

RESEARCH ARTICLE

Open Access

Transcriptome analysis of callus from *Picea balfouriana*

Qingfen Li, Shougong Zhang and Junhui Wang*

Abstract

Background: *Picea likiangensis* var. *balfouriana* (Rehd. et Wils.) Hillier ex Slavin (also known as *Picea balfouriana*) is an ecologically and economically important conifer that grows rapidly under optimum conditions and produces high-quality wood. It has a wide geographic distribution and is prevalent in southwest and eastern regions of China. Under suboptimal conditions, *P. balfouriana* grows slowly, which restricts its cultivation. Somatic embryogenesis has been used in the mass propagation of commercial species. However, low initiation rates are a common problem and the mechanisms involved in the induction of somatic embryogenesis are not fully understood. To understand the molecular mechanisms regulating somatic embryogenesis in *P. balfouriana*, high-throughput RNA-seq technology was used to investigate the transcriptomes of embryogenic and non-embryogenic tissues from three *P. balfouriana* genotypes. We compared the genes expressed in these tissues to identify molecular markers with embryogenic potential.

Results: A total of 55,078,846 nucleotide sequence reads were obtained for the embryogenic and non-embryogenic tissues of *P. balfouriana*, and 49.56% of them uniquely matched 22,295 (84.3%) of the 26,437 genes in the *Picea abies* genome database (Nature 497: 579-584, 2013). Differential gene expression analysis identified 1,418 differentially expressed genes (false discovery rate <0.0001; fold change ≥ 2) in the embryogenic tissues relative to the non-embryogenic tissues, including 431 significantly upregulated and 987 significantly downregulated genes. KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis revealed that the most significantly altered genes were involved in plant hormone signal transduction, metabolic pathways (starch and sucrose metabolism), and phenylalanine metabolism.

Conclusions: We found that the initiation of embryogenic tissues affected gene expression in many KEGG pathways, but predominantly in plant hormone signal transduction, plant-pathogen interaction, and starch and sucrose metabolism. The changes in multiple pathways related to induction in the *P. balfouriana* embryogenic tissues described here, will contribute to a more comprehensive understanding of the mechanisms involved in the initiation of somatic embryogenesis. Additionally, we found that somatic embryogenesis receptor kinase (*SERK*), arabinogalactan proteins, and members of the WUS-related homeobox protein family may play important roles and could act as molecular markers in the early stage of somatic embryogenesis, as reported previously.

Keywords: *Picea balfouriana*, Somatic embryogenesis, Embryogenic tissue, Non-embryogenic tissue

Background

Picea balfouriana grows in a specific type of forest ecosystem in southwest China and is an eastern species on the Tibetan plateau. For rapid growth and production of high-quality wood, many plantations selected *P. balfouriana* as one of the main species for afforestation. Seedlings of *P. balfouriana* are produced primarily by sexual propagation;

however, the seeds set late in the season, and the seedlings grow very slowly during the early phases.

Propagation by somatic embryogenesis (SE) is the primary technology used for many conifer biotechnology products, including the development of transgenic trees. The SE system for most spruce (*Picea* spp.) and some pine (*Pinus* spp.) species is sufficiently refined for commercial use; however, for many economically important conifer species, the initiation frequency of SE is insufficient for commercial applications. In recent years, much research has been focused on the development and optimization of protocols for the induction and maturation of somatic

* Correspondence: wangjh@caf.ac.cn

State Key Laboratory of Forest Genetics and Tree Breeding, Research Institute of Forestry, Chinese Academy of Forestry, Number 1 of Dongxiaofu in Haidian District, Beijing, China

embryos of coniferous species [1]. The developmental stage of zygotic embryos, the contents of the growth or culture medium, the type and concentration of plant growth regulators, and culture condition factors such as pH, agar type, and nitrogen level can affect the generation of embryogenic tissue (ET). Therefore, understanding and balancing these factors is crucial for increasing the initiation rate of SE. A complete understanding of the genes involved in the development of ET and non-embryogenic tissue (NET) will improve the understanding of the initiation process in embryogenic tissue.

Cellular identity and function are determined by the transcriptome, i.e. the complete repertoire of expressed RNA transcripts. Transcriptome profiling is a powerful method that has been used widely to assess the relative importance of gene products in cells, tissues, organisms, or under different conditions. Next-generation sequencing (NGS) technology is a cutting-edge approach for high-throughput sequence determination, which has dramatically improved the efficiency and speed of gene discovery [2,3]. For example, Illumina sequencing technology generates more than one billion bases of high-quality DNA sequence per run at less than 1% of the cost of capillary-based methods [4]. Thus, NGS has significantly accelerated and improved the sensitivity of gene-expression profiling, and is expected to boost collaborative and comparative genomics studies [5,6]. Illumina sequencing of transcriptomes for organisms with completed genomes confirmed that the relatively short reads produced by this methodology can be assembled effectively and used for gene discovery and comparisons of gene expression profiles [7,8]. Despite the obvious potential, NGS methods have not yet been used in *P. balfouriana*.

The high-throughput RNA-seq methodology involves whole-transcriptome shotgun sequencing, in which mRNA or cDNA is mechanically fragmented to produce overlapping short fragments that cover the entire transcriptome [9-11]. By mapping the RNA-seq reads to a reference genome, gene assignments and large-scale functional annotation of genes can be carried out. Further, compared with DNA microarray data, RNA-seq data can be used to detect and quantify RNAs that are expressed at very low levels [12]. Additionally, Gene Ontology (GO) analysis of the annotated genes can lead to a deeper understanding of the functions of the genes in cells, thereby providing sensitive and accurate profiling of the transcriptome and a description of gene function that resembles the biology of the cell [13].

Very few studies reporting the mechanism of initiation during SE have been published so far. The main objective of the present study was to identify differentially expressed genes in ETs and NETs, independent of genotype, through RNA-seq analysis and mapping to the reference *Picea abies* genome [14]. The results of

this work may provide fundamental insights into the early SE in *P. balfouriana*.

Results

RNA-seq library sequencing

Six *P. balfouriana* RNA-seq libraries were produced and sequenced; namely, the ETs and NETs from three genotypes. In each library, 11.7–12.5 million clean reads were obtained after the low-quality reads were filtered out (Table 1). Among the clean reads, 8.4–10.9 million (71.24–87.31%) were mapped to known genes (Table 1).

Approximately 19,000 genes (short read mapped to reference genome) were detected in each library, accounting for more than 70% of the 26,437 genes in the *P. abies* reference genome database [14] (Table 2). Overall, 22,295 unique genes were expressed among the six libraries (Additional file 1), which gives a mapping coverage of 84.3%, conferring high confidence in the RNA-seq results. The mean coverage of all genes was above 60%, with the highest coverage at 99.96%. GO terms were assigned to the mapped genes and an enrichment analysis of the GO terms showed that intracellular organelle part, anion binding, and primary metabolic process were dominant in the cellular component, molecular function, and biological process categories, respectively. Some of the genes assigned primary metabolic process terms were associated with KEGG pathways involved in purine metabolism, RNA polymerase, pyrimidine metabolism, inositol phosphate metabolism and phenylalanine, and tyrosine and tryptophan biosynthesis (Additional file 2).

Uniform genes with consistent fold-changes in ETs and NETs from the three genotypes

After two-factor analysis of variance, 4,734 genes with significantly different effects for “embryonic vs. non-embryonic” ($P < 0.05$) and insignificant effects for “genotype” ($P > 0.05$) were selected (Additional file 3). Then, the selected genes were analyzed using DESeq (version 2.14) to test for differentially expressed genes (DEGs) in ETs and NETs, independent of genotype (false discovery rate (FDR) < 0.001 , fold change ≥ 2). Ultimately, 1,418 DEGs between ET and NET were identified, among which 431 were upregulated and 987 were downregulated (Additional file 4). Some of the DEGs were associated with SE (germin-like proteins, abscisic acid receptor, cytochrome P450, and chitinases). About one-quarter of the genes were upregulated by ≥ 2 -fold, and about 1,000 genes were downregulated by ≥ 2 -fold.

Oxidoreductase activity, cell wall, and apoplast were the most enriched GO terms, while biosynthesis of secondary metabolites, metabolic pathways, pentose and glucuronate interconversion, starch and sucrose metabolism, phenylpropanoid biosynthesis, plant hormone signal transduction, flavonoid biosynthesis, phenylalanine metabolism, ether lipid

Table 1 Mapping of *Picea balfouriana* RNA-seq library reads to the *Picea abies* reference genome database

Classification	ET Genotype1	NET Genotype 1	ET Genotype 2	NET Genotype 2	ET Genotype 3	NET Genotype 3
Raw reads	12,519,224	11,703,983	12,150,702	11,734,531	12,150,562	12,280,320
Base pairs	613,419,76	573,495,167	595,384,398	574,992,019	595,377,538	601,735,680
%	100.00	100.00	100.00	100.00	100.00	100.00
Mapped reads	5,713,055	5,044,017	5,590,502	4,359,900	5,519,889	5,456,021
%	45.63	43.10	46.01	37.15	45.43	44.43
Perfect match	3,923,702	3,401,759	3,865,351	2,947,585	3,870,249	3,674,461
%	31.34	29.06	31.81	25.12	31.85	29.92
<=2 bp mismatch	1,789,353	1,642,258	1,725,151	1,412,315	1,649,640	1,781,560
%	14.29	14.03	14.20	12.04	13.58	14.51
Unique match	5,374,869	47,721,812	5,262,180	4,081,332	5,214,938	5,117,424
%	42.93	40.34	43.31	34.78	42.92	41.67
Multi-position match	338,186	322,205	328,322	278,568	304,951	338,597
%	2.70	2.75	2.70	12.42	2.51	2.76
Unmapped reads	6,806,169	6,659,966	6,560,200	7,374,631	6,630,673	6,824,299
%	54.37	56.90	53.99	62.85	54.57	55.57

ET, embryogenic tissue; NET, non-embryogenic tissue.

metabolism, and propanoate metabolism were the most overrepresented KEGG pathways among genes related to embryogenic competence (Additional file 5). These findings will be very important for further understanding the intracellular signaling mechanisms of early SE in conifers.

Expression of housekeeping genes in the different tissue samples of *P. balfouriana*

The NormFinder algorithm uses a model-based approach to estimate alterations in the expression of housekeeping genes [15]. The algorithm also estimates variations across subgroups and avoids the artificial selection of co-regulated genes. The results of the NormFinder analysis of the genes detected in the six libraries show that WS0109_C05 [TAIR: At1g29260] and 18S rRNA were predicted to be the most stable housekeeping genes (Table 3).

BestKeeper measures stability using a pair-wise correlation analysis of all pairs of candidate genes and calculates the geometric mean of the best candidates [16,17]. A preliminary analysis based on the inspection of raw Ct values estimated the variation of four housekeeping genes (WS00912.B21_N13, WS0109_C05, 18S rRNA, and tubulin) to be compatible with an overall stability in gene expression (Table 4). The standard deviation values were <1. The four housekeeping genes were used

to calculate the BestKeeper index. BestKeeper allows a comparative analysis across internal reference genes by estimating correlations in the expression levels between all the possible candidates. Highly correlated control genes are combined into an index, and the pair-wise correlation between genes and the correlation between each gene and the index are also calculated. The results describe the consistency between the index and each internal reference gene. The four control genes tested in our analysis correlated well one with one another and with the NormFinder index (Table 3). The best correlation between one internal reference gene and the BestKeeper index was for WS0109_C05 ($r = 0.947$). The statistically significant correlation for WS0109_C05 with the BestKeeper index appeared to be consistent with the good performance of this housekeeping gene as assessed by NormFinder.

Quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR) analysis

The peroxisomal targeting signal receptor WS0109_C05 displayed constitutive expression in all the analyses performed using the *P. balfouriana* RNA-seq data; therefore, it was used as an internal control for the normalization of gene expression levels (Δ CT). Total RNA from the same six tissue samples that were used for the RNA-seq analysis

Table 2 Genes found in *Picea balfouriana* that uniquely match genes in the *Picea abies* reference genome

	ET Genotype 1	NET Genotype 1	ET Genotype 2	NET Genotype 2	ET Genotype 3	NET Genotype 3
Uniquely matched genes	18865	19256	19432	19143	19116	19287
Percentage	71.36%	72.84%	73.50%	72.41%	72.31%	72.95%

ET, embryogenic tissue; NET, non-embryogenic tissue.

Table 3 Ranking of *Picea balfouriana* candidate reference genes generated by NormFinder

Gene name	Stability value	Standard error
WS00912.B21_N13	0.403	0.268
WS0109_C05	0.282	0.311
18S rRNA	0.923	0.324
Tubulin	0.779	0.295

were used as templates for the qRT-PCR analysis (Table 5). For most of the genes, the transcript fold-ratios determined by qRT-PCR were approximately the same as those estimated from the RNA-seq data, indicating the reliability of the RNA-seq in *P. balfouriana*.

Discussion

Most early studies on the initiation mechanism of SE in *P. balfouriana* focused mainly on different genes between ET and NET from just one cell line. It has been suggested that genotype may have an important effect on initiation of ET. In this study, we demonstrated that many genes were influenced by genotype, so we filtered them out and selected the genes that were related only to embryogenic ability. Six RNA-seq libraries were created to analyze the DEGs in ET and NET, yielding a total of 22,295 genes. Among them, 9,988 (44.8%) and 13,135 (58.9%) were enriched in GO terms and KEGG pathways, respectively. Our transcriptome and gene expression profiling data will greatly enrich the current knowledge for the genus *Picea*, and will contribute to the database for conifer species. These data will also help promote research on the identification of novel genes related to embryogenic competence.

Several genes involved in the response to phytohormone and hormone-mediated signaling pathways in ETs were strongly upregulated (Additional file 4); for example, the genes that encode auxin-induced proteins. The initiation of embryogenesis in somatic tissues requires auxins,

especially in conifer species. Exogenous hormones and culture-related stress has been reported to play important roles in triggering SE, which presumably results from the expression of genes involved in cell division, cell polarization, and the regulation of signal transduction pathways. Several auxin-inducible genes that are involved in cellular changes, such as activation of cell division, have been cloned [18,19]. Bögre et al. [20] used auxin-responsive genes as molecular markers to distinguish embryogenic from non-embryogenic genotypes of *Medicago sativa*. Our results indicated that most of the DEGs were involved in plant-pathogen interaction, metabolic pathways (starch and sucrose metabolism), and plant hormone signal transduction (Additional file 5) and, therefore, they may play essential roles in the initiation of SE.

A number of genes encoding membrane-localized, leucine-rich, repeat receptor-like kinases, such as IMK2 [Swiss-Prot:Q9SCT4], were among the DEGs. The somatic embryogenesis receptor kinase (*SERK*) gene encodes a leucine-rich repeat receptor-like kinase that plays an important role in plant signaling pathways. The *SERK* gene was first identified in carrot (*Daucus carota*) suspension cultures where it was specifically expressed in cells that developed into somatic embryos [21]. *SERK* has been linked to SE in a number of species, including *Dactylis glomerata* [22], *Arabidopsis thaliana* [23], *Medicago truncatula* [24], sunflower (*Helianthus annuus*) [25], *Ocotea catharinensis* [26], *Citrus unshiu* [27], and *Theobroma cacao* [28]. There is evidence that *SERK* genes are required for embryo initiation in embryogenic cells [21,23,29]. In *D. glomerata*, and *A. thaliana*, *SERKs* are characteristic markers of embryogenic cell cultures and somatic embryos [21-23].

Several arabinogalactan proteins (AGPs) were upregulated in ETs compared with NETs, such as FLA7 [Swiss-Prot:Q9SJ81]. AGPs have been implicated in five fundamental cellular processes: cell proliferation, cell

Table 4 Statistical output from the BestKeeper analysis for the *Picea balfouriana* candidate reference genes

	WS00912.B21_N13	WS0109_C05	18SrRNA	Tubulin
N	6	6	6	6
G Mean [Ct]	28.03	27.57	13.12	24.37
A Mean [Ct]	28.06	27.58	13.13	24.4
Min [Ct]	26.21	26.38	12.33	22.62
Max [Ct]	29.78	28.79	13.81	25.68
SD [±Ct]	0.85	0.58	0.56	1.06
CV [% Ct]	3.02	2.09	4.26	4.35
Coeff. of corr. [r]	0.85	0.947	-0.014	0.605
P-value	0.032	0.004*	0.978	0.203

N, number of samples; G Mean [Ct], geometric mean of the Ct; A Mean [Ct], arithmetic mean of the Ct; Min [Ct] and Max [Ct], extreme values of the Ct; SD [±Ct], standard deviation of the Ct; CV [% Ct], coefficient of variance expressed as percentage on the Ct level. Coeff. of corr. [r], coefficient of correlation. *Indicates the best correlation between the control genes and the BestKeeper index.

Table 5 Comparison of RNA-seq data with expression data from qRT-PCR

NCBI gene ID	UniProt_ID	Description	Method	Expression level in ET	Expression level in NET	Fold change
840475	Swiss-Prot:Q9LP24	Leucine-rich repeat receptor-like protein kinase	RNA-seq	Log-cpm in ET 605.5	Log-cpm in NET 3640.6	log2 ratio (ET/NET) 2.6
			qRT-PCR	ET(Δ CT) 9.6	NET(Δ CT) 4.6	$2^{-\Delta\Delta$ CT) 2.1
817958	Swiss-Prot:Q6X7J4	WUSCHEL-related homeobox 9	RNA-seq	Log-cpm in ET 155.7412756	Log-cpm in NET 0.353625	log2 ratio (ET/NET) 8.8
			qRT-PCR	ET(Δ CT) 3.0	NET(Δ CT) 1.3	$2^{-\Delta\Delta$ CT) 3.4
831649	Swiss-Prot:Q8GY25	WUSCHEL-related homeobox 12	RNA-seq	Log-cpm in ET 136.2	Log-cpm in NET 1.8	log2 ratio (ET/NET) 6.3
			qRT-PCR	ET(Δ CT) 0.7	NET(Δ CT) 0.1	$2^{-\Delta\Delta$ CT) 1.5

Three biological repeats were reverse transcribed and amplified independently. The raw threshold cycle (Ct) values were normalized against the WS0109_C05 standard to generate the normalized Δ Ct values that were used to calculate the fold-change in expression in embryonic tissue (ET) and non-embryonic tissue (NET) of the three genotypes. Log-cpm, log-counts per million.

expansion, cell differentiation, programmed cell death, and cell-cell communication [30]. The involvement of AGPs in SE has been studied previously [31-37], and the role of AGPs in the establishment of embryogenic cell cultures and the influence of AGPs on cell development have also been examined. The presence of AGPs in seeds was found to increase the number of pro-embryogenic masses and the embryogenic potential [38-41]. Filonova et al. [42] reported the importance of intercellular communication for the acquisition of embryogenic competence, and observed that separation of Norway spruce cells by fractionation of suspension cultures inhibited the embryogenic process. AGPs restore differentiation potential after cell wall removal, and this restoration was reported to be more efficient with chitinase-cleaved forms of AGPs [36]. Based on these observations, we propose that complex interactions between cells and substances secreted in the medium of embryogenic cultures are essential to establish and maintain embryogenic competence in culture.

The qRT-PCR results confirmed the observed upregulation in ETs of two genes belonging to the WUS-related homeobox (*WOX*) family. Previous studies have reported the expression dynamics of *WOX* and shown that it is highly expressed during early somatic embryo development, but declines as the embryo matures [43-45]. Haecker et al. [43] found that *WOX8* and *WOX9* were expressed in *A. thaliana* during embryo development from the very early stages and throughout development. Several *WOX* genes, including *WOX2*, were reported to constitute potential markers of cell fate during early embryogenesis, suggesting that they might have important functions in early embryonic patterning [44]. Palovaara et al. [45] found that *WOX2* played a fundamental role in early somatic

embryo development in *P. abies*, possibly related to the regulation of cell division and/or differentiation in the embryos. It was also proposed that *WOX2* could be used as a molecular marker of embryogenic potential, but was not necessary for the regenerative capacity of cell lines [45]. Klimaszewska et al. [46] reported that *WOX2* was expressed exclusively in the early stages of SE, and might be useful as a marker of embryogenic potential. *WOX3*, *WOX8A*, and *WOX8B* were all significantly expressed in proembryogenic masses in *P. abies* rather than in somatic embryos [47].

Conclusions

Although ET and NET can be discriminated by visual inspection, and molecular markers of embryogenic potential are useful, bioinformatics tools provide a powerful approach to identify DEGs in ETs and NETs. Here, we found gene expression patterns in ETs from *P. balfouriana* that identified changes in multiple pathways related to early SE, such as plant hormone signal transduction, metabolic pathways (starch and sucrose metabolism), and phenylpropanoid biosynthesis. If these pathways could be modified to induce ETs with higher initiation rates that have competence to form normal somatic embryos they could be used to increase the initiation frequency of SE for commercial applications.

A large number of candidate genes, such as those that encode heat shock proteins and glutathione S-transferases, showed significant different expression in the ET libraries compared with their expression in the NET libraries; however, these results require further verification and characterization. Although the molecular functions of individual *P. balfouriana* genes and the associated signal

transduction pathways often remained largely unknown, the RNA-seq analysis has provided valuable information about the induction of SE in *P. balfouriana*, which could promote further investigations into the detailed physiological initiation mechanisms of SE. This analysis represents a starting point for detailed functional studies; however, further experiments are required to expand on these results and to define the complex interaction networks and molecular mechanisms responsible for the induction of SE in *P. balfouriana*.

Methods

Plant material

ETs and NETs were collected from three genotypes of *P. balfouriana* (Figure 1), and initiated and maintained on 1/2 LM basal medium [48], supplemented with $2.2 \text{ mg} \cdot \text{L}^{-1}$ 2,4-dichlorophenoxyacetic acid (2,4-D), $1.1 \text{ mg} \cdot \text{L}^{-1}$ BAP, $0.2 \text{ mg} \cdot \text{L}^{-1}$ KT, 1% sucrose, $1000 \text{ mg} \cdot \text{L}^{-1}$ casein hydrolysate, 1% sucrose, $500 \text{ mg} \cdot \text{L}^{-1}$ glutamine, and 0.2% gellan gum (Sigma, St. Louis, MO, USA).

RNA isolation

Total RNA was extracted immediately from fresh ET and NET samples, TRIZOL® reagent (Invitrogen, Carlsbad, CA, USA) was used to extract total RNA from 50 to 100 mg of tissue, according to the manufacturer's instructions. The RNA samples were incubated for 30 min at 37°C with 10 U of DNaseI (Takara, Dalian, China) to remove residual genomic DNA. The quality and quantity of the purified RNA were determined by measuring the absorbance at 260 nm and 280 nm (A260/A280) using a Nano-drop® ND-1000 spectrophotometer (LabTech, Holliston, MA, USA). The samples had an average RNA integrity number (RIN) of 8.9 based on a lab-on-chip analysis using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

RNA-seq library preparation and sequencing

After extracting total RNA from the six samples, oligo (dT) magnetic beads were used to enrich the mRNA by removing rRNA. Fragmentation buffer was used to break

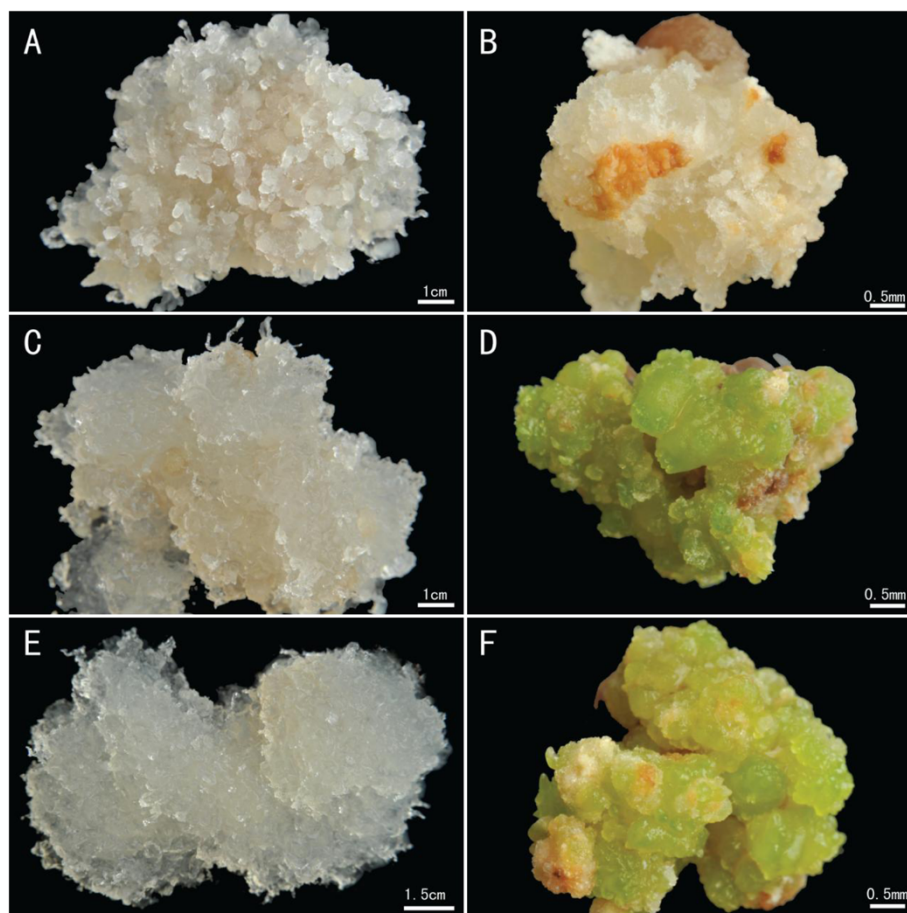


Figure 1 Embryonic tissue (ET) and non-embryonic tissue (NET) of three *Picea balfouriana* genotypes. (A) and (D) ET (bar = 1.0 cm) and NET (bar = 1.5 cm) of genotype 1, respectively; (B) and (E) ET (bar = 1.0 cm) and NET (bar = 1.5 cm) of genotype 2, respectively; (C) and (F) ET (bar = 1.0 cm) and NET (bar = 2.0 cm) of genotype 3, respectively.

the mRNA into short fragments (about 200-bp long), Then a random hexamer-primer was used to generate first-strand cDNA using the mRNA fragments as templates. Buffer, dNTPs, RNase H, and DNA polymerase I were added to synthesize the second-strand cDNA. A QiaQuick PCR extraction kit (Qiagen, Hilden, Germany) was used to purify the double-strand cDNA, which was then washed with EB buffer for end repair and poly(A) addition. Finally, sequencing adaptors were ligated onto the fragments. Agarose gel electrophoresis was used to purify the fragments, which were then enriched by PCR amplification. The library products were sequenced using a Illumina HiSeq™ 2000 system (Illumina, San Diego, USA).

Analysis and mapping of RNA-seq reads

The sequenced raw data were filtered to remove low-quality tags (reads with unknown nucleotides “N”), empty reads (no read sequence between the adaptors), and reads with only one copy number (reads that might have resulted from sequencing errors). SOAPaligner/soap2 (version2.21) [49] was used to align the remaining clean reads to the sequences in the *P. abies* reference genome database, allowing up to two base mismatches. At most, one continuous gap or two mismatches were allowed in the high-quality part of a read. The mapped clean reads were designated as unambiguous clean reads. For two-factor analysis of variance, the number of unambiguous clean reads for each gene was calculated and normalized to log-counts per million using the latest version of the limma package in R [50]. The RNA-seq data consisted of a matrix of read counts r_{gi} , for RNA samples $i = 1$ to n , and genes $g = 1$ to G . If R_i for the total number of mapped reads for sample i is defined as $R_i = \sum_{g=1}^G r_{gi}$, then the log-counts per million (log-cpm) value for each count can be calculated as $y_{ga} = \log_2 \left(\frac{r_{gi} + 0.5}{R_i + 1.0} \times 10^6 \right)$.

The counts were offset away from zero by adding 0.5 to avoid taking the log of zero, and to reduce the variability of log-cpm for low expression genes. The library size was offset by 1 to ensure that $(r_{gi} + 0.5)/(R_i + 1)$ was always less than 1 and greater than zero.

Differentially expressed genes between the ET and NET libraries

Two-factor analysis of variance for each gene, using two factors, “embryonic vs. non-embryonic” and “genotype”, was performed in R (version 2.15.) to detect genes that were differentially expressed between the ET and NET libraries, and not affected by genotype. DESeq was also used to detect the DEGs between the ETs and NETs [51]. Finally, genes with significant effects from “embryonic vs. non-embryonic” ($P < 0.05$) and insignificant effects from “genotype” ($P > 0.05$) in two-factor analysis of variance were

selected from the significantly changed genes detected by DESeq and considered as the DEGs between the ET and NET libraries. The FDRs of these DEGs were less than 0.001 and the absolute values of the \log_2 ratio were greater than 1 [52]. Briefly, if R DEGs are selected and S genes are genuinely differentially expressed while V genes are false positives, and if the error ratio ($Q = V/R$) is required to remain below a cutoff (e.g. 0.001), then the FDR can be calculated according to the Benjamini and Hochberg algorithm [52] as: $FDR = E(Q) = E\{V/(V+S)\} = E(V/R)$.

In this study, a FDR of ≤ 0.001 was used.

All the genes, including the DEGs, were used for the GO and KEGG ontology (KO) enrichment analyses. For the GO analysis, a corrected P-value of < 0.05 was used as the threshold to determine significant enrichment of the gene sets. For KO enrichment analysis, a Q-value ≤ 0.05 was used as the threshold to determine significant enrichment of the gene sets.

Validation of DEGs by qRT-PCR

Four reference housekeeping genes were selected for *P. balfouriana* and forward and reverse primers were designed as follows: WS00912.B21_N13 (5'-GTCGTG TGGATTGTCTCTGC-3'; 5'-ATGTATTCGAAGAGGA GGAATG-3'), WS0109_C05 (5'-AACTGTCATTGGA GTCCTGTAGG-3'; 5'-TGAGGGTTTGGGT AGTGA GA-3'), tubulin (PaTubF: 5'-CGTTACCTGCTGCCT GAG-3', PaTubR: 5'-G CTCTGTATTGCTGTGAAC C-3'), and 18S rRNA (Pa18SF:5'-CGGCGGATG TT GCTCTAAG-3'; Pa18SR:5'-TCTGTCAATCCTTACTA TGCTCTGG-3'). Those gene has been used as a reference in other studies [53,54]. The most stable gene (WS0109_C05) was selected as the reference gene. Details of the qRT-PCR analysis of spruce transcripts has been described previously [55]. In brief, total RNA (18 μ g) from each sample was treated with DNase I (Invitrogen). The absence of DNA in the treated RNA samples (10 ng) was confirmed by PCR using the WS0109_C05 primers. Next, RNA (three reactions of 4 μ g each per sample) was reverse transcribed, and cDNA (10 ng) was analyzed by PCR in a total volume of 20 μ L, in the presence of 10 μ L DyNAmo SYBR Green Mastermix (FinnZymes, Espoo, Finland) and 0.3 μ M each of a forward and a reverse primers. Primers of 20–24 nucleotides in length (usually located in the 3'-untranslated region) with a T_m of 60°C were designed to amplify a gene-specific, 160–200 bp fragment of the target cDNA. Reactions were carried out in triplicate in an ABI7900HT RT-PCR machine (ABI, Foster City, USA) with an initial step of 3 min at 95°C, followed by 40 cycles of 3 s at 95°C and 20 s at 60–63°C. Each cycle was followed by a data-acquisition step. After the last cycle and a final 10-min extension at 72°C, a melting curve was measured from 60 to 95°C with measurements every 0.2°C and holding for 1 s. Fluorescence was determined following each annealing and extension phase.

Availability of supporting data

The RNA-seq data supporting the results of this article are available at the NCBI under BioProject (<http://www.ncbi.nlm.nih.gov/bioproject/211928>) with accession number PRJNA211928.

Additional files

Additional file 1: Details of the 22,295 genes identified in six libraries of *Picea balfouriana*.

Additional file 2: Enriched GO terms and KEGG pathways for the 22,295 genes identified in six libraries of *Picea balfouriana*.

Additional file 3: Genes selected from two-factor analysis of variance as differentially expressed between the ET and NET libraries.

Additional file 4: Differentially expressed genes between the ET and NET libraries confirmed by DESeq.

Additional file 5: Enriched GO terms and KEGG pathways for the differentially expressed genes between the ET and NET libraries.

Abbreviations

AGP: Arabinogalactan proteins; ET: Embryogenic tissue; NET: Non-embryogenic tissue; FDR: False-discovery rate; *Picea balfouriana*: *Picea likiangensis* var. *balfouriana* (Rehd. et Wils.) Hillier ex Slavín; *SERK*: Somatic embryogenesis receptor-like kinase.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JW and SZ conceived and designed the study. QL performed the experiments and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Financial support for this study was provided by the Public Industrial and Special Scientific Research of Forestry, China (Project 201004009-1).

Received: 23 July 2013 Accepted: 30 June 2014

Published: 4 July 2014

References

1. Stasolla C, Kong L, Yeung EC, Thorpe TA: Maturation of somatic embryos in conifers: morphogenesis, physiology, biochemistry and molecular biology. *In Vitro Cell Dev Biol-Plant* 2002, **38**:93–105.
2. Schuster SC: Next-generation sequencing transforms today's biology. *Nat Methods* 2008, **5**:16–18.
3. Ansorge WJ: Next generation DNA sequencing techniques. *New Biotechnol* 2009, **25**:195–203.
4. Huang W, Marth G: EagleView: a genome assembly viewer for next-generation sequencing technologies. *Genome Res* 2008, **18**:1538–1543.
5. Blow N: Transcriptomics: The digital generation. *Nature* 2009, **458**:239–242.
6. t Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RHAM, de Menezes RX, Boer JM, van Ommen GJB, den Dunnen JT: Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res* 2008, **36**(21):e141.
7. Rosenkranz RG, Borodina T, Lehrach H, Himmelbauer H: Characterizing the mouse ES cell transcriptome with Illumina sequencing. *Genomics* 2008, **92**:187–194.
8. Hegedus Z, Zakrzewska A, Agoston VC, Ordas A, Rác P, Mink M, Spaik HP, Meijer AH: Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Mol Immunol* 2009, **46**:2918–2930.
9. Morozova O, Marra MA: Applications of next-generation sequencing technologies in functional genomics. *Genomics* 2008, **92**(5):255–264.
10. Mortazavi A, Williams BA, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008, **5**(7):621–628.
11. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M: The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 2008, **320**(5881):1344–1349.
12. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keefe S, Haas S, Vingron M, Lehrach H, Yaspo ML: A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 2008, **321**(5891):956–960.
13. Xiang LX, He D, Dong WR, Zhang YW, Shao JZ: Deep sequencing-based transcriptome profiling analysis of bacteria-challenged *Lateolabrax japonicus* reveals insight into the immune-relevant genes in marine fish. *BMC Genomics* 2010, **11**:472–493.
14. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Lexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Käller M, Luthman J, Lysholm F, Niittylä T, Olson Å, Rilakovic N, Ritland C, Rosselló JA, Sena J, et al: The Norway spruce genome sequence and conifer genome evolution. *Nature* 2013, **497**:579–584.
15. Andersen CL, Jensen JL, Orntoft TF: Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 2004, **64**:5245–5250.
16. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP: Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations. *Biotechnol Lett* 2004, **26**:509–515.
17. Spinsanti G, Panti C, Lazzeri E, Marsili L, Marsili L, Casini S, Frati F, Fossi CM: Selection of reference genes for quantitative RT-PCR studies in striped dolphin (*Stenella coeruleoalba*) skin biopsies. *BMC Mol Biol* 2006, **7**:32.
18. Ainley WM, Walker JC, Nagao RT, Key JL: Sequence and characterization of two auxin regulated genes from soybean. *J Biol Chem* 1988, **263**:10658–10666.
19. Takaahashi Y, Kuroda H, Tanaka T, Machida Y, Takebe I, Nagata T: Isolation of an auxin regulated gene cDNA expressed during the transition from G₀ to S phase in tobacco mesophyll protoplasts. *Proc Natl Acad Sci U S A* 1989, **86**:9279–9283.
20. Bögre L, Stefanov I, Abrahám M, Somogyi I, Dudits D: Differences in responses to 2,4-D-dichlorophenoxy acetic acid (2,4-D) treatment between embryogenic and non-embryogenic lines of alfalfa. In *Progress in Plant Cellular and Molecular Biology*. Edited by Nijkamp HJJ, van der Plas LHW, Vartrajk J. Boston: Kluwer Academic Publishers; 1990:427–436.
21. Schmidt EDL, Guzzo F, Toonen MAJ, de Vries SC: A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development* 1997, **124**:2049–2062.
22. Somleva MN, Schmidt EDL, de Vries SC: Embryogenic cells in *Dactylis glomerata* L. (Poaceae) explants identified by cell tracking and by SERK expression. *Plant Cell Rep* 2000, **19**:718–726.
23. Hecht V, Vielle-Calzada JP, Hartog MV, Schmidt EDL, Boutilier K, Grossniklaus U, de Vries SC: 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.
24. Nolan KE, Irwanto RR, Rose RJ: Auxin up-regulates *MtSERK1* expression in both *Medicago truncatula* root-forming and embryogenic cultures. *Plant Physiol* 2003, **133**:218–230.
25. Thomas C, Meyer D, Himber C, Steinmetz A: Spatial expression of a sunflower *SERK* gene during induction of somatic embryogenesis and shoot organogenesis. *Plant Physiol Biochem* 2004, **42**:35–42.
26. Santa-Catarina C, Hanai LR, Dornelas MC, Viana AM, Floh EIS: *SERK* gene homolog expression, polyamines and amino acids associated with somatic embryogenic competence of *Ocotea catharinensis* Mez. (Lauraceae). *Plant Cell Tissue Organ Cult* 2004, **79**:53–61.
27. Shimada T, Hirabayashi T, Endo T, Fujii H, Kita M, Omura M: Isolation and characterization of the somatic embryogenesis receptor-like kinase gene homologue (CitSERK1) from *Citrus unshiu* Marc. *Sci Hortic* 2005, **103**:233–238.
28. de Oliveira SM, Romano E, Yotoko KSC, Tinocoa MLP, Diasa BBA, Aragão FJL: Characterization of the cacao somatic embryogenesis receptor-like kinase (*SERK*) gene expressed during somatic embryogenesis. *Plant Sci* 2005, **168**:723–729.

29. Singla B, Khurana JP, Khurana P: **Characterization of three somatic embryogenesis receptor kinase genes from wheat, *Triticum aestivum*.** *Plant Cell Rep* 2008, **27**:833–843.
30. Shiota H, Satoh R, Watabe K, Harada H, Kamada H: **C-ABI3, the carrot homologue of the Arabidopsis ABI3, is expressed during both zygotic and somatic embryogenesis and functions in the regulation of embryo-specific ABA-inducible genes.** *Plant Cell Physiol* 1998, **39**:1184–1193.
31. Chapman A, Blervacq AS, Vasseur J, Hilbert JL: **Arabinogalactan-proteins in *Cichorium* somatic embryogenesis: effect of β -glucosyl Yariv reagent and epitope localization during embryo development.** *Planta* 2000, **211**:305–314.
32. Leļjak-Levanic D, Naana B, Jelaska MS: **Changes in DNA methylation during somatic embryogenesis in *Cucurbita pepo* L.** *Plant Cell Rep* 2004, **23**:120–127.
33. Stasolla C, Bozhkov PV, Chu TM, van Zyl L, Egertsdotter U, Suarez MF, Craig D, Wolfinger RD, Von Arnold S, Sederoff RR: **Variation in transcript abundance during somatic embryogenesis in gymnosperms.** *Tree Physiol* 2004, **24**:1073–1085.
34. Toonen MAJ, Schmidt EDL, Heo TH, Verhoeven HA, van Kammen A, de Vries SC: **Expression of the JIM8 cell wall epitope in carrot somatic embryogenesis.** *Planta* 1996, **200**:167–173.
35. van der Graa VE, Dulk-Ras AD, Hooykaas PJJ, Keller B: **Activation tagging of the *LEAFY PETIOLE* gene affects leaf petiole development in *Arabidopsis thaliana*.** *Development* 2000, **127**:4971–4980.
36. van Hengel AJ, Guzzo F, van Kammen A, de Vries SC: **Expression pattern of the carrot *EP3* endochitinase genes in suspension cultures and in developing Seeds.** *Plant Physiol* 1998, **117**:43–53.
37. von Recklinghausen IR, Iwanowska A, Kieft H, Mordhorst AP, Schel JHN, van Lammeren AAM: **Structure and development of somatic embryos formed in *Arabidopsis thaliana* pt mutant callus cultures derived from seedlings.** *Protoplasma* 2000, **211**:217–224.
38. Kreuger M, van Holst GJ: **Arabinogalactan proteins are essential in somatic embryogenesis of *Daucus carota* L.** *Planta* 1993, **189**:243–248.
39. Egertsdotter U, von Arnold S: **Importance of arabinogalactan proteins for the development of somatic embryos of Norway spruce (*Picea abies*).** *Physiol Plant* 1995, **93**:334–345.
40. Kreuger M, Postma E, Brouwer Y, van Holst GJ: **Somatic embryogenesis of *Cyclamen persicum* in liquid medium.** *Physiol Plant* 1995, **94**:605–612.
41. Saare-Surminski K, Preilb W, Knox JP, Lieberei R: **Arabinogalactan proteins in embryogenic and non-embryogenic callus cultures of *Euphorbia pulcherrima*.** *Physiol Plant* 2000, **108**:180–187.
42. Filonova LH, von Arnold S, Daniel G, Bozhkov PV: **Programmed cell death eliminates all but one embryo in a polyembryonic plant seed.** *Cell Death Differ* 2002, **9**:1057–1062.
43. 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.
44. Nardmann J, Zimmermann R, Durantini D, Kranz E, Werr W: ***WOX* gene phylogeny in Poaceae: a comparative approach addressing leaf and embryo development.** *Mol Biol Evol* 2007, **24**:2474–2484.
45. 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.
46. Klimaszcwska K, Overton C, Stewart D, Rutledge RG: **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.
47. Hedman H, Zhu TQ, von Arnold S, Sohlberg JJ: **Analysis of the *WUSCHEL-RELATED HOMEODOMAIN* gene family in the conifer *Picea abies* reveals extensive conservation as well as dynamic patterns.** *BMC Plant Biol* 2013, **13**:89.
48. Litvay JD, Verma DC, Johnson MA: **Influence of a loblolly pine (*Pinus taeda* L.) culture medium and its components on growth and somatic embryogenesis of the wild carrot (*Daucus carota* L.).** *Plant Cell Rep* 1985, **4**:325–328.
49. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J: **SOAP2: an improved ultrafast tool for short read alignment.** *Bioinformatics* 2009, **25**(15):1966–1967.
50. Law CW, Chen Y, Shi W, Smyth GK: **Voom! Precision weights unlock linear model analysis tools for RNA-seq read counts.** *Genome Biol* 2014, **15**:R29.
51. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**(10):R106.
52. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc. Series B-Methodological* 1995, **57**(1):289–300.
53. Friedmann M, Ralph SG, Aeschliman D, Zhuang J, Ritland K, Ellis BE, Bohlmann J, Douglas CJ: **Microarray gene expression profiling of developmental transitions in Sitka spruce (*Picea sitchensis*) apical shoots.** *J Exp Bot* 2007, **58**:593–614.
54. Phillips MA, D'Auria JC, Luck K, Gershenzon J: **Evaluation of Candidate Reference Genes for Real-Time Quantitative PCR of Plant Samples Using Purified cDNA as Template.** *Plant Mol Biol Rep* 2009, **27**:407–416.
55. Ralph S, Yueh H, Friedmann M, Zeznik JA, Nelson CC, Butterfield YSN, Kirkpatrick R, Liu J, Jones SJM, Marra MA, Douglas CJ, Ritland K, Bohlmann J: **Conifer defense against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large scale changes of the host transcriptome.** *Plant Cell Environ* 2006, **29**:1545–1570.

doi:10.1186/1471-2164-15-553

Cite this article as: Li et al.: Transcriptome analysis of callus from *Picea balsamifera*. *BMC Genomics* 2014 **15**:553.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

