Analysis of gene expression during the fruit set of tomato: 
A comparative approach

L. Pascual, J.M. Blanca, J. Cañizares 1, F. Nuez 1, * 

Instituto de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV), Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain

Received 14 May 2007; received in revised form 12 July 2007; accepted 17 July 2007
Available online 27 July 2007

Abstract

The study of tomato fruit set and early fruit development is of upmost importance due to their economic implications for crop production. Moreover, they are good subjects in the attempt to understand plant development control. In this work, we isolated more than 400 genes that are differentially expressed during these processes in tomato by using two complementary approaches: suppressive subtractive hybridization and genomic comparative analysis. We made a tomato flower library enriched in genes differentially expressed at 3 DPA when compared against anthesis. The library subtraction showed the high variability of genes and pathways implicated in this process. Using the AtGenExpress data from Arabidopsis, we detected 1879 genes differentially expressed during fruit set and early development. The expression of their orthologue tomato genes was tested by quantitative PCR, and more than 75% of the genes were also differentially expressed in tomato, meaning there is a conservation between these two species in spite of the significant differences in fruit morphology and development. We detected known and unknown fruit development genes and pathways, and this data will be a good source of information for future experiments about fruit set in tomato and Arabidopsis. We also showed that this transcriptomic comparative approach is very useful in identifying target genes of conserved processes in species where microarray facilities are not available, making the work with model species more profitable.

© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Tomato; Arabidopsis; Fruit set; SSH; Microarray; Fruit development

1. Introduction

The tomato (Solanum lycopersicon L.), like other crops, is grown to obtain its fruit, and its production is dependent upon fruit set and development among other factors. Fruit development is also a process of interest for the study of plant development regulation, and the tomato has been used as a model plant for climacteric fruit development [1]. Fruit development in tomato is divided in four phases: fruit set, cell division stage, cell expansion stage and ripening. In tomato, the fruit set can be defined as the restarting of cell division in the ovary after pollination or fecundation. The development of tomato fruit is independent of embryo development, and the linkage of these processes can be broken. Parthenocarpy, the production of fruits without seeds, is common in this species and can be caused by natural mutations, environmental factors or hormone treatments. The existence of a hormonal control in the fruit set is obvious and has been demonstrated by various studies reviewed in Ozga et al. [2] and Srivastava and Handa [3]. Gibberellins and auxins play a crucial role in this process in tomato, although it appears that other plant regulators might be implicated, such as cytokinins, ethylene and polyamines. The implication of these hormones has been shown by the measurement of endogenous levels in pollinated ovaries, in the unpollinated ovaries of parthenocarpic lines and by exogenous application. Natural and artificial mutants have demonstrated the existence of a genetic control in fruit set, but there is little known about this control. In tomato, there are several described mutations that produce parthenocarpic growth, such as pat, pat-2 and pat3/pat4 [4–9], but they have not been characterized at the molecular level. Several genes are also described as being implicated in fruit set; among others, Aux/IAA transcription factor IAA9 [10], TM29, a tomato SEPALLATA homologue [11] and the diageotropica gene [12].

---

* Corresponding author.
E-mail address: Fnuez@btc.upv.es (F. Nuez).

1 These authors have contributed equally to this work and should be regarded as last authors.
The panorama in other crop species is similar, and not much information is available.

As of this work, no global analysis of gene expression during fruit set has been published for tomato. Microarray or proteomic analyses have been used to analyze global expression during fruit development in tomato, but these analyses have not included the fruit set stage and are focused on the late stages of fruit development, from 7 DPA to ripening [13–15]. Digital northern is another approach for analyzing global expression, but the available tomato EST libraries do not include this stage of fruit development separately [13]. Therefore, more work is necessary in order to understand the genetic and hormonal control as well as the metabolic pathways of the fruit set in tomato and other crops, especially for the identification of the key genes and pathways. A suppressive subtractive hybridization (SSH) or SAGE that can be quick and useful approaches to create a library enriched in genes differentially expressed in any tissue, when there are no available EST libraries to perform a digital northern [16].

Arabidopsis is an excellent model for plant development and many data have been extrapolated to other species. The conservation of developmental and metabolic pathways, gene function and expression data between this species and other unrelated plants has been repeatedly reported. Arabidopsis genomic and transcriptomic data has been used as a tool for functional genomics in tomato [17]. Notwithstanding, these two species, with different types of fruits, seem too different to expect conservation. However, we are interested in fruit set, which may not be affected by fruit morphology. In fact, the Solanaceae family presents a high fruit variability which may not be affected by fruit morphology. In fact, the capsular and noncapsular Solanaceae fruits presents similarities with the Arabidopsis dehiscence process [19–21]. Good examples of this conservation are the genes implicated in carpel and ovule development or in ethylene signaling ([22,21] and references therein). In fact, comparative analyses of transcriptomic data between tomato, Arabidopsis and other species have already been done in late tomato fruit development and ripening, which shows the validity of this approach [13,23].

There are several microarray platforms for Arabidopsis that have been used extensively in different studies about development, stress, etc. These experiments have produced large amounts of data regarding gene expression that may be very useful in the study of developmental or metabolic pathways in this species. One of these projects, named AtGenExpress, consists in the hybridization of the ATH1 array from Affymetrix with a complete selection of developmental and stress samples, which includes the development of carpel and silique [24], and which has been used in several studies [25,26].

In this work, we isolated genes that are differentially expressed during the tomato fruit set by SSH, and we detected a high degree of correlation between the results of a suppressive subtractive hybridization of two stages of tomato fruit set and the AtGenExpress data. In order to increase the number of differentially expressed genes, we used the AtGenExpress data to detect genes regulated during Arabidopsis carpel and silique development, and we tested these genes in tomato by quantitative real-time PCR; in the end, more than 400 candidate genes were detected as being implicated in the fruit set of tomato.

2. Material and methods

2.1. Plant material

UC82 tomato plants were grown under greenhouse conditions (24 °C, 16 h L/D). Flowers were self-pollinated and the percentage of fruit set in the control plants was 90%. Flowers were collected at anthesis and each day until 7 days after anthesis (7 DPA), frozen in liquid nitrogen and stored at −80 °C to perform the library subtraction. In order to test the expression of the genes in the ovary, the flowers were collected and the ovaries extracted at 2 days before anthesis (pre-anthesis, when the stamens have reached their final size and turn from green to yellow) and at anthesis; flowers hand-pollinated at anthesis (percentage of fruit set 95%) were collected and the ovaries extracted at 1 DPA, and each day until 7 DPA and at 12 DPA, frozen in liquid nitrogen and stored at −80 °C.

2.2. Suppressive subtractive hybridization (SSH)

Total RNA from flowers at anthesis and 3 DPA was extracted as described by Bugos et al. [27]. Poly(A+) RNA was isolated using the mRNA Purification Kit (Amersham Pharmacia Biotech, Freiburg, Germany). The library subtraction was performed with the Clontech PCR-Select™ cDNA Subtraction Kit (BD Biosciences Clontech, Palo Alto, USA). The cDNAs were cloned with the InS/T/Aclone™ PCR Product Cloning Kit (Fermentas, Hanover, USA).

2.3. Sequence analysis

Clones were sequenced with the ABI PRISM 3100 Avant Genetic Analyzer (Applied Biosystems, Woolston, UK). Sequence trimming, clustering and assembly of the sequences were done with the Staden package (http://staden.sourceforge.net/staden_home.html). Sequences were submitted to the EMBL Nucleotide Sequence Database, and had been assigned the database accession numbers from AM495471 to AM495722. Similarity analysis was done by Blast [28] against the LeGI 11.0 tomato TIGR database. The orthologue analysis was done using in-house software. We did a tblastn comparison of the complete LeGI 11.0 tomato TIGR database against the TAIR6 Arabidopsis database as well as the reciprocal comparison. Genes were only labeled as orthologues when the two blasts gave reciprocal first hits [29] at an expect value of ≤E−15 and the next best match was of lower significance; there was a difference of ≥10 between expected scores. The GO annotation of the SSH library was done using the tomato TIGR database annotation.

2.4. RT and QPCR analysis

Total RNA was extracted with TRI Reagent (Sigma–Aldrich, Saint Louis, USA) following the manufacturer’s
instructions. The RNA was incubated with DNase I, RNase-free (Roche, Nonnenwald, Germany) to prevent DNA contamination; the absence of DNA was tested by quantitative PCR (QPCR) using actin primers and a genomic control. PCR was carried out in an ABI PRISM 7000 Sequence Detection System (Applied Biosystems) with Absolute SYBR Green ROX Mix (ABgene, Epsom, UK). First strand cDNA was synthesized from 1 μg of total RNA with oligo d(T) primer, using Expand Reverse Transcriptase (Roche). QPCR with specific primers was performed as mentioned and an actin control [30] was included in all plates to allow comparative analysis. The genes and control primers are presented as supplemental material (Supplemental Table S1). These primers were designed from the TIGR sequences with Primer Express 2.0 (Applied Biosystems) software. Three QPCR reactions from different RTs were done for each primer pair. To quantify each gene’s mRNA, we first calculated the amplification efficiency for each primer pair with DART-PCR 1.0 software [31] which was downloaded from http://www.gene-quantification.de/peirson-dart-version-1.pdf; second, relative gene expression levels were calculated using Liu and Saint’s algorithm [32], taking the actin as the reference gene; and finally, the average and standard error for the three samples were calculated.

2.5. Microarray analysis

ATH1 microarray gcRMA normalized Arabidopsis data from the AtGenExpress (Development) project [24] was downloaded from http://www.weigelworld.org/resources/microarray/AtGenExpress/. Three stages were chosen to search for differentially expressed genes: ATGE_37 (flower stage 12, carpels), ATGE_45 (flower stage 15, carpels) and ATGE_76 (silique stage 3). They were analyzed with the TIGR MeV version 3.1 package [33] which was downloaded from http://www.tm4.org/mev.html. To identify differentially expressed genes, we did a significance analysis of microarrays (SAM) [34] comparing the three samples two by two. The Δ of the SAMs were adjusted to a false discovery rate lower than 2% and a fold-change of 4 was forced. The analysis results were merged and the duplicate genes, as well as the probes that were found to be nonunique according to the ATH1 TAIR annotation, were eliminated. The cluster analysis of the genes with a tomato orthologue was done with SOTA [35] using the average expression of these stages, calculating the distances by Pearson correlation and using 7 clusters. Genes from each cluster were randomly selected to test their expression in tomato. To search for tomato orthologue genes whose expression levels remained constant throughout ATGE_37 (flower stage 12, carpels), ATGE_45 (flower stage 15, carpels), ATGE_76 (silique stage 3), ATGE_77 (silique stage 4) and ATGE_78 (silique stage 5), we performed a SAM analysis with five groups. From these constant genes we selected the ones that had a fold-change greater than 4 throughout the rest of the AtGenExpress experiment. The Arabidopsis GO analysis was done using the Babelomics suite [36,37] http://babelomics.bioinfo.cipf.es/.

3. Results

3.1. Suppressive subtractive hybridization of tomato flowers

The expression of the proliferating cell nuclear antigen (PCNA) was analyzed in flowers from anthesis to 7 DPA to determine when the cell cycle reactivation was detectable using QPCR, as that is the moment of the fruit set. PCNA was clearly activated 3 days after anthesis (3 DPA) (data not shown) in concordance with previous results [38,39]. This stage was therefore selected to develop a library enriched in genes differentially expressed at 3 DPA when we compared against anthesis. Two hundred fifty-two clones were sequenced and grouped in 181 contigs (Supplemental Table S2). The genes represented by contigs with two or more sequences were considered to be putative differentially expressed. There were 4 contigs with at least 5 sequences, 10 represented with 3 sequences and 25 with just 2 sequences. Table 1 shows all these contigs, as well as some genes only represented by one sequence whose expression was tested by QPCR. A blast against the LGI_091806 database from TIGR was done and the homologue TIGR tomato annotated genes were identified for most of the contigs, so we worked with the complete gene sequences from TIGR.

Subtraction effectiveness was tested by checking nine genes represented by different numbers of sequences by quantitative RT-PCR, using the starting RNA (Fig. 1A). We considered that a gene was differentially expressed when the expression at 3 DPA was at least 2-fold the anthesis expression. The gene represented by 10 sequences was confirmed as differentially expressed, 3 out of 4 genes with 3 sequences had also increased their expression, 1 out of 2 represented by 2 sequences and 1 out of 2 with only 1 sequence were also differentially expressed. This result confirms that the SSH was effective and that some of the genes represented by one sequence could be differentially expressed.

As it is impractical to check all the genes by QPCR, we analyzed the AtGenExpress data to confirm the rest of the genes obtained in the subtraction experiment. First, we detected the putative Arabidopsis orthologues using a reciprocal blast and labeling the tomato and the Arabidopsis genes as orthologues only when the first blast hit was the same in both directions. Thirty-two of the 181 tomato genes had a clear Arabidopsis orthologue. Second, we analyzed the expression of the orthologue Arabidopsis genes using the data from the AtGenExpress experiments ATGE_37 (flower stage 12, carpels), ATGE_45 (flower stage 15, carpels) and ATGE_76 (silique stage 3) which range from pre-anthesis to 4–5 DPA. Seven genes had a fold-change greater than 4 and were considered differentially expressed, which means a 22% of the genes with orthologue. This percentage is similar to the usual enrichment in differentially expressed genes in an SSH library. We tested these genes by QPCR between anthesis and 3 DPA with the RNA from flowers (Fig. 1B). Two primer pairs did not work properly, four genes had an expression at 3 DPA at least 4-fold greater than at anthesis, and the other one was between 2 and 4-fold greater at 3 DPA than at anthesis.
<table>
<thead>
<tr>
<th>Contig</th>
<th>No. of clones</th>
<th>Homologue TC</th>
<th>Arabidopsis orthologue</th>
<th>Blast annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_E11</td>
<td>10</td>
<td>TC180501*</td>
<td>UP(Q7/M222) LYCES (Q7/M222)</td>
<td>Probable proteinase inhibitor precursor, partial (92%)</td>
</tr>
<tr>
<td>D_H11</td>
<td>7</td>
<td>TC187752</td>
<td>Homologue to gb</td>
<td>AF036490.1 Acornus graminus large subunit</td>
</tr>
<tr>
<td>E_E11</td>
<td>6</td>
<td>AW040934</td>
<td>H3 histone H3R-11 (Oryza sativa) EGAD/131482/140407 histone</td>
<td></td>
</tr>
<tr>
<td>E_G04</td>
<td>5</td>
<td>TC170348</td>
<td>UP(IP22) LYCES (Q43710) Proteinase inhibitor type-2 TR8 precursor (Proteinase inhibitor type II TR8), complete</td>
<td></td>
</tr>
<tr>
<td>B_C11</td>
<td>3</td>
<td>TC187973</td>
<td>Homologue to UP(Q9ZR68) TOBAC (Q9ZR68) Aquaporin 1, partial (58%)</td>
<td></td>
</tr>
<tr>
<td>D_C05</td>
<td>3</td>
<td>TC170294</td>
<td>UP(METM) LYCES (P43282) S-adenosylmethionine synthetase 3 (S-adenosylmethionine synthetase 3)</td>
<td></td>
</tr>
<tr>
<td>C_G01</td>
<td>3</td>
<td>TC183252</td>
<td>UP(RL38) LYCES (P46291) 60S ribosomal protein L38, complete</td>
<td></td>
</tr>
<tr>
<td>D_H05</td>
<td>3</td>
<td>AA928470</td>
<td>Homologue to GB</td>
<td>DJQ19669.1 Solanum tuberosum ribosomal protein 6.1-like protein mRNA</td>
</tr>
<tr>
<td>B_H06</td>
<td>3</td>
<td>TC170179</td>
<td>Similar to UP(RL18) ARATH (P42791) 60S ribosomal protein L18, complete</td>
<td></td>
</tr>
<tr>
<td>E_G02</td>
<td>3</td>
<td>TC189949</td>
<td>Similar to UP(Q5W1I1) NIGGL (P42791) Histone</td>
<td></td>
</tr>
<tr>
<td>A_F02</td>
<td>3</td>
<td>TC170549*</td>
<td>UP(SYST) LYCES (Q27058) Systemin precursor, complete</td>
<td></td>
</tr>
<tr>
<td>A_F12</td>
<td>3</td>
<td>TC175253*</td>
<td>Homologue to UP(Q2VA97) SOLTU (Q2V9A7) Cyclolin ROC7-like, partial (97%)</td>
<td></td>
</tr>
<tr>
<td>B_E01</td>
<td>3</td>
<td>TC183105*</td>
<td>Homologue to UP(Q5TE3) TOBAC (Q45TE3) Senescence-associated protein, partial (34%)</td>
<td></td>
</tr>
<tr>
<td>A_A06</td>
<td>3</td>
<td>TC185802*</td>
<td>Similar to UP(Q38806) ARATH (Q38806) AHSP22.0 precursor (Heat shock protein 22.0) (T4A4.7 protein) (A4G10250), partial (51%)</td>
<td></td>
</tr>
<tr>
<td>D_B08</td>
<td>2</td>
<td>TC172422</td>
<td>Weakly similar to UP(Q9SW12) ARATH (Q9SW12) Acid phosphatase-like protein, partial (41%)</td>
<td></td>
</tr>
<tr>
<td>D_C06</td>
<td>2</td>
<td>TC189707</td>
<td>Homologue to UP(Q9XE59) SOLTU (Q9XE59) Phosphoglycerate mutase, partial (17%)</td>
<td></td>
</tr>
<tr>
<td>D_B10</td>
<td>2</td>
<td>TC175597</td>
<td>Similar to UP(Q8L934) ARATH (Q8L934) Nucleoid RNA-binding-like protein, partial (47%)</td>
<td></td>
</tr>
<tr>
<td>D_G11</td>
<td>2</td>
<td>TC189813</td>
<td>Similar to UP(Q2HE6) MEDTR (Q2HE6) Zinc finger, C2H2-type, partial (12%)</td>
<td></td>
</tr>
<tr>
<td>E_D09</td>
<td>2</td>
<td>BG129307</td>
<td>Extensin-like protein [Solanum tuberosum] PIR</td>
<td>T07176/T07176 extensin homologue – potato (fragment)</td>
</tr>
<tr>
<td>E_F04</td>
<td>2</td>
<td>TC176996</td>
<td>Homologue to UP(B2) DAUCA (P37707) B2 protein, partial (71%)</td>
<td></td>
</tr>
<tr>
<td>A_A02</td>
<td>2</td>
<td>TC169901</td>
<td>UP(Q5VL8) LYCES (Q5VL8) Beta-mannosidase enzyme, complete</td>
<td></td>
</tr>
<tr>
<td>A_E01</td>
<td>2</td>
<td>TC174179</td>
<td>Similar to UP(Q6YC3) VITV1 (Q6YC3) Wound-induced protein-like (Fragment), partial (31%)</td>
<td></td>
</tr>
<tr>
<td>E_C08</td>
<td>2</td>
<td>TC177353</td>
<td>Similar to UP(Q4G86) OLEEU (Q4G86) Beta-1,3-glucanase-like protein, partial (32%)</td>
<td></td>
</tr>
<tr>
<td>E_D05</td>
<td>2</td>
<td>TC172590</td>
<td>Similar to RF</td>
<td>NP_196542.1</td>
</tr>
<tr>
<td>A_E10</td>
<td>2</td>
<td>TC177340</td>
<td>Similar to UP(Q9VW5) DROME (Q9VW5) CG14687-PA (RE74890p), partial (12%)</td>
<td></td>
</tr>
<tr>
<td>B_B12</td>
<td>2</td>
<td>TC171658</td>
<td>Homologue to UP(Q307Z5) SOLTU (Q307Z5) Deacetylase-like protein, complete</td>
<td></td>
</tr>
<tr>
<td>B_C04</td>
<td>2</td>
<td>TC179011</td>
<td>Similar to RF</td>
<td>NP_195169.3</td>
</tr>
<tr>
<td>A_E02</td>
<td>2</td>
<td>TC170474</td>
<td>Homologue to UP(Q308A0) SOLTU (Q308A0) P40-like protein, partial (89%)</td>
<td></td>
</tr>
<tr>
<td>A_G07</td>
<td>2</td>
<td>TC169927</td>
<td>UP(Q8GT6) LYCES (Q8GT6) Xyloglucan-specific fungal endoglucanase inhibitor protein, complete</td>
<td></td>
</tr>
<tr>
<td>B_A11</td>
<td>2</td>
<td>TC170047</td>
<td>Homologue to UP(Q8GT7) SOLTU (Q8GT7) Tuber-induction protein (Fragment), partial (9%)</td>
<td></td>
</tr>
<tr>
<td>B_B06</td>
<td>2</td>
<td>TC170987</td>
<td>Similar to UP(Q94BU8) ARATH (Q94BU8) AT5g28050/ F15F15_120, partial (97%)</td>
<td></td>
</tr>
<tr>
<td>B_F05</td>
<td>2</td>
<td>BI210647</td>
<td>Similar to GP</td>
<td>55790949</td>
</tr>
<tr>
<td>B_H08</td>
<td>2</td>
<td>TC174895</td>
<td>Homologue to UP</td>
<td>Q33W76</td>
</tr>
<tr>
<td>D_A06</td>
<td>2</td>
<td>TC184407</td>
<td>Similar to UP(Q2PV0) 9MAGN (Q2PV0) Soluble inorganic pyrophosphatase, partial (82%)</td>
<td></td>
</tr>
<tr>
<td>D_A09</td>
<td>2</td>
<td>TC17Contigs tested by Q7701</td>
<td>Similar to UP</td>
<td>Q40380</td>
</tr>
</tbody>
</table>
To look for any biological processes over- or under-represented in our SSH, the percentages of each Gene Ontology (GO) term in the complete TIGR tomato database and in the genes picked by our SSH were calculated according to the TIGR annotation (Fig. 2). We detected a general enrichment in the library of genes related to protein biosynthesis, histones, nucleosome assembly, chromatin assembly or disassembly, chromosome organization, defense response, response to heat, response to unfolded protein, cell cycle, photosynthesis, spermatogenesis and ageing. But we also detected a decrease

<table>
<thead>
<tr>
<th>Contig</th>
<th>No. of clones</th>
<th>Homologue TC</th>
<th>Arabidopsis orthologue</th>
<th>Blast annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_F11</td>
<td>2</td>
<td>DV105017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_F11</td>
<td>2</td>
<td>TC172378</td>
<td>Homologue to UP:H2B_GOSHI (O22582) Histone H2B, partial (96%)</td>
<td></td>
</tr>
<tr>
<td>B_G11</td>
<td>2</td>
<td>TC189580</td>
<td>AT5G20720</td>
<td></td>
</tr>
<tr>
<td>F_E10</td>
<td>2</td>
<td>TC183832</td>
<td>UP</td>
<td>Q9M5A8_LYCES (Q9M5A8) Chaperonin 21 precursor, complete gb</td>
</tr>
<tr>
<td>B_G10</td>
<td>2</td>
<td>TC180002*</td>
<td>AT5G59845</td>
<td></td>
</tr>
<tr>
<td>F_B09</td>
<td>2</td>
<td>TC172546*</td>
<td>Homologue to UP</td>
<td>Q48Z4_SOLTU (Q48Z4) Snakin-1, complete</td>
</tr>
<tr>
<td>E_D04</td>
<td>1</td>
<td>TC180398*</td>
<td>Weakly similar to GB</td>
<td>AAC64192.1</td>
</tr>
<tr>
<td>F_C11</td>
<td>1</td>
<td>TC176776*</td>
<td>Weakly similar to GB</td>
<td>AAC63976.1</td>
</tr>
<tr>
<td>A_D05</td>
<td>1</td>
<td>TC177661*</td>
<td>AT2G34700</td>
<td></td>
</tr>
<tr>
<td>C_C02</td>
<td>1</td>
<td>TC171919*</td>
<td>AT4G35770</td>
<td></td>
</tr>
<tr>
<td>E_E05</td>
<td>1</td>
<td>TC170845*</td>
<td>AT5G06860</td>
<td></td>
</tr>
<tr>
<td>F_G01</td>
<td>1</td>
<td>TC174950*</td>
<td>AT1G18250</td>
<td></td>
</tr>
<tr>
<td>D_C11</td>
<td>1</td>
<td>BP892749*</td>
<td>AT1G54050</td>
<td></td>
</tr>
</tbody>
</table>

*a Contigs tested by QPCR.
in others, such as transport, regulation of transcription DNA-dependent, intracellular protein transport, electron transport, protein transport and small GTPase-mediated signal transduction.

3.2. Differential analysis of the Arabidopsis microarray data

Taking into account the good results obtained in the AtGenExpress data analysis, we searched for differentially expressed Arabidopsis genes and detected their tomato orthologues to increase the number of potentially regulated genes during fruit set. Moreover, we evaluated the correlation between the expression during tomato and Arabidopsis fruit set.

We used the data from the Arabidopsis AtGenExpress project, a wide analysis of Arabidopsis expression using the Affymetrix ATH1 arrays, to find genes differentially expressed in the Arabidopsis fruit set. We studied the mRNA carpel and silique hybridizations of three developmental stages around fruit set: stage 12 (pre-anthesis, Atge_37), stage 15 (feciundation, Atge_45) of carpels and stage 3 (early fruit, Atge_76) of siliques. The original data have triplicates of each hybridization, and the global data had been previously normalized [24].

To detect differentially expressed genes, we did a SAM comparing the three stages two by two using a fold-change equal to or greater than 4. The results of these analyses are shown in Table 2. In the stage 37 vs. stage 45 SAM, 2% of upregulated genes and 1% of downregulated genes were detected. In the stage 45 vs. stage 76 analysis we detected 3% of upregulated genes and 2% of downregulated ones, and the stage 37 vs. stage 76 analysis showed 5% of upregulated genes and 3% of downregulated genes. We merged the data and eliminated duplicates and the nonunique probes according to the ATH1 TAIR annotation, and found 8.3% to be differentially expressed genes.

The next step was the assignment of putative tomato orthologues. The analysis was done with the TAIR6 database from the Arabidopsis genomic project and the tomato LeGI database version 11.0 of TIGR. We assigned orthologues to 7216 array genes, 32% of the total microarray, and we detected 450 differential Arabidopsis genes with a putative tomato orthologue, 24% of the 1879 differentially expressed genes of Arabidopsis (Supplemental Table S3). In order to determine if these genes were also differentially expressed during tomato fruit set, we tested randomly selected genes by QPCR during tomato fruit set. To include genes with different expression patterns in Arabidopsis, we made a clustering of the 450 genes using the SOTA algorithm. Seven clusters were defined (Fig. 3): clusters 1 and 2 formed by genes whose expression increased from carpel stage 12 to silique stage 3; clusters 3, 4 and 7, which were formed by genes whose expression decreased from carpel

Table 2

<table>
<thead>
<tr>
<th>Analysis</th>
<th>SAM 1</th>
<th>SAM 2</th>
<th>SAM 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>Atge_37</td>
<td>Atge_45</td>
<td>Atge_37</td>
</tr>
<tr>
<td>Group B</td>
<td>Atge_45</td>
<td>Atge_76</td>
<td>Atge_76</td>
</tr>
<tr>
<td>Positive significant genes</td>
<td>398</td>
<td>632</td>
<td>1180</td>
</tr>
<tr>
<td>Negative significant genes</td>
<td>116</td>
<td>406</td>
<td>645</td>
</tr>
<tr>
<td>No-significant genes</td>
<td>22296</td>
<td>21769</td>
<td>20985</td>
</tr>
<tr>
<td>Fold-change</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Δ</td>
<td>3</td>
<td>1.99</td>
<td>1.55</td>
</tr>
<tr>
<td>Number of false significant genes (90%)</td>
<td>1.59</td>
<td>0.98</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Fig. 2. Percentage of the 16 most-represented Gene Ontology terms referring to biological processes in the tomato TIGR database and in the SSH library.
stage 12 to siliques stage 3; and clusters 5 and 6, whose genes had a peak at carpel stage 15. Around 10% of the genes from each cluster were randomly selected for further analysis.

3.3. Analysis of Arabidopsis differentially expressed genes in tomato

We tested the expression of these genes in tomato ovaries at pre-anthesis, anthesis each day until 7 DPA and at 12 DPA. From the genes selected, we tested only the 37 whose primer pairs amplified a unique product detected by dissociation protocol and with a detectable level of expression. According to the criteria employed in the microarray analysis, we considered the genes to be differentially expressed in tomato when the rate between the stage with the higher value and the stage with the lower was equal to or higher than 4. In Fig. 4, we present all the expression values classified by the Arabidopsis clustering and normalized by the lower expression value in each gene. In cluster 1, 6 out of 9 genes were differentially expressed; in cluster 2, 10 out of 12; in cluster 7, 3 out of 6; and in the rest of the clusters all genes tested were differentially expressed. To sum up, we detected 29 differentially expressed genes out of 37 (78% changing genes) with different patterns some over-expressed in pre-anthesis or anthesis and others activated at the last stages or with a peak in the middle stages. However, we did not find a clear relationship between the clusters built with the Arabidopsis data and the expression pattern in tomato, neither by making a clustering nor a PCA analysis with the tomato expression (data not shown).

To assess the efficiency of our method, we studied the correlation between tomato and Arabidopsis when working with constant genes. Firstly, we analyzed the expression of genes that could be considered to be non-differentially expressed throughout the Arabidopsis array hybridizations ATGE_37, ATGE_45, ATGE_76, ATGE_77 and ATGE_78 (carpels in stage 12, stage 15 and siliques in stage 3, stage 4 and stage 5). We chose more hybridizations to avoid the problem of stage correlation. Those constant genes were selected among the genes with a tomato orthologue by SAM. We selected genes with a fold-change lower than 0.5 between these stages, and greater than 4 between other experiments analyzed in the AtGenExpress development project. From the 704 constant genes with an orthologue, we selected 8 of them randomly. We tested those genes as above (Fig. 4). Six genes were also constant in tomato and two showed a significant difference of expression. The percentage of conservation was 75%, similar to that in the differentially expressed gene analysis.

3.4. Gene Ontology analysis

We also analyzed the Gene Ontology (GO) terms with the FatiGO program from the Babelomics suite. We represented the percentage of the GO terms higher than 1% in all the Arabidopsis genes, the differentially expressed genes, all the genes with a tomato orthologue and the differentially expressed genes with orthologues, and we did not detect any clear change in the percentages of all the genes and orthologues (Fig. 5). To identify which biological processes are implicated in fruit set, we extracted the GO terms that were over- or under-represented in our sets of genes with the FatiGO program. This program extracts the relevant GO terms for a group of genes with respect to a reference group by the
Fig. 4. Expression patterns of the genes tested in tomato ovaries by QPCR, normalized to actin and divided by the value of the lowest expression in each gene. Genes are classified according to the Arabidopsis clusters and constant genes. Expression values up to 40 are indicated.
application of a Fisher’s exact test that considers the multiple-
testing nature of the statistical contrast performed (Table 3).
Firstly, we compared the GO terms of all genes differentially
expressed during fruit set in Arabidopsis, according to our
SAM analysis, against the rest of the ATH1 array genes
(Supplemental Table S4). We found that the response to
different stimuli (gibberellic acid, auxin, jasmonic acid,
saliclyc acid and abscisic acid), response to heat, response
to desiccation, cell wall modification, regulation of transcription
and cell proliferation terms were over-represented in the
differentially expressed genes. In addition, ageing that had
been detected with the subtraction GO term analysis was also
significantly different in Arabidopsis. Secondly, we compared
the GO terms of all Arabidopsis genes differentially expressed
with tomato orthologues against the rest of the ATH1 array
genes with orthologues (Supplemental Table S5). The GO
terms obtained were redundant with the GO terms obtained in
the whole ATH1 array gene analysis, or clearly related to them.

4. Discussion

4.1. Suppressive subtractive hybridization of tomato
flowers

In this work, we identified more than 400 genes
differentially expressed during the fruit set in tomato, using
two complementary approaches: a suppressive subtractive
hybridization and a comparative genomic analysis. The
suppressive subtractive hybridization library (SSH) showed
the high variability of genes and pathways implicated in this
process. There was no redundancy in the sequences, except
some high-expressed transcripts, and several unique sequences
tested were actually differentially expressed between anthesis
and 3 DPA. Obviously, it is not possible with our number of
sequences to select the genes using only the sequence number
as criteria. To avoid this problem, we analyzed to see if they
were also regulated during Arabidopsis fruit set. Thirty-two
putative Arabidopsis orthologue genes of the SSH were
detected, and seven of them presented a fold-change greater
than 4 between carpel stages 12 and 15 and silique stage 3.
These results confirmed that genes represented by only one
sequence in our SSH library may be differentially expressed
during these two stages and that the library is a good source of
genes implicated in tomato fruit set.

4.2. Differential analysis of the Arabidopsis microarray
data

We detected 7 out of 32 Arabidopsis orthologue genes of the
SSH library that were also differentially expressed according
to the AtGenExpress data. This percentage was close to the SSH
enrichment, so we used the Arabidopsis data to find more genes
regulated in tomato during this process. The stages between
which we found more variable genes were the pre-anthesis
carpel (stage 12) and silique (stage 3) and we detected more
upregulated than downregulated genes throughout the process.
To study the expression of these genes in tomato, we did an
orthologue assignation, which was the limiting step as only 450
putative tomato orthologues were unambiguously assigned.

4.3. Comparative analysis between Arabidopsis and
tomato

To test the conservation between the two species, we
analyzed the expression of tomato orthologues during fruit set
and early fruit development (pre-anthesis to 12 DPA) by QPCR.
Thirty-seven genes were tested in tomato and 29 genes were
differentially expressed in tomato at these stages, representing
78%. The same analysis with no differentially expressed genes
gave the same results, as 75% of these genes were also constant
in tomato. This result proved that our analysis was valid and is
useful in detecting variable and constant genes in this process; it
Arabidopsis expressed genes with orthologues and the rest of Arabidopsis the mid-globular and early heart stages are reached in temporal patterns of early fruit development. As an example, different places due to the differences in the morphology and be implicated, but that they are used at different times and in tomato, the same metabolic or developmental pathways could show a great increase at 12 DPA, which shows their implication in the next phase of fruit development.

What is more, all clusters had upregulated and downregulated genes in tomato in similar percentages, close to a random distribution. This lack of correlation between the patterns may be due to the difference in the number of stages tested in Arabidopsis and tomato and to the trouble in assigning stage equivalents. Furthermore, we were analyzing a complex developmental process that includes different tissues, whose development may have a different temporal pattern, and which may also be independent, such as embryo and pericarp development. The results suggest that in Arabidopsis and tomato, the same metabolic or developmental pathways could be implicated, but that they are used at different times and in different places due to the differences in the morphology and temporal patterns of early fruit development. As an example, the mid-globular and early heart stages are reached in Arabidopsis at stage 3, 4–5 DPA [40], while in tomato the embryo does not reach these stages until 12–15 DPA, at the cell expansion stage [41,42].

The orthologue search was the bottleneck of this analysis. We had identified 1879 differentially expressed genes in Arabidopsis, but detected only 450 orthologues in tomato. The method used, which is based on the reciprocity of the first blast hits, rejects many genes which belong to a determinate family or are duplicated in the genome. We used this strict rule to find the most approximate valuation of the conservation between these two species without the bias caused by mistakes in orthologue assignation and because other approaches, such as phylogenetic analysis, were not viable in a global analysis due to a lack of public data. However, it may be possible to identify more orthologue genes in a less restrictive analysis focused on specific genes by using phylogenetic analysis or working with all possible detected orthologues or taking into account the expression data of the orthologues [43]. Furthermore, the number of differentially expressed genes can be increased using the usual criteria, a fold-change equal to 2, instead of the strict criteria we followed. A fold-change equal to or greater than 4 in Arabidopsis and tomato and to the trouble in assigning stage equivalents. However, it may be possible to identify more orthologue genes in a less restrictive analysis focused on specific genes by using phylogenetic analysis or working with all possible detected orthologues or taking into account the expression data of the orthologues [43]. Furthermore, the number of differentially expressed genes can be increased using the usual criteria, a fold-change equal to 2, instead of the strict criteria we followed. A fold-change equal to or greater than 4 in Arabidopsis, and also in tomato was used to avoid mistakes when evaluating the correlation between them.

In spite of the lack of a clear correlation in the expression pattern, we detected more than 400 genes differentially expressed in tomato fruit set and early development. These

<table>
<thead>
<tr>
<th>Biological process</th>
<th>ATH1 array (%)</th>
<th>ATH1 differentially expressed (%)</th>
<th>FDR p-value</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response to heat</td>
<td>0.38</td>
<td>2.22</td>
<td>4.16E-006</td>
<td>4</td>
</tr>
<tr>
<td>Cell wall modification</td>
<td>1.19</td>
<td>4.71</td>
<td>2.38E-005</td>
<td>7</td>
</tr>
<tr>
<td>Regulation of transcription, DNA-dependent</td>
<td>18.35</td>
<td>29.74</td>
<td>3.75E-004</td>
<td>8</td>
</tr>
<tr>
<td>Response to gibberellic acid stimulus</td>
<td>0.77</td>
<td>2.44</td>
<td>1.76E-003</td>
<td>5</td>
</tr>
<tr>
<td>Response to auxin stimulus</td>
<td>1.66</td>
<td>3.84</td>
<td>2.95E-003</td>
<td>5</td>
</tr>
<tr>
<td>Response to jasmonic acid stimulus</td>
<td>0.84</td>
<td>2.44</td>
<td>3.11E-003</td>
<td>4</td>
</tr>
<tr>
<td>Response to desicion</td>
<td>0.04</td>
<td>0.70</td>
<td>3.39E-003</td>
<td>5</td>
</tr>
<tr>
<td>Ribosome biogenesis</td>
<td>2.03</td>
<td>0.00</td>
<td>4.80E-003</td>
<td>7</td>
</tr>
<tr>
<td>Macromolecule catabolism</td>
<td>3.35</td>
<td>1.16</td>
<td>8.74E-003</td>
<td>5</td>
</tr>
<tr>
<td>Cell proliferation</td>
<td>0.09</td>
<td>0.78</td>
<td>9.00E-003</td>
<td>4</td>
</tr>
<tr>
<td>RNA metabolism</td>
<td>3.46</td>
<td>1.05</td>
<td>1.04E-002</td>
<td>6</td>
</tr>
<tr>
<td>Lipid transport</td>
<td>0.78</td>
<td>2.21</td>
<td>1.09E-002</td>
<td>5</td>
</tr>
<tr>
<td>Intracellular transport</td>
<td>4.42</td>
<td>1.98</td>
<td>1.35E-002</td>
<td>5</td>
</tr>
<tr>
<td>Stomatal complex morphogenesis</td>
<td>0.01</td>
<td>0.47</td>
<td>1.93E-002</td>
<td>5</td>
</tr>
<tr>
<td>Coenzyme metabolism</td>
<td>1.64</td>
<td>0.15</td>
<td>2.58E-002</td>
<td>6</td>
</tr>
<tr>
<td>Ageing</td>
<td>0.41</td>
<td>1.37</td>
<td>2.61E-002</td>
<td>3</td>
</tr>
<tr>
<td>Microsporogenesis</td>
<td>0.09</td>
<td>0.98</td>
<td>2.98E-002</td>
<td>7</td>
</tr>
<tr>
<td>Anthocyanin metabolism</td>
<td>0.24</td>
<td>3.92</td>
<td>3.33E-002</td>
<td>9</td>
</tr>
<tr>
<td>Response to salicylic acid stimulus</td>
<td>0.78</td>
<td>2.00</td>
<td>3.81E-002</td>
<td>4</td>
</tr>
<tr>
<td>Translation</td>
<td>2.34</td>
<td>0.39</td>
<td>4.48E-002</td>
<td>7</td>
</tr>
<tr>
<td>Response to abscisic acid stimulus</td>
<td>0.81</td>
<td>2.10</td>
<td>4.17E-002</td>
<td>5</td>
</tr>
</tbody>
</table>

**Table 3**

Comparison of Gene Ontology terms referring to biological processes with FatiGO+

<table>
<thead>
<tr>
<th>Biological process</th>
<th>Orthologues (%)</th>
<th>Orthologues differentially expressed (%)</th>
<th>FDR p-value</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response to temperature stimulus</td>
<td>0.90</td>
<td>5.29</td>
<td>4.97E−003</td>
<td>4</td>
</tr>
<tr>
<td>Regulation of transcription, DNA-dependent</td>
<td>15.22</td>
<td>35.36</td>
<td>6.98E−003</td>
<td>8</td>
</tr>
<tr>
<td>Protein biosynthesis</td>
<td>7.83</td>
<td>0.65</td>
<td>1.62E−002</td>
<td>6</td>
</tr>
<tr>
<td>Response to oxidative stress</td>
<td>0.90</td>
<td>4.33</td>
<td>3.35E−002</td>
<td>4</td>
</tr>
<tr>
<td>Response to pest, pathogen or parasite</td>
<td>2.20</td>
<td>6.73</td>
<td>4.73E−002</td>
<td>4</td>
</tr>
</tbody>
</table>

GO terms significantly different between Arabidopsis differentially expressed genes and the rest of Arabidopsis ATH1 array genes. Arabidopsis differentially expressed genes with orthologues and the rest of Arabidopsis ATH1 array genes with orthologues.
results proved that this approach is very useful in identifying genes that are implicated in conserved processes. This method will be especially useful when working with species that have no genomic or microarray facilities, such as many important crops. This work is another example that it is possible to use the plant model to study the global expression of a relative crop and not only the function of single genes. Future works in model plants and crops will show the conservation complexity of the most important developmental and physiological pathways, as has been repeatedly reported in the animal kingdom.

4.4. Functional analysis of the differentially expressed genes

To elucidate the possible pathway implicated in fruit set, we analyzed the GO terms of both experiments, the suppressive subtractive hybridization library and microarray data. There were several terms detected as differentially represented in both analyses, such as protein biosynthesis, response to heat, response to stresses and cell cycle. Other functions were detected only in the SSH library, such as that related with histone function (nucleosome assembly, chromatin assembly), photosynthesis and others. But it is necessary to bear in mind that the SSH was a comparative analysis between two stages of whole flower development, anthesis and 3 DPA, and the microarray analysis was focused on the carpel development. Moreover, in the SSH analysis we did not use any statistical method to detect the GO terms as the number of data was too small. In fact, in the Babelomics assay, there were also differences between the analyses of total Arabidopsis microarray genes and the genes with orthologues; these differences were clearly due to the lower number of genes with unambiguously assigned orthologues, as we did not detect any deviation in the differentially expressed terms when compared with the total microarray data and orthologue data.

Globally, we detected significant pathways, response to several hormone stimuli, cell wall modification, cell cycle, photosynthesis, regulation of transcription, lipid transport, ageing, etc. Many of these hormone response pathways have already been described as implicated in fruit set or fruit development of tomato, such as response to gibberellic acid, response to auxin, response to jasmonic acid and response to abscisic acid. Other differential processes implicated in fruit development have also been detected, such as cell cycle, cell wall modification [3] and photosynthesis [44]. Other terms, such as response to heat, temperature, abiotic stimuli and other stresses have been detected during other stages of fruit development [15,13]. We also detected an enrichment in transcription factors caused by the start of the transformation of tissues in the fruit set. These results showed that this gene collection may contain other pathways and genes whose implication for fruit set and development has to be experimentally probed, and which will be a good source of data for future works and analysis.

In this work, we isolated genes that are potentially implicated in tomato fruit set, detected a high degree of conservation in the expression between Arabidopsis and tomato, in spite of the clear differences between the fruits of these species, detected several pathways that may be implicated in this process and we tested a new approach to isolating differentially expressed genes. The high degree of conservation between Arabidopsis and tomato expression during fruit set and early fruit development proves that this orthologue approach can be useful in other biological processes or species, as it is a quick system for identifying candidate genes in species where there are no available microarray facilities.

Acknowledgements

J.C. is the recipient of a Ramon y Cajal Grant of the Ministerio de Educación y Ciencia. L.P. is the recipient of an FPU fellowship of the Ministerio de Educación y Ciencia. This work was supported by a project of the Generalitat Valenciana.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.plantsci.2007.07.006.

References

profiles are associated with early fruit tissue specialization in tomato, Plant Physiol. 139 (2) (2005) 750–769.


[26] I.K. Jordan, I.B. Rogozin, Y.I. Wolf, E.V. Koonin, Essential genes are


[29] I.K. Jordan, I.B. Rogozin, Y.I. Wolf, E.V. Koonin, Essential genes are


