In silico analysis of *Eucalyptus* thioredoxins

Aulus Estevão Barbosa and Paulo Marinho

*Universidade Federal do Rio Grande do Norte, Departamento de Biologia Celular e Genética, Natal, RN, Brazil.*

Abstract

The *Eucalyptus* Genome Sequencing Project (FORESTs), an initiative from the Brazilian ONSA consortium (Organization for Nucleotide Sequencing and Analysis), has achieved the sequencing of 123,889 EST clones from 18 different cDNA libraries. We have investigated the FORESTs data set to identify EST clusters potentially encoding thioredoxins (TRX). Two types of thioredoxin families described in plants, chloroplastic (TRXm/f/x/y) and cytosolic (TRX h), have been found in the transcriptome. Putative typical TRXs have been identified in fifteen clusters, four m-type, seven h-type, two f-type, one cluster for each x/y-types and one putative homologue of the TDX gene from *Arabidopsis thaliana*. One cluster presents an atypical active site WCMPS, different from the conserved WCGPC present in the other 15 clusters, and corresponds to a subgroup of cytosolic thioredoxins. Except in specific libraries from callus, roots, seedlings and wood tissues, thioredoxin deduced ESTs are found in all remaining libraries. According to the calculated frequencies of ESTs, chloroplastic thioredoxins are preferentially present in green tissues such as leaves whilst cytoplasmic thioredoxins are more general but demonstrate elevated frequencies in seedlings and flower tissues. TRX frequency patterns in the *Eucalyptus* transcriptome seem to indicate a good coherence with data from *Arabidopsis thaliana* gene expression.

Key words: Eucalyptus, transcriptome, thioredoxin.

Received: May 28, 2004; Accepted: February 22, 2005.

Introduction

Thioredoxins (TRXs) are ubiquitous disulfide reductases (14kDa average) possessing a characteristic active site WCGPC. The cysteine pair in the active center enables reversible thiol-disulfide exchange reactions and makes the enzyme activity in these proteins. TRX was discovered in *E. coli* and identified as the hydrogen donor for Ribonucleotide reductase (Laurent *et al.*, 1964). Since this discovery, many functions have been attributed to thioredoxins. They are considered to be the major factor responsible for maintaining proteins in their reduced state inside cells and are important agents in transcription factor activation, signaling, apoptosis and reduction of peroxiredoxins, among other processes (Arner and Holmgren, 2000). In plants, in contrast to animals, where a reduced number of genes exist, thioredoxins have been characterized as multigenic families, all encoded by nuclear genes, and classified according to their multigenic compartmentation. Chloroplastic, cytosolic and mitochondrial thioredoxins have been identified (for reviews see Meyer *et al.*, 1999; Schumann and Jacquot, 2000; Meyer *et al.*, 2002). Chloroplastic thioredoxins *m* and *f* were the first characterized and constitute active agents in photosynthesis reactions. They are reduced by a ferredoxin-dependent thioredoxin reductase (FTR) and activate NADP-malate dehydrogenase (NADP-MDH) and fructose-1,6-biphosphatase (FBPase) respectively (Buchanan, 1980; Jacquot *et al.*, 1997). Also in the chloroplast, two other thioredoxins have been described, TRXm (Mestres-Ortega and Meyer, 1999), and TRXY (Lemaire *et al.*, 2003a). TRXm, demonstrating similarity to TRXm, is equally activated by FTR and is the most efficient reductant of 2-Cys Peroxiredoxin (Collin *et al.*, 2003). The recently characterized TRXY, which was discovered in the green alga *Chlamydomonas reinhardtii* and also identified in *Arabidopsis*, is a highly conserved protein in photosynthetic organisms with no specific target described up to now (Lemaire *et al.*, 2003a). Cytosolic thioredoxins (TRXh, *h* stands for heterotrophic) are highly homologous to animal thioredoxins and constitute a system in which NADPH-thioredoxin reductases (NTR) are acting as reductants of thioredoxins (Jacquot *et al.*, 1994). They were characterized in *Arabidopsis thaliana* as a multigenic family (Rivera-Madrid *et al.*, 1995) and after the completion of the genome sequencing, comprise 8 genes in that plant (Gelhaye *et al.*, 2004a; Meyer *et al.*, 2002). Functional tests for TRXh indicate specific roles for these genes in response...
to oxidative stress (Netto, 2001), sulfate assimilation, and cell cycle progression (Mouahbe et al., 1998). Protein targets for this group of thioredoxins have been recently investigated both by biochemical and proteomic approaches (Verdoucq et al., 1999; Yamazaki et al., 2004; Marchand et al., 2004). These studies have confirmed the versatility of these proteins as reductant agents in plant cells. Mitochondrial plant thioredoxins (TRXo, o stands for organelle) have also been characterized in Arabidopsis thaliana (Laloi et al., 2001) and constitute a complete thioredoxin system composed of TRXo and the NADPH-dependent thioredoxin reductase NTRA. Recently, a specific form of TRXh has also been described in mitochondria (Gelhaye et al., 2004b).

The genetic characterization of plant thioredoxins in the last decade could be seen as a consequence of the Arabidopsis thaliana genome-sequencing program (Arabidopsis Genome Initiative, 2000). The establishment of plant transcriptomes and the massive sequence data generated in those systematic genomic approaches can also provide useful information on the levels of expression, structure and evolution of genes. In this paper we investigate the Eucalyptus transcriptome for the presence of thioredoxin gene families. The diversity of thioredoxin genes potentially encoded by the Eucalyptus genome is presented and the frequency of ESTs in the different libraries is discussed.

Material and Methods

Gene identification and sequence analysis

In this work we have investigated the FORESTs database (https://forests.esalq.usp.br/) to identify clusters potentially encoding thioredoxins. Analysis of the sequences (clusters) encoding putative thioredoxins has been performed in two ways. First, we have employed the BLAST (Altschul et al., 1997) service available on the FORESTs web page to align the Arabidopsis thaliana thioredoxin protein sequences against Eucalyptus clusters. The nucleotide clusters, representing the best hits after TBLASTN searches, were then analyzed for putative coding sequences using the ORF Finder program from NCBI (http://www.ncbi.nlm.nih.gov/). The putative amino acid sequences obtained have been formatted by the Translate tool from Infobiogen sequence analysis resources on the web (http://www.infobiogen.fr). These deduced sequences were then used for BLASTP alignments to identify the genes. Multiple sequence alignments using the CLUSTALW (Thompson et al., 1994) program from NPSA (http://npsa-pbil.ibcp.fr/cgi-bin/align_clustalw.pl) were then performed. The second method employed to obtain sequences, or just to confirm an analysis, was a direct search on the FORESTs web site using “thioredoxin” as the search keyword. This search gives a general view of the clusters related to the thioredoxins in the FORESTs data and the best hits resulting from automatic BLASTX alignment.

The number of reads encoding each specific thioredoxin cluster was then used to calculate the frequencies of reads in the libraries.

A Phylogenetic tree has been generated using CLUSTALX (Thompson et al., 1994) to create an alignment in phylip format; the amino acid sequences were then analyzed using the NEIGHBOR program from the PHYLIP package. PFAM distances were obtained using the PROTDIST program from the same interface. Bootstrap assessment of the tree in neighbor-joining analysis was performed with the SEQBOOT program and the tree was displayed using TREEVIEW.

Prediction of subcellular localization of proteins has been made using PREDOTAR program from http://genoplante-info.infobiogen.fr/predotar/ site.

cDNA libraries

Full description of cDNA libraries used in the FORESTs project can be found on the web site referenced above. It comprises 13 libraries made from different tissues of Eucalyptus grandis including the major organs (leaves, flowers, roots, stem) and five seedling libraries from E. grandis, E. globulus, E. saligna, E. urophylla and from E. camaldulensis grown in the dark. The cluster names contain information about the tissue, where the cDNAs were prepared, the laboratory responsible for the sequence submission and the plate and clone numbers.

Results and Discussion

The Arabidopsis thaliana genome-sequencing program has revealed the diversity of thioredoxin genes defining different families for these enzymes in plants. Over 20 genes have been identified in Arabidopsis and classified as chloroplastic, cytosolic and mitochondrial thioredoxins. Our in silico analysis of the Eucalyptus transcriptome confirms that diversity of gene families allows the identification of clusters that encode putative genes for chloroplastic and cytosolic thioredoxins. The transcriptome includes at least 15 clusters that encode putative typical thioredoxins and one partial clone that encodes a putative thioredoxin like protein homologous to the Arabidopsis thaliana Tetratricoredoxin (TDX) gene (Vignols et al., 2003).

BLASTP alignments of deduced amino acid sequences from EST clusters allowed us to find homologies to known proteins. Table 1 summarizes these results. For each Eucalyptus cluster, the most homologous protein has been described, as well as the percentage of amino acid identity between the two proteins. Also we have performed protein subcellular predictions to define more precisely the type of thioredoxin identified. On the basis of the best homologies with database proteins, we have classified four clusters for TRXm, seven clusters for TRXh, two clusters for TRXf and one cluster for each TRXx and TRXγ respec-
Putative TRXs m clusters demonstrates the best homology with thioredoxins from *Pisum sativum* and *Arabidopsis thaliana*, ranging from 53 to 76%. TRXs h clusters show homology with proteins from different plants coding sequence identities ranging from 55 to 78%. TRXs f clusters demonstrate the highest percentages of identity, 86% and 79%, to f thioredoxins from *Mesembryanthemum crystallinum* and *Arabidopsis thaliana* respectively. The identity of the TRXs cluster to the *Arabidopsis thaliana* TRXX was 67%. The cluster encoding TRXY, EGRFRT3019G02.g, was identified as a thioredoxin-like protein in *Arabidopsis thaliana* but may in fact correspond to a homologue of the AtTRXy1 (Lemaire et al., 2003a).

Subcellular localization predictions are also shown in Table 1. In general, the predictions agree with the type of thioredoxin found by the best homologies using BLASTP. This is the case in the defined chloroplastic thioredoxins, TRXf, TRXx and TRXy, here composed of the EGEQLV2006B06.g, EGRFRT3019G02.g clusters respectively. The same can be confirmed for the m thioredoxins predicted as plastidial proteins. For these clusters, high percentages are obtained be confirmed for the m thioredoxins predicted as plastidial encoding h type thioredoxins EGUTSL1040E09.g, EGR FR3T350D04.g, EGBMST2012A11.g, EGJFFB1008E10.g. Preliminary analysis of the FORESTs data has shown that most of the ESTs potentially encoding thioredoxins contain the coding sequence of the genes covering the ORF regions except for two clusters, EGEZFB1116E11.g and EGQHFB1231E06.g. If the *in silico* identification is compared with the *Arabidopsis* set of genes, we can say that a similar situation in terms of number and types of TRXs is observed, confirming the complexity of thioredoxin families also in *Eucalyptus* (Meyer et al., 2002). We show in Figure 1 a sequence alignment of the 15 *Eucalyptus* putative thioredoxins and the *E. coli* thioredoxin (Atrx). We can identify two conserved regions characteristic of thioredoxins, the active site WCGPC and the region corresponding to a hydrophobic surface of amino acids at different locations in the primary structure (Gly33, Pro34, Pro76, Gly92, and Ala93). This second region is implicated in the interaction of thioredoxins with other proteins (Rivera-Madrid et al., 1995). Specific domains or conserved amino acids can be also observed, for example,

**Table 1 - Identification of thioredoxin clusters.**

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Identification</th>
<th>Accession number</th>
<th>e-value</th>
<th>Identities</th>
<th>Predotar prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TRX m</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGBMST6205D10.g</td>
<td>thioredoxin m3 [A. thaliana]</td>
<td>gi</td>
<td>15226561</td>
<td>1e-32</td>
<td>68/127 (53%)</td>
</tr>
<tr>
<td>EGUTSL6218F08.g</td>
<td>thioredoxin m2 [P. sativum]</td>
<td>gi</td>
<td>15594012</td>
<td>2e-39</td>
<td>68/130 (52%)</td>
</tr>
<tr>
<td>EGEQLV2006B06.g</td>
<td>thioredoxin m [P. sativum]</td>
<td>gi</td>
<td>13880888</td>
<td>5e-48</td>
<td>88/115 (76%)</td>
</tr>
<tr>
<td>EGEZLV2004B03.g</td>
<td>thioredoxin m2 [P. sativum]</td>
<td>gi</td>
<td>15594012</td>
<td>1e-52</td>
<td>97/153 (63%)</td>
</tr>
<tr>
<td><strong>TRX x</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGJMF11092E10.g</td>
<td>thioredoxin [A. thaliana]</td>
<td>gi</td>
<td>15242525</td>
<td>1e-43</td>
<td>84/133 (63%)</td>
</tr>
<tr>
<td>EGUTSL1040E09.g</td>
<td>thioredoxin h [H. brasiliensis]</td>
<td>gi</td>
<td>14928460</td>
<td>1e-37</td>
<td>82/126 (65%)</td>
</tr>
<tr>
<td>EGQHFB1231E06.g</td>
<td>thioredoxin h [P. sativum]</td>
<td>gi</td>
<td>13624884</td>
<td>3e-09</td>
<td>35/63 (55%)</td>
</tr>
<tr>
<td>EGRFR3T350D04.g</td>
<td>thioredoxin H2 [I. batatas]</td>
<td>gi</td>
<td>33621080</td>
<td>3e-34</td>
<td>68/107 (63%)</td>
</tr>
<tr>
<td>EGBMST2108E08.g</td>
<td>thioredoxin h [P. sativum]</td>
<td>gi</td>
<td>13624884</td>
<td>4e-40</td>
<td>78/118 (66%)</td>
</tr>
<tr>
<td>EGBMST2012A11.g</td>
<td>thioredoxin h-type I[N. tabacum]</td>
<td>gi</td>
<td>2671214</td>
<td>9e-41</td>
<td>82/114 (71%)</td>
</tr>
<tr>
<td>EGJFFB1008E10.g</td>
<td>thioredoxin h-type [R. communis]</td>
<td>gi</td>
<td>11135282</td>
<td>2e-46</td>
<td>90/115 (78%)</td>
</tr>
<tr>
<td><strong>TRX f</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGEQFB1249F03.g</td>
<td>thioredoxin F precursor [M. crystallinum]</td>
<td>gi</td>
<td>11135152</td>
<td>1e-48</td>
<td>94/109 (86%)</td>
</tr>
<tr>
<td>EGEZLV1206H11.g</td>
<td>thioredoxin f2 [A. thaliana]</td>
<td>gi</td>
<td>1523734</td>
<td>6e-50</td>
<td>93/117 (79%)</td>
</tr>
<tr>
<td><strong>TRX y</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGJELV2261D10.g</td>
<td>thioredoxin x [A. thaliana]</td>
<td>gi</td>
<td>65396160</td>
<td>4e-42</td>
<td>86/127 (67%)</td>
</tr>
<tr>
<td><strong>TRX y</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGRFR3T3019G02.g</td>
<td>thioredoxin-like [A. thaliana]</td>
<td>gi</td>
<td>6143893</td>
<td>1e-44</td>
<td>91/109 (83%)</td>
</tr>
</tbody>
</table>
the KDEL motif or the Trp residue (W13 in *Eucalyptus*), both characteristic of thioredoxins.

Concerning the cytosolic TRXs, the variant of the active center WCPPC identified by Rivera-Madrid et al. (1995) in three TRXh of *Arabidopsis* (AtTRX3, AtTRX4, AtTRX5) has not been detected in our analysis. This is surprising because of the considerable level of expression of these genes in *Arabidopsis* compared to the other members of thioredoxin h family (Reichheld et al., 2002).

An atypical WCMPS active site has been found in the Cluster EGUTSL1040 (Figure 1). This cluster comprises 8 reads from 6 different libraries and demonstrates the highest homology score to a putative thioredoxin pseudogene described in *Hevea brasiliensis* harboring an active site WCIPS (Chow, 1999). It may in fact constitute a member of TRXh-like subgroup of thioredoxins described in poplar that share the same active site and that have been biochemically characterized (Gelhaye et al., 2003a). Also in the *Arabidopsis thaliana* genome, two genes present the active site CxxS, AtCxxS1 and AtCxxS2 which have WCLPS and WCIPS sites respectively (Gelhaye et al., 2003a). We have decided to keep the inferred protein sequence of this EST in our phylogenetic tree (Figure 2) where we can observe that it roots in the thioredoxins h group.

The phylogenetic tree also suggests the distribution of thioredoxins according to their gene families. We can observe that all the inferred ORFs encoding *Eucalyptus* thioredoxins are rooted in the tree beside their counterparts in *Arabidopsis thaliana*.

Expression of thioredoxins clusters

Putative thioredoxin genes of *Eucalyptus*, according to their calculated EST frequencies, are expressed differently in 13 of the 18 cDNA libraries from FORESTs data (Figure 3). No read encoding thioredoxin has been found in the CL2, RT6, SL8 and WD2 libraries. The other libraries demonstrate that thioredoxin reads vary according to the type of thioredoxin and the individual cluster. When we globally analyze the expression frequencies of TRXs clusters in *Eucalyptus* and compare these with published results using Northern blotting or RT-PCR techniques in other plants, we observe that the transcriptome can represent a deduced coherent overview of thioredoxin expression. This means that the tissue specificity of expression according to the group of TRXs seems to be in accord with the results described in literature. For example, the expression of putative chloroplastic thioredoxins in green tissues or the presence of TRXs transcripts preferentially in non-green tissues for the cytosolic group seems to coincide. Another interesting example is the important expression of TRXs in early seedling development, which is well documented in the literature (Kobrehel et al., 1992; Wong et al., 2002; Marx et al., 2003). Here we have observed high level of expression of thioredoxin genes in seedling libraries (Figure 3).

Chloroplastic thioredoxins

Chloroplastic TRXn is present in the *Eucalyptus* transcriptome with at least four putative clusters. The pres-
ence of their correspondent reads is observed in five libraries (Figure 3) composed mostly from green tissues (FB1, LV2, LV3, SL1, ST6). This result may be expected due to the function of TRXs in the photosynthesis process. The read frequencies restricted to the leaves, flowers and seedlings correspond to the major tissues where the expression of TRXm has been reported in Arabidopsis (Mestre-Ortega and Meyer, 1999). These authors have found, by Northern blotting, an expression profile of TRXm genes (Athm1, Athm2, Athm4) and TRXx (Athx) in seeds, roots, stem, leaves, flowers and callus tissues but with high RNA levels restricted to leaves, flower buds and seedlings. These tissues have allowed the cloning of cDNAs encoding TRXm in Eucalyptus. In the same way, TRXx cluster (EGJELV2261D10.g) demonstrates reads with elevated frequencies in the libraries from leaves (LV2) and seedlings (SL1).

The two clusters of TRX/ (EGEPFB1249F03.g, EGEZLV1206H11.g) present read frequencies and deduced expression following the TRXm reads. They have been detected in flower buds, leaves and seedling tissues in FB1, LV2, LV3 and SL1 libraries. Pagano et al. (2000) have demonstrated the expression of TRX/ in pea by RT-PCR and reported the presence of transcripts in leaves and also in non-photosynthetic organs like roots suggesting a role of TRX/ other than the FBPase modulation.

Finally we can remark on the specificity of TRXs m/f/x for the SL1 library. No EST has been found in the other five seedling libraries. The SL1 library differs from the others in that it represents cDNAs from seedlings culti-

\[ \text{Figure 2 - Phylogenetic tree of Eucalyptus putative thioredoxins and Arabidopsis thaliana thioredoxins.} \]

\[ \text{Arabidopsis thioredoxins:} \text{ AtTRXh1 (P29448), AtTRXh2 (S58123), AtTRXh3 (S58123), AtTRXh3 (S58118), AtTRXh4 (S58119), AtTRXh5 (S58120), AtTRX7 (AAD3916g), AtTRX9 (AAG 51343), AtTRXm1 (04873173), AtTRXm2 (AAFI5949), AtTRXm3 (AAFI5952), AtTRXm4 (Q9SSE6v), AtTRX1 (AAFI5952), AtTRXo1 (AAC 12840), AtTRXo2 (AAK8391t), AtTXy1 (AAF04439), AtTXy2 (AAAF58105). Eucalyptus clusters: TRXs h, EGRFRT3350D04.g, EGJHFB1081E08, EGBST2012A11g, EGBSM1201E08, EGBTSL1040E09, EGRX16g, EGRX6g.} \]

\[ \text{Eucalyptus clusters:} \text{ TRXs h, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.g, EGRFRT3350D04.
Figure 3 - Relative abundance of Eucalyptus thioredoxin ESTs in different cDNA libraries. Clusters with the same colors belong to the same group. TRXm, red; TRXh, blue; TRXf, yellow; TRXx, brown; TRXy, gray. cDNA libraries: BK1 - bark, alburnum, heartwood and pith from group. TRXent cDNA libraries. Clusters with the same colors belong to the same family (Nelson et al., 2003) in Arabidopsis (Rivera-Madrid et al., 1995). Among the thioredoxins expressed in plants, the cytosolic are the best characterized. Reichheld et al., (2002) have employed multiple approaches (Northern blotting, RT-PCR, GUS-fusion plants) in attempts to obtain information that could be useful to reveal the function of the 8 genes from Arabidopsis. They have shown that the expression occurs in most tissues studied and it seems clear that the expression is divergent depending on the gene studied. If we compare these results with the frequencies and tissue specificity of TRXh in the case of Eucalyptus, it is possible to say that it is comparable. However, experimental analysis should be conducted to confirm the data.

Analysis of the TDX cluster

We have identified the Cluster EGEZFB1116E11.g as a homologue of a TDX gene characterized by Vignols et al. (2003) in Arabidopsis thaliana. This gene encodes a singular bipartite protein composed by a carboxyl-terminal thioredoxin domain and the amino-terminal domain containing three tetratricopeptide repeats similar to those found in the Hip protein of rat and human (Hohfeld J et al., 1995). The TDX name has been attributed to Tetratricopeptide-domain-containing thioredoxin. Functional characterization of this gene reveals a disulfide reduction both in vivo and in vitro by the thioredoxin domain; whereas, the amino terminus presents an interaction with the yeast Hsp70 Ssb2 chaperone, a member of the heat shock protein family (Nelson et al., 1992). In this case, the redox domain will be acting as a redox switch that turns the complex with Ssb2 on and off.

Cluster EGEZFB1116E11.g found in the Eucalyptus transcriptome might be a partial cDNA encoding a homologue of TDX in Eucalyptus. It shows 69% homology with the Arabidopsis thaliana TDX in its thioredoxin domain. Figure 4 presents a multiple sequence alignment of the EGEZFB1116E11.g deduced protein and the TDX genes from Arabidopsis thaliana and Nicotiana tabacum. We can experimental data has to be made to confirm that assumption.

Cytosolic thioredoxins

We have identified seven clusters that, on the basis of the sequences and the phylogenetic analysis, could correspond to h thioredoxins. These clusters present an expression profile more distributed than chloroplastic or mitochondrial thioredoxins and are found preferentially in young and proliferating tissues from a wide range of libraries (Figure 3). We can note, for example, the presence of TRXx reads in the libraries of seedlings, flower buds, stems, bark and callus. They are absent in leaves. Some considerations can be given to the presence of TRXs transcripts in certain libraries. The presence of thioredoxins h in callus, for example, has been reported since the identification of the first genes in tobacco (Brugidou et al., 1993) and later in Arabidopsis (Rivera-Madrid et al., 1995). Among the thioredoxins expressed in plants, the cytosolic are the best characterized. Reichheld et al., (2002) have employed multiple approaches (Northern blotting, RT-PCR, GUS-fusion plants) in attempts to obtain information that could be useful to reveal the function of the 8 genes from Arabidopsis. They have shown that the expression occurs in most tissues studied and it seems clear that the expression is divergent depending on the gene studied. If we compare these results with the frequencies and tissue specificity of TRXh in the case of Eucalyptus, it is possible to say that it is comparable. However, experimental analysis should be conducted to confirm the data.

vated in the dark but exposed to light 3 h before RNA extract. The other libraries represent cDNAs from seedlings cultivated completely in the dark. This result could be related to the well-established role of chloroplastic thioredoxins as light-dependent enzymes in photosynthesis (Collin et al., 2003, Meyer et al., 2002).

In the case of TRXy, the read frequencies are restricted to leaves (LV1, LV2, LV3), flower buds (FB1) and differ from the other chloroplastic TRXs in a stem library (ST6). The expression of TRXy remains to be established in plants. The expression data available for TRXy is the result of an in silico analysis from ESTs of Chlamydomonas reinhardtii presented by Lemaire et al., (2003b). TRXy ESTs are found in the libraries prepared from cultures submitted to nutritional stress conditions and are poorly or undetected in normal culture conditions. In our case, the libraries that contain TRXy ESTs could be also considered a stress library, for example the LV3, which is prepared from leaves attacked by Thyrinteina for 7 days; however,
observe the very conserved region comprising the thioredoxin domain in the three proteins. The expression profile observed by the frequency analysis shows the presence of Cluster EGEZFB116E11.g restricted to the libraries from flower buds (FB1) and seedlings (SL4) at low levels.

Conclusions

In this paper we have shown that the Eucalyptus transcriptome encodes at least 15 putative thioredoxin ESTs and one cluster encoding a putative TDX gene. Two families of thioredoxins present in the transcriptome have been described: chloroplastic and cytosolic. The number of cDNAs and the type of thioredoxins encoded seems to be similar to the total number of genes found in Arabidopsis thaliana. Concerning the read frequencies, we can presume that the libraries that compose the Eucalyptus transcriptome provide an overview and a coherent expression profile of thioredoxins. Nevertheless, other putative thioredoxins and mainly thioredoxin-like transcripts present in the data bank remain to be studied.

Acknowledgments

We are very grateful to Yves Meyer and Jean-Philippe Reichheld (CNRS/ Université de Perpignan) for...
reviewing the manuscript. We also thank to Sarra Jamieson (Cambridge Institute for Medical Research) and Marcia Pedrini (UFRN-Brazil) for reading the paper. A.E.B was supported by Capes - Comissão de Aperfeiçoamento de Pessoal de Nível Superior/Brazil.

References


Associate Editor: Carlos F.M. Menck