

A new member of the chalcone synthase (CHS) family in sugarcane

Miriam G.G. Contessotto, Claudia B. Monteiro-Vitorello, Pilar D.S.C. Mariani and Luiz L. Coutinho*

Abstract

Sequences from the sugarcane expressed sequence tag (SUCEST) database were analyzed based on their identities to genes encoding chalcone-synthase-like enzymes. The sorghum (*Sorghum bicolor*) chalcone-synthase (CHS, EC 2.3.1.74) protein sequence (gi|12229613) was used to search the SUCEST database for clusters of sequencing reads that were most similar to chalcone synthase. We found 121 reads with homology to sorghum chalcone synthase, which we were then able to sort into 14 clusters which themselves were divided into two groups (group 1 and group 2) based on the similarity of their deduced amino acid sequences. Clusters in group 1 were more similar to the sorghum enzyme than those in group 2, having the consensus sequence of the active site of chalcone and stilbene synthase. Analysis of gene expression (based on the number of reads from a specific library present in each group) indicated that most of the group 1 reads were from sugarcane flower and root libraries. Group 2 clusters were more similar to the amino acid sequence of an uncharacterized pathogen-induced protein (PI1, gi|9855801) from the *S. bicolor* expressed sequence tag (EST) database. The group 2 clusters sequences and PI1 proteins are 90% identical, having two amino acid changes at the chalcone and stilbene synthase consensi but conserving the cysteine residue at the active site. The PI1 EST has not been previously associated with chalcone synthase and has a different consensus sequence from the previously described chalcone synthase of sorghum. Most of the group 2 reads were from libraries prepared from sugarcane roots and plants infected with *Herbaspirillum rubrisubalbicans* and *Gluconacetobacter diazotrophicans*. Our results indicate that we have identified a sugarcane chalcone synthase similar to the pathogen-induced PI1 protein found in the sorghum cDNA libraries, and it appears that both proteins represent new members of the chalcone and stilbene synthase super-family.

INTRODUCTION

Chalcone synthase (CHS) operates early in the biosynthetic pathway of flavonoids, secondary metabolites which play important roles in the interactions which occur between plants and their environment. Peters *et al.*, (1986) have shown that this synthase is involved in pigment formation, symbiosis, and plant defenses against pathogen attack and exposure to ultra-violet light. Shirley (1996) states that chalcone synthase is involved in flavonoid synthesis, where it catalyzes the condensation of acetate groups from malonyl-CoA with 4-coumaroyl-CoA to form naringenin-chalcone, leading to the formation of flavonols, flavonones, isoflavonoids and anthocyanins. Lanz *et al.* (1991) have pointed out that this enzyme is well-conserved among plants of different groups, and that it has a cysteine residue at amino acid 169 that is thought to be part of the 4-coumaroyl-CoA binding site and which is required for enzyme activity. Phytoalexins (which play a role in plant-microbe interactions) can be derived from the flavonoid pathway and activation of the plant defense-response can often be detected by checking for the accumulation of chalcone synthase mRNA or by measuring the activity of this enzyme after inoculating a plant with microorganisms (Cui *et al.*, 1996).

The chalcone synthase genes present a high degree of sequence similarity at the amino acid level and have been the object of numerous studies in dicotyledonous plants, where up to seven copies have been identified in several species (Koes *et al.*, 1987; 1989; Durbin *et al.*, 2000). In the monocotyledonous grasses, most of the genera studied (*i.e.* *Zea*, *Oryza*, *Hordeum* and *Secale*) have two copies of the chalcone synthase gene (Wienand *et al.*, 1986; Franken *et al.*, 1991; Rhode *et al.*, 1991), although seven copies have been identified in *Sorghum bicolor* by Lo and Nicholson (1999), (gi|5305906|, gi|5305908|, gi|5305910|, gi|5305912|, gi|5305914|, gi|5305916|, gi|5305918|). Durbin *et al.* (2000) have pointed out that chalcone synthase is suitable for studies of gene duplication and investigations on the origin of gene families.

The sugarcane genus, *Saccharum*, is complex and characterized by high polyploidy and frequent aneuploidy. The genome of modern sugarcane cultivars have a chromosome number between 70 and 120, the cultivars being derived essentially from interspecific hybridization involving different species *e.g.* *Saccharum officinarum*, *Saccharum barberi*, *Saccharum sinense* and the wild species *Saccharum spontaneum* and *Saccharum robustum*. The fact that the chalcone synthase gene family is not well characterized in graminea (Oberholzer *et al.*, 2000) and the complexity of

the sugarcane genome makes sugarcane an interesting system to study this gene family. In the work presented in this paper, we used sequence similarity and expression profiling to show that sugarcane has two groups of chalcone synthase sequences and that one of these groups represents a new member of the chalcone synthase family that may be induced by the interaction between plants and microbes.

MATERIAL AND METHODS

Sequence analysis and gene expression

The complete sugarcane expressed sequence tag (SUCEST) database (<http://sucest.lbi.dcc.unicamp.br>) with 291149 expressed sequence tags (ESTs) was searched with sorghum protein sequence (gi|12229613|) from Genbank (<http://www.ncbi.nlm.nih.gov/>).

All of the analyses were done using the cluster consensi database. Sequences were assembled at SUCEST using the fragment assembly program Phrap (Green, 1996) as described by Telles and Silva (2001). Comparisons of sequence homology between SUCEST and GenBank (<http://www.ncbi.nlm.nih.gov/>) sequences were performed using various basic local alignment search tool (BLAST) programs (*e.g.* tBLASTn) and multi-alignments were done at (<http://ca.expasy.org>) using the CLUSTALW program. The Pfam (<http://www.sanger.ac.uk/Pfam/>) and Prosite (<http://ca.expasy.org/prosite/>) protein family databases were used to recognize the domains.

The SUCEST database is made up of sequencing reads from clones of cDNA libraries prepared from the mRNA of a variety of tissues in different stages of development growing under different conditions, some of the libraries having as many as 18,000 reads. To investigate gene expression, the total number of sequences in a specific library was compared with the total number of reads in each cluster of a given library. We searched EST libraries from different sugarcane tissues, *i.e.* the flower (FL) libraries of 47,887 reads; the root (RT) libraries of 23,168 reads; the AD and HR libraries (containing a total of 22,320 reads) from plantlets infected with *Gluconacetobacter diazotrophicans* or *Herbaspirillum rubrisubalbicans*, respectively; the apical meristem (AM) libraries 20,078 reads; the seed (SD) library of 15,617 reads; the stem first internode (ST) library of 14,532 reads; the lateral bud (LB) library of 12,821 reads and the stem bark (SB) library of 11,313 reads.

RESULTS

Searching the SUCEST database for chalcone synthase cDNA

The *S. bicolor* chalcone synthase sequence (gi|12229613), of 401 amino acids, was compared to SUCEST database cluster consensi using the tBLASTn program. We found that 121 sugarcane ESTs showed high levels of sequence similarity to sorghum chalcone

synthase. Based on sequence identity we were able to form 14 clusters, which we divided into two groups, group 1 contained ten clusters (SCMCRT2085C06.g, SCEQRT1028G10.g, SCEZRT2018G01.g, SCQGRT1039D07.g, SCQSST3116F03.g, SCJFRT2059C02.g, SCVPRT2082E12.g, SCUTAM2006A07.g, SCUTSB1030C06.g, SCSFRT2072D11.g) and represents sugarcane chalcone synthase 1 (sugarcane CHS1) while group 2 contained four clusters (SCEQRT1029F12.g, SCCCRT2002C08.g, SCVPRT2073E08.g, SCEZHR1049A11.g) and represents sugarcane chalcone synthase 2 (sugarcane CHS2).

We found that the deduced amino acid sequence of sugarcane chalcone synthase 1 was 402 amino acids long and 97% identical to all seven *S. bicolor* chalcone synthases described (gi|5305906|, gi|5305908|, gi|5305910|, gi|5305912|, gi|5305914|, gi|5305916|, gi|5305918|), 95% identical to *Zea mays* chalcone synthase c2 (Franken *et al.*, 1991) and 93% identical to *Oryza sativa* chalcone synthase. We examined the 5' end nucleotide sequence of four full-length clusters and found differences of up to five base additions. All ten deduced open reading frames (ORFs) of the group 1 clusters presented differences in amino acid composition, suggesting that there may be ten different members of this enzyme in sugarcane.

Regarding sugarcane chalcone synthase 2, we found that the deduced amino acid sequence was 398 amino acids long and 81% identical to rice chalcone synthase. By searching the GenBank dEST database with the nucleotide sequence of sugarcane chalcone synthase 2 we found that the sugarcane enzyme shared sequence identity with a 596 base-pair (bp) long pathogen induced EST (PI1, GenBank sequence BE600826) from *Sorghum bicolor*. The sorghum EST has not been previously identified as chalcone synthase and we found that it has a different amino acid composition at the consensi signature compared with all chalcone synthases yet identified in sorghum. We found that the deduced amino acid sequence of PI1 was 90% identical to our sugarcane chalcone synthase 2. Two of our four full-length clusters were identical at the 5' end, while the other two had one base substitution and two base additions. All four deduced ORFs of the group 2 clusters had differences in amino acid composition, and it seems that in this case there may be less than four chalcone synthase 2 homologues.

Comparison of sugarcane chalcone synthases 1 and 2 with chalcone synthase genes from other organisms

All members of the chalcone and stilbene synthase families present conserved sequence features. For example, Lanz *et al.* (1991) have shown that the major domain of these enzymes is the well conserved active site with the consensus sequence R-[LIVMFYS]-x-[LIVM]-x-[QHG]-x-G-C-[FYNA]-[GAPV]-G-[GA]-[STAV]-x-[LIVMF]-[RAL], where C is the active site residue. When we aligned the active site of the sugarcane chalcone synthases 1 and 2 to

several other chalcone synthases (Table I), we found that the C residue at the center of the sequence aligned with the sorghum P11 protein and the bcsA protein from *Bacillus subtilis* (gi|16079263|), although there were two conserved changes in the P11 protein and sugarcane chalcone synthase 2 that did not match the conserved signature around the active residue. Jez and Noel *et al.*, 2000 have reported that three other residues (His³⁰⁷, Asn³⁴⁰ and Phe²¹⁹) form the catalytic center of chalcone synthase and are conserved in all chalcone synthase-like enzymes. We found that sugarcane chalcone synthases 1 and 2 all had the strictly conserved residues at the correct positions along with regions important for secondary structure organization (Table II).

Sequencing analysis and expression of sugarcane chalcone synthase 2

Figure 1 shows that the full-length cDNA of sugarcane chalcone synthase 2 is 1637bp (based on cluster analysis) with the translational start codon being at nucleotide 105 and the TGA stop codon at nucleotide 1301. This ORF has a G+C content of 59% and encoding a polypeptide of 398 amino acids. At the translational start the sequence presents the characteristic adenine residue at position -3 and cytosine residue at position +5 and has a putative polyadenylation signal (AATAAC) 124 nucleotides before the poly-adenine (poly A+) tail at base-pair position 1454-1460.

The sequence similarity and the conserved alterations at the consensus signature of sugarcane chalcone synthase

Table I - Alignment of sugarcane chalcone synthases (CHS) amino acid sequence with other chalcone synthases.

Gene	Alignment of the sugarcane chalcone synthase 159-176 amino acid region with regions from other chalcone synthases
<i>Saccharum officinarum</i> (CHS1)	159 RLMMYQQGCF AGGTVLR 176
<i>Oryza sativa</i>	158 RLMMYQQGCF AGGTVLR 175
<i>Arabidopsis thaliana</i>	160 RLMMYQQGCF AGGTVLR 177
<i>Sorghum bicolor</i>	159 RLMMYQQGCF AGGTVLR 176
<i>Sorghum cereale</i>	158 RLMMYQQGCF AGGTVLR 175
<i>Zea mays</i> (WHP1)	158 RLMMYQQGCF AGGTVLR 175
<i>Zea mays</i> (C2)	159 RLMMYQQGCF AGGTVLR 176
<i>Camellia sinensis</i>	154 RLMMYQQGCF-AGGTVLR 171
<i>Glycine max</i>	155 RYMMYQQGCF AGGTVLR 172
<i>Hordeum vulgare</i>	157 RLMMYQQGCF GGATVLR 174
<i>Saccharum officinarum</i> (CHS2)	159 <u>R</u> AMMY <u>H</u> QGCF AGG <u>M</u> VLR 176
<i>Sorghum bicolor</i> (P11)	159 <u>R</u> AMMY <u>H</u> QGCF AGG <u>M</u> VLR 176
<i>Bacillus subtilis</i>	135 RIPIWGLGCA GGASGLA 152
Consensus sequence	R.MMY.QGCF AGG.VLR

The common residue C, is shown in bold and the A, H, and M characteristic residues of CHS2 and P11 are underlined.

Table II - The five characteristic motifs of plant chalcone synthase (CHS) and the distance between them in sugarcane CHS 1 and 2. The different regions correspond to the loops of the secondary structure of the enzyme.

Motif	Amino acid positions and sequences of the five characteristic motifs of sugarcane CHS
Active site	159- R(L/A)MM(Y/H)QQGCF AGG(T/M)VLR -176
Region I	217- LFGDG - 222
Region II	302- FWVAHPGGPA - 312
Region III	338- GNMSSA - 344
Region IV	374- FGFGPGL - 381

Letters in bold-face type indicate amino acid residues which are strictly conserved in all members of chalcone and stilbene synthase family.

```

gtcttgaattcccgggtcgaccacgcgctccgcaaaaagcttgtt 45
ggccacagctagcttgtttagttagtctctgtgccagcagc 90
atccggtaggaaa 104
105 atgactgggaaggtaacattggaggcggtaggaaggcgcagcgc
M T G K V T L E A V R K A Q R
150 gccgagggatctgctgcggtgttggccattgggacggcgacacc
A E G S A A V L A I G T A T P
195 gcaaacctgcggtatcaggtgactaccpggactacttccgg
A N C V Y Q A D Y P D Y Y F R
240 gtcaccaagagcgaacaccttaccgacctcaagaaaaatcaag
V T K S E H L T D L K E K F K
285 aggatagtctacaagtcgatgattacgaagcgttacatgcat
R I C Y K S M I T K R Y M H L
330 actgagggtctctacaggagaacccaacatgagctcgtactcc
T E G F L Q E N P N M S S Y S
375 gcaccatccctagacgcacgcaggatctcctgatagaggagta
A P S L D A R Q D I L I E E V
420 cccaagctggcgcgccggcgccgagaaagcgtcaaaagatgg
P K L G A A A A E K A L K E W
465 ggccagccagcttccagatcacgcaacctgcttctgcaccacc
G Q P R S Q I T H L V F C T T
510 tctggcgtggacatgcctggcgccgactaccagctcatcaagta
S G V D M P G A D Y C G T I K L
555 ctcggtctatccccctctgtgaagcgagcgtatgtaccaccaa
L G L S P S V K R A M M Y H Q
600 ggttgettcgcccgggaatggtgctcgttgcgaaggacct
G C F A G G M V L R L A K D L
645 gccgagaacaaccgtggtgcccgggtgctcatcgtgctctgag
A E N N R G A R V L I V C S E
690 atcactgcggtcacattccgccccttcggagctcacccttgac
I T A V T F R G P S E S H L D
735 tccgttgttgccaagctctctcgtgacggcgagctgcggtg
S L V G Q A L F G D G A A A V
780 atcgtcgtgacagcccaagcagcagcagagggccatgttcc
I V G A D P S E P A E W P L F
825 cagctagtatcagcgagccagaccttccccaaactcggagggt
Q L V S A S Q T I L P N S E G
870 gccatcgaggccacctcctgaggtggggctcactccatctc
A I E G H L R E V G L T F H L
915 caggacagggttccacagcttatctccacgaacattgagcgttg
Q D R V P Q L I S T N I E R L
960 ctggaagacgcttccacgcgctagcctctctgattggaactcc
L E D A F T P L G I S D W N S
1005 atcttttgggtggcgaccctcctgagctcagccataactgacatg
I F W V A H P G G P A I L N M
1050 gtggaggccaaggctggccttgacaagcccgactlctgacacc
V E A K A G L D K A R L C A T
1095 cgccacatcctcgagagatggcaacatgtcaagcgcctgtgtc
R H I L A E Y G N M S S A C V
1140 ctcttcatcctgatgagatcgcaaaacagctgccaaaggacgga
L F I L D E M R N K S A K D G
1185 cacaccacaactggggaggcagtgagggtgggtgtcctctcggc
H T T T G E G M E W G V L F G
1230 tttggccccgctcaccgtcgagaccatgcttctcagagcgtt
F G P G L T V E T I V L Q S V
1275 cccatcaccacagtgacgcggcatga
P I T T V H A A *
cogttgtgcatcagtgcatgtagctccgctgacgcatggtcgtg 1346
gcaatcactctacagctactttacagttatcttggtttctcacc 1391
acaaaatgtactgctagaagcacttgcagtggttttgggtg 1436
ttgagctgtcttcttaccgaataaactatggtgtgtgctgaggat 1481
ctgtgctagtgttaaatggataaagcagtgacagtgcaatacag 1526
gaagtggggctgtaataattatagttatgtcaacaagagacaatg 1571
gaattgaaatggcaaaaaaaaaaaaaaaaaataaaaataaa 1613
    
```

Figure 1 - Possible full-length cDNA sequence of sugarcane chalcone synthase 2 (CHS 2). The translational start signal and polyadenylation signal are in bold-face and underlined.

2 and the sorghum P11 protein suggests that sugarcane chalcone synthase 2 may be a candidate for use in plant-microbe interaction studies.

There are sequenced reads from 83 EST clones in chalcone synthase 1 and from 38 EST clones in chalcone synthase 2. We analyzed the differential expression of chalcone synthase 1 and chalcone synthase 2 in different tissues based on the number of sequenced reads from a given library occurring within each of the enzymes (Figure 2). This analysis showed that these enzymes exhibit a tendency to differential expression. It can be seen in Figure 2 that 0.13% of root (RT) libraries and 0.04% of flower (FL) libraries are chalcone synthase 1. There is no sequencing reads of chalcone synthase 1 in the lateral bud (LB) library. There were more chalcone synthase 2 reads (0.06%) than chalcone synthase 1 reads (0.02%) in the AD and HR libraries which had been constructed from plantlets infected with *Gluconacetobacter diazotrophicans* (the AD library) or *Herbaspirillum rubrisubalbicans* (the HR library), two microbial endophytes associated with nitrogen fixation (Reinhold-Hurek and Hurek, 1998), and there were no chalcone synthase 2 reads in the flower library.

Chalcone synthase is a plant-specific polyketide synthase important for the biosynthesis of anti-microbial isoflavonoid phytoalexins, anthocyanin floral pigments and flavonoid inducers of *Rhizobium* nodulation genes (Dixon and Paiva, 1995; Long, 1989). Considering the number of sequenced clones of these libraries (AD1+HR1 = 22,320 and FL = 44,887) and the number of clones that are within each group, we have a three-fold increase in expression of the chalcone synthase 2 in sugarcane infected with microorganisms and no expression in flowers when comparing with chalcone synthase 1 (Figure 2). The work presented in this paper has shown that sugarcane chalcone synthase 2 is very similar to the pathogen-induced sorghum P11 protein, and we propose that sugarcane chalcone synthase 2 has a

different expression pattern and tissue specificity than sugarcane chalcone synthase 1 and that the expression pattern and tissue specificity of sugarcane chalcone synthase 2 is associated with the interaction between plants and bacteria. It is also probable that the same phenomenon exists in sorghum, with two distinct types of chalcone synthase being coded for in the genome of *Sorghum bicolor*.

DISCUSSION

Chalcone synthase which is important for the biosynthesis of secondary metabolites has been studied in various species, and some of them have several copies of this enzyme (Wienand *et al.*, 1986; Franken *et al.*, 1991; Rhode *et al.*, 1991). In our study we have characterized chalcone synthase-like enzymes of sugarcane and found that these enzymes can be divided into two different groups, each with several members.

The chalcone synthase group 1 is the well-conserved chalcone synthases found in plants and is highly homologous to sorghum chalcone synthases. In sorghum there are seven different genes (gi|5305907|, gi|5305909|, gi|5305911|, gi|5305913|, gi|5305915|, gi|5305917|, gi|5305919|). Working with maize, Franken *et al.* (1991) have described two chalcone synthase genes, c2 and Whp (white pollen), which are 94% identical and have the same active site as other chalcone synthases but different expression patterns and tissue specificity, the c2 gene being expressed mostly in developing kernels while the Whp gene is normally expressed in the tassel. According to our research, there may be ten different members of the sugarcane chalcone synthase 1 group which are expressed in the seeds, roots, flowers and tassels of sugarcane plants.

The sugarcane chalcone synthase 2 group may contain new members of the chalcone synthase family. We base this view on the changes, which we found in the chalcone family signature, being conserved in the sorghum EST deduced P11 protein. All the other features of the putative sugarcane chalcone synthase 2 sequences are the same as those which have been proposed to be essential for chalcone synthases, these features are conserved in sugarcane chalcone synthases 1, 2 and the sorghum P11 (BE600826) sequence.

From the distributions of sequencing reads within the chalcone synthase 1 and 2 groups it appears that members of these groups have different patterns of gene expression. In sugarcane, we found increased chalcone synthase 2 expression in plants infected with microorganisms. The presence of a similar pathogen-induced protein in sorghum cDNA libraries supports the view that in sugarcane, as in maize, chalcone synthase may be expressed for different purposes. It is also interesting to note that the gene sequencing study of Selman-Housein *et al.*, (1999) placed sugarcane closer to maize than to sorghum, but our sequence analysis of chalcone synthases indicates that sugarcane is closer to sorghum than to maize.

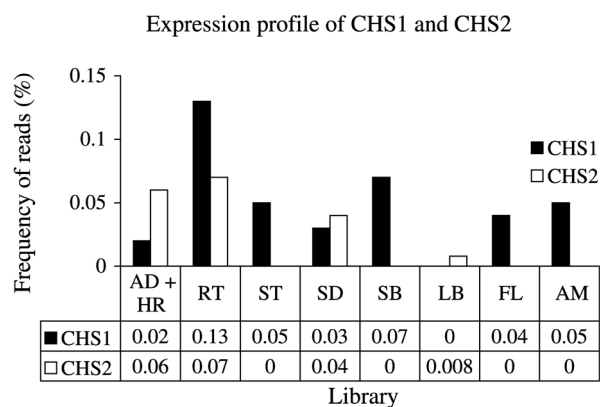


Figure 2 - Occurrence and frequency of chalcone synthase 1 (CHS1) and 2 (CHS2) in different sugarcane tissues. Library codes: AD+HR = plantlets infected with *Gluconacetobacter diazotrophicans* and *Herbaspirillum rubrisubalbicans* (22320 reads), RT = root (23168 reads), ST = stem first internode (14532 reads), SD = seed (15617 reads), SB = stem bark (11313 reads), LB = lateral bud (12821 reads), FL = flower (47887 reads), AM = apical meristem (20078 reads).

The research presented in this paper is the first report in grasses of differences in the sequence and expression patterns of chalcone synthase-like enzymes related plant-microbe interaction. To confirm the results presented in this paper the EST databases of sorghum and other plants could be examined to confirm whether or not the same chalcone synthase sequence and expression patterns that we found in sugarcane exist in other crops.

RESUMO

Seqüências do banco de dados SUCEST foram analisadas baseado em suas identidades com genes relacionados que codificam enzimas chalcone sintase (CHS). A seqüência da proteína CHS de sorgo (gi|12229613), foi usada para a busca de seqüências no banco de dados do SUCEST, que foram mais similares à CHS. Baseado na similaridade das seqüências de aminoácidos deduzidas, quatorze conjuntos formados por 121 seqüências, foram divididos em dois grupos. O primeiro grupo é mais similar à CHS de sorgo (EC 2.3.1.74) e apresenta a seqüência consenso do sítio ativo de chalcone e stilbene sintase. Análises da expressão gênica, baseada no número de seqüências de uma dada biblioteca presente em cada grupo, indicaram que neste caso, a maioria das seqüências são de bibliotecas preparadas de raiz e flores de cana-de-açúcar. O segundo grupo é mais similar à seqüência de aminoácidos deduzida de uma proteína, não caracterizada, induzida por patógeno (PI1, gi|9855801) do banco de dados de EST de *Sorghum bicolor*. Estas duas seqüências de proteínas são 90% idênticas e tem duas mudanças de aminoácidos no consenso padrão de chalcone e stilbene sintase, e conservam o resíduo de cisteína na posição do sítio ativo. Esta EST não foi associada a chalcone sintase antes, e apresenta diferenças na seqüência consenso da CHS previamente descrita de sorgo. A maioria das seqüências do segundo grupo são de bibliotecas de cana-de-açúcar preparadas de raiz e plantas infectadas com *Herbaspirillum rubrisubalbicans* e *Gluconacetobacter diazotrophicans*. Os resultados indicam que nós identificamos uma chalcone sintase, tal como a encontrada no banco de cDNA de sorgo em uma biblioteca induzida por patógeno, expressa em plantas de cana-de-açúcar, e que ambas as proteínas são novos membros da super família chalcone e stilbene sintase.

ACKNOWLEDGMENTS

This research was supported by a grant from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). The authors gratefully acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the masters fellowship conceded to M.G.G. Contessotto and the research productivity scholarship to

L.L. Coutinho and FAPESP for the post-doctoral fellowship to C.B. Monteiro-Vitorello.

REFERENCES

- Cui, Y., Magill, R., Frederiksen, R. and Magill, C. (1996). Chalcone synthase and phenylalanine ammonia-lyase mRNA levels following exposure of sorghum seedlings to three fungal pathogens. *Phys. Mol. Plant Path.* 49: 187-199.
- Dixon, R.A. and Paiva, N.L. (1995). Stress-induced phenylpropanoid metabolism. *Plant Cell* 7: 1085-1097.
- Durbin, M.L., McCaig, B. and Clegg, M.T. (2000). Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Mol. Biol.* 42: 79-92.
- Franken, P., Niesbach-Kloegen, U., Weydemann, U., Marechal-Drouard, L., Saedler, H. and Wienand, U. (1991). The duplicated chalcone synthase genes C2 and Whp (white pollen) of *Zea mays* are independently regulated; evidence for translational control of Whp expression by the anthocyanin intensifying gene. *EMBO J.* 10: 2605-2612.
- Green, P. (1996). <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>.
- Jez, J.M. and Noel, J.P. (2000). Mechanism of Chalcone synthase. pKa of the catalytic cysteine and the role of the conserved histidine in a plant polyketide synthase. *J. Biol. Chem.* 275: 39640-39646.
- Koes, R.E., Spelt, C.E., Mol, J.N.M. and Gerats, A.G.M. (1987). The chalcone synthase multigene family of *Petunia hybrida* V30 sequence homology chromosomal localization and evolutionary aspects. *Plant Mol. Biol.* 10: 159-169.
- Koes, R.E., Spelt, C.E., Van den Elzen, P.J.M. and Mol, J.N.M. (1989). Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. *Gene* (Amsterdam) 81: 245-257.
- Lanz, T., Tropf, S., Marner, F.-J. and Schroder, G. (1991). The role of cysteines in polyketide synthases. *J. Biol. Chem.* 15: 9971-9976.
- Lo, S.-C.C. and Nicholson, R.L. (1999). Molecular cloning of chalcone synthase genes from sorghum. (Unpublished) (<http://www.ncbi.nlm.nih.gov/>).
- Long, S.R. (1989). Rhizobium-legume nodulation - Life together in the underground. *Cell* 56: 203-214.
- Oberholzer, V., Durbin, M.L. and Clegg, M.T. (2000). Comparative genomics of chalcone synthase and Myb in the grass family. *Genes Genet. Syst.* 75: 1-16.
- Peters, N.K., Frost, J.W. and Long, S.R. (1986). A plant flavone, luteolin induces expression in *Rhizobium meliloti* nodulation genes. *Science* 233: 977-980.
- Reinhold-Hurek B. and Hurek T. (1998). Life in grasses: diazotrophic endophytes. *Trends Microbiol.* 6: 139-44.
- Rhode, W., Dorr, S., Salamini, F. and Becker, D. (1991). Structure of a chalcone synthase gene from *Hordeum vulgare*. *Plant Mol. Biol.* 16: 1103-1106.
- Selman-Housein, G., López, M.A., Hernández, D., Civardi, L., Miranda, F., Rigau, J. and Puigdomènech, P. (1999). Molecular cloning of cDNAs coding for three sugarcane enzymes involved in lignification. *Plant Science* 143: 163-171.
- Shirley, B.W. (1996). Flavonoid biosynthesis: 'new' function for an 'old' pathway. *Trends in Plant Science* 1: 377-382.
- Telles, G.P. and Silva, F.R. (2001). Trimming and clustering sugarcane ESTs. *Gen. Mol. Biol.* 24 (1-4): 17-23.
- Wienand, U., Weydemann, U., Niesbach-Kloegen, U., Peterson, P.A. and Saedler, H. (1986). Molecular cloning of the c2 locus of *Zea mays*, the gene coding for chalcone synthase. *Mol. Gen. Genetics* 203: 203-207.