Genetics and Molecular Biology, 30, 3 (suppl), 848-856 (2007)

www.sbg.org.br

Research Article

Citrus plastid-related gene profiling based on expressed sequence tag analyses

Tercilio Calsa Jr.* and Antonio Figueira

Laboratório de Melhoramento de Plantas, Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, SP, Brazil.

Abstract

Plastid-related sequences, derived from putative nuclear or plastome genes, were searched in a large collection of expressed sequence tags (ESTs) and genomic sequences from the *Citrus* Biotechnology initiative in Brazil. The identified putative *Citrus* chloroplast gene sequences were compared to those from *Arabidopsis*, *Eucalyptus* and *Pinus*. Differential expression profiling for plastid-directed nuclear-encoded proteins and photosynthesis-related gene expression variation between *Citrus sinensis* and *Citrus reticulata*, when inoculated or not with *Xylella fastidiosa*, were also analyzed. Presumed *Citrus* plastome regions were more similar to *Eucalyptus*. Some putative genes appeared to be preferentially expressed in vegetative tissues (leaves and bark) or in reproductive organs (flowers and fruits). Genes preferentially expressed in fruit and flower may be associated with hypothetical physiological functions. Expression pattern clustering analysis suggested that photosynthesis- and carbon fixation-related genes appeared to be up- or down-regulated in a resistant or susceptible *Citrus* species after *Xylella* inoculation in comparison to non-infected controls, generating novel information which may be helpful to develop novel genetic manipulation strategies to control *Citrus* variegated chlorosis (CVC).

Key words: Citrus variegated chlorosis; *in silico*; sweet orange, tangerine.

Received: July 21, 2006; Accepted: February 14, 2007.

Introduction

Citrus species, such as sweet orange, mandarin, lime and lemon, have a large economic and social importance in Brazil, the world largest exporter of concentrated orange juice and producer of fresh fruit. Due to its industrial importance, considerable investment in genomic research in Brazil has been directed toward *Citrus*, including sequencing the complete genome of the first plant pathogen *Xylella fastidiosa* (Simpson *et al.*, 2000), and the recent achievement of 182,529 expressed sequence tags (ESTs), derived from several *Citrus* tissues under various biotic or abiotic stresses, and from random genomic sequences (CitEST Database).

Crop yield potential is ultimately derived from photosynthesis, the process of converting solar energy into carbon backbones performed at thylakoid membranes of leaf chloroplasts (Taiz and Zeiger, 1998). Plastid is the generic name given to a group of specific plant cell organelles, which contain their own genome (plastome), ranging in size from 110 to 180 kbp depending on plant species. Plastids are maternally inherited in most crop species (Pyke, 1999), including *Citrus* (Moreira *et al.*, 2002). The double-stranded chloroplast DNA exhibits complex structural dynamics *in vivo* (Lilly *et al.*, 2001; Bendich, 2004), in contrast to the classic model of a closed circular genome. To date, more than 40 plants and algae plastomes have had their complete sequence complement determined (http://www.ncbi.nlm.nih.gov/Genomes). Plastome sequences have revealed a conserved structure even among distantly related taxa (Ogihara *et al.*, 2002; Calsa *et al.*, 2004), despite the occurrence of evolutionary events, such as plastid gene transfer to nucleus and/or mitochondria and functionally redundant gene loss (Sugiura, 2003).

Plastome sequences have been determined for a few tree species, including *Pinus* (Wakasugi *et al.*, 1994), *Eucalyptus* (Steane, 2005) and *Citrus* (Bausher *et al.*, 2006), opening new approaches for tree crop breeding, since plastids have been recognized as an interesting target for genetic engineering (Ruf *et al.*, 2001; Maliga, 2003; Bock and Khan, 2004). In *Citrus*, plastid DNA regions have usually been used as cytoplasm inheritance markers in somatic hybridization and cybrid development (Guo *et al.*, 2004; Takami *et al.*, 2004). Nuclear-encoded proteins with plastid

Send correspondence to Antonio Figueira. Laboratório de Melhoramento de Plantas, Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Av. Centenário 303, Caixa Postal 96, 13400-970 Piracicaba, SP, Brazil. E-mail: figueira@cena.usp.br.
*Current address: Laboratório de Genética Molecular, Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Recife, PE, Brazil.

activity have been investigated in *Citrus*, especially associated with oxidative stress response (Mullineaux *et al.*, 1998) and carotenoid biosynthesis (Tao *et al.*, 2005).

A plastid transcript termination signal consists of secondary structures containing short poly-A sequences (Rott *et al.*, 1998), which allow their eventual capture with oligo-dT primers, mainly from photosynthetic tissues, given the high ploidy and copy number of plastid genome in leaves. Plastid mRNAs may also be isolated under post-transcriptional control, because poly-adenylation is a degradation signaling mechanism (Hayes *et al.*, 1999). As already verified in EST, SAGE and MPSS plant databases from *Arabidopsis* and sugarcane (Meyers *et al.*, 2004; Robinson *et al.*, 2004; Calsa and Figueira, 2007), tens or even hundreds of sequences displayed identity or high similarity to plastome genes or regions. These sequences may be actual plastid transcripts, or derived from nuclear genes resulting from DNA transfer events to the nucleus, a proven and quantified process (Stegemann *et al.*, 2003). Potential *Citrus* plastome-encoded and plastid-directed nuclear-encoded expressed sequences in various organs were analyzed in this work. In addition, differential gene expression linked to photosynthesis from leaves infected or not with *Xylella fastidiosa*, the causal agent of *Citrus* variegated chlorosis (CVC) was analyzed.

Material and Methods

The *Citrus* sequence database (CitEST) was searched using the software GeneProject. Because the *Citrus* chloroplast genome sequence was not yet publicly available at the time of analysis, sequences potentially derived from the *Citrus* plastome were searched using the complete chloroplast genome sequences from *Arabidopsis thaliana*, *Eucalyptus globulus*, *Pinus thunbergii*, and *Pinus koraiensis* (GenBank AP000423; AY780259; D17510; and AY228468, respectively) through BlastN of the CitEST, accepting matches with E-value < 10^{-5}. Positive matching reads were clustered using an internal CAP3 within GeneProject and annotated by BlastX against public protein databases. Sequences related to nuclear genes encoding plastid-targeted products were recovered by keyword search "('chloroplast OR plastid') AND precursor."

Retrieved sequences were assessed considering the standard CitEST nomenclature for reads and cDNA libraries, derived from leaves, bark, fruits or flowers of *Citrus sinensis* (CS libraries). The analyzed leaf expressed sequences included libraries of non-infected *Citrus sinensis* 'Pera IAC' ('CS-100') or 30 days after inoculation with *Xylella fastidiosa* ('CS-102'); and non-infected *Citrus reticulata* 'Ponkan' ('CR-100') or 30 days after inoculation with *Xylella fastidiosa* ('CR-102'). Leaf ('C1') reads putatively associated to thylakoid membrane photosynthetic systems and primary carbon fixation were manually selected.

ESTs with identical or extremely similar putative annotations were counted, and their frequency was normalized for the total number of ESTs in each corresponding library, and expressed on a per thousand basis. The normalized frequencies of distinct libraries were statistically tested for differences based on Audic and Claverie (1997), considering significant p-value < 0.05. The normalized frequencies of contrasting libraries were also analyzed for expression pattern by hierarchical clustering and PlotCorr analyses using the Gene Expression Pattern Analysis (GEPAS; Herrero *et al.*, 2003; 2004) online tools. The putative transcripts were arbitrarily categorized into high expression (above 5 reads per thousand); medium expression (between 2 and 5 reads per thousand); and low expression (below 2 reads per thousand) before hierarchical clustering.

Results

Citrus plastome reads and partial shot-gun assembly

BlastN search of *Citrus* ESTs and random genomic sequences using the plastomes from *E. globulus*, *P. thunbergii*, *P. koraiensi* and *Arabidopsis* returned 362 significantly matched reads. From those, 155 (42.8%) were genomic sequences, while 207 were derived from ESTs from various organs, likely representing sequences transcribed from *Citrus* plastome of distinct plastid types (Table 1).

Plastome-matched *Citrus* sequences presented distinct similarity to the four queried species (Table 2). About one-fifth (19.3%) from all *Citrus* plastome-related sequences presented counterparts in Angiosperm dicotyledonous species (*A. thaliana* and *E. globulus*), while none was exclusively found to be similar with the analyzed Gymnosperm plastomes (*Pinus*). Conversely, the number of exclusive matches between *Citrus* and *Eucalyptus* was ten times higher than between *Citrus* and *Arabidopsis* (Ta-

Table 1 - *Citrus* expressed sequence tags (ESTs) and random genomic sequences with positive matches to other species' plastomes identified based on BlastN with an *E*-value < 10^{-5}.

Library	Genomic	Leaf	Bark	Fruit	Root	Seed
Total reads	3,465	82,585	14,031	62,003	4,185	3,368
Plastome-matched reads	155	133	9	51	2	11
Normalized frequency (%) in library	4.4	0.2	0.1	0.1	0.1	0.3
Percent of plastome-matched reads	42.8	36.7	2.5	14.1	0.6	3.0

ble 2). Noteworthy, no *Citrus* sequence presented a simultaneous match to *P. koraiensis* and *P. thunbergii*, or a joint match to *Pinus* and *Arabidopsis*. Additionally, only one *Citrus* EST presented an exclusive match to the *Arabidopsis* plastome.

The categorization of plastome-related *Citrus* sequences based on putative annotation (Table 3) revealed that the exclusive *Citrus-Arabidopsis* sequence had an unknown putative function (yet to be identified in any other plant transcriptome), and derived from a fruit cDNA library. Most exclusive *Eucalyptus*-matched sequences were associated to 'no hit' transcripts, although a few were categorized as 'hypothetical' or 'ribosomal proteins'. The highest proportion of the *Citrus* sequences with specific matches harbored simultaneous and exclusive similarity to *Arabidopsis* and *Eucalyptus*, both dicotyledonous Angiosperms, with lower resemblance to the Gymnosperm plastomes.

In an attempt to achieve a primary draft of the *Citrus* plastid DNA regions, the plastome-matched EST and genomic reads were assembled into clusters. This approach resulted in 65 contigs and 73 non-grouped sequences or singlets (Table 4). The sum of non-overlapping contigs reached 68,095 bp in size, while singlets altogether covered 55,695 bp. Since the reference *Citrus* chloroplast genome available to date comprises 160,129 bp (Bausher *et al.*, 2006), the *in silico* assembly covered around 77.3% of the plastome in a transcriptionally informative manner.

Citrus chloroplast-related nuclear genes

Through keyword search, a total of 19,246 *Citrus* expressed sequences were found to match known nuclear genes coding for precursor proteins targeted to plastid.

Table 2 - Number of *Citrus* sequences (reads) with significant positive match with the complete chloroplast genome sequences from four species (*Arabidopsis thaliana*, *Eucalyptus globulus*, *Pinus thunbergii*, or *Pinus koraiensis*), indicating distinct level of similarity.

Species plastome	Matched reads	Percent
Common to all 4 species	255	70.4
Arabidopsis + *Eucalyptus* + *P. koraiensis*	8	2.2
Arabidopsis + *Eucalyptus* + *P. thunbergii*	7	1.9
Arabidopsis + *Eucalyptus*	70	19.3
Eucalyptus + *P. koraiensis*	11	3.0
Eucalyptus	10	2.8
Arabidopsis	1	0.8
Total	362	100.0

Based on organ of origin and treatment, it was possible to define a general transcriptional profile for nuclear-encoded plastid-targeted gene products for the various plastid types, such as chloroplast, chromoplast or proplastid in sweet orange (*C. sinensis* 'Pera IAC'). There was no cDNA library prepared from sweet orange root tissues, but the identification of 104 expressed sequences derived from *C. limonia* 'Cravo' roots exposed or not to water deficit, perfectly aligning several putative functionally plastid-related genes (data not shown), indicated that plastome-related sequences might also occur in roots of sweet oranges.

Considering only *C. sinensis* transcripts, it was possible to analyze putative differential gene expression between organs/tissues and, consequently different plastid types, as well as to identify genes with an apparent preferential transcription in sweet orange flowers or fruits. Genes showing the lowest expression differences among leaf,

Table 3 - Distribution of putative annotation of reads matched to the four query species' plastome (*Arabidopsis thaliana*, *Eucalyptus globulus*, *Pinus thunbergii*, or *Pinus koraiensis*.).

Gene category annotation	*At*	*Eg*	*At, Eg*	*Eg, Pk*	*At, Eg, Pk*	*At, Eg , Pt*
No hit	1	5	19	1	0	0
Hypothetical protein	0	1	12	0	1	2
Ycf / ORF	0	0	3	0	1	1
ATPase	0	0	1	0	1	1
NADH-dehydrogenase	0	0	9	0	3	0
Photosystem I	0	0	1	0	0	0
Photosystem II	0	0	1	0	0	0
Cytochrome b6f	0	0	2	0	0	1
RNA polymerase	0	0	0	0	0	1
Ribosomal protein	0	1	6	10	1	0
Maturase	0	0	9	0	0	0
ACCase	0	0	0	0	1	0
Other	0	3	7	0	0	1
Total	1	10	70	11	8	7

At: *Arabidopsis thaliana. Eg*: *Eucalyptus globulus. Pk*: *Pinus thunbergii. Pt*: *Pinus koraiensis.*

Table 4 - Size and number of clusters and singlets from partial *in silico* shot-gun assembly of the *Citrus* plastid genome, from EST and genomic sequences with high similarity to other plastomes.

Size range (bp)	Clusters	Singlets
> 3,000	1	0
2,500 to 3,000	3	0
2,000 to 2,500	2	0
1,500 to 2,000	5	0
1,000 to 1,500	11	3
500 to 1,000	39	61
< 500	4	9
Total	65	73

bark, fruit and flower were putatively annotated as coding for alpha-1,4 glucan phosphorylase; ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) binding-protein; enolpyruvylshikimate-3-phosphate synthase (EPSPS); and cystathionine gamma-synthase (Figure 1A). The normalized frequencies for these genes presented the lowest standard errors among analyzed organs (not shown). On the other hand, the genes displaying the most variable expression were chlorophyll a/b binding proteins; photosystem I subunits; early light-induced protein; and chloroplast terpene synthase (Figure 1B). A sequence annotated as coding for photosystem subunit and another for the hypothetical chloroplast reading frame 19 displayed significantly higher preferential expression in vegetative organs (leaves and bark) than in reproductive organs (Figure 1C). On the other hand, the transcripts with the highest significant preferential expression in reproductive organs (flowers and fruits) were associated to inorganic pyrophosphatase; plastid terpene synthase; thiazole biosynthetic enzyme; and GcpE protein (Figure 1D). Additionally, significant preferential expression in fruits was observed for transcripts putatively encoding lipoxygenase C and isocitrate dehydrogenase (Figure 2A), while a significant preferential transcription in flowers was detected for a chloroplast translocon component, anthranilate synthase and alpha-glucan water dikinase (Figure 2B).

Chloroplast photosynthesis gene expression potentially affected by *Citrus* variegated chlorosis (CVC)

To detect potential relevant transcriptional variations associated with CVC occurrence, cDNA libraries from two

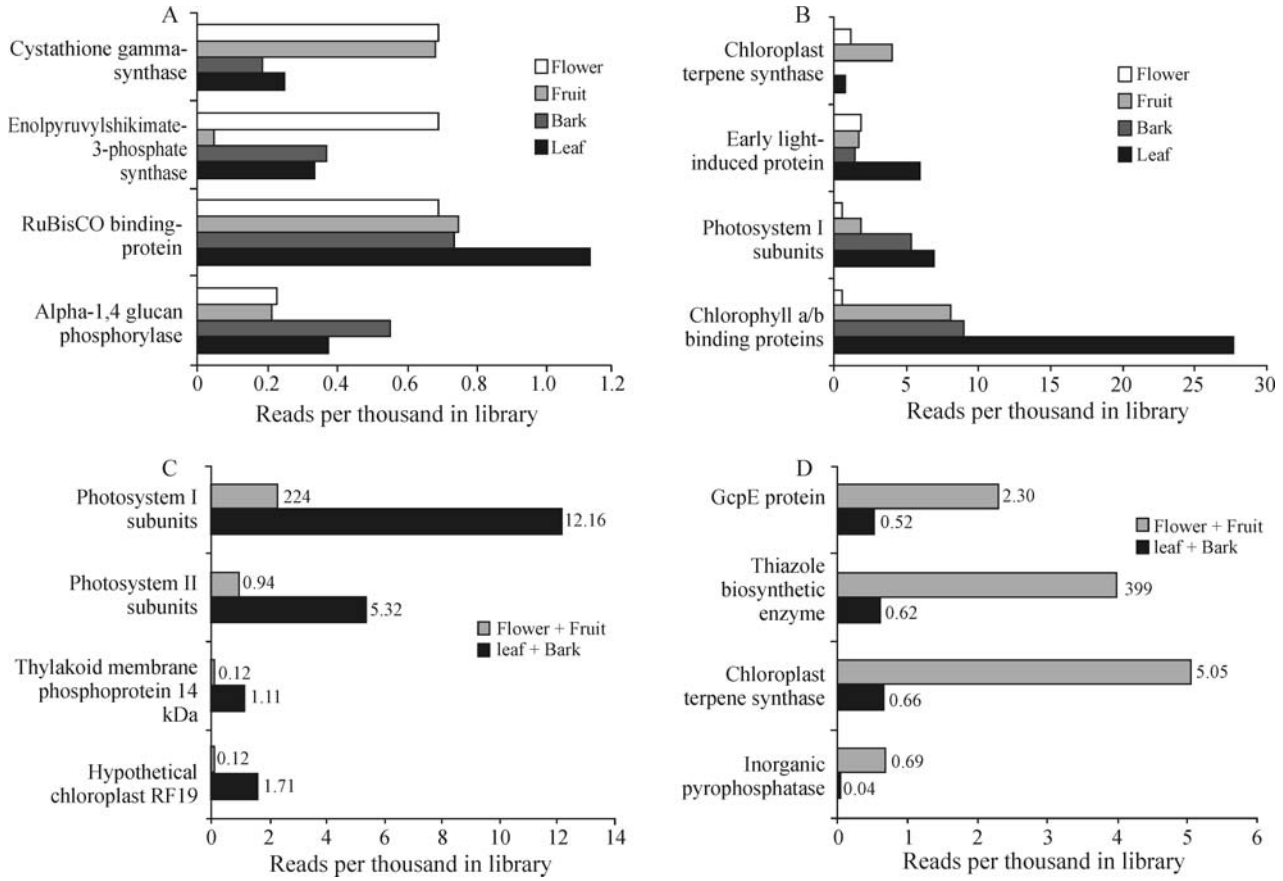

Figure 1 - Preferential expression of *Citrus sinensis* sequences in leaf, bark, fruit and flower-derived cDNA libraries based on normalized frequencies (A, B); and grouped into vegetative or reproductive organs (C, D). The sequences displayed significant differences in expression ($p < 0.05$) based on Audic and Claverie (1997), between vegetative and reproductive organs (C and D).

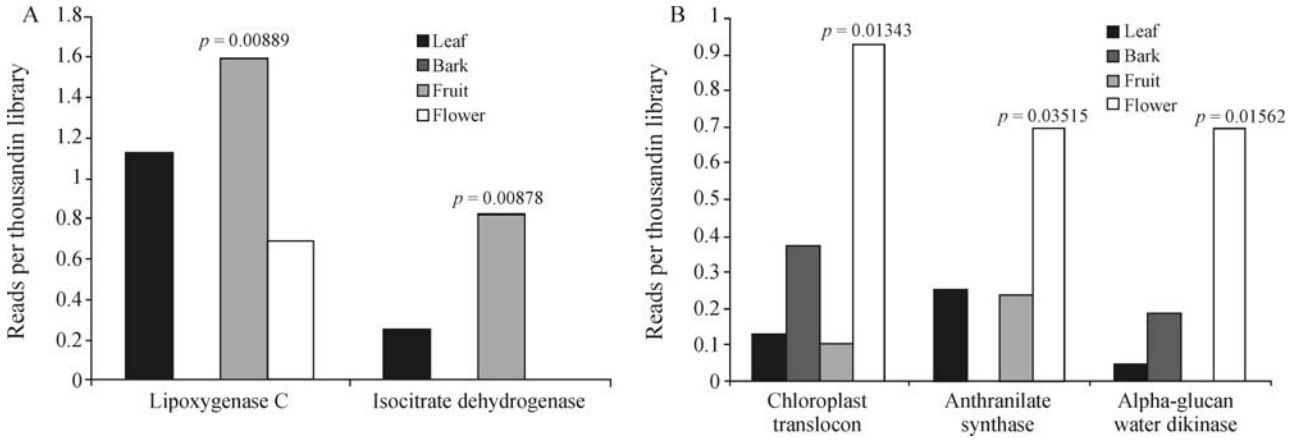

Figure 2 - Normalized expression of *Citrus sinensis* sequences with the highest significant ($p < 0.05$) preferential expression in fruits (A) or flowers (B), based on the test of Audic and Claverie (1997), when compared to the average expression over the other organs.

Citrus species contrasting in their response to *Xylella fastidiosa* (susceptible sweet orange or resistant tangerine), after inoculation or not, were compared for differential expression of photosystem- and carbon fixation-associated genes based on normalized frequencies. From putative annotation, 37 reads associated with thylakoid photosynthesis complex subunits, and 30 reads related with the carbon fixation cycle were identified (not shown). The putative transcripts were arbitrarily categorized into high expression (above 5 reads per thousand); medium expression (between 2 and 5 reads per thousand); and low expression (below 2 reads per thousand). This enabled a more sensitive analysis of hierarchical clustering according to normalized frequency expression (also known as virtual northern). Clustering expression patterns suggested transcriptional variations between *C. sinensis* and *C. reticulata*, infected or not by *Xylella fastidiosa*. Considering that sweet oranges are highly susceptible to CVC, while tangerines are considered to be standard tolerant, gene differential expression might indicate specific reaction to CVC. For example, chlorophyll a/b binding protein 1 (LHC-II type I CAB-1) appeared to be strongly induced after *Xylella* infection in tangerine, whereas the expression level did not appear to change in sweet orange following infection (Figure 3A). Ferredoxin NADP-reductase (leaf isozyme) and Thioredoxin M appeared to be induced in tangerine within the first 30 days after infection (Figure 3B). Among low-expression genes, it was noticed that transcript levels of ATP synthase beta, gamma and delta chains increased with infection in tangerine, while the opposite occurred in sweet orange (Figure 3C). Additionally, chlorophyllase I appeared to be slightly down-regulated in tangerine (Figure 3C).

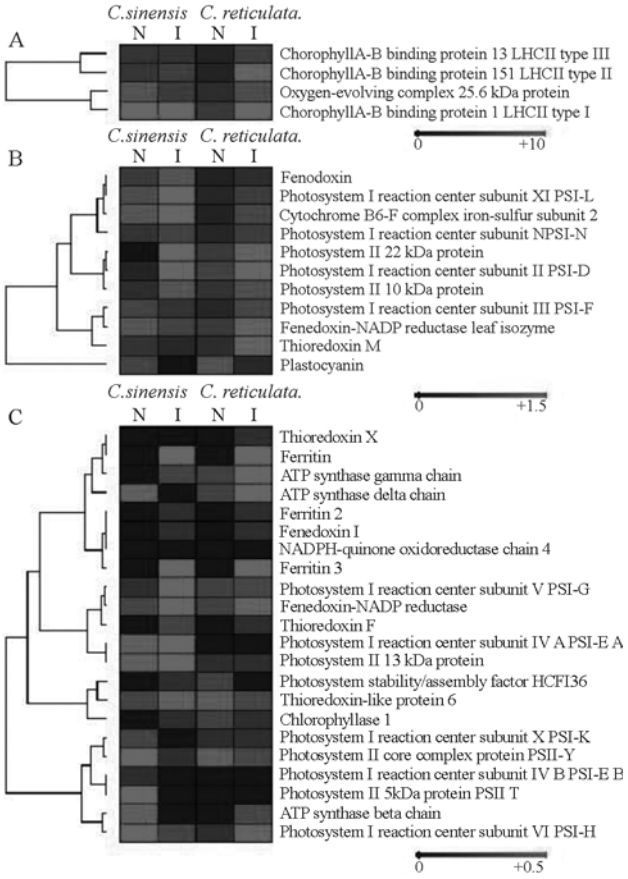

Figure 3 - Hierarchical clusterization by expression pattern of photosystems subunits- and Z scheme-related expressed sequences in *C. sinensis* and *C. reticulata*, infected (I) or not infected (N) with *Xylella fastidiosa*. Black represents 'no expression'; light gray represents 'maximum expression', with gray intensity proportionally representing intermediate expression levels.

Regarding transcripts associated with the carbon fixation cycle, minor differential expression was observed for carbonic anhydrase and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) in sweet orange and tangerine, with slight transcript accumulation after infection (Figure 4A). Ribulose 1,5-bisphosphate carboxylase (RuBisCO)-binding protein alpha subunit and pyruvate phosphate dikinase exhibited opposite expression profiles in both *Citrus* species, with an increase in transcripts in *C. sinensis* after infection with *Xylella* and a corresponding decrease in *C.*

reticulata (Figure 4B). Divergent patterns between both species were also observed for several low-expressed genes, such as ribulose-phosphate 3-epimerase; ribulose 1,5-bisphosphate carboxylase (RuBisCO) small chain 3; ribulose 1,5-bisphosphate carboxylase *N*-methyltransferase; fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase (Figure 4C).

Correlation analyses using PlotCorr indicated the most variable expression between *Xylella*-infected tangerine and sweet orange (Figure 5). Among the photosystem components, the oxygen-evolving complex 25.6 kD protein and chlorophyll a/b binding protein 151 type II showed a more genotype-specific expression, with both transcripts accumulating in infected *C. reticulata* (Figure 5A). Conversely, three genes related to the carbon fixation cycle were detected as presenting a more genotype-specific transcription (Figure 5B). Two (RuBisCO activase and transketolase) displayed a tangerine-preferential expression, while a RuBisCO small chain subunit was significantly more expressed in sweet orange (Figure 5B).

Discussion

In silico mining of plastid-related sequences in the CitEST database identified the presence of organelle-derived reads, with highly significant matches to plastome-specific regions. As expected, the plastome-specific transcripts were present at a low frequency, ranging from ca. 0.1% in ESTs to 4.4% in genomic libraries (Table 1). The data also suggested a stronger transcription of genes from chloroplast and chromoplast, respectively from shoots and fruits, associated with photosynthesis and fruit ripening (Table 1). From the plastome-matched *Citrus* sequences,

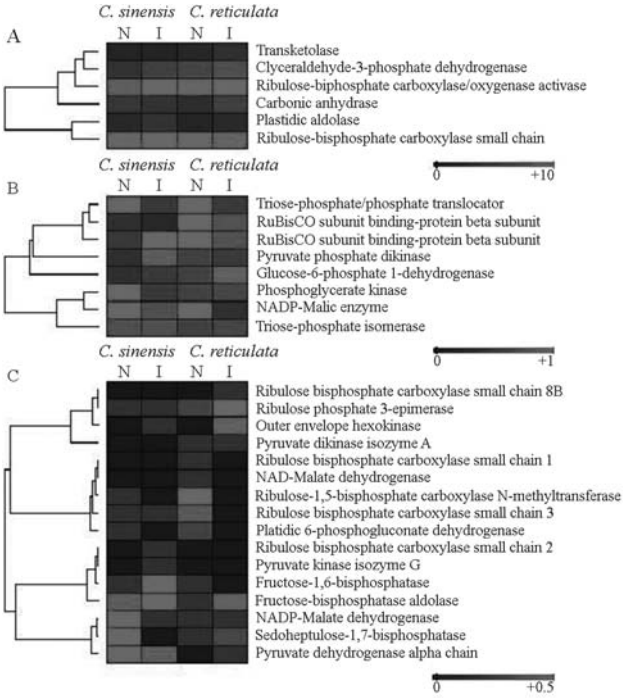

Figure 4 - Hierarchical clusterization by expression pattern of carbon fixation-related expressed sequences in *C. sinensis* and *C. reticulata*, infected (I) or not infected (N) with *Xylella fastidiosa*. Black represents 'no expression'; light gray represents 'maximum expression' with gray intensity proportionally representing intermediate expression levels.

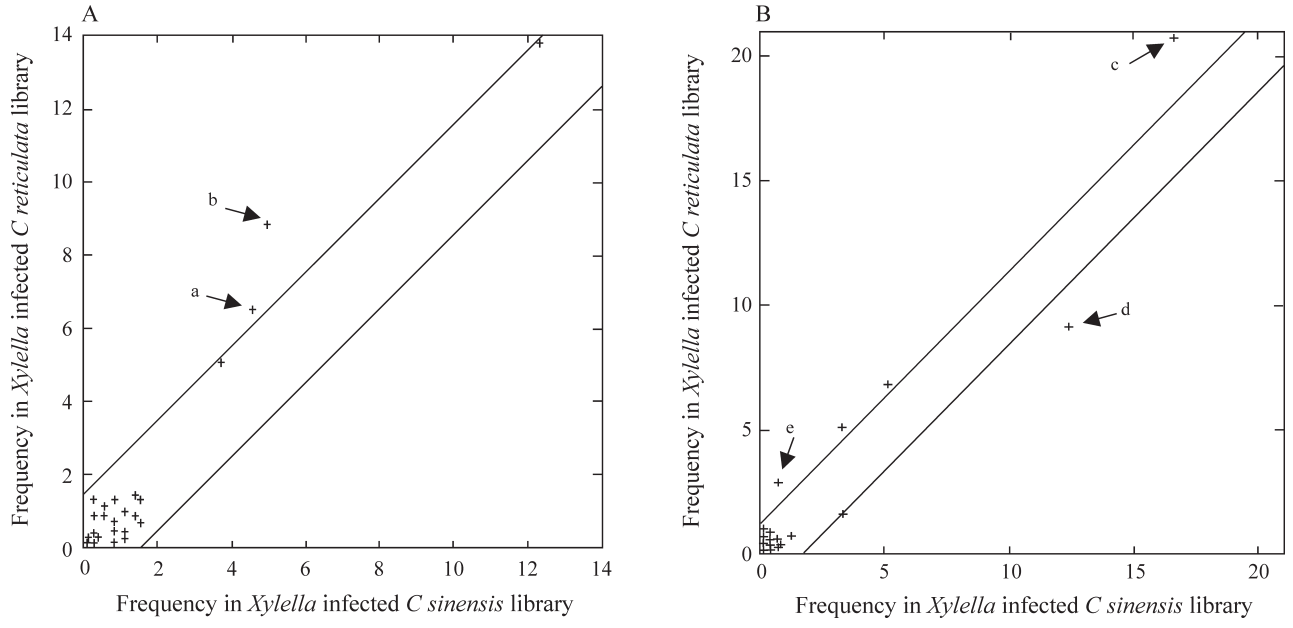

Figure 5 - PlotCorr analyses for comparison between *C. sinensis* (x-axis) and *C. reticulata* (y-axis) normalized expression for photosystems (A) and carbon fixation (B) related sequences. In (A) dots represent (a) oxygen-evolving complex 25.6 kD protein; (b) chlorophyll a/b-binding protein 151 LHCII type II CAB-151; in (B) (c) ribulose bisphosphate carboxylase/oxygenase activase; (d) ribulose bisphosphate carboxylase small chain; and (e) transketolase. Scales are in reads per thousand per library. Differential expression between libraries were significant only for ribulose bisphosphate carboxylase small chain (d) ($p = 0.0487$).

almost half derived from a genomic library made from leaves, likely reflecting the high copy number of chloroplast DNA in that organ.

The level of matching similarity between *Citrus* sequences and the other species plastid genome (Table 2) corroborated taxonomic classification, with more resemblance between *Citrus* and the other dicots (*Eucalyptus* and *Arabidopsis*), all from the same taxon (subclass Rosids). However, surprisingly, *Citrus* sequences were less similar to the *Arabidopsis* genome, despite the fact that both share closer ancestry between the level of order and subclass (Sapindales and Brassicales, respectively, both Eurosids II or Malvids), while *Eucalyptus* belong to the Myrtales order and is not part of the Eurosids I (Savolainen *et al.*, 2000). *Citrus* plastid sequences annotated as mRNA processing, ribosomal protein or others were more specifically related to *Eucalyptus*, suggesting that common life history traits might be important for similarity between these putative genes (Table 3). Genes putatively coding for plastid NADH-dehydrogenase, and those annotated as 'no hit' or 'hypothetical protein' comprised most of the Angiosperm specific-matches (Table 3), suggesting that these genes are more conserved between *Citrus* and the other dicotyledonous plastomes.

Clustering the plastome-matched genomic and EST reads resulted in a draft of *Citrus* plastid genome, theoretically covering around 77% of the reference chloroplast genome available (Bausher *et al.*, 2006). However, the contigs formed were unevenly distributed along the genome, since no cluster longer than 3 kbp was formed, and more than 90% of the contigs were smaller than 1.5 kbp (Table 4). This is likely due to the fact that large amounts of expressed sequences were used for assembly, with a significant lack of plastome intergenic regions in the libraries.

Transcriptional profiling of nuclear-encoded plastid-targeted proteins from various plastid types (chloroplasts, chromoplasts or proplastids) was achieved for leaf, bark, flower and fruit-expressed genes from *C. sinensis* 'Pera IAC', to presumably identify transcripts with minimum or maximum variation among organs and tissues, and consequently from distinct plastid types (Figure 1A-B). The least variable expressed sequences among sampled organs included putative genes involved in starch and amino acids biosynthesis (Figure 1A). Alpha-1,4-glucan phosphorylase has an essential role in plastid starch formation, and mobilization for sucrose and other polysaccharide precursors (Buchner *et al.*, 1996). RuBisCO-binding protein activity was unexpected in reproductive organs, but was apparently found expressed in flower proplastids and fruit chromoplasts, possibly involved in regulation of RuBisCO oxygenase activity, and the subsequent photorespiratory decarboxylating-like pathway required in reproductive or maturing organs. Enolpyruvylshikimate-3-phosphate synthase (EPSPS) and cystathionine gamma-synthase are vital to essential amino acids biosynthetic routes. An EPSPS gene has already been genetically inserted and expressed in transplastomic tobacco, which displayed increased resistance to glyphosate (Ye *et al.*, 2001). Together with cystathionine gamma-synthase, the first enzyme on the methionine biosynthesis pathway (Hacham *et al.*, 2006), EPSPS expression in several organs suggested a relatively constant metabolic requirement for amino acids in *Citrus*.

Based on normalized frequencies, the most variable expressed sequences among leaf, bark, fruit and flower samples included chlorophyll a/b binding proteins and photosystem I subunits associated genes, generally performing photosynthesis maintenance (Figure 1B). The expression levels of an early light-induced regulatory protein and terpene synthase were also highly variable among organs. The former is associated with phytochrome-mediated light perception, and it has been already correlated with acclimatization to low temperatures in *Poncirus* (Zhang *et al.*, 2005). Terpene synthase catalyzes a key step on sesquiterpene metabolism, and it has been described in *Citrus* due to its importance to typical citric flavor (Sharon-Asa *et al.*, 2003).

The most significantly induced genes in vegetative organs, in comparison to reproductive ones, were related to photosystem subunits, especially in leaves (Figure 1C). A transcript, annotated as plastid hypothetical frame ORF19, previously detected in other species but still without any associated function, also exhibited a vegetative-specific expression pattern, suggesting an apparent photosynthetic role for this gene. Conversely, significant differential expression in reproductive organs was detected for genes usually associated with secondary metabolic pathways (Figure 1D): transcripts encoding an inorganic pyrophosphatase, a vacuolar enzyme associated with fruit acidity, sugar accumulation and ripening (Marsh *et al.*, 2001); terpene synthase, a key enzyme for citric taste and flavor; thiazole biosynthesis enzyme, part of thiamine (B_1 vitamin) metabolism and fruit ripening (Jacob-Wilk *et al.*, 1997); and GcpE protein, involved in isoprenoid biosynthesis, known to participate in fruit maturation (Seemann *et al.*, 2006).

All genes with a reproductive organ-specific expression pattern were associated with fruit ripening (Figure 2A). A fruit-specific preference in expression for lipoxygenase C and isocitrate dehydrogenase was detected in the survey for organ-specific sequences. Both enzymes corroborate a fruit-specific function, since lipoxygenase C is linked with the development of volatile compounds, carotenoids and jasmonate responses in fruit (Rangel *et al.*, 2002), and plastid isocitrate dehydrogenase activity has been demonstrated to be associated with inorganic pyrophosphatase and accumulation of organic acids in fruit (Etienne *et al.*, 2002).

A translocon component from protein importing complexes from outer membranes of chloroplast and chromoplast (Summer and Cline, 1999) was identified as presenting flower-specific expression (Figure 2B), suggesting a potential extended activity to leucoplasts. The same was

observed for sequences associated to anthranilate synthase and alpha-glucan water dikinase, respectively involved in indol-alkaloid and terpenoid biosynthesis (Hong *et al.*, 2006) and in starch degradation (Baunsgaard *et al.*, 2005). The presence of these sequences suggested an intense leucoplastidic activity in *Citrus* flowers, especially concerning secondary products synthesis and reserve mobilization in non-photosynthetic tissues.

Clustering expression-patterns visually disclosed quantitative leaf transcriptional variation between *C. sinensis* (sweet orange) and *C. reticulata* (tangerine), whether infected or not with *Xylella fastidiosa* (Figure 3). Regarding the photosystems, it was observed that tangerine leaves, 30-days after inoculation, displayed an apparent accumulation of mRNAs related to ATP synthase subunits; chlorophyll a/b-binding proteins; ferredoxin NADP-reductase; and M-type thioredoxin, while a decrease in ATP synthase transcripts was detected in sweet orange. Maintenance of photosynthesis, suggested by differences in amounts of associated transcripts, might be a consequence of the lack of pathogen in the resistant species. Alternatively, these changes might be due to a metabolic distinction between CVC tolerant and susceptible species, with photosynthesis maintenance despite an occasional effect from toxins and/or catabolites released by *Xylella,* or from xylem clogging. Several genes associated with primary carbon fixation pathway in plastids also exhibited contrasting expression profiles (Figure 4). Transcripts encoding key enzymes responsible for a proper photosynthesis carbon assimilation, and subsequent synthesis of trioses and hexoses, transport or synthesis of more complex carbohydrates, as well as coding for regulatory proteins of these enzymes, appear to accumulate in tangerine infected leaves in comparison to an observed relative decrease in sweet orange leaves under the same conditions. Visually, chlorophyll a/b binding protein 151 type II, oxygen-evolving complex 25.6 kD protein, ribulose 1,5-bisphosphate carboxylase activase and transketolase appeared to be the most induced on infected *C. reticulata* leaves, while a ribulose 1,5-bisphosphate carboxylase small chain subunit transcript was abundant in *C. sinensis*.

The clarification of the role of the photosynthesis and carbon assimilation sequences on *Xylella fastidiosa* resistance in *Citrus* may open novel research opportunities. But, clearly further investigation is required to validate these findings and to distinguish between cause and effect of the detected differences in photosynthesis-related transcript accumulation in tangerine. In addition, plastome sequences may be helpful for developing plastid genetic engineering strategies in *Citrus*, promising for yield increase by photoassimilate enhancement or in fruit nutraceutical enrichment or development through secondary chromoplast metabolite manipulation.

Acknowledgements

The authors would like to thank the Citrus Biotechnology Laboratory, from the Centro de Citricultura “Sylvio Moreira” of the Instituto Agronômico de Campinas, for providing access to the CitEST database and necessary bioinformatic assistance. Support from CNPq and FAPESP was greatly appreciated.

References

Audic S and Claverie JM (1997) The significance of digital gene expression profiles. Genome Res7:986-995.

Baunsgaard L, Lutken H, Mikkelsen R, Glaring MA, Pham TT and Blennow A (2005) A novel isoform of glucan, water dikinase phosphorylates pre-phosphorylated alpha-glucans and is involved in starch degradation in *Arabidopsis*. Plant J 41:595-605.

Bausher MG, Mozoruk J, Lee S-B, Singh N, Jansen RK and Daniell H (2006) The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var ‘Ridge Pineapple’: Organization and phylogenetic relationships to other angiosperms. BMC Plant Biology 6:21.

Bendich AJ (2004) Circular chloroplast chromosomes: The grand illusion. Plant Cell 16:1661-1666.

Bock R and Khan MS (2004) Taming plastids for a green future. Trends Biotechnol 22:311-318.

Buchner P, Borisjuk L and Wobus U (1996) Glucan phosphorylases in *Vicia faba* L.: Cloning, structural analysis and expression patterns of cytosolic and plastidic forms in relation to starch. Planta 199:64-73.

Calsa Jr T, Carraro DM, Benatti MR, Barbosa AC, Kitajima JP and Carrer H (2004) Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. Curr Genet 46:366-373.

Calsa Jr T. and Figueira A (2007) Serial analysis of gene expression in sugarcane (*Saccharum* spp.) leaves revealed alternative C_4 metabolism and putative antisense transcripts. Plant Mol Biol 63:745-762.

Etienne C, Moing A, Dirlewanger E, Raymond P, Monet R and Rothan C (2002) Isolation and characterization of six peach cDNAs encoding key proteins in organic acid metabolism and solute accumulation: Involvement in regulating peach fruit acidity. Physiol Plant 114:259-270.

Guo WW, Prasad D, Cheng YJ, Serrano P, Deng XX and Grosser JW (2004) Targeted cybridization in citrus: Transfer of *Satsuma* cytoplasm to seedy cultivars for potential seedlessness. Plant Cell Rep 22:752-758.

Hacham Y, Schuster G and Amir R (2006) An *in vivo* internal deletion in the N-terminus region of *Arabidopsis* cystathionine gamma-synthase results in CGS expression that is insensitive to methionine. Plant J 45:955-967.

Hayes R, Kudla J and Gruissem W (1999) Degrading chloroplast mRNA: The role of polyadenylation. Trends Biochem Sci 24:199-202.

Herrero J, Al-Shahrour F, Diaz-Uriarte R, Mateos A, Vaquerizas JM, Santoyo J and Dopazo J (2003) GEPAS: A web-based resource for microarray gene expression data analysis. Nucleic Acids Res 31:3461-3467.

Herrero J, Vaquerizas JM, Al-Shahrour F, Conde L, Mateos A, Diaz-Uriarte JS and Dopazo J (2004) New challenges in

gene expression data analysis and the extended GEPAS. Nucleic Acids Res 32:W485-491.

Hong SB, Peebles CA, Shanks JV, San KY and Gibson SI (2006) Expression of the *Arabidopsis* feedback-insensitive anthranilate synthase holoenzyme and tryptophan decarboxylase genes in *Catharanthus roseus* hairy roots. J Biotechnol 122:28-38.

Jacob-Wilk D, Goldschmidt EE, Riov J, Sadka A and Holland D (1997) Induction of a *Citrus* gene highly homologous to plant and yeast thi genes involved in thiamine biosynthesis during natural and ethylene-induced fruit maturation. Plant Mol Biol 35:661-666.

Lilly JW, Havey MJ, Jackson SA and Jiang J (2001) Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. Plant Cell 13:245-254.

Maliga P (2003) Progress towards commercialization of plastid transformation technology. Trends Biotechnol 21:20-28.

Marsh K, Gonzalez P and Echeverria E (2001) Partial characterization of H-translocating inorganic pyrophosphatase from 3 citrus varieties differing in vacuolar pH. Physiol Plant 111:519-526.

Meyers BC, Vu TH, Tej SS, Ghazal H, Matvienko M, Agrawal V, Ning J and Haudenschild CD (2004) Analysis of the transcriptional complexity of *Arabidopsis thaliana* by massively parallel signature sequencing. Nat Biotechnol 22:1006-1011.

Moreira CD, Gmitter FG Jr, Grosser JW, Huang S, Ortega VM and Chase CD (2002) Inheritance of organelle DNA sequences in a citrus-poncirus intergeneric cross. J Hered 93:174-178.

Mullineaux PM, Karpinski S, Jimenez A, Cleary SP, Robinson C and Creissen GP (1998) Identification of cDNAS encoding plastid-targeted glutathione peroxidase. Plant J 13:375-379.

Ogihara Y, Isono K, Kojima T, Endo A, Hanaoka M, Shiina T, Terachi T, Utsugi S, Murata M, Mori N, *et al.* (2002) Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. Mol Genet Genomics 266:740-746.

Pyke KA (1999) Plastid division and development. Plant Cell 11:549-556.

Rangel M, Machado OL, da Cunha M and Jacinto T (2002) Accumulation of chloroplast-targeted lipoxygenase in passion fruit leaves in response to methyl jasmonate. Phytochemistry 60:619-625.

Robinson SJ, Cram DJ, Lewis CT and Parkin IA (2004) Maximizing the efficacy of SAGE analysis identifies novel transcripts in *Arabidopsis*. Plant Physiol 136:3223-3233.

Rott R, Liveanu V, Drager RG, Stern DB and Schuster G (1998) The sequence and structure of the 3'-untranslated regions of chloroplast transcripts are important determinants of mRNA accumulation and stability. Plant Mol Biol 36:307-314.

Ruf S, Hermann M, Berger IJ, Carrer H and Bock R (2001) Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. Nat Biotechnol 19:870-875.

Savolainen V, Fay MF, Albach DC, Backlund A, van der Bank M, Cameron KM, Johnson SA, Lledó MD, Pintaud JC, Powell M *et al.* (2000) Phylogeny of the eudicots: A nearly complete familial analysis based on *rbcl* gene sequences. Kew Bulletin 55:257-309.

Seemann M, Tse-Sum B, Wolff M, Miginiac-Maslow M and Rohmer M (2006) Isoprenoid biosynthesis in plant chloroplasts via the MEP pathway: Direct thylakoid/ferredoxin-dependent photoreduction of GcpE/IspG. FEBS Lett 580:1547-1552.

Sharon-Asa L, Shalit M, Frydman A, Bar E, Holland D, Or E, Lavi U, Lewinsohn E and Eyal Y (2003) *Citrus* fruit flavor and aroma biosynthesis: Isolation, functional characterization, and developmental regulation of *Cstps1*, a key gene in the production of the sesquiterpene aroma compound valencene. Plant J 36:664-674.

Simpson AJ, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LM, Araya JE, Baia GS, Baptista CS *et al.* (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* consortium of the organization for nucleotide sequencing and analysis. Nature 406:151-157.

Steane DA (2005) Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). DNA Res 12:215-220.

Stegemann S, Hartmann S, Ruf S and Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. Proc Natl Acad Sci USA 100:8828-8833.

Sugiura M (2003) History of chloroplast genomics. Photosynth Res 76:371-377.

Summer EJ and Cline K (1999) Red bell pepper chromoplasts exhibit *in vitro* import competency and membrane targeting of passenger proteins from the thylakoidal sec and DeltapH pathways but not the chloroplast signal recognition particle pathway. Plant Physiol 119:575-584.

Taiz L and Zeiger E (1998) Plant Physiology. 2nd edition. Sinauer Associates, Massachussets, 792 pp.

Takami K, Matsumara A, Yahata M, Imayama T, Kunitake H and Komatsu H (2004) Production of intergeneric somatic hybrids between round kumquat (*Fortunella japonica* Swingle) and 'Morita navel' orange (*Citrus sinensis* Osbeck). Plant Cell Rep 23:39-45.

Tao NG, Xu J, Cheng YJ and Deng XX (2005) Lycopene-epsilon-cyclase pre-mRNA is alternatively spliced in Cara Cara navel orange (*Citrus sinensis* Osbeck). Biotechnol Lett 27:779-782.

Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T and Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci USA 91:9794-9788.

Ye GN, Hajdukiewicz PT, Broyles D, Rodriguez D, Xu CW, Nehra N and Staub JM (2001) Plastid-expressed 5-enolpyruvylshikimate-3-phosphate synthase genes provide high level glyphosate tolerance in tobacco. Plant J 25:261-270.

Zhang CK, Lang P, Dane F, Ebel RC, Singh NK, Locy RD and Dozier WA (2005) Cold acclimation induced genes of trifoliate orange (*Poncirus trifoliata*). Plant Cell Rep 23:764-769.

Internet Resource

CitEST Database: http://biotecnologia.centrodecitricultura.br.
GEPAS: http://gepas.bioinfo.cipf.es.

Associate Editor: Alessandra Alves de Souza