## Transcriptome Profiling of Needle versus Cambium Tissues by EST Analysis of Subtracted cDNA Libraries in Black Spruce (*Picea mariana*)

**Ishminder K. Mann1,†, Christopher J. Friedline2,†, and Om P. Rajora1,3\***

1Forest Genetics and Biotechnology Group, Department of Biology, Life Sciences Centre, Dalhousie University, 1355 Oxford Street, Halifax, NS B3H 4J1 Canada

2Virginia Commonwealth University, Department of Biology, 1000 West Cary Street Room 126, Richmond, VA 23284 USA

3Canada Research Chair in Forest and Conservation Genomics and Biotechnology, Faculty of Forestry and Environmental Sciences, P.O. Box 44555, 28 Dineen Drive, University of New Brunswick, Fredericton, NB E3B 6C2, Canada

†Author contributed equally

\*Principal investigator and author for correspondence

Telephone: (506)

453-4501

Fax: (506) 453-3538

E-mail: [Om.Rajora@unb.ca](mailto:Om.Rajora@unb.ca)

**Abstract**

Leaves (called needles in conifers) and cambium play crucial roles in growth, development and survival of plants. A major portion of the planet’s biomass is produced by the cellular and metabolic activities in leaves and cambium. Despite being the primary sites for the key functional pathways in plants, very little is known about similarities and differences in gene expression patterns between leaves and cambium. We have identified genes expressed differentially in needle versus cambium tissues of black spruce (*Picea mariana*) via analysis of ESTs and dot blot hybridization of cDNA clones from subtracted cDNA libraries. Four subtracted cDNA libraries were constructed from needle and cambium tissues of the parents of a black spruce mapping population, using the suppression subtractive hybridization (SSH) method. A total of XXXX cDNA clones were partially sequenced, and a total of XXXX quality ESTs were obtained and analyzed. The SSH cDNA libraries uncovered transcripts involved in almost all major functional pathways, and revealed XXXX transcripts unique to black spruce and novel to the total spruce dbEST at NCBI. Functional distribution of the needle transcripts was remarkably different from that of the cambium transcripts. About XXX of the transcripts showed differential gene expression for their abundance between needles and cambium based on dot-blot hybridization. Genes involved in photosynthesis, energy and metabolism, disease and stress response, cell structure and cellular transport were differentially expressed. Genes involved in disease and stress response mechanisms, and cell structure were predominantly expressed in cambium tissues whereas those involved in photosynthesis, and energy and metabolism were predominantly expressed in needle tissues. Transcripts coding for ribulose-1,5-biphosphate carboxylase, and metallothionein-like proteins were most abundant in the needle, and cambium SSH cDNA libraries, respectively. We have developed an important spruce genomic resource.

**Keywords:** Differential gene expression in needles and cambium, Suppression subtractive hybridization, Functional distribution of transcripts in needles versus cambium, Novel spruce transcripts, Transcriptomics, Spruce genomic resource

# Introduction

Leaves and cambium play a vital role in the growth and development of plants. Leaves (called needles in conifers) are the primary site of photosynthesis, respiration and transpiration (Hopkins and Hunter 2004). They are the principal photosynthetic machinery and serve as the energy source by fixing the atmospheric carbon-dioxide. The fixed carbon is then converted to chemical energy, which is used for growth, development and the maintenance of cellular homeostasis (Hopkins and Hunter 2004). The cambium is a lateral meristem producing secondary growth, xylem, phloem and bark. In perennial woody plants, such as conifer and broadleaved trees, perennial cambial activity produces functional xylem and phloem on a regular basis, resulting in the production of wood. Therefore, a major portion of the planet’s biomass is produced by the cellular and metabolic activities in leaves (needles) and cambium. Despite leaves and cambium are the sites of major functional cellular pathways, very little is known about the similarities and differences in gene expression patterns between them. Transcriptome profiling of leaves versus cambium can help to better understand the cellular functions of these tissues as well as molecular mechanisms underlying growth, development and stress response in plants, including forest trees. Also, the sequencing of the transcripts from needles and cambium can provide an important genomic resource. Based on their known cellular functions, genes belonging to certain functional categories, such as photosynthesis, energy and metabolism, and cell structure are expected to be expressed differentially between leaves and cambium.

Tissue and organ-specific gene expression is a well-known phenomenon in plants. For example, in several woody and non-woody plant species, tissue-specific genes have been identified during their morphological and phenological developments, such as growth

and reproduction (Opsahl-Ferstad *et al*. 1997; Woo *et al*. 1995; Walden *et al.* 1999), embryogenesis (Heck *et al*. 1995; Nuccio and Thomas 1999), flowering (Sablowski and Meyerowitz 1998), and fruiting (Boss *et al*. 2001). Most of the previous studies on tissue and organ specific gene expression were based on single genes, including those in conifers (e.g., Hutchinson *et al*. 1999). However there is no report on differential gene expression or gene expression pattern comparisons between leaves and cambium tissues in any plant although expressed sequence tags (ESTs) have been obtained from cDNA libraries constructed from needle and cambium tissues in white spruce (*Picea glauca*) (Pavy *et al*. 2005).

Several methods have been used for identification of differentially expressed genes. These include differential analysis of library expression (Li *et al.* 2004), differential display (Liang *et al.* 1992), representational difference analysis (Lisitsyn *et al.* 1993), physical removal of common sequences (Akopian *et al.* 1995), serial analysis of gene expression (Velculescu *et al.* 1995), suppression subtractive hybridization (SSH, Diatchenko *et al.* 1996) and microarrays (Schena *et al.* 1995). Except for SSH, all other methods based on cDNA subtraction have a drawback of maintaining disproportionate concentrations of differentially expressed transcripts during subtraction. The microarray and SSH are powerful methods to identify differentially expressed genes. Although the microarray method is more informative than other methods, it requires availability of microarray chips. In conifers, microarray chips have recently been developed and used for gene expression profiling in response to wounding and insect feeding in Sitka spruce (*Picea sitchensis*) (Ralph *et al*. 2006), and during adventitious root development in

lodgepole pine (*Pinus contorta*) (Brinker *et al.* 2004). However, these microarrays are restricted to specific labs and are not publicly available for common use.

The SSH method generates an equalized representation of differentially expressed genes irrespective of their relative abundance in the genome (Diatchenko *et al.* 1996). It enables the construction of subtracted cDNA libraries using hybridization and suppression PCR in a single procedure, and enriches the differentially expressed mRNA in one transcript population relative to the other (Diatchenko *et al.* 1996). The SSH method not only allows reduction in the redundancy but also enrichment of rare transcripts in the library. The SSH method has been successfully used to identify differentially expressed genes involved in important biological and phenological processes in woody and non- woody plants, such as transcripts associated with early and late bud flushing in Norway spruce (*Picea abies*) (Yakovlev *et al.* 2006), genes expressed during xylogenesis in Eucalyptus (Foucart *et al.* 2006), and stress, disease and defense related genes in rice (Xiong *et al.* 2001, Wang *et al.* 2005) and *Arabidopsis* (Mahalingam *et al.* 2003).

Black spruce (*Picea mariana* (Mill.) B.S.P.) is a widespread transcontinental species of the North American boreal and temperate forests (Viereck and Johnston 1990), and has great ecological and economic importance. It is an early successional species and plays a major role in the functioning of forest ecosystems and in providing substantial carbon sink on the planet. Black spruce is one of the most important trees in Canada for the production of pulp and paper, and is the most important reforestation species in Canada (Morgenstern and Wang 2001). The Rajora lab has an ongoing research program on structural and functional genomics of black spruce and red spruce (*Picea rubens*) related to growth and adaptation to climate change. Like other conifers, genomics work in black spruce is challenging, primarily due to its relatively large genome size (1C= 15.8 pg; 2C =~31,000 mpb; Ohri and Khoshoo 1986; http://www.rbgkew.org.uk/cval/homepage.html) and long time taken for reaching reproductive maturity. The sequencing of the complete genome of black spruce or other conifers is not feasible at present merely because of their large genome size. Transcriptome profiling appears to be a viable way to identify genes involved in major processes in these plants, including growth, stress adaptation and defense. Identifying genes expressed differentially between needles and cambium in black spruce can help to better understand the mechanisms underlying growth and abiotic and biotic stress adaptation in this species in addition to providing an important genomic resource.

The objective of this study was to identify genes expressed differentially in needle versus cambium tissues in black spruce, using the SSH approach. Here we report on transcript profiling of needles versus cambium in black spruce with the snapshot of differentially expressed transcripts, based on EST analysis and dot blot hybridization of SSH cDNA libraries constructed using RNA isolated from needle and cambium tissues of the parents of a well-characterized mapping population.

# Material and methods

## Plant material and RNA isolation

The female (P32) and male (P40) parents of a well characterized black spruce mapping population, established in a genetic test at the Petawawa Research Forest (PRF) located in Chalk River, Ontario (46° N, 77° 30' W), were used as the RNA source. These parents were of approximate 30 years of age. Small shoot cuttings and twigs with current year’s new needles were collected from these trees on June 25, 2002 during the day and period of fresh growth of needles and cambium. The collected material was immediately frozen in liquid nitrogen and remained frozen for two days during transportation to Dalhousie University, Halifax. Then the plant material was transferred to a -85ºC freezer. Cambium was removed with forceps after peeling off the bark from the frozen stem cuttings with a scalpel. Needles were harvested from the frozen shoot cuttings. Total RNA was extracted from the needle and cambium tissues as described by Chang *et al.* (1993). Quality and quantity of the isolated RNA were determined using a spectrophotometer (SPECTRAmax PLUS), and the extracted RNA was of high quality (OD260/OD280=1.79-1.87). Quantity of the isolated RNA was higher for needles (~100 µg/gm) than for cambium (~50 µg/gm). The poly(A) RNA was purified using RNeasy Mini Kit (Qiagen Inc., Mississauga, ON, Canada).

## Construction of SSH cDNA libraries

Four SSH cDNA libraries were constructed (Table 1). Two SSH cDNA libraries were constructed each from needle and cambium tissues (one each from P32 and P40) to be considered as biological replicates. First-strand of cDNA and double stranded cDNA were

synthesized from 2 µg of poly(A) needle (tester population) and cambium (driver population) RNA using Smart PCR cDNA Synthesis Kit (Clontech Laboratories Inc., CA, USA). The cDNA was then size-fractioned followed by digestion with *Rsa*I. SSH was performed using the PCR-Select cDNA Subtraction Kit (Clontech Laboratories Inc., CA, USA) according to the manufacturer’s instructions. The final subtractive product is supposed to contain cDNA population expressed in needles. The same procedure was repeated by taking cambium cDNA as tester and needle cDNA as driver population to obtain the PCR product containing cDNA population expressed in cambium. The subtracted cDNA populations were cloned in T/A cloning vector pCR2.1 (Invitrogen Canada Inc., Mississauga, ON, Canada). TOP10 chemically competent *Escherichia coli* cells (Invitrogen Canada Inc. ON, Canada) were transformed with vectors containing cDNA inserts according to the manufacturer’s instructions. Plasmid DNA was isolated from the transformed cells grown overnight on Luria Broth agar plates containing antibiotic kanamycin (50 µg/ml) using QIAprep Spin Miniprep Kit (Qiagen Inc., Mississauga, ON, Canada).

Transformation efficiency was confirmed by amplifying the inserts from randomly chosen 10 to 15 clones from each SSH cDNA library. Each amplification reaction mixture contained hand picked clone and 2.5 µL 10X buffer, 2 µL MgCl2 (25 mM), 2 µL dNTP (2.5 mM each), 1 µL of nested primer 1 and nested primer 2R (10 µM) provided in PCR- Select cDNA Subtraction Kit, 17.875 µL of autoclaved, deionized distilled water, and 0.05 units of Taq (MBI Fermentas Inc., Burlington, ON, Canada). PCR was performed according to the following parameters: 95ºC for 5 min and 25 cycles of 95ºC for 10 s and 68ºC for 2 min, and final extension at 68ºC for 5 min. All four SSH cDNA libraries

displayed 100% transformation efficiency. The cDNA insert size was determined by electrophoresing the amplified cDNA inserts on 1.2% agarose gel containing ethidium bromide. The inserts ranged in size from 300 bp to >1000 bp, with an average of 700 bp.

## Sequencing and sequence analysis

A total of 7872 clones were hand picked from the four SSH cDNA libraries for sequencing (Table 2). This included 21 plates of 96 clones per plate for each of the P32C, P40N and P40C SSH cDNA libraries and 19 plates for the P32N SSH cDNA library. The sequencing was done at the Genome Atlantic Sequencing Platform located at the Institute for Marine Biosciences, National Research Council of Canada, Halifax, Canada. A total of 7232 clones were successfully sequenced (Table 2). These cDNA sequences were analyzed for their quality assessment using phred program and vector sequences were trimmed using cross\_match program with settings of minmatch 12 and minscore 20 (Ewing *et al.* 1998; Ewing and Green 1998).

Only the sequences with phred quality value of >20 and with the sequence length of greater than 100 bp were used for further analysis. The number of good quality sequences obtained after phred and cross\_match analyses ranged from 1298 to 1688, with a total of 5740 sequences for the four SSH cDNA libraries (Table 2).

## Differential gene expression analysis

Because a spruce DNA microarray chip was not available, we used EST analysis of SSH cDNA libraries and dot blot hybridization to examine differential gene expression between the needle and cambium tissues. These methods are complimentary to each other. EST

analysis method is based on EST similarities between the needle and cambium cDNA libraries, EST similarities with the available nucleotide and protein databases, annotation of transcripts to identify genes, and determining differential levels of representation of specific transcripts and their functional distribution in the cambium versus needle SSH cDNA libraries. The sequence similarities are based on the length of sequences available in the database and the test SSH cDNA libraries. Because a whole gene sequence may not be compared, the percentage of differentially expressed genes is likely to be biased from the EST analysis alone. Since the individual cDNA clones and whole cDNA population from a tissue are used in dot blot hybridization, this method will be more realistic to identify gene expressed specifically in a tissue and differentially in terms of their abundance. However, the dot blot hybridization method is quite time and labor intensive, and thus limited by the number of cDNA clones that can be analyzed. A combination of EST analysis and dot blot hybridization methods, as we used in the present study, can provide a more accurate estimate of differential gene expression.

## Sequence similarity analysis

Sequence similarity analysis was performed for the quality ESTs obtained from the four SSH cDNA libraries to (1) examine the percentage of similar and distinct (100 - % similar) ESTs among the four SSH cDNA libraries, (2) determine the similarities of ESTs with the total dbEST and *Picea* species dbEST available in Genbank of the National Centre for Biotechnology Information (NCBI), and with the black spruce ESTs obtained from a standard needle cDNA library, and (3) identify and annotate transcripts based on EST similarities with the non-redundant protein database.

Basic Local Alignment Search Tool (BLAST) for nucleotide sequences (BLASTN) was performed using blastall client from NCBI to determine the number and percentage of ESTs similar among the four SSH cDNA libraries. A threshold criterion of cutoff of e-1 was used so that the number of similar sequences is not underestimated. BLASTN analysis using blastall was also performed to determine the EST similarities with the ESTs developed from a standard black spruce needle cDNA library (Mann and Rajora, unpublished data), and dbEST available from all of the *Picea* species at NCBI (249,705 EST sequences, April 24, 2006), using a cutoff of e-5.

BLASTN and BLASTX analyses were conducted to determine the nucleotide and protein sequence similarities of ESTs from the four SSH cDNA libraries with the total dbEST and non-redundant (NR) protein database available at NCBI, using custom designed perl scripts. The perl scripts were designed so that the BLASTN and BLASTX results were obtained from only those sequences meeting the criteria of minimum bit score of 100 and alignment length of 100 bp. This stringent criterion was applied because the results from the BLASTX analysis were further used for gene annotation.

Transcripts from four the SSH cDNA libraries were annotated based on their BLASTX similarities. The expressed genes (transcripts coding for proteins of known function) were identified and then grouped into functional categories based on literature search, following the grouping of Kyoto Encyclopedia of Genes and Genomes (KEGG) website [(http://www.genom](http://www.genome.ad.jp/kegg/%3B)e[.ad.jp/kegg/;](http://www.genome.ad.jp/kegg/%3B) Kanehisa and Goto 2000). Gene transcript enrichment was calculated by comparing the relative number of annotated gene homologues in the needle versus cambium tissue SSH cDNA libraries. The enrichment factor for a specific gene or gene family in the needle versus cambium SSH libraries or

vice-versa was calculated as the ratio of the percentage of ESTs targeting that gene or gene family in the needle and cambium SSH libraries or vice-versa.

The ESTs of the most abundantly expressed transcripts (Ribulose-1,5-biphosphate carboxylase (RuBisCO) and its small subunit (RbcS) in the needle SSH libraries and metallothionein-like (MT) protein in the cambium SSH libraries) were further analyzed to

1. check whether the multiple ESTs targeting the same gene are the sequences of the same or different cDNA clones or whether the transcripts represent the same or different members of a multi-gene family, and (2) putatively identify the groups or classes of the multigene family the annotated genes belong to. The ESTs from those plant species that have sequences of the RuBisCO and its small subunit and MT protein (family 15; <http://www.expasy.org/cgi-bin/lists?metallo.txt)> family members available were

downloaded from the NCBI dbEST. Similarities of the ESTs from the black spruce SSH cDNA libraries targeting RuBisCO, RbcS and MT protein were determined with the downloaded dbEST. Different members of a multigene family in the black spruce SSH libraries and their putative classes were identified.

## Dot blot hybridization

Ninety-six transformed colonies from each of the four SSH cDNA libraries were used for the dot blot hybridization experiment. The 384 clones represented 157 genes belonging to 12 functional groups, including protein synthesis, energy and metabolism, disease and stress response, photosynthesis, protein destination and storage, cell structure, lipid biosynthesis and metabolism, secondary metabolism, signal transduction, transporters, growth and development, expressed and unknown protein, and 227 transcripts unique to

black spruce. The cDNA clones were selected on the basis of BLASTN and/or BLASTX similarity results for the P40N library and randomly for other three libraries. The cDNA sequences of the selected clones were PCR amplified from the transformed colonies as follows. Each amplification reaction mixture contained 1.5 µL (~20 ng) plasmid DNA, 2.5 µL 10X buffer, 2 µL MgCl2 (25 mM), 2 µL dNTP (2.5 mM each), 1 µL of nested primer 1 and nested primer 2R (10 µM), 16.375 µL of autoclaved double distilled water, and 0.05 units of Taq (MBI Fermentas Inc., Burlington, ON, Canada). PCR was performed according to the following parameters: 95ºC for 30 s and 25 cycles of 95ºC for 10 s and 68ºC for 2 min, and final extension at 68ºC for 5 min. The presence of a single amplified product was confirmed by electrophoresing the PCR products on 1.2% (w/v) agarose gels. Five µL of the PCR product from each positive clone was mixed with five µL of 0.6N NaOH. Then, one µL of the mixture was spotted onto a positively charged nylon membrane (Roche Diagnostics, Laval, Quebec) made in duplicates for each library (using one blot as control as it was hybridized with the cDNA population from the same tissue). The nylon membranes were placed on Whatmann 3MM paper presoaked with 10X saline sodium citrate (SSC). These membranes were exposed to ultraviolet light for 10 min to cross-link the DNA to the membrane. Fifteen µL (20 ng) of the unsubtracted cDNA was labeled from each cDNA library with digoxigenin-dUTP according to the manufacturer’s instructions using the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche Diagnostics, Laval, Quebec). The hybridization and detection were performed according to the manufacturer’s instructions. The membranes were pre-hybridized for 30 min and then hybridized for 12 h with the denatured labeled probes. The hybridized membranes were then washed with 0.5X SSC and 0.1% sodium dodecyl sulphate. After incubation in

blocking solution, antibody solution, washing solution, and detection solution, the membranes were subjected to immunological detection. Detection was performed by chemiluminescence with Chemiluminescent Substrate for Alkaline Phosphatase (CSPD) as a substrate. CSPD was applied to the membranes followed by an exposure of the X-Ray film at 20ºC for 15-20 minutes to visualize the hybridization intensity of transcripts present in needle and cambium tissues. Differences in the intensity of dot blot hybridization signals of the same clone with needle and cambium cDNA populations used as probe were used as indicative of the relative transcript levels. The number of clones showing differential gene expression in P32N, P32C, P40N and P40C SSH cDNA libraries were determined.

# Results

## EST similarities among needle and cambium SSH cDNA libraries

Twelve to 69% of the ESTs were found to be distinct among the four SSH cDNA libraries (Table 3). Such a high variability in the number of distinct sequences may be, among other factors, due to the differences in the cDNA subtraction efficiency. The percentage of similar ESTs was higher between the two needle or cambium SSH cDNA libraries than between the needle and cambium SSH cDNA libraries (Table 3). These results substantiate that the subtraction between the needle and cambium transcripts has occurred.

## EST similarities with the *Picea* species dbEST at NCBI

Based on a standard cutoff value of e-5, ESTs from the SSH cDNA libraries showed high similarities with the *Picea* dbEST available at NCBI (Table 4). About 0.4 to 14% ESTs, with a total of 256 from all four SSH cDNA libraries were found to be novel to the NCBI *Picea* dbEST. Sequence similarity of ESTs from the SSH cDNA libraries with those from a standard black spruce cDNA library varied from 46 to 90% (Table 4). A total of 1961 new black spruce transcripts were discovered in the SSH libraries compared to the standard needle cDNA library.

**EST similarities with the total dbEST and non-redundant protein database at NCBI** The percentage of black spruce ESTs from the four SSH cDNA libraries showing similarities with the available total dbEST at NCBI varied from 68 to 95%, (Table 2). The remaining 5 to 32% of the ESTs, with a total of 1135 ESTs were deemed to be novel. The novel ESTs are those sequences that did not show similarity to any other nucleotide

sequence in the NCBI database at a cutoff value of bit score of 100 and alignment length of 100 bp. The percentage of ESTs showing BLASTX similarities with the NR protein database ranged from 26.6 to 53.5, and the number of genes (transcripts coding for proteins with known functions) annotated ranged from 95 to 383 from the four SSH cDNA libraries (Tables 2, 5). The criterion used for the BLASTN and BLASTX similarity analyses here was different and more stringent than that used for the EST similarity analysis with the NCBI *Picea* dbEST or black spruce dbEST. This was done to minimize errors in gene annotation of the transcripts based on the above results.

## Gene annotation

A total of 570 different genes was identified from annotation of the transcripts from all four SSH cDNA libraries (Table 5). According to their functional roles, these annotated genes were involved in almost all major growth, development and metabolic pathways and belonged to 13 functional categories/classes (Table 5). The five most abundant functional categories of the transcripts included energy and metabolism, protein synthesis, expressed proteins, cell structure, and photosynthesis (Table 5). The genes involved in disease and stress response were also abundantly expressed. The transcripts of the genes involved in lignin biosynthesis (secondary metabolism) were also identified in both needle and cambium SSH cDNA libraries, and they did not show differential expression between the tissues. These include: 4-coumarate-CoA ligase (4CL) family protein, cinnamoyl-CoA reductase (CCR), cinnamoyl-CoA reductase family, caffeoyl-CoA O-methyltransferase (CCoA-OMT), phenylalanine ammonia-lyase (PAL), p-coumarate 3-hydroxylase. Two hundred new genes were identified in black spruce from the four SSH cDNA libraries in

addition to those identified previously from a standard cDNA library prepared using RNA extracted from needles (Mann and Rajora, unpublished data).

**Functional distribution of transcripts in needle and cambium SSH cDNA libraries** Functional distribution of the transcripts (ESTs) was substantially different between the needle and cambium tissue SSH cDNA libraries (Fig. 1). Among the ESTs obtained from the needle SSH cDNA libraries, genes involved in photosynthesis constituted the most abundant class, followed by genes involved in energy and metabolism, expressed proteins, and lipid biosynthesis and metabolism (Fig. 1A). In contrast, in the cambium SSH cDNA libraries, genes involved in the disease and stress response were the most abundant class of ESTs, followed by genes involved in cell structure, energy and metabolism, and protein synthesis (Fig. 1B). In the photosynthesis category, the most abundant transcript represented RuBisCO and its small subunit RbcS genes in the needle SSH libraries (Table 6), whereas in the disease and stress response category, the most abundant transcripts represented MT genes in the cambium SSH cDNA libraries. Both RuBisCO and MT genes belong to large multigene families.

Of the 337 ESTs representing RuBisCO and RbcS in the needle SSH cDNA libraries, 317 showed similarities with the larch (*Larix laricina*), and 20 with the Japanese black pine (*Pinus thurbergii*) RuBisCO sequences. The sequence similarities ranged from 65 to 98% over a long sequence length (49-167 amino acids). These 337 ESTs also showed similarities with RuBisCO and RbcS sequences from other plants but over a short sequence length (10–24 amino acids). All of the 337 ESTs had distinct sequences.

Of the 315 ESTs representing MT genes in the cambium SSH libraries, 310 showed 73-100% similarities with the white spruce MT sequence, and five 82.5-97.9% similarities

with the Norway spruce MT sequence, over a sequence length of 45 to 61 amino acids. Of

the 315 ESTs, 312 had distinct sequences and the remaining three ESTs shared the same

sequence. The conifer MT genes belong to MT Plant Type 3 of gene family (Binz and Kagi 2006; [http://www.expasy.org/cgi-bin/lists?metallo.txt).](http://www.expasy.org/cgi-bin/lists?metallo.txt)) We also checked whether the

ESTs representing MT genes in the cambium SSH libraries showed similarities to other classes of plant MT genes. This was done by excluding spruce MT sequences for the BLAST similarity analysis. The black spruce MT transcripts showed similarities with four other subdivisions of the MT multigene family: MT1 from wheat, MT2 from *Brassica napus*, *Cicer arietinum* and *Musa accuminata*, MT2A and MT2B from *Arabidopsis thaliana*. However, the sequence length showing similarities was quite low (4 to 15 amino acids).

**Differential gene transcript enrichment in needle and cambium SSH cDNA libraries** Comparison of the number of gene homologues identified in the needle versus cambium SSH cDNA libraries provided preliminary estimates of the enrichment of transcripts of specific genes in these libraries. The gene enrichment values based on 20 most abundant transcripts varied from 1.04 to 24.72 for the needle (Table 6) and 0.06 to 23.98 for the cambium (Table 7) SSH cDNA libraries. Of the 20 most abundant transcripts found in the needle SSH cDNA libraries, 12 represented the genes involved in photosynthesis, and energy and metabolism (Table 6). In cambium SSH cDNA library, 13 out of 20 abundant transcripts belonged to disease and stress response, energy and metabolism, cell structure

and protein synthesis functional categories (Table 7). Also transcripts of the genes belonging to the cellular transport were enriched by four folds in the needle SSH cDNA libraries.

## Dot blot hybridization

All 384 cDNA clones (96 each from one SSH cDNA library) showed positive hybridizations with the needle and cambium cDNA populations. However, about 10% of the clones showed differences in their hybridization intensities with the needle or cambium cDNA probes (Fig. 2; Table 8). These intensity differences were most likely due to the differences in the abundance of specific transcripts in needle versus cambium tissues. Most of these differentially expressed cDNA did not show any known BLASTX homologies/similarities. However, a number of these clones represented known genes expressed differentially between the needle and cambium tissues (Table 8). The dot blot hybridization results for these genes correspond with the abundance of their transcripts observed from the gene enrichment analysis.

# Discussion

## Differential representation of transcripts and differential expression of genes in cambium and needles

The functional distribution of transcripts and their enrichment in the needle versus cambium SSH cDNA libraries and dot blot hybridization results demonstrate that genes involved in photosynthesis, energy and metabolism, disease and stress response, cell structure and cellular transport are differentially expressed between the needle and cambium tissues of black spruce.

The distribution of ESTs based on their predicted functions shows that the genes involved in photosynthesis, energy and metabolism and cellular transport are more abundantly expressed in the needle than in cambium tissues, whereas the genes involved in disease and stress response, and cell structure are more abundantly expressed in cambium than in needle tissues of black spruce. These results are consistent with the respective functional roles of needles and cambium in plants. The dot blot hybridization results, albeit for a small number of genes, are consistent with the enrichment of differentially expressed gene transcripts between the needle and cambium SSH cDNA libraries. The transcripts of genes coding for MT proteins, and chitinases were down regulated in the needle tissues whereas those of RuBisCO, lipase, lipid transfer protein, and plasma membrane intrinsic protein 2A were down regulated in the cambium tissues. The dot blot hybridization demonstrates that about 10% (39) of the 384 transcripts (cDNA clones) were differentially expressed in terms of their abundance between needle and cambium tissues. Twenty-six of these differentially expressed transcripts had no known similarity/homology. They may represent rare or unique genes.

Transcripts coding for proteins involved in photosynthesis constitute 41% and 4% of the total ESTs, respectively, from the needle and cambium SSH cDNA libraries, compared to 13% from the standard cDNA library prepared using RNA extracted from needles of black spruce (Mann and Rajora unpublished data). Hence, these transcripts had been enriched in the needle SSH cDNA libraries by more than ten folds compared with cambium SSH cDNA libraries and by more than three folds compared with the standard cDNA library from needles. Transcripts involved in energy and metabolism were also enriched by more than two folds and those involved in cellular transport by four folds in the needle cDNA libraries compared with the cambium SSH cDNA libraries (Fig. 1) and the standard needle cDNA library.

The higher abundance of photosynthesis-related transcripts in black spruce needles is quite expected as needles are the primary site of photosynthesis. These results are consistent with similar observations for sunflower and petunia leaves (Fernández *et al*. 2003; Nagy *et al*. 1986). As might be expected, among transcripts of the genes involved in photosynthesis, RuBisCO was the most abundant class of transcripts in the needle SSH cDNA libraries (26%) compared to the cambium SSH cDNA libraries (2.3%), suggesting that RuBisCO genes are differentially expressed between the needle and cambium tissues of black spruce. These results are consistent with those in petunia, where RuBisCO genes were found to be most abundant in leaves (Nagy *et al*. 1986). RuBisCO is the most abundant protein in the leaves of higher plants (Gray and Kekwick 1974). This enzyme occupies a central position in photosynthesis and catalyzes the first major step in carbon fixation, converting inorganic atmospheric carbon dioxide into organic cellular constituents (Spreitzer and Salvucci 2002). It is composed of eight large and eight small

subunits. The small subunits of RuBisCO are encoded by a small multigene family of nuclear genes whereas the large subunits are coded by chloroplast genes (Spreitzer and Salvucci 2002). Our results suggest that the black spruce ESTs coded by RbcS belong to three members of the multigene family compared to six members identified in petunia (Dean *et al*. 1985) as well as common ice plant (*Mesembryanthemum crystallinum*) (DeRocher *et al.* 1993). Since the black spruce samples for construction of cDNA libraries were collected during the day, high abundance of RuBisCO is not surprising given that this enzyme is regulated by light. The other most abundant transcripts found in the needle SSH cDNA libraries were also coding for proteins involved in photosynthesis: photosystem I P700 apoprotein A2, chlorophyll a/b-binding protein Lhcb5, and Rubisco activase. The higher abundance of genes involved in photosynthesis, energy and metabolism, and cellular transport in needles of black spruce is consistent with the functional role of needles in photosynthesis, respiration, transpiration and cellular transport (Hopkins and Hunter 2004).

Conversely, in the cambium SSH cDNA libraries transcripts of the genes involved in disease and stress response and cell structure were most abundant, accounting for 52% of the ESTs. In comparison with the needle SSH cDNA libraries, enrichment of the transcripts of disease and stress responsive genes was more than 21 folds, and that of cell structure genes was three folds in the cambium SSH cDNA libraries (Fig.1). Clearly, our results suggest that the genes involved in disease and stress response and cell structure (including cell division) are more abundantly expressed in cambium than in needles in black spruce. Of course, abundant expression of the cell structure genes in cambium (vascular and cork) is quite expected because cambium is a lateral meristem and produces

secondary growth by cell division activities. Higher abundance of disease and stress responsive genes in cambium is a novel finding and suggests that cellular activities in cambium play a role in disease and stress response mechanisms in black spruce. These findings seem to be consistent with the role of cork cambium in the formation of bark in forest trees. Bark is an effective physical structural barrier against disease and pathogen access into stem.

The transcripts of genes coding for MT proteins were most abundant in the cambium SSH cDNA libraries (Table 7) and were expressed differentially between the needle and cambium tissues in black spruce. This suggests that the cellular activities occurring in the cambium tissue demands higher levels of expression of MT genes. The results are consistent with the abundance of transcripts of MT genes reported in two other spruce species. The MT protein encoding mRNAs were found to be the most abundant of all transcripts in the vegetative bud SSH cDNA libraries of the early- and late flushing Norway spruce families (Yakovlev *et al.* 2006), and were abundantly expressed during somatic embryogenesis of white spruce (Dong and Dunstan 1996). The MT proteins are known to be involved in metal detoxification in plants (Palmiter 1998; Cobbett and Goldsbrough 2002), and belong to a super multigene family (Robinson et al. 1993; Binz and Kagi 2006). Transcript profiling in rice seedlings suggested that the MT proteins might contribute to plant growth in addition to metal detoxification (Maatsumura et al. 1999). Other MT protein functions include cell death (Bhalerao *et al.* 2003), environmental stress (Etscheid *et al.* 1999), drought-stress (Reddy *et al.* 2002), senescence (Yakovlev *et al.* 2006), cell wall lignification and cell elongation (Omann *et al.* 1994; Yu *et al.* 1998). High abundance of transcripts representing MT genes in black spruce cambium, Norway

spruce vegetative buds and white spruce somatic embryogenesis suggest that MT-like proteins may be involved in growth and development in spruce and possibly in other conifers in addition to their known function in stress response and metal detoxification. In white spruce, the MT protein has been fully sequenced and consists of 60 amino acids (Dong and Dunstan 1996). All of the ESTs representing MT gene transcripts in black spruce showed very high (45 to 60 amino acid) sequence similarities with the white spruce or Norway spruce MT proteins. However, the sequence similarities were much lower with other plants. Therefore, all transcripts of MT genes in black spruce likely belong to plant MT-Type 3 of MT multigene family (Binz and Kagi 2006).

**Meristematic activity and similarity in expressed genes in needle and cambium tissues** Although 10% of the 384 transcripts showed differential expression for their abundance between the needle and cambium tissues based on dot blot hybridization, none of the transcripts was found to be specific to any of these tissues. These results, although limited to 384 transcripts, suggest that the same or similar genes were expressed in the needle and cambium tissues of black spruce at the time of sampling. This is most likely to be due to high meristematic activity taking place in both the needle and cambium tissues. At the Petawawa Research Forest where the sampled black spruce trees are located, the active growth period for black spruce new needles is considered to be from mid-May to mid- July, whereas that for cambium from mid-June to mid-September (John Major, personal communication). We collected stem cuttings and needles in the last week of June in 2002, when active growth and meristematic activity occur. It is worth noting that in conifers, matured needles have been found to have a cambial zone of 2-3 cell layers wide, which

seasonally produces secondary phloem in needles (Ewers and Aloni 1987). Meristematic activity has also been detected in leaf blades in *Arabidopsis* (Ha *et al.* 2003). Therefore, the presence of meristematic activity in both the tissues most likely explains the similar expression patterns of transcripts present in needle and cambium tissues. Another possible reason for the absence of gene transcripts expressed only in needles or cambium tissues could be inefficiency of subtraction during SSH library construction. However, large differences in the functional distribution of transcripts (Fig. 1), differential enrichment of transcripts (Tables 6, 7), and higher EST distinctness between the needle and cambium SSH cDNA libraries (Table 3) suggest that the subtraction has worked.

## SSH cDNA libraries for gene discovery and spruce genomic resource

A total of 256 (4.5%) transcripts showed no similarity with the spruce dbEST at NCBI, in spite of the fact 249,705 ESTs were available from various spruce species (April 24, 2006). Most of the existing *Picea* species dbEST is based on ESTs obtained from standard or normalized cDNA libraries. Obviously the SSH cDNA libraries can capture those transcripts, which are not captured in the standard and normalized libraries in spruce. The 256 novel transcripts identified in our study provide an important spruce genome resource. Thus, ESTs from SSH cDNA libraries can provide an important tool and resource for gene discovery in spruce and potentially in other plants.

The information generated from ESTs, gene annotation of transcripts, and differentially expressed transcripts/genes in this study provides an important genomic resource for various genomics studies and applications. These include gene discovery, identification of candidate genes involved in growth and development, single nucleotide

polymorphism (SNP) detection, candidate genes for association mapping, and microsatellite markers.

## Conclusions

This is the first known report on transcriptome profiling of needle/leaves versus cambium in plants. We have demonstrated that genes belonging to certain functional categories are differentially expressed in needle or cambium tissues of black spruce. Genes involved in photosynthesis and energy and metabolism are predominantly expressed in the needle tissues, whereas genes involved in cell structure, disease and stress response mechanisms are predominantly expressed in the cambium tissues. Transcripts coding for ribulose-1, 5- biphosphate carboxylase were most abundant in the needle, and transcripts coding for metallothionein-like proteins were most abundant in the cambium SSH cDNA libraries. The transcript profiles and differential gene expression between needle and cambium tissues of black spruce are consistent with the functional roles of these tissues. SSH cDNA libraries revealed 256 transcripts that were novel to the total spruce EST database at NCBI. Our study provides an important genomic resource for various structural, functional and population genomic studies and applications in spruce and potentially other conifers.

## Acknowledgements

The Principal Investigator of this study is O.P. Rajora. We thank John Major and Debby Barsi for collecting and transporting the plant material, Brent Higgins for helping in clone picking for cDNA libraries, and Paul Stathord for writing a custom perl script. I.K. Mann was partially supported by a Dalhousie University Graduate Scholarship. O.P. Rajora held

the Stora Enso Senior Chair in Forest Genetics and Biotechnology at Dalhousie University, which was supported by Stora Enso Port Hawkesbury Ltd., and holds the Senior Canada Research Chair in Forest and Conservation Genomics and Biotechnology at UNB, which is supported by the Canada Research Chair Program. The research was funded by an NSERC (Natural Sciences and Engineering Research Council of Canada) Strategic Project grant and Genome Canada funding to O.P. Rajora.

## References

Akopian AN, Wood JN (1995) Peripheral nervous system-specific genes identified by subtractive cDNA cloning. J Biol Chem; 270:21264-21270

Bhalerao R, Keskitalo J, Sterky F Erlandsson R, Bjorkbacka H, Birve SJ, Karlsson J, Gardestrom P, Gustafsson P, Lundeberg J, Jansson S (2003) Gene expression in autumn leaves. Plant Physiol 131:430–442

Boss P. K., Vivier M., Matsumoto S., Dry I. B. and Thomas M. R. (2001) A cDNA from grapevine (*Vitis vinifera* L.), which shows homology to AGAMOUS and SHATTERPROOF, is not only expressed in flowers but also throughout berry development. Plant Mol Biol 45: 541–553

Binz P-A, Kagi JHR (2006) Classification of metallothionein. http://www.biochem. unizh.ch/mtpage/classif.html

Brinker M, van Zyl L, Liu W, Craig D, Sederoff RR, Clapham DH, von Arnold S (2004) Microarray Analyses of gene expression during adventitious root development in *Pinus contorta* Plant Physiol 135: 1526–1539

Chang S, Puryear J and Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11:113–116

Cobbett C, Goldsbrough P (2002) Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. Ann Rev Pl Biol 53:159–182

Dean C, van den Elzen P, Tamaki S, Dunsmuir P, Bedbrook J (1985) Differential expression of the eight genes of the petunia ribulose bisphosphate carboxylase small subunit multi-gene family. EMBO J 4:3055-3061

DeRocher EJ, Quigley F, Mache R, Bohnert HJ (1993) The six genes of the Rubisco small subunit multigene family from *Mesembryanthemum crystallinum*, a facultative CAM plant. Mol Gen Genet 239:450–462

Diatchenko L, Lau Y-FC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Pukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci USA 93:6025–6030

Dong J-Z, Dunstan DI (1996) Expression of abundant mRNAs during somatic embryogenesis of white spruce [*Picea glauca* (Moench) Voss]. Planta 199:459 466

Etscheid M, Klümper S, Riesner D (1999) Accumulation of metallothionein-like mRNA in Norway spruce under environmental stress. J Phytopathol 147:207-213

Ewers FW, Aloni R (1987) Seasonal secondary growth in needle leaves of *Pinus strobus*

and *Pinus brutia*. Am J Bot 74:980-987

Ewing B, Green P (1998) Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. Genome Res 8:186-194

Ewing B, Hillier L, Wendl M and Green P (1998) Base-calling of automated sequencer traces using *Phred.* I. Accuracy assessment. Genome Res 8:175-185

Fernández P, Paniego N, Lew S, Hopp HE, Heinz RA (2003) Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project. BMC Genomics 4:40

Foucart C, Paux E, Ladouce N, San-Clemente H, Grima-Pettenati J, Sivadon P (2006) Transcript profiling of a xylem vs phloem cDNA subtractive library identifies new genes expressed during xylogenesis in Eucalyptus New Phytologist 170:739 - June 2006

Gray JC, Kekwick RG (1974) The synthesis of the small subunit of ribulose 1,5- bisphosphate carboxylase in the French bean *Phaseolus vulgaris*. Eur J Biochem 44:491–500

Ha CM, Kim G-T, Kim BC, Jun JH, Soh MS, Ueno Y, Machida Y, Tsukaya H, Nam HG (2003) The BLADE-ON-PETIOLE 1 gene controls leaf pattern formation through the modulation of meristematic activity in Arabidopsis. Development 130: 161-172

Heck GR, Perry SE, Nichols KW, Fernandez DE (1995) AGL15, a MADS domain protein expressed in developing embryos. Plant Cell 7: 1271–1282

Hopkins, WG, Hunter NPA (2004) Introduction to Plant Physiology. John Wiley and Sons, Inc. Hoboken, NJ

Hutchison KW, Harvie PD, Singer PB, Brunner AF, Greenwood MS (1990) Nucleotide sequence of the small subunit of ribulose-1,5-bisphosphate carboxylase from the conifer *Larix laricina*. Plant Mol Biol 14: 281-284

Kanehisa M and Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes.

Nucleic Acids Res 28:27-30

Li H, Gu X, Dawson VL, Dawson TM (2004) Identification of calcium- and nitric oxide- regulated genes by differential analysis of library expression (DAZLE). Proc Natl Acad Sci USA 101:647-652

Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967-971

Lisitsyn N, Lisitsyn N, Wigler M (1993) Cloning the differences between two complex genomes. Science 259:946-951

Mahalingam R, Gomez-Buitrago A, Eckardt N, Nigam Shah N, Guevara-Garcia A, Day P, Raina R and Fedoroff NV (2003) Characterizing the stress/defense transcriptome of Arabidopsis. Genome Biol 4:R20

Matsumura H, Nirasawa S, Terauchi R (1999) Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). Plant J. 20: 719–726.

Morgenstern EK, Wang BSP (2001) Trends in forest depletion, seed supply, and reforestation in Canada during the past four decades. For Chron 6:1014–1021

Nagy F, Fluhr R, Morelli G, Kuhlemeier C, Poulsen C, Keith B, Boutry M, Chua N-H. (1986) The Rubisco small subunit gene as a paradigm for studies on differential gene expression during plant development. Phil Trans Royal Soc Lon 313:409-417

Nuccio ML, Thomas TL (1999) ATS1 and ATS3: two novel embryo-specific genes in

*Arabidopsis thaliana*. Plant Mol Biol 39: 1153–1163

Ohri D and Khoshoo TN (1986) Genome size in gymnosperms. Plant Sys Evol 153:119– 132

Omann F, Beaulieu N, Tyson H (1994) cDNA sequence and tissue-specific expression of an anionic flax peroxidase. Genome, 37: 137–147

Opsahl-Ferstad HG, Le Deunff E, Dumas C, Rogowsky PM (1997) ZmEsr, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. Plant J 12: 235–246

Palmiter RD (1998) The elusive function of metallothioneins. Proc Natl Acad Sci USA 95: 8428-8430

Pavy N, Paule C, Parsons L, Crow JA, Morency MJ, Cooke J, Johnson JE, Noumen E, Guillet-Claude C, Butterfield Y, Barber S, Yang G, Liu J, Stott J, Kirkpatrick R,Siddiqui A, Holt R, Marra M, Seguin A, Retzel E, Bousquet J and MacKay J (2005) Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. BMC Genomics 6:144

Ralph SG, Yueh H, Friedmann M, Aeschliman D, Zeznik JA, Nelson CC, Butterfield YS, Kirkpatrick R, Liu J, Jones SJ, Marra MA, Douglas CJ, Ritland K, Bohlmann J (2006) Conifer defence 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 29:1545-70

Reddy AR, Ramakrishna W, Chandrasekhar A, Nagabhushana I, Ravindra Babu P, Bonaldo MF, Soarrese MB, Bennetzen JL (2002) Novel genes are enriched in normalized cDNA libraries from drought stressed seedlings of indica rice (*Oryza sativa* L.cv.Nagina22) Genome 45: 204-211

Robinson NJ, Tommey AM, Kuske C, Jackson PJ (1993) Plant metallothioneins. Biochem J 295: 1–10

Sablowski RW, Meyerowitz EM (1998) A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA. Cell 92: 93–103

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Spreitzer RJ, Salvucci ME (2002) RUBISCO: Structure, Regulatory Interactions, and Possibilities for a Better Enzyme. Ann Rev Pl Biol 53: 449-475

Viereck LA, Johnston WF (1990) *Picea mariana* (Mill.) B.S.P. - Black spruce. *In* Vol. 1 Conifers. Silvics of North America. Eds. R.M. Burns and B.H. Honkala. US Dept Agric, Forest Service, Agriculture Handbook 654, Washington DC, pp 227-237

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484-487

Walden AR, Walter C, Gardner RC (1999) Genes expressed in *Pinus radiata* male cones include homologs to anther-specific and pathogenesis response genes. Plant Physiol 121:1103–1116

Wang X-L, He RF, He G-C (2005) Construction of suppression subtractive hybridization libraries and identification of brown planthopper-induced gene. J Plant Physiol 162:1254-1262

Woo HH, Brigham LA, Hawes MC (1995) Molecular cloning and expression of mRNAs encoding H1 histone and an H1 histone-like sequences in root tips of pea (*Psium sativum* L.). Plant Mol Biol 28: 1143–1147

Xiong L, Lee MW, Qi M, Yang Y (2001) Identification of defense-related rice genes by suppression subtractive hybridization and differential screening. Mol Plant- Microbe Interact 14:685–692

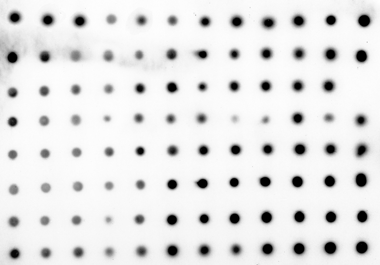
Yakovlev IA, Carl-Gunnar Fossdal, C-G, Johnsen Ø, Junttila O, Skrøppa T (2006) Analysis of gene expression during bud burst initiation in Norway spruce via ESTs from subtracted cDNA libraries. Tree Genet Gen 2: 39–52

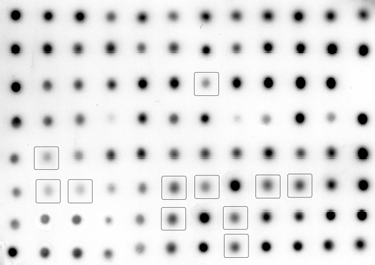
Yu LH, Umeda M, Liu JY, Zhao NM and Uchimiya H (1998) A novel MT gene of rice plants is strongly expressed in the node portion of the stem. Gene 206: 29

**Figure 2** Dot blot hybridization results showing differential expression of transcripts from needle tissue SSH cDNA library, P32N. Dot blots of cDNA clones from P32N cDNA library (A) hybridized to needle cDNA probe, and (B) hybridized to cambium cDNA probe (clones in squares show differential gene expression in terms of transcript abundance: c07,e02, f02, f03 clones have similarity to Ribulose-1, 5- bisphosphate carboxylase, plasma membrane intrinsic protein 2A, Lipase and Ribulose-1, 5- bisphosphate carboxylase, respectively and clones f06, f07, f09, f10, g06, g08 and h08 did not show any BLASTX similarity).

**01 02 03 04 05 06 07 08 09 10 11 12**

# a





**b c d e f g h**