

## RESEARCH ARTICLE

Genotype and transcriptome effects on somatic embryogenesis in *Cryptomeria japonica*Ayako Izuno<sup>1\*</sup>, Tsuyoshi E. Maruyama<sup>1</sup>, Saneyoshi Ueno<sup>1</sup>, Tokuko Ujino-Ihara<sup>1</sup>, Yoshinari Moriguchi<sup>2</sup>**1** Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Tsukuba, Ibaraki, Japan, **2** Graduate School of Science and Technology, Niigata University, Niigata, Japan\* [izuno@ffpri.affrc.go.jp](mailto:izuno@ffpri.affrc.go.jp)**OPEN ACCESS****Citation:** Izuno A, Maruyama TE, Ueno S, Ujino-Ihara T, Moriguchi Y (2020) Genotype and transcriptome effects on somatic embryogenesis in *Cryptomeria japonica*. PLoS ONE 15(12): e0244634. <https://doi.org/10.1371/journal.pone.0244634>**Editor:** Vijay Kumar, Lovely Professional University, INDIA**Received:** July 31, 2020**Accepted:** December 14, 2020**Published:** December 29, 2020**Copyright:** © 2020 Izuno et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.**Data Availability Statement:** Raw sequence data are available from DDBJ (accession number: DRA010541). De novo cDNA assembly obtained in this study is deposited in DDBJ (accession number: ICQR01000001–ICQR01031595).**Funding:** This work was supported in part by grants from the Ministry of Agriculture, Forestry and Fisheries of Japan and NARO Bio-oriented Technology Research Advancement Institution to Y. M. (No. 28013B and No. 28013BC). The funders had no role in study design, data collection and

## Abstract

Somatic embryogenesis (SE), which is *in vitro* regeneration of plant bodies from somatic cells, represents a useful means of clonal propagation and genetic engineering of forest trees. While protocols to obtain calluses and induce regeneration in somatic embryos have been reported for many tree species, the knowledge of molecular mechanisms of SE development is still insufficient to achieve an efficient supply of somatic embryos required for the industrial application. *Cryptomeria japonica*, a conifer species widely used for plantation forestry in Japan, is one of the tree species waiting for a secure SE protocol; the probability of normal embryo development appears to depend on genotype. To discriminate the embryogenic potential of embryonal masses (EMs) and efficiently obtain normal somatic embryos of *C. japonica*, we investigated the effects of genotype and transcriptome on the variation in embryogenic potential. Using an induction experiment with 12 EMs each from six genotypes, we showed that embryogenic potential differs between/within genotypes. Comparisons of gene expression profiles among EMs with different embryogenic potentials revealed that 742 differently expressed genes were mainly associated with pattern forming and metabolism. Thus, we suggest that not only genotype but also gene expression profiles can determine success in SE development. Consistent with previous findings for other conifer species, genes encoding leafy cotyledon, wuschel, germin-like proteins, and glutathione-S-transferases are likely to be involved in SE development in *C. japonica* and indeed highly expressed in EMs with high-embryogenic potential; therefore, these proteins represent candidate markers for distinguishing embryogenic potential.

## Introduction

Somatic embryogenesis (SE), which involves *in vitro* development of the bipolar plant body from somatic cells, is an effective means of clonal propagation of forest trees and promote breeding in plantation forestry and the conservation of valuable tree species. The SE-based method has advantages over conventional methods of clonal propagation, such as cutting, grafting, and coppicing, in terms of efficiency and species preservation. Once a proper

analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

protocol for culturing is established, SE-based propagation enables indoor production of any number of clonal plants. As juvenile plants derived from somatic embryos can regenerate even after cryopreservation, the long-term preservation of genetic resources and stable supply of saplings is feasible with SE-based propagation. Currently, SE-based propagation is applied on a commercial scale for the forestry of limited number of conifer species in genera such as *Abies*, *Larix*, *Picea*, *Pinus*, and *Pseudotsuga* [1, 2].

The molecular mechanisms underlying SE development remain to be fully elucidated and thus, developing SE protocols has been difficult for many tree species. Based on the gene expression assays and recent genome-wide studies, some gene families that encode SE receptor kinase, leafy cotyledon (LEC), and wuschel (WUS) proteins likely have important functions in SE development [3, 4]. However, it is unclear whether the genetic mechanisms controlling SE development are common across species. Furthermore, the importance of epigenetic regulation, which can also modulate gene reprogramming and determine the embryogenic state of culture cells [5, 6], requires further investigation. Elucidating the molecular mechanisms underlying SE development will help improve current SE protocols, which have often been established through trial and error.

In Japan, *Cryptomeria japonica* D. Don (Cupressaceae) is one of the most important conifer species in the plantation forestry, comprising 4.5 million ha (~45%) of plantation stands [7]. Conventional breeding has depended on cutting of selected trees with superior traits; however, SE-based propagation could improve *C. japonica* breeding programs with, for example, genetic engineering. One of the promising applications of genetic engineering in *C. japonica* involves establishing male-sterile varieties, which would benefit the ~30% of the Japanese population who are allergic to pollens from *C. japonica* plantations [8]. Genetic engineering with SE is theoretically possible in *C. japonica* because protocols to induce and maintain somatic embryos have been developed [9–12] and transformation with somatic embryos has been reported [13]. Accumulated genetic data for *C. japonica* [e.g., 14–21], as well as a future genome assembly, will also support genetic engineering of varieties with desirable traits.

We recently found a substantial variation in the embryogenesis of calluses among *C. japonica* siblings (i.e., different genotypes originated from identical parents) even with the optimum protocols for the *C. japonica* SE system [10, 11]. Embryogenic potential cannot be distinguished until somatic embryos actually start developing; such unpredictable variability means that many different genotypes must be maintained and this hinders the industrial application of SE-based propagation for *C. japonica*. In the present study, to improve the SE protocols for *C. japonica*, we investigated (1) whether the variability in embryogenesis is genetically determined and (2) whether genetic expression profiles differ between embryonal masses with high-embryogenic and low-embryogenic potential. Our results indicate that not only genotypes but also gene expression profiles are associated with the variation in the embryogenic potential. We report some transcriptome markers to discriminate the embryogenic potential in the *C. japonica* SE system.

## Materials and methods

### Cell lines

We used six embryogenic *C. japonica* cell lines, namely SSD-18, 73, 100, 113, 182, and 352, which were obtained from zygotic embryos excised from immature seeds in July 2016. The seeds were collected from a hybrid offspring produced by ‘Shindai 3’ (maternal) and ‘Suzu 2’ (paternal). The detailed procedures used to obtain embryonal masses (EMs) were described in [22]. The EMs were maintained and proliferated every two weeks with EM medium (S1a

[Table](#)). Embryogenesis was confirmed at least eight times with the maturation media ([S1b Table](#)) before this study began.

## Maturation

In August 2019, 12 EMs (100 mg fresh weight each) per cell line were transferred to Petri dishes (three EMs per dish) with the maturation media ([S1b Table](#)). The Petri dishes were maintained in darkness at 25°C. At 5, 7, and 9 weeks after the maturation, the number of normal somatic embryos was recorded and compared between and within cell lines by Tukey's multiple comparison test.

## RNA isolation

We investigated the gene expression profiles of five EMs from SSD-100, in which we found marked differences in somatic embryos maturation. At 5, 7, and 9 weeks after maturation, 100 mg of tissue was preserved in 200  $\mu$ l of CTAB buffer at -80°C until RNA extraction. As these EMs originated from an identical cell line (i.e., genotype), sequence polymorphisms in transcripts would not occur among samples. Therefore, the difference in the number of reads mapped on transcript sequences likely reflects differences in expression levels.

Immediately after the frozen tissue solutions had melted at room temperature, the tissues were macerated with a zirconium ball using TissueLyser II (Qiagen). After shaking at 30 Hz for 30 s, the sample tubes were placed on ice to avoid RNA degradation. This procedure was repeated four times. The sample tubes were then incubated at 65°C for 10 min with an additional 200  $\mu$ l of CTAB buffer (containing 2% CTAB, 40 mM EDTA, 100 mM Tris-HCl (pH 8.0), 2% PVP, 1.4 M NaCl, and 2% 2-mercaptoethanol). Subsequently, the solutions were mixed with a half volume of chloroform/isoamyl alcohol and then centrifuged at 13,000 g for 15 min. After collection of approximately 300  $\mu$ l of the aqueous phase in a new tube, the RNA molecules were precipitated with a quarter volume of 10M LiCl at -20°C for 2 h and subsequently centrifuged at 20,000 g and 4°C for 25 min. The pellet was purified using the SV Total RNA Isolation System (Promega) following the manufacturer's protocol.

## Transcriptome sequencing

The RNA-Seq libraries were prepared using TruSeq Stranded mRNA Library Prep (Illumina) and sequenced with 2  $\times$  100 bp technology in the NovaSeq 6000 System (Illumina). The sequence quality of raw reads was controlled using Trimmomatic (ver. 0.39) [23]. The clean read pairs were aligned on the reference cDNA sequences CJ3006NRE (49,758 transcripts of 39,229 genes expressed in mature leaves, inner bark, and male flowers; [24]) using bowtie2 (ver. 2.3.4.1) [25]. The number of reads was counted for each gene using RSEM (ver. 1.3.2) [26].

We gathered unmapped read pairs, which constituted ~20% (range: 15.7%–26.0%) of input read pairs per sample, and created a new assembly using Trinity (ver. 2.8.4) [27]. Open reading frames (ORFs) in the new transcript assembly were identified using TransDecoder (ver. 5.5.0; <https://github.com/TransDecoder/TransDecoder/wiki>). Using NCBI BLAST-P (ver. 2.7.1), the homology sequences of the ORFs were searched against 45,644 isoform sequences for plants reposted in SwissProt (downloaded on March 30, 2020) [28]. Using hmmscan in HMMER (ver. 3.2.1; <http://hmmer.org/>), protein domains in the ORFs were identified by reference to 18,197 Pfam-A domains (ver. 33.0; downloaded on March 30, 2020 [29]). Based on these two annotations, the coding sequences in the new assembly were predicted using TransDecoder (ver. 5.5.0). Short transcript sequences (< 500 bp) were removed. The unmapped reads were aligned on the new transcript assembly using bowtie2 (ver. 2.3.4.1) and the read coverage per

gene was counted using RSEM (ver. 1.3.2). The read count data for CJ3006NRE and the new transcript assembly were integrated into a single matrix.

### Differentially expressed genes between EMs with high-embryogenic and low-embryogenic potential

Based on the observed number of embryos, we recognized two out of five EMs (namely, C1 and C5) to have high embryogenic (HE) potential and the remaining (namely, C3, C4, and C6) to have low-embryogenic (LE) potential. In C1 and C5, the number of embryos increased from the 5th to 7th week and decreased from the 7th to 9th week, indicating that the development of new somatic embryos was active at the 5th and 7th weeks but had decelerated by the 9th week. Therefore, we searched genes significantly expressed in HE samples (C1 and C5) at the 5th and 7th weeks relative to expression in the LE samples (C3, C4, and C6). We selected genes with more than one count per million reads in at least two samples. Read count data were normalized based on the trimmed mean of M values normalization method [30]. Differentially expressed genes between HE and LE samples were identified using the *glmQLFit* and *glmQLFTest* functions implemented in the R package edgeR (ver. 3.30.1) [30]. Genes with less than 0.05 false discovery rate and more than 2 log-fold-change were considered significant.

Gene functions enriched in the target genes were identified using the R package topGO with the “elim” algorithm (ver. 2.40.0) [31]. Gene Ontology (GO) terms with  $p < 0.01$  were considered to be significant. For enrichment analysis, we annotated the reference cDNA sequences CJ3006NRE with the SwissProt plant isoform sequence data (the same data used to annotate the new assembly of unmapped reads; see above) and then obtained GO terms (S3 Table).

### Validation of differentially expressed genes by quantitative PCR

Among the differentially expressed genes identified by RNA-Seq, the expression levels of six genes (S5 Table) were validated by quantitative PCR (qPCR). Alpha-tubulin (AT1G50010 in the *Arabidopsis thaliana* genome) was selected for a reference housekeeping gene, because it is suitable for qPCR assays during somatic embryo development in conifer species [32].

Based on the reciprocal best hits between CJ3006NRE and TAIR10, a *C. japonica* transcript CJt088185 was assumed to be the most homologous with AT1G50010 and used for the subsequent primer design.

We first identified exon-exon junctions on template transcript sequences based on the alignment with the homologous gene sequences in TAIR10 (S5 Table). Primer pairs over the exon-exon junctions were designed using Primer-BLAST [33], with 80–150 bp of product size, 57–63°C (optimal: 60°C) of annealing temperature, 17–25 bp (optimal: 20 bp) of primer length, and 45–60°C of primer GC content. Primer specificity was checked through *in silico* PCR with a total of 81,353 *C. japonica* transcript sequences used in this study (S3 and S4 Tables).

We performed qPCR for a total of 10 EM samples, which are composed of five EMs of SSD-100 collected at the 5th and 7th week after maturation and used for RNA-Seq. After removing genomic DNA from ca. 500 ng of total RNA using gDNA Remover, cDNA was synthesized using ReverTra Ace qPCR RT Master Mix (TOYOBO). Each of six target and one reference loci was amplified from the cDNA template in triplicate of 10 µl reaction solution including 5–10% of cDNA, 1×KOD SYBR qPCR Mix (TOYOBO), and 0.2 µM each of forward and reverse primers. Using LightCycler 480 (Roche), the intensity of SYBR Green I was measured during PCR, which was composed of an initial step at 98°C for 2 min and 40 cycles of denaturation at 98°C for 10 sec, annealing at 60°C for 10 sec, and extension at 68°C for 30 sec. After the amplification step, a denaturation at 95°C for 5 sec and an annealing at 65°C for 1 min followed, then a melting curve analysis was carried out from 65°C to 95°C with continuous

fluorescent measurements at 5°C interval. The absence of primer dimers was confirmed by the reactions without cDNA (i.e., negative controls). The mean of cycle thresholds (Ct) was identified using 2nd derivative maximum method for each reaction, then relative expression level among the 10 samples was calculated for each locus with delta-delta Ct method, in which the expression level was normalized with that in C4 from the 7th week. The difference in gene expression level between EMs was examined using Tukey's multiple comparison test.

## Results

### SE development

The number of embryos developed from the 5th to 9th week varied between cell lines. SSD-18 and 73 showed higher capacities for developing somatic embryos than the other four cell lines during 5–9 and 7–9 weeks after maturation, respectively ( $p < 0.01$ , Fig 1a–1c). In some cell lines, namely SSD-100 and 113, the mean number of embryos developed during the experiment period was significantly different among EMs within cell lines ( $p < 0.01$ , S1 Fig).

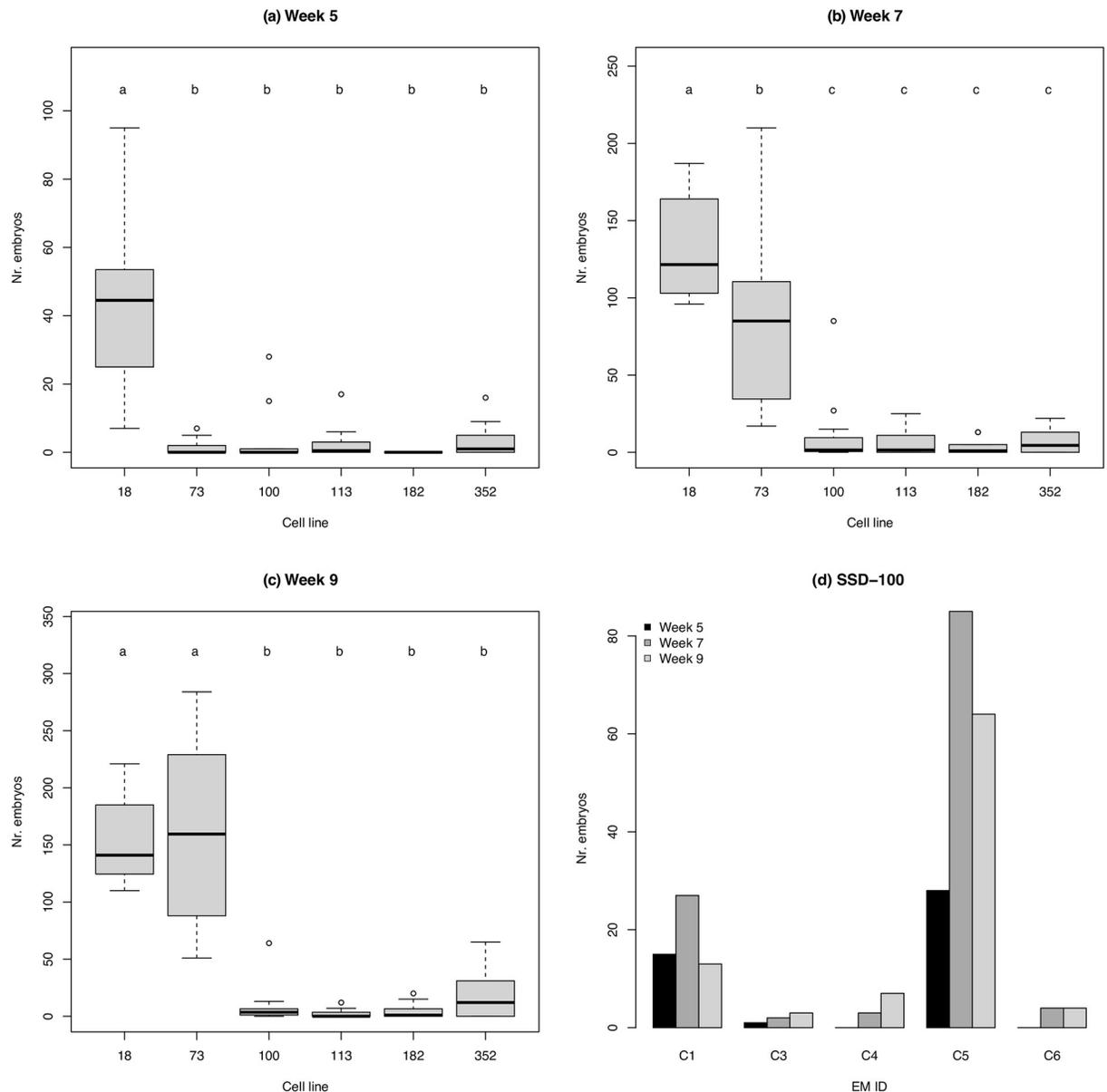
We found the five EMs from SSD-100 to be suitable to assay gene expressions because they showed a marked difference in the capacity to develop somatic embryos in spite of their shared genotype (Fig 2). C1 and C5 developed more than 10 embryos throughout the experimental period, whereas C3, C4, and C6 did not (Fig 1d). Because we collected some normal somatic embryos at each time point, the number of embryos could subsequently decrease unless embryogenesis was actively maintained. In C1 and C5, new somatic embryos appeared at the 5th and 7th week, but did not appear at the 9th week. Based on these observations, we considered C1 and C5 at the 5th and 7th week to have high embryogenic potential.

### Transcriptome sequencing

We obtained  $20\text{--}37 \times 10^6$  read pairs ( $4.1\text{--}7.5 \times 10^9$  bases) per sample (S2 Table). Raw data are deposited on DDBJ DRA (Accession number: DRA010541). After filtering,  $3.98\text{--}7.33 \times 10^9$  bases remained per sample and 74–84% of these were properly mapped onto CJ3006NRE. *De novo* assembly of unmapped reads generated 204,943 transcript sequences. Based on a homology search against SwissProt and Pfam-A, 78,415 isoform sequences of 25,364 genes were predicted. After removing short (< 500 bp) sequences, 31,595 isoform sequences of 12,073 genes were used for subsequent analyses. This assembly is deposited in DDBJ (Accession number: ICQR01000001–ICQR01031595) and the annotation is shown in S4 Table.

### Differentially expressed genes between EMs with high-embryogenic and low-embryogenic potential

At the 5th and 7th week, 742 genes were significantly upregulated in HE samples compared with LE samples. The 742 genes upregulated in HE samples included some candidate genes encoding such as leafy cotyledon (LEC; *g28713*; S3 Table), wuschel (WUS; *g27218* and *g34977*; S3 Table), and germin-like protein (GLP; *g31725* and *g38183*; S3 Table), as well as glutathione-S-transferase (GST; seven genes including *g3323* and *TRINITY\_DN4815\_c0\_g1*; S3 and S4 Tables). Consistent with the developmental stages in our experiment, we found ten genes encoding late embryogenesis abundant proteins. GO enrichment analysis showed that the 742 genes were significantly involved in the development of the plant body, especially specifications of polarity, and metabolic processes including the biosynthesis of flavonoids, gibberellins, and brassinosteroids (Table 1).

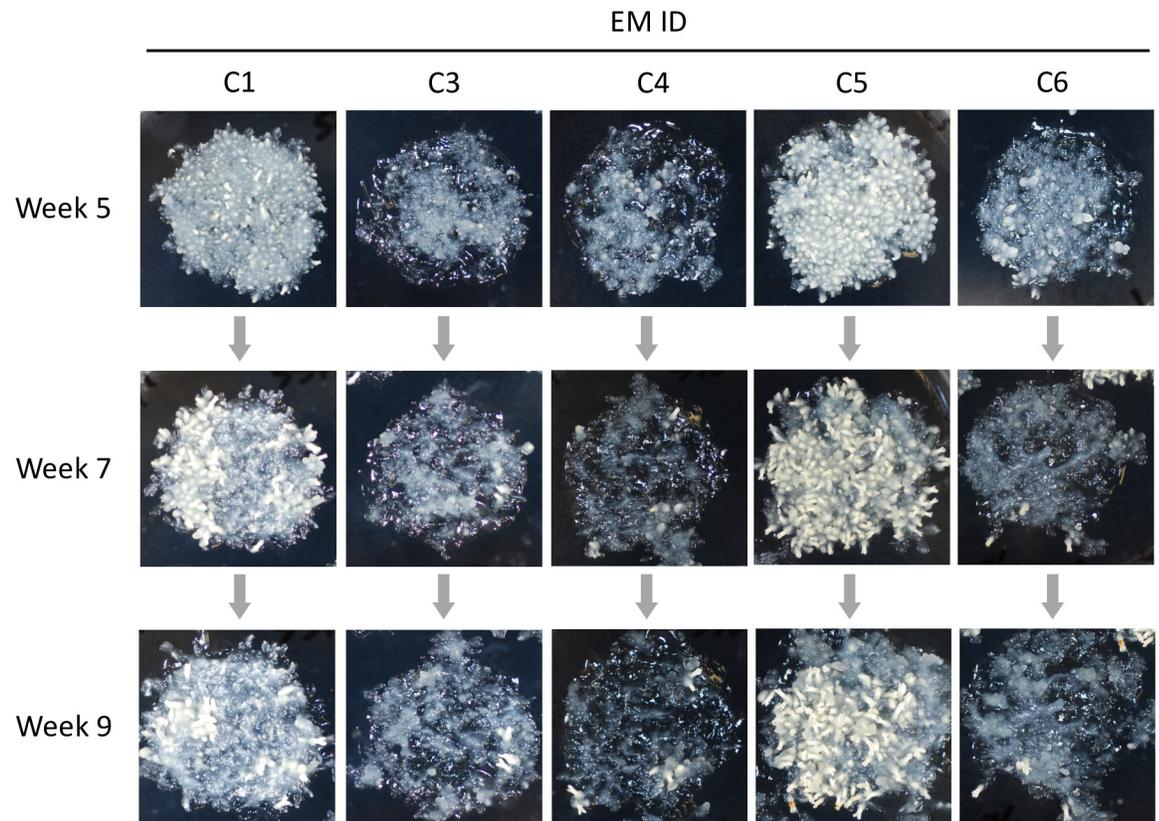


**Fig 1.** (a–c) The number of embryos observed in six *Cryptomeria japonica* cell lines at (a) 5, (b) 7, and (c) 9 weeks after maturation. Twelve embryonal masses (EMs) were cultured per cell line. Means with different letters are significantly different (Tukey's multiple comparison test,  $p < 0.01$ ). (d) The number of embryos observed in five EMs of the cell line SSD-100 at 5, 7, and 9 weeks after maturation. These EMs were used for RNA-Seq.

<https://doi.org/10.1371/journal.pone.0244634.g001>

### Validation of differentially expressed genes by quantitative PCR

qPCR assays showed that HE samples (C1 and C5) more highly expressed six candidate genes compared to LE samples (C3, C4, and C6) except for the C5 at the 5th week, where the expression levels of three genes, namely *g28713* (LEC), *g27218* (WUS), and *g34977* (WUS), were not significantly different from LE samples (Fig 3). Clear difference in expression levels between HE and LE samples was particularly found at the 7th week, consistent with the increased number of somatic embryos (Figs 1d and 2).



**Fig 2. The five *Cryptomeria japonica* embryonal masses (EMs) induced from an identical cell line SSD-100 at 5, 7, and 9 weeks after maturation.** The capacity to develop somatic embryos are different between EMs.

<https://doi.org/10.1371/journal.pone.0244634.g002>

## Discussion

Our results showed that the embryogenesis of *C. japonica* largely differs among cell lines (i.e., genotypes). Although we examined six cell lines with embryogenic potential, only two (SSD-18 and 73) showed stable embryogenesis (Fig 1a–1c). This is consistent with the varying potential of cell lines for forming embryos reported in other conifer species such as *Pinus radiata* [34]. Besides the variation among cell lines, we showed that embryogenesis differs within cell lines (S1 Fig). Given that all the EMs within cell lines share the same genomic background, embryogenic potential may also depend on gene expression regulation of relevant genes. Intrinsic factors, such as the totipotency of cells transplanted to maturation media, and extrinsic factors, including the concentration of the medium available for EMs, may affect the signaling pathways of gene expressions. The variation we observed in embryogenesis between and within cell lines confirms that a substantial number of cell lines as well as replicated EMs should be maintained to ensure a stable supply of *C. japonica* somatic embryos. These findings critically indicate the necessity of a means to identify embryogenic potential at certain developmental stages.

We found that gene expression profiles clearly differed between the HE and LE samples at the 5th and 7th weeks after maturation, during which somatic embryos developed in abundance. GO enrichment analysis suggested that genes upregulated in HE samples were involved in the pattern forming and metabolism required to develop somatic embryos (Table 1). In the 742 upregulated genes, we found some candidate genes: leafy cotyledon (LEC), wuschel

Table 1. Over-represented Gene Ontology (GO) terms in the 742 genes upregulated in high embryogenic *Cryptomeria japonica* embryonal masses.

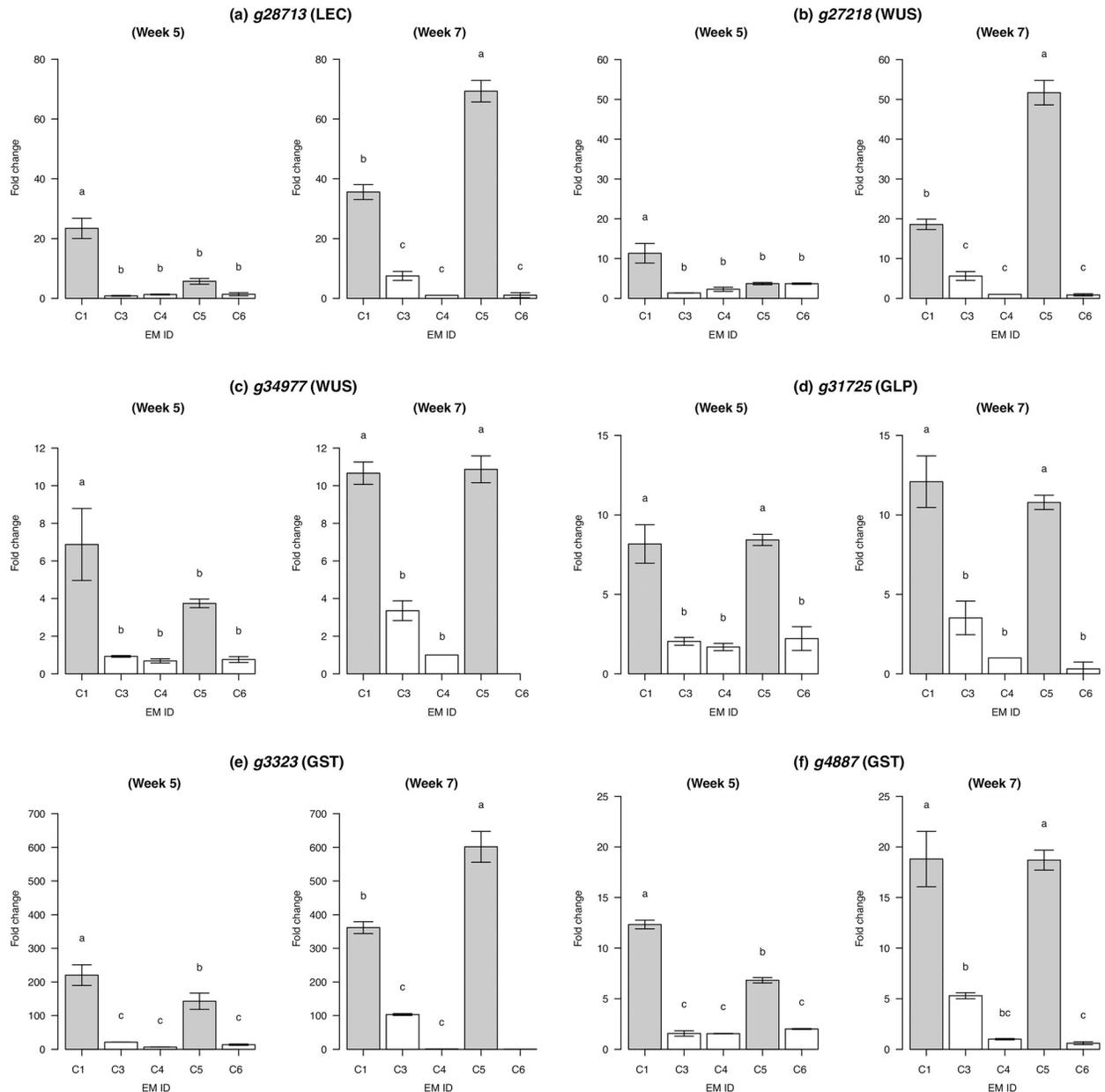
Category	GO ID	Term	p
Development			
	GO:0010158	Abaxial cell fate specification	2.E-05
	GO:0080060	Integument development	3.E-04
	GO:1902183	Regulation of shoot apical meristem development	8.E-04
	GO:0009943	Adaxial/abaxial axis specification	3.E-03
	GO:0010073	Meristem maintenance	4.E-03
	GO:0009944	Polarity specification of adaxial/abaxial axis	4.E-03
	GO:0099402	Plant organ development	7.E-03
	GO:1901342	Regulation of vasculature development	8.E-03
Metabolic process			
	GO:0009813	Flavonoid biosynthetic process	3.E-09
	GO:0009686	Gibberellin biosynthetic process	3.E-07
	GO:0016131	Brassinosteroid metabolic process	5.E-07
	GO:0016042	Lipid catabolic process	8.E-05
	GO:0006110	Regulation of glycolytic process	2.E-04
	GO:1901959	Positive regulation of cutin biosynthetic process	3.E-04
	GO:0009699	Phenylpropanoid biosynthetic process	7.E-04
	GO:0010345	Suberin biosynthetic process	8.E-04
	GO:0033473	Indoleacetic acid conjugate metabolic process	9.E-04
	GO:0042744	Hydrogen peroxide catabolic process	1.E-03
	GO:0006694	Steroid biosynthetic process	1.E-03
	GO:0016132	Brassinosteroid biosynthetic process	2.E-03
	GO:2000762	Regulation of phenylpropanoid metabolic process	2.E-03
	GO:0046463	Acylglycerol biosynthetic process	3.E-03
	GO:0042761	Very long-chain fatty acid biosynthetic process	5.E-03
	GO:0009690	Cytokinin metabolic process	7.E-03
	GO:0046189	Phenol-containing compound biosynthetic . . .	7.E-03
	GO:0010430	Fatty acid omega-oxidation	7.E-03
	GO:0006571	Tyrosine biosynthetic process	8.E-03
	GO:0009696	Salicylic acid metabolic process	8.E-03
	GO:1900378	Positive regulation of secondary metabolite biosynthetic process	8.E-03
Response to stimulus			
	GO:0009744	Response to sucrose	5.E-04
	GO:0043481	Anthocyanin accumulation in tissues in response to UV light	6.E-04
	GO:0009725	Response to hormone	2.E-03
	GO:1901700	Response to oxygen-containing compound	3.E-03
	GO:0033993	Response to lipid	3.E-03
	GO:0009958	Positive gravitropism	9.E-03
Transport			
	GO:0080170	Hydrogen peroxide transmembrane transport	6.E-06
	GO:0015840	Urea transport	1.E-05
	GO:0071577	Zinc ion transmembrane transport	3.E-04
	GO:0008643	Carbohydrate transport	5.E-03
	GO:0006833	Water transport	9.E-03
Others			
	GO:0010268	Brassinosteroid homeostasis	5.E-08
	GO:0070207	Protein homotrimerization	2.E-06

(Continued)

Table 1. (Continued)

Category	GO ID	Term	p
	GO:0010431	Seed maturation	3.E-05
	GO:0036290	Protein trans-autophosphorylation	3.E-04

<https://doi.org/10.1371/journal.pone.0244634.t001>



**Fig 3. Relative expressions of putatively embryogenesis-related genes in *Cryptomeria japonica*.** Expressions of six genes encoding (a) leafy cotyledon (LEC), (b, c) wuschel (WUS), (d) germin-like protein (GLP), and (e, f) glutathione-S-transferase (GST) were quantitated. For each gene, expression levels were normalized with that in C4 from the 7th week. Means with different letters are significantly different (Tukey's multiple comparison test,  $p < 0.01$ ). Samples with high embryogenic potential (C1 and C5) were indicated in gray. In C6 from the 7th week, the expressions of (c) *g34977* and (e) *g3323* were not quantified due to the low expressions.

<https://doi.org/10.1371/journal.pone.0244634.g003>

(WUS), germin-like protein (GLP), and glutathione-S-transferase (GST) genes. LEC genes are required for the specification of cotyledon identity and the completion of embryo maturation [35]. The WUS gene family acts to maintain the stem cell population in shoot apical meristems [36, 37], root apical meristems [38], and cambial meristems [39, 40]. Both LEC and WUS proteins play central roles in the auxin signaling pathway and regulate somatic cells for embryogenic development [41]. In *Arabidopsis thaliana*, LEC genes are activated by WUS genes to promote SE [42]. The high expression of LEC and WUS genes during SE development has also been confirmed in other conifer species (*Picea* and *Pinus* species; [34, 43, 44]). The GLP gene family is thought to mediate the initiation of embryo germination in wheat [45], *Arabidopsis* [46], and cotton [47]. In *Pinus* and *Picea*, GLPs likely function during the early stages of SE induction, as suggested by proteomic and transcriptomic studies [48–50]. GST enzymes are known to act under stress conditions and to minimize oxidative damage on cells. The transcripts have been detected in somatic embryos in *Chicorium* [51] and wheat [52].

According to a review by Mahdavi-Darvari et al. [3], LEC, WUS, GLP, and GST genes are good candidate markers for embryogenic potential. As our qPCR results validated the different expressions between HE and LE samples (Fig 3), the primers (S5 Table) can be used as appropriate markers to discriminate embryogenic potential in *C. japonica*. Identifying the expression profiles of these genes in the early stages of SE development will further inform the effectiveness of these genes as markers. Besides the focus on gene expression, studies on epigenetic [53] and proteome processes [54] could provide additional insights to improve the efficient collection of quality somatic embryos.

## Supporting information

**S1 Table. Chemical composition of (a) maintenance-proliferation medium and (b) maturation medium.**

(DOCX)

**S2 Table. The number of RNA-Seq reads processed.**

(DOCX)

**S3 Table. Gene annotations for CJ3006NRE.**

(XLSX)

**S4 Table. Gene annotations for the transcript sequence assembly of reads not aligned on the CJ3006NRE.**

(XLSX)

**S5 Table. Primer information for quantitative PCR.**

(XLSX)

**S1 Fig. The number of somatic embryos obtained in each embryonal mass at the 5th, 7th, and 9th week after maturation.** Means with different letters are significantly different (Tukey's multiple comparison test,  $p < 0.01$ ).

(TIF)

## Acknowledgments

We thank Shizue Tanaka for her assistance with laboratory work and Satoko Hirayama for providing seeds used for SE induction.

## Author Contributions

**Conceptualization:** Ayako Izuno, Tsuyoshi E. Maruyama, Saneyoshi Ueno.

**Data curation:** Ayako Izuno.

**Formal analysis:** Ayako Izuno.

**Funding acquisition:** Yoshinari Moriguchi.

**Investigation:** Ayako Izuno.

**Methodology:** Ayako Izuno, Tokuko Ujino-Ihara.

**Resources:** Tsuyoshi E. Maruyama, Saneyoshi Ueno.

**Visualization:** Ayako Izuno.

**Writing – original draft:** Ayako Izuno.

**Writing – review & editing:** Ayako Izuno, Tsuyoshi E. Maruyama, Saneyoshi Ueno, Tokuko Ujino-Ihara, Yoshinari Moriguchi.

## References

1. Klimaszewska K, Trontin J-F, Becwar MR, Devillard C, Park Y-S, Lelu-Walter M-A. Recent progress in somatic embryogenesis of four *Pinus* spp. *Tree and Forest Science and Biotechnology*. 2007; 1: 11–25.
2. Lelu-Walter M-A, Thompson D, Harvengt L, Sanchez L, Toribio M, Pâques LE. Somatic embryogenesis in forestry with a focus on Europe: State-of-the-art, benefits, challenges and future direction. 2013; 9: 883–899. <https://doi.org/10.1007/s11295-013-0620-1>
3. Mahdavi-Darvari F, Noor NM, Ismanizan I. Epigenetic regulation and gene markers as signals of early somatic embryogenesis. *Plant Cell Tiss Organ Cult*. 2015. <https://doi.org/10.1007/s11240-014-0615-0>
4. Plomion C, Bastien C, Bogeat-Triboulot M-B, Bouffier L, Déjardin A, Duplessis S, et al. Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science*. 2nd ed. 2016; 73: 77–103. <https://doi.org/10.1007/s13595-015-0488-3>
5. Noceda C, Salaj T, Pérez M, Viejo M, Cañal MJ, Salaj J, et al. DNA demethylation and decrease on free polyamines is associated with the embryogenic capacity of *Pinus nigra* Arn. cell culture. *Trees*. 2009; 23: 1285.
6. Zhang S, Zhou J, Han S, Yang W, Li W, Wei H, et al. Four abiotic stress-induced miRNA families differentially regulated in the embryogenic and non-embryogenic callus tissues of *Larix leptolepis*. *Biochem Biophys Res Commun*. 2010; 398: 355–360. <https://doi.org/10.1016/j.bbrc.2010.06.056> PMID: 20599742
7. Forestry Agency. Statistical handbook of forest and forestry. Ministry of Agriculture, Forestry and Fisheries; 2014. pp. 8–9.
8. Baba K, Nakae K. The national epidemiological survey of allergic rhinitis in 2008—comparison between 1998 and 2008. *Prog Med*. 2008; 28: 2001–2012.
9. Ogita S, Ishikawa H, Kubo T, Sasamoto H. Somatic embryogenesis from immature and mature zygotic embryos of *Cryptomeria japonica* I: Embryogenic cell induction and its morphological characteristics. *Journal of wood science*. Dordrecht: Springer, Dordrecht; 1999; 45: 87–91.
10. 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 Biotechnology*. Japanese Society for Plant Cell and Molecular Biology; 2000; 17: 281–296. <https://doi.org/10.5511/plantbiotechnology.17.281>
11. Igasaki T, Sato T, Akashi N, Mohri T, Maruyama E, Kinoshita I, et al. Somatic embryogenesis and plant regeneration from immature zygotic embryos of *Cryptomeria japonica* D. Don. *Plant Cell Reports*. 2003; 22: 239–243. <https://doi.org/10.1007/s00299-003-0687-5> PMID: 14586550
12. Maruyama E, Hosoi Y. Polyethylene glycol enhance somatic embryo production in Japanese cedar (*Cryptomeria japonica* D. Don). *Propagation of Ornamental Plants*. 2007; 7: 57–61.
13. Taniguchi T, Ohmiya Y, Kurita M, Tsubomura M, Kondo T. Regeneration of transgenic *Cryptomeria japonica* D. Don after *Agrobacterium tumefaciens*-mediated transformation of embryogenic tissue. *Plant Cell Reports*. 2008; 27: 1461–1466. <https://doi.org/10.1007/s00299-008-0569-y> PMID: 18542965

14. Futamura N, Ujino-Ihara T, Nishiguchi M, Kanamori H, Yoshimura K, Sakaguchi M, et al. Analysis of expressed sequence tags from *Cryptomeria japonica* pollen reveals novel pollen-specific transcripts. *Tree Physiology*. 2006; 26: 1517–1528. <https://doi.org/10.1093/treephys/26.12.1517> PMID: 17169891
15. Moriguchi Y, Totsuka S, Iwai J, Matsumoto A, Ueno S, Tsumura Y. Pyramiding of male-sterile genes in *Cryptomeria japonica* D. Don with the aid of closely linked markers. *Tree Genetics & Genomes*. 2017; 13: 61.
16. Moriguchi Y, Uchiyama K, Ueno S, Ujino-Ihara T, Matsumoto A, Iwai J, et al. A high-density linkage map with 2560 markers and its application for the localization of the male-sterile genes ms3 and ms4 in *Cryptomeria japonica* D. Don. *Tree Genetics & Genomes*. 2016; 12: 57.
17. Moriguchi Y, Ujino-Ihara T, Uchiyama K, Futamura N, Saito M, Ueno S, et al. The construction of a high-density linkage map for identifying SNP markers that are tightly linked to a nuclear-recessive major gene for male sterility in *Cryptomeria japonica* D. Don. *BMC Genomics*. 2012; 13: 95. <https://doi.org/10.1186/1471-2164-13-95> PMID: 22424262
18. Ueno S, Moriguchi Y, Uchiyama K, Ujino-Ihara T, Futamura N, Sakurai T, et al. A second generation framework for the analysis of microsatellites in expressed sequence tags and the development of EST-SSR markers for a conifer, *Cryptomeria japonica*. *BMC Genomics*. 2012; 13: 136. <https://doi.org/10.1186/1471-2164-13-136> PMID: 22507374
19. Uchiyama K, Ujino-Ihara T, Ueno S, Taguchi Y, Futamura N, Shinohara K, et al. Single nucleotide polymorphisms in *Cryptomeria japonica*: their discovery and validation for genome mapping and diversity studies. *Tree Genetics & Genomes*. 2012; 8: 1213–1222.
20. Mishima K, Hirao T, Tsubomura M, Tamura M, Kurita M, Nose M, et al. Identification of novel putative causative genes and genetic marker for male sterility in Japanese cedar (*Cryptomeria japonica* D. Don). *BMC Genomics*. 2018; 19: 277. <https://doi.org/10.1186/s12864-018-4581-5> PMID: 29685102
21. Hasegawa Y, Ueno S, Matsumoto A, Ujino-Ihara T, Uchiyama K, Totsuka S, et al. Fine mapping of the male-sterile genes (MS1, MS2, MS3, and MS4) and development of SNP markers for marker-assisted selection in Japanese cedar (*Cryptomeria japonica* D. Don). Chen C, editor. *PLoS ONE*. Public Library of Science; 2018; 13: e0206695. <https://doi.org/10.1371/journal.pone.0206695> PMID: 30439978
22. Maruyama E, Hosoi Y, Ueno S, Oonishi N, Totsuka S, Iwai J, et al. Somatic embryogenic cell induction from seeds of pollen-free sugi (*Cryptomeria japonica*) produced at the Niigata prefecture. *Kanto Shinrin Kenkyu*. 2017; 68: 41–44.
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. Oxford University Press; 2014; 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
24. Wei F-J, Ueno S, Ujino-Ihara T, Saito M, Tsumura Y, Higuchi Y, et al. Inspecting abundantly expressed genes in male strobili in sugi (*Cryptomeria japonica* D. Don) via a highly accurate cDNA assembly. *bioRxiv*. 2020;: 2020.04.21.054320.
25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. SN; 9: 357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
26. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011; 12: 323. <https://doi.org/10.1186/1471-2105-12-323> PMID: 21816040
27. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013; 8: 1494–1512. <https://doi.org/10.1038/nprot.2013.084> PMID: 23845962
28. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucl Acids Res*. 2017; 45: D158–D169. <https://doi.org/10.1093/nar/gkw1099> PMID: 27899622
29. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*. 2016; 44: D279–D285. <https://doi.org/10.1093/nar/gkv1344> PMID: 26673716
30. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009; 26: 139–140. <https://doi.org/10.1093/bioinformatics/btp616> PMID: 19910308
31. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. 2nd ed.
32. de Vega-Bartol JJ, Santos RR, Simões M, Miguel CM. Normalizing gene expression by quantitative PCR during somatic embryogenesis in two representative conifer species: *Pinus pinaster* and *Picea abies*. *Plant Cell Reports*. 2013; 32: 715–729. <https://doi.org/10.1007/s00299-013-1407-4> PMID: 23529547
33. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012; 13: 134. <https://doi.org/10.1186/1471-2105-13-134> PMID: 22708584

34. Bravo S, Bertin A, Turner A, Sepúlveda F, Jopia P, Parra MJ, et al. Differences in DNA methylation, DNA structure and embryogenesis-related gene expression between embryogenic and non embryogenic lines of *Pinus radiata* D. don. *Plant Cell Tiss Organ Cult*. Springer Netherlands; 2017; 130: 521–529. <https://doi.org/10.1007/s11240-017-1242-3>
35. Gaj MD, Zhang S, Harada JJ, Lemaux PG. Leafy cotyledon genes are essential for induction of somatic embryogenesis of *Arabidopsis*. *Planta*. 2005; 222: 977–988. <https://doi.org/10.1007/s00425-005-0041-y> PMID: 16034595
36. Mayer KFX, Schoof H, Haecker A, Lenhard M, Jürgens G, Laux T. Role of WUSCHEL in Regulating Stem Cell Fate in the *Arabidopsis* Shoot Meristem. *Cell*. Elsevier; 1998; 95: 805–815. [https://doi.org/10.1016/s0092-8674\(00\)81703-1](https://doi.org/10.1016/s0092-8674(00)81703-1) PMID: 9865698
37. Schoof H, Lenhard M, Haecker A, Mayer KFX, Jürgens G, Laux T. The Stem Cell Population of *Arabidopsis* Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. *Cell*. 2000; 100: 635–644. PMID: 10761929
38. Sarkar AK, Luijten M, Miyashima S, Lenhard M, Hashimoto T, Nakajima K, et al. Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. *Nature*. 2007; 446: 811–814. <https://doi.org/10.1038/nature05703> PMID: 17429400
39. Hirakawa Y, Kondo Y, Fukuda H. TDIF Peptide Signaling Regulates Vascular Stem Cell Proliferation via the WOX4 Homeobox Gene in *Arabidopsis*. *Plant Cell*. 2010; 22: 2618. <https://doi.org/10.1105/tpc.110.076083> PMID: 20729381
40. Ji J, Strable J, Shimizu R, Koenig D, Sinha N, Scanlon MJ. WOX4 Promotes Procambial Development. *Plant Physiol*. 2010; 152: 1346. <https://doi.org/10.1104/pp.109.149641> PMID: 20044450
41. Wójcik AM, Wójcikowska B, Gaj MD. Current Perspectives on the Auxin-Mediated Genetic Network that Controls the Induction of Somatic Embryogenesis in Plants. *IJMS*. 2020; 21: 1333–19. <https://doi.org/10.3390/ijms21041333> PMID: 32079138
42. Wang X, Niu Q-W, Teng C, Li C, Mu J, Chua N-H, et al. Overexpression of PGA37/MYB118 and MYB115 promotes vegetative-to-embryonic transition in *Arabidopsis*. *Cell Research*. 2009; 19: 224–235. <https://doi.org/10.1038/cr.2008.276> PMID: 18695688
43. Hedman H, Zhu T, Arnold von S, Sohlberg JJ. Analysis of the WUSCHEL-RELATED HOMEBOX gene family in the conifer *Picea abies* reveals extensive conservation as well as dynamic patterns. 2013; 13: 89. <https://doi.org/10.1186/1471-2229-13-89> PMID: 23758772
44. Li Q, Zhang S, Wang J. Transcriptome analysis of callus from *Picea balfouriana*. *BMC Genomics*. 2014; 15: 553. <https://doi.org/10.1186/1471-2164-15-553> PMID: 24993107
45. Thompson EW, Lane BG. Relation of protein synthesis in imbibing wheat embryos to the cell-free translational capacities of bulk mRNA from dry and imbibing embryos. *J Biol Chem*. 1980; 255: 5965–5970. PMID: 7380848
46. Membré N, Berna A, Neutelings G, David A, David H, Staiger D, et al. cDNA sequence, genomic organization and differential expression of three *Arabidopsis* genes for germin/oxalate oxidase-like proteins. *Plant Molecular Biology*. 1997; 35: 459–469. <https://doi.org/10.1023/a:1005833028582> PMID: 9349269
47. Kim HJ, Triplett BA. Cotton fiber germin-like protein. I. Molecular cloning and gene expression. *Planta*. 2004; 218: 516–524. <https://doi.org/10.1007/s00425-003-1133-1> PMID: 14648117
48. Domon JM, Dumas B, Lainé E, Meyer Y, David A, David H. Three glycosylated polypeptides secreted by several embryogenic cell cultures of pine show highly specific serological affinity to antibodies directed against the wheat germin apoprotein monomer. *Plant Physiol*. 1995; 108: 141–148. <https://doi.org/10.1104/pp.108.1.141> PMID: 7784502
49. Neutelings G, Domon JM, Membré N, Bernier F, Meyer Y, David A, et al. Characterization of a germin-like protein gene expressed in somatic and zygotic embryos of pine (*Pinus caribaea* Morelet). *Plant Molecular Biology*. 1998; 38: 1179–1190. <https://doi.org/10.1023/a:1006033622928> PMID: 9869423
50. Lippert D, Zhuang J, Ralph S, Ellis DE, Gilbert M, Olafson R, et al. Proteome analysis of early somatic embryogenesis in *Picea glauca*. *Proteomics*. 2005; 5: 461–473. <https://doi.org/10.1002/pmic.200400986> PMID: 15627954
51. Galland R, Blervacq A-S, Blassiau C, Smagghe B, Decottignies J-P, Hilbert J-L. Glutathione-S-Transferase is Detected During Somatic Embryogenesis in Chicory. *Plant Signal Behav*. Landes Bioscience; 2007; 2: 343–348. <https://doi.org/10.4161/psb.2.5.4652> PMID: 19516999
52. Singla B, Tyagi AK, Khurana JP, Khurana P. Analysis of expression profile of selected genes expressed during auxin-induced somatic embryogenesis in leaf base system of wheat (*Triticum aestivum*) and their possible interactions. *Plant Molecular Biology*. 2007; 65: 677–692. <https://doi.org/10.1007/s11103-007-9234-z> PMID: 17849219
53. Uddenberg D, Valladares S, Abrahamsson M, Sundström JF, Sundås-Larsson A, Arnold von S. Embryogenic potential and expression of embryogenesis-related genes in conifers are affected by

treatment with a histone deacetylase inhibitor. *Planta*. 2011; 234: 527–539. <https://doi.org/10.1007/s00425-011-1418-8> PMID: 21541665

54. Morel A, Teyssier C, Trontin J-F, Eliášová K, Pešek B, Beaufour M, et al. Early molecular events involved in *Pinus pinaster* Ait. somatic embryo development under reduced water availability: transcriptomic and proteomic analyses. *Physiol Plant*. 2014; 152: 184–201. <https://doi.org/10.1111/ppl.12158> PMID: 24460664