Genetics and Molecular Biology, 28, 3 (suppl), 520-528 (2005)

www.sbg.org.br

Research Article

Identification and *in silico* expression pattern analysis of *Eucalyptus* expressed sequencing tags (ESTs) encoding molecular chaperones

Thiago C. Cagliari[1,2#], Ana O. Tiroli[1,2#], Júlio C. Borges[1,2] and Carlos H.I. Ramos[1,2]

[1]*Laboratório Nacional de Luz Síncrotron, Centro de Biologia Molecular Estrutural, Campinas, SP, Brazil.*
[2]*Departamento de Bioquímica, Instituto de Biologia, Universidade Estadual de Campinas. Campinas, SP, Brazil.*

Abstract

Expressed Sequence Tags (ESTs) sequencing provides reliable and useful information concerning gene expression patterns in the genomic context. Our group used bioinformatics to identify and annotate 5'EST-contigs belonging to the molecular chaperones within the *Eucalyptus* Genome Sequencing Project Consortium (FORESTs) database. We found that 1,959 5'EST-contigs, or approximately 1.6% of the total 5'EST-contigs, encoded chaperones, emphasizing their biological importance. About 55% of the chaperones that we found were Hsp70 chaperones and its co-chaperones, 18% were Hsp90 chaperones, 15% were Hsp60 and its co-chaperone, 8% were Hsp100 chaperones, and 4% were Small Hsps. We also investigated the digital expression profile of the chaperone genes to gain information on gene expression levels in the different libraries and we found that molecular chaperones may have differential expression. The results discussed here give important hints about the role of chaperones in *Eucalyptus* cells.

Key words: chaperones, heat shock proteins, genome, expressed sequence tags.

Received: May 28, 2004; Accepted: March 17, 2005.

Introduction

Environmental temperature fluctuations and other forms of stress can provoke partial or complete protein unfolding. This unfolding induces the exposition of hydrophobic residues that are usually buried in the native state, leading to intracellular protein aggregation and its dreadful consequences. Molecular chaperones are able to bind to the exposed hydrophobic residues in unfolded proteins, preventing incorrect folding and aggregation (for recently reviews see Wang *et al*. 2004, and Borges and Ramos, 2005). Due to their proper response to stress conditions, chaperones were first classified as heat shock proteins (Hsp), however some of them are also constitutively expressed because they are required for substrate protein maturation, protein transport, and other functions (Ellis and Hartl, 1996; Fink, 1999). Molecular chaperones are divided into families according to their approximate molecular weight (Fink, 1999; Borges and Ramos, 2005). The main features of each family, such as state of oligomerization, presence of ATPase activity, and interaction with co-chaperones, which are proteins that somehow assist their activities, are described in Table 1.

Send correspondence to Carlos H.I. Ramos. Caixa Postal 6192, 13084-971 Campinas, SP, Brazil. E-mail: cramos@lnls.br.
[#]These authors contributed equally to this work.

Advances in large-scale cDNA sequencing provided a reliable source of information concerning quality and quantity of expressed mRNAs under very diverse developmental and environmental conditions. This information can lead to new insights on specific protein importance and activity in response to environmental conditions sensed by the organism. EST genome projects indicate trustworthy patterns of gene expression on specific tissues or even in entire organisms (Mayer and Mewes, 2002). Through the use of bioinformatics tools, such as annotation by homology search, valuable information about specific target molecules may be accessed.

Usually, plants are more exposed to unfavorable environmental changes than animals because of their sessile existence. Therefore, knowledge about the network of stress proteins in plants is of great interest not only to improve the agriculture production but also to enhance our understanding about the function of the chaperones and the protein folding process. Here we describe the *in silico* search and identification of molecular chaperone transcripts in the Brazilian *Eucalyptus* Genome Sequencing Project Consortium (FORESTs – https://forests.esalq.usp.br/) by EST homology comparison between sequences available at public databases and sequences generated by FORESTs. Assuming that the whole 5'EST set from FOR-

Table 1 - Molecular chaperones classification.

Chaperone family	Co-chaperone	Molecular mass (kDa)	n-mers	Comments
Hsp100	Unknown	≈100	6 or 7	ATPase activity; recover proteins from aggregates, form proteolytic complexes
Hsp90	HIP, HOP, p23 and others	≈90	2	Weak intrinsic ATPase activity; prevent protein aggregation; interact with a large set of proteins like transcription factors and protein kinase
Hsp70	Hsp40, GrpE and others	≈70	1 /2 /2	ATPase activity; bind nascent peptide chains to help many processes such as folding, transport through membranes, degradation, escape from aggregation
Hsp60 (GroEL)	Hsp10 (GroES)	≈60	14/7	ATPase activity; assist protein to fold efficiently
SmHsp	Unknown	15-30	9 to24	ATP-independent, involved in thermotolerance and escape from aggregation by keeping proteins in a refoldable state

ESTs database is in agreement with the real *Eucalyptus* mRNA expression profile, a reasonable amount of putative sequences related to molecular chaperones were identified, indicating the importance of these proteins to the cell. A digital northern blot was also performed showing that molecular chaperones may have differential expression. The relevance of these findings is also discussed.

Materials and Methods

The EST library preparation, sequencing, clustering, and other important data are available at the FORESTs home page (https://forests.esalq.usp.br) and in accompanying manuscripts in this issue. The strategy to annotate molecular chaperones in the FORESTs project was as follows. The translated amino acid sequences of specifics mRNAs from chaperones and stress-related proteins were chosen as described by Borges *et al*. (2001). The translated sequences were compared with the cluster consensus generated by the FORESTs database utilizing the software program tBLASTn (Altschul *et al*., 1997), which is basically an alignment research tool (Figure 1). The FORESTs cluster that had a statistically significant match (*i.e.* an E-value lower than 1e-5) was considered to be a molecular chaperone. The homology between protein sequences was analyzed using the alignment program LALIGN (http://www.ch.embnet.org/software/LALIGN_form). In order to determine the amount of sequences associated with molecular chaperone genes, the data were expressed as a percentage of all the 5'EST belonging to annotated chaperone families in the FORESTs. The annotation process was improved by increasing the number of data mining rounds, which was done by submitting the predicted protein family encountered in the first round to the following rounds (Figure 1). Each additional round increases the probability of finding a cluster that was incorrectly excluded in previous rounds.

A digital northern blot was also performed in order to determine the chaperone expression levels for each FORESTs library. The libraries with fewer than one thousand sequence tags were not considered in our analysis because the abundance of cDNA not picked up among one thousand clones is unlikely (about 5% chance) to be larger than 3.7/1000 (Audic and Claverie, 1997). The molecular chaperones expression level was calculated by dividing the number of the 5'EST sequences annotated to a specific chaperone in each particular library, by the total number of 5'EST sequenced in that library. The resulting number allows a quantitative and qualitative expression analysis of the molecular chaperones in different *Eucalyptus* tissues.

Results and Discussion

The data mining approach used here found 1,959 5'ESTs belonging to the molecular chaperone proteins category. The FORESTs database contains a total of 123,889 5'ESTs, which means that about 1.6% of the 5'ESTs sequenced by the FORESTs consortium were associated to chaperone genes (Table 2). Considering that the number of 5'EST in each library is related to the number of mRNAs

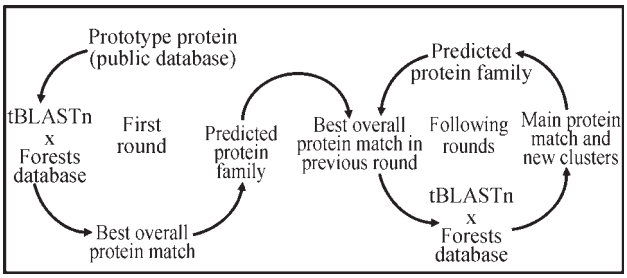

Figure 1 - Data mining and annotation strategy flowchart. Translated amino acid sequences of specific mRNAs were chosen from chaperones and stress-related proteins as described by Borges *et al*. (2001). The prototype sequences were compared with the first level cluster consensus generated by the *Eucalyptus* Genome Sequencing Project Consortium (FORESTs) using the basic local alignment search tool tBLASTn program (Altschul *et al*., 1997). The tBLASTn program was used to compare amino acid sequences from public databases with the FORESTs database. The FORESTs cluster sequences with an E-value lower than 1e-5 were considered to belong to the chaperone category. To improve the prediction accuracy, matches found in the first round of mining went through a second or even more rounds of mining, allowing precise identification of more matches and/or new Clusters not previously found.

Table 2 - The main chaperone families and classes, their cellular location and the mining results. Our data mining approach found 1,959 5'ESTs belonging to the molecular chaperones category, which is about 1.6% of the total 5'ESTs sequenced by the FORESTs consortium.

Molecular chaperones			Mining results		
Family	Class	Intracellular location	Annotated clusters	Annotated 5' EST sequences	Total Annotated 5' EST sequences (family only)
Small Hsp	Class I	cytoplasm	24	69	82
	Class II	cytoplasm	-	-	
	Class III	mitochondria	1	2	
	Class IV	chloroplast	3	11	
Hsp60 /TriC[1]	TCP-1[2]	mMitochondria	3	58	266
		chloroplast	3	89	
		cytoplasm	9	100	
	Hsp60-like	-	-	19	
Chaperonin-like co-chaperone	Hsp10	chloroplast	4	27	27
Hsp70	Hsc70	cytoplasm	21	272	495
	Mt-Hsp70	mitochondria	3	25	
	Cp-Hsp70	chloroplast	3	56	
	Bip[3]/Grp78	ER[4]	5	89	
	Hsp110	-	5	53	
Hsp70 co-chaperones	Hsp40	cytoplasm, mitochondria, ER	91	568	578
	GrpE[5]	mitochondria	4	10	
Hsp90	Hsp82	cytoplasm	10	279	360
	Grp94	ER	1	41	
	Cp-Hsp82	chloroplast	2	25	
	Hsp90-like	-	13	15	
Clp[6]	Hsp100/ClpB	cytoplasm,	4	13	151
	ClpA/C	mitochondria	13	53	
	ClpX	chloroplast	10	85	
		-			
Total			232	1,959	1,959 (1.6%)

[1]TriC: TCP ring Complex; [2]TCP1: Chaperonin-containing T-Complex; [3]Bip: Binding protein; [4]ER: endoplasmic reticulum; [5]GrpE: GroP-like gene E; [6]Clp: Caseinolytic protease.

produced by the cell and therefore to the plant gene expression profile, the chaperone genes have a remarkable expression level in *Eucalyptus* cells. This result is expected because the cell requires proper protein folding to be viable. About 55% of all 5'EST annotated as belonging to the molecular chaperones category were homologous to the Hsp70 system (Hsp70 and its co-chaperones), 18% were homologous to the Hsp90 family, 13% were homologous to the Hsp60/TCP-1 family, 8% to the Hsp100 family, 4% to the Small Hsp family, and about 2% to the Hsp10 family (Figure 2).

Hsp70 and co-chaperones

Hsp70 acts as a pivot chaperone, interacting with other chaperones and exchanging unfolded polypeptides with them. Hsp70 transports proteins to the Hsp60 system for *de novo* folding (Siegers *et al*., 1999), participates in aggregates recovery together with Hsp100 chaperones (Mogk *et al*.,1999), cooperates with the small Hsps to refold heat-denatured polypeptides (Lee and Vierling, 2000), and controls the activity of regulatory proteins as heat-shock factors, the transcriptional factors of Hsps genes (Morimoto, 1998). In plants, members of the Hsp70 family are expressed in response to environmental or abiotic stress conditions such as heat, cold, drought, etc. (Guy and Li, 1998; Sung *et al*., 2001). Despite the determination of thermotolerance acquisition upon Hsp70 overexpression, which also enhanced salt and water tolerance (Sugino *et al*., 1999; Alvim *et al*., 2001), the cellular mechanisms of Hsp70 upon stress are not fully understood (Wang *et al*., 2004).

Hsp70 requires two or more cofactors, and the most important among them are the co-chaperones Hsp40 and GrpE. Sequences related to Hsp70 and Hsp40 were responsible for more than half of the 5'EST identified as chaperones in the *Eucalyptus* (Figure 2). Hsp70 sequences comprise about 25% of all identified chaperone ESTs and Hsp40 sequences comprise about 30% (Figure 2). The approximated 1:1 expression profile between Hsp70 and Hsp40 in *Eucalyptus* is in agreement with the results obtained from the analyses of 5EST libraries from other organisms (Borges *et al*., 2001). Hsp70 and Hsp40 chaperones were highly expressed in *Eucalyptus*, a result also

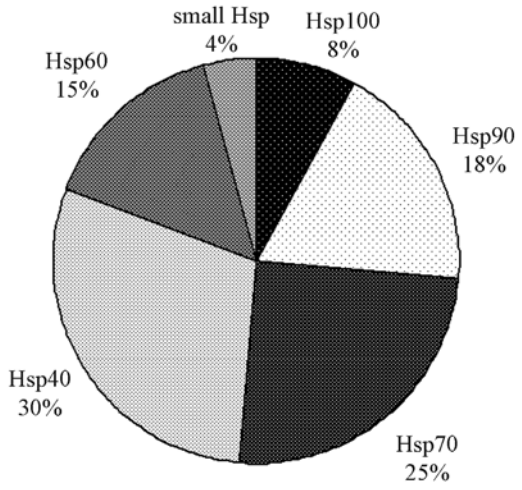

Figure 2 - Relative abundance of the chaperone families in *Eucalyptus*. The number of reads of each chaperone family was normalized as the percentage of the total chaperones annotated. The Hsp60 family slice represented here also contains the Hsp10 family, which represented nearly 2% of all annotated chaperones in FORESTs.

observed for *Arabidopsis* (Miernyk, 2001), sugar cane (Borges *et al*., 2001) and human genome databases. The data from the FORESTs analysis reinforced the importance of the Hsp70 and Hsp40 proteins for the protein-folding homeostasis in all organisms.

Chaperones from the Hsp70 family are classified in three subfamilies based on their nucleotide dissociation properties: the DnaK subfamily, the Hsc70 subfamily, and the HscA subfamily (Brehmer *et al*., 2001). Chaperones belonging to the three Hsp40 subfamilies are able to bind all Hsp70 subfamily members, and have different functions and structures (Lu and Cyr, 1998; Borges *et al*., 2005), which is in agreement with the high diversity of this protein family in the cell. Unlike Hsp40, GrpE binds only to the DnaK subfamily (Brehmer *et al*., 2001), which may explain why GrpE is less expressed than Hsp40 in *Eucalyptus* (Table 1) and also in sugar cane (Borges *et al*., 2001). The lower expression level of GrpE genes compared to Hsp40 genes is probably caused by the compartmentalization of the first, since it was described only in prokaryotes, and in both mitochondria and chloroplast eukaryotic organelles (Netzer and Hartl, 1998). Another explanation for this lower expression is the restricted and transient functions of GrpE. GrpE participates in the folding reaction only during the release of ADP, while Hsp70 and Hsp40 remain bound to the substrate protein during the whole process (Hartl, 1996).

Hsp70 homologues from all cellular compartments were found in FORESTs database, but the cytoplasmic Hsp70 (Hsc70 subfamily) possesses 55% of all reads related to Hsp70 family or 10% of all reads related to chaperones at the FORESTs database (Table 2). It is characteristic of a single plant species to display multiple genes for Hsc70 encoding proteins that are more than 90% identical to each other at the primary structure (Boston *et al*., 1996), which is probably due to the fact that most of the nascent proteins are produced in the cytoplasm. Hsc70 has also the most abundant expression (about 65%) among Hsp70 proteins in sugar cane cells (Borges *et al*., 2001), pointing to a universal pattern of expression for this protein in plants. These results are very important because Hsc70 plays an important role in the net response to low temperature in sugar cane (Nogueira *et al*., 2003) and it may be playing the same role in *Eucalyptus* (see discussion below). The 37 Hsp70 Clusters found within FORESTs are far more diverse than the 18 genes found in *Arabidopsis* (Lin *et al*., 2001) and the 12 found in spinach (Guy and Li, 1998). This difference cannot be solely explained by genetic distance. The Hsp70 Clusters encountered may be related to diverse mRNA processing and even clustering process failures, since FORESTs is an EST genome project.

The Hsp70 and Hsp40 families were encountered in all the FORESTs libraries examined (Figure 3), which indicates their relevance for protein folding in various tissues as indicated by several works with these proteins (Boston *et al*., 1996; Fink, 1999). Worthwhile mentioning is the high expression identified in the *E. grandis* seedling tissues cultivated in the dark (SL7 library) (Figure 3). This is an im-

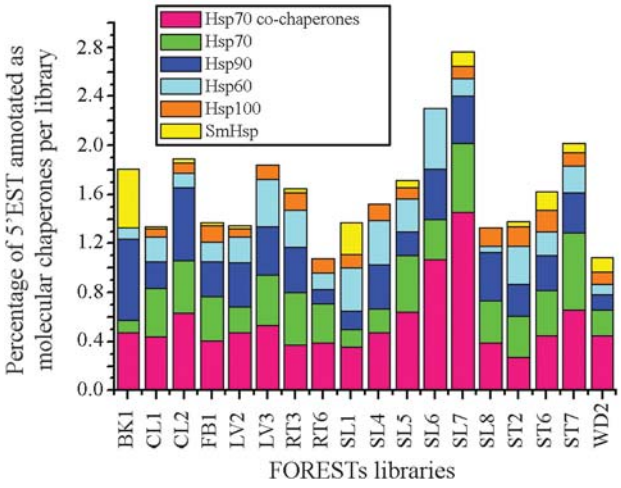

Figure 3 - Expression of the chaperone families in the different tissues of eucalyptus. The percentages represent the number of 5'EST associated to chaperone genes in each library. The colors indicate the contribution of each chaperone family. The FORESTs libraries are: BK1 (bark, duramen, pith, and alburnum); CL1 (*E. grandis* callus formed in the dark), CL2 (*E. grandis* callus formed in the light); FB1 (flower buds, flowers, and fruits); LV2 (leaves of efficient and inefficient trees in Phosphorus and Boron utilization); LV3 (leaves colonized by the *Thyrinteina* grub for seven days); RT3 (roots from seedlings); RT6 (roots from trees resistant and susceptible to frost); SL1 (*E.grandis* seedlings cultivated in the dark and exposed to light for three hours before the RNA extraction); SL4 (*E. globulus* seedlings cultivated in the dark); SL5 (*E. saligna* seedlings cultivated in the dark); SL6 (*E. urophylla* seedlings cultivated in the dark); SL7 (*E. grandis* seedlings cultivated in the dark); SL8 (*E. camaldulensis* seedlings cultivated in the dark); ST2 (stem from six month seedlings susceptible to hydric deficit); ST6 (seedlings stem susceptible to hydric deficit); and WD2 (*E. grandis* wood).

portant result from which the relevance of the chaperones for the cell function maintenance is verified (see discussion below).

Hsp60 and Hsp10

Chaperones belonging to the Hsp60 family play an important role in the cell because they assist the newly synthesized and newly translocated proteins to achieve native conformation, and function as pairs of stacked rings with seven subunits each (Borges and Ramos, 2005). Hsp60 proteins are necessary for the folding of many chloroplastic proteins such as Rubisco, the enzyme that makes up about 50% of the total chloroplast protein content and is responsible for the CO_2 fixation reaction.

Approximately 15% of the total reads annotated as molecular chaperones belonged to the Hsp60 family and its co-chaperone Hsp10 (Figure 2). The abundance of ESTs annotated as belonging to the Hsp60 family was almost equally distributed among the three intracellular compartments: mitochondria, chloroplast and cytoplasm. The expression of cytoplasmic Hsp60 (subfamily TCP-1) was the most significant, with 38% of the total 5'EST annotated as Hsp60 chaperone family, followed by other Hsp60 proteins located in the chloroplast (33%), and mitochondria (22%) (Table 2). The higher expression of the subfamily TCP-1 compared to the other Hsp60 subfamilies may be explained because the folding of newly synthesized or imported proteins assistance occurs mainly in the cytoplasm (Grantcharova *et al*., 2001). It is noteworthy that all the three Clusters related to chloroplastic Hsp60, one cluster related to the mitochondrial Hsp60 and four Clusters belonging to the TCP-1 subfamily had the sequence coding for the whole protein, reinforcing the confidence of the annotation results showed here.

The number of *Eucalyptus* Clusters related to mitochondrial Hsp60 and TCP-1 were quite similar to the number of mitochondrial Hsp60 and TCP-1 genes found in the *Arabidopsis* genome (Hill and Hemmingsen, 2001). Both organisms have 9 TCPs and *Arabidopsis* has 4 genes related to mitochondrial Hsp60 while 3 mitochondrial Hsp60 Clusters were annotated in the FOREST. However, only three Clusters encoding plastid Hsp60 members were found in *Eucalyptus* while *Arabidopsis* has 6 genes belonging to this category (Hill and Hemmingsen, 2001). This difference may be explained by a hypothetical very low expression of these genes in *Eucalyptus*, which would result in the absence of their ESTs in the final library, or by discrepancy between *Arabidopsis* predicted Cpn60 genes and *Eucalyptus* hypothetical Cpn60 genes, with the relatively low identity between these proteins increasing the difficulty in finding new Cpn60 orthologues (Hill and Hemmingsen, 2001). The nineteen Hsp60-like sequences annotated for *Eucalyptus* (Table 2) may be highly divergent Cpn60 genes which were not identified due to low homology.

The Hsp60 family was encountered in all the FOREST libraries examined (Figure 3), which indicates its relevance for protein folding in various tissues as indicated by many works with these proteins. The higher expression levels of the Hsp60 family were registered for the SL6 (*E. urophylla* seedlings) and for the RT3 (seedlings root) libraries (Figure 3), which represent tissues in development, suggesting that Hsp60 members are important to *Eucalyptus* developmental stages (see discussion below). This is compatible with the fact that mutated species of *Arabidopsis* chloroplasts Cpn60α exhibit defects in embryo and seedlings development (Apuya *et al*., 2001). The expression of Hsp60 genes in leaves colonized by the *Thyrinteina* grub for seven days is also worthy of note (LV3 library), and may indicate a differential expression of Hsp60 under biotic stress conditions.

Hsp90

Hsp90 is a dimeric chaperone that interacts with more than 40 protein substrates including transcription factors, protein kinases and unrelated proteins (Pearl and Prodromou, 2001), and forms cytoplasmic chaperone complexes or foldosomes with others chaperones, co-chaperones, and cellular proteins (Zhao *et al*., 2005). The main task of Hsp90 is to assist protein folding (Buchner, 1999) but it also influences other activities related to signal transduction, cell cycle control, and protein degradation and trafficking (Pearl and Prodromou, 2001). Hsp90 has important functions in plant cells. It acts as a buffer for cryptic genetic variations in *Arabidopsis* (Queitsch *et al*., 2002), which indicates an important role in morphological evolution, and is a key player in biotic stress because it interacts with resistance gene products (Hubert *et al*., 2003; Liu *et al*., 2004; Lu *et al*., 2003).

The Hsp90 family represented about 18% of all 5'EST annotated to molecular chaperone category in *Eucalyptus* (Figure 2). This result was in agreement with the fact that Hsp90 is one of the most abundantly expressed cytoplasmic proteins in several organisms, even in the absence of stress (Boston *et al*., 1996; Buchner, 1999; Borges *et al*., 2001). The cytoplasmic Hsp82 is by far the most expressed protein within the Hsp90 chaperone family with 78% of the total 5'EST belonging to this family and more than 10% of all 5'EST belonging to the chaperone category (Table 2). Hsp82 has also the most abundant expression (about 70%) among Hsp90 proteins in sugar cane cells (Borges *et al*., 2001), pointing to a universal pattern of expression for this protein in plants. This result is very important because Hsp82 is likely to be involved in a net response to low temperature in sugar cane (Nogueira *et al*., 2003) and it may be playing the same role in *Eucalyptus*. Since Hsc70, another protein involved in response to low temperature in sugar cane (Nogueira *et al*., 2003) also had more than 10% of all 5'EST belonging to the chaperone category (Table 2, and discussion above), we suggest that these two proteins are

likely to be directly involved in the response to low temperatures in eucalyptus, and for extension to several other plants. Therefore, we suggest that Hsc70 and Hsp82 are good candidates for *in vivo* studies, in order to better characterize the role of these proteins in plant tolerance under stress conditions.

The *Arabidopsis* genome presents four cytoplasmic Hsp90 genes while the FORESTs database presented ten cytoplasmic Hsp90 related Clusters. This difference may be attributed to alternative RNA splicing, not detected in the *Arabidopsis* genome project. *Arabidopsis* has one Hsp90 gene for endoplasmic reticule, mitochondria, and chloroplast, while our data mining found one and two Hsp90 Clusters for endoplasmic reticule and chloroplast respectively, and failed to find an Hsp90 cluster in mitochondria. The absence of a mitochondrial Hsp90 in *Eucalyptus* is intriguing since plant Hsp90 sequences are very conserved, sharing 60-70% identity with animal and yeast Hsp90 representatives (Krishna and Gloor, 2001). However, the *Eucalyptus* mitochondrial Hsp90 gene was probably present among the 13 Hsp90-like Clusters found but which were not annotated due to their low confidence since the average of the sequences in each cluster was about one.

The 5'EST related to the Hsp90 family was found in all libraries (Figure 3), which is in agreement with its indispensable necessity for viability of eukaryotic organisms (Buchner, 1999). In general, the number of transcripts related to the Hsp90 family was very significant among all the libraries, but CL2 (callus) and BK1 (bark, duramen, pith, and alburnum) libraries exhibited an especially high number of Hsp90 transcripts while RT6 (root) and WD2 (wood) presented a lower number of transcripts (Figure 3). The important Hsp90 expression observed in tissues representing distinct developmental steps, such as adult (BK1) and callus (CL2), underlined the importance of this chaperone family for diverse conditions of tissue development in *Eucalyptus grandis* (see discussion below). Tissues under growth conditions demand, theoretically, more expression of Hsp90 because it activates many kinases that regulate the mitotic cell cycle (Schulte *et al*., 1996).

Hsp100 family

Hsp100 proteins are specialized molecular machines that promote protein disaggregation and/or protein degradation, removing potentially harmful polypeptides arising from misfolding, denaturation or aggregation (Mogk *et al*., 1999). The Hsp100 family, also named Clp (caseinolytic protease), solubilizes the aggregate protein and releases it in a conformation that can be refolded by the Hsp70 chaperone system (Wang *et al*., 2004).

The Hsp100 family represented approximately 8% of the total 5'EST found to be molecular chaperones (Figure 2) and about 56% of these belonged to the ClpX subfamily (Table 2). ClpX belongs to the Class II of Clp ATPases (Gottesman *et al*., 1990 and 1993), which has only one ATP-binding domain that has significant homology with the second ATP-binding domain, from Class I Clp ATPases (Horwich, 2002). This low expression of Hsp100 proteins was spread almost undistinguishably throughout all the libraries investigated (Figure 3). The lack of heat shock induction, which is required for expression of many Hsp100 genes (Schirmer *et al*., 1994), may explain the lower expression of Hsp100 representatives compared to other families. Unlike other chaperone families, Hsp100 neither promotes the initial folding nor prevents the aggregation of proteins (Wang *et al*., 2004). For instance, the disassemble of protein aggregates by Hsp101 takes place mainly during the post-stress phase (see Agarwal and colleagues, 2003, for an example in plants). In other words, the methodology applied for certain libraries (*i.e.* exposition to light during three hours) was not the most effective in triggering Hsp100 differential expression in *Eucalyptus* cells.

Small heat shock proteins

Small Hsp proteins have monomeric masses ranging from 12 to 43 kDa, are characterized by a α-crystallin domain in its C-terminus (De Jong *et al*., 1998), and are responsible for maintaining the solubility of unfolding proteins during heat stress (Kim *et al*., 2003). Small Hsps seem to form large heteroligomeric aggregates that undergo conformational changes with temperature increasing, promoting the display of hydrophobic amino acids residues that bind to proteins denaturated by heat (Ehrnsperger *et al*., 1998). Small Hsps are highly induced in mitochondrias and chloroplasts by oxidative stress (Lee and Vierling, 2000).

Of all the chaperone families investigated here, Small Hsp was the least abundant (Figure 2), which is in accordance with most small Hsps not being detected in the vegetative tissues, under normal growth conditions (Sun *et al*., 2002), although they are capable of being produced and accumulated in response to environmental stress and developmental stimuli. Only one of its classes (cytoplasmic Class I) accounted for about 90% of the Clusters found (Table 2). These results are consistent with those from the digital expression analysis of small Hsps in sugar cane cells (Borges *et al*., 2001) and may indicate a universal pattern of expression for this family in plants.

The expression profile analysis of the Small Hsp family in *Eucalyptus* tissues was marked by the high expression of this protein found in the BK1 library (Figure 3). This library originated from bark tissues that are in general exposed to daily stress conditions, which indicated the participation of Small Hsps in the folding process of synthesized protein in this tissue. Our analyses identified a very small or even lack of expression of Small Hsps in leaf (LV2 and LV3) and root (RT3 and RT6) libraries (Figure 3). These results are in accordance with previous works that show that Small Hsps are not detected in the absence of

stress in leaf or root tissues (DeRocher *et al*., 1991; Hsieh *et al.*, 1992).

Small Hsps are usually found in reproductive organs during several stages of plant development (Waters *et al*., 1996) and after heat (or other) shock induction (DeRocher *et al*., 1991), in agreement with the high levels of Small Hsp 5'ESTs found in libraries (ST6 and ST7) originating from *Eucalyptus* stem tissues susceptible to hydric deficit (Figure 3). Similar results were found for other plants such as sunflowers and *A. thaliana*, in which an increase in Small Hsp expression correlates with osmotic stress (Sun *et al*., 2002). The dehydration can contribute to aggregation caused by the increasing concentration of proteins inside the cells. Thus, the presence of molecular chaperones, especially Small Hsps, is important because they bind partially denatured proteins and prevent their aggregation (Ehrnsperger *et al*., 1998). The high diversification of plant Small Hsps shows a molecular adaptation to stress conditions that seems to be exclusively from plants. Therefore, the abundance and diversity of small Hsp family suggest their important function in plant stress tolerance (Sun *et al*., 2002; Wang *et al*., 2003).

Tissue expression analysis

The digital northern blot analysis performed here is based on the assumption that the number of EST clones is directly related to the abundance of messenger RNA and therefore to the protein expression level in each library (Audic and Claverie, 1997). Although there is general agreement that variations may exist, this approach is generally accepted as a reasonable approximation to the true protein level of expression. The expression profile of the molecular chaperone mRNAs provided by the digital northern blot indicated that chaperones were differently expressed in the *Eucalyptus* tissues (Figure 3). The libraries with the highest amount of 5'EST sequences annotated as chaperones were SL7 (*E. grandis* seedlings), SL6 (*E. urophylla* seedlings), and ST7 (stem tissue), with 2.8%, 2.3%, and 2% of sequences, respectively. The libraries with the lowest amount of 5'EST sequences annotated as chaperones, approximately 1% each, were RT6 (root) and WD2 (wood).

The SL libraries (4, 5, 6, 7, and 8) represent five different species of *Eucalyptus* seedlings submitted to the same experimental conditions – totally cultivated in the dark: *E. grandis* (2.8% of the total 5'EST from SL7 library were related to molecular chaperone category), *E. urophylla* (2.3% of the total 5'EST from SL6), *E. saligna* (1.7% of the total 5'EST from SL5), *E. globulus* (1.5% of the total 5'EST from SL4) and *E. camaldulensis* (1.3% of the total 5'EST from SL8) (Figure 3). These results show that the expression of chaperones was species-dependent, one species with twice as much chaperone expression as the other. As far as we know, a comparison of stress resistance among these species has not yet been experimentally performed. It would be interesting to find out if this character is associated with the chaperone gene expression profile found here.

There are other observations regarding the general expression of chaperone proteins in the *Eucalyptus* that deserve to be mentioned. Two different libraries originating from *E. grandis* seedlings were created. The SL1 library originated from seedlings cultivated in the dark and exposed to light for three hours before RNA extraction. The SL7 library originated from seedlings totally cultivated in the dark. About 2.8% of the total ESTs of the SL7 library were associated to chaperone genes, but only 1.4% were found in the SL1 library (Figure 3). These results showed that these tissues had differential chaperone expression that was dependent on the exposition to light during their development. Closer inspection of the expression profile of these two libraries showed that the increase in chaperone expression is mainly due to the increase in the expression of Hsp70 and its co-chaperones (Figure 3). There are two other libraries, originating from callus tissues, which differ from their treatment with regard to exposition to light. The CL2 library originated from callus exposed to light and the CL1 library originated from callus tissues developed in the absence of light. About 1.9% of the total ESTs of the CL2 library were associated to chaperone genes, whereas about 1.3% were found in the CL1 library (Figure 3). These results showed that these tissues also had differential chaperone expression that was dependent on the exposition to light during their development. A close inspection of the expression profile of these two libraries showed that the increase in chaperone expression is mainly due to the increase in the expression of Hsp70 and Hsp90 genes (Figure 3). Hsp70 and Hsp90 are protein families involved with basic cellular process that do not necessarily concern protein folding (for reviews see Buchner, 1999; Jensen and Johnson, 1999; Borges and Ramos, 2005) and that explains their high expression not only during stress conditions but also whenever high cellular activities take place. The diverse function of these two proteins may be the explanation for one tissue presenting higher chaperone expression when exposed to light than in dark conditions, and another tissue presenting an opposite behavior. Since seedlings are tissues in development, they may need the "over-assistance" of chaperones in basic cell functions (Wang *et al*., 2004). In the case of the callus tissue, the chaperones may be necessary for decreasing the effects caused by external conditions that stressed the cell. Therefore, since these two functions are difficult to distinguish and they may be taking place at the same time, more experimental work has to be done in order to support the hypothesis discussed above.

By comparing the library LV3, which was originated from leaf tissues that were colonized by the *Thyrinteina* grub, wich a leaf library not exposed to these insects (LV2) we found an increased level of chaperone expression (1.9% *vs.* 1.5% of the total ESTs associated to chaperone genes)

(Figure 3). This result suggested that this biotic stress condition could affect the chaperone levels in *Eucalyptus* cells. The increase in chaperone gene expression was mainly restricted to Hsp70 and Hsp60 families, which are mainly involved in the folding of newly synthesized proteins. The likely explanation is that the expression of defense proteins increases in these conditions and Hsp70 and Hsp60 are need for their proper folding (Wang *et al*., 2004). It would be interesting to cross the data found here with the data collected for the defense proteins expression profile to search for expression patterns. In addition, Hsp90 has been implicated in biotic stress tolerance in plants, since it interacts with proteins which confer resistance to pathogens (Liu *et al*., 2004; Lu *et al*., 2003). Our analyses showed that Hsp90 expression was not significantly altered by the presence of *Thyrinteina* grub, indicating that the nature of the biotic stress influences Hsp90 expression.

Figure 3 also shows other libraries that have an increased expression profile of proteins associated to chaperones: roots from seedlings (RT3 had 1.8% of the total ESTs annotated as chaperones), stem tissues from six-month seedlings susceptible to hydric deficit (ST2 had 1.6%) and from seedlings susceptible to hydric-deficit (ST6 had 2.0%). Since these tissues are either under development (and therefore require an increased rate of protein synthesis and activation) or facing stress conditions, we believe it is important to point out the increased expression of chaperones that are requested for nascent protein proper folding (Hsp60 family) or protein activation (Hsp90 family) as expected for members of theses families (Boston *et al*., 1996; Buchner, 1999).

Final Considerations

The data generated by the Brazilian *Eucalyptus* Genome Sequencing Project Consortium (FORESTs) allowed the evaluation of the most relevant molecular chaperone gene expression levels within a library context, which helped to give clues about the molecular response to diverse physiological and environmental conditions. The genes belonging to the chaperones category had a considerable expression in *Eucalyptus*, which was expected because these proteins have high importance to the cell function. The cytoplasm compartment displays the most representative expression profile concerning molecular chaperones when compared to others. The Hsp70 and Hsp40 families were the most abundant, followed by the Hsp90 family, the Hsp60 family, the Hsp100 family, and the Small Hsp family. In conclusion, we found that stress-related proteins are abundantly expressed and diverse in *Eucalyptus*.

Acknowledgments

We would like to thank the Fundação de Amparo do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for their financial support. TCC and JCB wish to thank FAPESP for fellowships and AOT is grateful to CNPq for a fellowship.

References

Agarwal M, Sahi C, Katiyar-Agarwal S, Agarwal S, Young T, Gallie RR, Sharma VM, Ganesan K and Grover A (2003) Molecular characterization of rice Hsp 101: Complementation of yeast Hsp104 mutation by disaggregation of protein granules and differential expression in *indica* and *japonica* rice types. Plant Mol Biol 51:545-553.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, 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.

Alvim FC, Carolino SM, Cascardo JC, Nunes CC, Martinez CA, Otoni WC and Fontes EP (2001) Enhanced accumulation of BiP in transgenic plants confers tolerance to water stress. Plant Physiol 126:1042-1054.

Apuya NR, Yadegari R, Fischer RL, Harada JJ, Zimmerman JL and Goldberg RB (2001) The *Arabidopsis* embryo mutant *schlepperless* has a defect in the chaperonin-60 gene. Plant Physiol 126:717-730.

Audic S and Claverie J (1997) The significance of digital gene expression profiles. Gen Res 7:986-995.

Borges JC, Fischer H, Craievich AF and Ramos CHI (2005) Low resolution structural study of two human Hsp40 chaperones in solution: DjA1 from subfamily A and DjB4 from subfamily B have different quaternary structures. J Biol Chem 280, in press.

Borges JC, Peroto MC and Ramos CHI (2001) Molecular chaperone genes in the sugarcane expressed sequence database (SUCEST). Genetics and Molecular Biology 24:85-92.

Borges JC and Ramos CHI (2005) Protein folding assisted by chaperones. Protein Pep Lett 12:257-261.

Boston RS, Viitanen PV and Vierling E (1996) Molecular chaperones and protein folding in plants. Plant Mol Biol 32:191-222.

Brehmer D, Rüdiger S, Gässler CS, Klostermeier D, Packschies L, Reinsten J, Mayer MP and Bukau B (2001) Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange. Nat Struct Biol 8:427-432.

Buchner J (1999) Hsp90 & Co. - A holding for folding. Trends Biochem Sci 24:136-141.

De Jong WW, Caspers GJ and Leunissen JA (1998) Genealogy of the alpha-crystallin – Small heat-shock protein superfamily. Int J Biol Macromol 22:151-62.

DeRocher AE, Helm KW, Lauzon LM and Vierling E (1991) Expression of a conserved family of cytoplasmic low molecular weight heat shock proteins during heat stress and recovery. Plant Physiol 96:1038-1047.

Ehrnsperger M, Gaestel M and Buchner J (1998) Structure and function of small heat shock proteins. In: Fink AL and Goto Y (eds) Molecular Chaperones in the Life Cycle of Proteins. Dekker, New York, pp 533-575.

Ellis RJ and Hartl FU (1996) Protein folding in the cell: Competing models of chaperonin function. FASEB J 10:20-26.

Fink AL (1999) Chaperone mediated protein folding. Physiol Ver 79:425-449.

Gottesman S, Clark WP, de Crecy-Lagard V and Maurizi MR (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*: Sequence and *in vivo* activities. J Biol Chem 268:22618-22626.

Gottesman S, Squires C, Pichersky E, Carrington M, Hobbs M, Mattick JS, Dalrymple B, Kuramitsu H, Shiroza T and Foster T (1990) Conservation of the regulatory subunit for the Clp ATP-dependent protease in prokaryotes and eukaryotes. Proc Natl Acad Sci USA 87:3513-3517.

Grantcharova V, Alm EJ, Baker D and Horwich AL (2001) Mechanisms of protein folding. Curr Opin Struct Biol 11:70-82.

Guy CL and Li QB (1998) The organization and the evolution of the spinach stress 70 molecular chaperone gene family. Plant Cell 10:539-556.

Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381:571-580.

Hill JE and Hemmingsen SM (2001) *Arabidopsis thaliana* type I and II chaperonins. Cell Stress Chap 6:190-200.

Horwich A (2002) Protein folding in the cell. In: Richards FM, Eisenberg DS and Kuriyan J (eds) Advances in Protein Chemistry. Academic Press, California, pp 175-180.

Hsieh M-H, Chen J-T, Jinn T-L, Chen Y-M and Lin C-Y (1992) A class of soybean low molecular weight heat shock proteins. Immunological study and quantitation. Plant Physiol 99:1279-1284.

Hubert DA, Tornero P, Belkhadir Y, Krishna P, Takahashi A, Shirasu K and Dangl JL (2003) Cytosolic Hsp90 associates with and modulates the *Arabidopsis* RPM1 disease resistance protein. EMBO J 22:5679-5689.

Jensen RE and Johnson AE (1999) Protein translocation: Is Hsp70 pulling my chain? Curr Biol 9:779-782.

Kim R, Lai L, Lee HH, Cheong GW, Kim KK, Wu Z, Yokota H, Marqusee S and Kim SH (2003) On the mechanism of chaperone activity of the small heat-shock protein of *Methanococcus jannaschii*. Proc Natl Acad Sci USA 14:8151-8155.

Krishna P and Gloor G (2001) The Hsp90 family of proteins in *Arabidopsis thaliana*. Cell Stress Chap 6:238-246.

Lee GJ and Vierling E (2000) A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. Plant Physiol 122:189-197.

Lin BL, Wang JS, Liu HC, Chen RW, Meyer Y, Barakat A and Delseny M (2001) Genomic analysis of the Hsp70 superfamily in *Arabidopsis thaliana*. Cell Stress Chap 6:201-208.

Liu Y, Burch-Smith T, Schiff M, Feng S and Dinesh-Kumar SP (2004) Molecular chaperone Hsp90 associates with resistance protein N and its signaling proteins SGT1 and Rar1 to modulate an innate immune response in plants. J Biol Chem 279:2101-2108.

Lu Z and Cyr DM (1998) Protein folding activity of Hsp70 is modified differentially by the Hsp40 co-chaperones Sis1 and Ydj1. J Biol Chem 273:27824-27830.

Lu R, Malcuit I, Moffett P, Ruiz MT, Peart J, Wu AJ, Rathjen JP, Bendahmane A, Day L and Baulcombe DC (2003) High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. EMBO J 22:5690-5699.

Mayer K and Mewes HW (2002) How can we deliver the large plant genomes? Strategies and perspectives. Curr Opin Plant Biol 5:173-177.

Miernyk JA (2001) The J-domain proteins of *Arabidopsis thaliana*: An unexpectedly large and diverse family of chaperones. Cell Stress Chaperones 6:209-218.

Mogk A, Tomoyasu T, Goloubnoff P, Rüdger S, Röder D, Langen H and Bukau B (1999) Identification of thermolabile *Escherichia coli* proteins: Prevention and reversion of aggregation by DnaK and ClpB. EMBO J 18:6934-6949.

Morimoto RI (1998) Regulation of heat shock transcriptional response: Cross talking between a family of transcription factors, molecular chaperones and negative regulators. Genes Dev 12:3788-3796.

Netzer WJ and Hartl FU (1998) Protein folding in the cytosol: Chaperonin-dependent and independent mechanisms. Trends Biochem Sci 23:68-73.

Nogueira FTS, De Rosa Jr VE, Menossi M, Ulian EC and Arruda P (2003) RNA expression profiles and data mining of sugarcane response to low temperature. Plant Physiol 132:1811-1824.

Pearl LH and Prodromou C (2001) Structure, function, and mechanism of the Hsp90 molecular chaperone. Adv Protein Chem 59:157-186.

Queitsch C, Sangster TA and Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. Nature 417:618-624.

Schirmer EC, Lindquist S and Vierling E (1994) An *Arabidopsis* heat-stress protein complements a thermotolerance defect in yeast. Plant Cell 6:1989-1909.

Schulte TW, Blagosklonny MV, Romanova L, Mushinski JF, Monia BP, Johnston JF, Nguyen P, Trepel J and Neckers LM (1996) Destabilization of Raf-1 by geldanamycin leads to disruption of the Raf-1-MEK-mitogen-activated protein kinase-signalling pathway. Mol Cell Biol 16:5839-5845.

Siegers K, Waldmann T, Leroux MR, Grein K, Shevchenko A, Schiebel E and Hartl FU (1999) Compartmentation of protein folding *in vivo*: Sequestration of non-native polypeptide by the chaperonin-GimC system. EMBO J 18:75-84.

Sugino M, Hibino T, Tanaka Y, Nii N, Takabe T and Takabe T (1999) Overexpression of DnaK from a halotolerant cyanobacterium *Aphanotece halophytica* acquires resistance to salt stress in transgenic tobacco plants. Plant Sci 146:81-88.

Sun W, Van Montagu M and Verbruggen N (2002) Small heat shock proteins and stress tolerance in plants. Bioch et Bioph Acta 1577:1-9.

Sung DY, Vierling E and Guy CL (2001) Comprehensive expression profile analysis of the *Arabidopsis* Hsp70 gene family. Plant Physiol 126:789-800.

Wang W, Vinocur B and Altman A (2003) Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. Planta 218:1-14.

Wang W, Vinocur B and Altman A (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic response. Trends Plant Sci 9:244-252.

Waters ER, Lee GJ and Vierling E (1996) Evolution, structure and function of the small heat shock proteins in plants. J Exp Bot 47:325-338.

Zhao R, Davey M, Hsu YC, Kaplanek P, Tong A, Parsons AB, Krogan N, Cagney G, Mai D, Greenblatt J, Boone C, Emili A and Houry WA (2005) Navigating the chaperone network: An integrative map of physical and genetic interactions mediated by the Hsp90 chaperone. Cell 120:715-27.

Associate Editor: Marcio de Castro Silva Filho