



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 (TRX $m/f/x/y$) 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, TRX x (Mestres-Ortega and Meyer, 1999), and TRX y (Lemaire *et al.*, 2003a). TRX x , demonstrating similarity to TRX m , is equally activated by FTR and is the most efficient reductant of 2-Cys Peroxiredoxin (Collin *et al.*, 2003). The recently characterized TRX y , 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 (TRX h , 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 (Gelhay *et al.*, 2004a; Meyer *et al.*, 2002). Functional tests for TRX h indicate specific roles for these genes in response

to oxidative stress (Netto, 2001), sulfate assimilation, and cell cycle progression (Mouaheb *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 (TRX_o, *o* stands for organelle) have also been characterized in *Arabidopsis thaliana* (Laloi *et al.*, 2001) and constitute a complete thioredoxin system composed of TRX_o and the NADPH-dependent thioredoxin reductase NTRA. Recently, a specific form of TRX_h 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 databank proteins, we have classified four clusters for TRX_m, seven clusters for TRX_h, two clusters for TRX_f and one cluster for each TRX_x and TRX_y respec-

tively. Putative TRXs *m* clusters demonstrate 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 TRXx cluster to the *Arabidopsis thaliana* TRXx was 67%. The cluster encoding TRXy, EGRFR T3019G02.g, was identified as a thioredoxin-like protein in *Arabidopsis thaliana* but may in fact correspond to a homologue of the AfTRXy1 (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 EGEPFB1 249F03.g, EGEZLV1206H11.g, EGJELV2261D10.g, and EGRFRT3019G02.g clusters respectively. The same can be confirmed for the *m* thioredoxins predicted as plastidial proteins. For these clusters, high percentages are obtained using the Predotar program that predicts a plastidial localization. Cluster EGEQLV2006B06.g, here kept in Table 1 in the *m* type group because of the homology to *m* thioredoxin, is actually predicted by TargetP (Emanuelsson *et al.*,

2000) as a mitochondrial protein. Those clusters that have been predicted “elsewhere” by Predotar have also been submitted to TargetP program. This analysis (data not shown) indicates a cytoplasmic localization for the clusters encoding *h* type thioredoxins EGUTSL1040E09.g, EGR FRT3350D04.g, EGABSL7210E08.g, EGBMST2012A 11.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.

Clusters	Identification	Accession number	e-value	Identities	Predotar prediction
TRX <i>m</i>					
EGBMST6205D10.g.	thioredoxin m3 [<i>A. thaliana</i>]	gi 15226561	1e-32	68/127 (53%)	0.60 plastid
EGUTST6218F08.g	thioredoxin m2 [<i>P. sativum</i>]	gi 15594012	2e-39	68/130 (52%)	0.78 plastid
EGEQLV2006B06.g	thioredoxin m [<i>P. sativum</i>]	gi 1388088	5e-48	88/115 (76%)	0.96 elsewhere
EGEZLV2004B03.g	thioredoxin m2 [<i>P. sativum</i>]	gi 15594012	1e-52	97/153 (63%)	0.95 plastid
TRX <i>h</i>					
EGJMF1092E10.g	thioredoxin [<i>A. thaliana</i>]	gi 15242525	1e-43	84/133(63%)	0.20 possible plastid
EGUTSL1040E09.g	thioredoxin h [<i>H. brasiliensis</i>]	gi 4928460	3e-37	82/126 (65%)	0.97 elsewhere
EGQHFB1231E06.g	thioredoxin h [<i>P. sativum</i>]	gi 13624884	3e-09	35/63 (55%)	no terminal met
EGRFRT3350D04.g	thioredoxin H2 [<i>I. batatas</i>]	gi 33621080	3e-34	68/107 (63%)	0.98 elsewhere
EGABSL7210E08.g	thioredoxin h [<i>P. sativum</i>]	gi 13624884	4e-40	78/118 (66%)	0.97 elsewhere
EGBMST2012A11.g	thioredoxin h-type 1[<i>N. tabacum</i>]	gi 267124	9e-41	82/114 (71%)	0.98 elsewhere
EGJFFB1008E10.g	thioredoxin h-type [<i>R. communis</i>]	gi 11135282	2e-46	90/115 (78%)	0.96 elsewhere
TRX <i>f</i>					
EGEPFB1249F03.g	thioredoxin F precursor [<i>M. crystallinum</i>]	gi 11135152	1e-48	94/109 (86%)	0.97 plastid
EGEZLV1206H11.g	thioredoxin f2 [<i>A. thaliana</i>]	gi 15237341	6e-50	93/117 (79%)	0.97 plastid
TRX <i>x</i>					
EGJELV2261D10.g	thioredoxin x [<i>A. thaliana</i>]	gi 6539616	4e-42	86/127 (67%)	0.83 plastid
TRX <i>y</i>					
EGRFRT3019G02.g	thioredoxin-like [<i>A. thaliana</i>]	gi 6143893	1e-44	91/109 (83%)	0.89 plastid

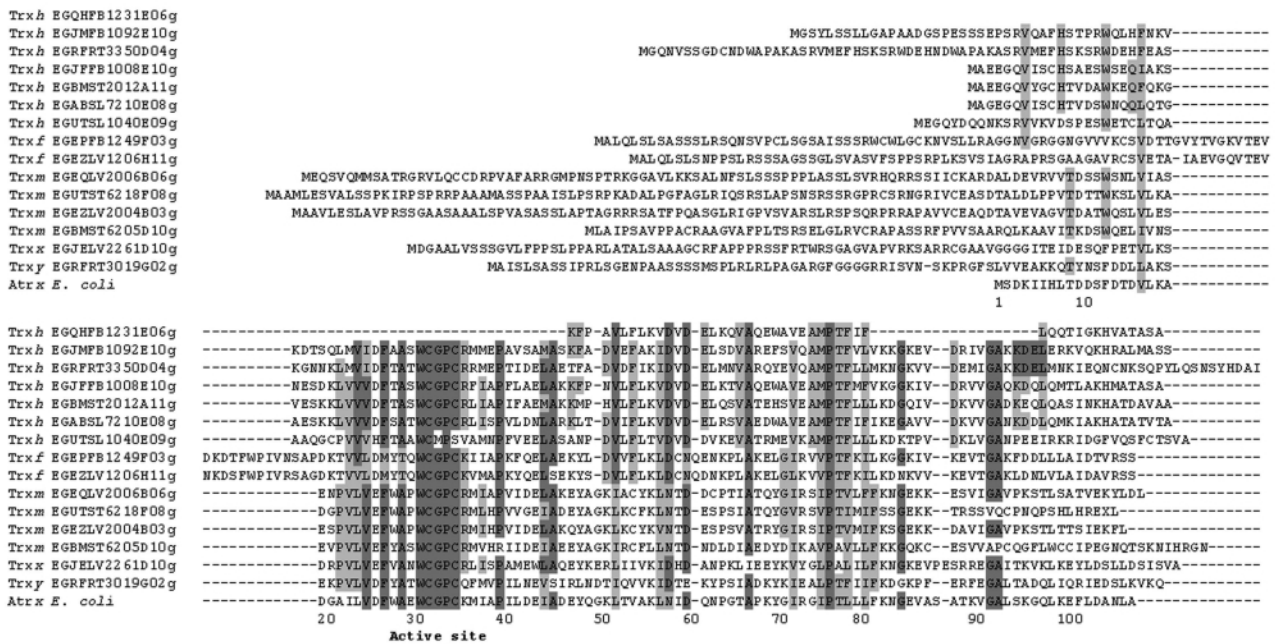


Figure 1 - Multiple sequence alignment of deduced amino acid sequences of *Eucalyptus* thioredoxins and *E. coli* thioredoxin. TRXs *h*, EGQHFB1231 E06.g, EGJMFB1092E10g, EGRFRT3350D04g, EGJFFB1008E10g, EGBMST2012A11g, EGABSL7210E08g, EGUTSL1040E09g; TRXs *f*, EGE PFB1249F03g, EGEZLV1206H11g; TRXs *m*, EGEQLV2006B06g, EGUTST6218F08g, EGEZLV2004B03g, EGBMST6205D10g; TRXx, EGJEL V2261D10g; TRXy, EGRFRT3019G02g, Atr. *E. coli* thioredoxin (M54881).

the KDEL motif or the Trp residue (W13 in *Eucalyptus*), both characteristic of *h* 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 (Gelhay *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 (Gelhay *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 library 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 thioredoxins genes in seedling libraries (Figure 3).

Chloroplastic thioredoxins

Chloroplastic TRXm is present in the *Eucalyptus* transcriptome with at least four putative clusters. The pres-

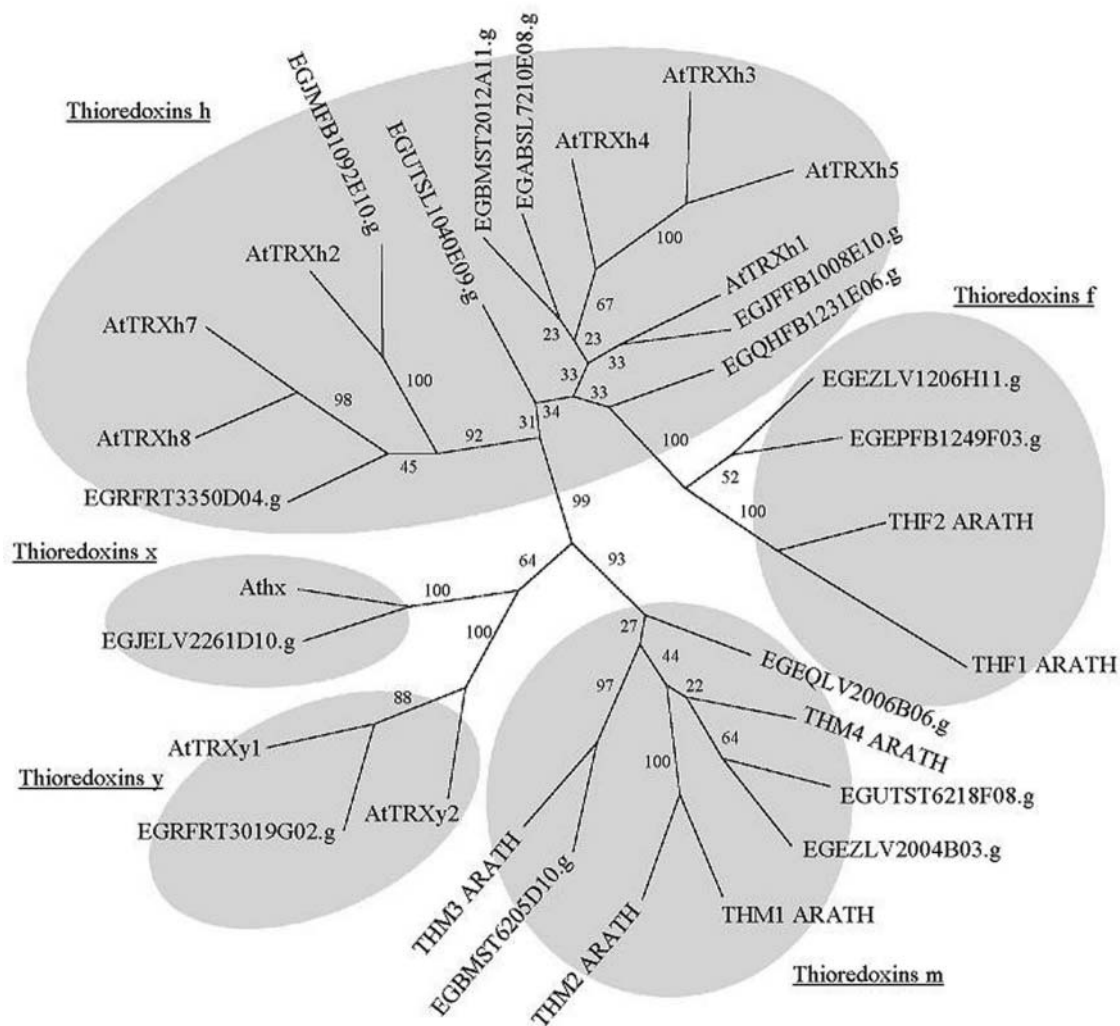


Figure 2 - Phylogenetic tree of *Eucalyptus* putative thioredoxins and *Arabidopsis thaliana* thioredoxins. *Arabidopsis* thioredoxins: AtTRXh1 (P29448), AtTRXh2 (S58123), AtTRXh3 (S58123), AtTRXh3 (S58118), AtTRXh4 (S58119), AtTRXh5 (S58120), AtTRX7 (AAD39316), AtTRX9 (AAG 51343), AtTRXm1 (0487137), AtTRXm2 (AAF15949), AtTRXm3 (AAF15952), AtTRXm4 (Q9SSEV6), AtTRXx (AAF15952), AtTRXo1 (AAC128 40), AtTRXo2 (AAK83918), AtTRXy1 (AAF04439), AtTRXy2 (AAM91085). *Eucalyptus* clusters: TRXs *h*, EGQHFB1231E06.g, EGJMFB1092E10.g, EGRFRT3350D04.g, EGJFFB1008E10.g, EGBMST2012A11.g, EGABSL7210E08.g, EGUTSL1040E09.g; TRXs *f*, EGEPFB1249F03.g, EGEZLV1206 H11.g; TRXs *m*, EGEQLV2006B06.g, EGUTST6218F08.g, EGEZLV2004B03.g, EGBMST6205D10.g; TRXx, EGJELV2261D10.g; TRXy, EGRFRT 3019G02.g.

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 *m* 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 (EGJEL

V2261D10.g) demonstrates reads with elevated frequencies in the libraries from leaves (LV2) and seedlings (SL1). The two clusters of TRXf (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 TRXf 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 TRXf 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-

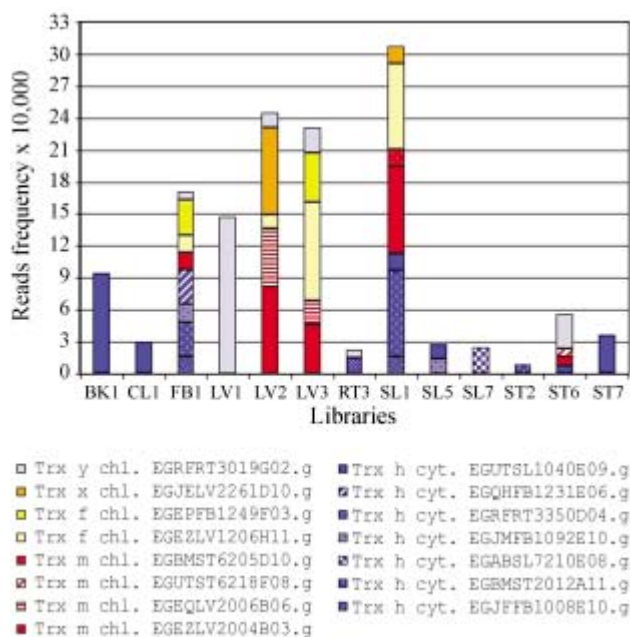


Figure 3 - Relative abundance of *Eucalyptus* thioredoxin ESTs in different cDNA libraries. Clusters with the same colors belong to the same group. TRX_m, red; TRX_h, blue; TRX_f, yellow; TRX_x, brown; TRX_y, gray. cDNA libraries: BK1 - bark, album, heartwood and pith from *E. grandis* 8 years old; CL1 - callus of *E. grandis* grown in the presence of light; FB1 - flower buds, flowers and fruits; LV1 - leaves from plantlets; LV2 - leaves from adult trees deficient in phosphorus and boron, sensitive to rust and susceptible to canker; LV3 - leaves colonized by *Thyrinteina* for 7 days; RT3 - roots from greenhouse plants; SL1 - seedlings from *E. grandis* grown in the dark and exposed to light for 3 h before RNA extractions; SL4 - seedlings from *E. globulus* grown in the dark; SL5 - seedlings from *E. saligna* grown in the dark; SL6 - seedlings from *E. urophylla* grown in the dark; SL7 - seedlings from *E. grandis* grown in the dark; ST2 - stem from plants susceptible to water deficit (0.6 to 2.0 kb insert); ST6 - stem from plants susceptible to water deficit (0.8 to 3.0 kb insert); ST7 - stem from trees resistant and susceptible to hoarfrost.

vated in the dark but exposed to light 3 h before RNA extraction. 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 TRX_y, 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 TRX_y remains to be established in plants. The expression data available for TRX_y is the result of an *in silico* analysis from ESTs of *Chamydomonas reinhardtii* presented by Lemaire *et al.*, (2003b). TRX_y 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 TRX_y ESTs could be also considered a stress library, for example the LV3, which is prepared from leaves attacked by *Thyrinteina* for 7 days; however,

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 TRX_h 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 TRX *h* 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 EGEZFB116E11.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 EGEZFB116E11.g deduced protein and the TDX genes from *Arabidopsis thaliana* and *Nicotiana tabacum*. We can

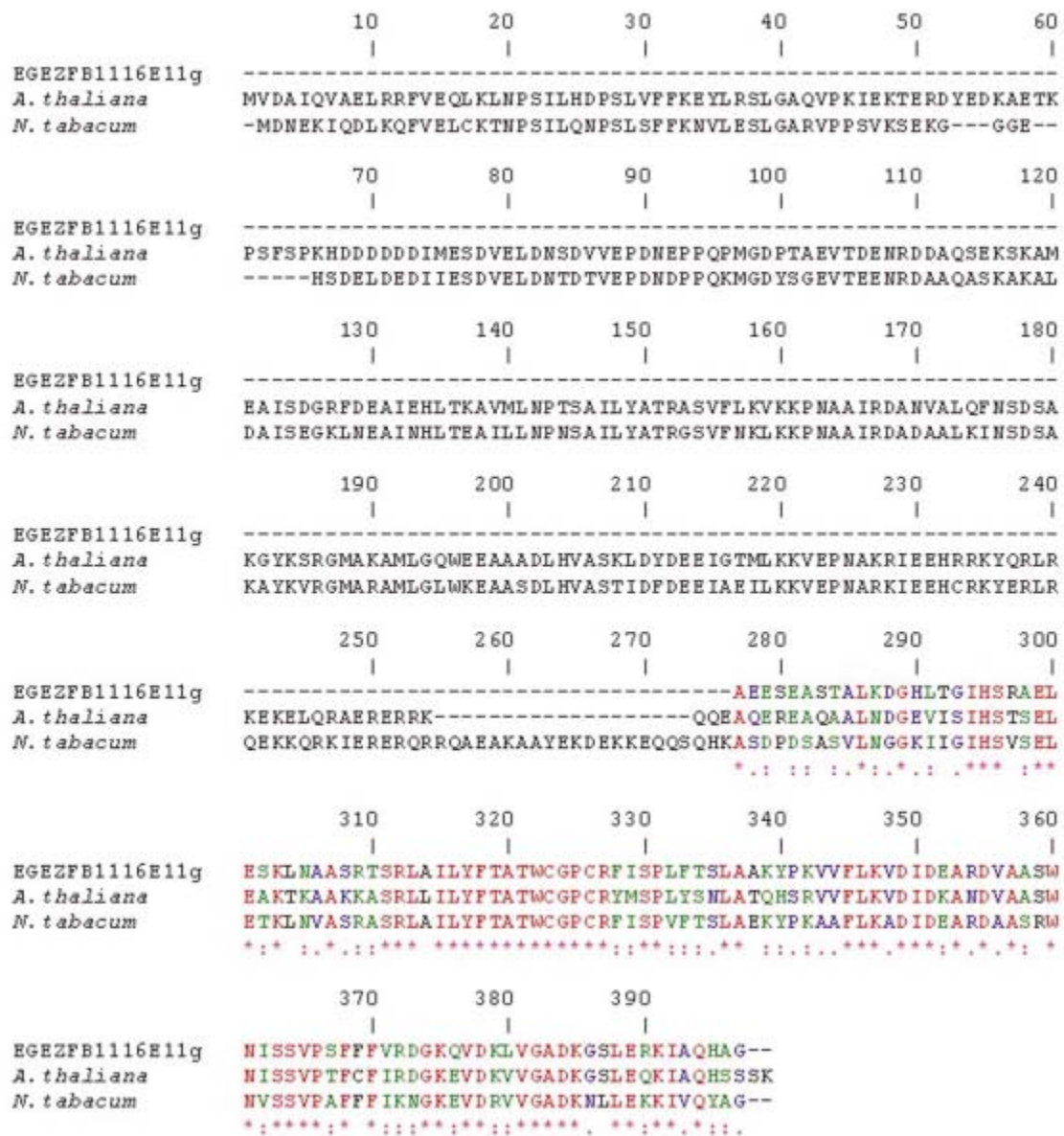


Figure 4 - Multiple sequence alignment of TDX proteins. Deduced amino acid sequences of *Arabidopsis thaliana* (AAL5485), *Nicotiana tabacum* (AAL54858) TDX proteins and the partial clone EGEZFB1116E11.g.

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 EGEZFB1116E11.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

- Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 409:796-815.
- Arner ESJ and Holmgren A (2000) Physiological functions of thioredoxin and thioredoxin reductase. *Eur J Biochem* 267:6102-6109.
- Altschul SF, Madden TL, Schäffer AA, Zhang ZZ, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.
- Balmer Y, Koller A, del Val G, Manieri W, Schurmann P and Buchanan BB (2003) Proteomics gives insight into the regulatory function of chloroplast thioredoxins. *Proc Natl Acad Sci U SA* 100:370-5.
- Brugidou C, Marty I, Chartier Y and Meyer Y (1993) The *Nicotiana tabacum* genome encodes two cytoplasmic thioredoxin genes which are differently expressed. *Mol Gen Genet* 238:285-93.
- Buchanan BB (1980) Role of light in the regulation of chloroplast enzymes. *Annu Rev Plant Physiol* 31: 341-374.
- Chow KS (1999) Cloning of a thioredoxins h cDNA from *Hevea brasiliensis*. *Plant Physiol* 120:339.
- Collin V, Issakidis-Bourguet E, Marchand C, Hirasawa M, Lancelin JM, Knaff DB and Miginiac-Maslow M (2003) The *Arabidopsis* plastidial thioredoxins: New functions and new insights into specificity. *J Biol Chem* 278:23747-23752.
- Emanuelsson O, Nielsen H, Brunak S and von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence *J Mol Biol* 300:1005-1016.
- Gelhaye E, Rouhier N and Jacquot JP (2003a) Evidence for a subgroup of thioredoxins h that requires GSH/Grx for its reduction. *FEBS Letters* 555:443-448.
- Gelhaye E, Rouhier N, Vlamis-Gardikas, Girardet JM, Sautiere PE, Sayzet M, Martin F and Jacquot JP (2003b) Identification and characterization of a third thioredoxins h in poplar. *Plant Physiol Biochem* 41:629-635.
- Gelhaye E, Rouhier N and Jacquot JP (2004a) The thioredoxins h system of higher plants. *Plant Physiol Biochem* 42:265-71.
- Gelhaye E, Rouhier N, Gerard J, Jolivet Y, Gualberto J, Navrot N, Ohlsson PI, Wingsle G, Hirasawa M, Knaff DB, Wang H, Dizengremel P, Meyer Y and Jacquot JP (2004b) A specific form of thioredoxin h occurs in plant mitochondria and regulates the alternative oxidase. *Proc Natl Acad Sci USA* 101:14545-14550.
- Hohfeld J, Minami Y and Hartl FU (1995) Hip, a novel cochaperone involved in the eukaryotic Hsc70/Hsp40 reaction cycle. *Cell* 83:589-98.
- Jacquot JP, Rivera-Madrid R, Marinho P, Kollarova M, Le Marechal P, Miginiac-Maslow M and Meyer Y (1994) *Arabidopsis thaliana* NADPH thioredoxin reductase. cDNA characterization and expression of the recombinant protein in *Escherichia coli*. *J Mol Biol* 235:1357-63.
- Jacquot JP, Lancelin JM and Meyer Y (1997) Thioredoxin: Structure and function in plant cells. *News Phytol* 136:543-570.
- Kobrehel K, Wong JH, Balogh A, Kiss F, Yee BC and Buchanan B (1992) Specific reduction of wheat storage proteins by thioredoxins h. *Plant Physiol* 99:919-924.
- Laloi C, Rayapuram N, Chartier Y, Grienenberger JM, Bonnard G and Meyer Y (2001) Identification and characterization of a mitochondrial thioredoxin system in plants. *Proc Natl Acad Sci U SA* 20:14144-14149.
- Laurent TC, Moore EC and Reichard P (1964) Enzymatic synthesis of deoxyribonucleotides. Isolation and characterization of thioredoxin, the hydrogen donor of *Escherichia coli*. *J Biol Chem* 239:3436-3444.
- Lemaire SD, Collin V, Keryer A, Quesada M and Miginiac-Maslow M (2003a) Characterization of thioredoxins y, a new type of thioredoxins identified in the genome of *Chlamydomonas reinhardtii*. *FEBS Letters* 543:87-89.
- Lemaire SD, Collin V, Keryer E, Issakidis-Bourguet, Lavergne D and Miginiac-Maslow M (2003b) *Chlamydomonas reinhardtii* a model organism for the study of the thioredoxins family. *Plant Physiol Biochem* 41:513-521.
- (2004) New targets of *Arabidopsis* thioredoxins revealed by proteomic analysis. *Proteomics* 4:2696-706.
- Marx C, Wong JH and Buchanan BB (2003) Thioredoxin and germinating barley: Targets and protein redox changes. *Planta* 216:454-60.
- Mestres-Ortega D and Meyer Y (1999) The *Arabidopsis thaliana* genome encodes at least four thioredoxins m and a new prokaryotic-like thioredoxin. *Gene* 240:307-16.
- Meyer Y, Verdoucq L and Vignols F (1999) Plant thioredoxins and glutaredoxins: Identity and putative roles. *Trends Plant Sci* 4:388-394.
- Meyer Y, Vignols F and Reichheld JP (2002) Classification of plant thioredoxins by sequence similarity and intron position. *Methods Enzymol* 347:394-402.
- Mouaheb N, Thomas D, Verdoucq L, Monfort P and Meyer Y (1998) *In vivo* functional discrimination between plant thioredoxins by heterologous expression in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 95:3312-3317.
- Nelson RJ, Ziegelhoffer T, Nicolet C, Werner-Washburne M and Craig EA (1992) The translation machinery and 70 kd heat shock protein cooperate in protein synthesis. *Cell* 71:97-105.
- Netto LES (2001) Oxidative stress in sugarcane. *Genet and Molec Biol* 24:93-102.
- Pagano EA, Chueca A and Lopez-Gorge J (2000) Expression of thioredoxins f and m, and of their targets fructose-1,6-biphosphatase and NADP-malate dehydrogenase, in pea plants grown under normal and light/temperature stress conditions. *Journal of Experimental Botany* 51:1299-1307.
- Reichheld JP, Mestres-Ortega D, Laloi C and Meyer Y (2002) The multigenic family of thioredoxin h in *Arabidopsis thaliana* : Specific expression and stress response. *Plant Physiol Biochem* 40:685-690.
- Rivera-Madrid R, Mestres D, Marinho P, Jacquot JP, Decottignies P, Miginiac-Maslow M and Meyer Y (1995) Evidence for five divergent thioredoxin h sequences in *Arabidopsis thaliana*. *Proc Natl Acad Sci U SA* 92:5620-5624.

- Schumann P and Jacquot JP (2000) Plant thioredoxins systems revisited. *Ann Rev Plant Physiol* 51:371-4000.
- Thompson LD, Higgins DG and Gibson TJ (1994) CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positive-specific gap penalties and weight matrix choice. *Nucleic Acid Res* 22:4673-4680.
- Verdoucq L, Vignols F, Jacquot JP, Chartier Y and Meyer Y (1999) *In vivo* characterization of a thioredoxin h target protein defines a new peroxiredoxin family. *J Biol Chem* 274:19714-22.
- Vignols F, Mouaheb N, Thomas D and Meyer Y (2003) Redox control of Hsp70-Co-chaperone interaction revealed by expression of a thioredoxin-like *Arabidopsis* protein. *J Biol Chem* 278:4516-23.
- Yamazaki D, Motohashi K, Kasama T, Hara Y and Hisabori T (2004) Target proteins of the cytosolic thioredoxins in *Arabidopsis thaliana*. *Plant Cell Physiol* 45:18-27.
- Wong JH, Kim YB, Ren PH, Cai N, Cho MJ, Hedden P, Lemau PG and Buchanan BB (2002) Transgenic barley grain overexpressing thioredoxin shows evidence that the starchy endosperm communicates with the embryo and the aleurone. *Proc Natl Acad Sci USA* 99:16325-30.

Associate Editor: Carlos F.M. Menck