

Document downloaded from:

<http://hdl.handle.net/10251/83446>

This paper must be cited as:

García Lor, A.; Garcia Martinez, J.L.; Perez Amador, MA. (2012). Identification of ovule and seed genes from Citrus clementina. *Tree Genetics and Genomes*. 8(2):227-235.
doi:10.1007/s11295-011-0435-x.



The final publication is available at

<http://doi.org/10.1007/s11295-011-0435-x>

Copyright Springer Verlag (Germany)

Additional Information

Authors:

Andres Garcia-Lor¹, Jose Luis Garcia-Martinez, and Miguel A Perez-Amador

Title:

Identification of ovule and seed genes from *Citrus clementina*

Affiliation:

Instituto de Biología Molecular y Celular de Plantas (IBMCP).

Universidad Politécnica de Valencia (UPV) - Consejo Superior de Investigaciones Científicas (CSIC).

Ciudad Politécnica de la Innovación, edificio 8E. Ingeniero Fausto Elio s/n
46022 Valencia, Spain

Current address (A G-L):

Instituto Valenciano de Investigaciones Agrarias (IVIA). Ctra. Moncada-Náquera km 4.8.

Moncada, 46113 Valencia, Spain

Corresponding Author:

Miguel A Perez-Amador

Phone: 34-963877723; e-mail:mpereza@ibmcp.upv.es

Abstract

Seedlessness is a highly desirable trait in fresh fruit. Citrus varieties, such as Clementine mandarin and other related species, show parthenocarpic fruit development without seeds due to self-incompatibility. In spite of that, these fruits frequently contain seeds as a result of cross-pollination by insects with compatible pollen from other citrus cultivars grown nearby. To solve this problem using a biotechnological approach we aim at the destruction of ovules and/or seeds by directing the expression of a toxic gene using the promoter of an ovule and/or seed specific gene. With the purpose of isolating this kind of genes we constructed two cDNA libraries from ovules and seeds at different developmental stages of the Clementine mandarin (*Citrus clementina* cv. Clemenules). A total of 1014 ESTs from the ovule library and 1042 ESTs from the seed library were generated, with a novelty percentage of 27% and 36% among the Spanish Citrus Functional Genomic Project (CFGP) ESTs database, respectively. Quantitative PCR analysis confirmed nearly specific expression in ovule and/or seed of two genes, *TRANSPARENT TESTA16* (*CcTT16*) and *TRANSPARENT TESTA7* (*CcTT7*). Expression of these two genes is restricted to early seed development, and is localized in the embryo sac and endothelium. The promoters of those genes may be useful to genetically engineer citrus species to avoid seed formation in fruits of commercial varieties.

Keywords:

Ovule, Seed, EST, Citrus, Clementine (*Citrus clementina*)

Introduction

Parthenocarpy, the development of fruit without seeds (seedless), is a desirable trait for many commercial fruits (Koltunow 1993). Natural parthenocarpy due to spontaneous mutations, that ultimately alter hormone synthesis and/or response, has been described for several species (Talon et al. 1992; Fos et al. 2000; Fos et al. 2001; Olimpieri et al. 2007). Parthenocarpy can also be induced by application of plant growth regulators in species like grape, tomato and eggplant (Schwabe and Mills, 1981). In addition, plants with parthenocarpic fruit capacity have been developed using different strategies, including traditional breeding. However, traditional breeding, is especially difficult in woody perennial plants due to their long juvenile period and high heterozygosity. Alternatively, direct molecular strategies have resulted in successful parthenocarpy, like the expression of chimeric genes (*DefH9-iaaH*) that promote ectopic auxin synthesis in placenta and ovules in different plant families (Mezzeti et al. 2004; Constantini et al. 2007).

In some cases, fruits of parthenocarpic cultivars may contain seeds due to different reasons. The presence of seeds in ready-to-harvest fruits is considered as a negative quality trait which results in a decrease in the commercial value of the crop, mostly for those with hard seeds such as Citrus (mainly orange, mandarines and grapefruit). In Citrus, several self-incompatible species produce parthenocarpic fruits in the absence of fertilization (Koltunow, 1993). In spite of that, a percentage of these fruits may develop seeds due to cross-pollination with compatible pollen from other citrus varieties due to insects (mainly honey bees). The growth of different varieties in the vicinity of the orchards makes it almost impossible to avoid spurious seed development. Therefore, a new strategy is necessary to obtain truly parthenocarpic fruits with a total absence of undesired seeds.

One way to prevent seed development in fruits with parthenocarpic capacity would be to express an enzymatic gene under the control of a seed specific promoter to cause cell disruption of seed tissues. As a result, the seeds would be destroyed or ablated as they develop inside the fruit. That strategy has already been used in different species, including citrus (Koltunow, 1993). For instance, promoters of the seed storage gene from soybean have been used to direct the expression of the RNase Barnase from *Bacillus amyloliquefaciens* (Hartley, 1988) and to cause seed death and decrease seed size in tobacco and *Arabidopsis* (Koltunow et al. 1998). However, some cell death was also detected in other tissues as the expression of the gene is not seed-specific. Other example is the transformation of embryogenic calluses from

Ponkan mandarin (*C. reticulata*) and regeneration of plants with a construct that express the Barnase driven by an anther tapetum-specific promoter (*TA29*) from tobacco (Li et al. 2002).

Therefore, the isolation of highly ovule- and seed-specific genes from the species to be modified is extremely important. The activity of their promoters has to fit several characteristics, such as tissue specificity and timing of expression. Tissue specificity would ensure that no other developmental or response processes would be affected by the chimeric gene. Furthermore, to effectively destroy or limit seed development, the activity of a promoter should be high at early stages of seed development, completely destroying the seed or significantly affecting its development, making them undetectable for the consumer. Finally, the use of promoter regions isolated from the same species (intragenic) to be genetically modified would be a desirable strategy, in comparison with the use of exogenous genetic elements (transgenic) (Rommens et al. 2007).

To avoid the presence of seeds in citrus fruits in a biotechnological manner, we have initiated a project to obtain ovule- and/or seed-specific genes of citrus by the generation and sequencing of cDNA libraries. The promoters of these genes could be used to direct the expression of a protein that functions to abort seed development, like Barnase (Hartley, 1988), or a gibberellin (GA) 2-oxidase that inactivates gibberellins (Singh et al. 2010). Alternatively, the promoter could direct the expression of a gene able to promote parthenocarpy like *iaaM* or *iaaH* (Rotino et al. 2005). These constructs would then be used to transform *Citrus* plants to generate fruits without seeds.

In this work we first characterized *Citrus clementina* cv. Clemenules seed development by light microscopy. RNA was isolated from the desired developmental stages and used to produce cDNA libraries and isolate clones. Several candidate genes have been tested for its ovule/seed specific expression. Two clones, *CcTT16*, a MADS-box homologous to *TRANSPARENT TESTA16*, and *CcTT7*, a flavonoid 3'-monooxygenase/oxygen binding *CYP75B1* homologous of *TRANSPARENT TESTA7*, were selected. qPCR and mRNA *in situ* hybridization proved that the expression of these genes was highly specific of ovule and seed.

Material and Methods

Plant material

Citrus clementina cv. Clemenules trees were located at the IVIA research station (www.ivia.es) in Moncada, Valencia. Flowers at anthesis were emasculated and pistils were immediately pollinated manually with pollen of Fortune (a hybrid between Clementine mandarin cv. Fino and Dancy mandarin), compatible with Clemenules. Fruits were harvested at 30, 45, 60, 75, 90, 120 and 180 days after pollination (dap) and seeds were isolated. Unfertilized ovules were also harvested at anthesis.

Fixation and inclusion in paraffin

Ovules and seeds were fixed o/n at 4°C in 5% p-formaldehyde in 0.2 M phosphate buffer pH 7.2. Samples were then dehydrated in a series of EtOH:H₂O solutions (15%, 30%, 50%, 70%, 96%, and 100%) 30 min each at 4°C, and o/n at 4°C in 100% EtOH. Next, samples were incubated in a series of solutions of HistoClear (National Diagnostics) and EtOH at 1:2, 1:1, 2:1 (v/v), one hour each at room temperature, followed by incubation o/n in 100% HistoClear. The next day, samples were incubated in HistoClear:paraffin at ratios 1:2, 1:1, and 2:1 (v/v), 4-5 h each at 58°C, and o/n in 100% paraffin. Paraffin was changed three times and the paraffin blocks with the samples were made.

Sections (7 µm thick) were generated and deposited onto glass slides covered with 3-aminopropyltriethoxysilane (TESPA) (Sigma). Slides were washed for 10 min in 100% HistoClear and sections were rehydrated through a series of EtOH:H₂O (100%, 70%, 50% 30%, and finally water), 10 min each. Sections were stained in 0.02% toluidine blue and observed under an optical microscope Eclipse E600 (Nikon).

RNA Extraction

For the construction of cDNA libraries, total RNA was extracted according to Bugos et al. (1995). Half-gram of samples was grinded in a mortar with liquid N₂. Powder was homogenized in 4 ml of extraction buffer (100 mM Tris-HCl pH 9.0, 200 mM NaCl, 15 mM EDTA pH 8.0, 0.5% sarkosyl, and 8 µl/ml β-mercaptoethanol), and phenol-extracted by adding 4 ml phenol, 0.8 ml chloroform:isoamyl alcohol (24:1), and 280 µl of 3 M sodium acetate pH 5.2. The mix was incubated 15 min on ice and centrifuged 10 min at 10,000 rpm. The supernatant was phenol-extracted with 4 ml of phenol:chloroform:isoamyl alcohol (25:24:1), and spun 10 min at 10,000 rpm. One volume of isopropanol was added to the supernatant, followed by incubation for 2 h on ice and centrifugation 10 min at 10,000 rpm. The pellet was

washed with 70% EtOH and resuspended in 1 ml of DEPC-treated H₂O. RNA was precipitated with 0.5 ml of LiCl 8M o/n at -20°C. RNA was collected by centrifugation and resuspended in 20 µl DEPC-treated H₂O. Poly(A⁺) RNA was isolated from total RNA using the Oligotex mRNA kit (Qiagen).

For quantitative PCR (qPCR) total RNA was isolated from leaves, internode, fruit pulp, roots, flower bud and pericarp using an RNeasy Plant Mini Kit (Qiagen), and RNA was treated with RNase-free DNase set (Qiagen), according to the manufacturer's instructions.

RNA was quantified using a Nanodrop ND-1000 spectrophotometer.

Construction of a full-length cDNA library and EST sequencing

To generate full-length cDNA library we used the method described previously in Marques et al. (2009), which uses the SMART™ PCR cDNA Synthesis Kit (Clontech) and a modified pENTR-1A vector (Marques et al. 2009). cDNA was generated with RNA from unfertilized ovules and seeds at different developmental stages (30, 45, 60, 75, and 90 dap).

Subtracted cDNA library was carried out using as a driver cDNA generated from an equal mixture of RNA from leaf, internode, pericarp, flower bud, flavedo and root. Tester cDNA was the same used for the full-length cDNA libraries. Normalization was done using the DSN (Duplex-Specific Nuclease) of the Trimmer cDNA normalization kit (Evrogen, Moscow) as previously described (Marques et al. 2009).

Clones were selected in LB-kanamycin media, and plasmid DNA was purified by alkaline lysis method, using the PerfectPrep (Eppendorf) kits. Sequencing was carried out from the 5' end of cDNA clones with the pENTR-F oligo (5'-GGCTTTAAAGGAACCAATTCA-3'), using an ABI 3100 capillary sequencer (Applied Biosystems) with fluorescent dye terminator technology. EST sequences were deposited in the HarvEST database (<http://harvest.ucr.edu>).

EST analysis, annotation and Gene Ontology classification, *in silico* expression analysis, and sequence comparison analysis

EST processing and assembly was performed by using EST2uni (<http://bioinf.comav.upv.es/est2uni>) (Forment et al. 2008). ESTs were compared against the EST database generated by the Spanish Citrus Functional Genomic Project (CFGP) (<http://bioinfo.ibmcp.upv.es/genomics/cfgpDB>) (Forment et al. 2005). The Gene Ontology

(GO) annotation of the *Arabidopsis* most similar protein was used for annotation of the new citrus unigenes. GO analysis was carried out at the Gene Ontology Consortium (www.geneontology.org/GO).

Functional enrichment analysis was done with the FatiGO software (www.fatigo.org) (Al-Shahrour et al. 2005) at the Babelomics web page (Al-Shahrour et al. 2006), by means of a Fisher's exact test. An *in silico* gene expression analysis (Meta-Profile Analysis Anatomy) of the new citrus unigenes was done with the Genevestigator *Arabidopsis* database (www.genevestigator.com), a reference expression database and meta-analysis system, using the *Arabidopsis* orthologs.

Sequence alignment was performed with the ClustalW2 software available at EMBL-EBI (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Phylogenetic approach was performed using Parsimony with the PHYLIP program (phylemon.bioinfo.cipf.es). An ordinary parsimony analysis was performed using the matrix given by the PROTPARS from the PHYLIP software. A majority rule and non rooted consensus tree was obtained.

Quantitative PCR (qPCR)

Total RNA (1.5 µg) was used for cDNA synthesis using TaqMan^R Reverse transcription Kit (Applied Biosystems) with oligo-dT primers, at a final volume of 20 µl. Each PCR was performed in triplicate with 10 µl of SYB^R Green Master Mix reagent (Applied Biosystems), 1 µl of cDNA solution and 300 nM of each primer, in a final volume of 20 µl, using an ABI PRISM^R 7000 Sequence detection System (Applied Biosystems). PCR amplification protocol was 2 min at 50°C, 10 min at 95°C, and 40 cycles of 15 s at 95°C and 1 min at 60°C. Primary data were analyzed using the SDS 2.0 software (Applied Biosystems). Gene-specific primers were designed using Primer ExpressTM v2.0 software (Applied Biosystems) (Supplemental Table 1). A template-free control was included for gene amplification. Post-amplification dissociation curves indicated that a single PCR product was amplified by each primer combination.

The citrus actin mRNA (*CX289161*) was amplified as housekeeping control gene for the analysis of expression in different tissues. C_T values of *CcTT7* and *CcTT16* were normalized to the C_T value of actin, and relative expression was calculated as described in Dorcey et al. (2009). Analysis of *CcTT16* and *CcTT7* expression during seed development was done using a calibration curve with a 10-fold serial dilution of the ovule RNA sample (Hashimoto et al.

2004). Total RNA amounts were accurately quantified and were used to normalize mRNA levels, as described by Alos et al (2006).

in situ mRNA hybridization

Localization of *CcTT16* and *CcTT7* mRNAs was determined by *in situ* hybridization as described by Jackson (1992). To generate the riboprobes, *CcTT16* and *CcTT7* cDNA clones were inserted in pBlueScript SK⁻ (Stratagene). Antisense probes and sense control probes were generated by the digestion with *KpnI* and *NotI*, respectively. Probes were labeled with digoxigenin and immuno-detected with an alkaline-phosphatase-conjugated anti-digoxigenin antibody. Alkaline phosphatase was detected by the 5-bromo-4-chloro-3-indolyl-phosphate/nitroblue tetrazolium procedure. A microscope (Diaphot-TMD, Nikon) was used for sample visualization under phase contrast.

Sequence analysis

Clones C37005H06 and C36004D04 were fully sequenced using internal primers, and complete sequences were deposited in the GeneBank database with accession numbers HQ634393 and HQ634392, respectively.

Alignments and phylogenetical analyses of the *CcTT7* gene were done using the flavonoid 3',5'-hydroxylases from *Gossypium hirsutum* (AAP31058.1), *Populus trichocarpa* (XP_002314004.1), *Cyclamen persicum* (ACX37698.1), *Vitis vinifera* (BAE47007.1), *Vinca major* (ACZ63205.1), *Catharanthus roseus* (CAA09850.1), *Camellia sinensis* (AAY23287.1), *Rhododendron x pulchrum* (BAH98132.1), *Viola x wittrockiana* (BAF93855.1), and *Arabidopsis thaliana AtTT7* (NP_196416.1). Alignments and phylogenetical analyses of the *CcTT16* gene was done with the MADS-box protein from *Vitis vinifera* (XP_002276139.1), *Populus trichocarpa* (XP_002318890.1), *Antirrhinum majus DEFH21* (Q8RVL4.1), *Oryza sativa* (AAY46447.1), *Petunia hybrida FBP24* (Q9ATE5.1), *Zea mays ZMM17* (NP_001105130.1), *Brassica napus* (ABY59774.1), *Sorghum bicolor* (XP_002453370.1), *Ginkgo biloba* (BAD93174.1), *Gnetum gnemon GGM13* (Q9XGJ4.1), *Ricinus communis* (XP_002513703.1), and *Arabidopsis thaliana AtTT16* (NP_974823.1).

Results

Characterization of seed development in *C. clementina*

Since the aim of this work was to isolate seed specific genes from *C. clementina*, we needed first to analyze seed development in that species to select the appropriate developmental stages for library construction. For this, ovaries at the anthesis stage were pollinated with pollen from the self-compatible cultivar Fortune. Fruits were harvested at different time points and weight of fruits and seeds were measured. Seed weight changed very little during the first month but increased rapidly afterwards, reaching a maximum at the end of fruit ripening, around 6 months after pollination (Fig. 1). Fruit growth followed a similar pattern, although its weight reached a maximum later than that for seeds (data not shown).

Histological analysis indicated that seed growth was slow from pollination to 45 dap (Fig. 2). At this stage, the endosperm had already filled the nucellar space. The embryo appeared later, and was only detectable at the globular stage at 60 dap, when seed weight had already increased 10-fold relative to ovule weight at anthesis. Afterwards, the embryo developed rapidly and went through the heart and torpedo stages. By 90 dap the cotyledons have enlarged and replace most of the endosperm within the seed. During the next three months, seed weight increased, the cotyledons accumulated reserve nutrient substances, and the mature seed was finally formed.

Isolation of Citrus ovule and seed genes

We generated two libraries enriched with full length cDNA clones, constructed from equivalent quantities of RNA from seeds at five different stages of development (between 30 and 90 dap), corresponding with early stages of embryo development. First we generated a pilot full-length cDNA library using a strategy developed in our laboratory, by combining the Clontech SMART™ PCR cDNA Synthesis Kit and directional cloning in the modified pENTR-1A plasmid (Marques et al. 2009). Out of 96 clones sequenced from this library, only 8 were not represented in the database of the CFGP (data not shown). To increase the proportion of new clones not previously isolated, we next generated a subtracted library enriched in seed specific cDNA (SeedClem library). We used the same cDNA as tester. Driver cDNA for the normalization/subtraction was generated from RNA from different tissues (roots, leaves, internodes, floral buds and pericarp of developing fruits). Removal of abundant and/or ubiquitous genes (presented in both driver and tester cDNA populations) was carried out using

the digestion with the nuclease DSN (Zhulidov et al. 2004; Marques et al. 2009). After normalization-subtraction, 64 clones were sequenced and we found that 12 of them were not present at the CFGP database, representing a 19% of novelty. As a result, we sequenced a total of 1226 clones, generating 1042 high quality ESTs, which represent 796 unigenes clustered in 249 singletons and 547 contigs when compared with the CFGP database (Table 1). From this, 287 (36%) corresponded to new genes not represented in the CFGP. The high percentage of novelty indicated that the new collection of ESTs generated complemented the CFGP database with potential seed-specific genes.

A new library, made from RNA from unfertilized ovules (OvuClem library) was generated using the same strategy as the SeedClem library (Table 1). Over a thousand clones were sequenced, generating 1014 ESTs, grouped in 758 genes (186 singletons and 572 contigs) representing 27% of novelty (203 unigenes) when compared against the CFGP database. In total, both libraries contributed 490 new genes not present previously in the CFGP database.

All genes isolated from the SeedClem and OvuClem libraries were functionally classified in comparison with the CFGP database, using the GO annotation of the *Arabidopsis* orthologs for each gene (Supplemental Figure 1). Overall, GO Biological Process categories of the three sets of genes show similar distributions. Next, the new citrus genes isolated were analyzed to search for enrichment of GO categories against the whole CFGP database. Several GO categories were enriched among ovule unigenes (Table 2 and Supplemental Table 2), while none of them were enriched among seed unigenes. GO categories significantly enriched in ovule genes were the Embryonic development, specifically Fruit (GO 0010154) and Seed development (GO 0048316); Modification-dependent proteolysis (GO 0019941); Amino acid and derivative metabolic process, in the Flavonoid biosynthetic pathway (0009813); and the Responses to osmotic (Salt) stress (0009651), Oxidative stress (0006979) and Response to the metal ion cadmium (GO 0046686) (Table 2).

To identify genes with potential seed- and ovule-specific expression, an *in silico* analysis was carried out using the *Arabidopsis* Genevestigator database (Zimmermann et al. 2004). The new citrus genes isolated in this work (490 unigenes) were first analyzed to identify the corresponding *Arabidopsis* orthologs; then the expression of these ortholog was analyzed with the meta-profile anatomy tool. Finally, we could retrieve data in the Genevestigator database from 90 and 55 genes from SeedClem and OvuClem libraries, respectively. Five genes from each ovule and seed libraries were selected based on their specific expression pattern in seedling, flower, silique and seed tissues of their *Arabidopsis* orthologs (Supplemental Table 3 and Supplemental Figures 2 and 3).

Expression of candidate genes in the ovule and during seed development

Clones selected by *in silico* analysis were analyzed for tissue specific expression by qPCR (Supplemental Figure 4). Seven tissues were analyzed: leaves, roots, internodes, flower bud, pericarp, pulp, and developing seeds. Two of the genes tested, C36004D04 isolated from the SeedClem library and C37005H06 from the OvuClem library, showed nearly specific expression in seeds (Figure 3). Expression levels of both genes in tissues other than seeds were extremely low or near the detection level. C36004D04 showed extremely low expression in roots and null expression in the other tissues, while C37005H06 showed low expression in flower buds and null expression in the other tissues.

Sequence analysis of C36004D04 and C37005H06 unigenes indicated that they corresponded to the *C. clementina* orthologs of the *Arabidopsis* *TRANSPARENT TESTA7* (*AtTT7 At5g07990*) and *TRANSPARENT TESTA16* (*AtTT16 At5g23260*) genes, respectively (Supplemental Figures 5 and 6). Protein sequences were 70% and 58% similar to their *Arabidopsis* orthologs and were named *CcTT7* and *CcTT16*. *AtTT16* belongs to a class of transparent testa proteins with the BSISTER (ABS) MADS domain (Nesi et al. 2002). *AtTT7* encodes the P450-dependent monooxygenase enzyme with flavonoid 3'-hydroxylase (F3'H) activity involved in flavonoid biosynthesis (Schoenbohm et al. 2000). Phylogenetic studies were performed using a maximum parsimony approach to confirm their similarity to different MADS-Box proteins and different flavonoid genes belonging to the P450 superfamily, respectively, from different plant species (Supplemental Figure 7).

Next, a time course expression analysis was performed for these new genes using six different stages of development: unfertilized ovules at anthesis and seeds at 30, 45, 60, 75 and 90 dap. Both *CcTT7* and *CcTT16* genes showed high expression levels in ovules and in seeds during early development, between 0 and 45 dap (Figure 4). The expression decreased afterwards to reach low levels towards 90 dpa.

mRNA *in situ* hybridization was used to determine the cellular/tissue expression of *CcTT7* and *CcTT16* (Figure 5). Both genes were expressed in the ovary at anthesis and seed at all stages studied with a similar pattern. In unfertilized ovaries, expression was located at the centre of the ovule, corresponding with the gametophyte. In 60-day-old seeds, expression was still localized in the embryo. Later, expression of both genes was restricted to the endothelium layer. Based on their expression pattern, the promoters of these two genes are good candidates to be used to specifically direct the expression of heterologous proteins in ovules and seeds

during their early development. The low expression detected in other tissues (roots and floral buds for *CcTT7* and *CcTT16*, respectively) may result in undesirable effects, which have to be carefully tested upon genetic transformation. Given juvenility in citrus, the effects on root tissue, if any, should be the first to be observed.

Discussion

Seed development of *C. clementina* followed a sigmoid pattern, similar to that of seeds of other Citrus cultivars (Koltunow et al. 1995). Our analysis allowed us to identify the stages of development appropriate for RNA isolation, i.e. when seeds contain low amount of reserve substances but have high contribution of tissues like the nucella, endosperm and embryo. Therefore, for generation of enriched cDNA libraries with the purpose of isolating ovule and/or young seed specific genes, in addition to unfertilized ovules we also selected seeds at relatively early stages of development, between 30 and 90 dap. The ovule and seed EST collections generated contained more than 2000 ESTs, which has enriched the CFGP ESTs database with new citrus genes isolated from two very relevant organs for plant development (ovules and seeds) not previously used. This information will contribute as novel resources for genomic studies and will be a useful tool for better understanding ovule and seed development in Citrus.

Out of the 490 new clones isolated, 10 genes were analyzed by qPCR, and two of them, *C37005H06 CcTT16* and *C36004D04 CcTT7*, corresponded to citrus orthologs of *AtTT16* and *AtTT7*, respectively, with highly specific expression in ovule and/or seed. *AtTT7* encodes a P450-dependent monooxygenase enzyme with flavonoid 3'-hydroxylase (F3'H) activity (Schoenbohm et al. 2000) involved in flavonoid biosynthesis. On the other hand, *AtTT16* encodes an *Arabidopsis* BSISTER (ABS) MADS domain protein (Nesi et al. 2002), involved in the accumulation of proanthocyanidin (PA), a particular subclass of flavonoids, in the endothelium of the seed coat and in the specification of endothelial cells. *AtTT16* is required to maintain normal transcription levels of *BANYULS* (*BAN*) and for normal endothelial cell shape and vacuolization, controlling the differentiation of all cells of the inner integument (Debeaujon et al. 2003; Nesi et al. 2002). Interestingly, it is not necessary for the differentiation of a small section of seed coat endothelium in the chalaza and micropyle ends of the seed, where other regulatory genes may participate. Therefore, the probable role of *CcTT16* is the specification of endothelial cell type in Clementine. The expression of *CcTT7* and *CcTT16* could suggest a role of these proteins in the distribution of auxin during seed development, as flavonoids have long been suspected to inhibit auxin transport (Peer et al. 2001). The isolation of *Arabidopsis*

mutants with altered seed coat color (*tt*, *transparent testa*; *ttg*, *transparent testa glabra*; and *ban*, *banyuls*) or altered PA levels has allowed the identification of genes required for the differentiation of the flavonoid-producing endothelial cells (Haughn et al. 2005). These proteins can be classified in two groups: a) proteins required for the biosynthesis and compartmentation of flavonoid compounds, which include TT3, TT4, TT5, TT6, TT7, BAN, TT12, TT19, TDS4/TT18 and AUTOINHIBITED H⁺-ATPase ISOFORM 10, and b) the transcription factors TT1, TT2, TT8, TT16, TTG1, and TTG2, which appear to have regulatory function.

Expression of *CcTT16* and *CcTT7* showed almost identical pattern, and it was very similar to the pattern described for their orthologues in *Arabidopsis*, *AtTT16* (Nesi et al. 2002) and *AtTT7* (Schoenbohm et al. 2000), respectively, and for other *transparent testa* genes, such as *AtTT12* (Debeaujon et al. 2001). Therefore, *CcTT16* and *CcTT7* are probably involved in the regulation of PA biosynthesis in *C. clementina* seed coats.

GO term analysis of genes isolated from ovule RNA showed enrichment of several processes relevant to the development of that organ. Interestingly, the terms Fruit and Seed development were enriched among genes isolated from ovules. This could be explained by the assignment of GO terms to genes. For instance, it has already been shown that several genes expressed during seed development have also significant expression in ovules prior to fertilization (Le et al. 2010). This points out to the need for a high quality GO term assignment to each gene. Other enriched term was the flavonoid biosynthetic process (GO 0009813), which includes the isolated genes *CcTT7* and *CcTT16*, (for *AtTT7 At5g07990* and *AtTT16 At5g23260*, respectively) (Table 2 and Supplemental Table 2).

To conclude, in addition of isolating 490 new citrus genes from ovules and young seeds of Clementine, thus enriching the CFGP ESTs database, in this work we report the isolation of two genes with almost specific expression in ovule/seed putatively involved in flavonoid biosynthesis (*CcTT7*) and its regulation (*CcTT16*). The isolation of their promoters should allow citrus transformation using constructions with those promoters and genes that code for enzymes capable of aborting the unfertilized ovules and young developing seeds. In any case, the possible undesirable effects in growth and development of roots or flower bud tissues that might have the very low levels of expression detected of *CcTT7* in roots and *CcTT16* in flower buds, respectively, have to be tested through genetic transformation.

Acknowledgement

The authors wish to thank Dr. L. Navarro and J. Juarez, from the Instituto Valenciano de Investigaciones Agrarias (Generalitat Valenciana), for the use of the Clementine orchard and Fortune pollen; Dr. J Forment and the Genomics Facility at the IBMCP for sequence analysis and maintenance of to the CFGP database; and Drs. F Tadeo and J Carbonell for critical reading of the manuscript. This work was supported by grants from the Conselleria de Agricultura, Pesca y Alimentacion (Generalitat Valenciana) and Spanish Ministerio de Ciencia y Tecnologia (research grant GEN2001-4885-C05). A. G-L. received a PhD fellowship from the Conselleria de Agricultura, Pesca y Alimentacion.

References

- Al-Shahrour F, Diaz-Uriarte R, Dopazo J (2005) Discovering molecular functions significantly related to phenotypes by combining gene expression data and biological information. *Bioinformatics* 21:2988-2993
- Al-Shahrour, F, Minguez P, Tarraga J, Montaner D, Alloza E, Vaquerizas JMM, Conde L, Blaschke C, Vera J, Dopazo J (2006) BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Res* 34:W472-W476
- Alos E, Cercos M, Rodrigo MJ, Zacarias L, Talon M (2006) Regulation of color break in citrus fruits. Changes in pigment profiling and gene expression induced by gibberellins and nitrate, two ripening retardants. *J Agric Food Chem* 54:4888-4895
- Bugos RC, Chiang VL, Zhang XH, Campbell ER, Podila GK, Campbell WH (1995) RNA isolation from plant tissues recalcitrant to extraction in guanidine. *Biotechniques* 19:734-737
- Costantini E, Landi L, Silvestroni O, Pandolfini T, Spena A, Mezzetti B (2007) Auxin synthesis-encoding transgene enhances grape fecundity. *Plant Physiol* 143:1689-1694
- Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, Caboche M, Lepiniec L (2003) Proanthocyanidin-accumulating cells in *Arabidopsis* testa: regulation of differentiation and role in seed development. *Plant Cell* 15:2514-2531

- Dorcey E, Urbez C, Blazquez MA, Carbonell J, Perez-Amador MA (2009) Fertilization-dependent auxin response in ovules triggers fruit development through the modulation of gibberellin metabolism in Arabidopsis. *Plant J* 58:318-332
- Debeaujon I, Peeters AJ, Léon-Kloosterziel KM, Koornneef M (2001) The TRANSPARENT TESTA12 gene of Arabidopsis encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell* 13:853-571
- Forment J, Gadea J, Huerta L, Abizanda L, Agusti J, Alamar S, Alos E, Andres F, Arribas R, Beltran JP, Berbel A, Blazquez MA, Brumos J, Canas LA, Cercos M, Colmenero-Flores JM, Conesa A, Estables B, Gandia M, Garcia-Martinez JL, Gimeno J, Gisbert A, Gomez G, Gonzalez-Candelas L, Granell A, Guerra J, Lafuente MT, Madueno F, Marcos JF, Marques MC, Martinez F, Martinez-Godoy MA, Miralles S, Moreno P, Navarro L, Pallas V, Perez-Amador, MA, Perez-Valle J, Pons C, Rodrigo I, Rodriguez PL, Royo C, Serrano R, Soler G, Tadeo F, Talon M, Terol J, Trenor M, Vaello L, Vicente O, Vidal C, Zacarias L, Conejero V (2005) Development of a citrus genome-wide EST collection and cDNA microarray as resources for genomic studies. *Plant Mol Biol* 57:375-391
- Forment J, Gilabert F, Robles A, Conejero V, Nuez F, Blanca JM (2008) EST2uni: an open, parallel tool for automated EST analysis and database creation, with a data mining web interface and microarray expression data integration. *BMC Bioinformatics* 9:5
- Fos M, Nuez F, Garcia-Martinez JL (2000) The gene *pat-2*, which induces natural parthenocarpy, alters the gibberellin content in unpollinated tomato ovaries. *Plant Physiol* 122:471-480
- Fos M, Proaño K, Nuez F, Garcia-Martinez JL (2001) Role of gibberellins in parthenocarpic fruit development induced by the genetic system *pat-3/pat-4* in tomato. *Physiol Plant* 111:545-550
- Hashimoto JG, Beadles-Bohling AS, Wiren KM (2004) Comparison of RiboGreen and 18S rRNA quantitation for normalizing real-time RT-PCR expression analysis. *Biotechniques* 36:54-60
- Haughn G, Chaudhury A (2005) Genetic analysis of seed coat development in Arabidopsis. *Trends Plant Sci* 10:472-477
- Hartley RW (1988) Barnase and barstar. Expression of its cloned inhibitor permits expression of a cloned ribonuclease. *J Mol Biol* 202:913-915

- Jackson DP (1992) *In situ* hybridization in plants. In: Bowles DJ, Gurr SJ, Pherenson M (eds). Molecular Plant Pathology: A Practical Approach, pp 163-174. Oxford University Press, Oxford, UK
- Koltunow AM (1993) Isolation and construction of genes to control seed production in Citrus. In: Hyashi T, Omura M, Scott NS (eds) Techniques on Gene Diagnosis and Breeding, pp. 101-108. Fruit Tree Research Station, Tsukuba, Japan
- Koltunow AM, Soltys K, Nito N, McClure S (1995) Anther, ovule, seed and nucellar embryo development in *Citrus sinensis* cv Valencia. Can J Bot 73:1567-1582
- Koltunow AM, Brennan P, Bond JE, Barker SJ (1998) Evaluation of genes to reduce seed size in Arabidopsis and tobacco and their application to Citrus. Mol Breeding 4:235-251
- Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamuro JK, Harada JJ, Goldberg RB (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proc Natl Acad Sci USA 107:8063-8070
- Li DD, Shi W, Deng XX (2002) Agrobacterium-mediated transformation of embryogenic calluses of Ponkan mandarin and the regeneration of plants containing the chimeric ribonuclease gene. Plant Cell Rep 21:153-156
- Marques MC, Alonso-Cantabrana H, Forment J, Arribas R, Alamar S, Conejero V, Perez-Amador MA (2009) A new set of ESTs and cDNA clones from full-length and normalized libraries for gene discovery and functional characterization in citrus. BMC Genomics 10:428
- Mezzetti B, Landi L, Pandolfini T, Spena A (2004) The DefH9-iaaM auxin-synthetisizing gene increases plant fecundity and fruit production in strawberry and raspberry. BMC Biotechnology 4:4
- Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M, Lepiniec L (2002) The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell 14:2463-2479
- Olimpieri I, Siligato F, Caccia R, Mariotti L, Ceccarelli N, Soressi GP, Mazzucato A (2007) Tomato fruit-set driven by pollination or by the parthenocarpic fruit allele are mediated by transcriptionally regulated gibberellin biosynthesis. Planta 226:877-888

- Peer AW, Brown DE, Tague BW, Muday GK, Taiz L, Murphy AS (2001) Flavonoid accumulation patterns of transparent testa mutants of *Arabidopsis*. *Plant Physiol* 126: 536-548
- Rommens CM, Haring MA, Swords K, Davies HV, Belknap WR (2007) The intragenic approach as a new extension to traditional plant breeding. *Trends Plant Sci* 12:397-403
- Rotino GL, Acciarri N, Sabatini E, Mennella G, Scalzo RL, Maestrelli A, Molesini B, Pandolfini T, Scalzo, J, Mezzetti B, Spena A (2005) Open field trial of genetically modified parthenocarpic tomato: seedlessness and fruit quality. *BMC Biotechnology* 5:32
- Schoenbohm C, Martens S, Eder C, Forkmann G, Weisshaar B (2000) Identification of the *Arabidopsis thaliana* flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme. *Biol Chem* 381:749-753
- Schwabe WW, Mills JJ (1981) Hormones and parthenocarpic fruit set: a literature survey. *Hort Abstracts* 51:661-698
- Singh DP, Filardo FF, Storey R, Jermakow AM, Yamaguchi S, Swain SM (2010) Overexpression of a gibberellin inactivation gene alters seed development, KNOX gene expression, and plant development in *Arabidopsis*. *Physiol Plant* 138:74-90
- Talon M, Zacarias L, Primo-Millo E (1992) Gibberellins and parthenocarpic ability in developing ovaries of seedless mandarins. *Plant Physiol* 99:1575-1581
- Zimmermann, P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR: *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol* 136:2621-2632
- Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA, Shagin DA (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res* 32:e37

Table 1. Characterization of citrus ovule and seed cDNA libraries

| Library | Clones | High-quality ESTs | Singletons | Contigs | Unigenes | Redundancy (%) | Library-specific unigenes | Novelty (%) |
|----------|--------|-------------------|------------|---------|----------|----------------|---------------------------|-------------|
| SeedClem | 1226 | 1042 | 249 | 547 | 796 | 24 | 287 | 36 |
| OvuClem | 1056 | 1014 | 186 | 572 | 758 | 25 | 203 | 27 |
| Total | 2282 | 2056 | 435 | 1027 | 1462 | 29 | 490 | 34 |

Table 2. Gene Ontology (GO) Biological Process categories significantly overrepresented in ESTs from the OvuClem library compared to the CFGP EST database

| GO name and accession | GO level | OvuClem (%) | CFGP (%) | Adjusted <i>P</i> -value ¹ |
|--|----------|-------------|----------|---------------------------------------|
| Response to metal ion (0010038) | 5 | 5.80 | 2.43 | 0.00162 |
| Response to cadmium ion (0046686) | 6 | 5.31 | 2.17 | 0.00264 |
| Embryonic development (0009790) | 3 | 4.64 | 2.14 | 0.01855 |
| Embryonic development ending in seed dormancy (0009793) | 4 | 4.48 | 1.88 | 0.01027 |
| Fruit development (0010154) | 5 | 5.14 | 2.26 | 0.00821 |
| Seed development (0048316) | 6 | 5.14 | 2.15 | 0.00379 |
| Proteolysis involved in cellular protein catabolic process (0051603) | 7 | 4.98 | 2.46 | 0.03976 |
| Modification-dependent protein catabolic process (0019941) | 8 | 4.98 | 2.42 | 0.02434 |
| Cellular amino acid and derivative metabolic process (0006519) | 4 | 7.63 | 3.60 | 0.00162 |
| Cellular amino acid derivative biosynthetic process (0042398) | 6 | 2.99 | 0.96 | 0.00769 |
| Cellular amino acid biosynthetic process (0008652) | 6 | 3.15 | 1.13 | 0.01093 |
| Phenylpropanoid metabolic process (0009698) | 6 | 2.32 | 0.83 | 0.04969 |
| Phenylpropanoid biosynthetic process (0009699) | 7 | 2.16 | 0.63 | 0.01806 |
| Flavonoid biosynthetic process (0009813) | 8 | 1.49 | 0.28 | 0.01093 |
| Response to osmotic stress (0006970) | 4 | 5.80 | 2.17 | 0.00078 |
| Response to salt stress (0009651) | 5 | 5.31 | 2.03 | 0.00158 |
| Response to oxidative stress (0006979) | 4 | 3.15 | 1.29 | 0.03985 |

¹, Adjusted *P*-value from Fisher's exact test corrected for multiple hypothesis testing (<0.05) are shown

Figure Legend

Fig. 1 Time course of seed (●) and fruit (○) fresh weight during development of *C. clementine* fruits. Seeds and fruits were harvested at different time-points after pollination at anthesis with pollen from the self-compatible Fortune clementine. Mean \pm SE are shown.

Fig. 2 Analysis of seed development in *C. clementina* at anthesis (A), and at 30 (B), 45 (C), 60 (D), 75 (E), 90 (F), 105 (G), 120 (H), and 180 (I) days after pollination with pollen from Fortune clementine. co, cotyledon; en, endosperm; es, embryo sac; m, micropyle; n, nucellus; ne, nucellar embryo; o, ovary; ov, ovule; vn, vascular network; ze, zygotic embryo. Scale bars are shown in each panel.

Fig. 3 Expression levels of genes *C36004D04* (*CcTT7*) and *C37005H06* (*CcTT16*) in different tissues of *C. clementine*, relative to actin. L, leaf; R, root, I, inflorescence; FB, flower bud, P, petiole; FP, fruit pulp; S, developing seeds (a mix of equal amount of RNA from seeds harvested at 30, 45, 60, 75 and 90 dpa). Three independent samples were assayed with similar results. Mean \pm SD of three technical replica are shown.

Fig. 4 Expression levels of genes *C37005H06* (*CcTT16*) and *C36004D04* (*CcTT7*) during seed development of *C. clementine*. Seeds were harvested at different time-points after pollination at anthesis with pollen from the self-compatible Fortune clementine. Three independent samples were assayed with similar results. Mean \pm SD of three technical replica for each biological replica are shown.

Fig. 5 Localization of the expression of genes *C37005H06* (*CcTT16*) (A to D) and *C36004D04* (*CcTT7*) (E to H) in unfertilized ovaries at anthesis (A and E) and during seed development (B, C, D, F, G, and H) of *C. clementine*. Seeds were analyzed at 60 (B and F), 75 (C and G) and 90 (D and H) dap with pollen from Fortune clementine. co, cotyledons; en, endothelium; es, embryo sac; ge, globular embryo; ov, ovule; ii, inner integument. Scale bars are shown in each panel.

Supplemental Figure 1 Functional categorization of genes isolated from the ovule (758 unigenes, 684 with *Arabidopsis* ortholog) and seed (796 unigenes, 707 with *Arabidopsis* ortholog) libraries compared with the CFGP database (34895 unigenes, 24780 with *Arabidopsis* ortholog). Annotation were classified based on the GO Biological Process categories. The percentage of genes in each GO are shown.

Supplemental Figure 2 *In silico* Meta-Profile analysis of ovule gene expression using Geneinvestigator. Reproductive tissues and seed are highlighted by horizontal red rectangles. Five genes were identified (asterisks). The *Arabidopsis* ortholog gene to the citrus C37005H06 *CcTT16* (At5g23260, microarray probe 249851_at) is highlighted by a vertical green rectangle. Orthologs for the other 4 selected genes (Supplemental Figure 4) are highlighted by a grey rectangle: At1g67820, probe 245194_at; At4g21190, probe 254427_at; At5g10510, probe 250426_at; and At4g30850, probe 253590_at).

Supplemental Figure 3 *In silico* Meta-Profile analysis of seed expression using Geneinvestigator. Reproductive tissues and seed are highlighted by horizontal red rectangles. Five genes were identified (asterisks). The *Arabidopsis* ortholog gene to the citrus C36004D04 *CcTT7* (At3g49240, microarray probe 252305_at) is highlighted by a vertical green rectangle. Orthologs for the other 4 selected genes (Supplemental Figure 4) are highlighted by a grey rectangle: At5g14050, probe 250222_at; At5g66540, probe 247046_at; At3g14690, probe 258094_at; and At5g07990, probe 250558_at).

Supplemental Figure 4 Expression levels of seed (C36) and ovule (C37) genes selected by the *in silico* meta-profile analysis using Geneinvestigator in different tissues of *C. clementine*. L, leaf; R, root, I, inflorescence; FB, flower bud, P, petiole; FP, fruit pulp; S, seed. Three independent samples were assayed with similar results. Mean \pm SD of three technical replica are shown.

Supplemental Figure 5 Sequence alignment of *C. clementine* C36004D04 *CcTT7* with its orthologs *Gossypium hirsutum* (AAP31058.1), *Populus trichocarpa* (XP_002314004.1), *Cyclamen persicum* (ACX37698.1), *Vitis vinifera* (BAE47007.1), *Vinca major* (ACZ63205.1), *Catharanthus roseus* (CAA09850.1), *Camellia sinensis* (AAY23287.1), *Rhododendron x pulchrum* (BAH98132.1), *Viola x wittrockiana* (BAF93855.1), and *Arabidopsis thaliana* AtTT7

(NP_196416.1). Alignment was carried out with the Clustalw2 software at EMBL-EBI (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>).

Supplemental Figure 6 Sequence alignment of *C. clementine* C37005H06 CcTT16 and its orthologs from *Populus trichocarpa* (XP_002318890.1), *Arabidopsis thaliana* AtTT16 (NP_974823.1), *Brassica napus* (ABY59774.1), *Ricinus communis* (XP_002513703.1), *Vitis vinifera* (XP_002276139.1), *Antirrhinum majus* DEFH21 (Q8RVL4.1), *Zea mays* ZMM17 (NP_001105130.1), *Sorghum bicolor* (XP_002453370.1), *Oryza sativa* (AAY46447.1), *Gnetum gnemon* GGM13 (Q9XGJ4.1), *Petunia hybrida* FBP24 (Q9ATE5.1), and *Ginkgo biloba* (BAD93174.1). Alignment was carried out with the Clustalw2 software at EMBL-EBI (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Relevant motifs and domains are indicated.

Supplemental Figure 7 Phylogenetic relationships of C36004D04 CcTT7 and related flavonoid 3',5'-hydroxylase proteins (A) and C37005H06 CcTT16 and related MADS domain proteins (B). The consensus most-parsimonious trees are shown.