclerenchyma fibers are little known in plants, especially for extraxylar fibers (leaf fibers), as leaf fibers are classified in Agave [51,73]. From that is known of interfascicular fibers in *A. thaliana*, it is assumed that fiber lignification is a cell autonomous process because fibers necessary require cell death to achieve their structural support function [52,74]. However, leaf fibers have a different ontogeny compared to interfascicular fibers [74,75]; thus, events leading to cellulose and lignin formations are remained to be uncovered fully in monocots. Although our dissection technique of fibers is not enough precise to discard contamination of xylem elements, our qPCR analysis for some key genes of cellulose and lignin clearly showed an enhanced transcriptional expression in tissue surrounding to fibers and vascular bundles, suggesting: (a) *AtqCESA7* and *AtqCAD5* are enzymes working in SCW formation, and (b) cooperative or partial cooperative lignification is underlying leaf fiber development in *A. tequilana*. Agave genome sequencing will help to know with accuracy the gene families' size and the evolution history of genes related to cell wall metabolism. Likewise, further research will be required to uncover the molecular mechanisms underlying fiber development in Agave species, which seem to be an excellent model plant to study sclerenchyma tissue.

4. Materials and Methods

4.1. Data Mining for Cell Wall-Related Genes

To identify genes involved with cellulose and lignin biosynthesis (Figures S1 and S2, respectively), we mined data of eight transcriptomes from roots, stem apical meristem, and juvenile and adult leaves of *A. tequilana*, sequenced by HiSeq 2000 (Ilumina) and previously reported [9] (GenBank SRP019885; Table S5). Raw data of all these short-read cDNA libraries were used for a novo transcriptome assembly and functional annotation. Before assembly, quality control of sequenced reads was carried out. Raw

reads were filtered for high-quality sequences using the qualityControl.py.script from Next-Generation Sequencing (NGS) toolkits (https://github.com/Czh3/NGSTools/blob/master/qualityControl.py, accessed on 30 August 2021) with the default settings. Overlapping reads were merged into a single longer read using Seqprep program (https://github.com/jstjohn/SeqPrep, accessed on 5 September 2021). The resulting reads were used for assembly by Trinity (V2.4.0) [76], and the unigenes obtained were trimmed from polyA/T tails, ends rich in undetermined bases, and low-complexity sequences by using SeqClean software (http://compbio.dfci.harvard.edu/tgi/, accessed on 10 September 2021). Frameshift errors were corrected in the predicted open-reading frames by AlignWise pipeline [77]. Later, a non-redundant dataset was generated using the BLASTClust program (https://www.ncbi.nlm.gov/Web/Newsltr/Spring04/blastlab.html, accessed on 12 September 2021). Finally, the coding regions corresponding to the unigenes generated were translated to proteins. Using the BBH (Best BLAST Hit) method, homologous proteins were identified to proteins translated from A. tequilana. As reference proteins, we used complete genomes sequenced of five different species plants, which were strategically selected as representatives of specific clades along the phylogenetic distribution of angiosperm plants: *Vitis vinifera* within the order Vitals, *Amborella trichopoda* within Amborellaceae, *S. lycopersicum* within Solanales, *Z. mays* within Poales, and *A. thaliana* within Brassicales.

4.2. Orthologous Groups Identification and Protein Subcellular Localization Prediction

The closest orthologous and paralogous genes were identified using the Markov clustering algorithm [78], an inflation value of 1.5 [79], and the OrthoMCL software v. 2.0.9 [80]. For the formation of the orthologous groups, a minimum length of 30 aa was considered in all the proteins included in the analysis. Predicted proteins were analyzed for subcellular localization using the WoLF PSORT tool available in www.wolfpsort.hgc.pj (accessed on 14 September 2021).

4.3. Sequence Alignment and Phylogenetics of AtqCESA and AtqCAD Families

To identify structural and functional motifs, multiple alignments of AtqCESA and AtqCAD protein sequences were performed using ClustalW with default parameters [81]. Multi-species phylogenetic trees for both types of proteins were then constructed using MEGA 6.06 [82] with the Neighbor-joining method (1000 boot-strap trials, the Poisson model) based on the previous multi-alignments. CESA and CAD protein sequences of *Z. mays*, *O. sativa*, *A. officinalis*, *A. thaliana*, *Populus trichocarpa*, *T. aestivum*, *B. vulgaris*, *Eucalyptus grandis*, *G. max*, *Gossypium hirsutum*, *H. vulgare*, and *S. tuberosum*, included in the phylogenetic analyses, were obtained from the UniProt database (www.uniprot.org/help/uniprotkb). Identity/similarity analysis was carried out for all the CESA and CAD proteins identified in *A. tequilana*. For this, pairwise determinations were calculated with Sequence Manipulation Suite (SMS), available at https://www.bioinformatics.org/sms2/ident_sim.html (accessed on 18 September 2021) [83].

4.4. In Silico Gene Expression Analysis of Cellulose and Lignin Biosynthesis

Our approach to propose robust sets of closest orthologous genes encoding enzymes for cellulose and lignin was carried out an expression profiling analysis after orthology prediction, using data of eight RNA-seq transcriptomes [16] [June Simpson lab. Unpublished data] obtained from a wide range of *A. tequilana* tissues and organs, enriched with PCWs (meristem, anthers, ovules, tepals, and pistils) and SCWs (roots, stems, and leaves). Normalized data of these RNAseq transcriptomes were used to get tissue expression profiles using RSEM software [84]. RSEM uses an expectation-maximization algorithm, which produces an array with the expression profiles in transcripts per million (TPM). Heatmaps were created based on the expression levels using the heatmap.2 function of the gplots library of the statistical package R version 3.6.1 (https://www.r-project.org) (accessed on 21 September 2021).

4.5. Samples Collection and Processing

For lignin and structural carbohydrate quantifications and gene expression analysis, *A. tequilana* leaves samples were collected and processed as described immediately. Adult leaves (third full-expanded leaf) were collected from 4-year plants in Rancho Agropecuaria Santa Genóveva, Campeche México (19°33′03.0″ N 90°01′10.9″ W). On the lab-bench, three types of samples were obtained from leaves: pooled samples of the whole leaf, cells surrounding fibers, and fibers. For pooled samples of the whole leaf, three fractions (16 cm^2) were dissected from the basal (C), medium (B), and distal (A) parts, frozen with liquid nitrogen (N$_2$), and kept at −80 °C until processing. These three leaf fractions were used to get a landscape of the whole leaf, as reported previously in Agaves [9]. For isolation of cells surrounding fibers, leaves were longitudinally segmented with a sterilized knife, and then fibers were detached one by one and immediately frozen with N$_2$ and kept at −80 °C until processing. The presence of surrounding cells still to fibers was verified by propidium iodide staining (0.2 μg ul^{-1}) (Sigma-Aldrich, St. Louis, MO, USA) and visualized by laser scanning confocal microscopy (514 nm line for excitation, Carl Zeiss, Jena, Germany). Fibers isolated after grounding with N$_2$ were washed deeply with tap water and dried to 65 °C.

4.6. Monosaccharide and Lignin Composition

Monosaccharide and lignin profiles were quantified in whole leaves and dried fibers. The leaf fractions A, B, and C, isolated as described above, were grounded separately using N$_2$ and then lyophilized. A pooled sample representing the whole leaf was prepared using equal amounts of these fractions. Whole leaves and dried fibers were then analyzed by triplicate using high-performance anion exchange chromatography with a pulsed amperometric detector (HPAEC-PAD). Starch-free cell walls from whole leaf and fibers were prepared as reported [85,86]. Hydrolysates were diluted with deionized water and then filtered through a nylon membrane of 0.22 mm before injection to a Dionex ICS-3000 (Thermo, Waltham, MA, USA) equipped with a CarboPac PA-100 guard column (4 mm × 50 mm) and a CarboPac PA-100 analytical column (4 mm × 250 mm) (Thermo, Waltham, MA, USA). Parameters used for monosaccharide separation were those previously reported [87]. Lignin composition (total content and monomers) was determined by pyrolysis gas chromatography/mass spectrometry (py-MBMS) as previously reported [86,88]. Briefly, processed samples were pyrolyzed (PY-2020 iS, Frontier Labs) for 2 min at 500 °C to produce volatile compounds. NIST8492 (*P. deltoids*) and aspen (*P. tremuloides*) analytical standards were included in each run of the experiment for quality control. Volatile compounds were analyzed for lignin using the MBMS (Extrel Max-1000, Super-Sonic, Pittsburgh, PA, USA), and raw data were processed through UnscramberX 10.1 software.

4.7. Histochemistry and Microscopy

For histochemical detection of lignin and cellulose associated with extraxylary fibers, samples of the leaf fraction C described above were processed as reported [86]. Chemical-fixed samples in 4% paraformaldehyde (*w/v*) (Sigma-Aldrich, St. Louis, MO, USA) were dehydrated in ethanol series and then embedded in JB4- water-soluble resin (Electron Microscopy Sciences, Hatfield, PA, USA). Cross sections, obtained with a Leica EM UC7 ultramicrotome (Leica Microsystems, Wetzlar, Germany), were used for histochemical analyses. Cellulose was detected by staining with 0.1% Direct Red 23 (Sigma-Aldrich, St. Louis, MO, USA) and imaged with a Leica SP8-iPhox inverted microscope (excitation 562 nm; emission 580–615 nm) (Leica Microsystems, Wetzlar, Germany), while lignin was detected by autofluorescence. Blue or blue-green-stained extraxylary fibers and xylem vessels were revealed using a 0.02% Toluidine Blue O solution (Sigma-Aldrich, St. Louis, MO, USA). For SEM analysis, dry fibers were sputter-coated with a gold layer and visualized using a JEOL JSM-6360LV (Tokyo, Japan).

4.8. RT-qPCR Analyses

For gene expression analysis, reverse transcription–quantitative real-time PCR (RT-qPCR) assays were carried out evaluating two kinds of samples: whole leaf and cells surrounding fibers. Ultra-frozen leaf fractions A, B, and C, obtained as described above, were processed separately for RNA isolation. Fibers with cells attached were deeply grounded in N_2, then the fibers were removed, and the tissue left on the mortar was used for further analysis. Extraction and purification of total RNA were carried out following the manufacturer's instructions (QIazol 79306, RNeasy MinElute Cleanup kit 74204, QIAGEN, Germantown, MD, USA). For the whole leaf, we prepared a pooled sample using equal amounts of total RNAs from fractions A, B, and C. Three biological replicates were processed for each type of sample, and five minipreps for each biological replicate. cDNA templates were synthetized using 1.7 µg of purified total RNA, oligodT primers (10 µM), and SuperScript III (18080093, Invitrogen/TermoFisher, Waltham, MA, USA). For real-time quantitative PCR (qPCR) analysis, based on the in silico expression patterns of the transcriptomes analyzed, we selected some key biosynthetic genes for cellulose (AtqSUS2, AtqCESA1, and AtqCESA7) and lignin (AtqCCR2 and AtqCAD5) to quantify their expression levels. Gene-specific oligonucleotides used in this study are listed in Table S6. qPCR reactions (20 µL) contained: 8 µL de SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA), 1.5 µL of template cDNA (cDNA diluted 1: 10), 2 µL of the mix of specific oligonucleotides (1 µL of each one, at 10 µM), and 8.5 µL of ultrapure water (Invitrogen/Life Technologies, Waltham, MA, USA). Reactions were run in a StepOne Real-Time PCR System equipped with StepOne Software v2.3 (Applied Biosystems, Foster City, CA, USA). Cycling conditions were as follows: one cycle at 95 °C for 10 min, 38 cycles of 95 °C for 15 s, and one cycle at 60 °C for 1 min, linked to a default dissociation stage program to detect non-specific amplification. *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE* (*GAPDH*) was used as the internal reference gene [89], and relative expression levels were determined using the comparative quantification method $2^{-\Delta\Delta Ct}$ [90].

5. Conclusions

Our results identified sets of orthologous for biosynthetic genes for cellulose and lignin pathways in *A. tequilana*. Sequence analyses and phylogeny reflected a high evolutionary conservation between all the Agave isoforms retrieved, and those from monocots and eudicots included in this study, which could have biological roles in cell wall biogenesis and metabolism as suggested by their differential expression patterns found. In sclerenchyma fiber development, the high expression of AtqCAD5 in parenchyma cells associated to extraxylary fibers suggest the existence in Agave of a cooperative lignification mechanism similar to that known for xylem tissue.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants11111496/s1, Figure S1: Biosynthetic pathway for cellulose in plants; Figure S2: Biosynthetic pathway for lignin in plants; Figure S3: AtqCESA proteins have conserved transmembrane domains; Figure S4: Protein alignment for AtqCAD; Table S1: Expression levels of genes involved in cellulose biosynthesis in *Agave tequilana* in several organs and tissues sampled; Table S2: Expression levels of genes involved in lignin biosynthesis in *Agave tequilana* in several organs and tissues sampled; Table S3: Percentages of identity/similarity between amino acid sequences of orthologous *CESA* genes of *A. tequilana* (*AtqCESA*), *A. officinalis* (*AoCESA*), and *A. thaliana* (*AtCESA*); Table S4: Percentages of identity/similarity between amino acid sequences of orthologous *CAD* genes of *A. tequilana* (*AtqCAD*), *O. sativa* (*OsCAD*), and *A. thaliana* (*AtCAD*); Table S5: cDNA libraries used in this study, previously reported by Gross et al., 2013 and deposited in Genbank (SRP019885); Table S6: Gene-specific oligonucleotides used for qPCR analysis. Supplementary Materials: DNA sequences for Atq orthologous for cellulose and lignin biosynthesis.

Author Contributions: L.F.M.-L., D.C.M.-V., J.L.V.-A., E.I.-L., A.G.R., M.B., J.Z.T., P.A., I.L.-R., F.A.-C. and M.G.L. Investigation. L.F.M.-L., E.B.G.-C., G.T., F.A.-C. and J.M.S. Formal analysis. L.F.M.-L., E.B.G.-C., J.S. and F.A.-C. Writing—original draft preparation. L.F.M.-L. and F.A.-C. Conceptualiza-

References

1. Lampugnani, E.R.; Khan, G.A.; Somssich, M.; Persson, S. Building a plant cell wall at a glance. *J. Cell Sci.* **2018**, *131*, jcs207373. [CrossRef] [PubMed]
2. Pauly, M.; Keegstra, K. Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.* **2008**, *54*, 559–568. [CrossRef] [PubMed]
3. Silva, C.O.G.; Vaz, R.P.; Filho, E.X.F. Bringing plant cell wall-degrading enzymes into the lignocellulosic biorefinery concept. *Biofuel Bioprod. Biorefin.* **2018**, *12*, 277–289. [CrossRef]
4. Prasad, R.K.; Chatterjee, S.; Mazumder, P.B.; Gupta, S.K.; Sharma, S.; Vairale, M.G.; Datta, S.; Dwivedi, S.K.; Gupta, D.K. Bioethanol production from waste lignocelluloses: A review on microbial degradation potential. *Chemosphere* **2019**, *231*, 588–606. [CrossRef]
5. Liu, Q.; Luo, L.; Zheng, L. Lignins: Biosynthesis and biological functions in plants. *Int. J. Mol. Sci.* **2018**, *19*, 335. [CrossRef]
6. Zarra, I.; Revilla, G.; Sampedro, J.; Valdivia, E.R. Biosynthesis and regulation of secondary cell wall. In *Progress in Botany*; Cánovas, F.M., Lüttge, U., Leuschner, C., Risueño, M.-C., Eds.; Springer International Publishing: Cham, Switzerland, 2020; Volume 81, pp. 189–226. [CrossRef]
7. Davis, S.C.; Dohleman, F.G.; Long, S.P. The global potential for *agave* as a biofuel feedstock. *GCB Bioenergy* **2011**, *3*, 68–78. [CrossRef]
8. Escamilla-treviño, L.L. potential of plants from the genus *agave* as bioenergy crops. *Bioenergy Res.* **2012**, *5*, 1–9. [CrossRef]
9. 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. [CrossRef]
10. Corbin, K.R.; Byrt, C.S.; Bauer, S.; DeBolt, S.; Chambers, D.; Holtum, J.A.M.; Karem, G.; Henderson, M.; Lahnstein, J.; Beahan, C.T.; et al. Prospecting for energy-rich renewable raw materials: *Agave* leaf case study. *PLoS ONE* **2015**, *10*, e0135382. [CrossRef]
11. Lu, M.L.; Wyman, C.E. Elucidation of native California *Agave americana* and *Agave deserti* biofuel potential: Compositional analysis. *PLoS ONE* **2021**, *16*, e0252201. [CrossRef]
12. Avila de Dios, E.; Gomez Vargas, A.D.; Damián Santos, M.L.; Simpson, J. New insights into plant glycoside hydrolase family 32 in *agave* species. *Front. Plant Sci.* **2015**, *6*, 594. [CrossRef]
13. 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. [CrossRef] [PubMed]
14. Cervantes-Pérez, S.A.; Espinal-Centeno, A.; Oropeza-Aburto, A.; Caballero-Pérez, J.; Falcon, F.; Aragón-Raygoza, A.; Sánchez-Segura, L.; Herrera-Estrella, L.; Cruz-Hernández, A.; Cruz-Ramírez, A. Transcriptional profiling of the CAM plant *Agave salmiana* reveals conservation of a genetic program for regeneration. *Dev. Biol.* **2018**, *442*, 28–39. [CrossRef] [PubMed]
15. 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. [CrossRef] [PubMed]
16. de Dios, E.A.; Delaye, L.; Simpson, J. Transcriptome analysis of bolting in *A. tequilana* reveals roles for florigen, MADS, fructans and gibberellins. *BMC Genom.* **2019**, *20*, 473. [CrossRef]
17. 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. [CrossRef]

18. Sarwar, M.B.; Ahmad, Z.; Rashid, B.; Hassan, S.; Gregersen, P.L.; De la O Leyva, M.L.; Nagy, I.; Asp, T.; Husnain, T. De novo assembly of *Agave sisalana* transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Sci. Rep.* **2019**, *9*, 396. [CrossRef]

19. Trigo Raya, F.; Marone Pupke, M.; Carvalho, L.M.; Rabelo, S.C.; de Paula, M.S.; Zaneli Campanari, M.F.; Freschi, L.; Sampaio Mayer, J.L.; Ribeiro Ferreira Silva, O.R.; Mieczkowski, P.; et al. Extreme physiology: Biomass and transcriptional profiling of three abandoned *Agave cultivars*. *Ind. Crops Prod.* **2021**, *172*, 114043. [CrossRef]

20. Penning, B.W.; Hunter, C.T., III; Tayengwa, R.; Eveland, A.L.; Dugard, C.K.; Olek, A.T.; Vermerris, W.; Koch, K.E.; McCarty, D.R.; Davis, M.F.; et al. Genetic resources for maize cell wall biology. *Plant Physiol.* **2009**, *151*, 1703–1728. [CrossRef]

21. Wang, L.; Guo, K.; Li, Y.; Tu, Y.; Hu, H.; Wang, B.; Cui, X.; Peng, L. Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *BMC Plant Biol.* **2010**, *10*, 282. [CrossRef]

22. Courtial, A.; Soler, M.; Chateigner Boutin, A.L.; Reymond, M.; Mechin, V.; Wang, H.; Grima-Pettenati, J.; Barriere, Y. Breeding grasses for capacity to biofuel production or silage feeding value: An updated list of genes involved in maize secondary cell wall biosynthesis and assembly. *Maydica* **2013**, *58*, 67–102.

23. Penning, B.W.; McCann, M.C.; Carpita, N.C. Evolution of the cell wall gene families of grasses. *Front. Plant Sci.* **2019**, *10*, 1205. [CrossRef] [PubMed]

24. Penning, B.W.; Shiga, T.M.; Klimek, J.F.; SanMiguel, P.J.; Shreve, J.; Thimmapuram, J.; Sykes, R.W.; Davis, M.F.; McCann, M.C.; Carpita, N.C. Expression profiles of cell-wall related genes vary broadly between two common maize inbreds during stem development. *BMC Genom.* **2019**, *20*, 785. [CrossRef]

25. Zhao, K.; Lin, F.; Romero-Gamboa, S.P.; Saha, P.; Goh, H.-J.; An, G.; Jung, K.-H.; Hazen, S.P.; Bartley, L.E. Rice genome-scale network integration reveals transcriptional regulators of grass cell wall synthesis. *Front. Plant Sci.* **2019**, *10*, 1275. [CrossRef]

26. Abraham-Juarez, M.J.; Martinez-Hernandez, A.; Leyva-Gonzalez, M.A.; Herrera-Estrella, L.; Simpson, J. Class I KNOX genes are associated with organogenesis during bulbil formation in *Agave tequilana*. *J. Exp. Bot.* **2010**, *61*, 4055–4067. [CrossRef] [PubMed]

27. Abraham Juárez, M.J.; Hernández Cárdenas, R.; Santoyo Villa, J.N.; O'Connor, D.; Sluis, A.; Hake, S.; Ordaz-Ortiz, J.; Terry, L.; Simpson, J. Functionally different PIN proteins control auxin flux during bulbil development in *Agave tequilana*. *J. Exp. Bot.* **2015**, *66*, 3893–3905. [CrossRef]

28. Liepman, A.H.; Wightman, R.; Geshi, N.; Turner, S.R.; Scheller, H.V. Arabidopsis a powerful model system for plant cell wall research. *Plant J.* **2010**, *61*, 1107–1121. [CrossRef]

29. Li, S.; Bashline, L.; Lei, L.; Gu, Y. Cellulose synthesis and its regulation. *Arab. Book* **2014**, *12*, e0169. [CrossRef]

30. Maloney, V.J.; Park, J.Y.; Unda, F.; Mansfield, S.D. Sucrose phosphate synthase and sucrose phosphate phosphatase interact in planta and promote plant growth and biomass accumulation. *J. Exp. Bot.* **2015**, *66*, 4383–4394. [CrossRef]

31. Stein, O.; Granot, D. Plant fructokinases: Evolutionary, developmental, and metabolic aspects in sink tissues. *Front. Plant Sci.* **2018**, *9*, 1–12. [CrossRef]

32. Bilska-Kos, A.; Mytych, J.; Suski, S.; Magoń, J.; Ochodzki, P.; Zebrowski, J. Sucrose phosphate synthase (SPS), sucrose synthase (SUS) and their products in the leaves of *Miscanthus × giganteus* and *Zea mays* at low temperature. *Planta* **2020**, *252*, 23. [CrossRef] [PubMed]

33. Muhlbach, H.; Schnarrenberger, C. Properties and intracellular distribution of two phosphoglucomutases from spinach leaves. *Planta* **1978**, *141*, 65–70. [CrossRef] [PubMed]

34. Periappuram, C.; Steinhauer, L.; Barton, D.L.; Taylor, D.C.; Chatson, B.; Zou, J. The plastidic phosphoglucomutase from Arabidopsis. A reversible enzyme reaction with an important role in metabolic control. *Plant Physiol.* **2000**, *122*, 1193–1200. [CrossRef] [PubMed]

35. Maceda-López, L.F.; Villalpando-Aguilar, J.L.; García-Hernández, E.; Ávila de Dios, E.; Andrade-Canto, S.B.; Morán-Velázquez, D.C.; Rodríguez-López, L.; Hernández-Díaz, D.; Chablé-Vega, M.A.; Trejo, L.; et al. Improved method for isolation of high quality total RNA from *Agave tequilana* Weber roots. *3 Biotech* **2021**, *11*, 75. [CrossRef]

36. Lacombe, E.; Hawkins, S.; Van Doorsselaere, J.; Piquemal, J.; Goffner, D.; Poeydomenge, O.; Boudet, A.-M.; Grima-Pettenati, J. Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: Cloning, expression and phylogenetic relationships. *Plant J.* **1997**, *11*, 429–441. [CrossRef]

37. Kaur, S.; Dhugga, K.S.; Gill, K.; Singh, J. Novel structural and functional motifs in cellulose synthase (CesA) genes of bread wheat (*Triticum aestivum*, L.). *PLoS ONE* **2016**, *11*, e0147046. [CrossRef]

38. Scavuzzo-Duggan, T.R.; Chaves, A.M.; Singh, A.; Sethaphong, L.; Slabaugh, E.; Yingling, Y.G.; Haigler, C.H.; Roberts, A.W. Cellulose synthase 'class specific regions' are intrinsically disordered and functionally undifferentiated. *J. Integr. Plant Biol.* **2018**, *60*, 481–497. [CrossRef]

39. Nawaz, M.A.; Lin, X.; Chan, T.-F.; Imtiaz, M.; Rehman, H.M.; Ali, M.A.; Baloch, F.S.; Atif, R.M.; Yang, S.H.; Chung, G. Characterization of cellulose synthase A (CESA) gene family in Eudicots. *Biochem. Genet.* **2019**, *57*, 248–272. [CrossRef]

40. Vergara, C.E.; Carpita, N.C. β-D-Glycan synthases and the CesA gene family: Lessons to be learned from the mixed-linkage (1→3),(1→4)β-D-glucan synthase. In *Plant Cell Walls*; Carpita, N.C., Campbell, M., Tierney, M., Eds.; Springer: Dordrecht, The Netherlands, 2001. [CrossRef]

41. Beeckman, T.; Przemeck, G.K.H.; Stamatiou, G.; Lau, R.; Terryn, N.; De Rycke, R.; Inzé, D.; Berleth, T. Genetic complexity of cellulose synthase a gene function in Arabidopsis embryogenesis. *Plant Physiol.* **2002**, *130*, 1883–1893. [CrossRef]

42. Xu, L. Overexpression of a new cellulose synthase gene ('PuCesA6') from Ussuri poplar ('*Populus ussuriensis*') exhibited a dwarf phenotype in transgenic tobacco. *Plant Omics* **2014**, *7*, 54.

43. Cao, S.; Cheng, H.; Zhang, J.; Aslam, M.; Yan, M.; Hu, A.; Lin, L.; Ojolo, S.P.; Zhao, H.; Priyadarshani, S.V.G.N.; et al. Genome-wide identification, expression pattern analysis and evolution of the Ces/Csl gene superfamily in pineapple (*Ananas comosus*). *Plants* **2019**, *8*, 275. [CrossRef] [PubMed]

44. Tang, R.; Zhang, X.Q.; Li, Y.H.; Xie, X.M. Cloning and *in silico* analysis of a cinnamyl alcohol dehydrogenase gene in *Pennisetum purpureum*. *J. Genet.* **2014**, *93*, 145–158. [CrossRef] [PubMed]

45. Youn, B.; Camacho, R.; Moinuddin, S.G.A.; Lee, C.; Davin, L.B.; Lewis, N.G.; Kang, C.H. Crystal structures and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org. Biomol. Chem.* **2006**, *4*, 1687–1697. [CrossRef] [PubMed]

46. Guo, D.-M.; Ran, J.-H.; Wang, X.-Q. Evolution of the cinnamyl/sinapyl clcohol dehydrogenase (CAD/SAD) gene family: The emergence of real lignin is associated with the origin of *bona fide* CAD. *J. Mol. Evol.* **2010**, *71*, 202–218. [CrossRef] [PubMed]

47. Bukh, C.; Nord-Larsen, P.H.; Rasmussen, S.K. Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in *Brachypodium distachyon*. *J. Exp. Bot.* **2012**, *63*, 6223–6236. [CrossRef]

48. 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. [CrossRef]

49. Bomati, E.K.; Noel, J.P. Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell.* **2005**, *17*, 1598–1611. [CrossRef]

50. Cheng, X.; Li, M.; Li, D.; Zhang, J.; Jin, Q.; Sheng, L.; Cai, Y.; Lin, Y. Characterization and analysis of CCR and CAD gene families at the whole-genome level for lignin synthesis of stone cells in pear (*Pyrus bretschneideri*) fruit. *Biol. Open.* **2017**, *6*, 1602–1613. [CrossRef]

51. Evert, R.F. *Esau's Plant Anatomy*, 3rd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006.

52. Barros, J.; Serk, H.; Granlund, I.; Pesquet, E. The cell biology of lignification in higher plants. *Ann. Bot.* **2015**, *115*, 1053–1074. [CrossRef]

53. Lledías, F.; Gutiérrez, J.; Martínez-Hernández, A.; García-Mendoza, A.; Sosa, E.; Hernández-Bermúdez, F.; Dinkova, T.D.; Reyes, S.; Cassab, G.I.; Nieto-Sotelo, J. Mayahuelin, a type I ribosome inactivating protein: Characterization, evolution, and utilization in phylogenetic analyses of *Agave*. *Front. Plant Sci.* **2020**, *11*, 573. [CrossRef]

54. Appenzeller, L.; Doblin, M.; Barreiro, R.; Wang, H.; Niu, X.; Kollipara, K.; Carrigan, L.; Tomes, D.; Chapman, M.; Dhugga, K.S. Cellulose synthesis in maize: Isolation and expression analysis of the cellulose synthase (*CesA*) gene family. *Cellulose* **2004**, *11*, 287–299. [CrossRef]

55. Wang, D.; Qin, Y.; Fang, J.; Yuan, S.; Peng, L.; Zhao, J.; Li, X. A missense mutation in the zinc finger domain of *OsCESA7* deleteriously affects cellulose biosynthesis and plant growth in rice. *PLoS ONE* **2016**, *11*, e0153993. [CrossRef] [PubMed]

56. Mazarei, M.; Baxter, H.L.; Li, M.; Biswal, A.K.; Kim, K.; Meng, X.; Pu, Y.; Wuddineh, W.A.; Zhang, J.-Y.; Turner, G.B.; et al. Functional analysis of cellulose synthase *CesA4* and *CesA6* genes in switchgrass (*Panicum virgatum*) by overexpression and RNAi-mediated gene silencing. *Front. Plant Sci.* **2018**, *9*, 1114. [CrossRef]

57. Zhang, W.; Zhang, S.; Lu, X.; Li, C.; Liu, X.; Dong, G.; Xia, T. Tissue-specific transcriptome analysis reveals lignocellulose synthesis regulation in elephant grass (*Pennisetum purpureum* Schum). *BMC Plant Biol.* **2020**, *20*, 528. [CrossRef] [PubMed]

58. Guillaumie, S.; Goffner, D.; Barbier, O.; Martinant, J.-P.; Pichon, M.; Barrière, Y. Expression of cell wall related genes in basal and ear internodes of silking brown-midrib-3, caffeic acid O-methyltransferase (COMT) down-regulated, and normal maize plants. *BMC Plant Biol.* **2008**, *8*, 71. [CrossRef] [PubMed]

59. Jardim-Messeder, D.; da Franca Silva, T.; Fonseca, J.P.; Junior, J.N.; Barzilai, L.; Felix-Cordeiro, T.; Pereira, J.C.; Rodrigues-Ferreira, C.; Bastos, I.; da Silva, T.C.; et al. Identification of genes from the general phenylpropanoid and monolignol-specific metabolism in two sugarcane lignin-contrasting genotypes. *Mol. Genet. Genom.* **2020**, *295*, 717–739. [CrossRef]

60. Hirano, K.; Aya, K.; Kondo, M.; Okuno, A.; Morinaka, Y.; Matsuoka, M. *OsCAD2* is the major *CAD* gene responsible for monolignol biosynthesis in rice culm. *Plant Cell Rep.* **2012**, *31*, 91–101. [CrossRef]

61. Ferreira, S.S.; Simoes, M.S.; Carvalho, G.G.; De Lima, L.G.A.; Svartman, R.M.A.; Cesarino, I. The lignin toolbox of the model grass *Setaria viridis*. *Plant Mol. Biol.* **2019**, *101*, 235–255. [CrossRef]

62. Lv, L.; Sun, G.; Xie, J.; Zang, X.; Hu, Y.; Duan, J. Determination of chromosomal ploidy in *Agave* ssp. *Afr. J. Biotechnol.* **2009**, *8*, 5248–5252.

63. Simpson, J.; Martínez Hernández, A.; Abraham Juárez, M.J.; Delgado Sandoval, S.; Sánchez Villarreal, A.; Cortés Romero, C. Genomic resources and transcriptome mining in *Agave tequilana*. *GCB Bioenergy* **2011**, *3*, 25–36. [CrossRef]

64. Tamayo-Ordóñez, M.C.; Rodriguez-Zapata, L.C.; Narváez-Zapata, J.A.; Tamayo-Ordóñez, Y.J.; Ayil-Gutiérrez, B.A.; Barredo-Pool, F.; Sánchez-Teyer, L.F. Morphological features of different polyploids for adaptation and molecular char-acterization of CC-NBS-LRR and LEA gene families in *Agave* L. *J. Plant Physiol.* **2016**, *195*, 80–94. [CrossRef] [PubMed]

65. Vanholme, R.; Cesarino, I.; Rataj, K.; Xiao, Y.; Sundin, L.; Goeminne, G.; Kim, H.; Cross, J.; Morreel, K.; Araujo, P.; et al. Caffeoyl Shikimate Esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science* **2013**, *341*, 1103–1106. [CrossRef] [PubMed]

66. Ha, C.M.; Escamilla-Trevino, L.; Yarce, J.C.S.; Kim, H.; Ralph, J.; Chen, F.; Dixon, R.A. An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. *Plant J.* **2016**, *86*, 363–375. [CrossRef] [PubMed]

67. Vogt, T. Phenylpropanoid biosynthesis. *Mol. Plant* **2010**, *3*, 2–20. [CrossRef] [PubMed]

68. Tohge, T.; Watanabe, M.; Hoefgen, R.; Fernie, A.R. Shikimate and phenylalanine biosynthesis in the green lineage. *Front. Plant Sci.* **2013**, *4*, 62. [CrossRef]

69. Xue, J.-S.; Zhang, B.; Zhan, H.; Lv, Y.-L.; Jia, X.-L.; Wang, T.; Yang, N.-Y.; Lou, Y.-X.; Zhang, Z.-B.; Hu, W.-J.; et al. Phenylpropanoid derivatives are essential components of sporopollenin in vascular plants. *Mol. Plant* **2020**, *13*, 1644–1653. [CrossRef]

70. Subramanian, S.S.; Nair, A.G.R. Chlorogenin and kaempferol glycosides from the flowers of *Agave americana*. *Phytochemistry* **1970**, *9*, 2582. [CrossRef]

71. Almaraz-Abarca, N.; Delgado-Alvarado, E.A.; Hernández-Vargas, V.; Ortega-Chávez, M.; Orea-Lara, G.; Cifuentes-Díaz de León, A.; Ávila-Reyes, J.A.; Muñiz-Martínez, R. Profiling of phenolic compounds of somatic and reproductive tissues of *Agave durangensis* Gentry (*Agavaceae*). *Am. J. Appl. Sci.* **2009**, *6*, 1076–1085. [CrossRef]

72. Chen, Z.; Hong, X.; Zhang, H.; Wang, Y.; Li, X.; Zhu, J.-K.; Gong, Z. Disruption of the cellulose synthase gene, *AtCesA8/IRX1*, enhances drought and osmotic stress tolerance in Arabidopsis. *Plant J.* **2005**, *43*, 273–283. [CrossRef]

73. Müssig, J.; Slootmaker, T. Types of fibre. In *Industrial Application of Natural Fibres: Structure, Properties, and Technical Applications*; Müssig, J., Ed.; John Wiley & Sons: Chichester, UK, 2010.

74. Smith, R.A.; Schuetz, M.; Roach, M.; Mansfield, S.D.; Ellis, B.; Samuels, L. Neighboring parenchyma cells contribute to Arabidopsis xylem lignification, while lignification of interfascicular fibers is cell autonomous. *Plant Cell.* **2013**, *25*, 3988–3999. [CrossRef]

75. Chernova, T.E.; Gorshkova, T.A. Biogenesis of plant fibers. *Russ. J. Dev. Biol.* **2007**, *38*, 221–232. [CrossRef]

76. 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. Biotechnol.* **2011**, *29*, 644–652. [CrossRef] [PubMed]

77. Evans, T.; Loose, M. AlignWise: A tool for identifying protein-coding sequence and correcting frame-shifts. *BMC Bioinforma.* **2015**, *16*, 376. [CrossRef] [PubMed]

78. Van Dongen, S. Performance criteria for graph clustering and Markov cluster experiments. In *Technical Report INS-R 0012*; National Research Institute for Mathematics and Computer Science: Amsterdam, The Netherlands, 2000; Available online: http://ftp.cwi.nl//CWIreports/INS/INS-R0012.pdf (accessed on 1 April 2022).

79. Enright, J.A.; Van Dongen, S.; Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **2002**, *30*, 1575–1584. [CrossRef] [PubMed]

80. Li, L.; Stoeckert, C.J.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome. Res.* **2003**, *13*, 2178–2189. [CrossRef]

81. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [CrossRef] [PubMed]

82. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [CrossRef]

83. Stothard, P. The sequence manipulation suite: Javascript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **2000**, *28*, 1102–1104. [CrossRef]

84. 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]

85. Morais de Carvalho, D.; Martínez-Abad, A.; Evtuguin, D.V.; Colodette, J.L.; Lindström, M.E.; Vilaplana, F.; Sevastyanova, O. Isolation and characterization of acetylated glucuronoarabinoxylan from sugarcane bagasse and straw. *Carbohydr. Polym.* **2017**, *156*, 223–234. [CrossRef]

86. Morán-Velázquez, D.C.; Monribot-Villanueva, J.L.; Bourdon, M.; Tang, J.Z.; López-Rosas, I.; Maceda-López, L.F.; Villalpando-Aguilar, J.L.; Rodríguez-López, L.; Gauthier, A.; Trejo, L.; et al. Unravelling chemical composition of *Agave* spines: News from *Agave fourcroydes* Lem. *Plants* **2020**, *9*, 1642. [CrossRef] [PubMed]

87. Mellado-Mojica, E.; Seeram, N.P.; López, M.G. Comparative analysis of maple syrups and natural sweeteners: Carbohydrates composition and classification (differentiation) by HPAEC-PAD and FTIR spectroscopy-chemometrics. *J. Food Compos. Anal.* **2016**, *52*, 1–8. [CrossRef]

88. Sykes, R.; Yung, M.; Novaes, E.; Kirst, M.; Peter, G.; Davis, M. High-throughput screening of plant cell-wall composition using pyrolysis molecular beam mass spectroscopy. In *Biofuels: Methods and Protocols*; Mielenz, J.R., Ed.; Humana Press: Totowa, NJ, USA, 2009; Volume 581, pp. 169–183. [CrossRef]

89. Sarwar, M.B.; Ahmad, Z.; Anicet, B.A.; Sajid, M.; Rashid, B.; Hassan, S.; Ahmed, M.; Husnain, T. Identification and validation of superior housekeeping gene(s) for qRT-PCR data normalization in *Agave sisalana* (a CAM-plant) under abiotic stresses. *Physiol. Mol. Biol. Plants* **2000**, *26*, 567–584. [CrossRef]

90. Schmittgen, T.D.; Livak, K.J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **2008**, *3*, 1101–1108. [CrossRef] [PubMed]