Research Article

Towards the identification of flower-specific genes in *Citrus* spp.

Marcelo Carnier Dornelas[1], Raquel Luciana Boscariol Camargo[2], Irving Joseph Berger[2]
and Marco Aurélio Takita[2]

*[1]Departamento de Fisiologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas,
Campinas, SP, Brazil.*
[2]Centro APTA Citrus Sylvio Moreira, Instituto Agronômico de Campinas, Cordeirópolis, SP, Brazil.

Abstract

Citrus sinensis is a perennial woody species, for which genetic approaches to the study of reproductive development are not readily amenable. Here, the usefulness of the CitEST Expressed Sequence Tag (EST) database is demonstrated as a reliable new resource for identifying novel genes exclusively related to *Citrus* reproductive biology. We performed the analysis of an EST dataset of the CitEST Project containing 4,330 flower-derived cDNA sequences. Relying on bioinformatics tools, sequences exclusively present in this flower-derived sequence collection were selected and used for the identification of *Citrus* putative flower-specific genes. Our analysis revealed several *Citrus* sequences showing significant similarity to conserved genes known to have flower-specific expression and possessing functions related to flower metabolism and/or reproductive development in diverse plant species. Comparison of the *Citrus* flower-specific sequences with all available plant peptide sequences unraveled 247 unique transcripts not identified elsewhere within the plant kingdom. Additionally, 49 transcripts, for which no biological function could be attributed by means of sequence comparisons, were found to be conserved among plant species. These results allow further gene expression analysis and possibly novel approaches to the understanding of reproductive development in *Citrus*.

Key words: development, flower-specific genes, reproduction, transcription factors.

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

Introduction

Understanding flowering and reproduction of perennial plant species is not only a fundamental concern of plant biology but is also of practical interest in agriculture. The identification of genes involved in flowering and reproduction of perennial plants could greatly contribute to improvements in breeding and in establishing alternative techniques to obtain interesting agronomic traits. Nevertheless, our current knowledge of molecular pathways controlling flower development comes mostly from studies of the herbaceous model plant *Arabidopsis thaliana* reviewed by Komeda (2004) and Kramer and Hall (2005). In recent years, gene expression analysis using genomic tools has become a powerful resource for the unraveling of flowering-related genes in other non-model plant species (Dornelas and Rodriguez, 2001; Izawa *et al.*, 2003; Dornelas and Rodriguez, 2004; 2005; Hecht *et al.*, 2005; Laitinen *et al.*, 2005; Dornelas and Rodriguez, 2006; Dornelas *et al.*, 2006).

Send correspondence to Marcelo C. Dornelas. Departamento de Fisiologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Caixa Postal 6109, 13083-970 Campinas, SP, Brazil. E-mail: dornelas@unicamp.br.

Sequencing of cDNA stretches to generate expressed sequence tags (ESTs) have proven to be a powerful, economical and rapid approach to identify genes that are preferentially expressed in certain tissue or cell types of multicellular organisms (Adams *et al.*, 1992; McCombe *et al.*, 1992; Newman *et al.*, 1994). When ESTs are generated from non-normalized cDNA libraries, gene expression patterns can be inferred from the relative abundance of these tags among different libraries (Ewing *et al.*, 1998). The availability of a significant EST database from a certain plant species offers the possibility of studying gene expression for different tissues and organs. This approach, associated with microarray techniques, has implicated 724 genes in *Arabidopsis* floral development (Hu *et al.* 2003). A genome-wide microarray study of the *Arabidopsis* male gametophytic transcriptome identified 992 pollen-specific transcripts (Honys and Twell, 2003). Similarly, with the use of both cDNA and oligonucleotide arrays on the *Arabidopsis* floral homeotic mutants *apetata*1, *apetala*2, *apetala*3, *pistillata* and *agamou*s, 1,453 genes were identified to be specifically or at least predominantly expressed in one type of floral organ (Wellmer *et al.*, 2004).

Recently, the use of a genomic approach, based on *in silico* EST sequence analysis, has been used to identify flower-specific genes in diverse non-model plant species (Figueiredo *et al.*, 2001; Dornelas and Rodriguez, 2004; Forment *et al.*, 2005; Hecht *et al.*, 2005; Laitinen *et al.*, 2005; Dornelas and Rodriguez, 2006; Dornelas *et al.*, 2006). Here we report on the screening of the CitEST Expressed Sequence Tag (EST) database for identifying novel genes exclusively related to sweet orange (*Citrus sinensis* L. Osbeck) reproductive biology. As *Citrus* species are generally perennial woody plants, for which genetic approaches to the study of reproductive development are not readily amenable, we believe the results presented here will allow novel approaches to the understanding of reproductive development in *Citrus*.

Material and Methods

Expressed Sequence Tags (ESTs) were generated by the CitEST Project from diverse *Citrus* species and different tissues. Nevertheless, a single flower-derived library was produced, containing cDNAs from mature flowers and flower buds at different developmental stages, of sweet orange (*Citrus sinensis*, var. Pêra IAC). Information concerning the construction of libraries, sequencing, sequence clustering and nomenclature can be found in other papers of this journal issue. In this paper, we have adopted as a cluster name the first sequence that was picked to form such a cluster.

The identification of flower-specific sequence clusters was based upon the rule that each cluster must contain only reads derived from the cDNA library made with floral tissues (library CS00-C5-003). For this purpose, custom-made scripts in Perl programming language were used to query the CitEST database. These scripts were designed to cluster all reads of the database and identify the tissue origin (library) of the reads inside each cluster. If a cluster was entirely formed by reads from the flower library, it was selected. Additionally, all singletons from the flower library (isolated sequences that did not form clusters with any other sequence) were also selected.

The putative identity of each *Citrus* flower-specific cluster/singleton was established by performing BLAST (Altschul *et al.*, 1997) searches against the GenBank databases (Benson *et al.*, 2000). A putative *Arabidopsis* ortholog was attributed to each *Citrus* flower-specific cluster/singleton, by querying the Arabidopsis Genome Initiative dataset using the TAIR BLAST 2.2.8 algorithm, considering e-values better than e^{-10}. Alternatively, the Plant Genome Database (Dong *et al.*, 2005) and the TIGR Plant Gene Indices (Quackenbush *et al.*, 2000) were also searched for the identification of putative homologs of the *Citrus* flower-specific sequences in other plants. Additionally, sequences were functionally characterized according to the MIPS Funcat (Mewes *et al.*, 2004). The ten most expressed sequences (*i.e.* the clusters composed by the largest number of reads) classified as "unknown function" were further investigated for the presence of conserved motifs by querying the Pfam database (Bateman *et al.*, 2004). A double *in silico* hybridization strategy, combining the likelihood algorithm (R-statistics) proposed by Stekel *et al.*, (2000) to compare multiple libraries at once and the P-statistics described by Audic and Claverie (1997), was used to identify differentially expressed clusters among *Citrus* tissues. All statistically significant flower-specific clusters from the R-statistics were concluded as upregulated, flower-specific expressed sequences. Additional and simultaneous 2x2 P-statistics significance of those flower-specific clusters in contrast to the leaf and fruit datasets was interpreted as a strong suggestion of tissue specificity.

Results

Identifying *Citrus* flower-specific sequences

A single cDNA library derived from flower tissues was sequenced within the frame of the CitEST Project. A total of 4,330 valid EST sequences were produced from this library, which represents 2.5% of the total number of valid sequences produced. After assembly and selection of the clusters containing only flower-derived transcripts and/or single flower-derived sequences that did not form clusters, 1,012 putative flower-specific sequences were found. These represent 23% of the number of flower-derived sequences and only 0.5% of the total number of ESTs in the CitEST database. These putative flower-specific sequences could be organized in 133 clusters (31% of the ESTs) and 696 singletons. Of the putative flower-specific clusters, 97% contained two or three ESTs indicating that the library is extremely non-redundant and that more flower-specific sequences might be obtained with further sequencing of other cDNA clones derived from this library.

Functional annotation of *Citrus* putative flower-specific sequences

Analysis of the putative flower-specific sequences and their derived peptide sequences allowed a tentative annotation of their biological functions to be performed (Table 1). Functional assignments were calculated from BLASTX, performed for the functionally annotated proteins from the *Arabidopsis* genome (the MIPS Funcat). In total, 23.6% of the *Citrus* putative flower-specific sequences were homologous to *Arabidopsis* proteins of known function, with another 15.3% similar to *Arabidopsis* proteins of unknown or unclear function. The remaining 43.9% of *Citrus* putative flower-specific sequences without a BLAST match above the threshold were designated as "unknowns". Nevertheless, comparisons of the *Citrus* sequences to more broad databases, such as GenBank and/or other plant genome/EST sequence collections, such as the TIGR Plant Gene Indices (Quackenbush *et al.*, 2000), al-

Table 1 - Functional classification of the *Citrus* putative flower-specific transcripts.

Match *Arabidopsis* MIPS[a]	56.1
Functionally annotated[b]	23.6
Ambiguous/unknown function	15.3
Metabolism	9.8
Cellular organization	7.3
Signal transduction	4.6
Cell defense	4.3
Transcription	3.8
Energy	3.6
Protein destination	3.3
Cellular biogenesis	3.0
Protein synthesis	2.8
Transport facilitation	2.5
Transport mechanisms	2.3
Cell growth	1.9
Development	1.6
No match to *Arabidopsis* MIPS[c]	43.9

[a]The appropriate BLAST algorithms were used to query functionally annotated *Arabidopsis* proteins (MIPS Funcat) and the results were filtered using the expectation value of e^{-10}. This results in *Citrus* clusters or singletons that correspond to a functionally annotated protein or to putative novel genes.
[b]The functionally annotated group represents proteins of known function, but also includes sequences of categories of unknown or unclear functions (MIPS Funcat codes 98 and 99).
[c]Sequences that cannot be assigned to a functional class were listed separately from the sequences to which a putative function could be assigned. Note that more than one function can be assigned to the same sequence, so the sum of assignments is greater than the whole.

Table 2 - Comparison of the *Citrus* flower-specific EST collection with other sequence collections of whole genomes, partial genomes or large EST projects.

	Matches to database (%)[a]	Unique to database[b]
Arabidopsis thaliana proteome	596 (58.9)	13
Oryza sativa proteome	184 (18.2)	0
Populus trichocarpa genome	475 (47.3)	3
Medicago truncatula BAC collection	139 (13.8)	2
Pooled Asterid EST sequences	418 (41.4)	6
Pooled Eurosid EST sequences	580 (57.4)	16
Pooled Monocot EST sequences	197 (19.5)	1
GenBank protein sequence database	605 (59,8)	8
Unassigned *Citrus* flower-specific ESTs		247

[a]*Citrus* flower-specific clusters and singleton sequences were queried in the databases using the appropriate BLAST algorithm (Altschul *et al.*, 1997), and the results were filtered arbitrarily at e^{-10}. The number of *Citrus* sequences that can be mapped to the query collection is shown along with this value, expressed as a percentage of all *Citrus* flower-specific sequences.
[b]Also shown is the number of sequences that are unique to the queried database and have no homolog elsewhere within the experiment.

lowed further identification and annotation of the *Citrus* putative flower-specific transcripts (Table 2).

Comparison of *Citrus* putative flower-specific sequences with other plant genomes and EST databases

The *Citrus* putative flower-specific sequence collection was placed in comparative context with other plant species. As references, the annotated *Arabidopsis* and rice protein collections were included along with the draft *Populus trichocarpa* genome and all available *Medicago truncatula* BAC sequences. To identify putatively lineage-specific sequences, TIGR Plant Gene Indices (Quackenbush *et al.*, 2000) clustered EST collections were pooled to form Eurosid, Asterid, and monocot collections. While 58.9% of *Citrus* putative flower-specific sequences have a counterpart in the annotated *Arabidopsis* proteome (since the *Arabidopsis* GenBank database is more redundant than MIPS), only 18% of sequences have a match within the rice genome (Table 2). When the comparison is restricted to *Populus*, 47.3% of the *Citrus* sequences have a match (Table 2). Of the 1,012 *Citrus* putative flower-specific se-

quences, 247 sequences (24%) do not have a match to a known sequence. Of these, 3% can be excluded as short sequences or sequences likely to represent untranslated regions (UTRs, with less than 10% of coding potential). This results in 163 (16%) *Citrus* putative flower-specific sequences not observed elsewhere within the plant kingdom. Nevertheless, we cannot discard the possibility that some of these sequences may be artifacts or products of contamination, thus further expression pattern analysis (*e.g.*, by *in situ* hybridization experiments) is needed to confirm this observation.

Identification of upregulated sequences as candidates to the most probable flower-specific genes

When sequences are generated from non-normalized cDNA libraries, the relative abundance of ESTs reflects the gene expression patterns in terms of up- and downregulated genes (Ewing *et al.*, 1998). As the library of flower tissues made by the CitEST Project is non-normalized, we applied statistical algorithms to predict upregulated, thus statistically supported, stronger candidates to flower-specific genes in *Citrus* flower tissues.

In silico hybridization of all CitEST-derived clusters of *C. sinensis* flower, leaf and fruit ESTs have pointed to all 133 exclusive flower clusters being statistically supported upregulated genes in that tissue (R-value > 5.0, data distribution not shown; Stekel *et al.*, 2000). Additional and simultaneous significance of upregulated flower clusters, in comparison to the other tissues resulting from 2x2 *in silico* hybridizations (P-value < 0.05; Audic and Claverie, 1997), was interpreted as a strong suggestion of tissue specificity.

Only ten clusters from the first hybridization were found to fit that second condition (it includes only clusters with 4 or more ESTs). Corresponding to the ten most abundant expressed sequences in the tissue, they are hereinafter considered the most probable candidates to flower-specific genes in *Citrus*.

Characterization of genes upregulated during *Citrus* flower development

Indeed, most of the genes represented by the ten most abundant Citrus putative flower-specific ESTs encode proteins related to the biosynthesis of flower-specific products found in other species (Table 3). Seven out of the ten most expressed *Citrus* flower-specific genes could have a putative function attributed to them, based on sequence comparisons with proteins for which a function was previously attributed experimentally in model plants.

The most abundant *Citrus* flower-specific transcript, CS00-C5-003-027-G10-CT, is a putative homolog to the *Arabidopsis* flower-specifc *AtMYB21* (Table 3), which encodes a transcription factor belonging to the large MYB family (Shin *et al.*, 2002). Transgenic *Arabidopsis* plants overexpressing *AtMYB21* have shorter stems, narrower petals and malformed carpels (Shin *et al.*, 2002). This gene is conserved in other plant species such as *Gerbera* and *Pisum* and always shows a flower-specific pattern of expression (Uimari and Strommer, 1997; Elomaa *et al.*, 2003), and is involved in the biosynthesis of phenylpropanoids, including anthocyanin (Elomaa *et al.*, 2003) and other phlobaphene pigments (Uimari and Strommer, 1997). In accordance with this biological role, the expression of *AtMYB21* is regulated by light-signaling components such as COP1

and is ectopically expressed in cop1 mutants (Shin *et al.*, 2002).

CYP79A2, a member of the *Arabidopsis* cytochome P450 family, is the most probable homolog to the CS00-C5-003-015-F09 cluster (Table 3), the second most expressed *Citrus* putative flower-specific transcript. It is expressed preferentially in carpels and is involved in converting L-phenylalanine to phenylacetaldoxime, and is probably related to hormone homeostasis, as knock-out mutants show increased levels of cytokinins and overproliferation of carpels (Wittstock and Halkier, 2000; Tantikanjana *et al.*, 2004).

Three out of the ten most abundant *Citrus* putative flower-specific transcripts encode methyltransferases (Table 3). The analysis of expression patterns coupled with biochemical characterization showed that these carboxyl methyltransferases are involved either in floral scent biosynthesis or in plant defense responses (Effmert *et al.*, 2005). Clusters CS00-C5-003-026-G09-CT and CS00-C5-003-009-A05-CT encode putative benzenoid carboxyl methyltransferases which synthesize methyl esters, the constituents of aromas and scents of many plant species. The top ten BLAST hits for these clusters are S-adenosyl-1-methionine:benzoic acid carboxyl methyltransferases expressed exclusively in petals and are key regulators of flower scent production in roses, petunia, snapdragon, *Clarkia* and *Stephanotis* (Dudareva *et al.*, 2000; Lavid *et al.*, 2002; Effmert *et al.*, 2005; Scalliet *et al.*, 2006). On the other hand, CS00-C5-003-051-F10-CT encodes a putative 7-methylxanthine methyltransferase, similar to one of the caffeine-synthases which are specifically expressed in

Table 3 - Characterization of the ten most expressed *Citrus* putative flower-specific transcripts.

CitEST cluster[a]	reads[b]	homology[c]	e-value[d]	*Arabidopsis*[e]	Putative function[f]
CS00-C5-003-027-G10-CT	8	MYB-family transcription factor	$2e^{-62}$	At3g27810	regulates phenylpropanoid biosynthesis
CS00-C5-003-015-F09-CT	7	cytochrome P450	$3e^{-94}$	At5g05260	biosynthesis of benzylglucosinolate
CS00-C5-003-026-G09-CT	7	SAM-dependent methyltransferase	$2e^{-32}$	At4g36470	biosynthesis of methyl benzoate
CS00-C5-003-028-B11-CT	7	STAR-related membrane protein	$2e^{-94}$	At1g64720	unknown
CS00-C5-003-051-F10-CT	5	caffeine synthase	$1e^{-44}$	At5g66430	biosynthesis of xanthine derivatives
CS00-C5-003-030-D01-CT	5	no hit	-	-	-
CS00-C5-003-016-D04-CT	5	lipid transfer protein	$7e^{-66}$	At5g52130	tapetum development
CS00-C5-003-049-D12-CT	4	UDP-glucuronic acid decarboxylase	$1e^{-69}$	At5g59290	stamen development
CS00-C5-003-009-A05-CT	4	orcinol O-methyltransferase	$7e^{-68}$	At4g35160	biosynthesis of orcinol dimethyl ether
CS00-C5-003-038-A01-CT	4	glycine-rich membrane protein	$1e^{-75}$	At4g30420	unknown

[a]Each cluster name is derived from the code of the first EST taken to start building such a cluster and it has no relation to how representative this sequence is of the whole cluster consensus sequence.
[b]Number of EST sequences used to form the referred cluster.
[c]The results refer to the best hit when using BLASTx algorithm to query the non-redundant protein sequence dataset of GenBank.
[d]The results presented refer to the best hit when using BLASTx algorithm to query the functionally annotated *Arabidopsis* proteins (MIPS Funcat).
[e]Taking into account the presence of conserved protein domains and the expression patterns and/or experimentally determined functions reported for the putative homologs of these sequences in *Arabidopsis* or other plant species.

young floral buds, flowers and fruits, specially in the endosperm tissue (Mizuno *et al.*, 2003).

The other three out of the ten most abundant *Citrus* putative flower-specific transcripts encode proteins whose functions remain unknown, even if their sequences are conserved among different plant species (CS00-C5-003-028-B11-CT and CS00-C5-003-038-A01-CT; Table 3) or the querying of their sequences in the databases produced non-significative hits (CS00-C5-003-030-D01-CT; Table 3). Although their biological roles remain unknown, putative homologs to clusters CS00-C5-003-028-B11-CT and CS00-C5-003-038-A01-CT could be found within the *Arabidopsis* proteome and they encode a START-related membrane protein and a glycine-rich cell wall protein, respectively. START (STeroidogenic Acute Regulatory-related lipid Transfer) proteins are generally involved in the transport of phosphatidylcholine-derivatives and may be implicated in signal transduction (Bateman *et al.*, 2004). Phosphatidylcholine, phosphatidylethanolamine and phosphatidic acid are major components of the polar lipid fraction of pollen grains (Caffrey *et al.*, 1987). On the other hand, glycine-rich proteins are among the most expressed anther-specific genes of *Arabidopsis* (Rubinelli *et al.*, 1998; Oliveira *et al.*, 1993) and have been involved in stamen development in lily, where they are expressed exclusively in anther tissues (Mousavi *et al.*, 1999).

Two other putative stamen- or anther-specific contigs were found among the most abundant *Citrus* flower-specific ESTs: CS00-C5-003-016-D04-CT and CS00-C5-003-049-D12-CT. *In situ* hybridization studies have demonstrated that FIL1, a protease inhibitor/lipid-transfer protein which is the *Antirrhinum* putative homolog to CS00-C5-003-016-D04-CT, is expressed exclusively in stamens and petals (Nacken *et al.*, 1991). Likewise, *ATA7*, the putative homolog of this same gene in *Arabidopsi*s, is tapetum-specific (Rubinelli *et al.*, 1998). On the other hand, CS00-C5-003-049-D12-CT encodes a putative nucleotide sugar epimerase (UDP-glucuronic acid decarboxylase). A homolog to this protein has been implicated in stamen development in cucumber and is strongly expressed in bisexual and male cucumber flowers, but its transcripts are absent from female flowers and vegetative tissues (Terefe and Tatlioglu, 2005).

Typically, the cDNAs encoding transcription factors upregulated in flowers are MADS-box genes, which are known regulators of flower-organ development (see review by Irish and Litt, 2005). Accordingly, we have found among the *Citrus* flower-specific transcripts (namely CS00-C5-003-003-A04-CT), a putative homolog to the *APETALA1* gene (with an e-value of $4e^{-78}$), which encodes a flower-specific MADS-box transcription factor, responsible for the determination of sepal and petal identities (Mandel *et al.*, 1992). Additionally, a Citrus putative homolog to a SHAGGY-like kinase ASK (Dornelas *et al.*, 1998; 2000), which is exclusively expressed in pollen and

ovules (Dornelas *et al.*, 2000; Wellmer *et al.*, 2004), was also found among the flower-specific transcripts (*i.e.* CS00-C5-003-063-B12-CT, e-value of $2e^{-82}$).

Discussion

The accumulation of sequence data from taxonomically diverse species evidently benefits the functional analysis of the corresponding genes in all experimental systems, as well as the understanding of plant evolution both from a phylogenetic and a mechanistic perspective (Cronk, 2001; Albert *et al.*, 2002; Frohlich, 2003). Our previous work demonstrates that sequence comparisons, in combination with phylogenetic analyses, reveal functionally related gene groups, but also produce predictions for gene duplication and functional diversification during the evolution of plant reproductive development (Dornelas and Rodriguez, 2001; 2004; 2005; 2006; Dornelas *et al.*, 2006). Comparison of the *Citrus* EST data with the available plant genomes and pooled EST collections, representing evolutionary distinct lineages within the plant kingdom, nicely demonstrates the high potential for the discovery of novel genes. As the genus *Citrus* belongs to the core Eucotyledons, specifically to the Eurosids clade, it was expected that comparisons of the *Citrus* ESTs to those of other Eurosids would produce a greater degree of similarity (thus a greater chance of finding putative homologs) than in comparison to those of Asterids or even to those of Monocotiledonous plants. It was also striking that the degree of similarity found among *Citrus* flower-derived sequences and those of other woody species, such as those from *Populus* (Table 2) or those of apple and peach (data not shown), was frequently greater than when comparing *Citrus* sequences with other herbaceous plants. This could indicate that reproductive development of woody perennials would share, at least to some extent, particular motifs not found among herbaceous-derived proteins. This observation has also been reported for other woody species such as *Eucalyptus* (Dornelas and Rodriguez, 2005) and apple (Newcomb *et al.*, 2006).

Among the *Citrus* putative flower-specific transcripts, we found several clusters showing significant sequence similarity with known floral organ-specific genes, which encode proteins that control flower development and metabolism in a number of species. These results indicate that the selection criteria applied here were suitable for the identification of flower-expressed transcripts. Future analysis of the expression patterns of these previously uncharacterized genes by techniques such as *in situ* hybridization might bring further support to this conclusion. Additionally, it would be of great interest to validate and correlate the double *in silico* hybridization strategy herein adopted with *ex-silico* expression patterns. Strict correlation would suggest an efficient *in silico* approach for allowing tissue-specific gene discovery and for directing validation experiments to more limited and probable positive targets.

This has been accomplished recently for genes expressed during *Arabidopsis* early seed development (Becerra *et al.*, 2006).

Many of the *Citrus* transcripts upregulated in floral tissues encode proteins involved in specific steps of flower metabolism and only a few were putatively involved in early steps of floral evocation or flower meristem differentiation. This bias towards late-expressing genes can be explained by the fact that the RNA preparations used for the cDNA library construction were, because of the differences in size of young and old floral buds, strongly enriched for RNA from older buds. Therefore, genes that are expressed during early stages of flower development might have been too diluted in the RNA samples, and thus are underrepresented in the CitEST database.

The majority of the genes that were predicted to be floral organ-specific were assigned to the stamen, whereas only very few were assigned to the organs of the perianth. This difference in number is likely because of key developmental events, such as the formation of pollen and ovules that occur during late stages of flower development in the reproductive organs. In addition, the reproductive organs contain many different tissues and cell types, whereas the anatomy of sepals and petals appears to be less complex (Wellmer *et al.*, 2004). Thus, the observation that the number of genes expressed in stamens is generally greater than the number of genes expressed in other floral organs has also been reported for *Arabidopsis* (Wellmer *et al.*, 2004) and *Gerbera* (Laitinen *et al.*, 2005).

We have also identified a large number of genes that are specifically expressed or predominantly abundant in *Citrus* floral tissues which have not yet been characterized in detail, even in model plants such as *Arabidopsis*. Nevertheless, in *Arabidopsis*, the targeted inactivation of genes has become a very powerful approach for functional analysis. RNA interference can be used to induce loss-of-function phenotypes (Chuang and Meyerowitz, 2000), and T-DNA insertion lines are available for many genes (Alonso *et al.*, 2003). Thus, the function of the putative flower-specific genes identified in *Citrus* can now be systematically studied by reverse genetics in heterologous model systems such as *Arabidopsis*. On the other hand, more than 200 *Citrus* putative flower-specific transcripts showed no significant homology to any other plant transcript, which strongly suggests putative functions for them in aspects of flower organ development and/or metabolism that are particular to *Citrus* species. Thus, these sequences could encode putative novel regulators of flower development and metabolism, yet to be described.

Based on the observations above, we conclude that we have successfully uncovered *Citrus* transcripts that are putatively flower-specific. Our results also indicate that spatially limited expression of several genes may be part of *Citrus* flower development and metabolism, as has been demonstrated for model plants (Wellmer *et al.*, 2004). Our data additionally provide a rich source of target genes for reverse genetics approaches and candidates for flower-specific markers.

Acknowledgments

References

Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C and Venter JC (1992) Sequence identification of 2,375 human brain genes. Nature 355:632-634.

Albert VA, Oppenheimer DG and Lindqvist C (2002) Pleiotropy, redundancy and the evolution of flowers. Trends Plant Sci 7:297-301.

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653-657.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research 25:3389-3402.

Audic S and Claverie JM (1997) The significance of digital gene expression profiles. Genome Res 710:986-995.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, *et al.* (2004) The Pfam protein families database. Nucleic Acids Res 32:D138-D141.

Becerra C, Puigdomenech P and Vicient CM (2006) Computational and experimental analysis identifies *Arabidopsis* genes specifically expressed during early seed development. BMC Genomics 7:38.

Benson DA, Karlsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA and Wheeler DL (2000) GenBank. Nucleic Acid Res 28:15-18.

Caffrey M, Werner BG and Priestley DA (1987) A crystalline lipid phase in a dry biological system: Evidence from X-ray diffraction analysis of *Typha latifolia* pollen. Biochim Biophys Acta 921:124-134.

Chuang CF and Meyerowitz EM (2000) Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 97:4985-4990.

Cronk QCB (2001) Plant evolution and development in a post-genomic context. Nature Rev Gen 2:607-619.

Dong Q, Lawrence CJ, Schlueter SD, Wilkerson MD, Kurtz S, Lushbough C and Brendel V (2005) Comparative plant genomics resources at plantGDB. Plant Physiol 139:610-618.

Dornelas MC and Rodriguez APM (2001) A genomic approach to elucidating grass flower development. Genet Mol Biol 24:69-76.

Dornelas MC and Rodriguez APM (2004) Identification of differentially expressed genes during reproductive development

in sugarcane (*Saccharum* sp) by the analysis of expressed sequence tags. Flowering Newsletter 37:40-45.

Dornelas MC and Rodriguez APM (2005) Identifying *Eucalyptus* expressed sequence tags related to *Arabidopsis* flowering-time pathway genes. Braz J Plant Physiol 17:255-266.

Dornelas MC and Rodriguez, APM (2006) Evolutionary conservation of genes controlling flowering pathways between *Arabidopsis* and grasses. In: Teixeira da Silva JA (ed) Floriculture, Ornamental and Plant Biotechnology, v. 4. 1st edition. Global Science Books, London, pp 272-279.

Dornelas MC, Lejeune B, Dron M and Kreis M (1998) The *Arabidopsis* Shaggy-related protein kinase (ASK) gene family: Structure, organization and evolution. Gene 212:249-257.

Dornelas MC, Tsai SM and Rodriguez, APM (2006) Expressed sequence tags of genes involved in the flowering process of *Passiflora* spp. In: Teixeira da Silva JA (ed) Floriculture, Ornamental and Plant Biotechnology, v. 1. 1st edition. Global Science Books, London, pp 483-488.

Dornelas MC, Van Lammeren AAM and Kreis, M (2000) *Arabidopsis thaliana* SHAGGY-related protein kinases (AtSK11 and 12) function in perianth and gynoecium development. Plant J 21:419-429.

Dudareva N, Murfitt LM, Mann CJ, Gorenstein N, Kolosova N, Kish CM, Bonham C and Wood K (2000) Developmental regulation of methyl benzoate biosynthesis and emission in snapdragon flowers. Plant Cell 12:949-961.

Effmert U, Saschenbrecker S, Ross J, Negre F, Fraser CM, Noel JP, Dudareva N and Piechulla B (2005) Floral benzenoid carboxyl methyltransferases: From *in vitro* to *in planta* function. Phytochem 66:1211-1230.

Elomaa P, Uimari A, Mehto M, Albert VA, Laitinen RA and Teeri TH (2003) Activation of anthocyanin biosynthesis in *Gerbera hybrida* (Asteraceae) suggests conserved protein-protein and protein-promoter interactions between the anciently diverged monocots and eudicots. Plant Physiol 133:1831-1842.

Ewing B, Hillier L, Wendl MC and Green P (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res 8:175-185.

Figueiredo RC, Brito MS, Figueiredo LHM, Quiapin AC, Vitorelli PM, Silva LR, Santos RV, Molfetta JB, Goldman GH and Goldman MHS (2001) Dissecting the sugarcane expressed sequence tag (SUCEST) database: Unraveling flower-specific genes. Genet Mol Biol 24:77-84.

Forment J, Gadea J, Huerta L, Abizanda L, Agusti J, Alamar S, Alos E, Andres F, Arribas R and Beltran JP (2005) Development of a citrus genome-wide EST collection and cDNA microarray as resources for genomic studies. Plant Mol Biol 573:375-391.

Frohlich MW (2003) An evolutionary scenario for the origin of flowers. Nature Rev Gen 4:559-566.

Hecht V, Foucher F, Ferrandiz C, Macknight R, Navarro C, Morin J, Vardy ME, Ellis N, Beltran JP and Rameau C (2005) Conservation of *Arabidopsis* flowering genes in model legumes. Plant Physiol 137:1420-1434.

Honys D and Twell D (2003) Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiol 132:640-652.

Hu W, Wang Y, Bowers C and Ma H (2003) Isolation, sequence analysis, and expression studies of florally expressed cDNAs in *Arabidopsis*. Plant Mol Biol 53:545-563.

Irish VF and Litt A (2005) Flower development and evolution: Gene duplication, diversification and redeployment. Curr Opin Genet Dev 154:454-460.

Izawa T, Takahashi Y and Yano M (2003) Comparative biology comes into bloom: Genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. Curr Opin Plant Biol 6:113-120.

Komeda Y (2004) Genetic regulation of time to flower in *Arabidopsis thaliana*. Annu Rev Plant Biol 55:521-535.

Kramer EM and Hall JC (2005) Evolutionary dynamics of genes controlling floral development. Curr Opin Plant Biol 8:13-18.

Laitinen RA, Immanen J, Auvinen P, Rudd S, Alatalo E, Paulin L, Ainasoja M, Kotilainen M, Koskela S, Teeri TH, *et al.* (2005) Analysis of the floral transcriptome uncovers new regulators of organ determination and gene families related to flower organ differentiation in *Gerbera hybrida* (Asteraceae). Genome Res 15:475-486.

Lavid N, Wang J, Shalit M, Guterman I, Bar E, Beuerle T, Menda N, Shafir S, Zamir D, Adam Z, *et al.* (2002) O-methyltransferases involved in the biosynthesis of volatile phenolic derivatives in rose petals. Plant Physiol 129:1899-1907.

Mandel MA, Gustafson-Brown C, Savidge B and Yanofsky MF (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA1. Nature 360:273-277.

McCombe WR, Adams MD, Kelley JM, Fitzgerald MG, Utterback TR, Khan M, Dubnick M, Kerlavage AR, Venter JC and Fields C (1992) *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologs. Nature Genetics 1:124-131.

Mewes HW, Amid C, Arnold R, Frishman D, Guldener U, Mannhaupt G, Munsterkotter M, Pagel P, Strack N, Stumpflen V, *et al.* (2004) MIPS: Analysis and annotation of proteins from whole genomes. Nucleic Acids Res 32:D41-D44.

Mizuno K, Okuda A, Kato M, Yoneyama N, Tanaka H, Ashihara H and Fujimura T (2003) Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee (*Coffea arabica* L.). FEBS Lett 534:75-81.

Mousavi A, Hiratsuka R, Takase H, Hiratsuka K and Hotta Y (1999) A novel glycine-rich protein is associated with starch grain accumulation during anther development. Plant Cell Physiol 40:406-416.

Nacken WK, Huijser P, Beltran JP, Saedler H and Sommer H (1991) Molecular characterization of two stamen-specific genes, tap1 and fil1, that are expressed in the wild type, but not in the deficiens mutant of *Antirrhinum majus*. Mol Gen Genet 229:129-136.

Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EH, Allan AC, Beuning LL, Bowen JH, Gera E, Jamieson KR, Janssen BJ, *et al.* (2006) Analyses of expressed sequence tags from apple. Plant Physiol 141:147-166.

Newman T, De Bruijn FJ, Green P, Keegstra K, Kende H, Mcintosh L, Ohlrogge J, Raikhel N, Somerville S and Thomashow M (1994) Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol 106:1241-1255.

Oliveira DE de, Franco LO, Simoens C, Seurinck J, Coppieters J, Botterman J and Van Montagu M (1993) Inflorescence-specific genes from *Arabidopsis thaliana* encoding glycine-rich proteins. Plant J 3:495-507.

Quackenbush J, Liang F, Holt I, Pertea G and Upton J (2000) The TIGR Gene Indices: Reconstruction and representation of expressed gene sequences. Nucleic Acids Res 28:141-145.

Rubinelli P, Hu Y and Ma H (1998) Identification, sequence analysis and expression studies of novel anther-specific genes of *Arabidopsis thaliana*. Plant Mol Biol 37:607-619.

Scalliet G, Lionnet C, Le Bechec M, Dutron L, Magnard JL, Baudino S, Bergougnoux V, Jullien F, Chambrier P, Vergne P, *et al.* (2006) Role of petal-specific orcinol O-methyltransferases in the evolution of rose scent. Plant Physiol 140:18-29.

Shin B, Choi G, Yi H, Yang S, Cho I, Kim J, Lee S, Paek NC, Kim JH, Song PS, *et al.* (2002) AtMYB21, a gene encoding a flower-specific transcription factor, is regulated by COP1. Plant J 30:23-32.

Stekel DJ, Git Y and Falciani F (2000) The comparison of gene expression from multiple cDNA libraries. Genome Res 10:2055-2061.

Tantikanjana T, Mikkelsen MD, Hussain M, Halkier BA and Sundaresan V (2004) Functional analysis of the tandem-duplicated P450 genes SPS/BUS/CYP79F1 and CYP79F2 in glusinolate biosynthesis and plant development by Ds transposition-generated double mutants. Plant Physiol 135:840-848.

Terefe D and Tatlioglu T (2005) Isolation of a partial sequence of a putative nucleotide sugar epimerase, which may involve in stamen development in cucumber (*Cucumis sativus* L.). Theor Appl Genet 111:1300-1307.

Uimari A and Strommer J (1997) Myb26: A MYB-like protein of pea flowers with affinity for promoters of phenylpropanoid genes. Plant J 12:1273-1284.

Wellmer F, Riechmann JL, Alves-Ferreira M and Meyerowitz EM (2004) Genome-wide analysis of spatial gene expression in *Arabidopsis* flowers. Plant Cell 16:1314-1326.

Wittstock U and Halkier BA (2000) Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. J Biol Chem 275:14659-1466.

Internet Resources

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

Arabidopsis Genome Initiative, http://www.arabidopsis.org/Blast/ (February 10, 2006)

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

Plant Genome Database, http://www.plantgdb.org/ (February 10, 2006)

TIGR Plant Gene Indices, http://www.tigr.org/tdb/tgi/plant.shtml (February 10, 2006)

MIPS Funcat, http://mips.gsf.de/projects/funcat (March 19, 2006)

Pfam database, http://www.sanger.ac.uk/Software/Pfam/ (February 10, 2006)