Journal of Experimental Botany, Vol. 61, No. 14, pp. 3847–3864, 2010
doi:10.1093/jxb/erq196 Advance Access publication 23 August, 2010
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses

Frank Bedon[1,2], Claude Bomal[1], Sébastien Caron[1], Caroline Levasseur[3], Brian Boyle[1], Shawn D. Mansfield[4], Axel Schmidt[5], Jonathan Gershenzon[5], Jacqueline Grima-Pettenati[2], Armand Séguin[3] and John MacKay[1,*]

[1] Centre d'Étude de la Forêt, Université Laval, Québec (QC), G1V A06, Canada
[2] UMR UPS/CNRS 5546, Pôle de Biotechnologies Végétales, 24 chemin de Borde Rouge, BP42617, Auzeville Tolosane, 31326 Castanet Tolosan, France
[3] Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Québec (QC), G1V A06, Canada
[4] Canada Research Chair in Wood and Fibre Quality, Department of Wood Science, University of British Columbia, 4030-2424 Main Mall, Vancouver (BC), V6T 1Z4, Canada
[5] Max Planck Institute for Chemical Ecology, Hans-Knoell-Str.8, Beutenberg-Campus, D-07745 Jena, Germany

* To whom correspondence should be addressed. E-mail: jmackay@rsvs.ulaval.ca

Received 21 December 2009; Revised 9 June 2010; Accepted 10 June 2010

Abstract

Transcription factors play a fundamental role in plants by orchestrating temporal and spatial gene expression in response to environmental stimuli. Several *R2R3-MYB* genes of the *Arabidopsis* subgroup 4 (Sg4) share a C-terminal EAR motif signature recently linked to stress response in angiosperm plants. It is reported here that nearly all Sg4 *MYB* genes in the conifer trees *Picea glauca* (white spruce) and *Pinus taeda* (loblolly pine) form a monophyletic clade (Sg4C) that expanded following the split of gymnosperm and angiosperm lineages. Deeper sequencing in *P. glauca* identified 10 distinct Sg4C sequences, indicating over-represention of Sg4 sequences compared with angiosperms such as *Arabidopsis*, *Oryza*, *Vitis*, and *Populus*. The Sg4C MYBs share the EAR motif core. Many of them had stress-responsive transcript profiles after wounding, jasmonic acid (JA) treatment, or exposure to cold in *P. glauca* and *P. taeda*, with *MYB14* transcripts accumulating most strongly and rapidly. Functional characterization was initiated by expressing the *P. taeda MYB14* (*PtMYB14*) gene in transgenic *P. glauca* plantlets with a tissue-preferential promoter (cinnamyl alcohol dehydrogenase) and a ubiquitous gene promoter (ubiquitin). Histological, metabolite, and transcript (microarray and targeted quantitiative real-time PCR) analyses of PtMYB14 transgenics, coupled with mechanical wounding and JA application experiments on wild-type plantlets, allowed identification of *PtMYB14* as a putative regulator of an isoprenoid-oriented response that leads to the accumulation of sesquiterpene in conifers. Data further suggested that PtMYB14 may contribute to a broad defence response implicating flavonoids. This study also addresses the potential involvement of closely related Sg4C sequences in stress responses and plant evolution.

Key words: Gene family expansion, gymnosperms, isoprenoid metabolism, MYB transcription factors, microarray RNA profiling, *Picea glauca*, plant evolution, stress response, terpenes, tissue-specific expression.

Introduction

Since their divergence ~300 million years ago (Mya) (Magallón and Sanderson, 2005), gymnosperm and angiosperm plants have diversified their strategies to cope with changing environmental conditions, competing plants, potential pests, and foraging animals (Brooker, 2006; Agrawal, 2007). As a result of their early divergence and adaptations to harsh environments, gymnosperms have acquired unique traits that distinguish them from herbaceous

and woody angiosperms. For example, defence mechanisms have been described in conifer trees such as pines and spruces which can be constitutive or induced in response to biotic and abiotic stresses and lead to the formation of physical barriers (i.e. formation of traumatic resin ducts and calcium oxalate crystals, and cell wall lignification) and/or to the synthesis of phenolic compounds, or volatile and non-volatile terpenoid compounds (for a review, see Keeling and Bohlmann, 2006). Several genes and proteins related to these responses have been identified, but gene regulators that enable plants to cope with environmental challenge remain largely uncharacterized outside of angiosperm plant systems.

Regulation of gene expression plays a fundamental role in plant response to environmental stimuli. Transcription factors (TFs) belonging to the ERF, bZIP, and WRKY families have been linked to a suite of mechanisms leading to defence and stress responses (Singh *et al.*, 2002; Fujita *et al.*, 2006). A few members of the R2R3-MYB family have also been implicated in plant stress responses to cold (Agarwal *et al.*, 2006), UV (Jin *et al.*, 2000), and wounding (Taki *et al.*, 2005), but R2R3-MYBs have primarily been shown to regulate plant secondary metabolism (Vom Endt *et al.*, 2002). Furthermore, they are well known as positive and negative regulators of biosynthetic enzymes required for the production of phenylpropanoids (Legay *et al.*, 2007; Bomal *et al.*, 2008), flavonoids (Grotewold, 2005), and benzenoids (Verdonk *et al.*, 2005).

Sequence analyses of available plant genomes have shown that R2R3-MYBs form one of the largest and most diverse families of TFs in plants. For instance, between 109 and 192 R2R3-MYBs were identified in *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, and *Vitis vinifera* genomes (Jia *et al.*, 2004; Yanhui *et al.*, 2006; Matus *et al.*, 2008; Wilkins *et al.*, 2009). Despite large-scale gene discovery in conifer trees (e.g. Kirst *et al.*, 2003; Pavy *et al.*, 2005; Ralph *et al.*, 2007), relatively few *R2R3-MYB* genes have been described in conifers or gymnosperms. The full-length cDNA sequence for 18 different putative *R2R3-MYB* genes, previously reported in *Pinus taeda* (loblolly pine) and *Picea glauca* (white spruce), were shown to have diverse transcript profiles (Bedon *et al.*, 2007). Analyses of this relatively limited set of sequences suggested that several *R2R3-MYB* gene duplications post-date the angiosperm–gymnosperm split. DNA-binding domains (DBDs) were highly conserved between gymnosperm and angiosperm R2R3-MYBs, and many of the C-terminal motifs described in angiosperms (Kranz *et al.*, 1998) could be found in conifers. However, conifer-specific C-terminal amino acid motifs also suggested possible functional divergence between closely related MYBs from the major plant phylla. A more extensive characterization of the conifer *R2R3-MYB* gene family and the associated putative functional domains may aid in elucidating potential functional conservation and divergences.

Plant R2R3-MYBs are classified into different subgroups based on DNA binding affinity (Romero *et al.*, 1998) and C-terminal amino acid motifs (Kranz *et al.*, 1998; Matus *et al.*, 2008). Some C-terminal motifs are required for transcriptional activation (Li *et al.*, 2006), repression (Jin *et al.*, 2000), or protein interactions (Zimmermann *et al.*, 2004). For example, the R2R3-MYBs of subgroup 4 (Sg4) (Kranz *et al.*, 1998) or C2 repressor motif clade (Matus *et al.*, 2008) harbour the core $^{D}/_{E}LNL^{D}/_{N}L$ from the ERF-associated amphiphilic repression (EAR) motif proposed to play key roles in defence and stress responses (Ohta *et al.*, 2001; Weigel *et al.*, 2005). For example, the Sg4 gene *AtMYB4* was shown to regulate UV stress and wound responses in *Arabidopsis* (Jin *et al.*, 2000; Taki *et al.*, 2005). It was shown that several conifer sequences including the *P. glauca PgMYB5*, *10*, and *13*, and the *P. taeda PtMYB14* gene products share the core EAR motif (Bedon *et al.*, 2007). This led to the hypothesis that these conifer MYBs may play a role in defence, although such a role had yet to be supported by functional evidence.

Gain-of-function experiments using constitutive expression constructs in tobacco, *Arabidopsis*, and spruce have recently shed light on the functions of conifer MYBs. The *P. taeda* genes *PtMYB1* (Patzlaff *et al.*, 2003*a*; Bomal *et al.*, 2008), *PtMYB4* (Patzlaff *et al.*, 2003*b*), and *PtMYB8* (Bomal *et al.*, 2008) have been linked to phenylpropanoid metabolism, and *PtMYB1* and *PtMYB4* to ammonium assimilation (Gomez-Maldonado *et al.*, 2004). However, non-constitutive expression may be more informative considering unforseen effects that may arise when investigating closely related sequences. For example, tissue-specific expression has been shown to be an efficient approach to investigate the function of TFs (Zhang, 2003).

The present report describes an experimental approach driven by the discovery and analysis of a relatively large number of closely related Sg4 *R2R3-MYB* sequences in *P. glauca* and *P. taeda*. Transcript accumulation following mechanical wounding, exposure to cold, and jasmonic acid (JA) application was consistent with a role for several of the sequences in responses to environmental stimuli or stresses. The *P. taeda PtMYB14* was selected as a candidate gene for overexpression in *P. glauca* (white spruce) using the tissue-preferential *cinnamyl alcohol dehydrogenase* promoter (*CADpro*; Bedon *et al.*, 2009) and the constitutive *ubiquitin* promoter (*UBIpro*; Christensen *et al.*, 1992). Comparative analyses of histological, metabolite, and transcriptional phenotypes resulting from these constructs were discussed in relation to the putative involvement of *MYB14* in isoprenoid and flavonoid metabolism in conifers.

Materials and methods

DNA cloning

Conifer Sg4 *R2R3-MYB* sequences were isolated by using 3′-rapid amplification of cDNA ends (RACE; SMART RACE cDNA Amplification Kit, BD Biosciences Clontech, CA, USA) from *P. glauca* (Pg) (needles, mechanically wounded seedling stems) and *P. taeda* (Pt) (seedlings, non-wounded stems and bark). Total RNAs were isolated as in Chang *et al.* (1993). An Sg4-specific and partially degenerate primer (5′-TGGCGYTCSCTTCC-CAAGGCC-3′ where Y=C, T, or U, and S=C or G) was designed based on the WRSLPKAA amino acid sequence from *PtMYB14*,

PgMYB5, *PgMYB10*, and *PgMYB13* (Bedon et al., 2007), *PtMYB1* and *PtMYB4* (Patzlaff et al., 2003a and b), and *PmMBF1* (Xue et al., 2003). Touchdown PCR for the nested 3′-RACE reaction used a DNA engine PTC-225 thermal cycler (Biorad, Hercules, CA, USA) with five cycles of two steps at 94 °C for 30 s and 72 °C for 3 min, five cycles of three steps at 94 °C for 30 s, 70 °C for 30 s, and 72 °C for 3 min, and 25 cycles of three steps at 94 °C for 30 s, 68 °C for 30 s, and 72 °C for 3 min. Amplification products were separated on a 1% agarose gel, gel extracted (Gel Extraction Kit, Qiagen, Mississauga, CA, USA), ligated to pCR2.1, and transformed into TOPOF′ competent cells (TA cloning Kit, Invitrogen, Carlsbad, CA, USA) prior to sequencing.

Sequence comparisons and phylogenetic tree construction

Phylogenetic tree construction used 45 different R2R3-MYB cDNA sequences from *P. taeda*, *P. glauca*, and several angiosperms (Kranz *et al.*, 1998; Karpinska *et al.*, 2004; Bedon *et al.*, 2007; Matus *et al.*, 2008; Wilkins *et al.*, 2009), and two R1R2R3-MYBs (AtMYB3R4 and PttMYB3R) as an outgroup. Several alignment and phylogenetic construction methods were tested (see methods and results in Supplementary Table S1 available at *JXB* online). Clustal W sequence alignments (Thompson *et al.*, 1994) were obtained with the MEGA 4.0.2 software (Tamura *et al.*, 2007), with parameters set as follows: gap opening penalty/gap extension penalty of 25/1 for pairwise alignment and 5/1 for multiple alignment by using the blosum protein weight matrix for amino acid sequences; and 15/6 for both pairwise and multiple alignment with IUB (International Union of Biochemistry) DNA weight matrix for nucleotide sequences (transition weight of 0.5). Two Neighbor–Joining trees were constructed by MEGA 4.0.2 software using (i) amino acid sequences with the Jones–Taylor–Thornton substitution model (Gamma parameter of 1.0 and pairwise gap deletion); and (ii) nucleotide sequences with the p-distance model including transitions and transversion (all codon positions selected and complete deletion). Both bootstrap consensus trees were infered from 1000 replicates.

The MEME analysis software (Bailey and Elkan, 1994) was used to identify C-terminal conserved motifs of 4–6 amino acids in 42 Sg4 sequences with the occurrence parameter for a single motif set at zero or one motif. Amino acid sequence similarities (Fig. 1C) were obtained with Clustal W (blosum matrix, gap open, 25; gap extend, 1) in the Bioedit software (Version 6.0.7). Pairwise sequence similarities between *Pt*/*PgMYB14* and *PgMYB15* were obtained with the Smith–Waterman algorithm in the Emboss package (Matrix: EBLOSUM62, gap penalty, 3.0; extend penalty, 0.1).

Plasmid constructs and stable transformation of white spruce

Constructs were generated for constitutive and tissue-preferential expression of the complete coding sequence of the *P. taeda PtMYB14* and of the *uidA β-glucuronidase* (*GUS*) reporter gene. The maize *UBIpro* (constitutive overexpression) (Christensen *et al.*, 1992) and the *P. glauca CADpro* (tissue-preferential overexpression) (Bedon *et al.*, 2009) were cloned into the pMJM expression vector (Bomal *et al.*, 2008). The *CADpro* is a 1163 bp genomic DNA fragment of upstream flanking sequence of the *CAD* gene (Bedon *et al.*, 2009). Stable *Agrobacterium tumefaciens* transformation of the four constructs into *P. glauca* embryogenic tissues (genotype Pg653), selection of the transgenic lines (independent transformation event), and production of transgenic spruce plantlets were as described (Bomal *et al.*, 2008).

Histology, and histochemical and fluorometric GUS assays

Tissue samples (hypocotyls and roots) were fixed and paraffin embedded as described in Bomal *et al.* (2008). Semi-thin paraffin-free sections were stained in Sharman's safranin O/orange G/tannic acid as per Sharman (1943). For X-gluc staining, samples were prepared essentially as described in Hawkins *et al.* (1997). Briefly, samples were pre-treated for 30 min in cold 90% acetone to facilitate substrate penetration and prevent transgene induction. Samples were then rinsed twice in 100 mM potassium phosphate buffer (pH 7.0) and incubated for 4 h in 5-bromo-4-chloro-3-indolyl-P-D-glucuronic acid in the dark at 37 °C until blue colour had developed. All observations used an Axioskop microscope (Zeiss, Jena, Germany) fitted with a digital camera. MUG (4-methylumbelliferyl glucuronide) fluorometric enzymatic assays were as described in Côté and Rutledge (2003) and used from four biological replicates of 25 hypocotyls (with cotyledons) per line.

Determination of starch content, and analysis of anthocyanins

For starch content determination, soluble metabolites were extracted by overnight incubation at –20 °C with methanol:chloroform:water (12:5:3, v/v/v). The sample was centrifuged, the supernatant removed, and the remaining pellet washed twice with fresh methanol:chloroform:water (12:5:3). The dried residual pellet was weighed, and then hydrolysed using 4% sulphuric acid at 121 °C for 4 min. The liberation of glucose, representing starch, was quantified directly by anion exchange high-performance liquid chromatography (HPLC; Dionex, Sunnyvale, CA, USA) on a DX-600 equipped with a Carbopac PA1 column and an electrochemical detector (Coleman *et al.*, 2006).

For anthocyanin analysis, ~20 mg of tissues were suspended in 1.5 ml of methanol:HCl (95:5, v/v), pulverized at high power for 20 s in a cell disrupter (FastPrep), and incubated for 4 h at 40 °C in a constant temperature heating block. Following extraction, the homogenate was centrifuged for 10 min at 13 000 rpm; the supernatant was dried on a Speedvac and the pellet was resuspended in 200 µl of HPLC grade methanol. Detection and identification of anthocyanins was achieved by liquid chromatography–mass spectrometry (LC-MS) on an HP 1100 LC-MSD-Trap XCT plus. The methanolic extracts were separated on an SB C-18 Zorbax Rapid Resolution 4.6×150 mm 3.5 µm column at 40 °C at a flow rate of 1 ml min^{-1}, using a linear gradient from 95% solvent A and 5% solvent B to 75% solvent A and 30% solvent B over 37 min. Solvent A was water with 0.2% formic acid and solvent B was acetonitrile with 0.2% formic acid. Detection was monitored with an HP 1100 photodiode array detector with the reference wavelength set to 520 nm, while mass determination was achieved by electrospray ionization (ESI) in negative ion polarity.

Terpene extraction and analysis

Terpene extractions were based on procedures of Martin *et al.* (2002). All steps were carried out using 2 ml vials with a teflon-coated screw cap (Hewlett-Packard, Palo Alto, CA, USA). Samples (~200 mg) were submerged in 1.5 ml of *tert*-butyl methyl ether containing 150 µg ml^{-1} isobutylbenzene and 200 µg ml^{-1} pimaric acid as internal standards, and extracted for 14 h overnight with constant shaking at room temperature. To purify extracted terpenes from other organic acids, the ethereal extract was transferred to a fresh vial and washed with 0.3 ml of 0.1 M $(NH_4)_2CO_3$ (pH 8.0). Half of the extract used for diterpene resin acid analysis was methylated by adding 50 µl of 0.2 M *N*-trimethylsulphonium hydroxide in methanol (Macherey-Nagel GmbH, Germany) to 0.4 ml of the washed ethereal extract in a separate vial. After 2 h incubation at room temperature, the solvent was evaporated down to 100 µl under nitrogen, and stored at –20 °C. The other half of the extract used for monoterpene and sesquiterpene analysis was filtered through a Pasteur pipette column filled with 0.3 g of silica gel (Sigma, 60 Å) overlaid with 0.2 g of anhydrous $MgSO_4$. The column was washed with 1 ml of diethyl ether, and the combined eluant collected in a fresh vial, evaporated to ~100 µl, and stored at –20 °C.

Gas chromatography (GC)–MS analysis of monoterpenes and sesquiterpenes was carried out with a Hewlett-Packard 6890 GC-MSD system, using a DB-WAX column (0.25 mm×0.25 µm×30 m,

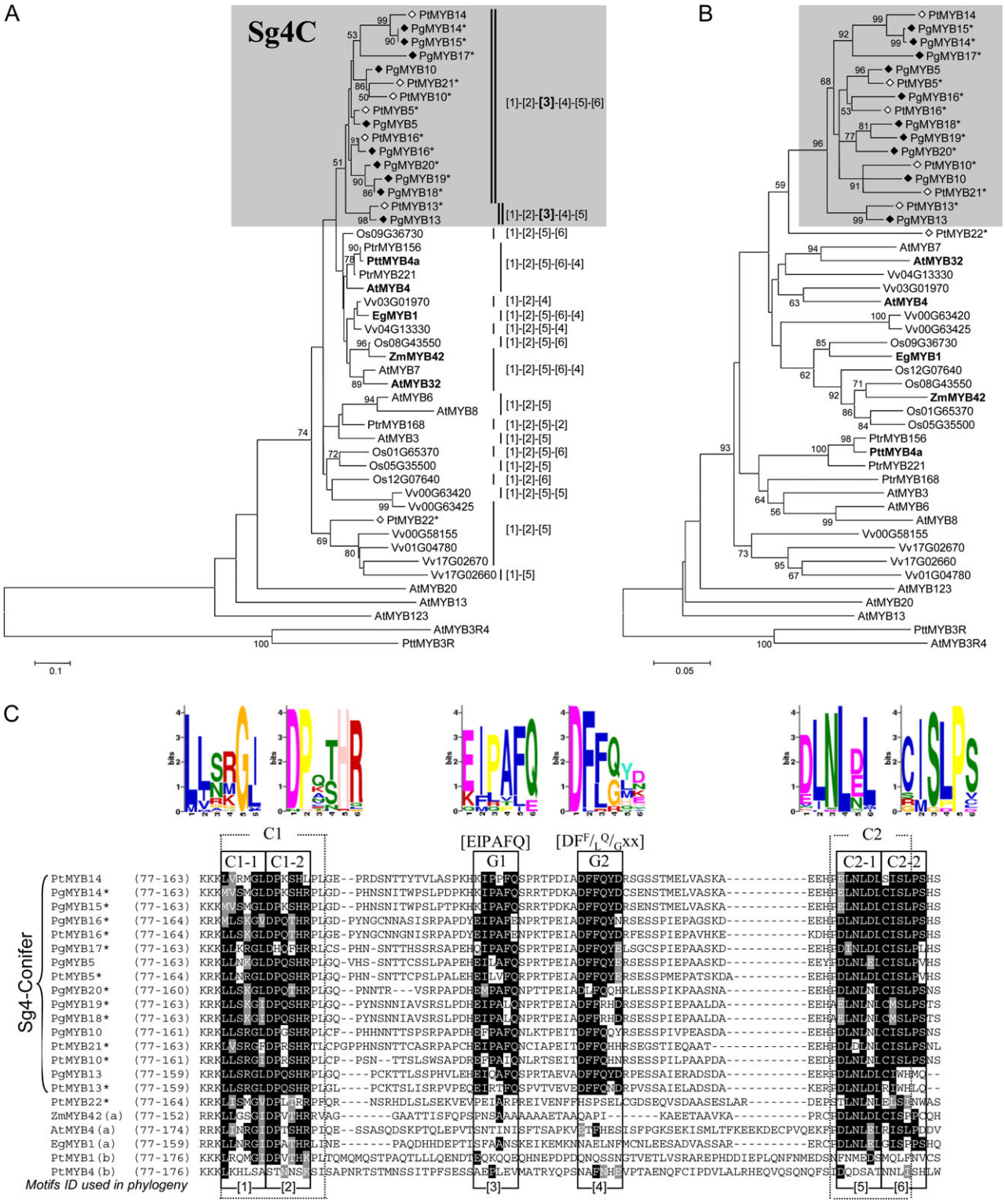

Fig. 1. Sequence analyses of angiosperm and conifer Sg4 R2R3-MYBs define the conifer specific subclade Sg4C (grey boxes) and identify associated amino acid motifs. (A and B) Rooted Neighbor–Joining trees were obtained with MEGA 4 software (Tamura *et al.*, 2007) and Clustal W alignments of (A) the amino acid sequence (WRSLPKAAG in R2 to the predicted stop codon) or (B) the MYB DBD nucleotides (see Materials and methods). Bootstrap values >50% are shown. Other alignment methods and tree construction algorithms gave consistent results and are detailed in Supplementary Table S1 at *JXB* online. **Picea glauca* (filled lozenges) and *Pinus taeda* (open lozenges) sequences reported here; known repressors of phenylpropanoid/lignin pathways are in bold. AtMYB3R4 and PttMYB3R are the outgroup, and AtMYB13, 20, and 123 (not of Sg4) are landmarks. Bars indicate the evolutionary distance as a percentage. (C) Amino acid motifs in the Sg4C C-termini ([1]–[6]) (MEME software). Top: MEME motif logos; bit scores indicate the information content for each position. Bottom: Clustal W alignment of predicted amino acid sequences and conserved motifs (shaded): G1 (EIPAFQ) and G2 (DF$^{F}/_{L}{}^{Q}/_{G}$xx)

J W Scientific, Folsom, CA, USA). Injections (1 µl of ethereal extract) were made with an injector temperature of 220 °C. An initial temperature of 40 °C was held for 3 min and increased to 80 °C at 3 °C min^{-1}, and then increased to 180 °C at 5 °C min^{-1}, held for 5 min, and ramped at 15 °C min^{-1} to 240 °C. The helium flow was a constant 1 ml min^{-1}. For compound identification, the MS detector was operated using a mass range of 40–350 for monoterpenes and sesquiterpenes, and of 40–550 for diterpenes. All samples were measured in the single ion modus, with masses 91, 93, 136, and 161 *m/z* for the detection of monoterpenes and sesquiterpenes, and masses 239, 241, 314, and 316 *m/z* for diterpenes. Spectra were collected under standard conditions (electron impact ionization at 70 eV).

Analysis of diterpenes was performed on the same GC-MS instrument, with an HP-5 column (0.25 mm×0.25 µm×30 m, Hewlett-Packard), and helium as carrier gas with the same flow rate as above. Injections used 1 µl of concentrated, derivatized ethereal extract with an injector temperature of 220 °C. GC-MS split ratios were 1:10 with an injector temperature of 220 °C. The initial instrument temperature of 120 °C was increased to 150 °C at 1 °C min^{-1}, and to 280 °C at 5 °C min^{-1} (6 min hold).

GC-MS-generated peaks were quantified using Hewlett-Packard Chemstation software. Isobutylbenzene was the internal standard for monoterpenes and sesquiterpenes, and methylated pimaric acid was an internal standard to calculate diterpene concentrations. For quantitative analysis of monoterpenes and sesquiterpenes, the MS detector was operated in the single ion mode monitoring ions at 91, 93, and 161 *m/z*, corresponding to the internal standard, monoterpenes, and sesquiterpenes, respectively. For diterpene quantification, those monitored were: 121, 135, 239, 241, and 316 *m/z*.

Identification of terpenes was based on retention time comparison and mass spectra with authentic standards (Aldrich Chemicals) or on mass spectra in the Wiley 275.L or National Institute of Standards and Technology 98.1MS libraries.

RNA extraction, cDNA synthesis, and RT-qPCR analysis

Tissue sample preparation, total RNA extraction, and cDNA synthesis were as decribed in Bomal *et al.* (2008). The real-time quantitative PCRs (RT-qPCRs) were assembled in a LightCycler® 480 Multiwell plate 384 (Roche, Basel, Switzerland) using a pipetting robot (EpMotion 5075, Eppendorf, Hamburg, Germany). The thermal cycling used a LightCycler® 480 (Roche), a 15 min activation period at 95 °C followed by 40 cycles (95 °C for 10 s, 55 °C for 60 s, and 72 °C for 30 s), and fluorescence readings taken at the end of each cycle. Melting curves were used to verify amplicon purity. Crossing point (Cp) values were determined with the LC480 software, and standard curves were used to transform Cp values into numbers of transcript molecules. Standard curves were based on dilution series covering five orders of magnitude (10^{-1}–10^{-6} ng μl^{-1}) prepared for each cDNA linearized with *Eco*RI or *Bam*HI, purified on Qiaquick columns (Qiagen), and verified on a bioAnalyser (model 2100, DNA 1000 LabChip kit, Agilent Technologies). Transcript levels were normalized against the transcript level of *ELONGATION FACTOR 1-α* (*EF1-α*) or *CELL DIVISION CYCLE 2* (*CDC2*).

Most of the DNA sequences used in gene expression experiments came from expressed sequence tag (EST) databases (Supplementary Table S5 at *JXB* online). In addition, cDNA fragments for nine different enzymes of the mevalonate (MVA) and the methylerythritol phosphate (MEP) pathways were isolated from *P. taeda* by PCR and used to produce standard curves for RT-qPCR. Primers were designed with sequences of the TIGR pine gene index (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=pine) with the Primer 3 software (Supplementary Table S4). PCR (50 µl reaction volume) used 0.2 µM of forward and reverse primers, 2 mM $MgSO_4$, 0.2 mM dNTPs at 10 mM each, 1 µl of pine cDNA, and 1 U of Platinum® Taq DNA polymerase High Fidelity (Invitrogen, CAT#11304-011). The cycling conditions were 94 °C for 15 min followed by 35 cycles of 94 °C for 30 s, 60 °C for 30 s, 68 °C for 2 min, and 68 °C for 5 min.

Microarray experiment and data analysis

The microarray experiment was designed in compliance with MIAME guidelines and the data are deposited in ArrayExpress. Tissues were isolated from 4-week-old somatic plantlets from two transgenic embryogenic lines per construct (lines 16 and 18 for *UBIproPtMYB14*, lines 10 and 14 for *CADproPtMYB14*) and an untransformed control line (wild type). Total RNAs were extracted from four biological replicates of 25 plantlets (hypocotyls and cotyledons) per line. Transcript profiling used a 9K spruce cDNA microarray, RNA indirect amplification, microarray hybridization, image analysis, and statistical analyses methods (Bomal *et al.*, 2008).

Eight hybridizations (two replicates per line in each dye) were used for each transgene construct versus wild-type comparison. Differentially expressed transcripts were identified in each comparison based on a *P*-value <0.01 (from the LIMMA), a log2 ratio ⩾|0.5| (1.41-fold change), and a false discovery rate (FDR) of 1% (Benjamini and Hochberg, 1995). A common set of genes was identified based on the overlap between the lists from each construct. The statistical analysis was enhanced by using a permutation test to evaluate the number of genes that would pass, under the null hypothesis, the conditions to be part of the ovelapping set by chance alone. For this, a set of permutated data files was obtained by assigning the original data set at random to one of the two possible treatments (i.e. wild type or transgenic) without replacement (including the correct assignment). Independent data analyses were performed with the 70 permutated data files for each construct type. All the 4900 possible pairwise combinations between the two comparisons were examined in R to identify potential overlaps occurring by chance alone. A maximum of five genes in the overlap of two sets was obtained with a mean number of 0.153 genes over the 4900 pairwise combinations, while the correct assignment in both comparison gave an overlap of 53 sequences.

Wounding, JA treatment, and exposure to cold in pine and spruce in vitro *plantlets*

Pinus taeda seeds were sterilized by soaking in ethanol (95%) for 5 min, sodium hypohlorite (2%) with Tween-80 (0.1%) for 10 min, and H_2O_2 (10%) for 5 min, then ethanol (70%) for 5 min, and rinsed three times for 10 min in sterile distilled water. Seeds were stratified by soaking for 24 h in sterile MilliQ water at room temperature in the dark and then scarified by cutting the radicular tegument, set to germinate on half-strength MS medium supplemented with 0.5% activated charcoal, and incubated at 25 °C under continuous light (40 µmol m^{-2} s^{-1}). *Picea glauca* plantlets were obtained from mature somatic embryos (Pg653) and germinated as described (Bomal *et al.*, 2008).

Treatments were applied to 4-week-old plantlets as follows. Mechanical wounding was achieved by pinching the hypocotyl with forceps and by cutting one-third of the cotyledons and needles with scissors. Wounded tissues were sampled for analysis at 0, 1.5, 6, and 24 h after wounding. JA (Sigma-Aldrich, Buchs, Switzerland) was dissolved in methanol and 100 µM solution was

are conserved in conifers; the C1 LIsrGIDPx$^{T}/_{S}$HRx$^{I}/_{L}$ (Kranz *et al.*, 1998) and the C2 motifs pdLNL$^{D}/_{E}$LxI$^{G}/_{S}$ (Kranz *et al.*, 1998, Jin *et al.*, 2000) are conserved in all taxa (see A). Numbers in brackets are the amino acids positions (predicted from the MYB14 full-length sequence). (a) Angiosperms Sg4 sequences. (b) Other subgroup conifer sequences.

sprayed onto the arial portion of unwounded plantlets; controls were sprayed with methanol alone. JA-treated hypocotyl and cotyledons were harvested for analysis at 0 h and 24 h after treatment. Exposure to cold consisted of incubating 4-week-old plantlets for 24 h in a growth cabinet at 4 °C, while the control plants were maintained at 23 °C. Three biological replicates for each of the above treatments were used in a randomized complete block design, with each replicate being comprised of three plantlets (pine seedlings in magenta boxes) or 10 plantlets (spruce somatic seedlings in Petri dishes).

Accession numbers

The EMBL/GenBank accession numbers are as follows. Ptine *MYB* genes: FJ469924 (*PtMYB5*), FJ469925 (*PtMYB10*), FJ469926 (*PtMYB13*), FJ469927 (*PtMYB16*), FJ469928 (*PtMYB21*), and FJ469929 (*PtMYB22*). Pg *MYB* genes: FJ469917 (*PgMYB14*), FJ469918 (*PgMYB15*), FJ469922 (*PgMYB16*), FJ469923 (*PgMYB17*), FJ469919 (*PgMYB18*), FJ469920 (*PgMYB19*), and FJ469921 (*PgMYB*20). The Pg *CADpro* sequence is FJ428229 (Bedon *et al.*, 2009). Other conifer sequences used in this report are: *PtMYB1* (AY356372), *PtMYB4* (AY356371, *PtMYB14* (DQ399056), *PgMYB5* (EF601068), *PgMYB10* (EF601073), *PgMYB13* (EF601076); *Oryza sativa* Os09G36730, Os08G43550, Os12G07640, Os01G65370, and Os05G35500; *Vitis vinifera* Vv04G13330 (GSVIVT00036753001/Vv4g15106462), Vv00G63420 (GSVIVT00038661001), Vv00G63425 (GSVIVT00038662001), Vv00G58155 (GSVIVT00022367001), Vv17G02670 (GSVIVT00017797001), Vv17G02660 (GSVIVT00017798001), Vv01G04780 (GSVIVT00030469001), Vv03G01970 (GSVIVT00036110001/Vv3g1569838/VvMYB4); *Populus trichocarpa PtrMYB156* (eugene3.00090345), *PtrMYB221* (estExt_fgenesh4_pg.C_LG_IV1453), *PtrMYB168* (eugene3.00190735); *Populus tremuloides*×*Populus trycocarpa PttMYB4a* (AJ567346), *PttMYB3R* (AJ567344); *Arabidopsis thaliana*: *AtMYB3R4* (NM_121189), *AtMYB4* (AT4G38620), *AtMYB7* (AT2G16720), *AtMYB3* (AT1G22640), *AtMYB6* (AT4G09460), *AtMYB8* (At1g35515), *AtMYB32* (At4g34990), *AtMYB123* (At5g35550), *AtMYB20* (At1g66230), *AtMYB13* (At1g06180); *Eucalyptus gunnii EgMYB1* (AJ576024); and *Zea mays*: *ZmMYB42* (AM156908).

Results

PtMYB14 *and related sequences define a large conifer subclade within Sg4 of R2R3-MYBs*

The isolation of several new cDNA sequences closely related to *PtMYB14* (Bedon *et al.*, 2007) brings the number of Sg4 gene sequences to 10 in *P. glauca* and seven in *P. taeda* (Fig. 1). Six distinct partial cDNAs were obtained for each of these species (see Materials and methods), with predicted coding sequences ranging from 159 to 230 amino acids, from the WRSLPKAA motif in the DBD to the putative stop codon. The *PgMYB17* full-length cDNA sequence of 236 predicted amino acids was identified by EST mining (Pavy *et al.*, 2005) by full-length sequencing. The new sequences are numbered in continuation with a previous report (Bedon *et al.*, 2007).

Phylogenetic trees constructed with sequences from *Picea*, *Pinus*, and several angiosperms (*Arabidopsis*, *Oryza*, *Populus*, and *Vitis*) group the 17 conifer sequences with the Sg4 *MYB* genes (Fig. 1A, B). They show that conifers have many Sg4 sequences compared with most angiosperms. *Picea* has 10 distinct sequences compared with 3–8 in the *Arabidopsis*, *Oryza*, *Populus*, and *Vitis* genomes (Matus *et al.*, 2008; Wilkins *et al.*, 2009). Nearly all of the Sg4 conifer sequences formed a monophyletic subclade, refered to as Sg4C (Fig. 1A, B). Different alignment methods and tree construction algorithms were compared to validate the robustness of the Sg4C subclade (see Supplementary Table S1 at *JXB* online). The DBD sequences alone gave well conserved tree topologies, with the Sg4C subclade supported by high bootstrap values (without including *PtMYB22*) or moderate bootstrap values (with *PtMYB22*) (Fig. 1B; Supplementary Table S1 at *JXB* online). Analyses encompassing the DBD and the C-terminus clearly excluded *PtMYB22* from Sg4C and suggested C-terminal conservation among Sg4C sequences.

The degree of C-terminal conservation was shown by amino acid motif analysis (MEME software). Despite the highly variable C-terminus in MYBs, six conserved amino acids motifs were identified ([1]–[6]), two of which were uniquely conserved in sequence and location in Sg4C ([3] and [4], Fig. 1A, C). The Sg4 sequences from all taxa harboured the C1 motif of unknown function and the C2 repressor domain containing the core from the EAR motif (motif C2-1, Fig. 1A, C) that is characteristic of Sg4 (Kranz *et al.*, 1998) also known as the C2 repressor motif clade (Matus *et al.*, 2008). The G1 and G2 motifs located between C1 and C2 appear to be a signature of the Sg4C MYBs (Fig. 1A). The G1 motif was unique to Sg4C members, while the G2 motif occurs in some Sg4 angiosperms but at a different location.

Sg4C transcript levels are modulated in response to stresses

The accumulation of Sg4C transcript was monitored with gene-specific RT-qPCR plantlets subjected to mechanical wounding, JA application, and exposure to cold (Fig. 2). It was hypothesized that Sg4C *MYB* expression may be modulated by stresses because several sequences containing the EAR repressor motif have been implicated in responses to stresses, including drought, cold, salt, and oxidative stress (Ohta *et al.*, 2001; Weigel *et al.*, 2005; Kazan, 2006, and references therein). In *Arabidopsis*, the Sg4 gene *AtMYB4* negatively regulated the formation of UV protectant sinapate esters and its expression was modulated by UV-B (Jin *et al.*, 2000).

A time course evaluation of the response to mechanical wounding indicated differential accumulation among Sg4C transcripts (Fig. 2A). The transcripts of *PtMYB14* in pine, and *PgMYB14* and *PgMYB15* in spruce accumulated strongly and rapidly (after 90 min), and decreased after a few hours. The spruce *PgMYB16* and *PgMYB18* also accumulated transiently, but after 6 h and less strongly than the *PgMYB14* and *PgMYB15* transcripts. Many of the other Sg4C transcripts were transiently down-regulated after wounding (after 90 min) and some were unaffected. In JA-treated plantlets, *PtMYB14* and *PtMYB13* transcripts increased 14-fold and 2-fold, respectively, while *PgMYB14*

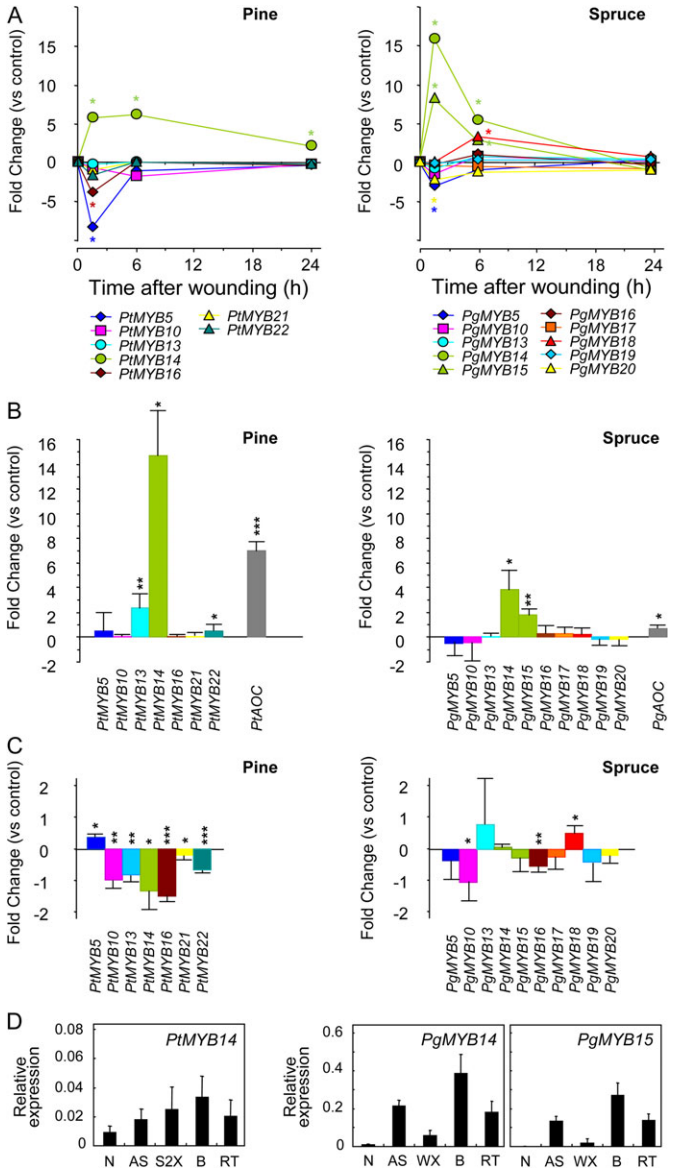

Fig. 2. Stress-responsive transcript profiles of Sg4C sequences in *P. taeda* and *P. glauca* determined by RT-qPCR. (A–C) Transcript accumulation (fold change relative to controls) in 4-week-old plantlets in response to (A) mechanical wounding (after 0, 1.5, 6, and 24 h), (B) jasmonic acid (after 24 h) and (C) exposure to cold (24 h at 4 °C); *AOC*, *allene oxide cyclase*. (D) Transcript accumulation of *PtMYB14*, *PgMYB14*, and *PgMYB15* in 2-year-old trees. N, needle; AS, apical stem (young elongating shoot); S2X, stem differentiating secondary xylem; WX: whole stem xylem; B, bark; RT, root tip. Data were from three biological replicates and were normalized relative to transcript levels of *EF1-α* (A–C) or both *EF1-α* and *cdc2* (D). Significant Student's test at 0.05 (*), 0.01 (**), or 0.001 (***).

and *PgMYB15* transcripts increased 4-fold and 2-fold (Fig. 2B). Transcripts of *allene oxide cyclase* (*AOC*), a reporter gene for the JA pathway (Wasternack, 2007), accumulated in JA-treated plantlets. Following an exposure to cold, the level of many Sg4C transcripts decreased slightly, and only *PgMYB18* transcripts increased marginally (Fig. 2C).

The most strongly induced sequences, *PtMYB14* and the closest spruce homologues *PgMYB14* and *PgMYB15*, were analysed in more detail. The predicted amino acid sequence of PtMYB14 shares 82.4% and 81.1% identity with PgMYB14 and PgMYB15, respectively. The spruce MYB14 and MYB15 sequences share 90.2% amino acid identity and 95% nucleotide identity. In 2-year-old untreated trees (Fig. 2D), *PtMYB14* RNA transcripts accumulated preferentially in the bark and were detected at low levels in all tissues tested. The transcripts *PgMYB14* and *PgMYB15* accumulated preferentially in the bark, as well as in the root tip and the apical portion of the stem. Based on these data, *PtMYB14* was selected for functional analysis.

Constitutive and tissue-preferential overexpression of PtMYB14 *alters development*

Two different promoters were employed to express PtMYB14 in stably transformed *P. glauca*, as a heterologous conifer expression system. The *P. glauca CADpro* was used for tissue-preferential expression (Bedon *et al.*, 2009) and the maize *UBIpro* drove constitutive expression (Christensen *et al.*, 1992). The *CADpro* was shown to give xylem and bark preferential expression in young trees, and *CAD* transcripts also accumulate in response to mechanical wounding in these tissues (Bedon *et al.*, 2009). The *CAD* transcript profile overlaps with those of *PtMYB14* and *PgMYB14*, but *CAD* transcripts accumulate at higher levels.

The cellular and tissue activity of each promoter was shown in stably transformed *P. glauca* plantlets harbouring promoter–GUS constructs (Fig. 3). Histochemical GUS staining indicated strong enzyme activity in all cells with the *UBI* promoter (Fig. 3A) and acitivity localized to differentiating tracheids and xylem ray cells with the *CAD* promoter (Fig. 3B). Quantitative enzyme assays comparing the two constructs were consistent with the large differences in transgene transcript accumulation, but showed that each construct gave GUS activity levels that varied little between the different organs (Fig. 3C).

Constitutive overexpression of *PtMYB14* (*UBIproPtMYB14*-OE) produced a strong phenotype affecting plantlet morphology (Fig. 3D, E). A few weeks after germination, the hypocotyls and cotyledons were hypertrophic and red pigmentation had accumulated compared with controls, but root development was normal (Fig. 3E). Approximately 10 weeks after germination, plantlets withered and were unsuitable for transfer to soil. Histological observation of *UBIproPtMYB14*-OE hypocotyls showed that the vasculature was disorganized and had relatively few tracheids with more variable diameters (Fig. 3H, asterisk), and the parenchyma was expanded and contained inner cells accumulating numerous starch grains (Fig. 3H, black arrows) and outer cells accumulating phenolic compounds (Fig. 3K, white arrows). Polyphenolic parenchyma cells (white arrowheads), seen in the *UBIproPtMYB14*-OE plantlets (Fig. 3H, K) but not in controls (Figure 3G, J),

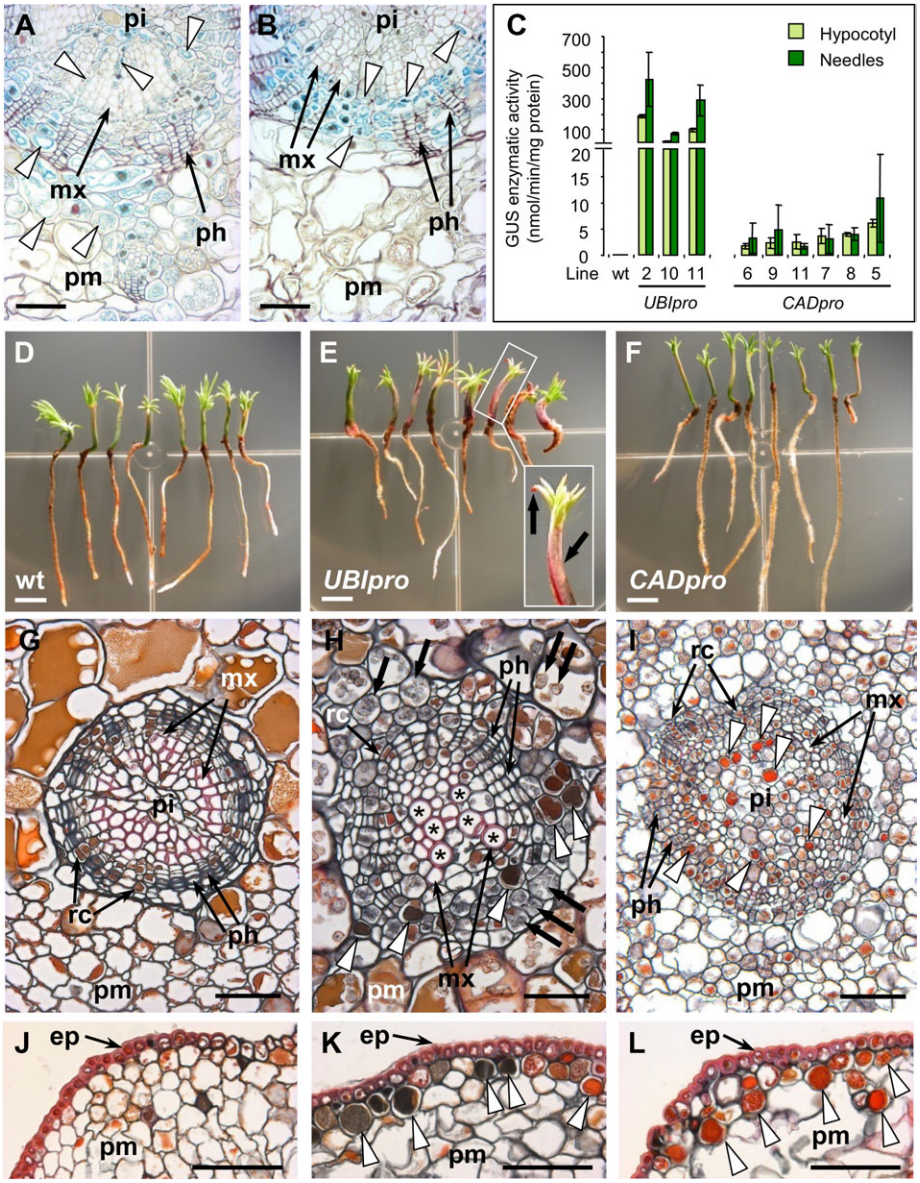

Fig. 3. Promoter analysis and transgenic phenotypes with constitutive and tissue-preferential overexpression of *PtMYB14*. (A and B) Analysis of *UBIproGUS* and *CADproGUS* in 10 μm cross-sections from paraffin-embedded hypocotyls. (A) GUS staining (white arrow) in parenchyma, differentiating xylem, and ray cells with the *UBIproGUS* construct and (B) in differentiating xylem and ray cells with the *CADproGUS* construct. (C) Enzymatic MUG assays: means (±SE) for three biological replicates from the wild type (wt) and independent transgenic lines. (D–F) Morphology of 6-week-old *in vitro* plantlets. (D) Wild-type; (E) *UBIpro PtMYB14*-OE with hypertrophic hypocotyl and cotyledons with red pigmentation (black arrows in inset); (F) *CADproPtMYB14*-OE. (G–L) Cross-sections of 5 μm in paraffin-embedded hypocotyls of wild-type (G, J), *UBIproPtMYB14*-OE (H, K), and *CADproPtMYB14*-OE (I, L) hypocotyls showing polyphenolic parenchyma cells (white arrowhead) in both transgenics within the vasculature (H, I) and in outer subepidermal layers (K, L). Disorganized vasculature with large tracheids (*); starch grain accumulation in inner parenchyma cells (black arrows) in *UBIproPtMYB14* transgenics (H). Bars: 50 μm (A, B), 5 mm (D–F), 100 μm (G–L). pm, parenchyma; ph, phloem; mx, metaxylem; pi, pith; rc, ray cell; ep, epidermis.

are part of constitutive and inducible defences in conifers (Franceschi *et al.*, 2005).

The *CADproPtMYB14*-OE plantlets were morphologically more similar to the controls (Fig. 3F); they developed nearly normal hypocotyls, although a slight increase in pigmentation was noted. Histological observations clearly revealed altered cellular organization and development (Fig. 3G): the inner parenchyma cells accumulated starch grains (Fig. 3I) but fewer than in the *UBIproPtMYB14*-OE plantlets (Fig. 3H). Within the vasculature, the *CADproPtMYB14*-OE tracheids were narrower and had thinner cell walls than the wild type; and polyphenolic parenchyma cells were observed within the vasculature, the pith, and the outer parenchyma (Fig. 3I, L).

The two constructs gave contrasting *PtMYB14* transgene transcript levels and differentially impacted upon transcript accumulation of other Sg4C sequences (Fig. 4). Relative to the endogenous *PgMYB14*, the *UBIpro* construct resulted

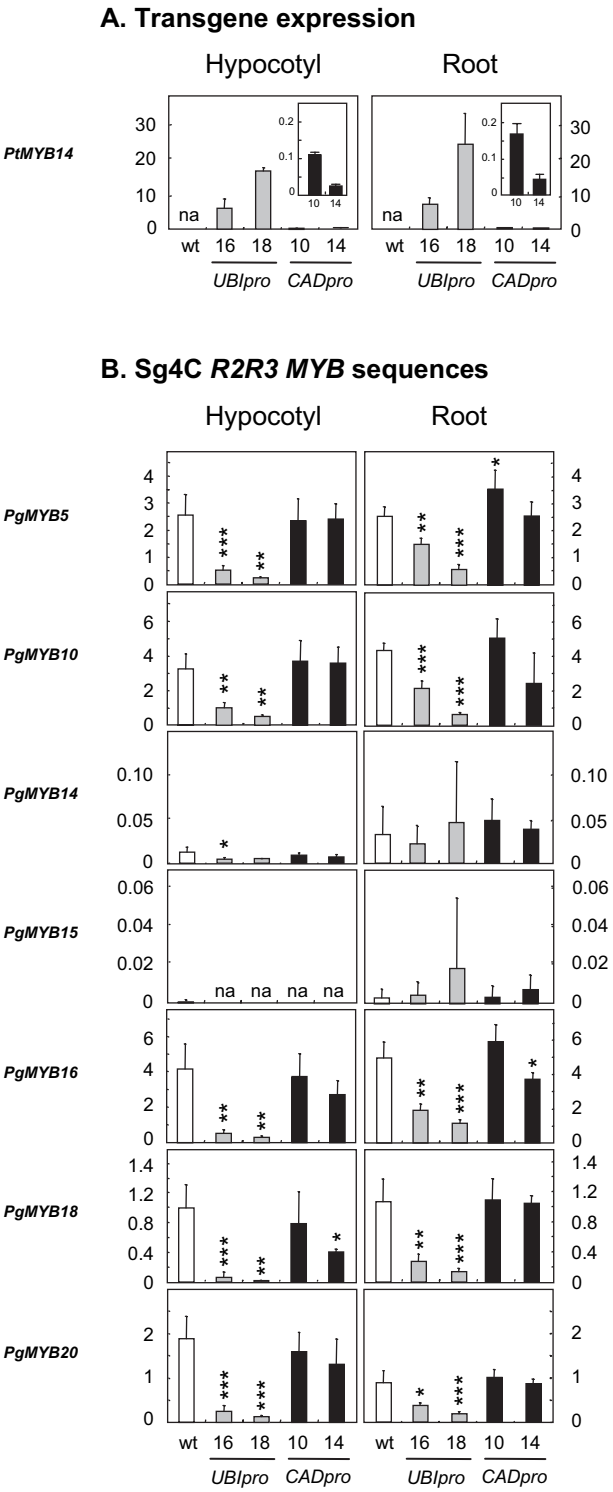

Fig. 4. Transcript accumulation of the *PtMYB14* transgene and other Sg4C sequences in hypocotyls and roots of 4-week-old wild-type and transgenic plantlets. Mean (±SE) of RT-qPCR determinations in control (wt), *UBIproPtMYB14*-OE (lines 16 and 18), and *CADproPtMYB14*-OE (lines 10 and 14) plantlets of four biological replicates, normalized relative to *EF1-α* transcripts. na, not amplified. Significant Student's test at 0.05 (*), 0.01 (**), or 0.001 (***).

in a 300-fold increase in expression and the *CADpro* construct gave a 3-fold increase (Fig. 4A). The other Sg4C sequences tested were all down-regulated with the *UBIpro* construct (4- to 14-fold), but not significantly with the *CADpro* construct (Fig. 4B).

Constitutive and tissue-preferential overexpression of PtMYB14 *has overlapping impacts on the spruce transcriptome*

As described above, partly overlapping phenotypes affecting hypocotyl development were observed with constitutive and tissue-preferential overexpression of *PtMYB14*. Microarray transcript profiles (9K custom cDNA microarray) were obtained for *UBIproPtMYB14* and *CADproPtMYB14* plantlets and compared in order to assess the impact of both types of overexpression upon gene expression relative to the controls.

The constructs resulted in very different numbers of gene sequences being statistically different from the controls (Supplementary Fig. S1A at *JXB* online). A total of 517 and 78 of these sequences had a fold change of ⩾1.4, for *UBIpro* and *CADpro*, respectively. The proportion of up- and down-regulated sequences also varied with the construct, i.e. 59–41% for the *UBIpro* lines and 94–6% for the *CADpro* lines. Based on GO and KEGG annotations, 8.9% and 13.5% of these misregulated sequences were assigned to the 'biotic/abiotic stress' category for *UBIpro* and *CADpro* constructs, respectively (Supplementary Fig. S1B). BlastX results against Uniref100, AGI10.0, PGI5.0, and SGI1.0 indicated that several other sequences appear to be related to stress and defence responses, thus increasing the latter proportions to 34% and 41% for the *UBIpro* and *CADpro* construct, respectively (Supplementary Table S2 at *JXB* online).

Despite these apparent differences, there was a significant overlap among the differentially expressed genes between the *UBIproPtMYB14* and *CADproPtMYB14* lines which was comprised of a total of 53 common sequences (Table 1). The probability of such an overlap occurring by chance was estimated by permutation analysis to be $<2.04\times10^{-4}$. These include 49 sequences with increased transcript abundance compared with the control, and only three genes that were down-regulated, namely a heat shock protein (BT101895), an expansin, and an extensin-like protein. Approximately two-thirds of the genes misregulated with *CADproPtMYB14* were also misregulated in a similar manner with the *UBIproPtMYB14* construct, although the fold change ratios were generally higher in the *UBIpro*-overexpressers (Table 1). The genes in common and their putative roles in metabolism or in stress and defence response are described below.

RT-qPCR analyses were performed to validate the results of the microarray analyses. A set of 24 genes related to cell wall biogenesis, defence response, as well as primary and secondary metabolism was tested and confirmed for both types of transgenics plantlets (Table 1, and Supplementary Table S2 at *JXB* online).

Table 1. Overlapping set of differentially expressed genes in *UBIproPtMYB14* and *CADproPtMYB14* spruce plantlets[a]

Sequence[b] (cDNA model)	GenBank accession no.[b]	Annotation[c]				*UBIproPtMYB14*		*CADproPtMYB14*	
		Predicted function	Accession no.	E-value	AP[d]	FC[e]	*P*-value[f]	FC	*P*-value
GO:0006629 Lipid metabolic process; GO:0016042 Lipid catabolic process									
GQ03221_G05	BT110413	Acyl-CoA thioesterase	Q8GYW7	3E-17	JA	1.5	1.2E-03	1.5	9.7E-03
GQ04104_D12	BT119330	Acyl-CoA synthetase	Q946Z2	5E-46	JA	1.8	6.9E-05	1.5	9.0E-03
GQ02802_D14	BT103793	Acyl-CoA synthetase (4CL-like)	Q9M0X9	5E-49	JA	2.0	1.9E-07	1.7	1.8E-04
GQ0195_E17	BT102487	Lipoxygenase 1 (9-lipoxygenase)	Q76I22	6E-77	JA	2.5	1.1E-03	3.1	1.2E-03
GQ03307_D15	BT111900	Phytoalexin-deficient 4-1 protein (PAD4)	Q2TNK3	1E-13	SA	1.9	3.6E-07	1.5	1.0E-03
GQ04007_D22	BT118519	EDS1, lipase class 3	Q65GA1	0E+00	SA	2.3	4.8E-07	2.3	1.4E-05
GQ01307_L15	CK441554.2*	EDS1, lipase class 3	Q8LL12	3E-20	SA	1.5	2.5E-06	1.3	2.8E-03
GQ03708_J16	BT115907	EDS1-like, lipase, active site	Q1S9M3	7E-20	SA	2.2	4.1E-08	1.4	6.2E-03
GO:000829 Isoprenoid biosynthetic process; GO:0008202 Steroid metabolic process; GO:0009813 Flavonoid biosynthetic process									
GQ03709_L23	BT115978	Acetoacetyl-CoA thiolase	Q2L8A7	3E-46	MVA	1.5	1.6E-05	1.6	5.5E-05
GQ03013_M13	BT103082	3-Hydroxy-3-methylglutaryl-CoA-synthase	P93773	0E+00	MVA	1.4	2.6E-03	1.6	3.4E-04
GQ03804_B11	BT116659	Mevalonate disphosphate decarboxylase	Q5UCT8	3E-75	MVA	2.0	2.8E-07	1.6	3.5E-04
GQ04108_J07	BT119670	E,E-α-farnesene synthase	Q675K8	2E-13	TP	2.6	2.7E-05	2.6	3.4E-04
GQ03508_H04	BT103172	Zeatin *O*-glucosyltransferase 1	Q7Y232	8E-35		2.1	2.2E-05	1.9	1.0E-03
WS03219_O18	DR550970.1	Thiohydroximate *S*-glucosyltransferase	Q4EVY6	1E-28		1.3	1.8E-03	1.5	3.8E-04
GQ0206_A18	BT103070	Chalcone synthase	Q2ENB1	0E+00	FV	1.4	2.4E-03	1.5	6.6E-03
GQ04013_C16	BT109912	Cinnamoyl-CoA reductase-like	Q9M0B3	2E-29		1.4	2.9E-03	1.4	9.0E-03
GO:0008152 Metabolic process									
GQ0205_O01	BT103059	NAD-dependent malate oxidoreductase	P37221	5E-101	TCA	1.3	5.8E-03	1.6	5.5E-05
GQ03005_P01	BT106595	ATP:citrate lyase b-subunit	Q93YH3	4E-121	TCA	1.4	3.8E-03	1.4	9.2E-03
GQ03719_C02	BT116465	ATP:citrate lyase b-subunit	Q9AXR6	3E-78	TCA	1.6	1.1E-03	1.7	2.0E-03
GQ03230_I01	BT102565	Cystathionine γ-synthase	P55217	9E-69		1.3	5.8E-04	1.4	9.7E-04
GQ02802_I24	BT103811	Fructokinase, putative	Q9C524	4E-41		1.5	1.1E-05	1.4	1.0E-03
GO:0006810 Transport									
GQ03810_L15	BT101863	ABC transporter, PRD-like	Q8GU87	1E-35	FV	2.2	1.2E-05	1.9	1.0E-03
GQ03810_L15	BT101863	ABC transporter, PDR-like	Q0IRX8	1E-30	FV	2.3	3.4E-06	1.8	1.7E-03
GQ03238_I01	BT111495	Amino acid permease	Q9FKS8	1E-16		2.2	2.5E-07	1.6	8.0E-04
GQ04005_E15	DV983142.2*	Sulphate transporter (SULTR1;3)	Q9AT12	5E-58		1.7	3.7E-05	1.5	5.4E-03
GO:0006118 Electron transport									
GQ03709_B14	BT115936	UDP-glucose dehydrogenase	Q6RK07	0E+00		1.5	1.2E-05	1.3	5.1E-03
GQ03510_P10	BT114233	FAD-binding domain-containing protein	Q10RL8	1E-50	FV	1.6	6.7E-06	1.3	7.7E-03
GQ03413_I24	BT113565	Cytochrome P450 (CYP720B1)	Q50EK4	7E-31	TP	1.9	1.7E-04	2.2	2.6E-04
GQ03508_P12	BT114142	Flavonoid 3′,5′-hydroxylase (CYP76C7)	O04773	7E-40	FV	1.9	2.5E-06	2.0	2.3E-05
GO:0009664 Cell wall organization and biogenesis									
GQ02808_E03	BT100811	Expansin	Q84UT0	1E-69		–2.9	3.4E-06	–1.8	9.0E-03
GQ0024_A07	BT100525	Extensin-like protein	Q9SPM0	2E-50		–2.1	4.9E-10	–1.3	9.1E-04
GQ03902_B01	BT117517	Chitinase class II	Q6E6M9	3E-94	JA	3.9	1.4E-08	2.2	2.2E-04
GQ03803_J20	BT116631	Chitinase class IV (Chia4-Pa1)	Q6WSR9	2E-66		1.9	1.2E-07	1.4	2.0E-03
GQ03908_A22	BT117700	MAP kinase 6	Q39026	1E-109	SA	1.5	1.4E-04	1.8	4.6E-05
GQ0083_E11	CK438292.2*	Receptor-like protein kinase	Q8LP72	1E-27		2.9	9.5E-08	2.3	4.6E-05

Table 1. *Continued*

Sequence[b] (cDNA model)	GenBank accession no.[b]	Annotation[c]				*UBIproPtMYB14*		*CADproPtMYB14*	
		Predicted function	Accession no.	E-value	AP[d]	FC[e]	*P*-value[f]	FC	*P*-value
GO:0045449 Regulation of transcription									
GQ03311_L15	BT112195	Zinc finger CCCH-type protein	Q10EL1	7E-05		1.5	5.7E-03	1.6	8.0E-03
GQ02902_P24	BT106103	Parathymosin-like	Q6ZFI5	2E-21		1.5	1.3E-04	1.4	7.1E-03
GO:0006950 Response to stress; GO:0006952 Defense response; GO:0009611 Response to wounding									
GQ02905_A21	BT106211	Heat shock protein (HSP70)	Q03683	5E-08		1.4	3.3E-04	1.4	5.3E-03
GQ0132_K16	BT101895	Heat shock protein (HSP17)	Q9ZSZ0	3E-03		–1.5	1.4E-03	–1.5	6.6E-03
GQ04105_K15	BT119437	Glycine-rich protein (Grp94)	Q8H6B6	8E-96		1.5	7.4E-04	1.5	4.3E-03
GQ04110_H09	BT119814	Glutathione *S*-transferase, Tau class	Q4PNY9	5E-57		2.2	2.7E-06	1.7	1.3E-03
GQ0204_N24	BT103013	Glutathione *S*-transferase, Tau class	O65056	3E-73		2.1	8.6E-07	1.6	2.3E-03
GQ02818_G19	BT105094	Dehydrin	Q3ZDL1	0E+00		1.4	1.4E-04	1.5	3.8E-04
GQ03702_K04	BT115610	ADR1-like 1, CC-NBS-LRR prot.	Q1L6F3	3E-77		1.4	2.2E-03	1.6	9.7E-04
GQ03308_B18	BT111949	ADR1-like 1, NBS-LRR protein	Q2VWL9	9E-18		2.7	1.3E-06	1.7	6.2E-03
GO:0006508 Proteolysis									
GQ02810_F11	BT104435	Serine carboxypeptidase-like 46	Q1EP77	3E-117		1.5	2.0E-05	1.5	3.3E-04
GQ03811_C08	BT117017	Serine carboxypeptidase-like 28	Q9SFB5	3E-47		1.3	6.2E-03	1.5	1.2E-03
Miscellaneous									
GQ03519_E01	BT114678	Plasma membrane-associated protein	Q1ECE0	1E-34		1.5	7.4E-04	1.6	5.0E-04
GQ04104_H01	BT119347	Alternative oxidase 1b	Q84V46	0E+00		1.3	6.2E-04	1.5	3.4E-04
GQ03310_I01	BT112130	Expressed protein	O82812	1E-19		1.5	4.8E-04	1.4	1.0E-02
GQ03102_D11	BT107053	DNAJ heat shock family protein	Q9SJZ7	7E-21		–1.4	2.6E-03	1.5	4.1E-03
GQ03237_M12	BT111442	No hit	N/A	N/A		1.7	5.5E-06	1.4	4.3E-03
PtMYB14	DQ399056	Transgene	ABD60279	0E+00		9.1	7.3E-08	2.0	3.2E-02

[a] The 53 shared genes were found in each of the constructs by analysing four replicates from two independent lines in each construct (*P*-value <0.01, and log2 ratio ⩾|0.5|, FDR of 1%). A permutation analysis using random assignment gave an average number of 0.153 overlapping genes occurring by chance and a probability of obtaining 53 genes $<2.04\times10^{-4}$ (1/4900).
[b] Representative cDNA clone from Arborea EST and full-length insert cDNA database (www.Arborea.ulaval.ca); accession numbers for cDNA clones GenBank, except where marked *; these are for ESTs (dbEST). IDs in bold indicate that the microarray result was validated by RT-qPCR.
[c] Functional annotation methods are as described in Pavy *et al.* (2005); accession numbers are for UniProt (www.uniprot.org/uniprot)
[d] AP, associated pathway: FV, flavonoid; TCA, tricarboxylic acid cycle; TP, terpenoid.
[e] Fold change (FC) relative to the wild type. (+) are over-represented and (–) are under-represented in transgenics.
[f] Student's t-test level of confidence.

Misregulation of transcripts related to metabolic processes, to defence, and to stress responses in PtMYB14-OE spruce

The overlapping transcript profile data indicate that *PtMYB14* overexpression most clearly impacted gene sequences linked to isoprenoid/terpenoid, flavonoid, pyruvate, and citrate [tricarboxylic acid (TCA)] metabolism, as well as the lipid catabolic process leading to JA biosynthesis (Table 1). The common set of up-regulated transcripts included several sequences that were most similar to genes of the cytosolic MVA pathway leading to isoprenoid formation, such as *acetoacetyl-CoA thiolase* (*AACT*), *3-hydroxy-3-methylglutaryl-CoA-synthase* (*HMGS*), and *mevalonate disphosphate decarboxylase* (*MVD1*), as well as an *α-farnesene synthase* (*AFS*) from the downstream pathway leading to sesquiterpenes, and a *chalcone synthase* (*CHS*) involved in flavonoid biosynthesis (Table 1). Congruent with these data, several transcripts associated with electron transport were misregulated, including a *cytochrome p450 monooxygenase*, a *flavonoid 3′,5′-hydroxylase*, and an *FAD-binding domain*-containing protein. Transcripts accumulated for *ATP-binding cassette (ABC) family transporters (PDR-like)*, defined as important partners in processes such as pigment accumulation, detoxification, and oxidation damage (Theodoulou, 2000), for *NAD-dependent malate oxidoreductase* and *ATP:citrate lyase* that could be linked to pyruvate and the citrate cycle (TCA), and for *lipoxygenase* (*LOX*), *acyl-CoA*

synthase, and *acyl-CoA thioesterase* linked to α-linolenic acid metabolism leading to JA biosynthesis.

A large proportion of the remaining misregulated sequences common to both constructs could be linked to defence and stress responses. These included sequences coding for activated disease resistance (ADR) 1 proteins (Table 1), defined by functional domains such as leucine-rich repeat (LRR), Toll-interleukin-1 receptor-like (TIR), coiled-coil (CC), receptor-like kinase, and NBD (nucleotide binding) which have been implicated in defence response and plant resistance (Fluhr, 2001). Other misregulated stress-responsive sequences, included those coding for dehydrin as well as glycine-rich (GRP) and heat-shock (HSP) proteins, in addition to glutathione *S*-transferases (Table 1). The strong up-regulation of wound-responsive class II and IV chitinases (Hietala *et al.*, 2004) was common to both constructs, together with the accumulation of transcripts encoding serine carboxypeptidase hydrolytic enzymes shown to function as acyltransferase in secondary metabolism (Lehfeldt *et al.*, 2000). Transcripts of reporter genes from the salicylic acid (SA) response such as *EDS1-like* and *phytoalexin deficient 4* (*PAD4*) (Wiermer *et al.*, 2005), as well as transcripts of *MAP kinase 6* associated with transcriptional regulation and signal transduction in the JA, SA, and ethylene response in *Arabidopsis* (Fujita *et al.*, 2006), also accumulated with both constructs (Table 1).

PtMYB14 *overexpression stimulated terpene and anthocyanin accumulation in spruce plantlets*

In light of the overlapping transcript profiles, the increased pigmentation, the presence of polyphenolic cells, and the accumulation of starch grains in parenchyma cells observed with *PtMYB14*-OE constructs, the aim was to gather further information on the metabolite content of transgenic plantlets.

Extensive identification and quantification of monoterpenoids, sesquiterpenoids, and diterpenoids, known plant defence coumpounds (Martin *et al.*, 2002; Keeling and Bohlman, 2006; Phillips *et al.*, 2007), was performed by GC-MS (Fig. 5, Supplementary Table S3 at *JXB* online). The diversity of sesquiterpenes was clearly increased following both tissue-preferential and constitutive expression of PtMYB14 in spruce plantlets (Fig. 5). Longifolene, α-bergamotene, germacrene D, and two unidentified sesquiterpenes were detected uniquely in the transgenics (Fig. 5). In addition, the amount of most of the detected sesquiterpenes increased significantly in *UBIproPtMYB14*-OE plantlets relative to controls (Supplementary Table S3 at *JXB* online). The total amount of monoterpenes also significantly increased 3-fold in *UBIproPtMYB14*-OE but not in the *CADproPtMYB14*-OE plantlets compared with controls (Fig. 5, Supplementary Table S3). Diterpene accumulation was not substantially affected in the transgenics plantlets.

The *UBIproPtMYB14*-OE plantlets also accumulated large amounts of anthocyanin *O*-glucosides, including delphinidin 3-*O*-glucoside, cyanidin 3-*O*-glucoside, petunidin 3-*O*-glucoside, peonidin 3-*O*-glucoside, and malvidin 3-*O*-glucoside in *UBIpro* (Supplementary Fig. S2A at *JXB* online). The *CADproPtMYB14*-OE plantlets only gave increased accumulation of cyanidin 3-*O*-glucoside. Starch content determinations in *UBIproPtMYB14*-OE plantlets showed a 2-fold increase compared with controls, but no increase was recorded in any of the *CADproPtMYB14*-OE lines (Supplementary Fig. S2B).

MYB14 *transcripts accumulated coordinately with those of isoprenoid/terpenoid-related genes following mechanical wounding*

In light of microarray and metabolite determination results (*PtMYB14*-OE plantlets), it was postulated that *PtMYB14* could contribute to the regulation of the isoprenoid and terpenoid pathways, and JA metabolism during defence responses (Table 1, Figure 6, Supplementary Table S3 at *JXB* online). Mechanically wounded seedlings (Fig. 2A) were analysed for accumulation of isoprenoid and terpenoid transcripts encoding all of the enzymes from the cytosolic MVA, and plastidial MEP pathways, together with those coding for prenyltransferases and terpene synthases. Transcript accumulation was observed for two reporter genes, a *defensin* (*DEF*) for wounding (Pervieux *et al.*, 2004) and an *AOC* for the JA pathway (Wasternack, 2007) (Fig. 6).

A clear wound response was observed for several isoprenoid and terpenoid pathway transcripts in both pine and spruce when considering cytosolic and plastidial routes. Transcripts for MVA pathway enzymes (AACT, HMGS, and MVD1) increased in both species; the *HMGS* RNA transcripts rapidly increased 15-fold and 5-fold in pine and spruce, respectively (Fig. 6). Only transcripts of the first MEP pathway enzyme, DXS, accumulated, giving 3-fold and 6-fold increases, respectively. Transcripts also accumulated for prenyltransferases (FPPS and GGPPS) and AFS from downstream pathways leading to sesquiterpenes. These transcript accumulation profiles were slightly delayed or simultaneous with *PgMYB14* and *PtMYB14*.

Discussion

In response to developmental and environmental cues, plants activate specific parts of their metabolism to produce a variety of primary and secondary metabolites that contribute to adaptative traits. TFs orchestrate the temporal and spatial expression of the numerous genes related to these complex metabolic routes. Several transcriptional regulators were shown to modulate phenylpropanoid, flavonoid, and benzenoid accumulation (Grotewold, 2005), but only a few have been linked to the regulation of isoprenoid metabolism (Xu *et al.*, 2004; Broun *et al.*, 2006; Memelink and Gantet, 2007). This report identified the pine *R2R3-MYB* gene *PtMYB14* as a putative partner in an isoprenoid-oriented response that leads to sesquiterpene accumulation in conifers. It also broadly illustrates the potential involvement of closely related Sg4C sequences in stress responses and plant evolution.

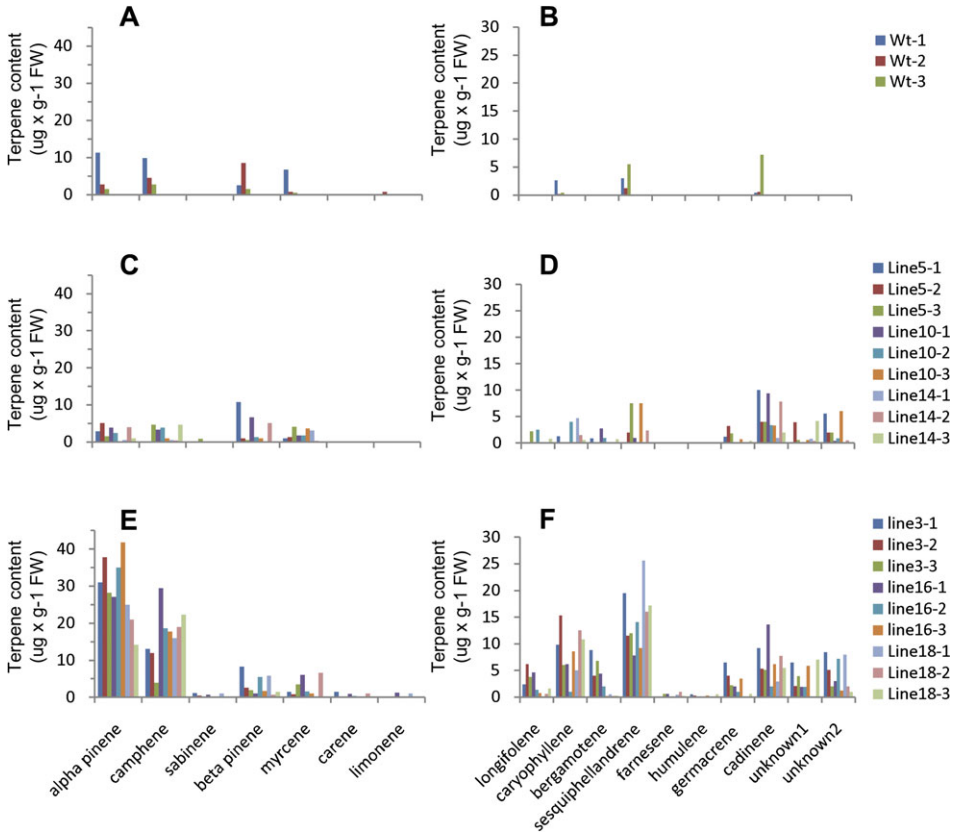

Fig. 5. Mono- and sesquiterpenoids in wild-type and PtMYB14-OE transgenic spruce (raw data). Monoterpenes (A, C, E) and (B, D, F) sesquiterpenes in wild-type (A, B), *CADproPtMYB14* (C, D), and *UBIproPtMYB14* (E, F) plantlets in three lines per transgenic and three replicates per line. Each value is a mean of a technical duplicate.

Overlapping transcription, histological, and metabolite phenotypes identified by tissue-specific and constitutive PtMYB14 overexpression approaches

Combining both constitutive and tissue-preferential overexpression strategies was an efficient way to define traits potentially related to PtMYB14 activity. The comparative microarray analyses provided a robust subset of 53 sequences that were mainly up-regulated (49 out of 53) in both types of *PtMYB14* overexpressors (Table 1). It is proposed that this set of shared misregulated sequences most probably includes genes and pathways directly linked to the *PtMYB14* activity, because their transcripts were modulated similarly (Table 1) despite the very different levels and spatial distributions of *PtMYB14* transgene expression observed with the two strategies (Fig. 3A–C). Although the *CADproPtMYB14*-OE plantlets had fewer misregulated sequences, the majority of them (around two-thirds) were part of the common set. The two types of transgenics also shared common histological and metabolite traits that were considered to be a consequence of PtMYB14 activity. They included (i) polyphenolic parenchyma cells, which are components of constitutive and inducible defence response in conifers (Franceschi *et al.*, 2005), observed within the vasculature, the pith, and the outer parenchyma (Fig. 3H, I); (ii) altered vascular organization and tracheid development compared with wild-type plantlets (Fig. 3); and (iii) an overlapping modulation in the terpene (Fig. 5, Supplementary Table S3 at *JXB* online) and anthocyanin profiles (Supplementary Fig. S2). These shared or overlapping phenotypes, i.e. transcript, histological, and metabolite profiles, are proposed to indicate candidate genes and metabolic processes that are influenced by PtMYB14 activity, as opposed to non-specific, or random effects leading to pleiotropic changes.

The phenotypes uniquely observed in the *UBIproPtMYB14* overexpressors were considered as potential pleiotropic effects (Fig. 3, Supplementary Fig. S2 at *JXB* online). They include additional effects on transcript profiles of genes belonging to other secondary metabolism pathways (Supplementary Table S1). For instance, many of the sequences that were down-regulated with the *UBIproPtMYB14* construct were linked to shikimate, monolignol, and benzenoid biosynthetic pathways (Supplementary Table S1). Other R2R3-MYBs have been implicated in the transcriptional control of phenylpropanoid (Zhong and Ye, 2010) and benzenoid metabolism through the upstream shikimate pathway (Verdonk *et al.*, 2005). This transcript profile was consistent with the additional modifications in the metabolite profile (i.e. monoterpenes, anthocyanins, and starch) that were observed when PtMYB14 was driven by *UBIpro* (Fig. 5, Supplementary Fig. S2). Due to the highly

conserved DBD observed among conifer MYBs (Bedon *et al.*, 2007), the genes that were uniquely down-regulated with *UBIpro* may reflect indirect effects caused by the high level of MYB expression, such as non-specific binding to gene promoters. The *PtMYB14* transcript level was ~100-fold higher for the *UBIproPtMYB14* than for the *CADproPtMYB14* construct when normalized to *CDC2* gene transcripts. Such a dosage-linked effect would influence most strongly the expression of genes with similar *cis*-regulatory motifs, as suggested in *Arabidopsis* (Jin *et al.*, 2000). Differences in the effects of the two overexpression strategies may also be due to spatial expression of the transgene. A second mechanism by which MYB overexpression may repress gene expression is by competing with cognate TFs (squelching) and thus sequestrating components of the transcriptional machinery away from *cis*-regulatory DNA elements (Gill and Ptashne, 1988). All the Sg4C sequences described herein share a signature motif similar to the basic helix–loop–helix (bHLH) interaction site identified by Zimmermann *et al.* (2004) in *Arabidopsis*. Thus, non-specific effects may entail interactions with others proteins, such as bHLH and WD40 which have been identified as important MYB partners in flavonoid pathways (Vom Endt *et al.*, 2002; Ramsay and Glover, 2005). Finally, based on the differential expression of other Sg4C transcript levels observed between the *UBIpro*- and *CADpro*-driven overexpression of *PtMYB14* (Fig. 4B), one can argue that strong and constitutive *PtMYB14* overexpression had a co-suppressing effect on other Sg4C members while it was not the case for moderate and tissue-preferential *PtMYB14* overexpression. If other Sg4C members also play a role in regulating defence and stress responses in conifers, as suggested by expression profiles (Fig. 2A–C), it might be expected that their down-regulation would affect gene expression and, in part, explain the large difference in misregulated sequences observed between *UBIpro*- and *CADpro*-driven overexpression of *PtMYB14*.

Overexpression phenotypes are consistent with a role for PtMYB14 *in the regulation of isoprenoid metabolism*

Isoprenoids are involved in numerous cellular processes and serve as hormones, defensive agents, membrane constituents, components of signal transduction networks, mating pheromones, and photoprotective agents (Sacchettini and Poulter, 1997). Among the isoprenic-derived compounds (Lichtenthaler, 1999; Lange *et al.*, 2000), sesquiterpenes and sterols are predominantly generated from the cytosolic MVA pathway, while monoterpenes, diterpenes and carotenoids are synthesized via the plastidic MEP pathway (Wu *et al.*, 2006). A few transcription factors have been linked to the regulation of isoprenoid metabolism in angiosperm plants (Broun *et al.*, 2006; Memelink and Gantet, 2007). To our knowledge, no MYB has been directly linked to the regulation of isoprenoid pathways.

The overlapping transcription and metabolite phenotypes observed upon PtMYB14 overexpression in spruce are clearly consistent with a role in the regulation of isoprenoid metabolism. Several of the up-regulated transcripts were related to early steps in isoprenoid biosynthesis in the cytosolic MVA pathway, such as *AACT*, *HMG-S*, and *MVD1* (Table 1). An AFS in the downstream sequiterpene pathway was also up-regulated (Table 1; Supplementary Table S2 at *JXB* online). Furthermore, the metabolite profiles clearly point to the accumulation of more diverse volatile oleoresins such as sesquiterpenes (Fig. 5 and Supplementary Table S3). Similarly, mechanical wounding in wild-type spruce and pine plantlets led to the coordinated accumulation of *Pt/PgMYB14* transcripts with those of *AACT* and *HMGS* from the MVA pathway, and prenyltransferase (FPPS) and terpene synthase (AFS) of the sesquiterpene pathway (Fig. 6). In Norway spruce, the induction of genes encoding key steps in the MEP pathway (i.e. DXS) was stimulated by mechanical wounding and by methyl jasmonate (MeJA) (Philipps *et al.*, 2007). Here, the wounding experiment impacted both MVA and MEP pathway transcripts (Fig. 6), but the microarray analyses of PtMYB14-overexpressors did not detect any differentially expressed MEP pathway transcripts (Table 1; Supplementary Table S2). Therefore, it was postulated that *Pt/PgMYB14* may be preferentially involved in isoprenoid metabolism through the cytosolic (MVA) rather than the plastidic (MEP) route, although it may be part of a complex signalling network that controls the wound response in conifers.

The accumulation of transcripts encoding flavonoid pathway enzymes (Table 1) and of cyanidin-*o*-glucoside (Supplementary Fig. S2 at *JXB* online) in both transgenics also suggested that PtMYB14 may have a role in the regulation of flavonoid metabolism. Anthocyanins are reported to act as UV-B protectants in addition to taking part in antioxidative defence (Steyn *et al.*, 2002), but their accumulation at particular developmental stages can be deleterious or interfere with normal development (Chawla *et al.*, 1999). Besseau *et al.* (2007) reported that flavonoids including anthocyanins can also interfere with plant growth and the development of tracheary elements in *Arabidopsis*. In spruce, both of the PtMYB14 overexpression approaches clearly altered vascular tissue organization (Fig. 3) and the common set of misregulated trancripts included sequences related to cell wall organization and biogenesis (i.e. expansin, extensin, and chitinases) (Table 1). This experimental evidence suggests two possible explanations. The PtMYB14 acts on genes associated with the anthocyanin pathway in specific tissues during development. Alternatively, PtMYB14 may be part of a complex defensive programme, including terpenoid and anthocyanin production. Further experiments are needed to test these hypotheses.

The differential gene family expansion of Sg4C R2R3-MYBs highlights the distinct evolutionary trajectories of angiosperm and gymnosperm plants

Several conifer sequences form a monophyletic subclade (Sg4C) within Sg4 of plant R2R3-MYBs previously defined by Kranz *et al.* (1998). Evidence was also presented that conifers have accumulated more Sg4 sequences than angiosperm genomes sequenced to date. Ten Sg4C members were

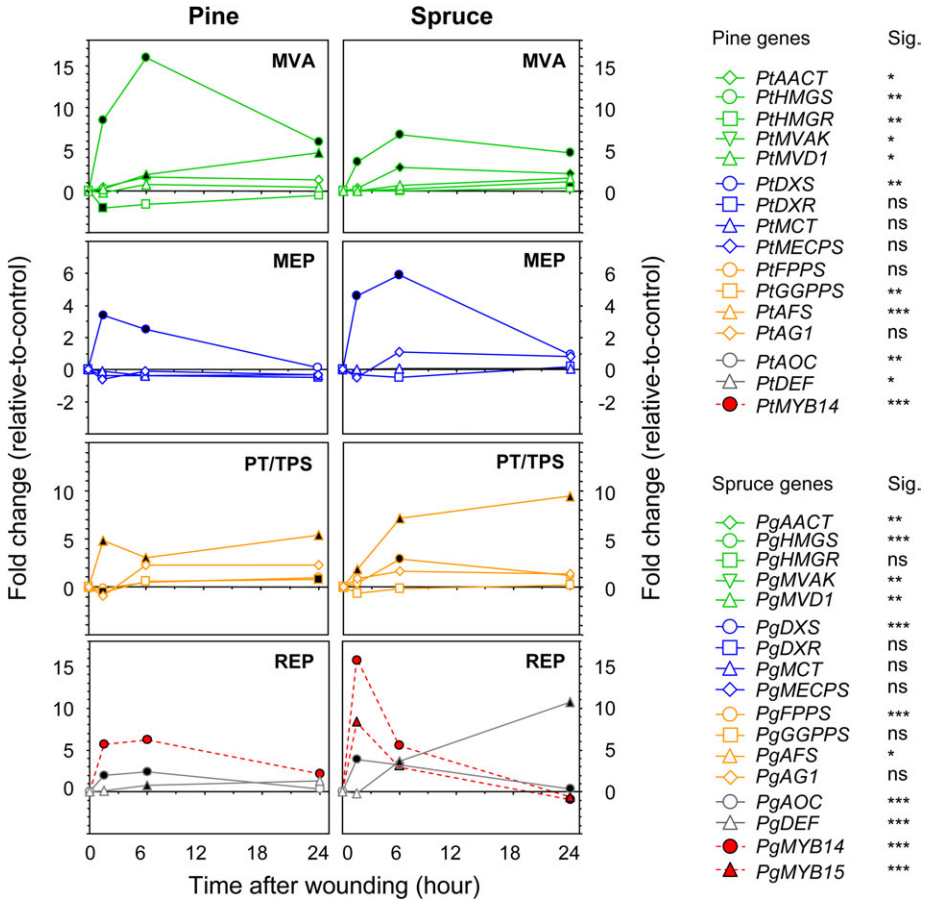

Fig. 6. Time course accumulation of isoprenoid- and terpenoid-related transcripts in response to mechanical wounding in wild-type plantlets. Samples are the same as in Fig. 2. The black-filled shapes indicate fold changes (relative to controls) that are statistically significant according to Student's test at 0.05 (*), 0.01 (**), or 0.001 (***). ns, non significant.

identified in *P. glauca*, whereas the angiosperms have between three and eight sequences. Considering that cDNA isolation is non-exhaustive, the Sg4C inventory probably remains incomplete. Interestingly, the grapevine (*V. vinifera*) genome has a total of eight Sg4 MYBs, six of which form a monophyletic clade (Matus *et al.*, 2008), representing an angiosperm Sg4 gene family structure that is analogous to the Sg4C in conifers.

Several analyses of expressed gene sequences provide evidence of monophyletic gene family amplifications that occurred after the angiosperm–gymnosperm split, as reported here for Sg4C MYBs. Specific examples include the *P. taeda* arabino-galactan proteins (Zhang *et al.*, 2000) and cellulose synthases (CesA) (Nairn and Haselkorn, 2005), and the *Picea* spp. KNOX-1 gene family (Guillet-Claude *et al.*, 2004) and dirigent proteins (Ralph *et al.*, 2007). This latter protein family and the expansin superfamily (Sampedro and Cosgrove, 2005) provide striking examples of entire subfamilies or subclades that are specific to conifers. Some of these studies have also pointed to reduced complexity of gene family structure in conifers. For example, conifer sequences have been found in only one of the three subbranches in the KNOX-1 gene family, although all of the conifers examined had more gene duplications within this branch (Guillet-Claude *et al.*, 2004). Taken together, these data clearly illustrate the evolutionary trajectories of conifer gene families as being distinct from angiosperms.

The gymnosperms account for four of the five divisions/phyla of seed plants (spermatophytes). A very deep split exists between the gymnosperms which appeared ~300 Mya and the more recent angioperms estimated to be ~190 million years old (Magallón and Sanderson, 2005). While conifers (coniferales) appeared relatively early in the evolution of gymnosperms, the pinaceae family and its major genera, including those of *Pinus* and *Picea*, are estimated to have radiated ~90–100 Mya. The phylogenetic trees of plant Sg4 sequences include multiple sequences within both the *Picea* and *Pinus* subfamilies (Fig. 1, Supplementary Table S1 at *JXB* online). This enables it to be postulated that many of the gene duplication events that gave rise to Sg4C actually occurred between 100 and nearly 300 Mya. The relatively short branches in the Sg4C subclade may not suggest such an ancient time frame for these gene duplications; however, conifers are very long lived and notorious for their slow rate of evolution (Savolainen and Pyhäjärvi, 2007). Further investigations such as the estimation of rates of substitution and molecular clock analyses are needed to shed light on this question.

The maintenance of duplicated genes within Sg4C over tens of millions of years suggests that selective advantages

have been derived over time from the additional gene copies. Such advantages may arise in the shorter term from enhanced expression, or with time from sequence divergence resulting in modified biochemical functions or biological roles. Expression studies in *Arabidopsis* indicated that highly homologous sequences may have different expression profiles under particular conditions or treatments (Yanhui *et al.*, 2006). Not surprisingly, widespread duplication has been previously invoked as an important mechanism for *MYB* gene evolution leading to the acquisition of new functions, as well as enabling plant adaptation to new environments (Grotewold, 2005). For example, duplication of the MYB regulator *p1* and *p2* genes in maize from the ancestral *p* gene resulted in two different expression profiles related to flavonoid regulation. The *p1* gene is expressed in the kernel pericarp, and *p2* is expressed in the developing anther and silk (Zhang *et al.*, 2000).

Transcript expression data suggest that the Sg4C sequences have diverged functionally to some extent. A survey of transcript profiles across different tissues in spruce (Bedon *et al.*, 2007; Fig. 2D), as well as changes in transcript accumulation levels upon wounding, JA application, and exposure to cold (Fig. 2A–C), clearly indicate differential regulation. The majority (but not all) of Sg4C were shown to respond to at least one of the stress treatments tested. Their stress responsiveness and the evidence supporting the involvement of Pt/PgMYB14 in the regulation of isoprenoid metabolism (Fig. 6) lead us to postulate that Sg4C sequences have contributed to the adaptation of gymnosperms and conifers to changing environments throughout evolution.

Supplementary data

Supplementary data are available at *JXB* online.

Figure S1. Number and functional grouping of genes differentially expressed in *UBIpro*- and *CADproPtMYB14*-OE spruce plantlets.

Figure S2. Accumulation of anthocyanins and starch in wild-type and *PtMYB14*-OE transgenic spruce.

Table S1. Phylogenetic analysis of the conifer, and angiosperm subgroup 4 R2R3-MYB sequences with differents methods of alignments and phylogeny.

Table S2. Functional grouping of genes differentially expressed in *UBIpro*- and *CADproPtMYB14*-OE spruces.

Table S3. Mono-, di-, and sesquiterpenoids in *PtMYB14*-OE transgenic spruce (raw data).

Table S4. Primers sequences used to generate DNA matrices for RT-qPCR analyses in pine.

Table S5. Primer sequences used for RT-qPCR in pine and spruce

Acknowledgements

The authors thank R. R. Sederoff (North Carolina State University, Raleigh, NC) for the *P. taeda* cDNA clones, N. Dallaire (Laval University) for assistance with RT-qPCR analyses, J. E. K. Cooke (University of Alberta) for microarray development, L. Tremblay (Canadian Forest Service) for assistance in tissue culture, and Stéphane LeBihan and Colleen Nelson (British-Columbia Microarray Facility) for aid with microarray development and manufacture. The authors thank the Max Planck Society for funds, to AS and JG, which supported the analytical equipment and supplies used for the terpene analysis. Funding was received from Genome Canada and Génome Québec to JM and AS for the Arborea project and from NSERC of Canada to JM.

References

Agarwal M, Hao Y, Kapoor A, Dong C-H, Fujii H, Zheng X, Zhu J- K. 2006. A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *Journal of Biological Chemistry* **281,** 37636–37645.

Agrawal AA. 2007. Macroevolution of plant defense strategies. *Trends in Ecology & Evolution* **22,** 103–109.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings/ International Conference on Intelligent Systems for Molecular Biology* **2,** 28–36.

Bedon F, Grima-Pettenati J, Mackay J. 2007. Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (Picea glauca). *BMC Plant Biology* **7,** 17.

Bedon F, Levasseur C, Grima-Pettenati J, Séguin A, MacKay J. 2009. Sequence analysis and functional characterization of the promoter of the Picea glauca cinnamyl alcohol dehydrogenase gene in transgenic white spruce plants. *Plant Cell Reports* **28,** 787–800.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57,** 289–300.

Besseau S, Hoffman L, Geoffroy P, Lapierre C, Pollet B, Legrand M. 2007. Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *The Plant Cell* **19,** 148–162.

Bomal C, Bedon F, Caron S, *et al.* 2008. Involvement of *Pinus taeda MYB1* and *MYB8* in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative *in planta* analysis. *Journal of Experimental Botany* **59,** 3925–3939.

Brooker RW. 2006. Plant–plant interactions and environmental change. *New Phytologist* **171,** 271–284.

Broun P, Liu Y, Queen E, Schwarz Y, Abenes MA, Leibman M. 2006. Importance of transcription factors in the regulation of plant secondary metabolism and their relevance to the control of terpenoid accumulation. *Phytochemistry Reviews* **5,** 27–38.

Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11,** 113–116.

Chawla HS, Cass LA, Simmonds JA. 1999. Developmental and environmental regulation of anthocyanin pigmentation in wheat tissues

transformed with anthocyanin regulator genes. *In Vitro Cellular & Developmental Biology–Plant* **35,** 403–408.

Christensen AH, Sharrock RA, Quail PH. 1992. Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Molecular Biology* **18,** 675–689.

Coleman HD, Ellis DD, Gilbert M, Mansfield SD. 2006. Up-regulation of sucrose synthase and UDP-glucose pyrophosphorylase impacts plant growth and metabolism. *Plant Biotechnology Journal* **4,** 87–101.

Côté C, Rutledge RG. 2003. An improved MUG fluorescent assay for the determination of GUS activity within transgenic tissue of woody plants. *Plant Cell Reports* **21,** 619–624.

Fluhr R. 2001. Sentinels of disease. Plant resistance genes. *Plant Physiology* **127,** 1367–1374.

Franceschi VR, Krokene P, Christiansen E, Krekling T. 2005. Anatomical and chemical defenses of conifer bark against bark beetles and other pests (Tansley Review). *New Phytologist* **167,** 353–376.

Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, Shinozaki K. 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Current Opinion in Plant Biology* **9,** 436–442.

Gill G, Ptashne M. 1988. Negative effect of the transcriptional activator GAL4. *Nature* **334,** 721–724.

Gomez-Maldonado J, Avila C, Torre F, Canas R, Canovas FM, Campbell MM. 2004. Functional interactions between a glutamine synthetase promoter and MYB proteins. *The Plant Journal* **39,** 513–526.

Grotewold E. 2005. Plant metabolic diversity: a regulatory perspective. *Trends in Plant Science* **10,** 57–62.

Guillet-Claude C, Isabel N, Pelgas B, Bousquet J. 2004. The evolutionary implications of *knox-I* gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence. *Molecular Biology and Evolution* **21,** 2232–2245.

Hawkins S, Samaj J, Lauvergeat V, Boudet A, Grima-Pettenati J. 1997. Cinnamyl alcohol dehydrogenase: identification of new sites of promoter activity in transgenic poplar. *Plant Physiology* **113,** 321–325.

Hietala AM, Kvaalen H, Schmidt A, Jøhnk N, Solheim H, Fossdal CG. 2004. Temporal and spatial profiles of chitinase expression by Norway spruce in response to bark colonization by *Heterobasidion annosum*. *Applied and Environmental Microbiology* **70,** 3948–3953.

Jia L, Clegg MT, Jiang T. 2004. Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies *indica* and *japonica* genomes. *Plant Physiology* **134,** 575–585.

Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C. 2000. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *EMBO Journal* **19,** 6150–6161.

Karpinska B, Karlsson M, Srivastava M, Stenberg A, Schrader J, Sterky F, Bhalerao R, Wingsle G. 2004. MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen. *Plant Molecular Biology* **56,** 255–270.

Kazan K. 2006. Negative regulation of defence and stress genes by EAR-motif-containing repressors. *Trends in Plant Science* **11,** 109–112.

Keeling CI, Bohlmann J. 2006. Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytologist* **170,** 657–675.

Kirst M, Johnson AF, Baucom C, Ulrich E, Hubbard K, Staggs R, Paule C, Retzel E, Whetten R, Sederoff R. 2003. Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **100,** 7383–7388.

Kranz HD, Denekamp M, Greco R, *et al*. 1998. Towards functional characterization of the members of the *R2R3-MYB* gene family from *Arabidopsis thaliana*. *The Plant Journal* **16,** 263–276.

Lange BM, Rujan T, Martin W, Croteau R. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences, USA* **97,** 13172–13177.

Legay S, Lacombe E, Goicoechea M, Brière C, Séguin A, Mackay J, Grima-Pettenati J. 2007. Molecular characterization of *EgMYB1*, a putative transcriptional repressor of the lignin biosynthetic pathway. *Plant Science* **173,** 542–549.

Lehfeldt C, Shirley AM, Meyer K, Ruegger MO, Cusumano JC, Viitanen PV, Strack D, Chapple C. 2000. Cloning of the *SNG1* gene of Arabidopsis reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism. *The Plant Cell* **12,** 1295–1306.

Li J, Yang X, Wang Y, Li X, Gao Z, Pei M, Chen Z, Qu LJ, Gu H. 2006. Two groups of MYB transcription factors share a motif which enhances trans-activation activity. *Biochemical and Biophysical Research Communications* **341,** 1155–1163.

Lichtenthaler HK. 1999. The 1-deoxy-d-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **50,** 47–65.

Magallón SA, Sanderson MJ. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evolution* **59,** 1653–1670.

Martin D, Tholl D, Gershenzon J, Bohlmann J. 2002. Methyl jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of Norway spruce stems. *Plant Physiology* **129,** 1003–1018.

Matus JT, Aquea F, Arce-Johnson P. 2008. Analysis of the grape *MYB R2R3* subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. *BMC Plant Biology* **8,** 83.

Memelink J, Gantet P. 2007. Transcription factors involved in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Phytochemistry Reviews* **6,** 353–362.

Nairn CJ, Haselkorn T. 2005. Three loblolly pine *CesA* genes expressed in developing xylem are orthologous to secondary cell wall CesA genes of angiosperms. *New Phytologist* **166,** 907–915.

Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M. 2001. Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *The Plant Cell* **13,** 1959–1968.

Patzlaff A, McInnis S, Courtenay A, *et al*. 2003*b*. Characterisation of a pine MYB that regulates lignification. *The Plant Journal* **36,** 743–754.

Patzlaff A, Newman LJ, Dubos C, Whetten RW, Smith C, McInnis S, Bevan MW, Sederoff RR, Campbell MM. 2003*a*. Characterisation of *Pt*MYB1, an R2R3-MYB from pine xylem. *Plant Molecular Biology* **53,** 597–608.

Pavy N, Laroche J, Bousquet J, Mackay J. 2005. Large-scale statistical analysis of secondary xylem ESTs in pine. *Plant Molecular Biology* **57,** 203–224.

Pervieux I, Bourassa M, Laurans F, Hamelin R, Séguin A. 2004. A spruce defensin showing strong antifungal activity and increased transcript accumulation after wounding and jasmonate treatments. *Physiological and Molecular Plant Pathology* **64,** 331–341.

Philipps MA, Walter MH, Ralph SG, *et al*. 2007. Functional identification and differential expression of 1-deoxy-d-xylulose 5-phosphate synthase in induced terpenoid resin formation of Norway spruce (*Picea abies*). *Plant Molecular Biology* **65,** 243–257.

Ralph SG, Jancsik S, Bohlmann J. 2007. Dirigent proteins in conifer defense II: extended gene discovery, phylogeny, and constitutive and stress-induced gene expression in spruce (*Picea* spp). *Phytochemistry* **68,** 1975–1991.

Ramsay NA, Glover BJ. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science* **10,** 63–70.

Romero I, Fuertes A, Benito MJ, Malpica JM, Leyva A, Paz-Ares J. 1998. More than 80 *R2R3 MYB* regulatory genes in the genome of *Arabidopsis thaliana*. *The Plant Journal* **14,** 273–284.

Sacchettini JC, Poulter CD. 1997. Creating isoprenoid diversity. *Science* **277,** 1788–1789.

Sampedro J, Cosgrove DJ. 2005. The expansin superfamily. *Genome Biology* **6,** 242.

Savolainen O, Pyhäjärvi T. 2007. Genomic diversity in forest trees. *Current Opinion in Plant Biology* **10,** 162–167.

Sharman BC. 1943. Tannic acid and iron alum with safranin and Orange G in studies of the shoot apex. *Stain Technology* **18,** 105–111.

Singh KB, Foley RC, Oñate-Sánchez L. 2002. Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology.* **5,** 430–436.

Steyn WJ, Wand SJE, Holcroft DM, Jacobs G. 2002. Anthocyanins in vegetative tissues: a proposed unified function in photoprotection. *New Phytologist* **152,** 349–361.

Taki N, Sasaki-Sekimoto Y, Obayashi T, *et al*. 2005. 12-Oxo-phytodienoic acid triggers expression of a distinct set of genes and plays a role in wound-induced gene expression in Arabidopsis. *Plant Physiology* **139,** 1268–1283.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24,** 1596–1599.

Theodoulou FL. 2000. Plant ABC transporters. *Biochimica et Biophysica Acta* **1465,** 79–103.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22,** 4673–4680.

Verdonk JC, Haring MA, van Tunen AJ, Schuurink RC. 2005. *ODORANT1* regulates fragrance biosynthesis in petunia flowers. *The Plant Cell* **17,** 1612–1624.

Vom Endt D, Kijne JW, Memelink J. 2002. Transcription factors controlling plant secondary metabolism: what regulates the regulators? *Phytochemistry* **61,** 107–114.

Wasternack C. 2007. Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Annals of Botany* **100,** 681–697.

Weigel RR, Pfitzner UM, Gatz C. 2005. Interaction of NIMIN1 with NPR1 modulates *PR* gene expression in Arabidopsis. *The Plant Cell* **17,** 1279–1291.

Wiermer M, Feys BJ, Parker JE. 2005. Plant immunity: the EDS1 regulatory node. *Current Opinion in Plant Biology* **8,** 383–389.

Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM. 2009. Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiology* **149,** 981–993.

Wu S, Schalk M, Clark A, Miles RB, Coates R, Chappell J. 2006. Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nature Biotechnology* **24,** 1441–1447.

Xu YH, Wang JW, Wang S, Wang JY, Chen XY. 2004. Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (1)-d-cadinene synthase-A. *Plant Physiology* **135,** 507–515.

Xue B, Charest PJ, Devantier Y, Rutledge RG. 2003. Characterization of a *MYBR2R3* gene from black spruce (*Picea mariana*) that shares functional conservation with maize *C1*. *Molecular Genetics and Genomics* **270,** 78–86.

Yanhui C, Xiaoyuan Y, Kun H, *et al*. 2006. The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology* **60,** 107–124.

Zhang JZ. 2003. Overexpression analysis of plant transcription factors. *Current Opinion in Plant Biology* **6,** 430–440.

Zhang P, Chopra S, Peterson T. 2000. A segmental gene duplication generated differentially expressed *myb*-homologous genes in maize. *The Plant Cell* **12,** 2311–2322.

Zhong R, Ye Z- H. 2010. Transcriptional regulation of lignin biosynthesis. *Plant Signaling & Behaviour* **4,** 1028–1034.

Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF. 2004. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal* **40,** 22–34.