Research Article

Boron transport in *Eucalyptus*. 2. Identification *in silico* of a putative boron transporter for xylem loading in eucalypt

Douglas Silva Domingues[1,$], Susi Meire Maximino Leite[1,#], Ana Paula Cazerta Farro[1],
Virgínia Elias Coscrato[1], Edson Seizo Mori[2], Edson Luiz Furtado[2], Carlos Frederico Wilcken[2],
Edivaldo Domingues Velini[2], Iraë Amaral Guerrini[2], Ivan Godoy Maia[1] and Celso Luis Marino[1]

[1]*Universidade Estadual Paulista, Instituto de Biociências, Departamento de Genética, Botucatu, SP, Brazil.*
[2]*Universidade Estadual Paulista, Faculdade de Ciências Agronômicas, Botucatu, SP, Brazil.*

Abstract

Boron (B) is a low mobility plant micronutrient whose molecular mechanisms of absorption and translocation are still controversial. Many factors are involved in tolerance to Boron excess or deficiency. Recently, the first protein linked to boron transport in biological systems, *BOR1*, was characterized in *Arabidopsis thaliana.* This protein is involved in boron xylem loading and is similar to bicarbonate transporters found in animals. There are indications that *BOR1* is a member of a conserved protein family in plants. In this work, FORESTS database was used to identify sequences similar to this protein family, looking for a probable *BOR1* homolog in eucalypt. We found five consensus sequences similar to *BOR1*; three of them were then used in multiple alignment analysis. Based on amino acid similarity and *in silico* expression patterns, a consensus sequence was identified as a candidate *BOR1* homolog, helping deeper experimental assays that could identify the function of this protein family in *Eucalyptus*.

Key words: Eucalyptus, boron, xylem loading, Expressed Sequence Tag (EST), *BOR1*.

Received: May 28, 2004; Accepted: April 20, 2005.

Introduction

Boron (B) is an essential nutrient for plant development, required mainly to maintain cell wall integrity. Boron deficiency or toxicity causes productivity losses in different crops around the world (Nuttall, 2000; Leite, 2003). Many problems of Boron deficiency or excess are linked to its low mobility in the majority of cultivated plants (Ruiz, 2001). Considering these facts, plants genes involved in Boron absorption capacity and transport are potential targets in plant breeding.

Until recently, the most commonly accepted theory for boron uptake was that boric acid only entered in root apoplast (extracellular space) by passive transport. However, Nuttall (2000), Dordas *et al.* (2000) and Dordas and Brown (2000) have shown that boron absorption can also occur by facilitated diffusion, through transmembrane channels - the aquaporins (Chrispeels *et al.*, 1999).

Once in the root cortical region, borate $(B(OH)^-_4)$ makes a radial movement to reach the xylem. The portion absorbed by passive transport diffusion travels across the root through simplast (intracellular region linked by plasmodesmata). The boron absorbed by apoplast first needs to enter the cell (simplast) to reach the xylem due to the Casparian band, an apoplast barrier in the endoderm. When these solutes enter the xylem, they return to the apoplast, since vase elements are made of dead cells.

The process in which a nutrient leaves simplast and enters the xylem through an ion-efflux channel is called xylem loading (Peres, 2002). This seems to be a key step in the accumulation of ions in shoots, as demonstrated for phosphate (Poirier *et al.*, 1991; Hamburger *et al.*, 2002) and potassium (Gaymard *et al.*, 1998).

BOR1, characterized by Takano *et al.* (2002), is the first protein linked to boron transport in biological systems and is related to boron xylem loading. Among the ten *BOR1* hypothetical transmembrane domains, Takano *et al.* (2002) found a difference of two amino acids in the second transmembrane domain of the putative protein expressed by *Arabidopsis* mutants which requires higher levels of boron.

Send correspondence to Celso Luis Marino. Universidade Estadual Paulista, Instituto de Biociências, Departamento de Genética, 18618-000 Botucatu, SP, Brazil. E-mail: clmarino@ibb.unesp.br.
#Universidade de Marília, Faculdade de Ciências Agrárias, Marília, SP, Brazil.
$Present address: Universidade de São Paulo, Instituto de Biociências, Departamento de Botânica, São Paulo, SP, Brazil.

Database searches show that proteins related to Cl[-]/HCO3[-] ion-exchangers and Na[+]/HCO3[−] co-transporters in yeast and mammals are similar to *BOR1*, but there is no other characterized protein similar to *BOR1* in plants. However, EST searches using *BOR1* as a query found many similar expressed sequences in monocots and dicots. These data, as suggested by Takano *et al.* (2002), indicate that *BOR1* is probably a member of a family of highly conserved membrane proteins in plants.

In this work, we show the first report of ESTs highly similar to *BOR1* in forest trees, using FORESTs database to identify and characterize the putative expression of ESTs related to *BOR1* family in eucalypt. We gave special emphasis to the identification of a consensus sequence that corresponds to a putative homolog of *BOR1* in *Eucalyptus.*

Material and Methods

ESTs database

We used FORESTs database (https://forests.esalq. usp.br) as a source of eucalypt ESTs. It was composed of 123,889 partial cDNA sequences from various *Eucalyptus* tissues, and grouped in 33,080 clusters. Descriptions of the cDNA libraries and sequence nomenclature are described in https://forests.esalq.usp.br.

Data mining of *BOR1*-related sequences

The amino acid sequence of the putative protein codified by *BOR1* (Access BAC20173 in Genbank) was used as a query to identify FORESTs clusters consensi in a tBLASTn search (Altschul *et al.*, 1997). The minimum criteria for annotation was an e-value lower than e-50. The sequences related to *BOR1* had been translated using the ESTScan tool (Iseli *et al.*, 1999) and revalidated by BLASTP. The sequences whose translation contained the transmembrane domain, linked to difference in boron absorption by Takano *et al.* (2002), were selected for multiple alignment analysis.

Multiple alignment

In multiple alignment, we used identified eucalypt sequences and the expressed sequences used by Takano *et al.* (2002) for comparison with *BOR1* gene. We extended these comparisons to include other expressed sequences from the Genbank dbEST and SUCEST databases (Vettore *et al.*, 2003). The expressed sequences from SUCEST and GenBank were also translated into putative proteins using ESTScan2 (Iseli *et al.*, 1999). The conserved domain of this protein family was identified using the MEME program (Bailey and Elkan, 1994). These domains and 10 amino-acids on each side of the conserved region were aligned using CLUSTALX (Thompson *et al.*, 1997). We used these data to draw a neighbor-joining dendrogram (Saitou and Nei, 1987) with a bootstrap test of 1000 replicates on MEGA2 software (Kumar *et al.*, 2001). It is important to note that this strategy was chosen because there was no EST that covered the full-length sequence of *BOR1* protein.

Data normalization

In the annotated *Eucalyptus* contigs, the read frequency from each tissue was normalized dividing the number of reads in each tissue by the total number of sequenced reads in the tissue, and the result multiplied by 100,000.

Inference of ESTs expression patterns

The normalized annotated contig data were used to study expression patterns in each tissue by hierarchical clustering (Eisen *et al.*, 1998). This was performed using a non-centered relational matrix and the average-linkage method, through Cluster program v2.20 (Eisen *et al.*, 1998). Data was displayed in TreeView program v 1.60 (http://rana.lbl.gov/EisenSoftware.htm). We also used the Audic and Claverie (1997) method to give statistical support to expression patterns based on the number of ESTs identified in different libraries; we considered $p < 0.05$ as the cutoff value to identify contigs with differential expression in any tissue.

Results

This study identified five eucalypt consensus sequences related to *BOR1* in FORESTs database (Table 1), according to established criteria. Three clusters have the transmembrane domain indicated by Takano *et al.* (2002) related to xylem loading boron transport deficiency.

Table 1 - Eucalypt EST contigs related to *BOR1* in tBLASTn search showing their respective e-value, identity, number of gaps and shared region with the query protein. Data were confirmed with translated sequences in BLASTp.

Contig	tBLASTn				BLASTp result
	e-value	Identity (amino acids)	Gaps (amino acids)	*BOR1* match (amino acids)	
EGEQRT3002E03.g	0.0	431/552 (78%)	24/552 (4%)	1-539	*BOR1*
EGEQRT3302A02.g	e-164	217/283 (76%)	2/283 (0%)	254-534	At3g62270
EGEQRT3102H04.g	e-109	152/292 (52%)	39/292 (13%)	377-635	At3g62270
EGEQFB1002H05.g	2e-91	223/404 (55%)	31/404 (7%)	39-635	At1g15460
EGEQBK1086D09.g	3e-65	128/160 (80%)	1/160 (0%)	1-159	*BOR1*

Figure 1 - Multiple alignment of the conserved domain in proteins that contained the second transmembrane domain of *BOR1*, with 10 amino acids added on each side of the common domain, was run in the Clustal X program. Genbank access (gi) of each sequence is indicated in each line. On the right is the analyzed region length.

In a BLASTP analysis of the hypothetical proteins codified by these clusters, all of them returned *BOR1* and/or related sequences in their first hit (Table 1).

Contigs EGEQRT330A02.g and EGEQRT3102H04.g, which did not contain the transmembrane domain, had *BOR1* as the 2nd best hit in BLASTP. They presented stronger similarity with a sequence in chromosome 3 of *Arabidopsis* whose function is still not characterized. From the sequences that contained the transmembrane domain, only contig EGEQFB1002H05.g did not have *BOR1* as a first match in BLASTP analysis. This contig had greater similarity to a sequence of chromosome 1 of *Arabidopsis* - also not characterized.

The conserved domain found with the aid of the MEME program in the deduced aminoacid sequences used in the multiple alignment analysis, corresponded to part of the second transmembrane region of *BOR1* protein. Incorporating 10 aminoacids at each end of the conserved region permitted evaluation of the transmembrane domain extension and its adjacent regions in the multiple alignment (Figure 1). We used this analysis to draw a dendrogram (Figure 2) that gives some indication of the possible relationship among members of the putative *BOR1* protein family. Contig EGEQRT300E03.g, which was preferentially expressed in roots according to the Audic and Claverie (1997) statistical test (Table 2), was grouped in the same clade as *BOR1*. Although the bootstrap value was lower than 50%, contig EGEQFB1002H05.g, found only in

the flower library, grouped with two hypothetical proteins of chromosome 1 of *Arabidopsis thaliana* (Figures 2 and 3). Curiously, contig EGEQBK1086D09.g, statistically more expressed in the BK (bark) library (Table 2), was grouped with two barley sequences.

Concerning the expression patterns, all the sequenced libraries presented ESTs related to the putative *BOR1* family, with the exception of stems susceptible and resistant to frost and water deficiency (Figure 4). If we consider the normalized data, these reads were more expressed in the roots and BK libraries. However, we must consider that the

Table 2 - Contigs with differential expression in FORESTs database, analyzed by the Audic and Claverie statistic test (1997), with $p < 0.05$. Legend: RT - root; SL - seedlings; ST - stem; BK - bark, xylem, heartwood and medulla; CL - callus; FB - flower bud; LV - leaves; WD - wood.

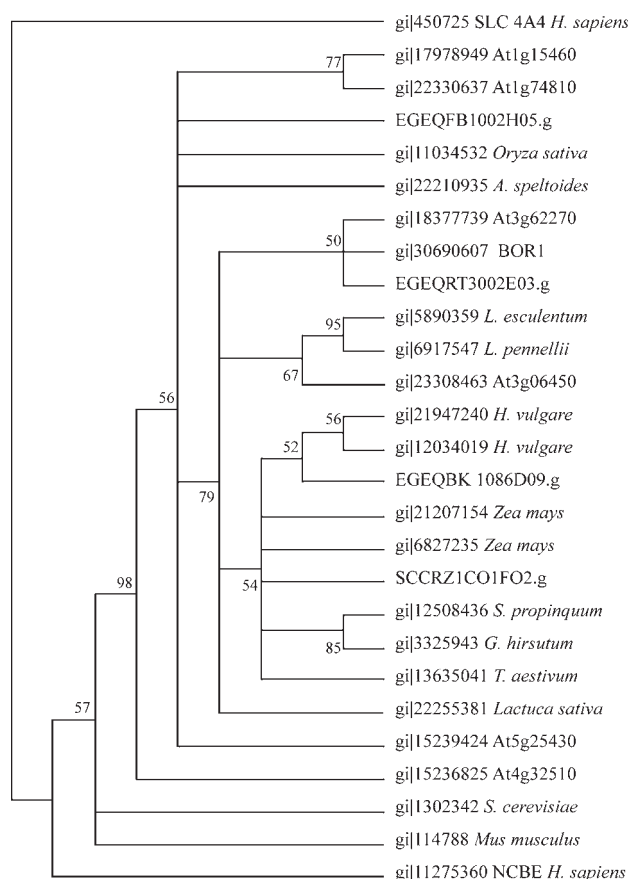| Cluster	Library differentially expressed	Significance
EGEQRT3002E03.g	RT > SL	0.03
	RT > ST	0.04
EGEQBK1086D09.g	BK > CL	0.02
	BK > FB	0.02
	BK > LV	0.02
	BK > RT	0.005
	BK > SL	0.008
	BK > ST	0.003
	BK > WD	0.02

Figure 2 - Neighbor-joining tree showing the relationship between hypothetical conserved domains in *BOR1*-like sequences and putative *Eucalyptus* proteins. The tree was drawn using MEGA2, with bootstrap of 1000 replicates. Only bootstrap values of 50% or higher are shown. Eucalypt sequences were represented as in FORESTs database (https://forests.esalq.usp.br) or SUCEST database (http://sucest.lad.ic.unicamp.br); other sequences were represented by their access in GenBank (GI) and the relevant species.
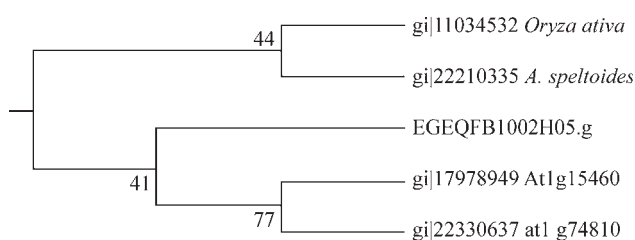

Figure 3 - Subtree detailing the relationship between contig EGEQFB 1002H05.g and hypothetical proteins of *Arabidopsis thaliana*. Bootstrap values were shown in percentages. Sequences were represented with their access in GenBank (GI) and the relevant species.

number of sequenced reads in the BK library was much lower than all other libraries, which could cause sampling deviation in expression patterns and reinforces the idea of a probable preferential expression in roots.

Discussion

Data analyses supported the hypothesis that the *Arabidopsis thaliana* xylem loading boron transporter is
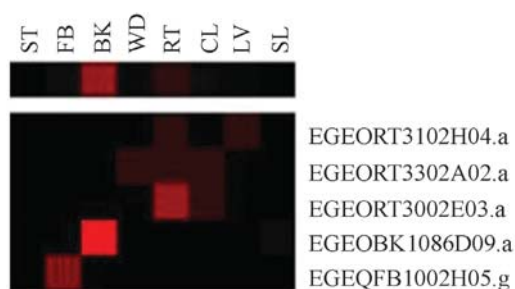

Figure 4 - Hypothetical expression pattern of sequences similar to *BOR1* found in FORESTs database, evaluated with normalized data. Each EST contig is represented in a line and each tissue with sequenced cDNAs in a column. Black means no expression, and different intensity red means the expression intensity of each read in a tissue. Legend: ST - stem; FB - flower bud; BK - bark, xylem, heartwood and medulla; WD - wood; RT - root; CL - callus; LV - leaves; SL - seedlings.

part of a conserved gene family in plants that codifies putative membrane proteins, probably related to anion transport. These proteins could be involved in other biological processes related to the efflux of boron or other ions. This is similar to phosphate xylem loading, as proposed by Hamburger *et al.* (2002).

The dendrogram and multiple alignment analysis distinguished animal and yeast proteins from the plant sequences, although it did not distinguish monocot from dicot plant sequences. This is probably due to the comparison methodology, which did not focus on distinguishing homologous from paralog genes - sequences that are produced by gene duplication and can be also related to *BOR1*, but have different functions (Gibas and Jambeck, 2002). Nevertheless, we observed some groupings that indicate a possible phylogenetic relationship; for example, the *Lycopersicon esculentum* and *L. pennellii* sequences that grouped with a chromosome 3 sequence of *A. thaliana* (At3g06450), and the sugarcane sequence which only has one aminoacid different from the maize sequences (Figure 1).

Eucalypt contigs that did not show the transmembrane region studied by Takano *et al.* (2002) are probably homologous to the At3g62270 protein, codified in chromosome 3 of *A. thaliana* and similar to *BOR1*, in accordance with the multiple alignment data. Of the three eucalypt contigs in the neighbor-joining tree (Figure 2), only EGEQBK1086D09.g did not confirm the BLASTP analysis. This cluster had *BOR1* as a first hit in BLASTP, but grouped in the dedrogram with *Hordeum vulgare* (barley) sequences. This discrepancy could be related to the conserved region selected for multiple alignment, which could have generated this artifact. Its expression pattern was also different from the other clusters (Figure 3), and possibly caused a statistical deviation in expression analyses. If we exclude this contig from the analysis, there is a preferential expression of *BOR1*-related contigs in root.

The dendrogram confirmed the local alignment (BLASTP) analyses with EGEQFB1002H05.g and EGEQ RT300E03.g consensus sequences (Figure 2). Considering that the bootstrap value was close to 50%, EGEQFB10

02H05.g, found only in the flower library, showed indications of being related to At1g15460 and At1g74810 proteins in *Arabidopsis*. In the same manner, the sequence that grouped better with *BOR1* in the dendrogram and BLASTP cluster was EGEQRT300E03.g. This is the most probable candidate to contain the partial sequence of the *BOR1* eucalypt homolog. Our analyses also suggest that the deduced protein of this contig may perform the same function as *BOR1* in *Arabidopsis thaliana*, based on the *in silico* analyses of expression patterns and the large *BOR1* region that is covered by the translation of this expressed sequence. Thus, experimental studies that involve the sequences assembled in this cluster could be important in characterizing the function of this putative protein.

Conclusion

The identification of eucalypt ESTs similar to *BOR1* reinforces the idea that this gene is a member of a conserved gene family in plants, expressed in different tissues, and involved in diverse biological processes. These data indicate the need for more detailed studies to understand the function of these ion transporters in different plant tissues, mainly in root. We also concluded that FORESTs database could be an initial source of information for studies that intend to follow this path.

Acknowledgments

References

Altschul SF, Madden TL, Schaffer 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.

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

Bailey TL and Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc 2nd Int Conf Intell Syst Mol Biol 28-36.

Chrispeels MJ, Crawford NM and Schroeder JI (1999) Proteins for transport of water and mineral nutrients across the membranes of plant cells. Plant Cell 11:661-675.

Dordas C and Brown PH (2000) Permeability of boric acid across lipid bilayers and factors affecting it. J Membr Biol 175:95-105.

Dordas C, Chrispeels MJ and Brown PH (2000) Permeability and channel-mediated transport of boric acid across plant membrane vesicles isolated from squash roots. Plant Physiol 124:1349-1362.

Eisen MB, Spellman PT, Brown PO and Botstein, D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863-14868.

Gaymard F, Pilot G, Lacombe B, Bouchez D, Bruneau D, Boucherez J, Michaux-Ferrière N, Thibaud J-B and Sentenac H (1998) Identification and disruption of a plant shaker-like outward channel involved in K$^+$ release into the xylem sap. Cell 94:647-655.

Hamburger D, Rezzonico E, Petétot JMC, Somerville C and Poirier Y (2002) Identification and Characterization of the *Arabidopsis* PHO1 gene involved in phosphate loading to the xylem. Plant Cell 14:889-902.

Iseli C, Jongeneel CV and Bucher P (1999) ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol 138-148.

Kumar S, Tamura K, Jakobsen IB and Nei M (2001) MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics 17:1244-1245.

Leite, SMM (2003) Comportamento de *Eucalyptus grandis* e híbridos "*urograndis*" submetidos à situações de suficiência e deficiência de boro. PhD Thesis, Instituto de Biociências, Universidade Estadual Paulista, Botucatu, Brazil.

Gibas, C and Jambeck P (2002) Desenvolvendo Bioinformática. Editora Campus, Rio de Janeiro, 440 pp.

Nuttall, CY (2000) Boron tolerance & uptake in higher plants. PhD Thesis, Department of Plant Sciences & Gonville and Caius College, University of Cambridge.

Peres, LEP (2002) Absorção e transporte de nutrientes pelas raízes. Revista Universa, 3:45-66.

Poirier Y, Thoma S and Schiefelbein J (1991) A mutant of *Arabidopsis* deficient in xylem loading of phosphate. Plant Physiol 97:1087-1093.

Ruiz JM (2001) Aquaporin and its function in boron uptake. Trends Plant Sci 6:95.

Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

Takano J, Noguchi K, Yasumori M, Kobayashi M, Gajdos Z, Miwa K, Hayashi H, Yoneyama T and Fugiwara T (2002) *Arabidopsis* boron transporter for xylem loading. Nature 420:337-340.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997) The CLUSTALX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876-4882.

Vettore AL, Silva FR, Kemper EL, Souza GM, Silva AM, Ferro MIT, Silva FH, Giglioti EA, Lemos MVF, Coutinho LL, Nóbrega MP, Carrer H, França SC, Bacci Jr M, Goldman MHS, Gomes SL, Nunes LR, Camargo LEA, Siqueira WJ, Van Sluys MA, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon MLPN, Ferro JA, Silveira HCS, Marini DC, Lemos EGM, Monteiro-Vitorello CB, Tambor JHM, Carraro DM, Roberto PG, Martins VG, Goldman GH, Oliveira RC, Truffi D, Colombo CA, Rossi M, Araújo PG, Sculaccio SA, Angella A, Lima MMA, Rosa VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SMZ, Nobrega FG, Menck CFM, Braga MDV, Telles GP, Cara FAA, Pedrosa G, Meidanis J and Arruda P (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Research 13:2725-2735.

Associate Editor: Luiz Eduardo Aranha Camargo