



## Article

# Isolation and Expression of Transcription Factors Involved in Somatic Embryo Development by Transcriptome Analysis of Embryogenic Callus of *Thuja koraiensis*

Chang Ho Ahn <sup>1,2</sup> , Jung Yeon Han <sup>2</sup>, Hyeong Soo Park <sup>3</sup>, Hyun Won Yoon <sup>1</sup>, Jung Won Shin <sup>1</sup> , Jeong Min Seo <sup>1</sup>, Hana Lee <sup>1</sup>, Yeoung Ryul Kim <sup>1</sup>, Saeng Geul Baek <sup>1</sup>, Jae Ik Nam <sup>1</sup>, Jung Min Kim <sup>1</sup> and Yong Eui Choi <sup>2,\*</sup>

<sup>1</sup> Division of Garden Material Research, Korea Institute of Arboretum Management, Sejong-si 30129, Republic of Korea

<sup>2</sup> Department of Forest Resources, College of Forest and Environmental Sciences, Kangwon National University, Chuncheon-si 24341, Republic of Korea

<sup>3</sup> Habitat Conservation Division, Korea National Park Service, Gurye-gun 57616, Republic of Korea

\* Correspondence: yechoi@kangwon.ac.kr

**Abstract:** *Thuja koraiensis* Nakai (Cupressaceae) is an endangered and ecologically important conifer endemic to Korea. Previously, we established a protocol for micropropagation in *T. koraiensis*, which involved somatic embryogenesis from embryogenic callus of *T. koraiensis*. However, the molecular mechanisms underlying somatic embryogenesis remain unclear. Herein, we performed transcriptomic analysis to identify somatic embryogenesis-related genes of *T. koraiensis* via Illumina RNA sequencing. We conducted de novo transcriptome assembly using a Trinity assembler, which produced 274,077 transcript contigs clustered into 205,843 transcripts (unigenes), with an average length of 825 base pairs. Of all the unigenes, 14.69%, 18.62%, and 7.4% had homologs in the Gene Ontology, NCBI Non-redundant Protein, and NCBI Nucleotide databases, respectively. Among these mRNA sequences, expression of putative embryogenesis-associated transcription factors, namely *BABYBOOM (BBM)*, *WUSCHEL-RELATED HOMEODOMAIN (WOX)*, and *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE (SERK)*, was analyzed during somatic embryogenesis. RT-PCR analysis revealed that *TkBBM*, *TkWOX*, and *TkSERK* were highly expressed in embryogenic callus and seedling roots, whereas their expression was reduced in seedling leaves. Our findings provide new insights into the roles of *BBM*, *WOX*, and *SERK* in somatic embryogenesis. Our results may serve as a reference for comparative transcriptome analysis in related species and further aid functional genomics studies in conifers.

**Keywords:** Cupressaceae; somatic embryogenesis; Illumina RNA-Seq; de novo assembly; RT-PCR



**Citation:** Ahn, C.H.; Han, J.Y.; Park, H.S.; Yoon, H.W.; Shin, J.W.; Seo, J.M.; Lee, H.; Kim, Y.R.; Baek, S.G.; Nam, J.I.; et al. Isolation and Expression of Transcription Factors Involved in Somatic Embryo Development by Transcriptome Analysis of Embryogenic Callus of *Thuja koraiensis*. *Horticulturae* **2023**, *9*, 131. <https://doi.org/10.3390/horticulturae9020131>

Academic Editors: Gianni Barcaccia, Alessandro Vannozzi, Fabio Palumbo and Silvia Farinati

Received: 1 December 2022

Revised: 2 January 2023

Accepted: 14 January 2023

Published: 19 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Gymnosperms first appeared more than 300 million years ago, and currently, approximately 630 conifer species exist on Earth [1]. Conifers are important species of ecological and economic value. Cupressaceae, also known as the cypress family, is distributed throughout the world and comprises >160 species in 32 genera [2]. Many members of the cypress family serve as timber sources and ornamentals and are crucial for the forest ecosystem. Korean arborvitae (*Thuja koraiensis* Nakai), a short and creeping evergreen shrub, which reaches a height of about 3 m, is only found in Korea, predominantly in the high mountains. Recently, *T. koraiensis* was designated rare and endangered due to climate change.

To date, many angiosperm genomes have been sequenced, particularly *Arabidopsis thaliana* [3] and poplar, a woody plant [4]; however, owing to the large size of gymnosperm genomes, reference genomes are still lacking for gymnosperms. Genome assembly has been conducted for only a few members of the cypress family, including *Cunninghamia lanceolata* [5,6], *Juniperus chinensis* [7], *Platycladus orientalis* [8,9], and *T. plicata* [10].

For *T. koraiensis*, Hou et al. [11] performed de novo genome assembly using restriction-associated DNA to develop SSR markers. Nonetheless, no studies have focused on transcriptome analysis in *T. koraiensis*.

Recently, we established a protocol for micropropagation in *T. koraiensis*, which involves somatic embryogenesis (SE) from embryogenic callus [12]. In SE from embryogenic callus, drastic changes take place in the transcriptome, and these changes promote embryogenic potential in plant cells. Among the various factors that are involved in this developmental switch, genes encoding transcription factors (TFs), which constitute sequence-specific DNA-binding proteins and regulate gene expression, are well known to play a central role in the developmental process. Although initiation of SE has been found in many conifers, the molecular mechanisms underlying SE remain unclear. Analyses of gene expression during SE can provide insights into its complex process [13].

TFs, namely *BABYBOOM* (*BBM*) and *WUSCHEL-RELATED HOMEODOMAIN* (*WOX*), and transmembrane kinase, namely *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE* (*SERK*), are key regulators of SE initiation and have been identified in both angiosperms and gymnosperms. *BBM* belongs to the *APETALA2* (*AP2*) subfamily of the *ETHYLENE-RESPONSE-FACTOR* (*ERF*) TFs [14] and is a key regulator of plant cell totipotency. Moreover, *BBM* has been shown to play a role in developmental pathways associated with cell proliferation and growth [15]. *WOX* forms a large family of genes in plants and has important functions during all stages of plant development [16]. Recently, several members of the *WOX* gene family were reported to be expressed in the early stages of embryogenesis. Haecker et al. [17] and Rupps et al. [18] reported that *WOX* genes were highly expressed during early somatic embryo development in *A. thaliana* and *Larix decidua*, respectively. Similarly, in *Picea abies*, *PaWOX2* expression was demonstrated to be the highest at the earliest stages of somatic embryo development, whereas low levels were observed in seedling tissues [19]. Moreover, *PaWOX2* expression has not been observed in non-embryogenic cell cultures. Therefore, *WOX* genes can serve as potential markers for identifying embryogenic potential. *SERK*, which encodes a leucine-rich repeat containing receptor-like kinase protein, plays a crucial role in the transduction of extracellular signals by phosphorylating intracellular target proteins during SE [20]. *SERK* is expressed in the embryogenic callus of carrots [20] and grapevine [21]. Expression of *SERK* genes in other plant species established a broader expression pattern extending to non-embryogenic, embryogenic, and vegetative tissues [22].

Due to its accuracy and low cost, high-throughput RNA sequencing (RNA-Seq) has been broadly used to analyze transcriptomes both quantitatively and qualitatively. Furthermore, the use of transcriptome sequencing technologies has increased with advancements in genomic technologies. For this reason and to gain insights on the functional genes involved in SE, in the current study, we performed de novo sequencing, assembly, and annotation of the genome sequence of *T. koraiensis* embryogenic callus using Illumina paired-end sequencing technology. Subsequently, on the basis of bioinformatics analysis of assembled transcriptome data, we selected potential embryogenesis-associated genes, namely *BBM*, *WOX*, and *SERK*, and analyzed their functions during SE. Taken together, our findings may serve as a comprehensive sequence resource for other related species and contribute to the understanding of the process of SE in *T. koraiensis*.

## 2. Materials and Methods

### 2.1. Plant Material and Initiation of Embryogenic Culture

In 2018, immature cones were collected from three open-pollinated *T. koraiensis* trees growing in Mt. Seorak (Gangwon province, Korea; at approximately 1590 m; 38°12'23'' latitude, 128°45'79'' longitude) during the northern hemisphere summer. Cones were dissected, immature seeds were extracted, and surfaces of whole seeds were disinfested as follows in the given order: 70% ethanol rinse for 1 min, sterile water rinse for 3 min, 50% Clorox rinse (approximately 4% sodium hypochlorite) for 1 min, and sterile water rinse five times. Seed coats were removed, and whole megagametophytes containing zygotic

embryos were cultured on a semisolid initiation medium (IM) [23], which was modified from an EM medium [24], in Petri dishes (60 × 15 mm). Embryogenic callus lines were maintained by transferring the calluses to fresh IM supplemented with 3% sucrose, 2.25 μM 2,4-dichlorophenoxyacetic acid, and 1.1 μM 6-benzyladenine every 2 weeks.

### 2.2. TruSeq Stranded mRNA and RNA-Seq

For Illumina sequencing, total RNA from embryogenic callus was isolated using Trizol reagent (MRC, Cincinnati, OH, USA) and RNeasy<sup>®</sup> Plant Mini Kit (QIAGEN, Hilden, Germany). Genomic DNA was removed from the total RNA using DNase I (Takara Bio, Shiga, Japan). Prior to cDNA synthesis, RNA quality and quantity were verified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and a NanoDrop<sup>™</sup> 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). RNA integrity was confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies) with a minimum RNA integrity number value  $\geq 7$ . Total RNA was converted into cDNA libraries using the TruSeq Stranded mRNA Sample Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. The cDNA libraries were analyzed for size distribution, quantitated by qPCR (Kapa Library Quantification Kit; Kapa Biosystems, Wilmington, MA, USA), normalized to 2 nmol/L, and then pooled in equimolar amounts for RNA-Seq. Sequencing runs were performed on the HiSeq 4000 (Illumina) according to the method previously described by Bodian et al. [25].

### 2.3. Sequence Trimming, De Novo Transcriptome Assembly, and Annotation

The raw reads were first trimmed to remove the low-quality reads using Trimmomatic software version 0.36 (<http://www.usadellab.org/cms/index.php?page=trimmomatic>, accessed on 26 June 2018) [26]. The trimmed reads for all samples were pooled and assembled using Trinity software (version r20140717, bowtie 1.1.2), which is generally utilized for the de novo reconstruction of transcriptomes, to obtain high-quality transcript sequences. For gene assembly, the longest contig of the assembled contigs was filtered and clustered into non-redundant transcripts using the CD-HIT-EST program (version 4.6). Subsequently, unigenes were obtained. For the functional annotation of these unigenes, we searched various public databases, including NCBI Nucleotide (NT; <http://www.ncbi.nlm.nih.gov/nucleotide/>, accessed on 25 June 2018), NCBI Non-redundant Protein (NR; <http://www.ncbi.nlm.nih.gov/protein/>, accessed on 16 June 2018), Protein Families (Pfam; <http://pfam.xfam.org/>, accessed on 25 June 2018), Gene Ontology (GO; <http://www.geneontology.org/>, accessed on 20 June 2018), Universal Protein Resource (UniProt; <http://www.uniprot.org/>, accessed on 26 June 2018), and EggNOG (<http://eggnogdb.embl.de/>, accessed on 15 June 2018), using BLASTN of NCBI BLAST (version 2.4.0+) and BLASTX of DIAMOND software (version 0.9.21; <http://github.com/bbuchfink/diamond>, accessed on 25 June 2018) with an *E*-value default cut-off of  $1.0 \times 10^{-5}$ . The abundance of the unigenes across samples was estimated using the RSEM algorithm (version v1.2.29, bowtie 1.1.2). The expression levels were calculated as read counts.

### 2.4. Gene Selection and Phylogenetic Analysis of Selected mRNA Sequences

The embryogenesis-associated genes, *BBM*, *WOX*, and *SERK*, were screened using the sequenced transcriptome and NR database. Full-length *BBM*, *WOX*, and *SERK* sequences of other species were obtained from Genbank Non-Redundant database using TBLASTN with TkBBM ("TkBBM1" and "TkBBM2"), TkWOX ("TkWOX1" and "TkWOX2"), and TkSERK ("TkSERK1", "TkSERK2", and "TkSERK3") proteins of *T. koraiensis* as queries. The sequence alignment of *T. koraiensis* proteins and other species (*BBM*, *WOX*, and *SERK* sequences) was compared via phylogenetic tree analysis using the neighbor-joining method with the MEGA v. 7.0.21 (DNA Star, Madison, WI, USA).

### 2.5. RT-PCR Analysis

To assess gene expression in *T. koraiensis* during SE, we collected embryogenic calluses at three different developmental stages and gathered somatic seedling leaves and roots. Total RNA was isolated from the samples using the RNeasy Plant Mini Kit (QIAGEN), and reverse transcription was conducted with the ImProm-II™ Reverse Transcription System (Promega, Madison, WI, USA). As per the RT-PCR, amplification conditions were adjusted as follows: 95 °C for 5 min; 35 cycles of 95 °C for 30 s, 57 °C for 30 s, and 72 °C for 1 min; and a final 7 min extension at 72 °C. The PCR-amplified products were analyzed using 1% (*w/v*) agarose gel electrophoresis, wherein the gel was previously stained with Red-Safe™ DNA dye (iNtRON Biotechnology, Sungnam, Korea) and imaged under GelDoc-ItTS2 Gel Imager (Analytik Jena US LLC, Upland, CA, USA). The intensity of the bands was quantified at least thrice using ImageJ software (National Institutes of Health, Bethesda, MD, USA). To assess RNA integrity, equivalent concentrations of cDNA were utilized; furthermore, actin was used as the internal control. The sequences of all RT-PCR primers are listed in Table S1.

## 3. Results and Discussion

### 3.1. RNA-Seq and De Novo Assembly

A cDNA library constructed from the total RNA extracted from embryogenic cells of *T. koraiensis* was sequenced on an Illumina sequencing platform after validation. To reduce bias in analysis, artifacts, such as low-quality reads, adaptor sequences, contaminant DNA, or PCR duplicates, were removed. Sequencing of *T. koraiensis* produced 250,474,062 reads, which were reduced to 246,686,422 after quality control (Table 1). After trimming to remove low-quality reads, the averages of Q20, Q30, and GC were 98.99, 96.87, and 42.86%, respectively. The de novo transcriptome assembly of *T. koraiensis* performed using the Trinity assembler produced a total of 214,470 transcripts, with an average contig length of 738 nt. Information regarding the number of read bases and the average contig length is shown in Table 2. In total, 205,843 unigenes were obtained, with an average contig length of 619 nt, an N50 of 931 nt, and an N90 of 256 nt. Among the obtained unigenes, 63,599 (30.9%) were  $\geq 500$  nt in length, and 29,113 (14.14%) were  $\geq 1000$  nt in length. The total length of the assembled sequence was 127,425,530 nt.

**Table 1.** The *T. koraiensis* raw sequenced data and trimmed sequenced data.

Types of Data	Total Number of Read Bases	Total Number of Reads	GC <sup>a</sup> (%)	Q20 <sup>b</sup> (%)	Q30 <sup>c</sup> (%)
Raw data statistics	25,297,880,262	250,474,062	42.90	98.55	96.06
Trimming data statistics	24,768,744,102	246,686,422	42.86	98.99	96.87

<sup>a</sup> GC content. <sup>b</sup> Ratio of bases that have a Phred quality score greater than or equal to 20. <sup>c</sup> Ratio of bases that have a Phred quality score greater than or equal to 30.

**Table 2.** Summary data of the assembled transcripts and unigenes in *T. koraiensis*.

Assembly	Transcripts	Unigenes
Total number of genes	214,470	205,843
GC <sup>a</sup> (%)	38.79	38.35
Contig N90 <sup>b</sup>	278	256
Contig N50 <sup>c</sup>	1345	931
Median contig length	371.0	340.0
Average contig length	738.06	619.04
Total assembled bases	202,286,040	127,425,530

<sup>a</sup> GC content. <sup>b</sup> At least 90% of the assembled transcript nucleotides are found in contigs with lengths  $\geq N90$ . <sup>c</sup> At least 50% of the assembled transcript nucleotides are found in with lengths  $\geq N50$ .

### 3.2. Functional Annotation

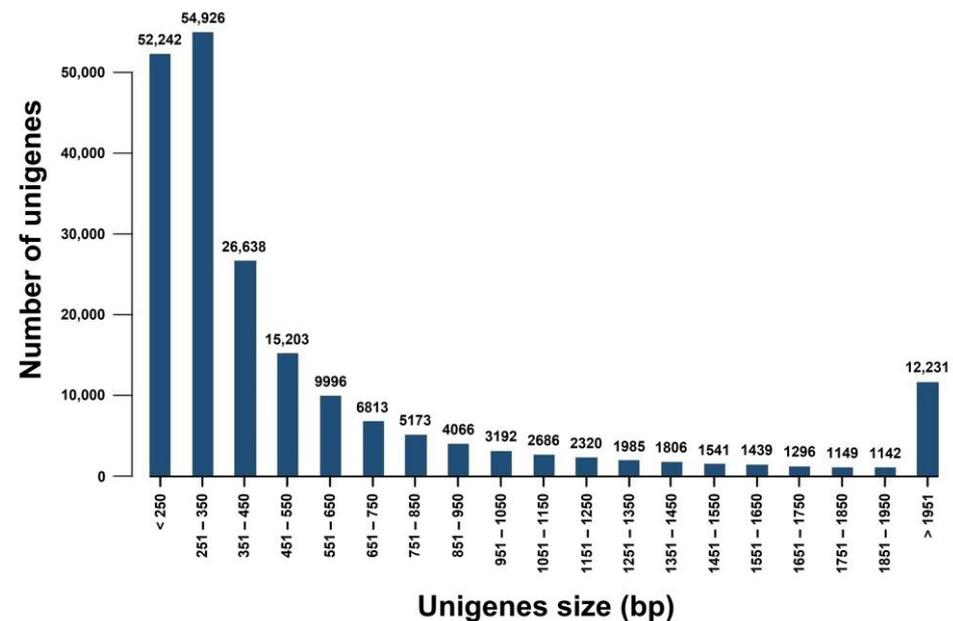
The unigenes of the transcriptome were annotated, and sequence similarity analysis was conducted using six databases: NT, NR, Pfam, GO, UniProt, and EggNOG. The greatest

number of annotations were retrieved from the NR database, i.e., 38,320 (18.62%), and the least were retrieved from the NT database, i.e., 15,237 (7.4%) (Table 3). Furthermore, 35,511 (17.25%) unigenes were annotated using EggNOG, and 27,659 (13.44%) unigenes were annotated using Pfam. Additionally, 26,274 (12.76%) unigenes were annotated using the UniProt databases. The size distribution of the unigenes is shown in Figure 1.

**Table 3.** Summary data of the annotation using various databases in *T. koraiensis*.

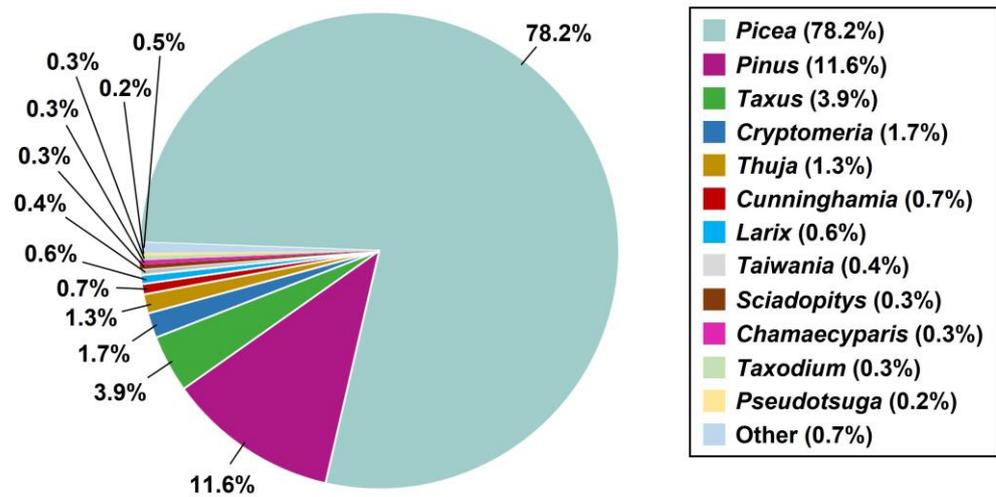
Annotation Database	The Number of Annotations	Annotation (%)
GO <sup>a</sup>	30,241	14.69
UniProt <sup>b</sup>	26,274	12.76
NR <sup>c</sup>	38,320	18.62
Pfam <sup>d</sup>	27,659	13.44
EggNOG	35,511	17.25
NT <sup>e</sup>	15,237	7.40
Overall	40,461	19.66

<sup>a</sup> Gene Ontology. <sup>b</sup> Universal Protein Resource. <sup>c</sup> NCBI non-redundant Protein. <sup>d</sup> Protein families. <sup>e</sup> NCBI Nucleotide. The annotations were blasted against public databases with a BLASTN and BLASTX, including NT, Pfam, GO, NR, UniProt, and EggNOG.



**Figure 1.** Length distribution of all unigenes.

The top-scoring BLASTX hits against the NR protein database showed that 78.2% of the unigenes were present in the genus *Picea*, 11.6% were present in the genus *Pinus*, and 3.9% were present in the genus *Taxus* (Figure 2). Moreover, in the cypress family (Cupressaceae), 1.7% of the unigenes were present in the genus *Cryptomeria*, and 1.3% and 0.7% of the unigenes were present in the genera *Thuja* and *Cunninghamia*, respectively. The most abundant transcript, with 5961 reads, was annotated as a hypothetical protein of *Citrus unshiu*. The 20 most abundant transcripts are listed in Table 4. Two unigenes matched the auxin transport protein in *Amborella trichopoda* and *Quercus suber*.



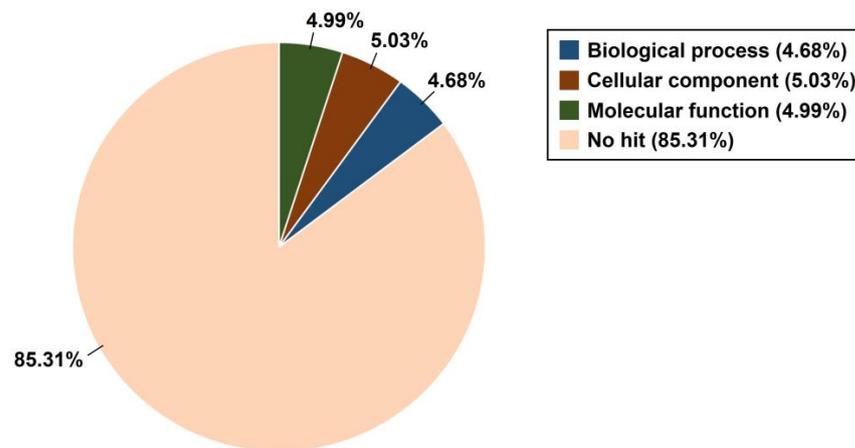
**Figure 2.** Distribution of the top BLASTX hits of conifers for unigenes in the NCBI Non-redundant database.

**Table 4.** Most abundant transcripts in the *T. koraiensis* transcriptome.

Transcript ID	Read	Target Description (Species)	Identity (%)	E Value
c121998_g2_i1	5961	Hypothetical protein ( <i>Citrus unshiu</i> )	70	$3 \times 10^{-127}$
c137033_g1_i3	5756	Hypothetical protein ( <i>Marchantia polymorpha</i> )	49	0
c119709_g1_i3	5619	Hypothetical protein ( <i>Physcomitrella patens</i> )	44	$2 \times 10^{-90}$
c125131_g1_i1	5427	Hypothetical protein ( <i>Plasmodium cynomolgi</i> strain B)	24	0.0006
c202221_g1_i1	5405	PREDICTED: Midasin ( <i>Ipomoea nil</i> )	71	$3 \times 10^{-10}$
c124649_g1_i2	5108	Auxin transport protein ( <i>Amborella trichopoda</i> )	60	0
c69094_g1_i1	5083	Auxin transport protein ( <i>Quercus suber</i> )	71	$5 \times 10^{-89}$
c115060_g1_i1	4787	Zinc finger protein ( <i>Macleaya cordata</i> )	56	0
c121372_g1_i1	4506	PREDICTED: A-kinase anchor protein 9-like isoform X5 ( <i>Musa acuminata</i> subsp. <i>malaccensis</i> )	24	$1 \times 10^{-8}$
c139228_g2_i8	4360	Uncharacterized protein ( <i>Amborella trichopoda</i> )	50	0
c137890_g3_i2	4321	Hypothetical protein ( <i>Marchantia polymorpha</i> )	53	$8 \times 10^{-191}$
c132000_g1_i1	4294	Peroxin/Ferlin domain ( <i>Macleaya cordata</i> )	54	0
c130259_g1_i3	4152	Hypothetical protein ( <i>Marchantia polymorpha</i> )	46	0
c138002_g5_i3	4128	Hypothetical protein ( <i>Vitis vinifera</i> )	32	$5 \times 10^{-166}$
c103923_g2_i1	4128	Hypothetical protein ( <i>Vitis vinifera</i> )	34	$7 \times 10^{-21}$
c130469_g2_i4	4105	PREDICTED: Chromatin structure-remodeling complex protein SYD isoform X1 ( <i>Nelumbo nucifera</i> )	55	0
c134892_g4_i2	3977	Hypothetical protein ( <i>Marchantia polymorpha</i> )	75	0
c87277_g1_i1	3961	Transformation/transcription domain associated protein ( <i>Trifolium pratense</i> )	37	0.00008
c134892_g6_i1	3896	PREDICTED: Transformation/transcription domain-associated protein-like ( <i>Nelumbo nucifera</i> )	76	$6 \times 10^{-120}$
c114560_g2_i1	3878	PREDICTED: Uncharacterized protein ( <i>Nicotiana tabacum</i> )	33	$3 \times 10^{-32}$

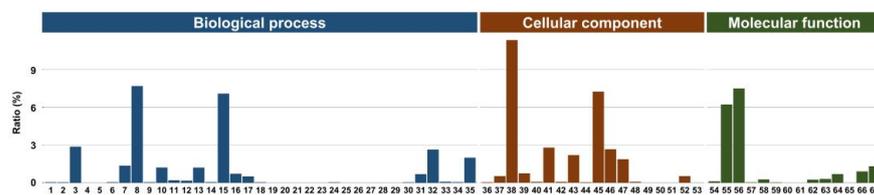
### 3.3. Functional Classification

For functional classification of *T. koraiensis* unigenes, the GO database was employed along with BLASTX of DIAMOND with an E-value cut-off of  $1.0 \times 10^{-5}$ . The GO terms belonging to biological processes, cellular components, and molecular functions were listed. Subsequently, 30,241 (14.69%) of the total unigenes were assigned GO terms, with 4.68% assignments from biological processes, 5.03% from cellular components, and 4.99% from molecular functions (Figure 3).



**Figure 3.** Distribution of the Gene Ontology terms for the three categories (biological process, cellular component, and molecular function).

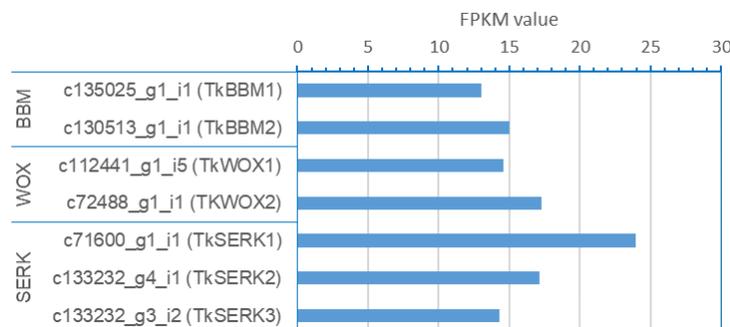
In addition, GO functional classification showed 67 functional groups (Figure 4). The category of biological process comprises a large number of unique sequences that are classified into 35 subcategories (Figure S1). The majority of the sequences in biological processes were involved in metabolic processes (35.61%), followed by those involved in cellular processes (13.09%). The unique sequences were grouped into 18 GO classifications under the cellular component category (Figure S2). Among them, cell part (45.99%) and organelle (22.3%) were the most represented cellular components. The molecular function category included 14 GO classifications (Figure S3). The two largest classifications of molecular function included catalytic activity (48.2%) and binding (33.45%).



**Figure 4.** Bar diagram of unigenes assigned to 67 Gene Ontology (GO) categories. GO terms of biological process: 1 behavior, 2 biological adhesion, 3 biological regulation, 4 carbon utilization, 5 cell killing, 6 cell proliferation, 7 cellular component organization or biogenesis, 8 cellular process, 9 detoxification, 10 developmental process, 11 growth, 12 immune system process, 13 localization, 14 locomotion, 15 metabolic process, 16 multicellular organismal process, 17 multi-organism process, 18 nitrogen utilization, 19 obsolete chloroplast ribulose biphosphate carboxylase complex biogenesis, 20 obsolete MAPK import into nucleus, 21 obsolete mitochondrial respiratory chain complex I biogenesis, 22 obsolete mitochondrial respiratory chain complex IV biogenesis, 23 obsolete mycelium development, 24 obsolete protein import into nucleus and docking, 25 obsolete RNA polymerase II complex import into nucleus, 26 obsolete RNA polymerase III complex import into nucleus, 27 obsolete transcription factor import into nucleus, 28 pigmentation, 29 presynaptic process involved in chemical synaptic transmission, 30 reproduction, 31 reproductive process, 32 response to stimulus, 33 rhythmic process, 34 signaling, and 35 unclassified. Cellular component: 36 cells, 37 cell junction, 38 cell part, 39 extracellular region, 40 extracellular region part, 41 membrane, 42 membrane-enclosed lumen, 43 membrane part, 44 nucleoid, 45 organelle, 46 organelle part, 47 protein-containing complex, 48 supramolecular complex, 49 symplast, 50 synapse, 51 synapse part, 52 unclassified, and 53 virion part. Molecular function: 54 antioxidant activity, 55 binding, 56 catalytic activity, 57 molecular carrier activity, 58 molecular function regulator, 59 nutrient reservoir activity, 60 obsolete cyclic pyranopterin monophosphate synthase activity, 61 protein tag, 62 signal transducer activity, 63 structural molecule activity, 64 transcription regulator activity, 65 translation regulator activity, 66 transporter activity, and 67 unclassified.

### 3.4. Identification and Sequence Analysis of TkBBM, TkSERK, and TkWOX

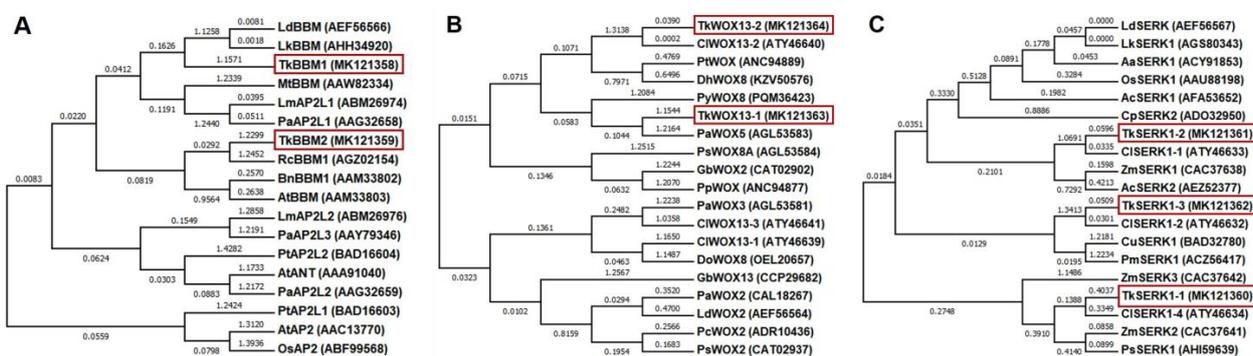
We sought to identify sequences in the transcriptome that encoded authentic *BBM*, *WOX*, and *SERK*. FPKM values of selected contigs for *BBM*, *WOX*, and *SERK* were shown in Figure 5.



**Figure 5.** FPKM values of selected contigs of *BBM*, *SERK*, and *WOX* in annotated transcriptome sequences of *T. koraiensis* embryonic callus.

Then, we selected full-length sequences of the coding regions for putative homologs of *BBM*, *WOX*, and *SERK* of *T. koraiensis*. The genes were referred to as *TkBBM* (*TkBBM1* and *TkBBM2*), *TkWOX* (*TkWOX1* and *TkWOX2*), and *TkSERK* (*TkSERK1*, *TkSERK2*, and *TkSERK3*). To examine the phylogenetic similarity between *T. koraiensis* and other plants, we compared the sequences of 57 highly conserved proteins of 28 species.

*BBM* has been reported to be involved in embryogenesis in various plant species, including *A. thaliana* [27], *Brassica napus* [27], *Glycine max* L. [28], *Capsicum annuum* L. [29], *L. decidua* [18], and *Zea mays* [30]. In this study, the open reading frame (ORF) of *TkBBM1* spanned 2325 base pairs (bp) and was predicted to encode a protein of 774 amino acids (aa). The ORF of *TkBBM2* spanned 1791 bp (596 aa). Phylogenetic tree analysis showed the evolutionary relationships between *TkBBM* (*TkBBM1* and *TkBBM2*) and 11 plant species (Figure 6A). *TkBBM* showed a high similarity with *BBM* from other plant species. *TkBBM1* was clearly grouped with *LdBBM* (*L. decidua*, Acc. AEF56566) and *LkBBM* (*Larix gmelinii* var. *olgensis* × *L. kaempferi aintegumenta*; Acc. AHH34920), whereas *TkBBM2* was observed to be neighboring *RcBBM1* (*Rosacanina*, Acc. AGZ02154). Rupps et al. [18] reported that *LdBBM* showed a similarity of 97.6% to its homolog, *LkBBM*. These findings are similar to our results.



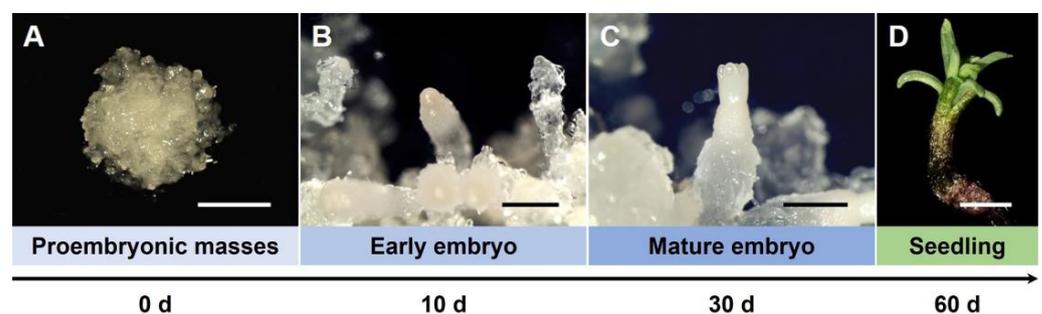
**Figure 6.** Phylogenetic tree of the putative *TkBBM* (*TkBBM1* and *TkBBM2*), *TkWOX* (*TkWOX1* and *TkWOX2*), and *TkSERK* (*TkSERK1*, *TkSERK2*, and *TkSERK3*) proteins based on the NCBI database and representatives characterizing each protein from broad plant lineages. (A) *TkBBM1* and *TkBBM2*; (B) *TkWOX1* and *TkWOX2*; (C) *TkSERK1*, *TkSERK2*, and *TkSERK3*.

Previously, Hedman et al. cloned *WOX* genes in *P. abies* and found that Norway spruce *WOX* genes are represented in all three major clades identified, and that the major diversification within the *WOX* gene family took place before the split between gymnosperms and angiosperms [31]. In this study, the ORF of *TkWOX1* spanned 705 bp and was predicted to encode a 234 aa protein. The ORF of *TkWOX2* spanned 963 bp (320 aa). As per *TkWOX* (*TkWOX1* and *TkWOX2*), eight gymnosperms and three angiosperms were clustered (Figure 6B). *TkWOX1* was closely related to *PaWOX5* (*P. abies*, Acc. AGL53583), and *TkWOX2* was observed to neighbor *CIWOX13-2* (*C. lanceolata*, Acc. ATY46640). According to Van der Graaff et al. [32], the *WOX* gene family was divided into three major clades, namely the modern, intermediate, and ancient clades.

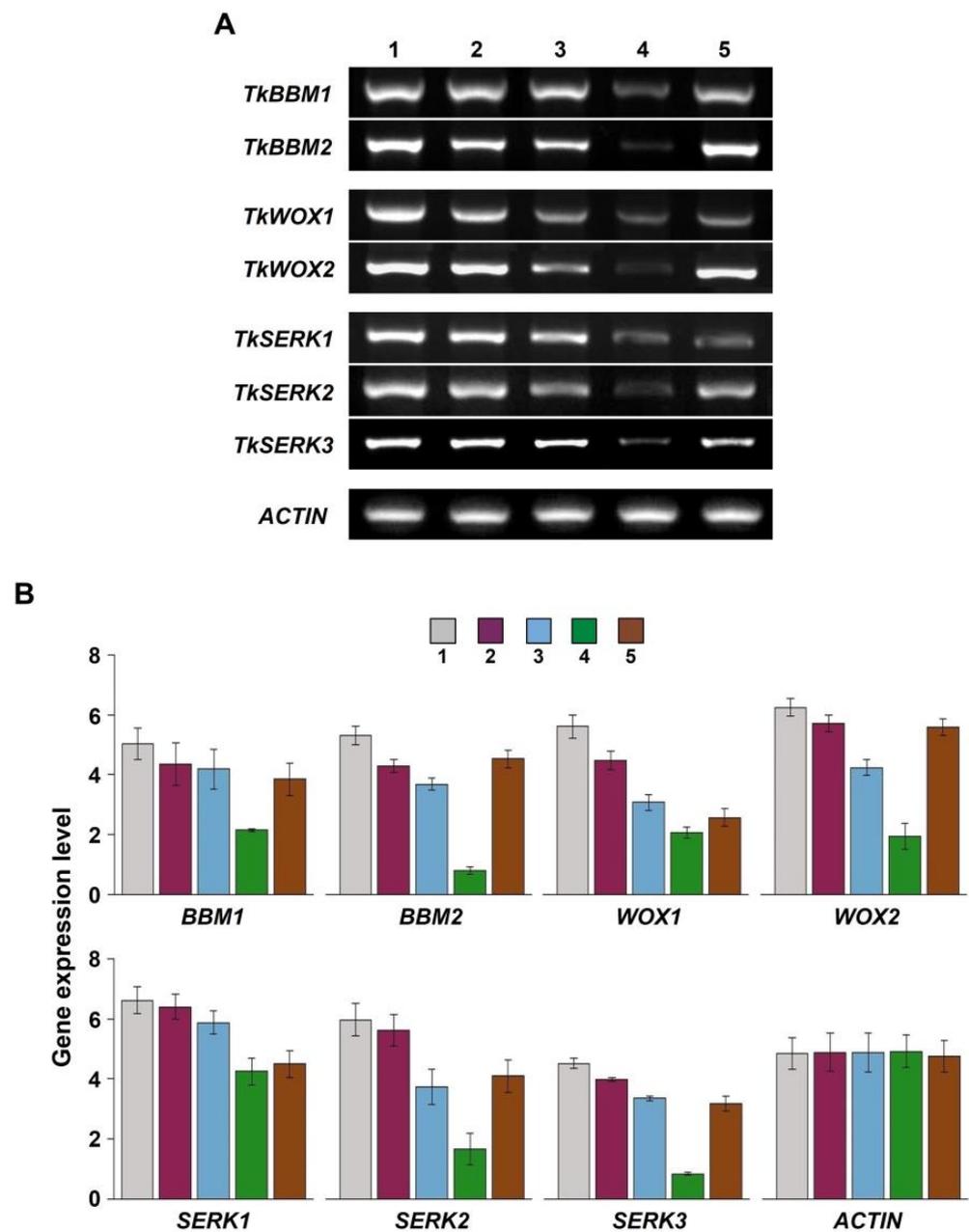
Five genes belonging to the *SERK* family in *A. thaliana* were identified. *SERK1* overexpression has been observed to result in a strong expression of embryogenic competence in transgenic lines [33]. Our results showed that the ORF of *TkSERK1* and *TkSERK2* comprised 1899 bp (632 aa) and 1827 bp (608 aa), respectively, and that the ORF of *TkSERK3* spanned 1845 bp, encoding a 614 aa protein. Expression of *LdSERK* in European larch (*L. decidua*) has been observed to be increased during late embryogenesis [18]. Our results showed that 11 plant species were found to cluster with the *TkSERK* genes (*TkSERK1*, *TkSERK2*, and *TkSERK3*) (Figure 6C). All three *TkSERK* protein sequences showed high similarity to *C. lanceolata* protein sequences (*CISERK1-1*, Acc. ATY46633; *CISERK1-2*, Acc. ATY46632; *CISERK1-4*, ATY46634).

### 3.5. Expression of *TkBBM*, *TkWOX*, and *TkSERK* during SE

For RT-PCR, embryogenic calluses at three different developmental stages and somatic seedling leaves and roots were collected (Figure 7). The expression of *TkBBM*, *TkWOX*, and *TkSERK* during SE was analyzed (Figure 8). In the present study, we observed that the expression of *TkBBM1* and *TkBBM2* were reduced during SE and in somatic seedling leaves. However, both genes showed high expression in somatic seedling roots. These results are in contrast with results reported for *L. decidua*. Rupps et al. [18] reported that the expression of *LdBBM* increased during embryogenesis. Li et al. [34] reported that *LkBBM* showed strong expression in the roots of *L. kaempferi* and *L. olgensis*. In addition, *BBM* groups with *PLT1* and *PLT2* in the AP2/ERF TF family have been reported to be expressed in the *A. thaliana* root [35]. The *BBM* gene was also identified as an auxin-inducible gene in *Medicago truncatula* roots [36]. Given our results and those of previous studies, we speculate that *TkBBM* plays an important role in regulating the development and growth of roots in *Thuja* spp.



**Figure 7.** Characteristic stages of somatic development of *T. koraiensis* embryos and seedling. (A) Proembryonic masses (PEMs) on maturation medium. Scale bar: 5 mm. (B) Early embryo on maturation medium after 10 days. Scale bar: 1 mm. (C) Fully mature cotyledonary embryo on maturation medium after 30 days. Scale bar: 2 mm. (D) Somatic seedling with the first true needles emerging and a taproot on germination medium after 60 days. Scale bar: 1 mm.



**Figure 8.** Expression analysis of the identified genes during the development of *T. koraiensis* somatic embryos. **(A)** RT-PCR analysis of somatic embryos of different explants. **(B)** Gene expression levels were analyzed, and intensity of bands was quantified by ImageJ densitometry analysis. Gray—proembryonic masses; purple—early embryo; blue—fully mature cotyledonary embryo; green—somatic seedling leaf; and brown—somatic seedling root.

*TkWOX1* and *TkWOX2* were observed to be highly expressed either in embryogenic calluses or at the early stage of somatic embryo development. The expression decreased in parallel with somatic embryo development. *TkWOX1* was similarly expressed in seedling leaves and roots, whereas *TkWOX2* showed faint expression in seedling leaves and strong expression in seedling roots. In the case of *P. abies*, *PaWOX8/9* was observed to be highly expressed in early embryo stages; however, a notably low expression was observed in the mature embryo stage [16]. Similar results were reported for *L. decidua* [18]. *TkSERK1*, *TkSERK2*, and *TkSERK3* were also observed to be highly expressed in early embryogenesis. As compared with the others, *TkSERK1* showed different expression patterns, i.e., uniform expression in seedling leaves and roots. Similarly, *TkSERK2* and *TkSERK3* showed similar

expression patterns. In particular, low expression was detected in seedling leaves. In contrast, *P. glauca* *SERK1* showed high expression in shoot bud [37]. Only few studies have reported the involvement of *SERK* genes in nodulation.

Owing to the high FPKM values of *TkBBMs*, *TkWOXs*, and *TkSERKs*, as seen in Figure 5, we speculate that *TkBBM*, *TkWOX*, and *TkSERK* are SE-related genes, possibly playing key roles in the SE of *T. koraiensis*. Our study serves as a valuable genetic resource for gene discovery, functional genomics, and comparative genomics in *T. koraiensis*. The *T. koraiensis* sequence resource reported in our study may serve as a reference for comparative transcriptome analysis in related species and aid functional genomics and association studies in conifers. In addition, the findings of the present study enhance our understanding of the transcriptome of *T. koraiensis* during SE.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/horticulturae9020131/s1>, Figure S1: Distributions of the GO terms for the biological processes; Figure S2: Distributions of the GO terms for the cellular processes; Figure S3: Distributions of the GO terms for the molecular processes; Table S1: List of primer sequences used for RT-PCR analysis.

**Author Contributions:** Conceptualization and funding acquisition, Y.E.C.; writing—original draft, C.H.A.; formal analysis, J.Y.H.; resources, H.S.P.; methodology, H.W.Y., J.W.S., J.M.S., H.L. and Y.R.K.; software, S.G.B., J.I.N. and J.M.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the R&D Program for Forest Science Technology provided by the Korea Forest Service (Korea Forestry Promotion Institute), grant number FTIS 2018131B10-1820-BB01 and 2021339A00-2123-CD02.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We would like to thank the National Park Research Institute of Korea for providing *T. koraiensis* seeds.

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

## References

- Nystedt, B.; Street, N.R.; Wetterbom, A.; Zuccolo, A.; Lin, Y.C.; Scofield, D.G.; Vezzi, F.; Delhomme, N.; Giacomello, S.; Alexeyenko, A.; et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* **2013**, *497*, 579–584. [[CrossRef](#)] [[PubMed](#)]
- Farjon, A. *A Monograph of Cupressaceae and Sciadopityes*; Royal Botanic Gardens, Kew: Kew, UK, 2005.
- Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **2000**, *408*, 796–815. [[CrossRef](#)] [[PubMed](#)]
- Tuskan, G.A.; Difazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, N.; Ralph, S.; Rombauts, S.; Salamov, A.; et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **2006**, *313*, 1596–1604. [[CrossRef](#)] [[PubMed](#)]
- Huang, H.H.; Xu, L.L.; Tong, Z.K.; Lin, E.P.; Liu, Q.P.; Cheng, L.J.; Zhu, M.Y. De novo characterization of the Chinese fir (*Cunninghamia lanceolata*) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. *BMC Genom.* **2012**, *13*, 648. [[CrossRef](#)]
- Wang, Z.; Chen, J.; Liu, W.; Luo, Z.; Wang, P.; Zhang, Y.; Zheng, R.; Shi, J. Transcriptome characteristics and six alternative expressed genes positively correlated with the phase transition of annual cambial activities in Chinese fir (*Cunninghamia lanceolata* (Lamb.) hook). *PLoS ONE* **2013**, *8*, e71562. [[CrossRef](#)]
- Hu, X.G.; Liu, H.; Zhang, J.Q.; Sun, Y.Q.; Jin, Y.Q.; Zhao, W.; El-Kassaby, Y.A.; Wang, X.R.; Mao, J.F. Global transcriptome analysis of *Sabina chinensis* (Cupressaceae), a valuable reforestation conifer. *Mol. Breed.* **2016**, *36*, 99. [[CrossRef](#)]
- Hu, X.G.; Liu, H.; Jin, Y.; Sun, Y.Q.; Li, Y.; Zhao, W.; El-Kassaby, Y.A.; Wang, X.R.; Mao, J.F. De novo transcriptome assembly and characterization for the widespread and stress-tolerant conifer *Platycladus orientalis*. *PLoS ONE* **2016**, *11*, e0148985. [[CrossRef](#)]
- Chang, E.; Zhang, J.; Deng, N.; Yao, X.; Liu, J.; Zhao, X.; Jiang, Z.; Shi, S. Transcriptome differences between 20- and 3,000-year-old *Platycladus orientalis* reveal that ROS are involved in senescence regulation. *Electron. J. Biotechnol.* **2017**, *29*, 68–77. [[CrossRef](#)]
- Shalev, T.J.; Yuen, M.M.S.; Gesell, A.; Yuen, A.; Russell, J.H.; Bohlmann, J. An annotated transcriptome of highly inbred *Thuja plicata* (Cupressaceae) and its utility for gene discovery of terpenoid biosynthesis and conifer defense. *Tree Genet. Genomes* **2018**, *14*, 35. [[CrossRef](#)]

11. Hou, L.; Cui, Y.; Li, X.; Chen, W.; Zhang, Z.; Pang, X.; Li, Y. Genetic evaluation of natural populations of the endangered conifer *Thuja koraiensis* using microsatellite markers by restriction-associated DNA sequencing. *Genes* **2018**, *9*, 218. [[CrossRef](#)]
12. Ahn, C.H.; Heo, K.; Park, H.S.; Choi, Y.E. In vitro propagation and cryopreservation of *Thuja koraiensis* Nakai via somatic embryogenesis. *In Vitro Cell. Dev. Biol.-Plant* **2019**, *55*, 605–614. [[CrossRef](#)]
13. Shi, X.; Zhang, C.; Liu, Q.; Zhang, Z.; Zheng, B.; Bao, M. De novo comparative transcriptome analysis provides new insights into sucrose induced somatic embryogenesis in camphor tree (*Cinnamomum camphora* L.). *BMC Genom.* **2016**, *17*, 26. [[CrossRef](#)]
14. Nole-Wilson, S.; Tranby, T.L.; Krizek, B.A. AINTEGUMENTA-like (AIL) genes are expressed in young tissues and may specify meristematic or division-competent states. *Plant Mol. Biol.* **2005**, *57*, 613–628. [[CrossRef](#)] [[PubMed](#)]
15. Passarinho, P.; Ketelaar, T.; Xing, M.; van Arkel, J.; Maliepaard, C.; Hendriks, M.W.; Joosen, R.; Lammers, M.; Herdies, L.; den Boer, B.; et al. BABY BOOM target genes provide diverse entry points into cell proliferation and cell growth pathways. *Plant Mol. Biol.* **2008**, *68*, 225–237. [[CrossRef](#)] [[PubMed](#)]
16. Zhu, T.; Moschou, P.N.; Alvarez, J.M.; Sohlberg, J.J.; von Arnold, S. WUSCHEL-related homeobox 8/9 is important for proper embryo patterning in the gymnosperm Norway spruce. *J. Exp. Bot.* **2014**, *65*, 6543–6552. [[CrossRef](#)] [[PubMed](#)]
17. Haecker, A.; Gross-Hardt, R.; Geiges, B.; Sarkar, A.; Breuninger, H.; Herrmann, M.; Laux, T. Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. *Development* **2004**, *131*, 657–668. [[CrossRef](#)]
18. Rups, A.; Raschke, J.; Rümmler, M.; Linke, B.; Zoglauer, K. Identification of putative homologs of *Larix decidua* to *Babyboom* (*bbm*), *Leafy cotyledon1* (*lec1*), *Wuschel-related homeobox2* (*wox2*) and *Somatic embryogenesis receptor-like kinase* (*serk*) during somatic embryogenesis. *Planta* **2016**, *243*, 473–488. [[CrossRef](#)]
19. Palovaara, J.; Hakman, I. Conifer WOX-related homeodomain transcription factors, developmental consideration and expression dynamic of WOX2 during *Picea abies* somatic embryogenesis. *Plant Mol. Biol.* **2008**, *66*, 533–549. [[CrossRef](#)]
20. Schmidt, E.D.L.; Guzzo, F.; Toonen, M.A.J.; De Vries, S.C. A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development* **1997**, *124*, 2049–2062. [[CrossRef](#)]
21. Schellenbaum, P.; Jacques, A.; Maillot, P.; Bertsch, C.; Mazet, F.; Farine, S.; Walter, B. Characterization of *VvSERK1*, *VvSERK2*, *VvSERK3* and *VvL1L* genes and their expression during somatic embryogenesis of grapevine (*Vitis vinifera* L.). *Plant Cell. Rep.* **2008**, *27*, 1799–1809. [[CrossRef](#)]
22. Lee, F.C.; Ong-Abdullah, M.; Ooi, S.E.; Ho, C.L.; Namasivayam, P. Cloning and characterization of *Somatic Embryogenesis Receptor kinase I* (*EgSERK I*) and its association with callus initiation in oil palm. *In Vitro Cell. Dev. Biol.-Plant* **2019**, *55*, 153–164. [[CrossRef](#)]
23. Ahn, C.H.; Choi, Y.E. In vitro clonal propagation and stable cryopreservation system for *Platyclusus orientalis* via somatic embryogenesis. *Plant Cell Tissue Organ Cult.* **2017**, *131*, 513–523. [[CrossRef](#)]
24. Maruyama, E.; Tanaka, T.; Hosoi, Y.; Ishii, K.; Morohoshi, N. Embryogenic cell culture, protoplast regeneration, cryopreservation, biolistic gene transfer and plant regeneration in Japanese cedar (*Cryptomeria japonica* D. Don). *Plant Biotechnol.* **2000**, *17*, 281–296. [[CrossRef](#)]
25. Bodian, D.L.; Solomon, B.D.; Khromykh, A.; Thach, D.C.; Iyer, R.K.; Link, K.; Baker, R.L.; Baveja, R.; Vockley, J.G.; Niederhuber, J.E. Diagnosis of an imprinted-gene syndrome by a novel bioinformatics analysis of whole-genome sequences from a family trio. *Mol. Genet. Genomic. Med.* **2014**, *2*, 530–538. [[CrossRef](#)] [[PubMed](#)]
26. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
27. Boutilier, K.; Offringa, R.; Sharma, V.K.; Kieft, H.; Ouellet, T.; Zhang, L.; Hattori, J.; Liu, C.M.; van Lammeren, A.A.M.; Miki, B.L.A.; et al. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* **2002**, *14*, 1737–1749. [[CrossRef](#)]
28. El Ouakfaoui, S.E.; Schnell, J.; Abdeen, A.; Colville, A.; Labbé, H.; Han, S.; Baum, B.; Laberge, S.; Miki, B. Control of somatic embryogenesis and embryo development by AP2 transcription factors. *Plant Mol. Biol.* **2010**, *74*, 313–326. [[CrossRef](#)]
29. Irikova, T.; Grozeva, S.; Denev, I. Identification of *Baby boom* and *Leafy cotyledon* genes in sweet pepper (*Capsicum annuum* L.) genome by their partial gene sequences. *Plant Growth Regul.* **2012**, *67*, 191–198. [[CrossRef](#)]
30. Salvo, S.A.G.D.; Hirsch, C.N.; Buell, C.R.; Kaeppler, S.M.; Kaeppler, H.F. Whole transcriptome profiling of maize during early somatic embryogenesis reveals altered expression of stress factors and embryogenesis-related genes. *PLoS ONE* **2014**, *9*, e111407. [[CrossRef](#)]
31. Hedman, H.; Zhu, T.; von Arnold, S.; Sohlberg, J.J. Analysis of the WUSCHEL-RELATED HOMEBOX gene family in the conifer *Picea abies* reveals extensive conservation as well as dynamic patterns. *BMC Plant Biol.* **2013**, *13*, 89. [[CrossRef](#)]
32. Van der Graaff, E.; Laux, T.; Rensing, S.A. The WUS homeobox-containing (WOX) protein family. *Genome Biol.* **2009**, *10*, 248. [[CrossRef](#)] [[PubMed](#)]
33. Hecht, V.; Vielle-Calzada, J.; Hartog, M.; Schmidt, E.; Grossniklaus, U.; de Vries, S. The *Arabidopsis somatic embryogenesis receptor kinase 1* gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiol.* **2001**, *127*, 803–816. [[CrossRef](#)] [[PubMed](#)]
34. Li, K.P.; Sun, X.M.; Han, H.; Zhang, S.G. Isolation, characterization and expression analysis of the BABY BOOM (*BBM*) gene from *Larix kaempferi* × *L. olgensis* during adventitious rooting. *Gene* **2014**, *10*, 551. [[CrossRef](#)] [[PubMed](#)]

35. Galinha, C.; Hofhuis, H.; Luijten, M.; Willemsen, V.; Blilou, I.; Heidstra, R.; Scheres, B. PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* **2007**, *449*, 1053–1057. [[CrossRef](#)] [[PubMed](#)]
36. Imin, N.; Nizamidin, M.; Wu, T.; Rolfe, B.G. Factors involved in root formation in *Medicago truncatula*. *J. Exp. Bot.* **2007**, *58*, 439–451. [[CrossRef](#)]
37. Klimaszewska, K.; Overton, C.; Stewart, D.; Rutledge, R.G. Initiation of somatic embryos and regeneration of plants from primordial shoots of 10-year-old somatic white spruce and expression profiles of 11 genes followed during the tissue culture process. *Planta* **2011**, *233*, 635–647. [[CrossRef](#)]

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