Available online at www.sciencedirect.com
ScienceDirect
Phytochemistry 68 (2007) 1975–1991
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

# Dirigent proteins in conifer defense II: Extended gene discovery, phylogeny, and constitutive and stress-induced gene expression in spruce (*Picea* spp.) ☆

Steven G. Ralph [a,1], Sharon Jancsik [a], Jörg Bohlmann [a,b,*]

[a] *Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4*
[b] *Departments of Botany and Forest Sciences, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4*

Received 2 December 2006; received in revised form 17 April 2007

**Abstract**

Analysis of expressed sequence tags (ESTs) and full-length (FL)cDNAs from species of spruce (*Picea* spp.) revealed a family of 35 unique dirigent proteins (DIR) and DIR-like proteins. Phylogenetic analysis indicates the spruce DIR and DIR-like genes cluster into three distinct subfamilies, DIR-a, DIR-b/d, and DIR-f, of a larger plant DIR and DIR-like gene family. Gene-specific primers were designed for 31 unique spruce DIR family genes, and closely related isoforms, and used to evaluate patterns of constitutive expression, as well as responses to herbivory by stem-boring insects (i.e., white pine weevil, *Pissodes strobi*) in bark tissue and defoliating insects (i.e., western spruce budworm, *Choristoneura occidentalis*) in green apical shoots. Furthermore, meta-analysis of microarray gene expression data obtained from a series of independent experiments using the same 16.7K cDNA array platform identified several distinct expression clusters of the spruce DIR transcriptome closely matching phylogenetic clusters of sequence similarity. Members of the DIR-a family, which also contains functionally characterized DIR from other plant species, are most prominent for their induced response to feeding by weevils on Sitka spruce bark.


*Keywords:* Sitka spruce; *Picea sitchensis*; Coniferaceae; Conifer genomics; Phenylpropanoid; Plant–insect interactions; White pine weevil; Western spruce budworm; cDNA microarray; Wood formation; Meta-analysis; Expressed sequence tag; Real-time PCR; Dirigent proteins

## 1. Introduction

Conifers are well known for their extensive phenolic and terpenoid secondary metabolism. While conifer terpenoids in the form of oleoresin secretions contribute primarily to defense and resistance against insects and pathogens (Keeling and Bohlmann, 2006), conifer phenolics have essential functions both as building blocks for wood formation and as chemical defenses. Under the umbrella of a large-scale spruce (*Picea* spp.) genome project, we are mining spruce expressed sequence tag (EST) and full-length (FL)cDNA sequences, as well as microarray gene expression data, for genes of conifer secondary metabolism (Ralph et al., 2006a). In a previous paper, we provided a first inventory and gene expression analysis of spruce dirigent (DIR) and DIR-like proteins (Ralph et al., 2006b). DIR proteins from *Forsythia suspensa* (Davin et al., 1997), *Podophyllum peltatum* (Xia et al., 2000), and *Thuja plicata* (western red cedar; Kim et al., 2002) have been identified biochemically to direct the stereospecific coupling of *E*-coniferyl alcohol to produce the lignan (+)-pinoresinol.

☆ The work was supported by Genome British Columbia, Genome Canada and the Province of British Columbia (Treenomix Conifer Forest Health grant to J.B.), and by the Natural Science and Engineering Research Council of Canada (NSERC, grant to J.B.). J.B. is an NSERC Steacie Memorial Fellow.
* Corresponding author. Address: Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4. Tel.: +1 604 822 0282; fax: +1 604 822 2114.
*E-mail address:* bohlmann@interchange.ubc.ca (J. Bohlmann).
[1] Address: Department of Biology, University of North Dakota, Grand Forks, ND, United States of America.

0031-9422/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2007.04.042

Since lignans may have defensive functions against insects or pathogens in species of spruce, we are interested in a comprehensive characterization of this gene family. Our previous analysis of spruce ESTs and FLcDNAs identified 19 unique DIR and DIR-like sequences (*PDIR1–PDIR19*) from Sitka spruce (*Picea sitchensis*), white spruce (*P. glauca*), and hybrid interior spruce (*P. glauca x engelmannii*) (Ralph et al., 2006b). These spruce DIR and DIR-like sequences cluster into two subfamilies, DIR-a and DIR-b, respectively, with other subfamilies containing angiosperm genes. Transcripts for members of the DIR-a family were found preferentially expressed in outer stem tissues of Sitka spruce saplings and were also strongly induced in Sitka spruce bark and xylem upon insect attack (feeding white pine weevils; *Pissodes strobi*) or mechanical wounding. Here we report an extended inventory and sequence analysis of the spruce DIR- and DIR-like gene family with the inclusion of 16 new members, *PDIR20–PDIR35*. A detailed quantitative real-time PCR expression analysis in constitutive as well as insect-induced Sitka spruce tissues is described for 31 genes and closely related isoforms. Finally, we provide a transcriptome meta-analysis of the spruce DIR- and DIR-like genes, which is based on data derived from several different microarray gene expression studies performed with a 16.7K spruce cDNA microarray platform.

## 2. Results and discussion

### 2.1. cDNA Cloning of DIR and DIR-like genes from spruce

In order to extend earlier efforts to identify DIR and DIR-like genes from conifers (Ralph et al., 2006b), a TBLASTN search of the spruce EST and FLcDNA databases of the Treenomix project (Ralph et al., 2006a) was performed using angiosperm and gymnosperm sequences representing each of the five previously identified dirigent protein subfamilies. This gene mining identified 158 new ESTs that were not present in an earlier version of the spruce database, thus extending the collection of spruce DIR and DIR-like sequences to 254 ESTs. An *in silico* comparison of these sequences, combined with complete insert sequencing of 27 clones, revealed 15 unique full-length DIR and DIR-like cDNAs (named *PDIR20* to *PDIR34* in accordance with the nomenclature of earlier work), as well as one new partial cDNA, *PDIR35* (Table 1 and Fig. 1). Combined with the 19 full-length or partial DIR and DIR-like cDNAs we have previously described (Ralph et al., 2006b), this now extends the family to 35 unique members in spruce.

Pairwise sequence similarities among predicted amino acids of the 35 spruce DIR cDNAs range from a high of 99.5% identity between *PDIR12* and *PDIR21* and a low of 17.6% between *PDIR18* and *PDIR26* (Fig. 2). There are several examples of closely related proteins sharing amino acid identity greater than 98% that may represent within-species or species-specific alleles of the same genes including the following pairs: *PDIR12* (white spruce) and *PDIR21* (Sitka spruce) at 99.5%, *PDIR23* and *PDIR24* (both Sitka spruce) at 99.4%, *PDIR31* and *PDIR30* (both Sitka spruce) at 99.4%, *PDIR2* (interior spruce) and *PDIR32* (Sitka spruce) at 98.5%, *PDIR1* (white spruce) and *PDIR20* (Sitka spruce) at 98.4%, *PDIR3* (interior spruce) and *PDIR7* (white spruce) at 98.4%, and *PDIR8* (interior spruce) and *PDIR33* (Sitka spruce) at 98.0% (Fig. 2). The predicted open reading frames (ORFs) for the 15 new FLcDNAs range from 164 (*PDIR30* and *PDIR31*) to 196 (*PDIR33*) amino acids, and have predicted *pI* values ranging from 5.01 (*PDIR29*) to 7.90 (*PDIR20*) (Table 1). The predicted molecular masses range from *ca.* 17.4 (*PDIR30*) to 21.7 (*PDIR32* and *PDIR33*) kDa.

Table 1
Clone ID, gene name, protein and transcript features of spruce DIR and DIR-like genes identified in this study[a]

| Clone ID | *DIR* nomenclature | cDNA library | cDNA length (bp) | ORF length (aa) | *pI* | MW (kDa) |
|---|---|---|---|---|---|---|
| WS02714_K21 | *PDIR20* | SS-IL-A-FL-14[b] | 907 | 186 | 7.90 | 20.1 |
| WS02730_M07 | *PDIR21* | SS-IL-A-FL-14[b] | 906 | 184 | 6.29 | 19.5 |
| WS0273_B17 | *PDIR22* | SS-IL-A-FL-14[b] | 955 | 174 | 6.83 | 18.8 |
| WS02812_D09 | *PDIR23* | SS-IB-A-FL-13[c] | 1247 | 168 | 6.79 | 18.1 |
| WS0297_K02 | *PDIR24* | SS-IB-A-FL-15[c] | 882 | 168 | 6.75 | 18.1 |
| WS0273_G20 | *PDIR25* | SS-IL-A-FL-14[b] | 772 | 175 | 6.72 | 18.9 |
| WS02913_F24 | *PDIR26* | SS-IB-A-FL-15[c] | 783 | 177 | 6.57 | 19.0 |
| WS02810_L21 | *PDIR27* | SS-IB-A-FL-13[c] | 758 | 176 | 5.88 | 19.1 |
| WS0273_I15 | *PDIR28* | SS-IL-A-FL-14[b] | 863 | 172 | 5.04 | 18.3 |
| WS02753_H08 | *PDIR29* | SS-IL-A-FL-14[b] | 840 | 172 | 5.01 | 18.3 |
| WS0282_G20 | *PDIR30* | SS-IB-A-FL-13[c] | 798 | 164 | 6.07 | 17.4 |
| WS02745_C03 | *PDIR31* | SS-IL-A-FL-14[b] | 927 | 164 | 6.07 | 17.5 |
| WS0297_I02 | *PDIR32* | SS-IB-A-FL-15[c] | 915 | 195 | 6.49 | 21.7 |
| WS0284_A11 | *PDIR33* | SS-IB-A-FL-13[c] | 926 | 196 | 6.09 | 21.7 |
| WS02741_I08 | *PDIR34* | SS-IL-A-FL-14[b] | 865 | 194 | 6.42 | 21.5 |
| WS02914_C22[d] | *PDIR35* | SS-IB-A-FL-15[c] | 556 | 148 | 5.86 | 15.8 |

[a] All clones derived from *Picea sitchensis* FB3-425 genotype. *PDIR1-PDIR19* have been described in Ralph et al. (2006b).
[b] Green apical shoot tissue harvested following *Choristoneura occidentalis* budworm herbivory.
[c] Bark tissue (with phloem and cambium attached) harvested following *Pissodes strobi* weevil herbivory.
[d] Partial cDNA.

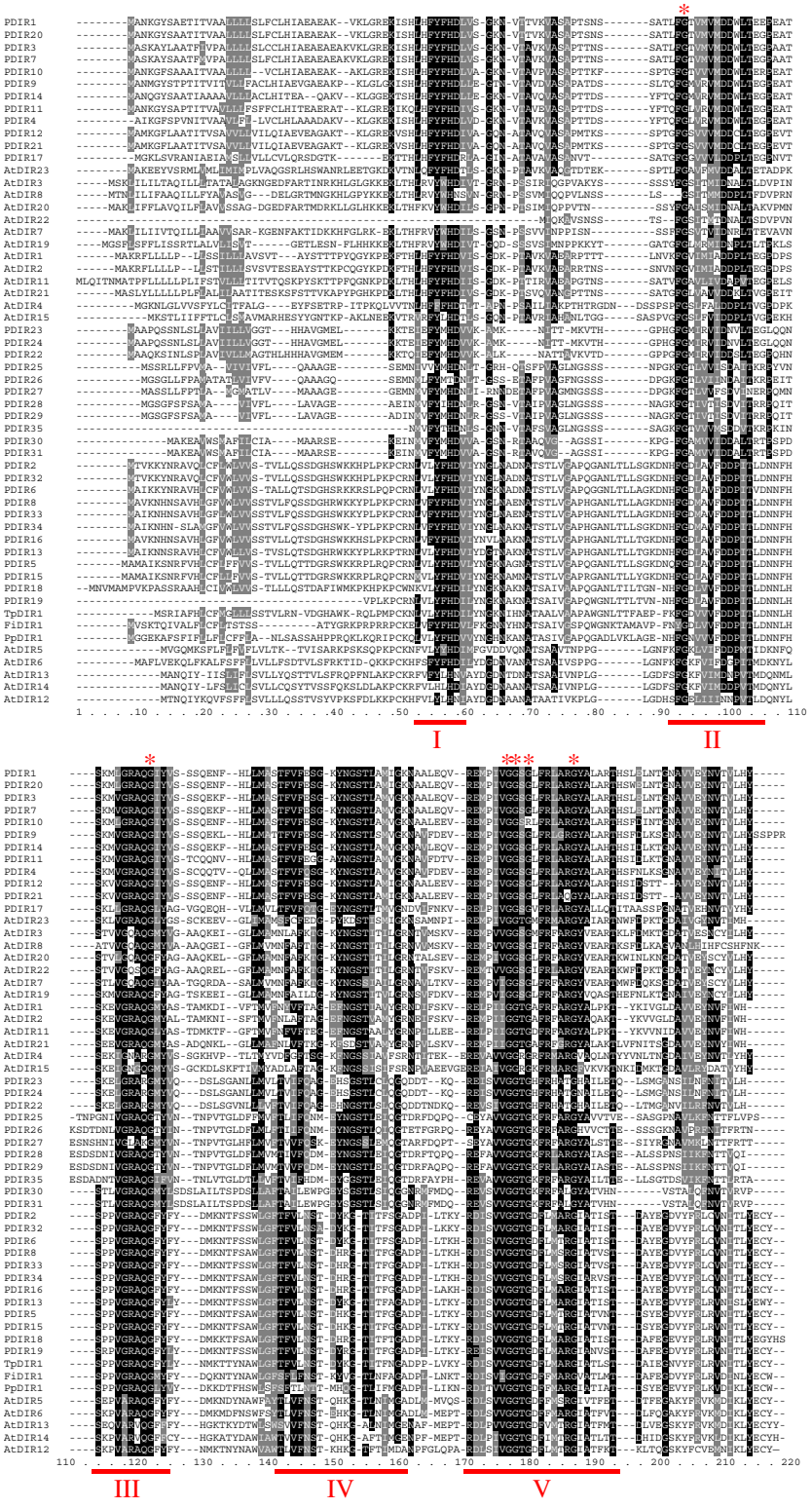


Fig. 1. Amino acid sequence alignment of the DIR family from spruce (P; *Picea* spp.) along with selected *Arabidopsis thaliana* (At) proteins and functionally characterized plant DIR proteins. Alignment generated using ClustalW (blosum matrix, gap open and gap extension penalties of 5 and 1.0, respectively) and Boxshade. Conserved similarity shading is based on 50% identity (black) and 50% similarity (gray). Amino acid motifs characteristic of the entire plant DIR family (see Section 4) are marked by red bars. Highly conserved amino acid residues present in more than 95% of the 150 plant DIR and DIR-like proteins analyzed in this study are marked with a red asterisk. For additional details of each gene see Table 2. PDIR18 is shown without the C-terminal 22 amino acids.

| | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | aa | ab | ac | ad | ae | af | ag | ah | ai |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDIR2 (a) | -- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR32 (b) | 98 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR6 (c) | 92 | 91 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR8 (d) | 88 | 88 | 86 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR33 (e) | 89 | 88 | 87 | 98 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR34 (f) | 86 | 86 | 84 | 94 | 94 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR16 (g) | 90 | 90 | 88 | 93 | 93 | 90 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR13 (h) | 88 | 87 | 86 | 85 | 85 | 83 | 87 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR5 (i) | 86 | 84 | 85 | 82 | 82 | 80 | 81 | 81 | -- | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR15 (j) | 85 | 84 | 85 | 82 | 83 | 80 | 82 | 82 | 96 | -- | | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR18 (k) | 67 | 67 | 65 | 65 | 64 | 63 | 66 | 66 | 66 | 66 | -- | | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR19 (l) | 69 | 68 | 68 | 66 | 67 | 68 | 67 | 68 | 67 | 66 | 61 | -- | | | | | | | | | | | | | | | | | | | | | | | |
| PDIR1 (m) | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 27 | 24 | 24 | 21 | 24 | -- | | | | | | | | | | | | | | | | | | | | | | |
| PDIR20 (n) | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 28 | 24 | 24 | 22 | 24 | 98 | -- | | | | | | | | | | | | | | | | | | | | | |
| PDIR3 (o) | 28 | 28 | 28 | 28 | 28 | 27 | 28 | 28 | 25 | 25 | 23 | 25 | 90 | 91 | -- | | | | | | | | | | | | | | | | | | | | |
| PDIR7 (p) | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 29 | 25 | 25 | 23 | 25 | 92 | 92 | 98 | -- | | | | | | | | | | | | | | | | | | | |
| PDIR10 (q) | 26 | 26 | 26 | 26 | 26 | 25 | 26 | 27 | 24 | 24 | 22 | 24 | 88 | 88 | 84 | 84 | -- | | | | | | | | | | | | | | | | | | |
| PDIR9 (r) | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 26 | 24 | 23 | 21 | 23 | 73 | 74 | 75 | 75 | 74 | -- | | | | | | | | | | | | | | | | | |
| PDIR14 (s) | 28 | 28 | 28 | 26 | 26 | 26 | 27 | 29 | 26 | 24 | 23 | 26 | 81 | 82 | 82 | 83 | 80 | 80 | -- | | | | | | | | | | | | | | | | |
| PDIR11 (t) | 25 | 26 | 26 | 25 | 24 | 25 | 25 | 27 | 24 | 23 | 22 | 24 | 77 | 77 | 76 | 76 | 75 | 74 | 78 | -- | | | | | | | | | | | | | | | |
| PDIR4 (u) | 25 | 25 | 25 | 24 | 25 | 25 | 24 | 26 | 24 | 22 | 21 | 24 | 77 | 78 | 76 | 76 | 79 | 75 | 79 | 79 | -- | | | | | | | | | | | | | | |
| PDIR12 (v) | 26 | 26 | 26 | 26 | 26 | 25 | 26 | 27 | 24 | 24 | 21 | 24 | 77 | 78 | 78 | 78 | 81 | 71 | 77 | 72 | 70 | -- | | | | | | | | | | | | | |
| PDIR21 (w) | 26 | 26 | 25 | 25 | 25 | 24 | 25 | 26 | 24 | 23 | 21 | 24 | 77 | 78 | 78 | 77 | 81 | 71 | 76 | 72 | 70 | 99 | -- | | | | | | | | | | | | |
| PDIR17 (x) | 28 | 28 | 26 | 27 | 27 | 28 | 26 | 28 | 26 | 25 | 24 | 26 | 52 | 53 | 53 | 53 | 54 | 50 | 53 | 50 | 52 | 53 | 53 | -- | | | | | | | | | | | |
| PDIR23 (y) | 20 | 20 | 19 | 20 | 20 | 19 | 21 | 20 | 20 | 20 | 18 | 20 | 30 | 31 | 30 | 30 | 32 | 28 | 31 | 29 | 30 | 32 | 32 | 33 | -- | | | | | | | | | | |
| PDIR24 (z) | 20 | 20 | 19 | 20 | 20 | 19 | 21 | 20 | 20 | 20 | 18 | 20 | 30 | 31 | 30 | 30 | 32 | 28 | 31 | 29 | 30 | 32 | 32 | 33 | 99 | -- | | | | | | | | | |
| PDIR22 (aa) | 23 | 23 | 21 | 23 | 23 | 22 | 24 | 22 | 22 | 22 | 20 | 22 | 32 | 34 | 33 | 33 | 35 | 32 | 33 | 33 | 33 | 33 | 33 | 34 | 80 | 79 | -- | | | | | | | | |
| PDIR25 (ab) | 21 | 21 | 20 | 20 | 20 | 20 | 20 | 21 | 21 | 20 | 20 | 23 | 28 | 28 | 27 | 28 | 28 | 27 | 27 | 26 | 28 | 28 | 28 | 31 | 30 | 30 | 31 | -- | | | | | | | |
| PDIR26 (ac) | 20 | 20 | 18 | 19 | 19 | 19 | 19 | 20 | 21 | 20 | 18 | 20 | 29 | 29 | 28 | 28 | 30 | 29 | 29 | 28 | 30 | 29 | 28 | 32 | 29 | 29 | 29 | 70 | -- | | | | | | |
| PDIR27 (ad) | 22 | 22 | 20 | 20 | 20 | 21 | 21 | 22 | 22 | 21 | 19 | 22 | 31 | 31 | 30 | 31 | 33 | 31 | 31 | 30 | 31 | 30 | 29 | 31 | 28 | 28 | 29 | 66 | 61 | -- | | | | | |
| PDIR28 (ae) | 23 | 23 | 22 | 22 | 22 | 23 | 22 | 22 | 22 | 22 | 22 | 24 | 34 | 34 | 33 | 33 | 33 | 33 | 34 | 32 | 34 | 31 | 31 | 36 | 32 | 32 | 31 | 67 | 65 | 61 | -- | | | | |
| PDIR29 (af) | 23 | 23 | 21 | 22 | 22 | 23 | 22 | 22 | 22 | 22 | 21 | 24 | 34 | 34 | 32 | 33 | 32 | 32 | 33 | 31 | 34 | 31 | 30 | 35 | 33 | 33 | 32 | 67 | 65 | 61 | 97 | -- | | | |
| PDIR35 (ag) | 22 | 22 | 21 | 22 | 22 | 22 | 22 | 22 | 22 | 21 | 19 | 27 | 30 | 30 | 31 | 31 | 29 | 29 | 30 | 29 | 29 | 29 | 28 | 31 | 28 | 28 | 29 | 55 | 53 | 53 | 61 | 62 | -- | | |
| PDIR30 (ah) | 25 | 25 | 25 | 24 | 25 | 24 | 24 | 25 | 23 | 23 | 20 | 26 | 29 | 30 | 30 | 30 | 30 | 28 | 30 | 30 | 30 | 32 | 32 | 30 | 32 | 32 | 34 | 40 | 39 | 35 | 39 | 39 | 38 | -- | |
| PDIR31 (ai) | 25 | 25 | 25 | 24 | 25 | 24 | 24 | 25 | 23 | 23 | 20 | 26 | 30 | 30 | 30 | 30 | 31 | 29 | 31 | 30 | 31 | 32 | 32 | 30 | 32 | 32 | 34 | 40 | 39 | 35 | 39 | 39 | 38 | 99 | -- |

DIR-a

DIR-b

DIR-f

Fig. 2. Sequence relatedness of spruce DIR and DIR-like proteins. Results from pairwise amino acid sequence comparisons, using complete open reading frames, shown as percent identity among members of the DIR-a, DIR-b and DIR-f subfamilies. Within subfamily comparisons are highlighted yellow, whereas comparisons between proteins of the subfamilies DIR-a and DIR-b, DIR-a and DIR-f, and DIR-b and DIR-f are highlighted blue, green and purple, respectively.

### 2.2. Extended phylogeny of the plant DIR family

To validate our previous analysis of evolutionary relationships among plant DIR and DIR-like genes (Ralph et al., 2006b) with the inclusion of all 35 spruce DIR and DIR-like proteins, we performed a comprehensive search of GenBank for DIR genes and identified 115 putative FLcDNAs or genomic ORF sequences (Table 2). The two largest DIR families, in addition to the spruce family, were identified from *Arabidopsis thaliana* (25 genes, Arabidopsis) and *Oryza sativa* (54 genes, rice) for which complete genome sequences exist. Supporting FLcDNAs were identified for 35 of these Arabidopsis and rice genes. The final number of DIR genes in the rice genome may change as ORF prediction and annotation continues to be refined. Among other gymnosperms, a set of nine genes from western red cedar and two genes from *Tsuga heterophylla* (western hemlock) were identified, and among other angiosperms an additional 25 DIR genes were identified from a variety of species. As with our previous analysis, no DIR genes were identified outside the seed plant division.

We next performed maximum likelihood analysis using all 150 DIR and DIR-like proteins identified in this study to analyze the evolutionary relationships between spruce and other gymnosperm and angiosperm DIR and DIR-like genes. Multiple protein sequence alignments were performed using the Dialign software (Morgenstern et al., 1998), which is optimized to identify regions of local similarity among divergent sequences. Alignments were then adjusted to define a conserved sequence of *ca.* 160 amino acids (i.e., an N-terminus up to 20 amino acids upstream from the conserved "HD" in motif I, to a C-terminus up to 20 amino acids downstream from the conserved C-terminal "T" in motif V, see Fig. 1 and below for motif designations). A phylogenetic tree was generated using the neighbor-joining algorithm (Fig. 3), which shows the presence of six distinct subfamilies (i.e., DIR-a, DIR-b/d, DIR-c, DIR-e, DIR-f, and DIR-g) within the extended DIR and DIR-like family.

The overall sequence identity between subfamilies is very low, such that the inclusion of proteins within the DIR and DIR-like families is determined by manual examination of sequence alignments for the presence of conserved motifs common to DIR and DIR-like proteins, as described in earlier work (Ralph et al., 2006b). Here we further refine these motifs by again using 50% amino acid identity as a threshold, applied now to all 150 DIR proteins. The identified motifs are highlighted in Fig. 1 and include: motif I, LchYhHD beginning at amino acid 43 (all positions relative to PDIR1); motif II, FGslsVhDDslT-G beginning at amino acid 74; motif III, SshVGRAQGhY beginning at amino acid 92; motif IV ThVFpsGcasGSTLslhG beginning at amino acid 117; and motif V, REhuVVGGTGcFphARG-Aph+T beginning at amino acid 143; where $c$ = charged, $h$ = hydrophobic, $s$ = small, $l$ = aliphatic, negative sign = negative charge, $p$ = polar, $a$ = aromatic, $u$ = tiny, and positive sign = positive charge. In search of critical residues potentially involved in the catalytic activity or regulation of DIR proteins, we also selected for amino acids conserved in 95% or more of the 150 DIR proteins. Interestingly, we identified six highly conserved glycine residues, contained within motifs II, III and V, and no other highly conserved amino acids. These residues represent important targets for future mutational analysis.

Twelve of the 35 spruce DIR and DIR-like proteins group into subfamily DIR-a (Fig. 3), including the new members PDIR32, PDIR33 and PDIR34, along with all other known gymnosperm DIR sequences and representatives from both monocot and dicot angiosperm species. Of

Table 2
Gymnosperm and angiosperm DIR and DIR-like genes represented in phylogenetic analyses

| Species | DIR nomenclature | FLcDNA Accession No. | Genomic ORF accession/AGI no. |
|---|---|---|---|
| Gymnosperms | | | |
| *Picea glauca* (white spruce) | *PDIR1* | ABD52112 | n.a. |
| | *PDIR7* | ABD52118 | n.a. |
| | *PDIR10* | ABD52121 | n.a. |
| | *PDIR12* | ABD52123 | n.a. |
| | *PDIR18* | ABD52129 | n.a. |
| *Picea sitchensis* (Sitka spruce) | *PDIR11* | ABD52122 | n.a. |
| | *PDIR15* | ABD52126 | n.a. |
| | *PDIR16* | ABD52127 | n.a. |
| | *PDIR17* | ABD52128 | n.a. |
| | *PDIR19* | ABD52130 | n.a. |
| | *PDIR20* | EF643700 | n.a. |
| | *PDIR21* | EF643701 | n.a. |
| | *PDIR22* | EF643702 | n.a. |
| | *PDIR23* | EF643703 | n.a. |
| | *PDIR24* | EF643704 | n.a. |
| | *PDIR25* | EF643705 | n.a. |
| | *PDIR26* | EF643706 | n.a. |
| | *PDIR27* | EF643707 | n.a. |
| | *PDIR28* | EF643708 | n.a. |
| | *PDIR29* | EF643709 | n.a. |
| | *PDIR30* | EF643710 | n.a. |
| | *PDIR31* | EF643711 | n.a. |
| | *PDIR32* | EF643712 | n.a. |
| | *PDIR33* | EF643713 | n.a. |
| | *PDIR34* | EF643714 | n.a. |
| | *PDIR35* | EF643715 | n.a. |
| *Picea glauca* × *engelmannii* (interior spruce) | *PDIR2* | ABD52113 | n.a. |
| | *PDIR3* | ABD52114 | n.a. |
| | *PDIR4* | ABD52115 | n.a. |
| | *PDIR5* | ABD52116 | n.a. |
| | *PDIR6* | ABD52117 | n.a. |
| | *PDIR8* | ABD52119 | n.a. |
| | *PDIR9* | ABD52120 | n.a. |
| | *PDIR13* | ABD52124 | n.a. |
| | *PDIR14* | ABD52125 | n.a. |
| *Tsuga heterophylla* (western hemlock) | *ThDIR1* | AAF25367 | n.a. |
| | *ThDIR2* | AAF25368 | n.a. |
| *Thuja plicata* (western red cedar) | *TpDIR1* | AAF25359 | n.a. |
| | *TpDIR2* | AAF25360 | n.a. |
| | *TpDIR3* | AAF25361 | n.a. |
| | *TpDIR4* | AAF25362 | n.a. |
| | *TpDIR5* | AAF25363 | n.a. |
| | *TpDIR6* | AAF25364 | n.a. |
| | *TpDIR7* | AAF25365 | n.a. |
| | *TpDIR8* | AAF25366 | n.a. |
| | *TpDIR9* | AAL92120 | n.a. |
| Angiosperms | | | |
| *Arabidopsis thaliana* (thale cress) | *AtDIR1* | n.a. | At5g42510 |
| | *AtDIR2* | AY093095 | At5g42500 |
| | *AtDIR3* | n.a. | At5g49040 |
| | *AtDIR4* | n.a. | At2g21110 |
| | *AtDIR5* | AK175255 | At1g64160 |
| | *AtDIR6* | BT002439 | At4g23690 |
| | *AtDIR7* | AK118030 | At3g13650 |
| | *AtDIR8* | n.a. | At3g13662 |
| | *AtDIR9* | BT010722 | At2g39430 |
| | *AtDIR10* | BT002889 | At2g28670 |
| | *AtDIR11* | AK176442 | At1g22900 |
| | *AtDIR12* | BT004016 | At4g11180 |
| | *AtDIR13* | BT009718 | At4g11190 |
| | *AtDIR14* | n.a. | At4g11210 |
| | *AtDIR15* | CNS0A3JZ | At4g38700 |

Table 2 (*continued*)

| Species | DIR nomenclature | FLcDNA Accession No. | Genomic ORF accession/AGI no. |
|---|---|---|---|
| | *AtDIR16* | BT008336 | At3g24020 |
| | *AtDIR17* | AK117592 | At3g58090 (CAB67637) |
| | *AtDIR18* | AY081267 | At4g13580 |
| | *AtDIR19* | AK117899 | At1g58170 |
| | *AtDIR20* | AY128336 | At1g55210 |
| | *AtDIR21* | n.a. | At1g65870 |
| | *AtDIR22* | BT015420 | At3g13660 |
| | *AtDIR23* | BT005788 | At2g21100 |
| | *AtDIR24* | CNS0A5KQ | At3g55230 |
| | *AtDIR25* | n.a. | At1g07730 |
| *Oryza sativa* (Japonica) (rice) | *OsDIR1* | AK109288 | Os07g0107300 |
| | *OsDIR2* | n.a. | Os07g0107500 |
| | *OsDIR3* | AK062751 | Os07g0107100 |
| | *OsDIR4* | AK108186 | Os08g0375400 |
| | *OsDIR5* | AK108922 | Os10g0398100 |
| | *OsDIR6* | n.a. | Os07g0617500 |
| | *OsDIR7* | n.a. | Os07g0617100 |
| | *OsDIR8* | n.a. | BAB89617 |
| | *OsDIR9* | AK108101 | Os07g0106900 |
| | *OsDIR10* | n.a. | Os03g0400200 |
| | *OsDIR11* | AK106022 | Os07g0636600 |
| | *OsDIR12* | n.a. | Os07g0636800 |
| | *OsDIR13* | n.a. | Os07g0637700 |
| | *OsDIR14* | n.a. | BAC19943 |
| | *OsDIR15* | AK108983 | Os07g0638500 |
| | OsDIR16 | n.a. | BAC16397 |
| | *OsDIR17* | n.a. | AAM74352 |
| | *OsDIR18* | n.a. | BAD25846 |
| | *OsDIR19* | AK121408 | Os03g0809000 |
| | *OsDIR20* | n.a. | AAO17346 |
| | *OsDIR21* | n.a. | BAB64642 |
| | *OsDIR22* | n.a. | BAD52647 |
| | *OsDIR23* | n.a. | Os10g0398600 |
| | *OsDIR24* | n.a. | BAD53304 |
| | *OsDIR25* | n.a. | AAM74358 |
| | *OsDIR26* | n.a. | AAM74346 |
| | *OsDIR27* | n.a. | BAD03849 |
| | *OsDIR28* | n.a. | BAD03720 |
| | *OsDIR29* | n.a. | BAD03711 |
| | *OsDIR30* | n.a. | BAD03854 |
| | *OsDIR31* | n.a. | Os12g0174700 |
| | *OsDIR32* | AK060077 | Os11g0180000 |
| | *OsDIR33* | AK063148 | Os11g0178800 |
| | *OsDIR34* | n.a. | AAX96293 |
| | *OsDIR35* | n.a. | Os11g0179100 |
| | *OsDIR36* | AK062657 | Os11g0179000 |
| | *OsDIR37* | AK101991 | Os12g0198700 |
| | *OsDIR38* | AK066682 | Os12g0247700 |
| | *OsDIR39* | n.a. | Os12g0228700 |
| | *OsDIR40* | n.a. | Os12g0199000 |
| | *OsDIR41* | n.a. | Os12g0227500 |
| | *OsDIR42* | n.a. | ABA94701 |
| | *OsDIR43* | AK119964 | Os11g0644700 |
| | *OsDIR44* | n.a. | BAB89759 |
| | *OsDIR45* | n.a. | Os11g0645400 |
| | *OsDIR46* | AK107902 | Os04g0666800 |
| | *OsDIR47* | AK111039 | Os11g0214400 |
| | *OsDIR48* | n.a. | Os11g0214900 |
| | *OsDIR49* | AK107711 | Os11g0215100 |
| | *OsDIR50* | n.a. | Os08g0349100 |
| | *OsDIR51* | n.a. | BAD89460 |
| | *OsDIR52* | n.a. | AAX96290 |
| | *OsDIR53* | n.a. | ABA93522 |
| | *OsDIR54* | n.a. | AAX96314 |

(*continued on next page*)

Table 2 (*continued*)

| Species | DIR nomenclature | FLcDNA Accession No. | Genomic ORF accession/AGI no. |
|---|---|---|---|
| *Hordeum vulgare* (barley) | *HvDIR1* | AAA87042 | n.a. |
| | *HvDIR2* | AAA87041 | n.a. |
| | *HvDIR3* | AAB72098 | n.a. |
| *Triticum aestivum* (wheat) | *TaDIR1* | AAC49284 | n.a. |
| | *TaDIR2* | AAM46813 | n.a. |
| | *TaDIR3* | BAA32786 | n.a. |
| | *TaDIR4* | AAR20919 | n.a. |
| *Saccharum officinarum* (sugarcane) | *SoDIR1* | AAR00251 | n.a. |
| | *SoDIR2* | CAF25234 | n.a. |
| | *SoDIR3* | AAV50047 | n.a. |
| *Pisum sativum* (pea) | *PsDIR1* | AAD25355 | n.a. |
| | *PsDIR2* | AAB18669 | n.a. |
| *Forsythia intermedia* (shrub) | *FiDIR1* | AAF25357 | n.a. |
| | *FiDIR2* | AAF25358 | n.a. |
| *Podophyllum peltatum* (mayapple) | *PpDIR1* | AAK38666 | n.a. |
| *Sorghum bicolor* (sorghum) | *SbDIR1* | AAM94289 | n.a. |
| | *SbDIR2* | ABI24164 | n.a. |
| *Gossypium barbadense* (cotton) | *GbDIR1* | AAS73001 | n.a. |
| | *GbDIR2* | AAY44415 | n.a. |
| *Zea mays* (corn) | *ZmDIR1* | AAF71261 | n.a. |
| *Sesamum indicum* (sesame) | *SiDIR1* | AAT11124 | n.a. |
| *Tamarix androssowii* (Tamarix) | *TanDIR1* | ABE73781 | n.a. |
| *Nicotiana benthamiana* (tobacco) | *NbDIR1* | BAF02555 | n.a. |
| *Arachis hypogaea* (peanut) | *AhDIR1* | AAZ20288 | n.a. |
| *Agrostis stolonifera* (bentgrass) | *AsDIR1* | AAY41607 | n.a. |

the six subfamilies identified in this phylogenetic analysis, DIR-a shows the highest degree of sequence conservation among representatives. To date, the ability to direct *in vitro* the stereoselective formation of lignans has only been demonstrated for members of the DIR-a family (Davin et al., 1997; Xia et al., 2000; Kim et al., 2002). Since biochemical functions have only been assigned to members of the DIR-a subfamily, we continue to designate proteins clustering within subfamilies DIR-b/d, DIR-c, DIR-e, DIR-f and DIR-g as DIR-like as suggested previously (Ralph et al., 2006b).

Twelve of the spruce DIR-like proteins group into subfamily DIR-b/d, including new members PDIR20 and PDIR21 (Fig. 3). No other DIR-like FLcDNA sequences from gymnosperm species were identified in public databases belonging to DIR-b/d. With the addition of a large number of new DIR and DIR-like sequences from the completed rice genome in the current phylogenetic analysis, we observed a change of topology of the DIR-d subfamily from our previous phylogenetic analysis (Ralph et al., 2006b). Several sequences from the previously distinct DIR-d cluster (i.e., AtDIR2, AtDIR11 and GbDIR1) have merged with the former DIR-b subfamily to form the new DIR-b/d subfamily, while rice DIR-like proteins from the former DIR-d subfamily (i.e., OsDIR1, OsDIR3, OsDIR4 and OsDIR9) now group as a separate subfamily in the current analysis that is designated DIR-g and consists solely of monocot species (i.e., rice and *Sorghum bicolor*). Overall sequence conservation among proteins in the DIR-b/d and DIR-g subfamilies is relatively low.

The remaining 11 spruce DIR-like proteins form a new tentative cluster, DIR-f, consisting exclusively of spruce proteins identified in this study (i.e., PDIR22-PDIR31, PDIR35; Fig. 3). The spruce DIR-like proteins in this cluster are considerably more divergent in sequence than among spruce proteins in the DIR-a or DIR-b/d clusters. Other than the sequence conservation observed in motifs I-V (Fig. 1), the overall sequence similarity between spruce DIR-f and spruce DIR-a or DIR-b/d proteins is relatively low. In addition, no angiosperm DIR-like proteins were observed to cluster with spruce DIR-f proteins, unlike subfamilies DIR-a and DIR-b/d where both angiosperms and gymnosperms cluster together. This suggests that the DIR-f clade may have evolved after the separation of angiosperms and gymnosperms, whereas the presence of DIR and DIR-like proteins from both plant divisions in the DIR-a and DIR-b/d clades suggests they share a common DIR ancestor.

As observed in our previous phylogenetic analysis (Ralph et al., 2006b), the DIR-c clade consists exclusively of angiosperm monocot DIR-like proteins (Fig. 3). Sequence conservation within the family is very low, and many members posses a unique *ca.* 140 amino acid extension at their C-terminus with high similarity to the jacalin-like domain common to lectin proteins that may contribute to defense against insects (Blanchard et al., 2001). Finally, the DIR-e subfamily consists of representatives from Arabidopsis and rice that are typically much larger than DIR and DIR-like proteins in other subfamilies, and possess an additional 150–200 amino acids at the N-terminus of unknown function (Fig. 3).

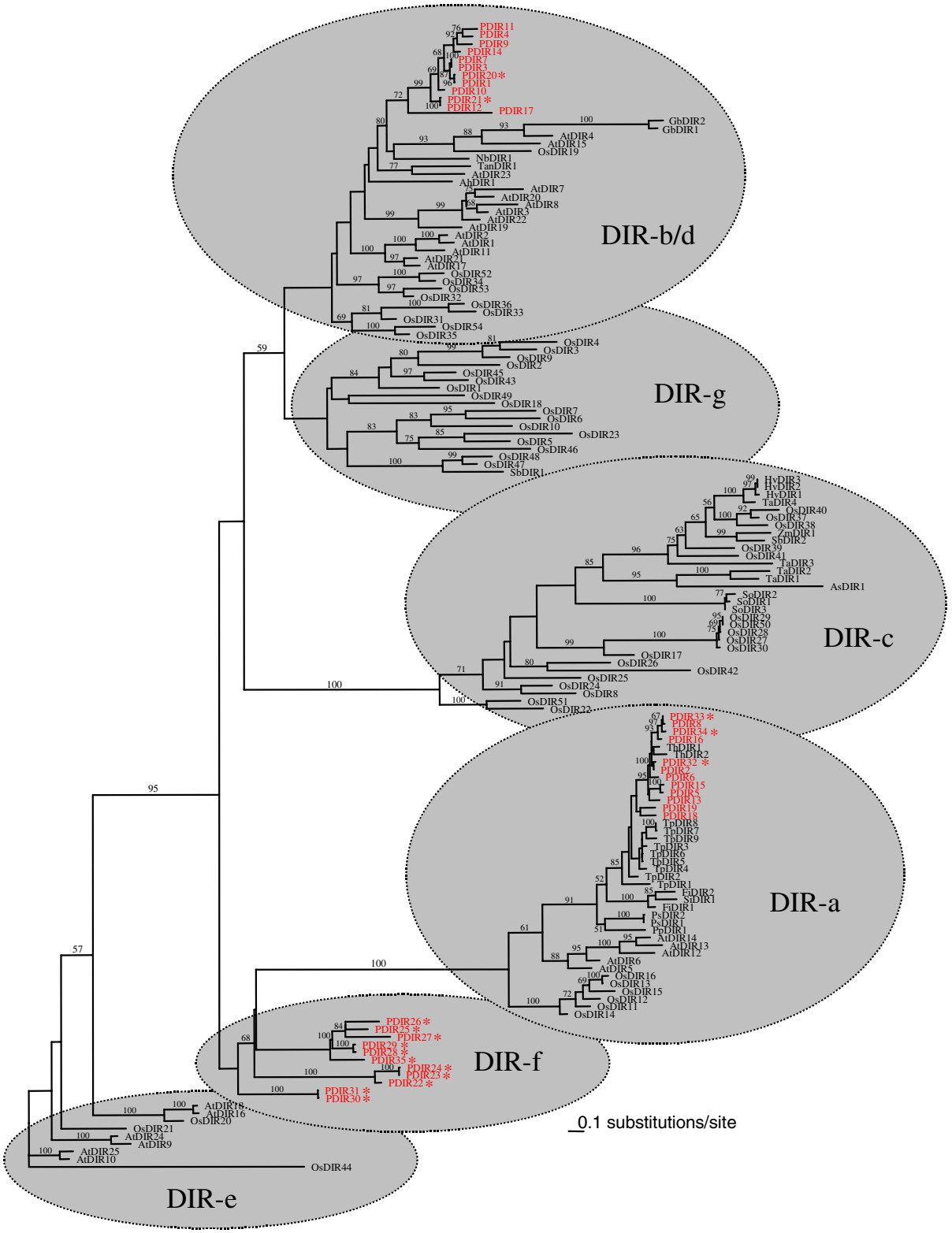


Fig. 3. Phylogenetic tree of plant DIR and DIR-like protein sequences. Amino acids of 150 DIR or DIR-like proteins were analyzed by maximum likelihood using Phyml. Bootstrap values are only provided at nodes with greater than 50% support. Maximum likelihood values represent percentages of 75 gamma-corrected replicates ($\log L = -27{,}775$). Subfamilies DIR-a, DIR-b/d, DIR-c, DIR-e, DIR-f and DIR-g are indicated by gray shading. Spruce DIR and DIR-like proteins are highlighted red, with new sequences identified in this study marked with an asterisk. DIR nomenclature is as follows: P, *Picea glauca*, *Picea sitchensis* or *P. glauca x engelmannii*; Th, *Tsuga heterophylla;* Tp, *Thuja plicata;* At, *Arabidopsis thaliana*; Os, *Oryza sativa;* Hv, *Hordeum vulgare*; Ta, *Triticum aestivum*; So, *Saccharum officinarum*; Ps, *Pisum sativum*; Fi, *Forsythia intermedia*; Pp, *Podophyllum peltatum*; Sb, *Sorghum bicolor*; Gb, *Gossypium barbadense*; Zm, *Zea mays*; Si, *Sesamum indicum*; Tan, *Tamarix androssowii*; Nb, *Nicotiana benthamiana*; Ah, *Arachis hypogaea*; and As, *Agrostis stolonifera*. For additional details see Table 2.

### 2.3. Constitutive expression profiles of the spruce DIR family assessed by real-time PCR

To extend our earlier analysis of gene expression patterns for select members of the DIR-a and DIR-b subfamilies (Ralph et al., 2006b), we designed 22 new gene-specific primer pairs based on nucleotide alignment of the 32 FLcDNA sequences (Table 3). These primer pairs represent 31 DIR and DIR-like transcripts, with primers designed to co-amplify transcripts in the case of nine pairs of closely related isoforms (note: despite several attempts, gene-specific primers could not be designed for *PDIR34*). The existence of three distinct DIR and DIR-like subfamilies, each containing as many as a dozen genes of relatively high sequence divergence, may suggest different roles of individual genes in constitutive or induced processes. In an attempt to illustrate possible differences in spatial patterns of RNA expression, the relative constitutive abundance of 22 different spruce DIR and DIR-like transcripts was quantified using real-time PCR in total RNA isolated from different stem tissues (cortex, phloem, cambium and xylem), young lateral shoots tips, and root tissues from Sitka spruce (Fig. 4a and b). We also measured DIR and DIR-like expression in developing somatic embryos of white spruce. Real-time PCR expression data were normalized to housekeeping gene control levels [i.e., eukaryotic translation initiation factor (TIF)5A]. For ease of comparison, constitutive expression data is presented on a standardized log scale.

Amongst all DIR and DIR-like genes, the three transcripts of highest abundance in each tissue were as follows: shoots (*PDIR3/7*, *PDIR1/20*, *PDIR12/21*), roots (*PDIR12/21*, *PDIR3/7*, *PDIR17*), cortex (*PDIR3/7*, *PDIR12/21*, *PDIR28/29*), phloem (*PDIR3/7*, *PDIR12/21*, *PDIR23/24*), cambium (*PDIR3/7*, *PDIR23/24*, *PDIR22*), xylem (*PDIR23/24*, *PDIR3/7*, *PDIR22*), and embryo (*PDIR3/7*, *PDIR12/21*, *PDIR10*) (Fig. 4a and b). Of the seven genes examined from the DIR-a subfamily, four (*PDIR2/32*, *PDIR6*, *PDIR8/33*, *PDIR16*) displayed a pattern of highest abundance in roots, with moderate levels also detected in shoots and cortex, and significantly lower abundance in the inner stem tissues phloem, cambium and xylem (Fig. 4a). Expression in embryos was low for *PDIR2/32* and *PDIR6*, with no detectable transcripts for *PDIR8/33* and *PDIR16*. The gradient of expression amongst the stem tissues [i.e., highest in outermost tissues (cortex) to lowest in innermost tissues (xylem)] is consistent with previous observations for *PDIR2*, *PDIR6* and *PDIR8* (Ralph et al., 2006b). Two other DIR-a family members, *PDIR5/15* and *PDIR13*, were expressed at very low levels relative to *TIF5A*, with highest abundance in roots and no detectable expression in phloem, cambium or xylem (Fig. 4a). Interestingly, *PDIR18* had a distinct expression profile among DIR-a genes, with ubiquitous expression that was highest in xylem and cambium.

Constitutive expression patterns of DIR-b/d subfamily members were notably divergent among the eight transcripts examined. Several genes showed ubiquitous expression at very high abundance (i.e., *PDIR1/20*, *PDIR3/7*, *PDIR10*, *PDIR12/21*), while *PDIR9* was ubiquitously expressed at moderate levels (Fig. 4a and 4b). This pattern was previously observed for *PDIR1*, *PDIR10* and *PDIR12* (Ralph et al., 2006b), and is suggestive of a possible role for these genes in a primary process or constitutive defense. In contrast, *PDIR14* was only marginally expressed in most tissues, with no detectable transcript in phloem and xylem,

Table 3
Primer sequences used for real-time PCR (5′ to 3′ orientation) and estimated PCR efficiencies

| Gene ID | Forward primer | Reverse primer | PCR efficiency (%) |
|---|---|---|---|
| *TIF5A* | GTGCCATCTTCACACAACTGC | CAGATTCAGTCAGCAGGCTAAC | 94.3 |
| *PDIR1/20* | TTCGATCAAATGGCAGTGTACTC | ACCATGACAAATATGCACAGAAG | 96.3 |
| *PDIR2/32* | AGTCTACTGATCTCTCCTGCTG | ATTATTTCACTCTACGCTAGCTG | 96.3 |
| *PDIR3/7* | ATTATGCGCATTCTGTAGTGATG | ACAGAAGACACGGTCACTAGAC | 96.4 |
| *PDIR5/15* | TGCATGAAGCTCCGAAATGCTG | TATATTGGATATGGCCTTCCAAC | 95.4 |
| *PDIR6* | CTACTGATCTTTCATGCATGTTC | TTCACTCTACCAATGCAGATTTG | 93.4 |
| *PDIR8/33* | TCCCTAATGTTATAGCGTGCAG | TCTCCATTGAGCATACATATTGG | 92.7 |
| *PDIR9* | GTCTTGATTGTATCATGGCTCTC | TGTCTCTGCAATGAAGCTTGAG | 93.2 |
| *PDIR10* | CCCGGCAGTGGTATATACATTG | GCAAGCAGCATTGGATACTGTC | 96.0 |
| *PDIR11* | TCTCGCTCGCACTCATTCCATC | CATTACGTAAAGAGTGTAGAAGG | 89.0 |
| *PDIR12/21* | TGCCGTTCCATCCATCTATCTC | GGTCATGATTAGACTGCCCTTG | 95.3 |
| *PDIR13* | AGTCTAGTGATCTCTCCTCCTG | TTTACTTGACGCTTACACCTTTG | 95.6 |
| *PDIR14* | TCATGTTGCGATGGAATATCAAC | TTTACCATACATCTCTCCTGGTC | 90.4 |
| *PDIR16* | GCTCTCCATTTCTATGTTCATTG | TTTGGGAGATTCATGCATATCTC | 83.5 |
| *PDIR17* | CCAGAGTCTTCATTCTGGGATC | ACATAGAGAGATCAATACACCAC | 89.3 |
| *PDIR18* | GAAATCTGTGCTAGCTTGACTTG | AGGAACCGCATTCTAAATTCGTG | 94.3 |
| *PDIR22* | TTATTCAGTGCTCAGAGTATGAC | CATATAATACTTGCCTGTCTGTG | 91.4 |
| *PDIR23/24* | ACAAAGCAGAGGGAGATATCAG | TTTGCTCCCATTGACAACTGTG | 95.7 |
| *PDIR25* | TGGCTATGCGGTTGTTACTGTG | ACAAGGTAACCTCTCCACAACG | 96.1 |
| *PDIR26* | CAAGAATGCTGTTCCCAGATTC | TAAGCCATACAAACGGAAACAAG | 94.5 |
| *PDIR27* | TTCACATCCCGTGCCTTATAAAG | TCAACGACACCACTAGGCAAAC | 95.5 |
| *PDIR28/29* | CTAGCACAGAAGCTCTTTCTAG | AAACTAACCGTCTGATAGGAACC | 95.1 |
| *PDIR30/31* | AGCATGTGTAGCCAAGATTTGAC | AGTTACACCATCTACCTATGCTG | 92.8 |

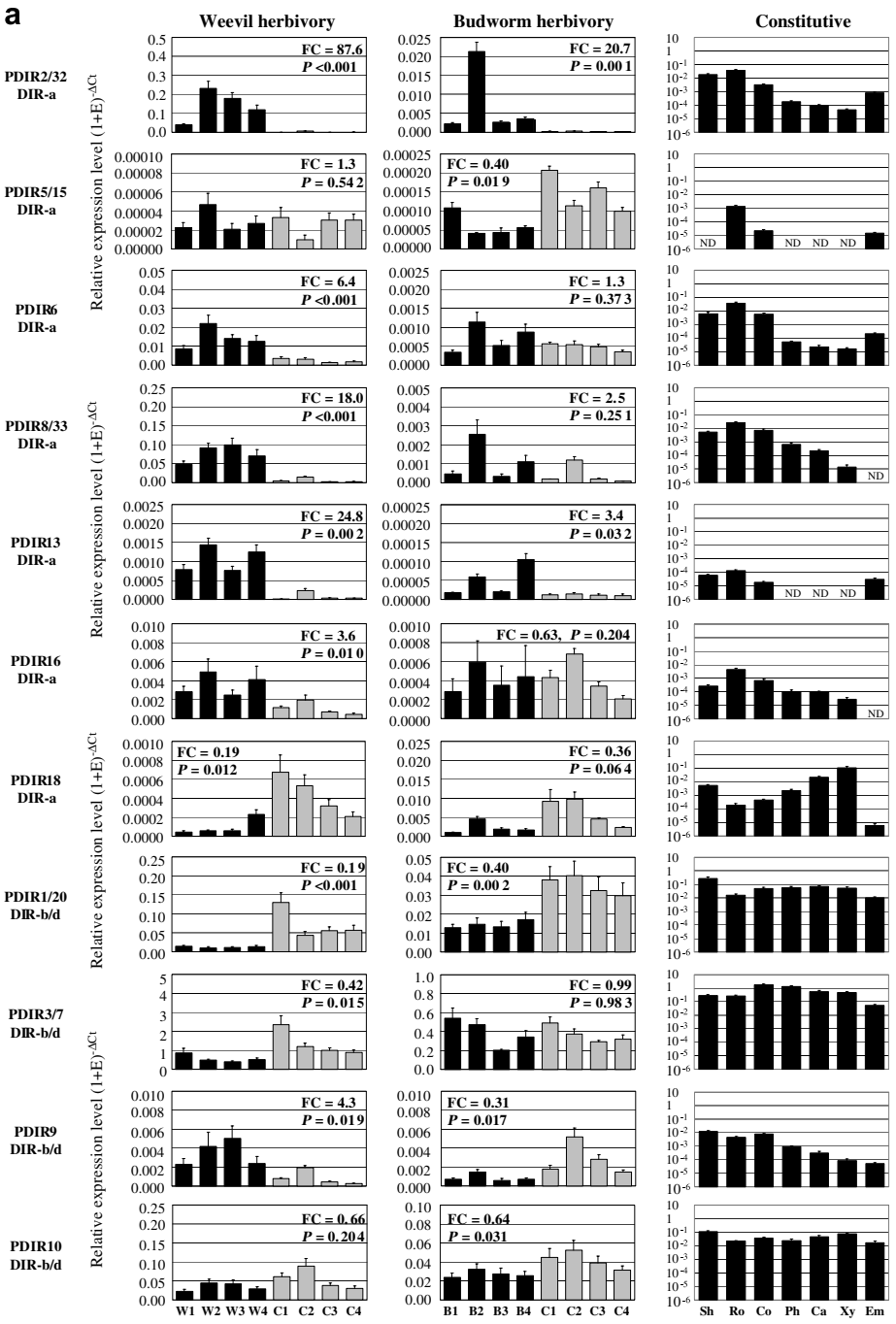


Fig. 4. Quantitative real-time PCR analysis of constitutive and insect-induced DIR and DIR-like transcript abundance in spruce tissues. Relative abundance of mRNA transcripts of 22 distinct DIR and DIR-like genes in bark tissues of individual Sitka spruce trees subjected to weevil herbivory (W) or left untreated as a control (C, left panel), and green leader tissues subjected to budworm herbivory (B) or left untreated as a control (center panel), 48 h after the onset of treatment. Analyses of insect-induced transcripts were performed with four independent biological replicates (W1–W4, C1–C4; B1–B4, C1–C4). Relative abundance of the same 22 transcripts in tissues collected from 3-year-old Sitka spruce trees harvested in May, as well as developing somatic embryos from white spruce is shown in the right panel. Tissues examined include lateral shoot tips (Sh), roots (Ro), cortex (Co), phloem (Ph), cambium (Ca), xylem (Xy) and embryos (Em). Values obtained by real-time PCR represent mean ± standard error of the mean [SEM; $n$ = three or more independent technical replicates, each consisting of RNA from individual trees (insect treatments) or pooled RNA from 10 biological replicates (constitutive expression)] normalized to eukaryotic translation initiation factor (*TIF5A*; EST IS0013_F24) expression in each tissue by subtracting the Ct value of each transcript, where $\Delta Ct = Ct_{transcript} - Ct_{TIF5A}$. Transcript abundance was obtained from the equation $(1 + E)^{-\Delta Ct}$, where $E$ is the PCR efficiency, as described by Ramakers et al. (2003). A transcript with a relative abundance of one is equivalent to the abundance of *TIF5A* in the same tissue. Statistical significance of expression differences was determined using a mixed-effect model (see Section 4). Abbreviations: ND, not detected; FC, fold-change.

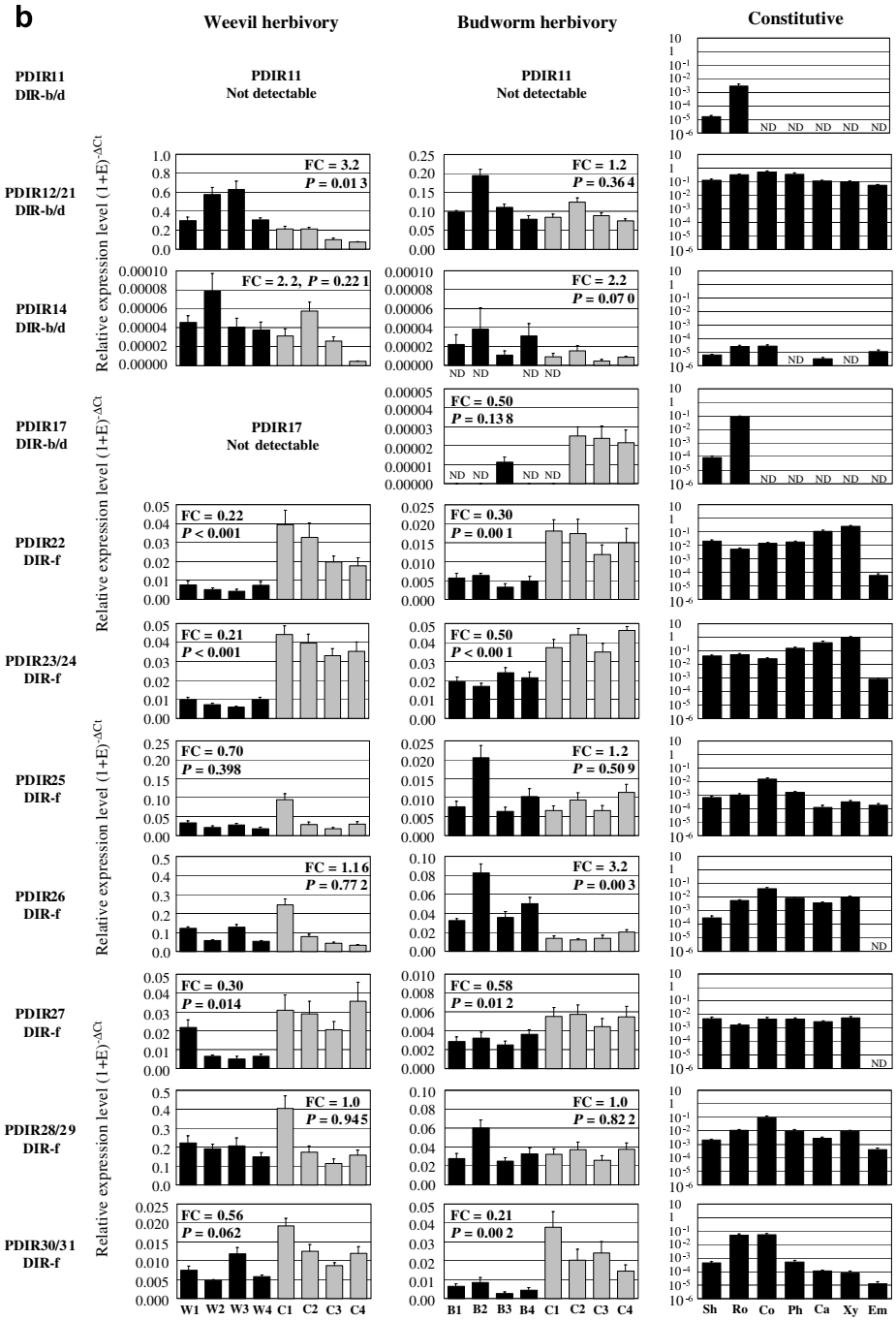


Fig. 4 (*continued*)

and *PDIR11* and *PDIR17* were predominantly expressed in root tissue, with low expression in shoots and no detectable transcript in other tissues (Fig. 4b). This finding of unique as well as partially overlapping patterns of expression among DIR-b/d subfamily members supports the possibility of differential regulation and tissue-specific roles.

Finally, examination of constitutive expression for the seven DIR-f subfamily members revealed ubiquitous expression at moderate to high abundance for all genes, with the exception of no detectable *PDIR26* and *PDIR27* transcripts in embryos (Fig. 4b). Several transcripts showed highest expression in cortex (i.e., *PDIR25*, *PDIR26*, *PDIR28/29*, *PDIR30/31*), while *PDIR22* and *PDIR23/24* were both most highly expressed in xylem and cambium tissues. In the absence of any information on biological function, we can only speculate of a possible role for these genes in a primary process or constitutive defense.

### *2.4. Expression profiles of the spruce DIR family in response to weevil and budworm herbivory assessed by real-time PCR*

Having established spatial patterns of expression for 22 DIR and DIR-like transcripts in spruce, we next examined these genes for a possible role in the defense response of

Sitka spruce against stem-boring weevils and defoliating budworms. In brief, 2-year-old Sitka spruce saplings were subjected on their stems or green apical shoots to feeding by adult white pine weevils or budworm larvae, respectively. Transcript profiles were monitored using real-time PCR 48 h after the onset of treatment and compared to untreated control tissues. Data are presented as transcript abundance normalized to the housekeeping gene *TIF5A*, with fold change relative to untreated control tissues indicated in each panel along with a measure of statistical significance ($P$ value from a mixed-effects model; Fig. 4a and b).

In bark tissue, 20 of 22 transcripts examined were detected, with *PDIR11* and *PDIR17* not present, consistent with the root- and shoot-specific patterns of constitutive expression for these genes (Fig. 4b). Amongst the DIR and DIR-like transcripts quantified in bark, a surprisingly high proportion (13 of 20) showed statistically significant changes in abundance in response to insect attack ($FC > 2$; $P < 0.05$). Of the seven transcripts induced following weevil feeding, increases ranged from 87.6-fold for *PDIR2/32* of DIR-a to 3.2-fold for *PDIR12/21* of DIR-b/d (Fig. 4a and b). The largest increases were observed among DIR-a subfamily members including 87.6-fold for *PDIR2/32*, 24.8-fold for *PDIR13*, 18.0-fold for *PDIR8/33* and 6.4-fold for *PDIR6*. These results are consistent with previous observations of strong induction of *PDIR2*, *PDIR6*, *PDIR8* and *PDIR13* in bark and xylem tissues 2–48 h following weevil feeding or mechanical wounding (Ralph et al., 2006b). Six transcripts were reduced in abundance in response to weevil feeding, with the largest changes observed for *PDIR18* of DIR-a (0.19-fold), *PDIR1/20* of DIR-b/d (0.19-fold), *PDIR23/24* of DIR-f (0.21-fold), and *PDIR22* of DIR-f (0.22-fold) (Fig. 4a and 4b). It is notable that all six DIR and DIR-like transcripts that were repressed following weevil feeding were also ubiquitously expressed at generally high levels in constitutive tissues.

In green apical shoots, we observed 21 of 22 DIR and DIR-like transcripts to be expressed (*PDIR11* was not detected), with nine genes showing differential expression following budworm feeding (Fig. 4a and b). Genes induced by budworm attack included *PDIR2/32* and *PDIR13* of DIR-a (20.7-fold and 3.4-fold, respectively), and *PDIR26* of DIR-f (3.2-fold), whereas repressed genes included *PDIR30/31*, *PDIR22* and *PDIR23/24* of DIR-f (0.21-fold, 0.30-fold and 0.50-fold, respectively), *PDIR5/15* of DIR-a (0.40-fold), and *PDIR9* and *PDIR1/20* of DIR-b/d (0.31-fold and 0.40-fold, respectively) (Fig. 4a and b).

### 2.5. Meta-analysis of expression profiles of the spruce DIR family assessed using a 16.7K cDNA microarray

The recent development and application of cDNA microarray platforms for spruce (Ralph et al., 2006a) now allows for the first meta-analysis of large-scale gene expression profiles generated from multiple experiments utilizing spruce. In an attempt to associate members of the spruce DIR and DIR-like families with biological processes, we conducted a meta-analysis using transcript expression data generated with the spruce 16.7K cDNA microarray, which contains 30 array elements annotated as DIR or DIR-like proteins. Expression data for these 30 elements was extracted from a series of experiments involving: (a) wounding and stem-boring insect herbivory treatments of bark tissue over time; (b) methyl jasmonate (MeJA) treatment of bark; (c) wounding alone, defoliating insect herbivory, and wounding plus insect regurgitant treatments of green apical leaders over time; the response in secondary xylem to compression (d) and opposite (e) wood formation over time; and (f) gene expression along a developmental gradient in apical leaders (Fig. 5; complete spruce DIR family microarray dataset provided in Supplemental Table I). When the expression data is clustered, the majority of array elements group according to their phylogenetic relationships. For example, all but one array element (WS0061_D02) from the DIR-a subfamily cluster together. These DIR-a genes show a pattern of strong induction in bark tissue following mechanical wounding or adult weevil herbivory; weak induction in bark following MeJA application; moderate induction in green apical leaders following wounding alone, herbivory by budworm larvae, or wounding plus the application of regurgitant from budworms; moderate down-regulation in secondary xylem from the compression and opposite sides of sapling stems following forced bending stress; and preferential expression in secondary xylem at the base of apical leaders compared to young, green bark at the apex of the leader.

Of the 11 DIR-b/d subfamily members on the spruce 16.7K array, all but one (i.e., WS0109_P06) group together with similar patterns of gene expression. In contrast to DIR-a genes, members of the DIR-b/d subfamily are only induced weakly, if at all, following stress treatments in bark or leader tissue, are weakly down-regulated in secondary xylem from compression and opposite sides of bent trees, and show moderate preferential expression in green apical shoots compared to secondary xylem at the leader base (Fig. 5). The divergent pattern of gene expression between genes of the DIR-a and DIR-b/d subfamilies reflects the considerable distance between these subfamilies in the phylogenetic tree (Fig. 3), and possibly different functions in different biological processes, where DIR-a genes are preferred candidates for future functional analysis in conifer defense.

Expression patterns for DIR-f genes are grouped into two clusters consisting of array elements representing *PDIR22* and *PDIR23* (Fig. 5), which cluster adjacent to each other in the phylogenetic tree (Fig. 3), and a set of nine elements with similarity to *PDIR25–PDIR29* plus *PDIR35*, which also form a distinct sub-clade within the larger DIR-f clade. One additional element of the DIR-f subfamily (i.e., WS01012_K12) has an expression pattern somewhat distinct from the other two clusters (Fig. 5), and also forms a separate branch of subfamily DIR-f of the phylogenetic tree (Fig. 3). Array elements representing the *PDIR25–PDIR29* plus *PDIR35* group show strong induction following MeJA

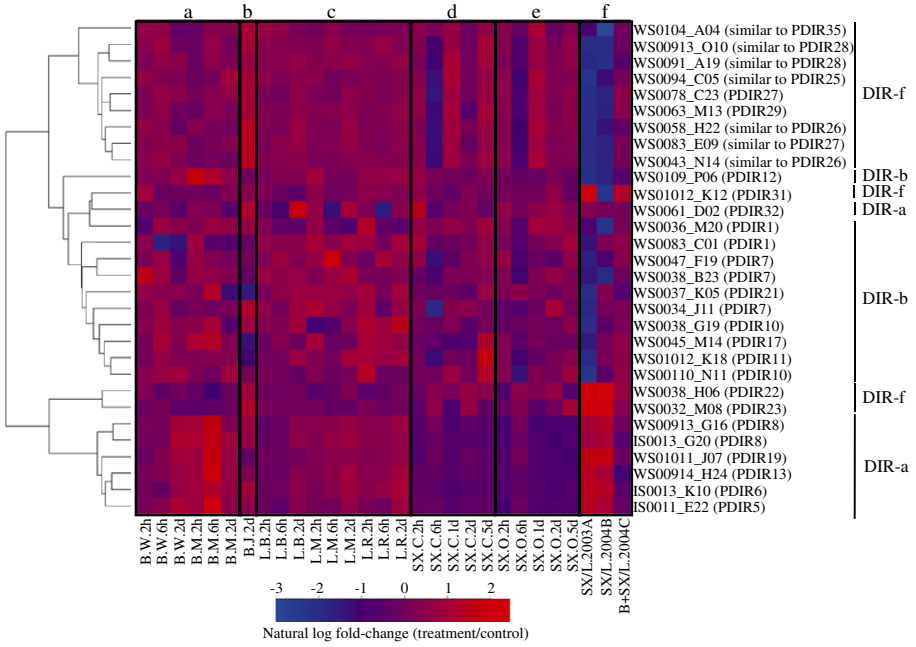


Fig. 5. Meta-analysis of DIR and DIR-like transcript expression data obtained with a spruce 16.7K cDNA microarray. Expression patterns for 30 expressed sequence tags (ESTs) representing at least 22 distinct DIR genes are presented as a heat map of fold-change expression relative to appropriate control tissue. A dendogram generated using the DIANA algorithm showing the relatedness of expression patterns is indicated to the left of the heat map. A color bar indicates fold-change expression differences on a natural log scale. An experiment code indicating the tissue, treatment, and time point is provided at the bottom of each column. Clone identifiers for each EST along with the closest matching full-length DIR protein are indicated to the right of each row. Experiments analyzed include: (a) bark tissue (B) of Sitka spruce (*Picea sitchensis*) saplings harvested 2 h, 6 h or 2 d following feeding by stem-boring weevils (W; *Pissodes strobi*) or mechanical wounding (M) by needle puncture, compared to bark from untreated control trees; (b) bark tissue of white spruce saplings (*P. glauca*) harvested 2 d following spraying with 150 mL of 0.01% (v/v) methyl jasmonate (J) resuspended in a 0.1% (v/v) Tween 20 solution, compared to bark from trees sprayed with Tween 20 only; (c) green apical leader tissue (L) of Sitka spruce saplings harvested 2 h, 6 h or 2 d following feeding by defoliating spruce budworms (B; *Choristoneura occidentalis*), mechanical wounding by scissors (M) alone, or wounding plus the application of budworm regurgitant (R), compared to leader tissue from untreated control trees; (d) secondary xylem tissue (SX) harvested from the compression zone (C) of white spruce saplings 2 h, 6 h, 1 d, 2 d or 5 d following forced bending at a 45° angle compared to secondary xylem from untreated control trees; (e) secondary xylem tissue harvested from the opposite zone (O) of the same saplings as in experiment (d); and (f) the woody base of apical leaders from sapling Sitka spruce trees enriched for secondary xylem by removal of the green bark tissue compared to the green tip of leaders from the same trees harvested in late summer of 2003(A), and repeated in 2004(B), as well as bark plus secondary xylem from the base of leaders compared to the green tip of leaders harvested from the same trees in 2004(C). For further details of experiments (a)–(e) see Section 4; details of experiment (f) are described in Friedmann et al. (2007).

treatment of bark, but weak, if any, induction in bark tissue following other stress treatments or in green leaders following wounding, insect herbivory or the application of insect regurgitant. In contrast to members of the DIR-a and DIR-b subfamilies, the *PDIR25–PDIR29* plus *PDIR35* cluster demonstrates a pattern of strong down-regulation in secondary xylem from compression and opposite sides of bent saplings 6 h following initiation of treatment, followed by strong up-regulation in the same tissues at 1 d and 5 d post-treatment. Transcripts of this gene cluster also demonstrate strong preferential abundance in the green apical leaders compared to secondary xylem at the leader base. In contrast, array elements representing *PDIR22* and *PDIR23* appear to be weakly down-regulated in bark and leader tissues following stress treatments, with the possible exception of MeJA treatment of bark where moderate up-regulation is observed (Fig. 5). *PDIR22* and *PDIR23* transcripts do not show consistent changes in abundance in secondary xylem following bending stress, and are much more abundant in secondary xylem at the leader base compared to green apical shoots (Fig. 5).

## 3. Conclusions

In summary, the extended gene family of DIR- and DIR-like proteins identified in the spruce EST and FLcDNA collections provides a rich resource for future functional analysis in the context of phenolic defense and possibly other metabolic processes in spruce. Based on phylogenetic analysis, gene-specific real-time PCR expression data and microarray expression meta-analysis, our future work in the context of weevil resistance in Sitka spruce will focus primarily on members of the spruce DIR-a subfamily.

## 4. Experimental

### 4.1. Plant and insect materials and treatment of trees

To monitor DIR expression in response to wounding or insect attack, Sitka spruce [*P. sitchensis*, clone FB3-425; derived from somatic embryogenesis (CellFor Inc., Vancou-

ver, BC, Canada)] trees were grown outside at the University of British Columbia (UBC) greenhouse as described by Ralph et al. (2006b) to a height of 60–70 cm. Adult white pine weevils (*Pissodes strobi*) were generously provided by Dr. Rene I. Alfaro (Pacific Forestry Centre, Canadian Forest Service, Victoria, BC, Canada). Weevils were reared from larvae in infested Sitka spruce shoots collected at natural infestation sites in British Columbia in 2002 and maintained on Sitka spruce shoots. Western spruce budworms (*Choristoneura occidentalis*) were obtained from the Canadian Forest Service and reared and maintained on artificial diet (Addy, 1969) at 25 °C, 50–60% humidity and 16/8 h (light/dark) photoperiod. Groups of mid-instar larvae were used in experiments. Two weeks prior to experiments in May (weevils) or July (budworms) 2003, 2-year-old trees were moved inside the UBC greenhouse and maintained as described by Ralph et al. (2006b). For mechanical wounding, the bark of five trees per time point was pierced, using an 18-gauge hypodermic needle, at 5 mm intervals on opposite sides along the entire length of the stem, and the apical leaders of five trees per time point had *ca.* 50% of needle tissue removed with scissors. For insect treatment, budworms and weevils were kept without food on moist filter paper for 48 h prior to placing them on trees. Weevils were caged under mesh bags on the upper two-thirds of the stem, and budworms were caged on the leader and uppermost crown of lateral shoots (10 weevils or budworms on each of five individual trees per time point). For regurgitant treatment, budworm larvae were maintained on Sitka spruce shoots for several days prior to harvesting gut contents and saliva by pinching behind the head and removing mouth contents with suction, and then stored at −80 °C prior to use. Approximately 20 μL of regurgitant was applied by brush to needles of each leader of five trees per time point that were previously wounded with scissors. For the untreated control for both the weevil and budworm experiments, an additional five trees per time point were covered with mesh bags, but otherwise left untreated. Trees were kept in separate treatment groups in a well-ventilated greenhouse. Five trees each for control, wounding alone, wounding plus budworm regurgitant, and insect treatments were harvested 2, 6 and 48 h after initiation of treatment. To harvest bark tissue, the upper two-thirds of the tree (excluding the green shoot tip) was cut longitudinally with a razor blade and the outer tissue was peeled off the woody inner stem tissues. To harvest leader tissue, only the young, light green portion of the leader was collected (apical 10–15 cm section). Tissues were harvested individually from each tree and separately flash frozen in liquid nitrogen and stored at −80 °C prior to RNA isolation. For real-time PCR analysis, total RNA was isolated from individual trees at the 48 h time point for both insect treatments and untreated controls. For cDNA microarray analysis, equal amounts of total RNA were combined from each tree for each treatment and time point.

To monitor DIR expression in response to methyl jasmonate treatment, white spruce [*P. glauca*, clone I-1026; derived from somatic embryogenesis (CellFor Inc., Vancouver, BC, Canada)] trees were maintained as described above for Sitka spruce. Five trees were each sprayed with 150 mL of 0.01% (v/v) methyl jasmonate (Sigma–Aldrich, St. Louis, MO, USA) dissolved in 0.1% (v/v) Tween20 (Sigma–Aldrich). As a control, an additional five trees were each sprayed with 150 mL of a 0.1% Tween 20 solution. Bark tissue was harvested individually from each tree 48 h following treatment as described above, followed by RNA isolation. Equal quantities of total RNA were pooled from each tree by treatment prior to cDNA microarray analysis.

To monitor DIR expression in response to stem bending, white spruce (clone I-1026) trees were maintained as described above for Sitka spruce. Ten trees per time point were subjected to bending stress by attaching one end of a string to the woody base of the apical leader, pulling the stem down to *ca.* a 45° angle without damaging, and tying the other end of the string to a firm support near the base of the potted plant. After removal of bark, secondary xylem tissue was harvested, using a razor blade, from a 5 cm region of stem centered at the point of maximum bending on the near-side (compression wood) and the far-side (opposite wood) of each tree. As a control, secondary xylem was harvested from an additional 10 untreated control trees at the same time of day. Tissue was pooled from each tree by treatment and time point prior to RNA isolation.

For tissue profiling of constitutive DIR and DIR-like expression, cortex, phloem, cambium and developing xylem were rapidly separated under a dissecting microscope (15 min per tree), along with young lateral shoots and roots, from 10 3-year-old Sitka spruce saplings in May 2004 and pooled by tissue. White spruce (clone I-1026) somatic embryos for profiling of constitutive DIR and DIR-like expression were generously provided by Dr. David Ellis (CellFor Inc.) and were grown from callus tissue on media supplemented with abscisic acid and indole-3-butyric acid for six weeks using standard procedures prior to harvest. All tissues used for real-time PCR analysis were flash frozen in liquid nitrogen and stored at −80 °C prior to total RNA isolation.

### 4.2. RNA isolation and cDNA synthesis

Total RNA was isolated from all tissues, except for somatic embryos, according to the protocol of Kolosova et al. (2004). Total RNA from somatic embryos was isolated using the RNAqueous Midi kit (Ambion, Austin, TX, USA) according to manufacturer's instructions. Total RNA was quantified and quality checked by spectrophotometer and agarose gel. RNA was also evaluated for integrity and the presence of contaminants using reverse transcription with Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) with an oligo d($T_{18}$) primer and $\alpha P^{32}$ dGTP incorporation. After removal of unincorporated nucleotides using gel filtration columns

(Microspin S-300 HR columns, Amersham Pharmacia Biotech, Buckinghamshire, UK) the resulting cDNA smear was resolved using a vertical 1% agarose alkaline gel and visualized using a Storm 860 phosphorimager (Amersham Pharmacia Biotech).

### 4.3. Real-time PCR gene expression and data analysis

For constitutive tissue expression profiling, 9 μg of total RNA from each tissue was treated with DnaseI (Invitrogen), divided into three aliquots of 3 μg each, and independent cDNA synthesis reactions were performed using Superscript II reverse transcriptase (Invitrogen) with an oligo d($T_{18}$) primer according to manufacturer's instructions. For analysis of induced tissues, 15 μg of total RNA per tree was treated with DNaseI, divided into three aliquots of 5 μg, and converted to cDNA in three independent reactions. Efficiency of each cDNA synthesis was assessed individually by gel electrophoresis prior to pooling cDNA templates by treatment and time point. Gene-specific PCR primers were designed (Table 3) using a stringent set of criteria including predicted melting temperature of $64 \pm 2$ °C, primer lengths of 20–24 nucleotides, guanine-cytosine contents of 40–60% and PCR amplicon lengths of 70–185 bp. Primer specificity (single product of expected length) was confirmed by analysis on a 2% agarose gel, by melting curve analysis, and by sequence verification of PCR amplicons for at least one PCR reaction per gene from both weevil- and budworm-induced tissues. To further evaluate the amplification performance of each primer pair, a 10-fold dilution series of corresponding DNA plasmids containing full-length transcripts ($10^{-2}$ to $10^{-4}$ ng template) was analyzed (i.e., to confirm PCR amplification in the absence of detectable PCR products across the constitutive tissue panel). Primers for spruce eukaryotic translation initiation factor (*TIF5A*; GenBank Accession DR448953; EST IS0013_F24) served as a quantification control. PCR was performed in optical 96-well plates with a DNA Engine Opticon2 continuous fluorescence detector (MJ Research, Waltham, MA, USA) using SYBR Green to monitor dsDNA synthesis. Reactions contained 7.5 μL DyNAmo HS SYBR Green qPCR kit master mix (Finnzymes, Espoo, Finland), 10 ng cDNA and 0.3 μM of each primer in a final volume of 15 μL. Reactions with the cDNA template replaced by nuclease-free $H_2O$ or 10 ng of non-reverse transcribed RNA were run with each primer pair as a control. The following standard thermal profile was used for all PCRs: 95 °C for 15 min; 40 cycles of 95 °C for 15 s, 60 °C for 30 s and 72 °C for 30 s; then 72 °C for 10 min. Fluorescence signal was captured at the end of each cycle and melting curve analysis was performed from 65 °C to 95 °C with data capture every 0.2 °C during a 1 s hold. Data were analyzed using the Opticon Monitor analysis software version 2.02 (MJ Research). Quantification of each transcript in each cDNA source consisted of at least three independent (different 96-well plates) technical replicates. To generate a baseline-subtracted plot of the logarithmic increase in fluorescence signal (ΔRn) versus cycle number, baseline data were collected between cycles 3 and 10. All amplification plots were analyzed with an Rn threshold of 0.003 to obtain Ct (threshold cycle) values. Transcript abundance was normalized to *TIF5A* by subtracting the Ct value of *TIF5A* from the Ct value of each transcript, where $\Delta Ct = Ct_{transcript} - Ct_{TIF5A}$. Average ± SEM Ct values for *TIF5A* in control and weevil-induced bark tissues, control and budworm-induced leader tissues, and constitutive tissues was $15.63 \pm 0.09$, $17.28 \pm 0.10$ and $17.48 \pm 0.36$, respectively. Transcript abundance was obtained from the equation $(1 + E)^{-\Delta Ct}$, where $E$ is the PCR efficiency, as described by Ramakers et al. (2003), which is derived from the log slope of the fluorescence versus cycle number in the exponential phase of each amplification plot and the equation $(1 + E) = 10^{slope}$. Average ± SEM PCR efficiency across all primer pairs was $93.58 \pm 0.65$ (Table 3). A transcript with a relative abundance of one is equivalent to the abundance of *TIF5A* in the same tissue.

In order to assess the biological response to weevil (W) or budworm (B) herbivory, a linear model for each gene containing a treatment effect for W or B minus untreated control (C) was fit using data from at least three independent technical replicates derived from each of the four weevil- or budworm-treated trees and four untreated control trees. W or B minus C effects, as well as biological and technical variation, were estimated using a mixed-effects model where the error term for the W or B minus C effect was computed by pooling the biological and technical variation. The ratio of the W or B minus C parameter estimate to the standard error was used to calculate a $t$ statistic and $P$ value.

### 4.4. Microarray hybridization and analysis

All hybridizations were performed using the spruce 16.7K cDNA microarray (Ralph S. and Bohlmann J., in preparation) using hybridization procedures as previously described (Ralph et al., 2006b). For weevil feeding and wounding of bark, total RNA was compared to untreated control bark for each time point using four technical replicate hybridizations for each comparison with dye flips (24 slides total). Total RNA from weevil feeding and wounded bark were also compared within treatments across the three time points, as well as between treatments at each time point, using two technical replicate hybridizations for each comparison with dye flips (18 slides total). For the methyl jasmonate treatment, total RNA was compared to Tween20 treated control bark using four technical replicate hybridizations with dye flips (four slides total). For budworm feeding, wounding alone, and wounding plus regurgitant treatment of leaders, total RNA was compared to untreated control leaders for each time point using four technical replicate hybridizations for each comparison with dye flips (36 slides total). Total RNA from budworm feeding, wounding, and regurgitant treatments was also compared within treatments across the three time points, as

well as between treatments at each time point, using two technical replicate hybridizations for each comparison with dye flips (36 slides total). For bending stress treatment, total RNA from secondary xylem of the compression and opposite zones was compared to untreated control secondary xylem tissue harvested at the same time of day for each time point using two replicate hybridizations for each comparison with dye flips (20 slides total). Two additional hybridizations were also performed for each time point to compare secondary xylem from compression and opposite zones directly (10 slides total). For the developmental gradient along apical shoots, a complete description of experimental treatments and microarray analysis is provided in Friedmann et al. (2007).

Slide scanning, background signal intensity correction, and data normalization were performed as described previously (Ralph et al., 2006a). For each experiment, the biological response to treatments was calculated using linear models containing a dye effect and a treatment effect (e.g., weevil herbivory at 2 h minus untreated control, compression wood at 1 d minus untreated control, etc.) as described previously (Ralph et al., 2006a). All statistical analyses were performed within the R statistical package (http://www.r-project.org/).

### 4.5. Isolation of spruce full-length DIR cDNA clones

A TBLASTN search of the Treenomix spruce EST database (www.treenomix.ca) containing 147,146 ESTs derived from 3′-end sequencing, was performed using publicly available plant DIR sequences. CAP3 sequence assembly software (Huang and Madan, 1999) was used to group ESTs into unique singletons and contigs (40 bp overlap, 95% identity). DIR clones in the pBluescript II SK (+) vector were identified in our cDNA library glycerol stocks, insert sized using PCR with −21M13 forward (5′-TGTAAAACGACGGCCAGT-3′) and M13 reverse (5′-CAGGAAACAGCTATGAC-3′) primers, and cDNA inserts were sequenced from both ends using the same primers. In this manner 15 unique spruce DIR FLcDNAs and one partial cDNA was obtained.

### 4.6. Sequence and phylogenetic analysis

Additional plant DIR genes were identified in a comprehensive search of GenBank using BLAST tools and the set of plant DIR previously described (Ralph et al., 2006b). This process was repeated with each newly identified set of plant DIR genes until no further sequences with significant similarity were identified. Predictions for *pI* and molecular mass were made using the entire ORF and the *pI*/MW tool at Expasy (www.expasy.org/tools/pi_tool.html). Amino acid multiple sequence alignments for spruce and Arabidopsis DIR proteins (Fig. 1) were made with ClustalW (www.ebi.ac.uk/clustalw/) and Boxshade (//bioweb.pasteur.fr/seqanal/interfaces/boxshade.html). All plant DIR sequences were aligned using Dialign (threshold = 0; Morgenstern et al., 1998; http://bioweb.pasteur.fr/seqanal/interfaces/dialign2-simple.html). Multiple sequence alignments were manually adjusted to a conserved set of *ca.* 160 amino acids prior to maximum likelihood analysis using Phyml v2.4.1 (Guindon and Gascuel, 2003) with the JTT (Jones et al., 1992) amino acid substitution matrix. The proportion of invariant sites as well as the alpha shape parameter were estimated by Phyml. Trees were generated using BIONJ (Gascuel, 1997), a modified neighbor-joining algorithm. SEQBOOT of the Phylip v3.62 package (Felsenstein, 1993; //evolution.genetics.washington.edu/phylip.html) was used to generate 75 bootstrap replicates, which were then analyzed using Phyml and the previously estimated parameters. CONSENSE, also from Phylip, was used to create a consensus tree. Treeview (Page, 1996) was used to visualize all trees.

## Acknowledgements

We thank Ms. Natalia Kolosova and Dr. Shawn Mansfield for collaboration leading to gene expression data for spruce trees treated by MeJA or stem-bending, respectively. We thank Dr. David Ellis, CellFor Inc., for spruce seedlings and somatic embryos; Dr. Rene I. Alfaro from the Canadian Forest Service for weevils and Mr. Bob McCron from the Canadian Forest Service for budworms; Dr. Kim Rensing for technical assistance with spruce tissue sectioning; Mr. Ian Cullis for somatic embryo propagation; Mr. Rick White for assistance with microarray and real-time PCR statistical analysis; and Mr. David Kaplan for greenhouse support.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.phytochem.2007.04.042.

## References

Addy, N.D., 1969. Rearing the forest tent caterpillar on an artificial diet. J. Econ. Entomol. 62, 270–271.

Blanchard, D.J., Cicek, M., Chen, J., Esen, A., 2001. Identification of β-glucosidase aggregating factor (BGAF) and mapping of BGAF binding regions on maize β-glucosidase. J. Biol. Chem. 276, 11895–11901.

Davin, L.B., Wang, H.B., Crowell, A.L., Bedgar, D.L., Martin, D.M., Sarkanen, S., Lewis, N.G., 1997. Stereoselective bimolecular phenoxy radical coupling by an auxiliary (Dirigent) protein without an active center. Science 275, 362–367.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package) Version 3.62. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Friedmann, M., Ralph, S.G., Aeschliman, D., Zhuang, J., Ritland, K., Ellis, B.E., Bohlmann, J., Douglas, C.J., 2007. Microarray gene expression profiling of developmental transitions in Sitka spruce (*Picea sitchensis*) apical shoots. J. Exp. Bot. 58, 593–614.

Gascuel, O., 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. 14, 685–695.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.

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

Kim, M.K., Jeon, J.H., Fujita, M., Davin, L.B., Lewis, N.G., 2002. The western red cedar (*Thuja plicata*) 8-8′ DIRIGENT family displays diverse expression patterns and conserved monolignol coupling specificity. Plant Mol. Biol. 49, 199–214.

Kolosova, N., Miller, B., Ralph, S., Ellis, B.E., Douglas, C., Ritland, K., Bohlmann, J., 2004. Isolation of high-quality RNA from gymnosperm and angiosperm trees. Biotechniques 36, 821–824.

Morgenstern, B., Frech, K., Dress, A., Werner, T., 1998. DIALIGN: finding local similarities by multiple sequence alignment. Bioinformatics 14, 290–294.

Page, R.D., 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.

Ralph, S.G., Yueh, H., Friedmann, M.F., Aeschliman, D., Zeznik, J.A., Nelson, C.C., Butterfield, Y.S.N., Kirkpatrick, R., Liu, J., Jones, S.J.M., Marra, M.A., Douglas, C.J., Ritland, K., Bohlmann, J., 2006a. 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. 29, 1545–1570.

Ralph, S., Park, J.Y., Bohlmann, J., Mansfield, S.D., 2006b. Dirigent proteins in conifer defense: gene discovery, phylogeny and differential wound- and insect-induced expression of a family of DIR and DIR-like genes in spruce (*Picea* spp.). Plant Mol. Biol. 60, 21–40.

Ramakers, C., Ruijter, J.M., Deprez, R.H.L., Moorman, A.F.M., 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci. Lett. 339, 62–66.

Xia, Z.Q., Costa, M.A., Proctor, J., Davin, L.B., Lewis, N.G., 2000. Dirigent-mediated podophyllotoxin biosynthesis in *Linum flavum* and *Podophyllum peltatum*. Phytochemistry 55, 537–549.