Genetics and Molecular Biology, 30, 3 (suppl), 832-840 (2007)

www.sbg.org.br

Research Article

A genomic approach to characterization of the *Citrus* terpene synthase gene family

Marcelo Carnier Dornelas and Paulo Mazzafera

Departamento de Fisiologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.

Abstract

Terpenes are a very large and structurally diverse group of secondary metabolites which are abundant in many essential oils, resins and floral scents. Additionally, some terpenes have roles as phytoalexins in plant-pathogen relationships, allelopathic inhibitors in plant-plant interactions, or as airborne molecules of plant-herbivore multitrophic signaling. Thus the elucidation of the biochemistry and molecular genetics of terpenoid biosynthesis has paramount importance in any crop species. With this aim, we searched the CitEST database for clusters of expressed sequence tags (ESTs) coding for terpene synthases. Herein is a report on the identification and *in silico* characterization of 49 putative members of the terpene synthase family in diverse *Citrus* species. The expression patterns and the possible physiological roles of the identified sequences are also discussed.

Key words: secondary metabolism, flavor, aroma, plant protection, terpene synthesis.

Received: July 21, 2006; Accepted: April 2, 2007.

Introduction

Terpenes are found widely distributed in the plant kingdom, from lichens and algae to higher plants. Although terpenes cover a wide range of compounds with diverse structure, the word terpene has been frequently associated with essential oils, which are volatile compounds belonging to some terpene classes. Terpenes have several applications in the food and cosmetic industry in the production of flavors and fragrances (Croteau *et al.*, 2000; Phillips *et al.*, 2006; Tholl, 2006).

Terpenes are classified according to the following classes: hemiterpenes (C_5H_8); monoterpenes ($C_{10}H_{16}$); sesquiterpenes ($C_{15}H_{24}$); diterpenes ($C_{20}H_{32}$); sesterterpenes ($C_{25}H_{40}$); triterpenes ($C_{30}H_{48}$); tetraterpenes ($C_{40}H_{64}$) and polyterpenes [$(C_5H_8)_n$]. Hemiterpenes are not found in appreciable amounts as free compounds but are usually bound to other non-terpene compounds, such as the purine base adesonine and the plant hormone zeatin (Goodwin, 1967; Croteau *et al.*, 2000; Phillips *et al.*, 2006; Tholl, 2006). Monoterpenes consist of two condensed basic units of isopentenyl pyrophosphate (IPP). Accordingly, sesquiterpenes have three, diterpenes four, sesterterpenes five, triterpenes six and tetraterpenes eight IPP molecules, respectively. Polyterpenes are all terpenes containing more than eight isoprene units, which include all natural rubbers (Goodwin, 1967; Croteau *et al.*, 2000).

Elucidation of the biochemistry and molecular genetics of terpenoid biosynthesis has made rapid progress in recent years (Rohdich *et al.*, 2005). The genes coding for the main enzymes involved in this biological process are being identified, and all the members of the *Arabidopsis* terpene synthase gene family have been characterized following the sequencing of the whole genome of this model plant (Aubourg *et al.*, 2002). Figure 1 shows the general scheme for the synthesis of the precursor molecules of the main terpene classes. The starting reaction is the isomerization of IPP to dimethylallyl pyrophosphate. Both compounds condense to form the first parental structure, geranyl-PP. Subsequently, the other structures are formed by sequential addition of IPP (Phillips *et al.*, 2006; Tholl, 2006). Two farnesyl-PP are condensed to form the parent molecule squalene, the precursor of triterpenes, and two geranyl-PP are condensed to form phytoene, the parent molecule for tetraterpenes (Goodwin, 1967; Croteau *et al.*, 2000; Rohdich *et al.*, 2005; Phillips *et al.*, 2006; Tholl, 2006).

The biological functions of the many terpene molecules in plants are linked not only to the biosynthesis of hormones, but also to protection against UV radiation and photo-oxidative stress. Additionally, terpenes are also related to thermal protection, pollinator attraction, membrane stabilization, resistance against insects and microorgan-

Send correspondence to Marcelo C. Dornelas. Departamento de Fisiologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Cidade Universitária Zeferino Vaz, Caixa Postal 6109, 13.083-970 Campinas, SP, Brazil. E-Mail: dornelas@unicamp.br.

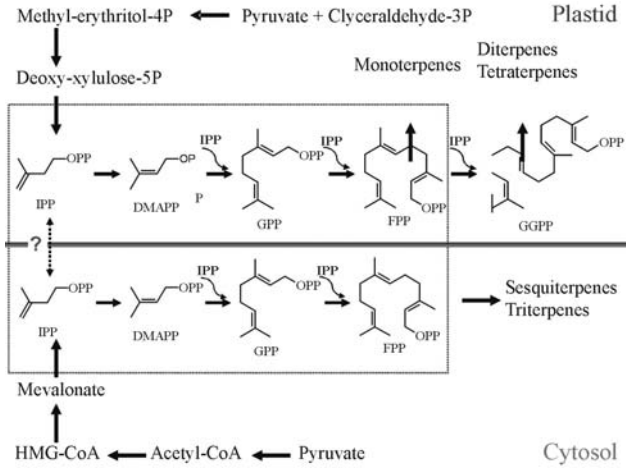

Figure 1 - General scheme of the biosynthesis of different terpene classes. The precursor IPP is biosynthesized in two different pathways, in plastids and cytosol, which involve similar reactions, but performed by specific enzymes in different compartments (dashed lines). In the cytosol, IPP will be the precursor of sequiterpenes and triterpenes, while monoterpenes, diterpenes and tetraterpenes are formed in plastids. HMG-CoA = hydroxymethyl glutaryl-CoA; IPP = isopentenyl pyrophosphate; DMAPP = dimethylallyl pyrophosphate; GPP = geranyl pyrophosphate; FPP = farnesyl pyrophosphate; GGPP = geranyl-geranyl pyrophosphate; GFPP = geranyl-farnesyl pyrophosphate.

isms, plant-plant signaling, etc. (Steele *et al.*, 1998; Trapp and Croteau, 2001; Copolovici *et al.*, 2005; Baldwin *et al.*, 2006; Keeling and Bohlmann, 2006).

In *Citrus* spp, terpene molecules belonging to different classes are produced especially in leaves, fruit epidermis (flavedo) and fruit juice. These terpenes have special economic interest, as they are the main components of *Citrus* essential oils and some of them (carotenoids) give the *Citrus* juice its special color. Additionally, carotenoids are well known to be important to human health. Several biotechnological approaches have been taken to increase such compounds in food (Botella-Pavia and Rodriguez-Concepcion, 2006). There are several reports on the composition of terpenes in several *Citrus* species, mainly regarding the essential oil composition (Ruberto and Rapisarda, 2002; Sawamura *et al.*, 2005; Verzera *et al.*, 2005). The aromatic components of citrus are classified in two categories: those present in the oil from the flavedo and juice, and those soluble in the water and components of the juice. Monoterpene d-limonene is the main component of oil from the flavedo, with concentrations over 85% of the oil fraction. In the Pêra variety of orange, the concentration may reach up to 93% while in the Tahiti variety of lime it ranges from 50 to 60%. In addition to d-limonene, other terpenes found in the flavedo oil fraction are linalool, geraniol, citronellol, α-terpineol, valencene, mircene, α-pinene, etc. (Ruberto and Rapisarda, 2002; Sawamura *et al.*, 2005; Verzera *et al.*, 2005).

Thus, due to the importance of *Citrus* terpenes to the Brazilian economy (Boteon and Neves, 2005), we undertook a genomic approach for the characterization of the *Citrus* terpene synthase gene family by analyzing the CitEST database of *Citrus* expressed sequence tags. We have identified 49 putative members of the terpene synthase family in diverse *Citrus* species and we suggest their possible biological roles based on their expression patterns and on sequence comparisons with other terpene synthases that have already been functionally characterized in other plant species.

Material and Methods

Searching *Citrus* EST homologs for terpene synthases

The clustered expressed sequence tags (ESTs) from the CitEST project database were used as a primary source of data for our analyses. These sequences were assembled from ESTs obtained from the sequencing of several *Citrus* spp. cDNA libraries, made from different tissues and various physiological states (see other papers in this issue for details on library construction and sequencing). Nucleotide and amino acid sequences from other terpene synthase genes were obtained from The National Center for Biotechnology Information (NCBI). Searches for terpene synthase sequences in the CitEST database were conducted using the tBLASTN module that compares the consensus amino acid sequence with a translated nucleotide sequences database (Altschul *et al.*, 1997). We used as a query a consensus terpene synthase sequence obtained by aligning all *Arabidopsis thaliana* terpene synthase protein sequences (Aubourg *et al.*, 2002). All sequences that exhibit a significant alignment (e-value lower than 10^{-5}) with the consensus were retrieved from the CitEST database. All retrieved sequences were then re-inspected for occurrence of terpene synthase conserved motives using the InterProScan.

Phylogenetic analysis and expression patterns of putative *Citrus* terpene synthases

Amino acid sequences were used for all the phylogenetic analyses. Sequence alignments were performed With ClustalX (Thompson *et al.*, 1994) using default parameters, but the final alignment was visually inspected and manually corrected. The MEGA software, version 2.0 (Kumar *et al.*, 2000) was used for the phylogenetic analysis. Average p-distances were high so the Poisson model was used to provide unbiased estimates of the number of substitutions between sequences. Phylogenetic trees were obtained using parsimony and/or genetic distance calculations. Neighbor-joining (Saitou and Nei, 1987) and Bootstrap (with 1000 replicates) trees were also constructed.

For each EST-contig, the frequency of reads in the selected libraries was calculated. This procedure requires a normalization that is accomplished by dividing the total number of reads in the specific library by the total number of reads in all libraries and then dividing the number of

reads of each EST-contig by the ratio found for each library. The results were cast in a matrix and a hierarchical clustering was performed, using the Cluster and Tree View programs (Eisen *et al.*, 1998). The pattern of gene expression was displayed as color-coded arrays of EST-contigs, using a color scale representing the number of reads from a specific library in each EST-contig.

Results

Identifying *Citrus* spp putative terpene synthase genes

Despite the fact that the whole genome of the model plant *Arabidopsis thaliana* had been completely sequenced four years ago, the function of only three, out of the 32 *Arabidopsis* terpene synthase genes (Aubourg *et al.*, 2002), has been described to date. These are the *AtTPSGA1* gene for copalyl diphosphate synthase (Sun and Kamiya, 1994) and the *AtTPSGA2* transcript which encodes kaurene synthase (Yamaguchi *et al.*, 1998), both of which are involved in the formation of gibberellins, and the *AtTPS10* transcript for myrcene/ocimene synthase (Bohlmann *et al.*, 2000) required for the formation of acyclic monoterpenes. In an initial attempt to identify the members of the *Citrus* terpene synthase gene family, we have performed an *in silico* screen of the CitEST database for putative terpene synthase sequences. This exhaustive sequence search detected 49 unique assembled *Citrus* spp EST sequences.

We started our analysis by detecting conserved sequence motifs within the putative *Citrus* terpene synthase proteins. In pairwise comparisons of all predicted *Citrus* terpene synthases to all *Arabidopsis* AtTPS proteins, overall sequence identity varies widely from 18% (LT33-C1-003-056-E04 and AtTPS21) to 91% (CS00-C1-100-123-A05 and AtTPGA1). These sequence comparisons allowed the construction of the distance-based tree shown in Figure 2. For this analysis, we considered all *Citrus* spp EST clusters found within the CitEST database (we used the name of the founder EST sequence to name the entire sequence cluster), all *Arabidopsis* TPS and all publicly available *Citrus* terpene synthase sequences available at GenBank (as of May 2006). We adopted the separation of the terpene synthases into classes, as suggested by Aubourg *et al.* (2002).

When the putative *Citrus* terpene synthase EST contigs were long enough to allow the complete encoded protein sequences to be deduced, their size ranged from 547 to 617 amino acids, which corresponds to the size of known monoterpene synthases, sesquiterpene synthases and diterpene synthases of secondary metabolism (Bohlmann *et al.*, 1998; Aubourg *et al.*, 2002). Variation in length within this class of terpene synthase could be attributed to the presence or absence of putative plastid transit peptides (Figure 1).

Most terpene synthases encoded by class-III genes contain variations of a conserved motif, RR(x)8 W, close to

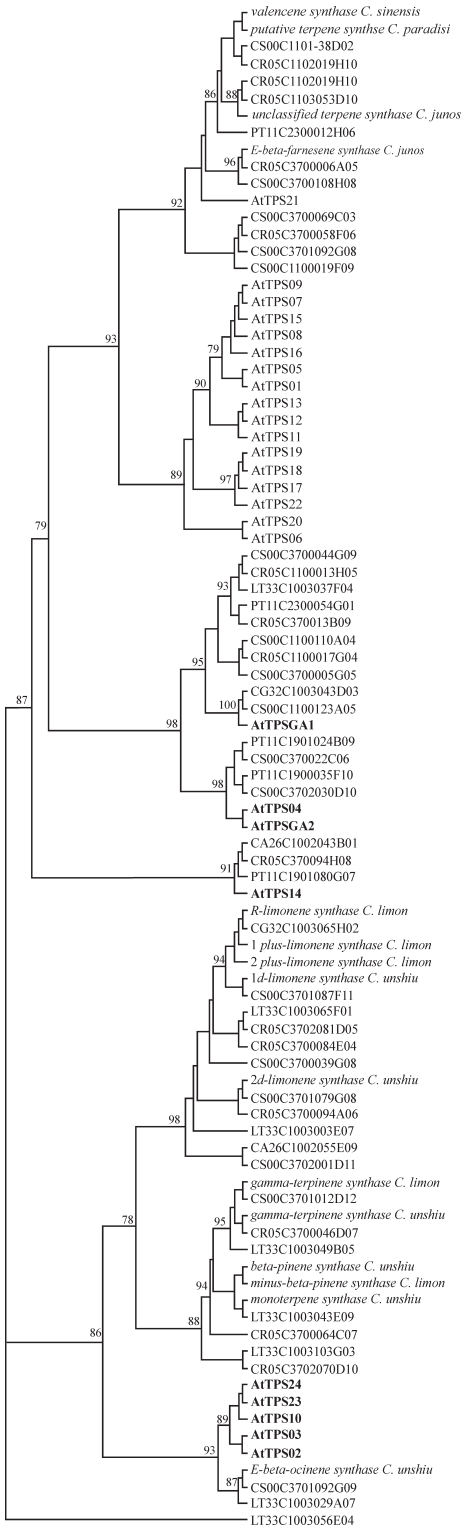

Figure 2 - Phylogenetic tree of the *Citrus* putative terpene synthase family members, the *Arabidopsis* terpene synthase family (AtPS) and representative *Citrus* terpene synthases of known function. The neighbor-joining tree was generated from an alignment of amino acid sequences. Nodes supported by bootstrap values higher than 75% are shown. For designation of the AtTPS genes see Aubourg *et al.* (2002). For characterized *Citrus* sequences see Sharon-Asa *et al.* (2003) and Shimada *et al.* (2004; 2005). CA: *Citrus aurantium*; CG: *C. aurantifolia*; CR: *C. reticulata*; CS: *C. sinensis*, LT: *C. latifolia*; PT: *Poncirus trifoliata*. Shadowed sequences contain predicted signaling peptides directing them to the plastid.

the N-terminus (Bohlmann *et al.*, 1998; Aubourg *et al.*, 2002). The RR(x)8 W motif is absolutely conserved in most *Citrus* sequences that resemble typical monoterpene synthases. There was a clade within this group that contained only *Arabidopsis*-derived sequences (AtTPS02, AtTPS 03, AtTPS 10, AtTPS 23 and AtTPS 24; see Figures 2 and 3) and it has been reported that these proteins contain variations on the RR(x)8 W motif; however, these variations have as yet unknown biological implications (Aubourg *et al.*, 2002).

Attributing putative functions to *Citrus* putative terpene synthases by sequence comparisons

Comparison of the predicted *Citrus* terpene synthase proteins with homologs of known function from other species allowed the identification of putative orthologues that evolved from a common ancestral gene by speciation and retained the same or similar biological function in different species during the course of evolution (Tatusov *et al.*, 1997; Huynen and Bork, 1998). The deduced *Citrus* terpene syn-

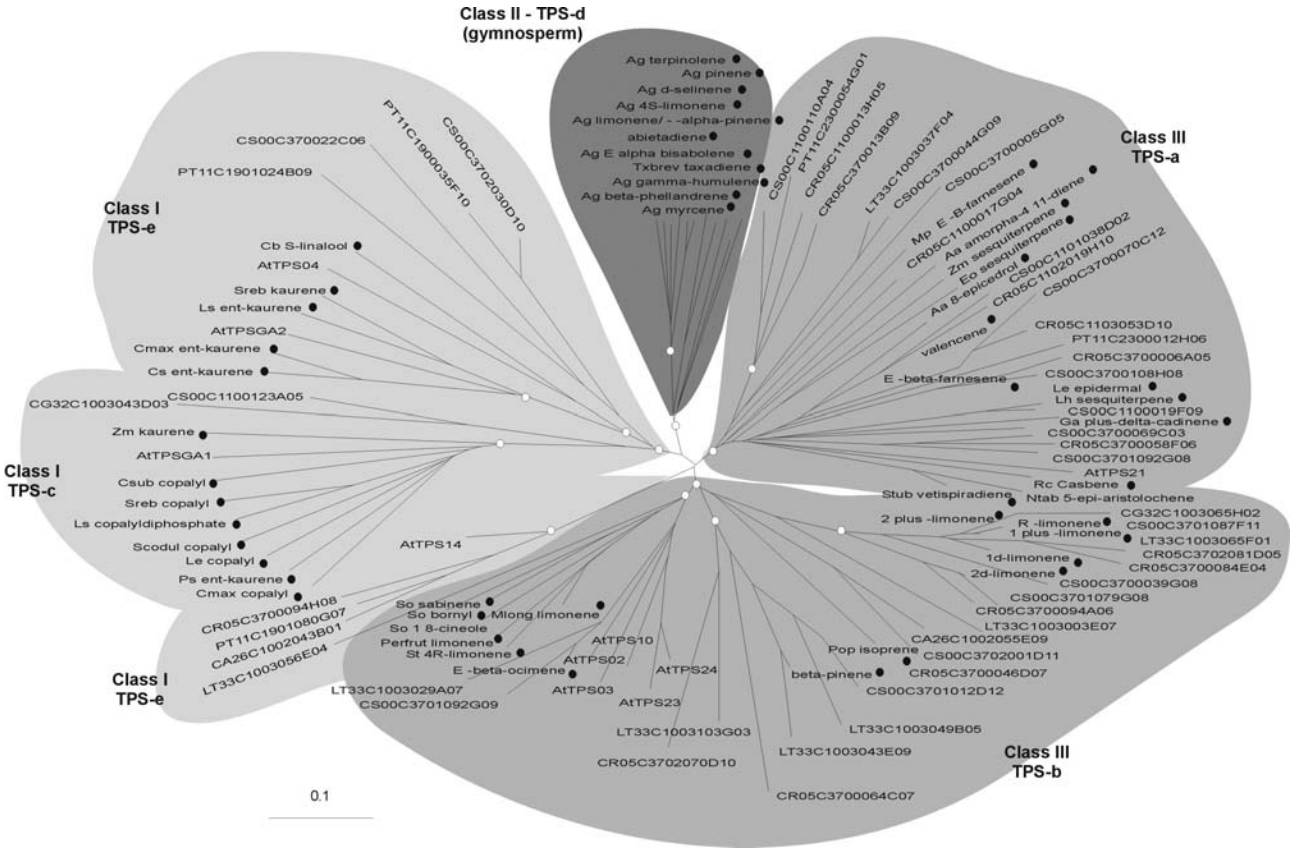

Figure 3 - Phylogenetic tree of the *Citrus* putative terpene synthase family members, the *Arabidopsis* terpene synthase family (AtPS) and representative terpene synthases of known function from diverse plant species. The neighbor-joining tree was generated from an alignment of 16 AtTPS and 43 documented terpene synthases from 25 different plant species. Nodes supported by bootstrap values higher than 700 out of 1000 replicates are marked in yellow. Subfamilies have previously been defined (Bohlmann *et al.*, 1998; Aubourg *et al.*, 2002). For designation of the AtTPS genes see Aubourg *et al.* (2002). CA: *Citrus aurantium*; CG: *C. aurantifolia*; CR: *C. reticulata*; CS: *C. sinensis*; LT: *C. latifolia*; PT: *Poncirus trifoliata*. Accession numbers for other TPS represented in the phylogenetic tree are: *Abies grandis* abietadiene synthase (U50768); *A.grandis* (E)-α-bisabolene synthase (AF006195); *A.grandis*-γ-humulene synthase (U92267); *A.grandis* (-)-4S-limonene synthase (AF006193); *A.grandis* (-)-limonene/(-)-alpha-pinene synthase (AF139207); *A.grandis* myrcene synthase (U87908); *A.grandis* β-phellandrene synthase (AF139205); *A.grandis* (-)-pinene synthase (U87909); A.grandis δ-selinene synthase (U92266); *A.grandis* terpinolene synthase (AF139206); *Artemisia annua* amorpha-4,11-diene synthase (AF138959); *A.annua* epi-cedrol synthase (AF157059); *Clarkia breweri* S-linalool synthase (U58314); *Croton sublyratus* copalyl diphosphate synthase (AB042424); *Cucumus sativus* ent-kaurene synthase (AB045310); *Cucurbita maxima* copalyl diphosphate synthase (AF049905); *C.maxima* ent-kaurene synthase (U43904); *Elaeis oleifera* sesquiterpene synthase (AF080245); Gossypium arboreum (+)-*d*-cadinene synthase XC14 (U23205); *Latuca sativa* copalyl diphosphate synthase (AB031204); *L.sativa* ent-kaurene synthase (AB031205); *Lycopersicon esculentum* copalyl diphosphate synth- ase (AB015675); *L.esculentum* germacrene C synthase (AF035630); *L.hirsutum* germacrene B synthase SSTHL1 (AF279455); *Mentha x piperita* (E)-β- farnesene synthase (AF024615); *Mentha longifolia* (-)-4S-limonene synthase (AF175323); *Nicotiana tabacum* 5-epi-aristolochene synthase (L04680), *Perilla frutescens* limonene synthase (D49368); *Pisum sativum* copalyl diphosphate synthase (U63652); *Populus alba x tremula* isoprene synthase (AJ294819); *Ricinus communis* casbene synthase (L32134); *Scoparia dulcis* copalyl diphosphate synthase (AB046689); *Schizonepeta tenuifolia* (+)-4R-limonene synthase (AF282875); *Salvia officinalis* (+)-bornyl diphosphate synthase (AF051900); *S. officinalis* (+)-sabinene synthase (AF051901); *S. officinalis* 1,8-cineole synthase (AF051899); *Solanum tuberosum* vetispiradiene synthase (AB022598); *Stevia rebaudiana* copalyl diphosphate synthase (AF034545); *S. rebaudiana* kaurene synthase (AF097310); *Taxus brevifolia* taxadiene synthase (U48796); *Zea mays* copalyl diphosphate synthase (L37750); *Z.mays* sesquiterpene synthase (AF296122).

thase proteins were compared with more than 40 terpene synthases from over 20 different species, including monocotyledonous and dicotyledonous species and gymnosperms, to determine sequence identity. A neighbor-joining tree was constructed based on multiple sequence alignment (Figure 3). Six subfamilies of the plant terpene synthase family, designated TPS-a through TPS-f, were previously defined based on clusters identified in the phylogeny (Bohlmann *et al.*, 1997; 1998; Aubourg *et al.*, 2002). Sequence relatedness places all *Citrus* putative terpene synthases into the previously defined angiosperm terpene synthase subfamilies (Figure 3). Most *Citrus* sequences cluster in the class III. The terpene synthases from this group contain all known sesquiterpene and diterpene synthases of the secondary metabolism from angiosperms. It is therefore most likely that the *Citrus* members of this group are also sesquiterpene synthases or diterpene synthases, rather than monoterpene synthases. The lack of transit peptides in some of the *Citrus* proteins from this group is reminiscent of previously characterized sesquiterpene synthases of the TPS-a group (Bohlmann *et al.*, 1998; Aubourg *et al.*, 2002).

Class I terpene synthases include AtTPS GA1 (Sun and Kamiya, 1994), which is a member of the TPS-c group of angiosperm copalyl diphosphate synthases. This class also includes the AtTPS GA2 enzyme (Yamaguchi *et al.*, 1998), which is a diterpene synthase of the TPS-e subfamily of kaurene synthases. These terpene synthases are involved in the biosynthesis of gibberellic acid and putative *Citrus* ortologues were found for both of them. The Class I sub-clade, which includes the highly divergent AtTPS14, also contains terpene synthases from three different *Citrus* species that share the conserved DDxxD motif. Finally, Class I also includes AtTPS04 of which the primary structure is reminiscent of that of linalool synthase from *Clarkia breweri* (Dudareva *et al.*, 1996).

Expression patterns of *Citrus* putative terpene synthases

To investigate whether the expression patterns of putative *Citrus* terpene synthases were biased towards a certain organ and/or tissue, we performed an *in silico* Northern in order to determine the relative abundance of putative terpene synthase-coding transcripts among different CitEST cDNA libraries. The results, shown in Figure 4, indicate a preferential accumulation of transcripts in leaves, and in the fruit flavedo, especially during the early stages of fruit development. These results are in agreement with observations of terpene accumulation and essential oil production by these tissues in *Citrus* plants (Ruberto and Rapisarda, 2002; Sawamura *et al.*, 2005; Verzera *et al.*, 2005).

Citrus putative terpene synthase sequences were present at extremely low frequencies (a single EST among all CitEST sequences, as in the case of PT11-C1-901-080-G07) or at relatively high proportions (more than one hundred ESTs, in the case of CR05-C3-700-046-D07), indicating that the terpene synthase family members are differentially expressed among *Citrus* species, as well as in different tissues and developmental stages (Figure 4).

Discussion

The duplication followed by divergence of terpene synthase genes are central to the biosynthesis of hundreds of basic terpenoid skeletons derived from only four prenyl-diphosphate intermediates of the isoprenoid pathway (Figure 1; Davis and Croteau, 2000). Evolution of a large terpene synthase family reflects, at the genetic level, the structural diversity of terpenoid natural products and their roles in ecological plant interactions. Earlier phylogenies of plant terpene synthases established characteristic features for this family: clustering of terpene synthases into at least six subfamilies; independent evolution of specific catalytic functions of terpene synthases in gymnosperms and angiosperms; the presence of a 200 amino acid motif in an ancestor of angiosperm and gymnosperm terpene synthases and divergence of terpene synthase genes involved in secondary metabolism as the consequence of gene duplication and functional diversification (Bohlmann *et al.*, 1998; Aubourg *et al.*, 2002).

In the present analysis, a protein phylogeny was combined with novel genomic information generated from mining the data provided by the CitEST database of Citrus expressed sequence tags. Due to the importance of *Citrus* terpenes in aroma, flavor and the juice industries, we set out to characterize the *Citrus* spp terpene synthase gene family. We successfully identified 49 putative *Citrus* terpene synthase coding sequences derived from six different species (Figure 2). In most cases, sequence comparison analyses revealed that some terpene synthases showed higher similarity among different *Citrus* species, indicating a possible conservation of their biological roles within these species. This observation also suggests that the different classes of terpene synthases diverged before *Citrus* speciation events and that the last ancestor of all *Citrus* species analyzed possessed at least one member of each known terpene synthases classes. For instance, CS00-C1-100-123-A05 showed higher similarity to CG32-C1-003-043-D03 from *C. aurantifolia* than to any other terpene synthase from *C. sinensis*.

The position of three conifer terpene synthase sequences within the present protein phylogeny (Figure 3) is in agreement with previous analyses of terpene synthase family evolution (Bohlmann *et al.*, 1998; Aubourg *et al.*, 2002). Nevertheless, more terpene synthase sequences from gymnosperms and lower nonvascular plants are required for a better understanding of their phylogenetic location relative to the large number of known angiosperm tepene synthases. On the other hand, the placement of all *Citrus* spp putative terpene synthases in the protein phylog-

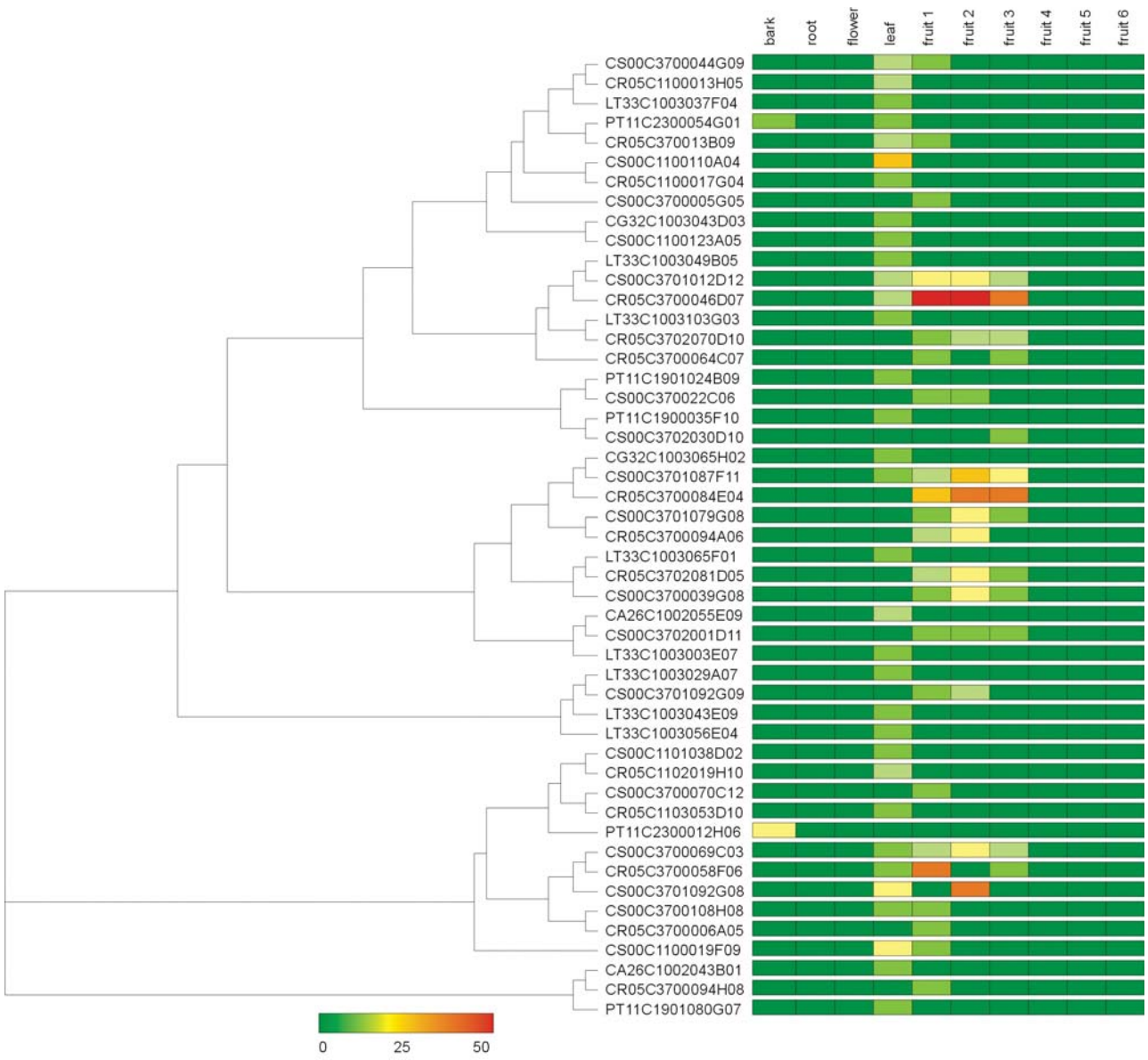

Figure 4 - Expression profiles of the 49 putative *Citrus* terpene synthase EST clusters in selected cDNA libraries from the CitEST database. Data represent the relative number of reads from a specific library in each EST cluster after normalization. Each EST cluster is represented by a single row, and each library is represented by a simple column. The cladogram on the left represents the relatedness of all *Citrus* sequences and was built according to their relative genetic distances (Saitou and Nei, 1987). For consistency of each clade, use other figures for comparison.

eny is well supported and might be indicative of their potential biological roles (Figure 3). For instance, the Class III sub-clade containing well-characterized limonene synthases also includes many clusters from five different *Citrus* species, which might indicate that these clusters code for putative limonene synthase orthologs from these species.

Sequence comparisons of all *Citrus* putative terpene synthases found in the CitEST database with terpene synthases from other plant species allowed the prediction of their putative preferential substrates and thus, suggest their potential biological roles (Figure 3). Among the *Citrus* putative terpene synthase coding sequences that we have found, most of them were preferentially expressed in the leaves and in the fruit flavedo, in agreement with their expected putative roles in terpene biosynthesis (Figure 4). Apart from being an important constituent of *Citrus* essential oils, some reports have described the activation of terpene synthase gene expression in defense responses against insects and pathogens. For example, in cotton, sesquiterpene phytoalexins are elicited in response to bacterial or fungal infection. Chen *et al.* (1995) observed that *Gossypium arboreum* cell suspension culture showed an increase of transcripts of a (+)-delta-cadinene synthase when challenged by a preparation from *Verticillium dahliae*. The authors concluded that such observation was consistent

with a role for this enzyme as the first step in the pathways leading to the biosynthesis of phytoalexins gossypol and lacinilene C in cotton. Accordingly, tent caterpillars feeding on leaves of hybrid poplar induced local and systemic emissions of (-)-germacrene D, (E)-β-ocimene, linalool, (E)-4,8-dimethyl-1,3,7-nonatriene, benzene cyanide, and (E,E)-α-farnesene (Arimura *et al.*, 2004b). This emission of volatile terpene compounds was correlated with an increase in transcription levels of a sesquiterpene synthase greatly induced in response to herbivory (Arimura *et al.*, 2004a). We found putative *Citrus* homologs to terpene synthases expressed preferentially in leaves (Figure 4). However, it remains speculative whether their biological function is related to the plant signaling system induced by pest and/or pathogen attack.

Transcripts of a valencene synthase were found to accumulate in fruits of *C. sinensis* only towards maturation, contributing to the accumulation of valencene (Sharon-Asa *et al.*, 2003). Curiously, *Citrus* fruits are non-climacteric but valencene and its synthase were induced by ethylene, indicating that this hormone may play a role at the final stages of *Citrus* fruit maturation.

Within the CitEST frame, six libraries were made from the flavedo tissue of fruits at different developmental stages, ranging from 1 to 9 cm diameter. Only two *C. reticulata* EST clusters, CR05-C3-700-084-E04 and CR05-C3-700-046-D07), showed a significant increase in EST abundance correlated with developmental progress up to the third developmental stage, fruits of 5 cm in diameter (Figure 4). The putative orthologs of these sequences in other *Citrus* species analyzed did not show this behavior, indicating that this up-regulation during fruit development may be a characteristic of *C. reticulata* fruit.

Cluster CS00-C3-700-022-C06 showed very high similarity to cycloartenol synthase, which converts oxidosqualene to cycloartenol, a pentacyclic isomer of the animal and fungal sterol precursor lanosterol. Plants cyclize oxidosqualene to cycloartenol as the initial sterol (Corey *et al.*, 1993). Other clusters/genes closely related within this same clade (Figure 3) are also apparently involved with steroid biosynthesis. Steroids are structural components of membranes and are very important for membrane fluidity. Brassinosteroids, a class of hormones, are also exclusively formed by the terpene metabolism (Croteau *et al.*, 2000). Thus it would be interesting to test whether the protein predicted to be coded by cluster CS00-C3-700-022-C06 is indeed involved with steroid biosynthesis in *Citrus*.

The cluster CA26-C1-002-055-E09 showed high similarity to an isoprene synthase from *Pueraria montana* (Sharkey *et al.*, 2005). Isoprene is formed from dimethylallyl-PP and its emission from leaves can significantly influence the surrounding atmosphere. Claeys *et al.* (2004) showed that photo-oxidation of isoprene emitted from plants in the Amazon was sufficient to influence the rain regime in the region. Additionally, isoprene emission has been reported to protect plants against high temperatures particularly during rapid temperature fluctuation periods (Velikova and Loreto, 2005). Future work on the putative substrate of the enzyme coded by cluster CA26-C1-002-055-E09 might help with the elucidation of its biochemical function.

We found many *Citrus* spp clusters showing high levels of similarity with limonene synthases and γ-terpinene synthases (Figure 3). We speculate that these transcripts might represent *Citrus* orthologs for limonene synthases, as limonene is the most abundant terpene in *Citrus* spp essential oils. Figure 4 shows that contigs CR05C3700084E04 and CR05C3700046D07, which showed high similarity to both limonene and γ-terpinene synthases, are highly expressed in the flavedo, especially during the first three stages of fruit development. This was also observed for d-limonene synthase (Shimada *et al.*, 2005) and γ-terpinene synthase (Shimada *et al.*, 2004) genes isolated from *C. unshiu.* The transcripts of these genes were reported to accumulate in the fruit peel at the early stages of fruit development (Shimada *et al.*, 2004; 2005).

It was surprising to discover that some putative *Citrus* terpene synthases such as those coded by *Poncirus* sequences PT11-C2-300-054-G01 and PT11-C2-300-012-H06 were apparently exclusively expressed in bark tissue (Figure 4). Nevertheless, it has been recently suggested that expression of limonene synthase may be related with induced oleoresinosis response against the white pine weevil attacking Sitka spruce trees (Byun-McKay *et al.*, 2006). Oleoresin is a mixture of turpentine (85% monoterpenes and 15% sesquiterpenes) and rosin (diterpene resin acids) that seal wounds and is toxic to both invading insects and their pathogenic fungal symbionts (Steele *et al.*, 1998). Of course, this attempt to attribute a putative function to the proteins predicted to be coded by sequences PT11-C2-300-054-G01 and PT11-C2-300-012-H06 remains speculative, but it would be interesting to find out whether their expression is indeed bark-specific.

Conclusions and Perspectives

This initial characterization of a large number of putative members of the *Citrus* spp terpene synthase gene family provides novel resources for research on terpene secondary metabolism in *Citrus* species. We report here the largest number of putative terpene synthase sequences ever published for a single plant genus. We have characterized 49 sequence contigs that might represent 49 different genes, although the exact number of members of the *Citrus* terpene synthase gene family will be established only when a complete genome sequence is available for *Citrus*. Together with the recent characterization of the complete terpene synthase gene family for the model plant *Arabidopsis* (Aubourg *et al.*, 2002) and the functional characterization of some of its members (Bohlmann *et al.*,

2000; van Poecke *et al.*, 2001), our findings add exciting new aspects to our concept of secondary metabolism in *Citrus* and open several new avenues for natural product research directed by genome analysis. The expression patterns of some of the *Citrus* putative terpene synthases presented here will be tested in future work by using a combination of *in situ* hybridization and functional experiments.

Acknowledgments

We would like to thank Marcos Machado for coordinating our efforts on the analysis of the data generated by the CITEST Project, Luiz Humberto Gomez for help in early stages of this work and CNPq and FAPESP (Brazil) for financial support. The authors would also like to thank CNPq for research fellowships.

References

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.

Arimura G, Huber DP and Bohlmann J (2004a) Forest tent caterpillars (*Malacosoma disstria*) induce local and systemic diurnal emissions of terpenoid volatiles in hybrid poplar (*Populus trichocarpa x deltoides*): cDNA cloning, functional characterization, and patterns of gene expression of (-)-germacrene D synthase, PtdTPS1. Plant J 37:603-616.

Arimura G, Ozawa R, Kugimiya S, Takabayashi J and Bohlmann J (2004b). Herbivore-induced defense response in a model legume: Two-spotted spider mites induce emission of (E)-β-ocimene and transcript accumulation of (E)-β-ocimene synthase in *Lotus japonicus*. Plant Physiol 135:1976-1983.

Aubourg S, Lecharny A and Bohlmann J (2002) Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. Mol Genet Genomics 267:730-745.

Baldwin IT, Halitschke R, Paschold A, Von Dahl CC and Preston CA (2006) Volatile signaling in plant-plant interactions: "Talking trees" in the genomics era. Science 311:812-815.

Bohlmann J, Meyer-Gauen G and Croteau R (1998) Plant terpenoid synthases: Molecular biology and phylogenetic analysis. Proc Natl Acad Sci USA 95:4126-4133.

Bohlmann J, Steele CL and Croteau R (1997) Monoterpene synthases from grand fir (*Abies grandis*): cDNA isolation, characterization, and functional expression of myrcene synthase, (-)-(4S)-limonene synthase, and (-)-(1S,5S)-pinene synthase. J Biol Chem 272:21784-21792.

Bohlmann J, Meyer-Gauen G and Croteau R (1998) Plant terpenoid synthases: Molecular biology and phylogenetic analysis. Proc Natl Acad Sci USA 95:4126-4133.

Bohlmann J, Martin D, Oldham NJ and Gershenzon J (2000) Terpenoid secondary metabolism in *Arabidopsis thaliana*: cDNA cloning, characterization, and functional expression of a myrcene/(E)-b-ocimene synthase. Arch Biochem Biophys 375:261-269.

Botella-Pavia P and Rodriguez-Concepcion M (2006) Carotenoid biotechnology in plants for nutritionally improved foods. Physiol Plant 126:369-381.

Boteon M and Neves EM (2005). Citricultura brasileira: Aspectos econômicos. In: Mattos Jr D, De Negri JD and Pompeu Jr J (eds) Citrus. Instituto Agronômico de Campinas, Campinas, pp 19-36.

Byun-McKay A, Godard KA, Toudefallah M, Martin DM, Alfaro R, King J, Bohlmann J and Plant AL (2006) Wound-induced terpene synthase gene expression in sitka spruce that exhibit resistance or susceptibility to attack by the white pine weevil. Plant Physiol 140:1009-1021.

Chen XY, Chen Y, Heinstein P and Davisson VJ (1995) Cloning, expression, and characterization of (+)-delta-cadinene synthase: A catalyst for cotton phytoalexin biosynthesis. Arch Biochem Biophys 324:255-266.

Claeys M, Graham B, Vas G, Wang W, Vermeylen R, Pashynska V, Cafmeyer J, Guyon P, Andreae MO, Artaxo P, *et al.* (2004) Formation of secondary organic aerosols through photo-oxidation of isoprene. Science 303:1173-1176.

Copolovici LO, Filella I, Llusia J, Niinemets U and Penuelas J (2005) The capacity for thermal protection of photosynthetic electron transport varies for different monoterpenes in *Quercus ilex*. Plant Physiology 139:485-496.

Corey EJ, Matsuda SPT and Bartel B (1993) Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen. Proc Natl Acad Sci USA 90:11628-11632.

Croteau R, Kutchan TM and Lewis NG (2000) Natural products In: Buchanan B, Gruissem W and Jones R (eds) Biochemistry and Molecular Biology of Plants. American Society of Plant Physiology, New York, pp 1250-1318.

Davis EM and Croteau R (2000) Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. Curr Topics Chem 209:53-95.

Dudareva N, Cseke L, Blanc VM and Pichersky E (1996) Evolution of floral scent in *Clarkia*: Novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. Plant Cell 8:1137-1148.

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.

Goodwin TW (1967) Terpenoids in Plants. Academic Press, New York, 326 pp.

Huynen MA and Bork P (1998) Measuring genome evolution. Proc Natl Acad Sci USA 95:5849-5856.

Keeling CI and Bohlmann J (2006) Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. New Phytol 170:657-675.

Kumar S, Tamura K, Jacobsen I and Nei N (2000) MEGA: Molecular Evolutionary Genetics Analysis, version 2.0.Pennsylvania and Arizona State University, University Park, Pennsylvania and Tempe, Arizona.

Phillips DR, Rasbery JR, Bartel B and Matsuda SPT (2006) Biosynthetic diversity in plant triterpene cyclization. Curr Opin Plant Biol 9:305-314.

Rohdich F, Bacher A and Eisenreich W (2005) Isoprenoid biosynthetic pathways as anti-infective drug targets. Biochem Soc Trans 33:785-791.

Ruberto G and Rapisarda P (2002) Essential oils of new pigmented citrus hybrids: *Citrus sinensis* L. Osbeck x *C. clementina* Hort. ex Tanaka. J Food Sci 67:2778-2780.

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

Sawamura M, Tu NTM, Yu XL and Xu BQ (2005) Volatile constituents of the peel oils of several sweet oranges in China. J Essen Oil Res 17:2-6.

Sharkey TD, Yeh S, Wiberley AE, Falbel TG, Gong D and Fernandez DE (2005) Evolution of the isoprene biosynthetic pathway in kudzu. Plant Physiol 137:700-712.

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.

Shimada T, Endo T, Fujii H, Hara M, Ueda T, Kita M and Omura M (2004) Molecular cloning and functional characterization of four monoterpene synthase genes from *Citrus unshiu* Marc. Plant Sci 166:49-58.

Shimada T, Endo T, Fujii H and Omura M (2005) Isolation and characterization of a new d-limonene synthase gene with a different expression pattern in *Citrus unshiu* Marc. Sci Hortic 105:507-512.

Steele CL, Katoh S, Bohlmann J and Croteau R (1998) Regulation of oleoresinosis in grand fir (*Abies grandis*) - Differential transcriptional control of monoterpene, sesquiterpene, and diterpene synthase genes in response to wounding. Plant Physiol 116:1497-1504.

Sun T and Kamiya Y (1994) The *Arabidopsi*s GA1 locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis. Plant Cell 6:1509-1518.

Tatusov RL, Koonin EV and Lipman DJ (1997) A genomic perspective on protein families. Science 278:631-637.

Tholl D (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Curr Opin Plant Biol 9:297-304.

Thompson JD, Higgins DG and Gibson, TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.

Trapp S and Croteau R (2001) Defensive resin biosynthesis in conifers. Annu Rev Plant Physiol Plant Molec Biol 52:689-724.

Van Poecke RMP, Posthumus MA and Dicke M (2001) Herbivore-induced volatile production by *Arabidopsis thaliana* leads to attraction of the parasitoid *Cotesia rubecula*: Chemical, behavioural and gene-expression analysis. J Chem Ecol 27:1911-1928.

Velikova V and Loreto F (2005) On the relationship between isoprene emission and thermotolerance in *Phragmites australis* leaves exposed to high temperatures and during the recovery from heat stress. Plant Cell Environ 28:318-327.

Verzera A, Trozzi A, Zappala M, Condurso C and Cotroneo A (2005) Essential oil composition of *Citrus meyerii* Y. Tan. and *Citrus medica* L. cv. Diamante and their lemon hybrids. J Agric Food Chem 53:4890-4894.

Yamaguchi S, Sun T-P, Kawaide H and Kamiya Y (1998) The GA2 locus of *Arabidopsis thaliana* encodes ent-kaurene synthase of gibberellin biosynthesis. Plant Physiol 116:1271-1278.

Internet Resources

CitEST Database, http://citest.centrodecitricultura.br/ (March 25, 2006).

National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/ (March 25, 2006).

BLAST, http://www.ncbi.nlm.nih.gov/BLAST/ (March 25, 2006).

InterProScan, http://www.ebi.ac.uk/InterProScan/ (March 28, 2006).

Associate Editor: Marco Aurélio Takita