

RESEARCH ARTICLE

Open Access

# Transcription profiling of fertilization and early seed development events in a solanaceous species using a 7.7 K cDNA microarray from *Solanum chacoense* ovules

Faiza Tebbji<sup>1,2</sup>, André Nantel<sup>1,2</sup>, Daniel P Matton<sup>1\*</sup>

## Abstract

**Background:** To provide a broad analysis of gene expression changes in developing embryos from a solanaceous species, we produced amplicon-derived microarrays with 7741 ESTs isolated from *Solanum chacoense* ovules bearing embryos from all developmental stages. Our aims were to: 1) identify genes expressed in a tissue-specific and temporal-specific manner; 2) define clusters of genes showing similar patterns of spatial and temporal expression; and 3) identify stage-specific or transition-specific candidate genes for further functional genomic analyses.

**Results:** We analyzed gene expression during *S. chacoense* embryogenesis in a series of experiments with probes derived from ovules isolated before and after fertilization (from 0 to 22 days after pollination), and from leaves, anthers, and styles. From the 6374 unigenes present in our array, 1024 genes were differentially expressed ( $\geq \pm 2$  fold change,  $p$  value  $\leq 0.01$ ) in fertilized ovules compared to unfertilized ovules and only limited expression overlap was observed between these genes and the genes expressed in the other tissues tested, with the vast majority of the fertilization-regulated genes specifically or predominantly expressed in ovules (955 genes). During embryogenesis three major expression profiles corresponding to early, middle and late stages of embryo development were identified. From the early and middle stages, a large number of genes corresponding to cell cycle, DNA processing, signal transduction, and transcriptional regulation were found. Defense and stress response-related genes were found in all stages of embryo development. Protein biosynthesis genes, genes coding for ribosomal proteins and other components of the translation machinery were highly expressed in embryos during the early stage. Genes for protein degradation were overrepresented later in the middle and late stages of embryo development. As expected, storage protein transcripts accumulated predominantly in the late stage of embryo development.

**Conclusion:** Our analysis provides the first study in a solanaceous species of the transcriptional program that takes place during the early phases of plant reproductive development, including all embryogenesis steps during a comprehensive time-course. Our comparative expression profiling strategy between fertilized and unfertilized ovules identified a subset of genes specifically or predominantly expressed in ovules while a closer analysis between each consecutive time point allowed the identification of a subset of stage-specific and transition-specific genes.

\* Correspondence: [dp.matton@umontreal.ca](mailto:dp.matton@umontreal.ca)

<sup>1</sup>Institut de recherche en biologie végétale, Département de sciences biologiques, Université de Montréal, 4101 rue Sherbrooke est, Montréal, Québec, H1X 2B2, Canada

Full list of author information is available at the end of the article

## Background

Angiosperm sexual reproduction starts when pollen is transferred from the anther to the stigma. After hydration and germination, the pollen tube carrying two sperm cells enters the embryo sac through the micropyle of the ovule and penetrates one of the synergid where it discharges its contents. One sperm cell fuses with the egg cell and the resultant zygote develops into the embryo. The central cell unites with the second sperm cell to form a triploid primary endosperm cell that develops into the endosperm [1-5]. Thus, the double fertilization event initiates the development of two interconnected multicellular structures, the diploid embryo and the triploid endosperm. Embryogenesis patterning starts with an asymmetric cell division that produces a small apical cell, that ultimately becomes the embryo, and a large basal cell (leading to the suspensor) that functions to provide nutrients from the endosperm to the growing embryo [6]. Embryogenesis in higher plants can be divided conceptually into three overlapping phases [7-9]. The first phase involves morphogenesis and pattern formation, during which the polar axis of the plant body is defined by specification of the shoot and root apices, and the embryonic tissue and organ systems are formed (pattern formation/cell proliferation/cell division). During the second phase (maturation), storage reserves start to accumulate in the maturing embryo, and, in the third phase, the embryo prepares for desiccation and enters a period of developmental arrest. Successful embryogenesis thus leads to seed development. In *Arabidopsis* embryogenesis is rapid, requiring only 14 days after pollination to produce the desiccated mature seed [10]. In solanaceous species, embryogenesis spans a much longer period. For example, in *Solanum phureja*, embryo maturation, excluding desiccation, is only completed 27 days after pollination [11].

Classification of gene expression patterns associated with specific stages of embryo and seed development and functional understanding of the encoded genes are critical for understanding the molecular and biochemical events associated with embryogenesis. Most of the knowledge concerning the genetic program activated during embryogenesis of higher plants comes from the study of mutants. Several mutants affecting embryonic development were identified in numerous plant species including *A. thaliana*, maize, rice, and tomato through various mutagenesis strategies [12-26]. These analyses enabled the identification and characterization of several genes that play key roles in plant embryonic development. However, mutagenesis strategies alone cannot identify all of the genes that are potentially involved in a biological process.

In the past few years, several differential screening techniques (including differential display, subtracted libraries, differential hybridization, etc.) have made it

possible to characterize genes differentially expressed during embryogenesis [27,28]. More recently, other methods such as transcriptional profiling have allowed us to visualize global changes in transcript abundance in a spatial, temporal or conditional way. Genome-wide transcription profiling is an important and powerful tool leading to the generation of testable hypotheses for novel processes not yet characterized at the molecular level. Its usefulness has been demonstrated in the investigation of transcriptional programs occurring in a variety of developmental processes such as fruit ripening, seed development, flower development, embryo development, defense response to pathogens, and the response to wounding [29-44]. Now that the technology has matured, microarray data are routinely correlated with other methods that measure RNA expression levels [45,46].

Recently we completed a medium scale EST sequencing project from normalized cDNA libraries made from *Solanum chacoense* ovules in order to identify signal transduction components involved in fertilization and early seed development [47]. *Solanum chacoense*, a diploid ( $2n = 2x = 24$ ) and self-incompatible close relative of the potato and tomato, produces a large numbers of easily isolated ovules inside a fleshy tomato-like fruit. It also bears tubers, albeit small, like all the members of the *Solanum* Petota section, and is also considered an important source of interesting genetic traits for its high resistance to more than 20 pests and diseases [48]. Flowers of *S. chacoense* are actinomorphic with a 5-lobed calyx and corolla, five stamens, and a 2-carpellate superior ovary. *S. chacoense* display homomorphic self-incompatibility of the gametophytic type involving S-RNases. The presence of a gametophytic self-incompatibility barrier was used to dissect the molecular events occurring at precise time points following pollination and fertilization since it enables synchronization of fertilization. These EST were used to construct cDNA microarrays consisting of 7,741 cDNAs spotted in duplicate, which were used to investigate the regulation of gene expression during fertilization and early seed development events from 0 to 22 days after pollination (DAP). Our aims were to: 1) identify genes expressed in a tissue-specific and temporal-specific manner; 2) define clusters of genes showing similar patterns of spatial and temporal expression; and 3) identify stage-specific or transition-specific candidate genes for further functional analyses.

## Methods

### *Solanum chacoense* cDNA clones and inserts amplification

We have previously used a subtraction screen on two cDNA libraries covering embryo development from

early fertilization events (6-12 hours post-fertilization) to ovules bearing late torpedo stage embryos as a method to produce an EST pool enriched for weakly expressed messenger RNAs. From roughly 50,000 colonies spotted on nylon membranes, 8000 colonies that displayed a hybridization signal corresponding to the lowest 20% were selected for further analyses. 7741 good sequences were obtained and these comprised 6374 unigenes [47]. Inserts of these cDNA clones were PCR amplified using flanking primers complementary to vector sequences upstream and downstream of the cDNA inserts. Plasmid templates (1-2 ng) were added to 100  $\mu$ l PCR mixture containing 0.2 mM of each nucleotide, 1  $\mu$ M of each primer, 1.5 mM of  $MgCl_2$  and 10 units of Taq DNA Polymerase. Inserts were amplified for 36 cycles (95°C for 30 sec, 54°C for 30 sec, 72°C for 2 min), with an initial denaturation at 95°C for 2 min and final extension at 72°C for 6 min. Two (2)  $\mu$ l of each reaction were separated on 1.5% agarose gels to confirm amplification quantity and quality. PCR products were purified using MultiScreen<sup>®</sup> PCR<sub>96</sub> filter plates (Millipore, Billerica, MA, USA), and lyophilized. In general, the amount of each PCR product was greater than 5  $\mu$ g and the average insert size was around 1000 bp.

#### cDNA microarray preparation

PCR products were resuspended in 10  $\mu$ l of 50% DMSO and arrayed from 384 well microtiter plates onto Ultra-GAPS<sup>™</sup> slides (Corning incorporated, Corning, NY, USA). A total of 7741 ESTs spotted in duplicate along with a variety of controls, including buffer-only spots and a dilution series of a plasmid harboring the *Can-dida albicans Ece1* gene.

#### Plant materials and RNA isolation

The diploid ( $2n = 2x = 24$ ) and self-incompatible wild potato species *Solanum chacoense* Bitt. was greenhouse grown with an average photoperiod of 14-16 h per day. The genotypes used were originally obtained from the USDA Agricultural Research Service, NRSP-6 Potato Genebank (Potato Introduction Station, Sturgeon Bay, WI, USA). Plant material was collected from female progenitor, *S. chacoense* genotype G4 ( $S_{12}$  and  $S_{14}$  self-incompatibility alleles) [49]. For fertilization-related events, the fully compatible *S. chacoense* genotype V22 ( $S_{11}$  and  $S_{13}$  self-incompatibility alleles) was used as pollen donor. Plants were hand pollinated and ovules were collected between 0 and 22 days after pollination (DAP) and used for RNA extraction and probe preparations. Leaf, style and anther tissues were collected from plant genotype G4 grown under the same conditions as above. All samples for RNA preparation were quick-frozen in liquid nitrogen and ground to a powder with a mortar and pestle. Total RNA was extracted using the TRIzol<sup>®</sup> Reagent according to the

manufacturer's protocol (Invitrogen, Burlington, ON, Canada). The yield and purity of RNA were assessed by determination of absorbance at both 260 nm and 280 nm. RNA was only used when the ratio  $Abs_{260\text{ nm}}/Abs_{280\text{ nm}}$  was higher than 1.7. RNA integrity was checked by both agarose gel 1% and with the RNA 6000 Nano Assay Kit and the Agilent 2100 Bioanalyzer. RNA from unfertilized ovules served as control.

#### Design of microarray experiments

To monitor the expression pattern from genes involved in fertilization and embryogenesis processes, flowers were hand-pollinated and ovules were isolated every two days during a 22 days period after pollination. Four independent biological replicates were produced from each time points. In addition, to isolate genes specifically or predominantly expressed in ovules, four biological replicates of leaf, anther, and style tissue mRNA preparations were individually hybridized against unfertilized ovule mRNAs and compared with the data obtained from unfertilized and fertilized ovules at various time points after pollination. To estimate reproducibility and to produce control data for statistical analysis, a large number of unfertilized ovules were isolated and separated between seven independent control groups. RNA from randomly selected pairs of control was hybridized on six microarrays.

#### Preparation of fluorescent probes

Thirty (30)  $\mu$ g of total RNA, 1.5  $\mu$ l oligo (dT)<sub>21</sub> (100 pmol/ $\mu$ l), and water to a volume of 36  $\mu$ l was denatured at 65°C for 10 min and cooled to room temperature for 5 min. The reverse transcription reaction was performed at 42°C for 2 h after addition of 3  $\mu$ l dNTP-minus dCTP (6.67 mM each), 1  $\mu$ l dCTP (2 mM), 4  $\mu$ l DTT (100 mM), 8  $\mu$ l 5 $\times$  first strand buffer (Invitrogen), 2  $\mu$ l of cyanine 3-dCTP (1 mM) or cyanine 5-dCTP (1 mM; Perkin Elmer-Cetus/NEN, Boston, MA, catalog # NEL999) and 2  $\mu$ l of SuperScript II (200 units/ $\mu$ l, Invitrogen). After incubation at 42°C, 0.05 units each of RNase A and RNase H (Invitrogen) were added and the reaction mix was incubated at 37°C for 30 min. The probes were then purified using CyScribe GFX Purification Kit (GE Healthcare Bio-Sciences Inc., Baie d'Urfé, QC, Canada) according to the manufacturer's instructions. Incorporation efficiency of the CyDye probes was measured using the ND-1000 Nano drop spectrophotometer (NanoDrop, Wilmington, DE, USA). In general, optimal amounts are 20 pmol of dye incorporated with an incorporation efficiency of between 20-50 labeled nucleotides per 1000 nucleotides. Probes were then air-dried.

#### Hybridization

Slides were prehybridized at 42°C for at least 1 hour, with 50  $\mu$ l of a solution containing 5 $\times$  SSC, 0.1% SDS

and 1% BSA. The two cDNA targets were pooled together, and mixed with the hybridization buffer to a volume of 50  $\mu$ l Dig Ease Hybridization buffer (Roche Applied Science, Mississauga, ON, Canada), 2.5  $\mu$ l tRNA (Roche Applied Science) and 2.5  $\mu$ l of Sonicated Salmon Sperm DNA (10 mg/ml; Invitrogen). The hybridization solution was heat denatured at 95°C for 3 min, cooled to room temperature, and applied onto the DNA microarray slides for overnight hybridization at 42°C. To account for the possibility of dye bias, half of the hybridizations were performed in the Cy3/Cy5 configuration, and half in the Cy5/Cy3 configuration. The microarray slides were covered with a 24  $\times$  60-mm glass coverslip (Fisher Scientific, Ottawa, ON, Canada). During all hybridization steps the hybridization chamber was kept at high humidity level. Immediately before hybridization, the DNA microarray slides were washed twice with 0.1 $\times$  SSC at room temperature for 5 min and once with water for 30 sec and centrifuged at 800 rpm for 3 min. The DNA microarray slides were kept dry for a minimal amount of time before hybridization. Afterward, slides were completely immersed in a large volume of washing buffer, and the coverslips were carefully removed before washing twice for 10 min at 42°C with 1 $\times$  SSC, 0.1% SDS, twice for 10 min at 37°C with 0.1 $\times$  SSC, 0.1% SDS, and finally, with quick consecutive washes in three 0.1 $\times$  SSC baths. DNA chips were air dried and stored protected from light until scanning.

#### Microarray scanning and data analysis

The DNA microarray slides were scanned with a ScanArray Lite microarray scanner (Perkin Elmer-Cetus, Wellesley, CA) at 10- $\mu$ m/pixel resolution. The fluorescence intensities were quantified with QuantArray software (Perkin Elmer-Cetus; versions 2.0 and 3.0). Microarray data normalization and analysis was performed in GeneSpring GX software version 7.3 (Agilent Technologies, Santa Clara, CA, USA). Raw intensities were normalized with a Lowess curve using 20% of the data to fit each point. To identify transcripts with a significant change in abundance, the fluorescence ratios from each time point were compared to the fluorescence ratios from 6 control hybridizations using the Welch t-test and the Benjamini and Hochberg False Discovery Rate. We selected genes with a p-value  $\leq$  0.01 and further restricted the lists to transcripts whose change in abundance was greater than 2-fold. The use of such a stringent gene selection method that uses both multiple testing correction and a fold-change cutoff did not significantly affect our conclusions. GeneSpring GX was also used for hierarchical clustering and to detect significant overlaps between gene lists. Functional categorization was performed according to the Gene Ontology (GO) Consortium and Arabidopsis consortium

information (<http://www.geneontology.org/>). Translated *S. chacoense* sequences were sorted into 15 functional categories by sequence comparison. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [50] and are accessible through GEO Series accession number GSE21552 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21552>).

#### Real-time PCR

cDNA samples were synthesized from 2  $\mu$ g of 2 independent preparations of total RNA using the SuperScript<sup>®</sup> II Reverse Transcriptase (Invitrogen). The RNA was denatured at 70°C for 10 min. The total volume was adjusted to 40  $\mu$ l by adding 8  $\mu$ l of 5 $\times$  RT buffer, 2 pmoles of oligo dT 25 primer, 4  $\mu$ l of 0.1 mM DTT, 2  $\mu$ l of 10 mM dNTP, 200 units of SuperScript<sup>®</sup> II Reverse Transcriptase and 19  $\mu$ l of water. The mixture was then incubated 1 hour at 42°C. The resulting first strand cDNA was used for real-time PCR amplification experiments. Ovule RNAs from three different developmental stages (4, 12, and 22 days after pollination) were compared. RNAs from unfertilized ovules were used as the calibration tissue. Duplicate quantitative assays for each tissue were performed with the SYBR Green Master mix (Invitrogen) according to the manufacturer's instructions. For real-time PCR amplification, the following PCR program was used: 50°C for 2 min, 95°C for 10 min, 95°C for 15 s, 60°C for 1 min; steps 3 and 4 were repeated 40 times in Mx4000<sup>®</sup> Multiplex QPCR System (Stratagene, La Jolla, CA, USA). The relative quantification analysis was performed using the comparative  $\Delta\Delta$ Ct method [51]. To evaluate the gene expression level, the results were normalized using an ubiquitin gene (DN977330) as control. Primer sequences used in real time quantitative PCR are summarized in Additional file 1.

#### Tissue fixation and optical microscopy observation

Ovules were fixed in FAA for 24 h at 4°C (50% ethanol, 1.35% formaldehyde, and 5% glacial acetic acid). Samples were then dehydrated in an increasing series of ethanol baths (from 30% to pure ethanol). Microscopic observations were taken on an AxioImager M1 microscope equipped with an Axio Cam MRm camera (Carl Zeiss Canada, Toronto, ON, Canada).

## Results

#### Microarray spotting and evaluation

We have previously generated expressed sequence tags (ESTs) derived from fertilized ovule cDNA libraries covering embryo development from zygote to late torpedo stages in *Solanum chacoense* Bitt [47]. The 7741 ESTs corresponded to weakly expressed mRNAs obtained through a subtraction selection screen that produced a

highly enriched unigene set (6374 unigenes) corresponding to 82% of the total ESTs sequenced. To provide a glimpse into the ovule transcriptome during embryogenesis in a solanaceous species and to identify a subset of ovule-specific and stage-specific genes comprising this transcriptome, these ESTs were used to construct amplicon-based cDNA microarrays. In the present study, we used a comparative expression profiling strategy to analyze gene expression profiles in ovules from 0 to 22 days after pollination.

For microarray production, the ESTs were amplified by PCR with vector-specific primers. The PCR products were purified, verified for single clone amplification, normalized, and arrayed onto aminosilane-coated glass surfaces. To increase the reliability of the detected signals, each PCR sample was spotted twice resulting in 15,482 data points for each array. Several controls were also spotted to verify the reliability of our hybridization. After each spotting runs, samples of printed arrays were visualized with a fluorescent dye to control evenness of DNA deposition and spot morphology.

Sixty-two (62) hybridizations corresponding to 15 experiments were performed (see material and methods for details). Twelve experiments consisted of the comparison of ovule tissues harvested at 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 and 22 days after pollination (DAP) with unfertilized ovules (control, 0 DAP). To determine the percentage of genes specifically or predominantly expressed in ovules, three experiments that compared RNA from unfertilized ovules to leaf, style, and anther tissues were also conducted. Due to the large amount of control tissue material needed for each time point comparison and for each replicate (unfertilized ovules corresponding to 0 DAP), the control RNA obtained were pooled from different replicates and used for comparison with the various time points and other tissues. Pooling RNA before labeling has the advantage of reducing the variation due to biological replication and sample handling [52]. To estimate this variation we also competitively hybridized cDNA probes from six additional control samples collected from unfertilized ovule against the pooled control. The six individual control samples were different from the pooled one that was used as the universal control in all our experiments. The correlation coefficient between the normalized intensities of the two channels was of 0.98, and only few spots showed occasional differences in intensities above 2-fold (Figure 1a). This very high correlation coefficient gives a strong indication that biological replication is not a significant source of variation. For each experiment, four biological replicates were collected from the same greenhouse and from a large number of plants. For each individual replicate, all the necessary material was collected at once. Each hybridization was performed against the pooled control RNA sample (unfertilized ovules, UO). To

account for the possibility of dye bias, half of the hybridizations were performed in the Cy3/Cy5 configuration and half in the Cy5/Cy3 configuration.

#### **Expression profiling of genes involved in fertilization and embryogenesis**

To monitor gene expression profiles during fertilization and early embryogenesis, total RNA samples from UO to 22 DAP were isolated, labeled, and hybridized to the 7.7 K microarray. A fraction of the ovules from each sample was set aside and fixed. These ovules were cleared and analyzed by differential interference contrast (DIC) microscopy to ensure that homogenous developmental stages were used in the various developmental time points used. Strict pollination timing was accomplished through manual pollinations with pollen from a fully compatible *S. chacoense* genotype on anthesis day. Fertilization occurs between 36 and 42 hours following pollination in *S. chacoense* [53]. Since *S. chacoense* is a self-incompatible species this prevents self-fertilization and eliminates the need for emasculation, which could induce a bias in gene expression through a wounding response. Furthermore, since anther dehiscence is not completed at the time of anthesis while the pistil is already fully receptive, this reduces the risk of self-pollen landing on the stigma that could trigger a self-incompatible response.

For each time point, four biological replicates were used, with half labeled as Cy3 cDNA probes and the other half as Cy5 cDNA probes. Using Anova testing, along with a Benjamini and Hochberg multiple testing correction algorithm, 1997 transcripts showed a statistically significant change in abundance ( $p \leq 0.01$ ) in at least one of the time points compared to the control hybridizations (Figure 2A). Amongst those, 1024 showed a greater than  $\pm 2$ -fold variation in expression in at least one of the time points (Additional file 2). These differentially expressed genes were grouped according to the similarity of their expression profiles using two dimensional hierarchical clustering (Figure 2B) [54]. These results clearly segregate the dataset amongst three major groups of genes that specify early (494 genes), middle (604 genes), and late (203 genes) stages of embryogenesis (Figure 2A and 2B). Embryonic stages as determined by ovule clearing and microscopical observation encompassing each time point are schematically represented (Figure 2A). From 0 to 6 DAP (early stage), embryo development corresponded to the post fertilization stages that include embryos from the zygotic stage to 8-celled embryo proper. From 8 to 16 DAP (middle stage) ovules mainly bore embryos from the 16 cell to heart stages, and from 18 to 22 DAP (late stage) embryo developmental stages encompassed torpedo, walking stick, and mature embryos.

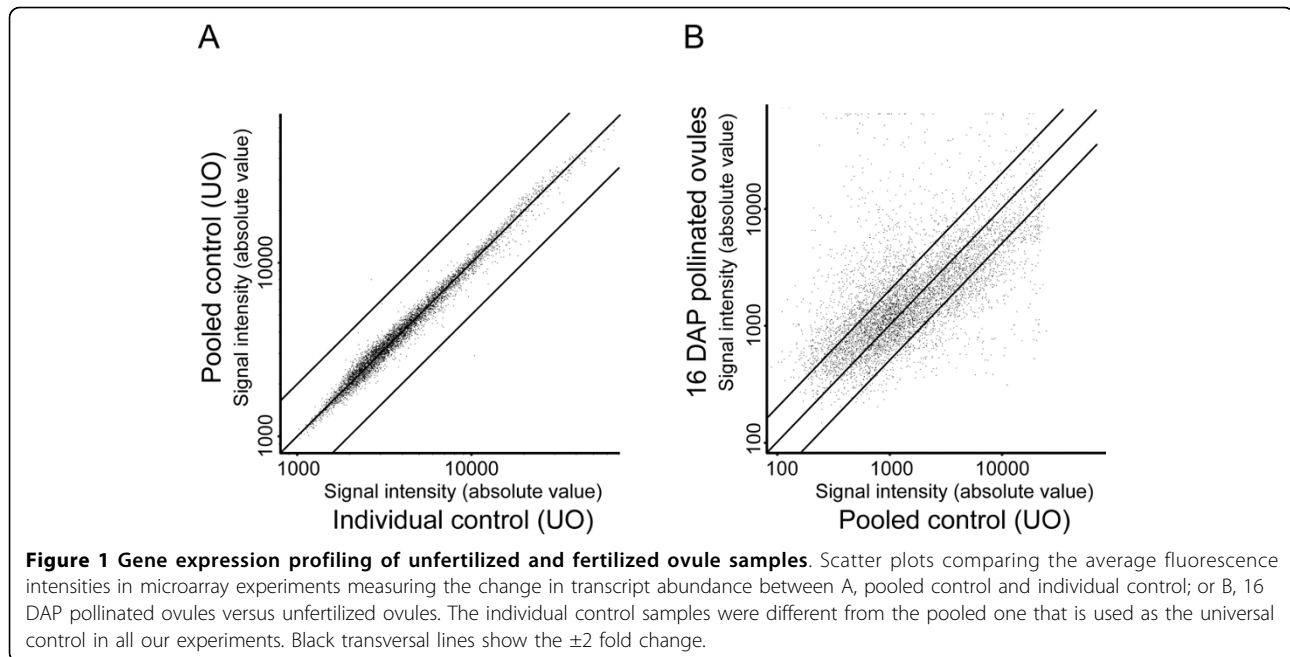


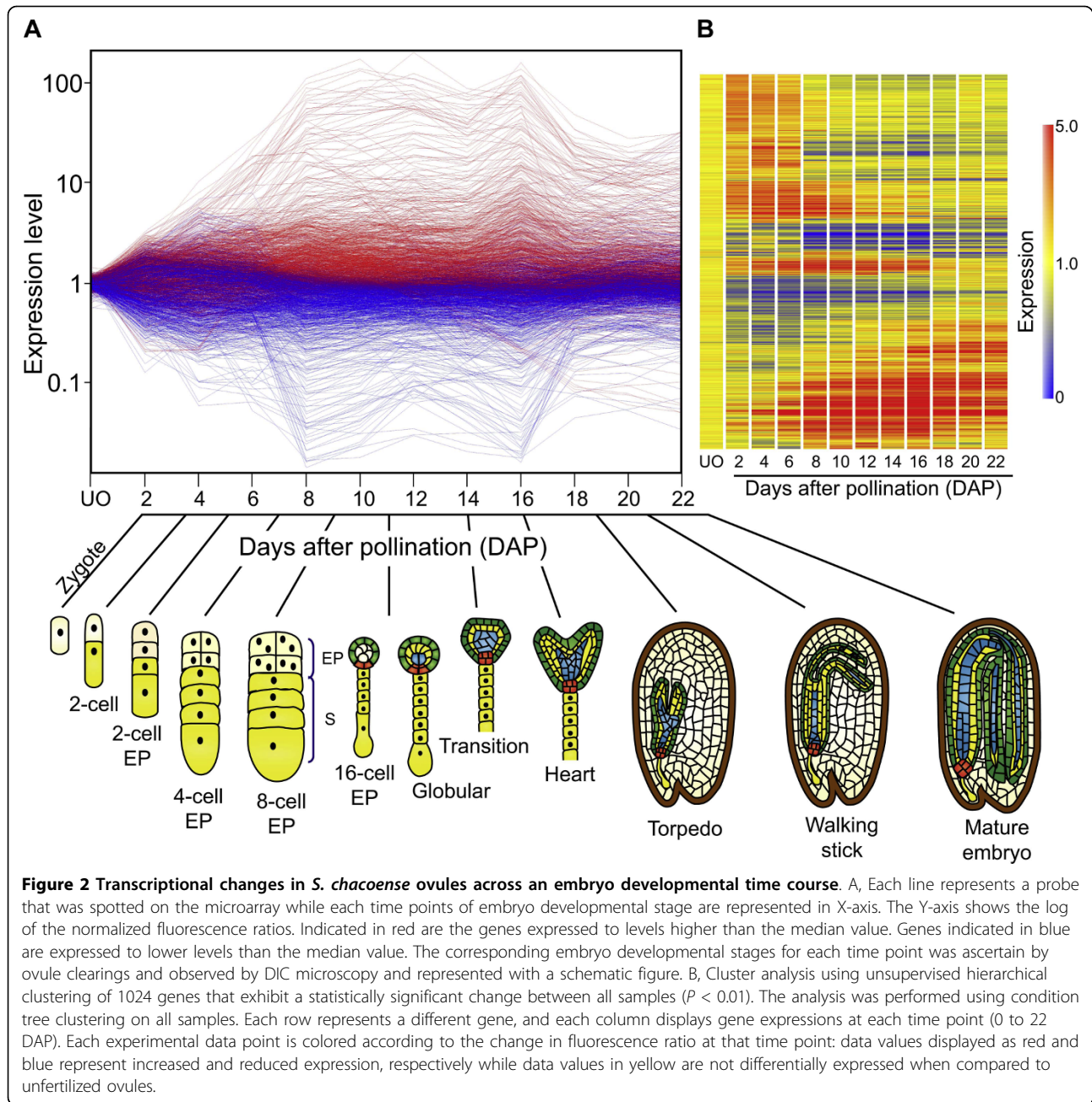
Figure 2 presents the global profiling during all stages of embryogenesis. In order to better illustrate the dynamic and magnitude of these changes in gene expression, and the distribution of the signal intensities obtained from each probe, an example of a scatter plot representing the data from a single time point comparing pollinated ovule (16 DAP) versus unfertilized ovule is shown in Figure 1B. This representation shows the global distribution of all the genes spotted on the microarray slide at 16 DAP vs. UO. The scatter plot showed a positive correlation between fertilized ovules at 16 DAP vs. unfertilized ovule for most of the genes, with the majority of the genes analyzed showing a lesser than  $\pm 2$  fold difference in expression when compared to UO across the full intensity range. Genes that show a greater than  $\pm 2$  fold expression difference (735 genes) are also evenly distributed across all intensity range up to 150 fold change with a group of genes that show signal saturation. Comparing each time point scatter plots against the UO experiment (Figure 1) clearly showed that, for each time point, a group of genes are significantly modulated. All time point scatter plot comparisons gave similar distributions (data not shown).

#### Analysis of genes specifically or predominantly expressed in ovule tissues

To determine the percentage of ovule-specific genes in our EST pool, RNA samples from leaf, anther, and style tissues of *S. chacoense* were compared to the UO pool sample. Each microarray experiment was performed four times using, in all cases, independently isolated

RNA samples as starting material (four biological replicates). Statistical analyses showed that 558 genes (out of all ESTs) showed a significant variation in transcript abundance ( $\geq \pm 2$ -fold coupled with  $p$  values  $\leq 0.01$ ) in at least one tissue (leaf, style, or anther) when compared to UO (Figure 3A). A Venn diagram analysis of modulated genes from leaves, anthers, and styles shows that, from the 558 genes, 3 genes were differentially expressed in all tissues compared to ovules; 322 genes are differentially expressed only in anthers; 10 genes are differentially expressed only in leaves; and 115 genes are differentially expressed only in styles. Four (4) genes are co-regulated in leaves and styles; 5 genes are co-regulated in leaves and anthers; and 99 genes are co-regulated in anthers and styles (Figure 3B).

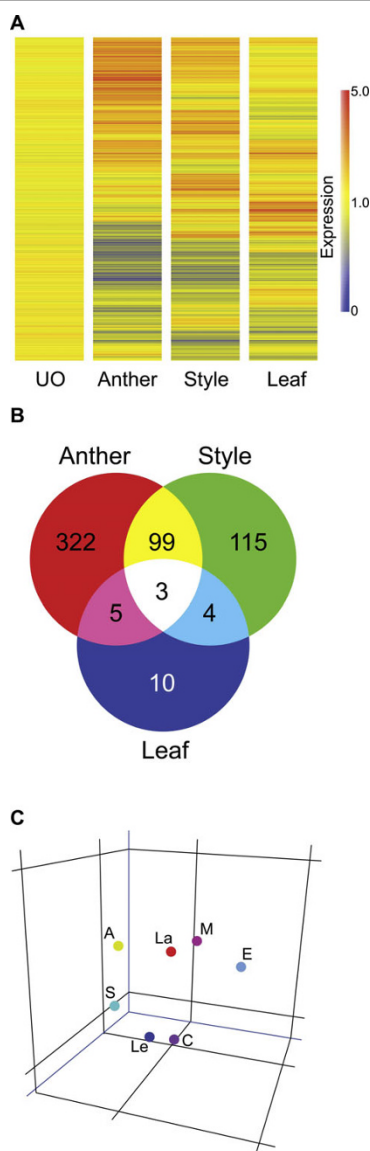
From these 558 genes, 262 genes were up-regulated ( $\geq 2$ -fold variation) and 296 genes were down-regulated ( $\leq 2$ -fold change) when compared to UO (Additional file 3). We thus considered that the transcripts corresponding to these 296 genes were predominantly expressed in UO. From the 262 transcripts that were predominantly expressed in styles, anthers, or leaves, 69 genes were also differentially expressed in ovules after pollination but their induction level was less than the one observed in the other tissues tested ( $< 2$ -fold). A principal components analysis of these genes clearly indicates that the subset of genes expressed in non-ovule tissues is quite different than those expressed in ovules (Figure 3C). Thus, these results confirm that a large number of the ESTs on our microarrays can be considered ovule-specific or ovule-predominant. Out of the 1024 genes that



showed more than  $\pm 2$ -fold variation in expression after fertilization, 955 (1024 - 69, 93.2%) are considered specifically or predominantly expressed in ovule tissues.

Amongst these genes, a small group (1.15%) was modulated by more than 100-fold during embryogenesis, 2.1% of the genes were modulated by  $\pm 50$  to 100-fold, 4.8% between  $\pm 20$  and 50-fold, and 6.1% between  $\pm 10$  and 20-fold (Additional file 4). The most highly expressed genes comprised almost exclusively genes involved in proteolysis and peak transcript accumulation occurred between 8 and 16 DAP (Figure 2A). Genes

involved in proteolytic function were also highly represented in the  $\geq 50$  to 100-fold and  $\geq 20$  to 50-fold up-regulated genes and covered various classes of proteases including serine-type peptidases (carboxypeptidases, subtilases or subtilisin-like serine protease), aspartic-type peptidases, and cysteine-type peptidases. Apart from genes involved in proteolysis, other highly up-regulated genes ( $\geq 10$ -100 fold) included numerous lipid transfer proteins (LTP) and non-specific lipid transfer proteins (nsLTP) that are involved in various biological processes including plant defense, pollen tube adhesion and



**Figure 3 Tissue-specific comparison between ovule, anther and leaf tissues.** A, Cluster analysis of the transcriptional profiles obtained from leaves, styles, and anthers. The clustering was performed using condition tree clustering on all samples. Each row represents a different gene, and each column displays gene expressions at different tissues (UO, style, leaf and anther). Data values displayed as red and blue represent elevated and reduced expression, respectively. A total of 601 genes with a statistically significant change in transcript abundance of  $\geq 2$ -fold in at least one tissue compared to unfertilized ovules ( $p$  value  $< 0.01$ ) are found. B, Venn diagram showing overlap between lists of genes differentially expressed in anther, style and leaf tissues. C, Principal Components Analysis (PCA) of various tissue types (A: Anthers; S: Styles; Le: Leaves; C: unfertilized ovules. E, M, L: early, middle and late stages of embryo development, respectively) based on the expression profile of 1376 genes including 558 that are differentially expressed in at least one tissue from styles, leaves and anthers and 1024 transcripts that are regulated in ovule which resulted in a clear differentiation distributed between 7 classifications.

guidance, embryo patterning and cell wall biogenesis [55-60], as well as proteinase inhibitors that are involved in plant defense responses and early seed development [61,62].

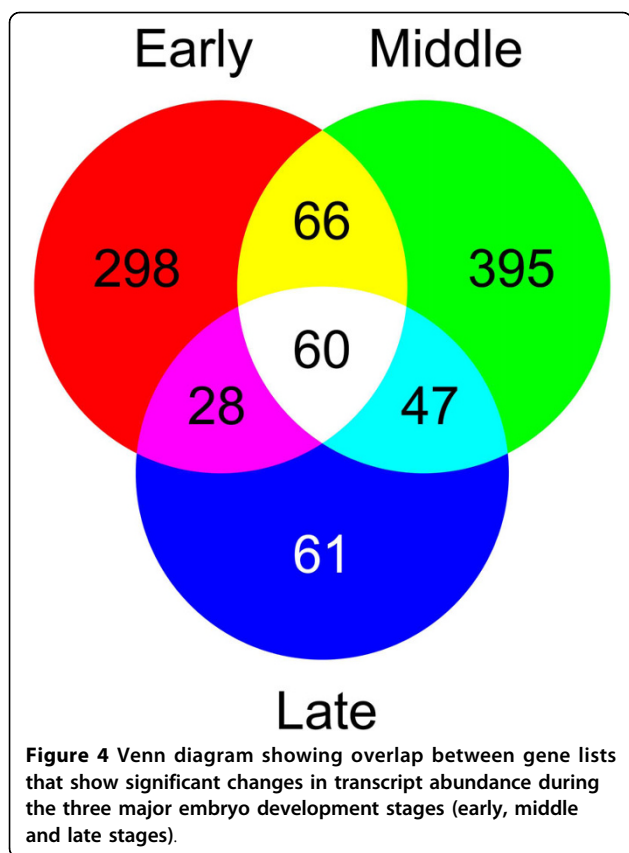
Amongst the genes found to be strongly repressed in the following combined categories (from  $\leq 10$  to 20,  $\leq 20$  to 50,  $\leq 50$  to 100-fold), most corresponded to genes classified as being involved in stress responses with a high representation of metallopeptidase inhibitor (MCPI) genes. One such MCPI was characterized by Martineau et al. (1991) as being highly expressed in anthesis stage ovaries in tomato (*Solanum lycopersicum*) while decreasing quite rapidly during fruit development, with an estimated 10-fold drop [63,64]. The three most similar *S. chacoense* orthologs to the tomato MCPI were also found in the 10 to 20-fold down-regulated genes (Additional file 4).

#### Functional classification of differentially expressed genes during fertilization and early seed development

The hierarchical gene clustering presented in Figure 2B showed that, following fertilization, the ovule-expressed gene pool could be clearly divided into three major groups that specified early, middle, and late stages of embryogenesis. A Venn diagram analysis of the genes predominantly expressed in fertilized ovules indicates that 298 genes are differentially expressed during the early developmental stage, 395 genes only during the middle stage, and 61 genes only during the late stage (Figure 4). Sixty-six (66) genes are differentially expressed in early and middle stages, 28 genes in early and late stages and 47 genes in middle and late stages. Sixty (60) genes are differentially expressed in all embryo development stages (Additional file 5). Since the ESTs used to build the microarray were isolated from ovules harvested at various time following pollination, they encompass all different developmental stages of the embryo. The microarray analysis was thus sensitive enough to separate them between the major developmental transitions. This also suggests a highly specific expression program for each separate stage, with little overlap between the three major stages defined.

To determine whether the differentially expressed genes were involved in similar biological processes, transcript profiling data was correlated with the Gene Ontology classification based on both the *S. chacoense* sequences and their closest orthologs in *Arabidopsis thaliana*. The functional categories of these genes are given in Figure 5. As expected, the largest category in every stage consisted of genes with unknown function or genes with no significant homology/no hit (29.8%), as observed in fully sequenced genomes [65]. For the early and middle stages, the largest functional categories included metabolism (49 and 83 genes respectively),





regulation of transcription (28 and 22 genes respectively), energy (11 and 18 transcripts respectively), signal transduction (14 and 12 transcripts respectively), protein fate (11 and 13 transcripts respectively) and proteins with binding functions (12 and 19 transcripts respectively).

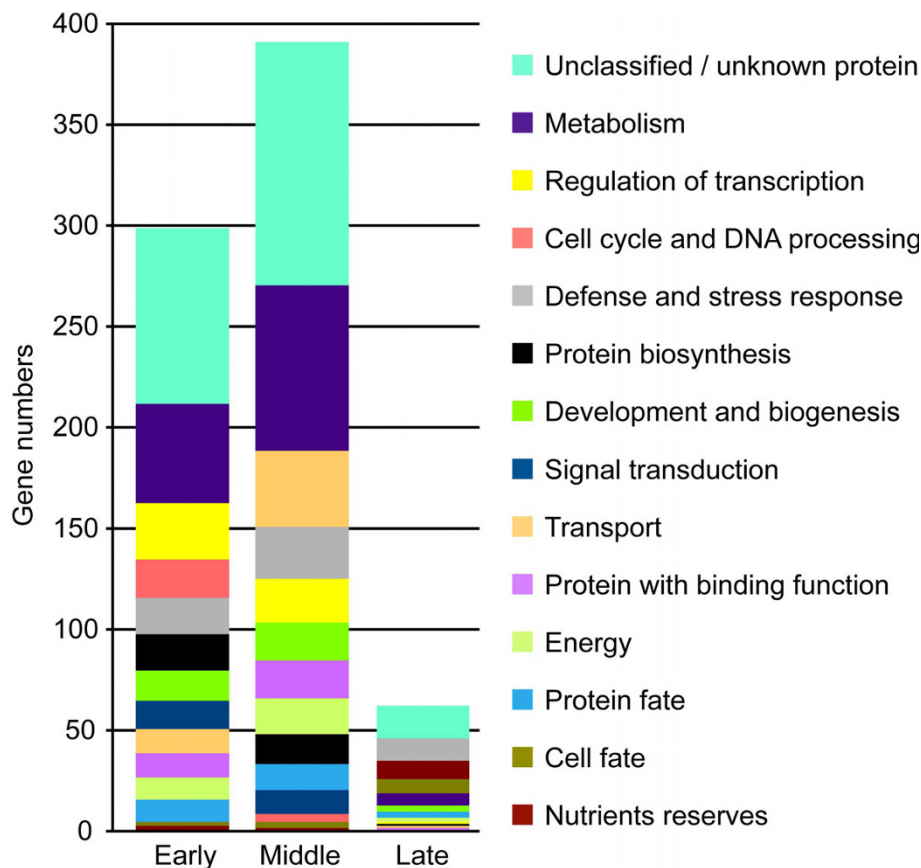
One functional category of genes that was highly over-represented in the early stage compared to the middle and late stages corresponded to genes related to the cell cycle & DNA processing category (19 in early stage, with 18 transcripts up-regulated, and 4 in middle stage with 3 genes up-regulated), probably reflecting the high cell division activity observed in the endosperm that precedes the development of the embryo. In solanaceous species, embryo division starts when the endosperm has reached 24-48 cells [11]. For the protein biosynthesis category, 18 genes were differentially expressed in early stage with a large proportion of up-regulated genes (14 transcripts). An inverse situation was found in the middle stage, where only 3 transcripts were up-regulated out of 15 genes for this functional category. Interestingly, for the protein fate group, 3 genes out of the 11 found were up-regulated in the early stage, while 9 genes out of 13 were up-regulated in the middle stage, suggesting a coordinate inverse regulation of protein

biosynthesis and fate during these two stages. For the following gene categories corresponding to development and biogenesis (15 in early and 19 in middle), and transport (12 in early and 38 in middle), the number of transcribed genes was higher in the middle stage when compared to the early stage.

Genes associated with nutrient reserves tend to specify the late stage, as expected. The defense and stress responses category was important in all three stages. Differentially expressed genes associated with the signal transduction and regulation of transcription classes were almost nonexistent in the late stage, comprising only 0.13% (1 out of 754 genes) of the transcripts while they comprised 5.57% (42 out of 754 genes) and 6.36% (48 out of 754 genes) in early and middle stages respectively.

#### Isolation of stage-specific genes

To characterize in more detail the transcriptional changes taking place between developmental stages, a different analysis was undertaken for each time point during embryogenesis from 0 to 22 DAP. A Volcano Plot with Student's t-test analysis was used to identify genes that were specifically expressed at only one time point when compared to UO (2 DAP vs UO, 4 DAP vs UO, etc) during embryogenesis, with a p value  $\leq 0.01$  and a  $\geq \pm 2$  fold cut-off. Depending on the time point chosen, from 217 to 735 genes were significantly modulated at any single time point and between 3 and 194 genes were specifically modulated during a single time point (Figure 6). Thus, these represent candidates for stage-specific or stage-predominant expressed genes. Of particular interest, four time points (2, 4, 8 and 16 DAP) had a high proportion (from 1/8 to 1/4) of their genes specifically modulated only at the observed time point. In the early zygotic phase at 2 DAP, 56 of 257 modulated genes (22%) were only expressed at this stage. At 4 DAP, when ovules bear embryos with a 2-cell embryo proper, 86 of 458 modulated genes (19%) were stage-specific. The last two time points enriched in stage-specific genes are found at the beginning and the end of the middle phase at 8 and 16 DAP, respectively. At 8 DAP, when ovules bear mostly embryos undergoing transition from a 4-cell to an 8-cell embryo proper 73 of 576 genes modulated genes (13%) were stage-specific while at 16 DAP, when ovules bear mostly heart-stage embryos, 194 of the 735 modulated genes (26%) were stage-specific. Thus, although the majority of the genes were differentially expressed during several embryo stages, genes differentially expressed only at a single time point were also found for all stages in variable numbers. Additional file 6 provides the list of these stage-specific genes. At 2 DAP, half the stage-specific modulated genes were classified as part of the functional category covering translation. At 4 DAP, we noted the



**Figure 5** Functional annotations of the 955 genes (passing the significance filter:  $P < 0.01$  and 2 fold cut-off) during the three major embryo developmental stages. The color keys show the designated gene class annotations.

accumulation of the stress and defense related genes including heat shock proteins (12.34%) and genes coding for proteins with binding function (9.87%). The major functional category represented at 8 DAP was related to energy (12.12%). A specific group of cell cycle regulation (3.03%) and chromosome organization and biogenesis (6.06%) were also found. Finally, the 16 DAP time point was specially enriched in three gene ontology (GO) categories: hormone related genes (4.84%), signal transduction (6.06%) and transport (6.06%).

#### Functional classification of differentially expressed genes in style, anther and leaf

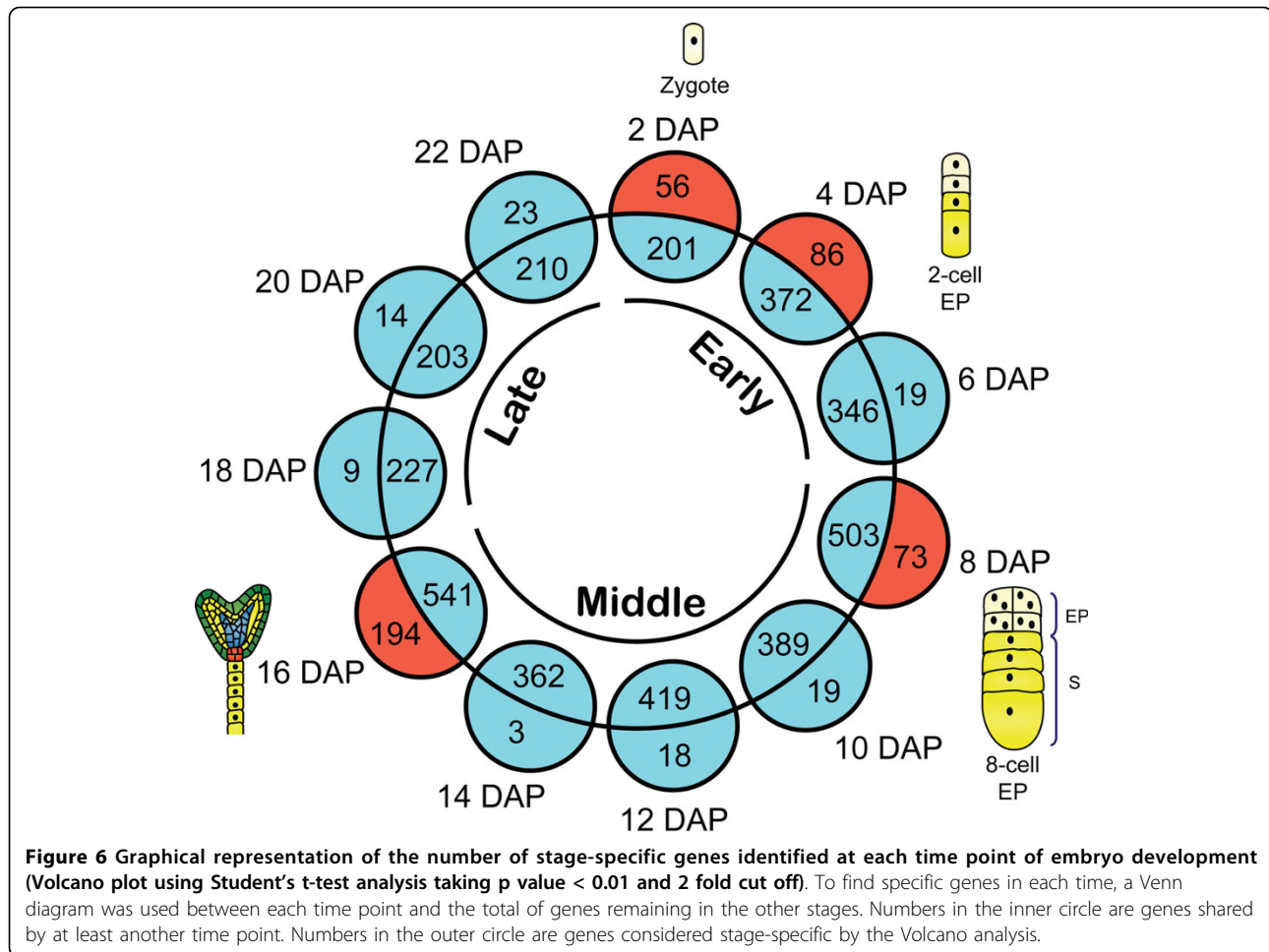
The 558 genes that were differentially expressed in other tissues when compared to UO were grouped into different functional categories (Additional file 3) according to their predicted gene products, based on the Gene Ontology (GO) Consortium through *S. chacoense* sequences and their closest orthologs in *Arabidopsis* (Figure 7). Apart from a large class of unknown and unclassified proteins (22.14%), the largest functional groups observed were protein synthesis (16.78%), metabolism (15.71%),

energy (7.14%), stress and defense related genes (9.46%), development and biogenesis (6.42%), transport (4.82%), regulation of transcription (4.28%), signal transduction (3.75%), proteins with binding function (3.57%), cell cycle and DNA processing (1.6%), and protein fate (1.25%).

Comparison of the categories found in these tissues and in ovules showed a completely different enrichment in functional categories (compare Figures 5 and 7). Whereas globally in ovule tissues (all stages taken together), we observed a large proportion of genes related to signal transduction and transcription, a very active cell cycle, and an up-regulation of genes associated with protein synthesis, in the other tissues tested these categories were less represented and even in the most represented category, protein biosynthesis, almost all genes were down-regulated (92 out of 94 genes).

#### Validation of microarray data by real-time PCR

In order to confirm the validity of differentially expressed genes identified by cDNA microarray analysis, we performed real time RT-PCR experiments on candidates from the three major groups of genes that specify early,



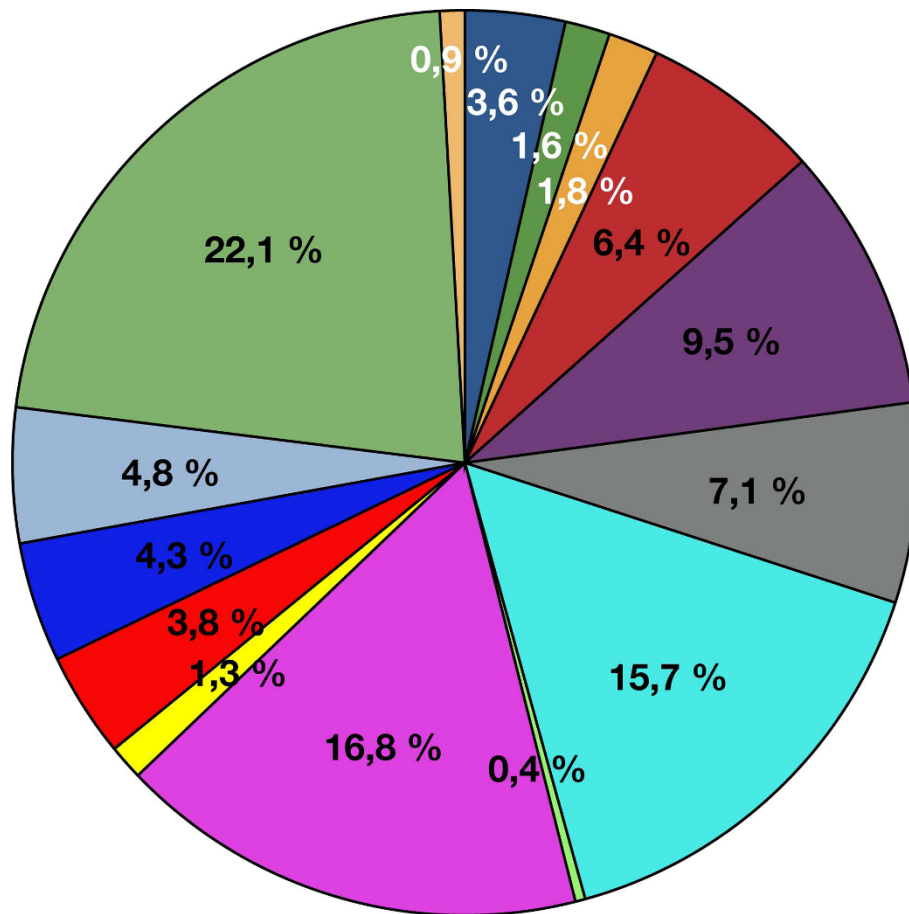
middle, and late stages of embryogenesis. Quantitative RT-PCR analyses were performed on RNA extracted from ovules four days after pollination corresponding to the early stage of embryo development; from ovules 12 days after pollination corresponding to the middle stage; and from ovules 22 days after pollination corresponding to the late stage. Unfertilized ovules were used as the calibration tissue and the ubiquitin amplicon as the internal control to normalize the data. Quantitative RT-PCR confirmed the expression pattern for all the nine genes chosen from the microarray analysis: DN980725, DN981910, DN978427, and DN983138 that specify the early stage of the embryo development; DN976716, DN983239, and DN978469 which specify the middle stage; DN979177 and DN976898 which specify the late stage (Figure 8). In all cases, the quantitative RT-PCR validation indicated that differential expression detected by microarray experiment was highly reliable.

### Discussion

Until recently, most methodologies used to study the molecular mechanism involved in plant embryogenesis

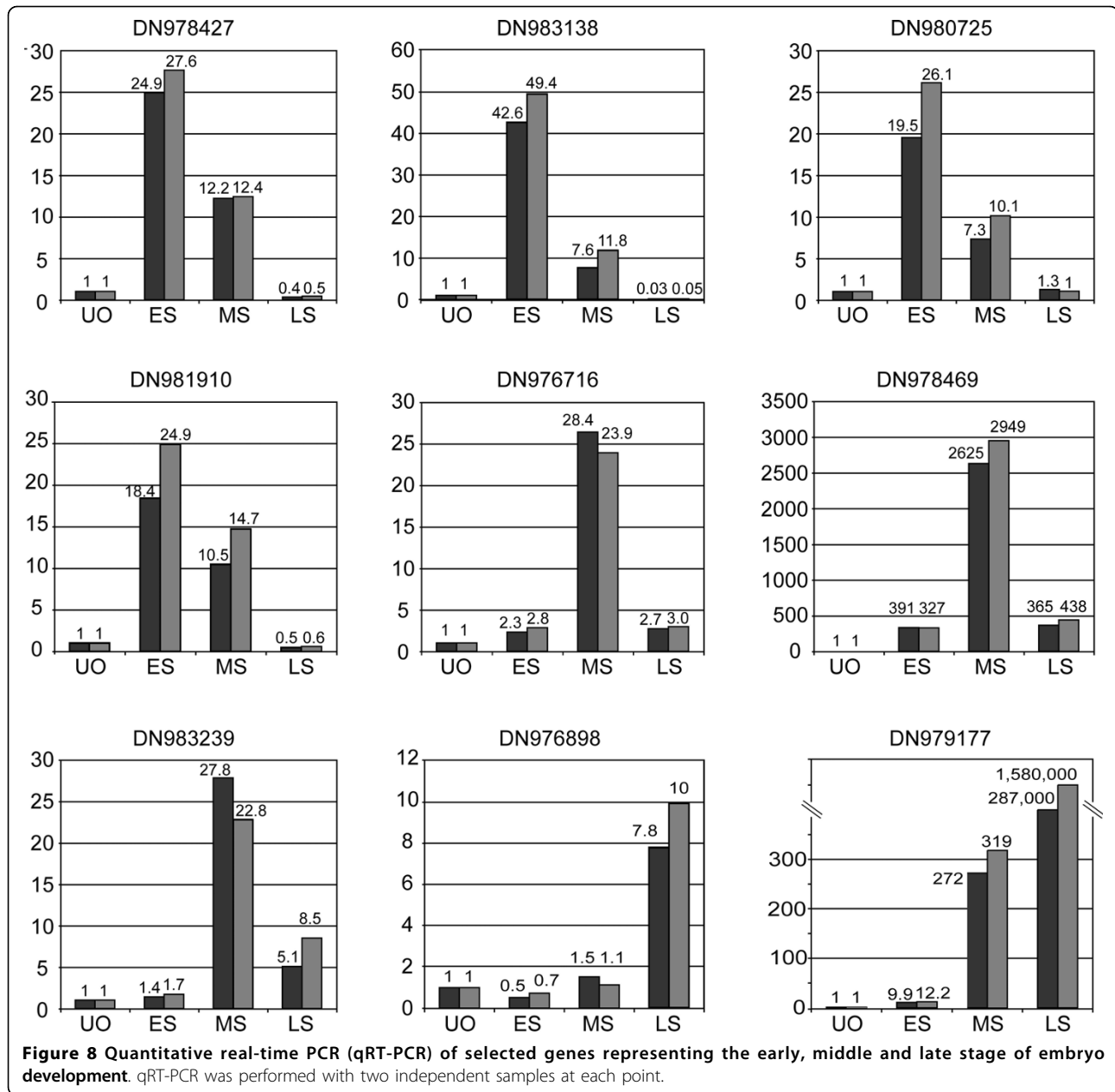
were based on mutagenesis approaches, and were mainly conducted in a few plant species, namely, *Arabidopsis*, rice and maize. These have allowed the characterization of numerous and informative mutants affecting early embryogenesis [19,21,22,66-73]. In recent years, candidate gene studies involving gene-profiling technologies have enabled the global visualization of spatial and temporal differential gene expression patterns, thus being complementary to mutagenesis approaches while also being able to characterize new key genes inaccessible through mutant screens. Only few such studies targeting plant embryogenesis have been conducted [36-38,42,43]. In this study, we have analyzed for the first time a large transcriptomic dataset of temporal gene expression in ovules from a solanaceous species, *S. chacoense*, by comparing the gene expression profiles of fertilized ovules versus unfertilized ovules.

Apart from slower embryo development, two other major differences between the model plant *A. thaliana* and *S. chacoense* should be considered. Both refer to early embryogenesis stages. Contrary to *A. thaliana*, which follows the Onagrad type, *S. chacoense* follows



- Protein with binding function
- Cell cycle and DNA processing
- Cell fate
- Development and biogenesis
- Defense and stress response
- Energy
- Metabolism
- Nutrients reserves
- Protein biosynthesis
- Protein fate
- Signal transduction
- Regulation of transcription
- Transport
- Unclassified and unknown protein
- Hormone related genes

**Figure 7** Distribution of 558 differentially expressed (by 2 fold-change,  $P < 0.01$ ) genes in anthers styles and leaves in various biological process categories. The color key shows the designated gene class annotations.



the Solanad type of early zygote divisions. In both cases, the first division of the zygote is transverse and asymmetrical, giving rise to two cells of different developmental fates. The smaller apical cell gives rise to the embryo proper, while the larger basal cell develops into the suspensor. In *A. thaliana* the first two divisions of the apical cell are longitudinal, while they are transversal in *S. chacoense* [74]. Endosperm also differs in whether nuclear divisions are followed by cellular divisions. Three variants are commonly recognized and classified as cellular, nuclear or helobial [75]. In cellular endosperms, as found in *S. chacoense*, cell-wall formation follows the first division of the primary endosperm nucleus

(PEN). In helobial endosperms, wall formation follows the first PEN division, producing two chambers, which vary in subsequent cellularization. In nuclear endosperms, as found in *A. thaliana*, walls do not develop between the free nuclei.

We have covered the whole embryo development process from zygote to mature embryo and have identified 955 genes ( $\geq \pm 2$  fold change,  $p \leq 0.01$ ) that are specifically or predominantly expressed in ovules compared to their expression profile in leaves, styles or anthers. Many differentially expressed genes encode proteins with putative regulatory functions, and most of them have not yet been characterized. Overall, in ovule

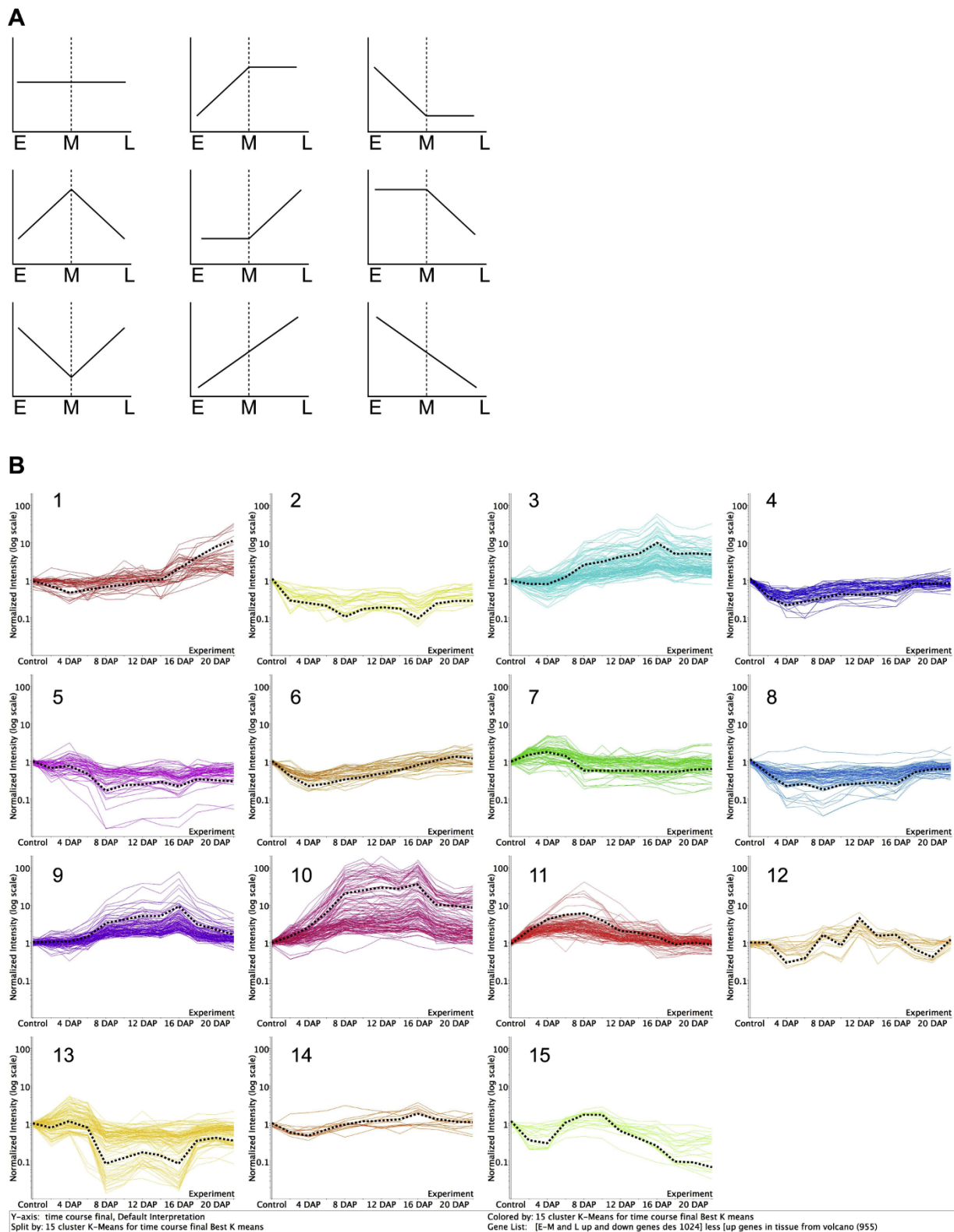
tissues, a large proportion of genes related to transport, signal transduction and regulation of transcription were modulated, and an up-regulation of genes associated with cell cycle and protein synthesis was noticed. In the other tissues tested (style, anthers and leaves), these categories were less represented and, even in the most represented category, protein biosynthesis, almost all genes were down-regulated (92 from 94 genes). The data obtained indicates that some gene categories are overrepresented in some tissues or organs. Interestingly, we found that expression of MADS-box, MYB, AP2-EREBP and YABBY transcription factor families were overrepresented in the transcriptional regulation groups during the early and middle stages of embryo development as well as in anthers tissues, but absent in the leaves and style. This suggests that these genes might play an important role in defining structural and functional identity of reproductive tissues. Consistent with this, many transcription factors from the above-mentioned gene families have been previously shown to play major roles during plant reproduction [36,76] and floral organ development [77-81]. Therefore, genes identified as being preferentially expressed in a given tissue probably have an important function in this tissue [26,82-85].

#### **Specific functional category enrichment in the three major embryonic developmental stages**

Differentially expressed genes in ovules were grouped according to the similarity of their expression profiles using unsupervised hierarchical clustering algorithms. Clustering results segregated the dataset amongst three major groups of genes that specified early (from zygote to 16-cell stage), middle (16-cell to heart stage) and late (heart to mature embryo) stages of embryogenesis (Figure 2). This result indicates a clear distinction in transcriptional profiles between early, middle and late stages of embryogenesis, as determined with the principal components analysis (Figure 3), the little overlap between the modulated genes (Figure 4), as well as the specific enrichment in the corresponding functional categories (Figure 5). K-means clustering using Pearson correlation was also used to separate the expression patterns during the embryo developmental time course. Considering that three major expression profiles are found, representing early, middle and late stages of embryonic development, a maximum of nine patterns can be expected (Figure 9a). Hennig et al. (2004) also proposed nine models of dynamic expression patterns for genes involved in reproductive development in *Arabidopsis* based on three different stages: before, during and after pollination [36]. Spencer et al. (2007) found that seven distinct expression patterns were present within the apical and basal section of the embryo, during the globular, heart, and torpedo stage embryos [37].

The two missing profiles are most probably the consequence of the absence of earlier stages, before the globular stage. When considering less global trends, as depicted in Figure 6 for stage-specific genes, a higher number of specific and slightly different patterns can be obtained, based on the 12 individual time points. Figure 9b depicts 15 such patterns that can be derived from the analysis of our time-course every second day from 0 to 22 DAP and that include the nine dynamic expression patterns representing early, middle and late stages of embryonic development. Although the majority of functional groups are shared between several stages, some of these clusters show specific functional categories enrichment (Additional file 6). The first cluster (cluster 1) includes genes involved in protein storage and specifies the late stages of embryogenesis. Cluster 7 corresponds to the early stage (2 and 4 DAP) and is enriched in genes classified in the protein biosynthesis category. Cluster 11 shows a peak at 8 DAP and is enriched in genes of the nucleosome assembly and cell cycle regulation group. Clusters 3 and 9 show a profile with a pronounced peak at 16 DAP, corresponding to the transition from heart to torpedo stages. These two clusters are enriched in functional categories corresponding to proteins with binding function, lipid metabolism, transport and genes implicated in development and biogenesis. Cluster 10 represents the transition of the early stage to middle (8 DAP) and middle to the late (16 DAP) and is characterized by a large number of genes involved in proteolysis, lipid metabolism and transport, development and protein folding.

Based on the global trend from the three major clusters specifying early, middle and late stages of embryogenesis, we found that the transcriptional pattern of some functional categories could also specify embryo developmental stages. These included metabolism, transcriptional regulation, cell cycle, protein biosynthesis, defense and stress response, development and cell growth, signal transduction, transport, proteins with binding function, protein fate and energy, which are represented in early and/or the middle stage. The late embryonic stage was characterized, in addition to some groups found in the early and middle stages, by the cell fate and nutrient reserves functional categories (Figure 5). A comparison with the closest study published by Spencer and collaborators (2007) in *Arabidopsis*, shows a larger representation of functional groups in our dataset. Spencer and colleagues found 1872 and 1226 differentially expressed genes in the apical and basal time course, respectively, using a p value of 0.05 for genes whose expression was significantly different from a value of 1. They then chose the 200 most highly expressed genes (100 in apical and 100 in basal tissues) between the developmental stages studied. In our



**Figure 9 K-means clustering using Pearson Correlation (GeneSpring version 7.2) during an embryo development time course. A,** Models of dynamic gene expression patterns representing all the nine different patterns found based on the three point time (early, middle and late). **B,** Clustering of specific differentially expressed genes in the ovules based on 15 clusters representing all the stages of embryonic development from 0 to 22 DAP. The black discontinues line represent the average of gene profile.

dataset, we used different criteria, a p value of 0.01 and cut off of  $\geq \pm 2$  fold in each time point compared to the same control. Our GO groups were assigned to all the genes that were specific and differentially expressed in the early (zygote to 16 cell), middle (globular to heart) and late (torpedo to mature embryo) stages of embryogenesis. While in the *Arabidopsis* published data [37], the transition from globular to heart stage is characterized by the up regulation of genes in functional groups corresponding to energy production, metabolism, protein biosynthesis, signal transduction and transcription, in our dataset, the middle stage, which include this transition, is enriched in the same functional groups except for the protein biosynthesis category which characterized the early embryonic stage (from zygote to almost the globular stage). In addition, other enriched functional categories were up-regulated in the middle stage including protein with binding function, development and cell growth, protein fate and transport. The late stage was characterized by the up-regulation of the nutrient reserve category. It was reported that the storage proteins typically accumulate only from the mid-maturation phase onwards, when cell division is completed and the basic form of the embryo has developed [10]. Given that the plant body pattern is already established and that the seedling continues to mature, it was expected to have similar transcriptional profile from torpedo to mature embryo and have an up regulation of nutrient reserves at all these late stages, which is indeed what we observed. Thus, during embryo development, some functional group characterized specific stages while other groups are shared between different embryogenesis stages. Indeed, the up regulation of genes related to cell cycle and protein biosynthesis characterized the early stage; protein with binding function, protein fate, cell-cell communication and transcription specified the middle stage; and reserve proteins were more abundant in late stage. Energy and transport functional groups were represented in the early and middle stages while stress and defense related genes were highly represented during all developmental stages.

Protein biosynthesis, including ribosomal proteins (RP) and components of the translation machinery, were up-regulated in the early embryonic stage but then down-regulated in the middle stage. Spencer and colleagues [37] found a similar trend but it appears to be delayed compared to our data as it shows a significant up-regulation of genes involved in protein synthesis between the heart stage and the torpedo stage in their apical embryo time course, and between the globular and the heart stage in their basal embryo time course. These embryo developmental stages would be included in our middle stage, where we observed a decrease in protein synthesis components when compared to our early stage. This

could be representative of the transition from early embryogenesis to the maturation and protein accumulation phase typical of the later embryogenesis stages [10]. Conversely, protein fate genes were down-regulated in the early stage but up-regulated in the middle stage. Genes belonging to the cell cycle category were exclusively up-regulated in the early stage. Up-regulation of the protein biosynthesis category specified the second and fourth day after pollination, afterward, these genes were either down-regulated or showed no differential expression during the later time points. This finding reflects the need to produce sufficient amounts of these essential proteins and the indispensability of protein synthesis in the actively developing seeds. Conversely, genes involved in protein degradation were overrepresented later in middle and late stages. Spencer et al. (2007), found the same enrichment order, while in the basal time course, from the globular to the heart stage, they showed the up regulation of protein biosynthesis, then the protein fate group was overrepresented from the heart to the torpedo stage and the protein biosynthesis group was absent. The over-expression of these particular categories of genes illustrates the high energetic requirements of embryo tissue and suggests that other regulatory mechanisms at the post-transcriptional and posttranslational levels are important during the embryogenesis. Consistent with this, other interesting genes were found including genes coding for ubiquitin, and proteinase inhibitors. Indeed, transcriptional and post-transcriptional control have been described for the proteinase inhibitors of soybean [86] and other legume seed proteins [87]; and ubiquitination was already found to be an important post-translational regulatory process of the unfertilized egg cell in wheat [88].

Examination of the cell cycle and DNA processing categories revealed a prevalence of gene up-regulation from 2 to 16 DAP. B-type cyclin-dependant kinases (CDKs) involved in the control of the G<sub>2</sub>/M progression in plants [89] as well as a large number of S-phase markers such as Histone 1, 2, 3 and 4, involved in DNA replication and modification [90-92] were significantly enriched and up-regulated. This is not unexpected since cell division in the embryo starts after fertilization and continues until the heart stage [8]. It has also been shown that protein synthesis and the cell cycle are associated at two levels [93]. On the one hand, protein synthesis is required for entry and progression through the cell cycle; on the other hand, protein biosynthesis is itself modified during cell division steps. This finding is in line with our data set. In the early stage, the up-regulation of RPs, translation and elongation factors coincide with the overexpression of histones. Markedly, in the middle stage, proteolysis genes represented by a large number of ubiquitin/proteasome-related genes, proteases, peptidases, and subtilases, were activated,



suggesting active degradation of proteins by proteolysis, while protein biosynthesis genes remain expressed at a basal, non-modulated level. Importance of ubiquitine/proteasome-related genes during embryogenesis has also been reported to strongly affect the progression of the *Arabidopsis* embryo from the globular stage onward [94]. Altogether, this suggests important post-translational protein regulation in cell cycle progression during embryogenesis.

The set of differentially expressed genes modulated during the early and middle phases also contained many signal transduction components and transcriptional regulators. Among them were 24 signaling proteins and 50 transcriptional regulators. In the signal transduction category, one calmodulin, most similar to the *Arabidopsis* CaM7 [95], and one calmodulin-like gene, most similar to the *AtCML24* [96] were isolated. Interestingly, three *Arabidopsis* SIP4 orthologs (SOS3 kinase-interacting protein 4 or CBL-interacting protein kinase 11 (CIPK11), also characterized in tobacco as ACRE216 - Avr9/Cf-9 rapidly elicited protein 216) [97,98] and an ortholog of the *Arabidopsis* CBL-interacting protein kinase 5 (CIPK5) were also isolated, suggesting an important role for protein phosphorylation and calcium signaling during sexual reproductive development. Of the protein kinases modulated during embryogenesis, only one coded for a receptor-like kinase, *ScORK11* [47]. Interestingly, from the EST pool that was used to make the cDNA microarray, we had previously isolated 30 different RLKs with 28 being predominantly expressed in ovary tissues or young developing fruits, and 23 being transcriptionally induced following fertilization [47]. This discrepancy highlights the difference in sensitivity of the methodologies used, e.g. microarray analysis (this study) vs. quantitative RT-PCR analysis [47]. A shaggy-like kinase gene similar to the *Arabidopsis* *ATSK41* [99], an IRE-Like AGC kinase gene [100] and an ortholog of the *Arabidopsis* RAB GTPase homolog *B1C* [101] were also isolated. Most of these have been previously shown to be involved in various stress responses and none had been shown to play a role in embryogenesis. Considering the high number of differentially expressed genes (100/955) classified as involved in defense and/or stress response during all stages of embryogenesis, this also emphasize their potential roles in response to pollination and during reproductive development.

Several genes implicated in embryonic patterning were found in the first two major stages. Amongst these, PROTODERMAL FACTOR 1 (*PDF1*) and TOPLESS-RELATED 1 (*TPR1*) are up-regulated in the early and in middle stage, respectively. *PDF1* is a predicted cell wall protein restricted to the protoderm of the embryo following tangential divisions between the 8 and 16-cell stages [102] and specifies the first element of the radial pattern [12]. Since the *S. chacoense* *PDF1* ortholog

shares the same expression pattern as the *Arabidopsis* *PDF1* with an expression peak at the 16-cell stage (2.24-fold at 8-cell and 3-fold at 16-cell stages) this protein most probably exerts the same function in solanaceous species. The *TPR1* gene is involved in maintenance of shoot fates during *A. thaliana* embryogenesis defining the apical pole of embryo [103]. Transcriptional activation of this protein in our study suggests that the *S. chacoense* *TPR1* ortholog is developmentally regulated during embryogenesis and might also act as a determinant of embryo polarity in solanaceous species.

Many members of the MADS box family are expressed at higher levels during embryogenesis [104]. Interestingly, a protein identified as an ortholog of the *A. thaliana* *AGL11/SEEDSTICK* gene was up-regulated during the middle stage. *AGL11/STK* is required for normal development of the funiculus, an umbilical-cord-like structure that connects the developing seed to the placenta, and for dispersal of the seeds when the fruit matures [105,106]. Furthermore, *AGL11/SEEDSTICK* is also expressed during embryogenesis in ovules as determined from publicly available microarray data (BAR, <http://www.bar.utoronto.ca>) with a high activity in stage 3 seeds (globular stage of embryo development). Then its expression decreases during the middle stage and increases again in the late stage of seed development. Rounsley et al. also showed that *AGL11/SEEDSTICK* expression is also maintained after pollination through late seed development [76]. MADS-box genes are known to act as homeotic selector genes determining floral organ identity and also as floral meristem identity genes [77,78,80,81]. Our profiling study highlights the potential role of MADS-box transcription factors during the early stages of reproductive development. Several others studies showed an important differential expression of transcription factors, including MADS-box, from various families during pollination, fertilization and seed development [36-40,107-109].

#### **Biotic and abiotic stress-responsive genes**

More than 100 genes out of the 955 differentially expressed genes are classified as potentially involved in defense and/or stress response during all stages. This suggests that these defense-related genes might also function in response to pollination and during embryo development, in addition to their contribution in defense against pathogens [110,111]. Among activated genes in our dataset, those corresponding to proteins of the subtilase family were induced by two- to 137-fold. An ortholog of the *ARA12* Subtilisin-like serine protease showed a two- to 73-fold change in *S. chacoense*. Gollack and colleagues showed by *in situ* hybridization that the *ARA12* subtilisin-like serine protease was present at higher level in pistils, ovules and anthers, but also

responded to stress and pathogen stimuli [112]. Several other pathogenesis-related proteins were also up-regulated after pollination during embryo development, including proteinase inhibitors, Thaumatin-like protein, PR10,  $\beta$ -1,3-glucanase-like protein and elicitor inducible chitinases. A strong increase in transcript levels, from 2- to 118-fold, of non-specific lipid transfer proteins (nsLTPs) has been found during the middle and late stages. These proteins are member of a superfamily of seed protein called prolamins that, in addition to storage, plays important roles in plant defense responses [113,114]. The nsLTPs have been reported in a number of plants, including solanaceous plants [115], and are involved in lipid transfer and in the formation of a protective covering of cutin and suberin layers over plant surfaces. Proteomics analysis of embryo and endosperm from germinating tomato seeds revealed a high level of nsLTPs in the tomato endosperm suggesting a role in the mobilization of lipids from the endosperm to the embryo and probably in defense against infection during germination [116]. Recently, Vriezen et al (2009) also reported 54 stress or defense-related genes that were expressed in tomato ovary tissues prior to fertilization with the vast majority (47/54) being down-regulated after fertilization. In our case, in 2 DAP ovaries compared to unfertilized ovaries, 54 stress or defense-response genes were regulated, with 34 down-regulated genes. Furthermore, when looking at the whole profile from 2 to 22 DAP, a much more dynamic and variable profile for these genes is observed, precluding any global generalization over the whole time-course. The activation of defense-related genes during embryogenesis is probably due to the fact that the plant protects embryo development by increasing the activation of defense and stress-related genes. Alternatively, some of these stress-related genes might exert other functions as for the POP2 gene that is also involved in pollen tube adhesion and growth [117,118] and in oxidative stress response [119], and for the HSP70-1 chaperone, involved in development and abiotic stress responses in *Arabidopsis* [120]. Altogether, this suggests an overlap between fertilization/embryogenesis and stress and defense response pathways. This was also reported by Lan et al. (2005) in rice, revealing a cross talk in genetic programs controlling pollination/fertilization and stress responses [121].

#### Identification of stage or transition specific candidate genes during embryo development

To identify stage-specific or transition-specific candidate genes we used a Volcano Plot analysis to identify genes that were specifically expressed at only one time point during embryogenesis with a p value  $\leq 0.01$  and a  $\geq \pm 2$  fold cut off. As mentioned before, 4 time points (2, 4, 8 and 16 DAP) had a high proportion of stage-specific or

stage-predominant candidate genes. At 16 DAP, when ovules bear mostly heart-stage embryos, 26% of modulated genes were stage-specific and were enriched in up-regulated genes from the signal transduction, transport, and hormone-related genes gene ontology (GO) categories, when compared to the GO profiles of the other time points. Of the genes orthologous to the *Arabidopsis* EMB genes required for normal embryo development [22], six (*EMB 2171* - 2 DAP, *EMB 1738* - 4 DAP, *EMB 2473* - 4 DAP, *EMB 1473* - 8 DAP, *EMB 1345* - 16 DAP, and *EMB 1990* - 16 DAP) were found to be stage-specific in the Volcano analysis. *EMB 2171* and *EMB 1473* code for the mitochondrial targeted 60S L17 RP and the chloroplastic 50S L13 RP, respectively. *Arabidopsis* mutant plants corresponding to these genes display embryos arrested at the globular stage (<http://www.seedgenes.org/index.html>). These two genes had been previously shown to be highly expressed during early embryo development in *S. chacoense* [122]. *EMB 1738* codes for the *CYP51G1* obtusifoliol 14- $\alpha$  demethylase gene, a gene involved in lipid signaling [123]. In *emb 1738* plants, two-thirds of the embryos are arrested at the early transition to heart stage [22]. *EMB 2473/MIRO1* codes for a GTPase required for embryogenesis that also affects pollen mitochondria morphology [124]. *MIRO1* mutants display embryos arrested at the zygotic to the 4-cell stage. *EMB 1345* codes for a WD40 protein similar to the CIA1 protein involved in iron/sulfur protein biogenesis [125], that interacts with the Willm's tumor suppressor protein WT1 [126]. In *emb 1345* mutants, only 3% of the analyzed plants displayed globular embryos while the remaining plants had no detectable embryos (<http://www.seedgenes.org/index.html>). The *EMB 1990* gene codes for an chloroplastic putative integral membrane protein [127]. In *emb 1990* mutants, two-thirds of the embryos are arrested at the globular stage while 13% have a cotyledon terminal phenotype (<http://www.seedgenes.org/index.html>).

Two families of auxin-regulated genes (the *Aux/IAA*, and *GH3* families) have been identified among the transcripts that were specifically found in 16 DAP heart-stage embryos. Auxins are important signaling molecules involved in many developmental processes in plants. Identification of embryo-defective mutants affected in the transport or the auxin signaling pathway indicates that auxins play various roles in coordinating the organization of the embryo by providing positional information [128]. During embryogenesis, it is particularly important in the specification of the apical cell after the first zygotic division [129], the formation of the root meristem from the hypophysis [130-132], and the establishment of bilateral symmetry [133]. Apical-to-basal transport of the hormone auxin is initiated at the early globular stage and plays a key role in regulation of many aspects of

embryonic pattern formation in plant [134,135]. In dicots, auxin is first found in the endosperm and only later is it detectable in the embryo itself with a special distribution starting with a maximal accumulation in the apical cell until the 32-cell-embryo stage. At later stages an inverse pattern of auxin maxima is observed [129,136,137]. Amongst the auxin-regulated genes found in our transcriptomic analysis, we noticed the up-regulation of the *GH3* gene, *YDK1* (*YADOKARI 1*). Takase et al (2004), found a dwarf mutant, named *ya do Kari 1-D* (*ydk1-D*), which had a T-DNA insertion proximal this *GH3* gene. The *ydk1-D* mutant is dominant and has a short hypocotyl, a short primary root, reduced lateral root number, and reduced apical dominance, suggesting that *YDK1* may function as a negative component in auxin signaling by regulating auxin activity [138]. A second gene of interest is *IAA13*, a member of the auxin response factors family which is required for the axialization of the embryo [128,139-143]. Weijers et al, showed that stabilization of BDL/IAA12 or its sister protein IAA13, prevents MP/ARF5-dependent embryonic root formation. A dominant mutation that renders IAA13 insensitive to auxin-dependent degradation results in elimination of embryonic axis: the cell at the base of the proembryo divides abnormally and never elongates or become organized in files. The seedlings produce a short peg in place of the hypocotyl and root and also show fused cotyledons with reduced vasculature. *IAA13* is initially expressed in the apical daughter of the zygote and in all cells of the proembryo, but it becomes restricted to the provascular cells by the midglobular stage [128]. Although the shoot apical meristem (SAM) gradually develops during embryogenesis, it appears as a distinct histological entity somehow late during embryogenesis [144], around the time where we observed peak expression of numerous auxin-related genes.

## Conclusion

Our analysis provides the first study, to our knowledge in a solanaceous species, of the transcriptional program that takes place during the early phases of plant reproductive development, including all embryogenesis steps. We used a comparative expression profiling strategy between fertilized and unfertilized ovules to provide an insight into the embryo transcriptome in *S. chacoense* and to identify a subset of genes comprising this transcriptome. Several potential regulators of fertilization and early seed development have been identified. We identified 1024 genes (955 genes were specifically expressed in ovules tissues when compared to others tissues) that were specifically regulated during these developmental stages, along with many highly stage-specific genes. Although the biological function of most genes remains to be determined, the identification of genes involved in reproductive processes

in solanaceous species that produces embryos much more slowly than model species like *A. thaliana*, will enable the selection of stage-specific genes that will deepen our understanding of complex stage transition processes as well as pinpoint potential targets for improving yield and seed quality by conventional breeding or biotechnological approaches.

## Additional material

**Additional file 1: Quantitative PCR primers.** Primer sequences used in the real time quantitative PCR expression analysis of selected genes representing candidates from the early, middle, and late stages of embryogenesis shown in figure 8.

**Additional file 2: Differential expression of 1024 genes amongst during embryogenesis.** Using Anova testing along with a Benjamini and Hochberg multiple testing correction algorithm, 1024 transcripts showed a statistically significant change in abundance ( $p$  value  $\leq 0.01$ ) and  $\pm 2$ -fold variation in expression in at least one of the time points during embryo development compared to the unfertilized ovule control.

**Additional file 3: Differential expression of genes in anthers, style, and leaf tissues compared to unfertilized ovules.** RNA samples from leaf, anther, and style tissues of *S. chacoense* were compared to the unfertilized ovules (UO) pool sample. Each microarray experiment was performed in quadruplicate using, in all cases, independently isolated RNA samples as starting material (four biological replicates). A total of 558 genes (262 genes were up-regulated and 296 genes were down-regulated) showed a significant variation in transcript abundance ( $\geq \pm 2$ -fold coupled with a  $p$  value  $\leq 0.01$ ) in at least one tissue (leaf, style, or anther) when compared to UO.

**Additional file 4: Differential expression of genes during embryo development showing a greater than  $\pm 10$ -fold change.** Using Anova testing, along with a Benjamini and Hochberg multiple testing correction algorithm, 135 transcripts showed a statistically significant change in abundance ( $p$  value  $\leq 0.01$ ) and  $\pm 10$ -fold variation in expression in at least one time point during embryo development compared to the unfertilized ovule control.

**Additional file 5: Differential expression of genes amongst the three major stages (early, middle, and late) of embryogenesis.** Using Anova testing along with a Benjamini and Hochberg multiple testing correction algorithm, 1024 genes were differentially expressed ( $\geq \pm 2$  fold change,  $p$  value  $\leq 0.01$ ) in fertilized ovules compared to unfertilized ovules. Only limited expression overlap was observed between these genes and the genes expressed in the other tissues tested, with the vast majority of the fertilization-regulated genes (955) being specifically or predominantly expressed in ovules. The hierarchical gene clustering analysis showed that, following fertilization, the ovule-expressed gene pool could be clearly divided into three major groups that specified early, middle, and late stages of embryogenesis. A Venn diagram analysis of the genes predominantly expressed in fertilized ovules indicated that 298 genes are differentially expressed during the early developmental stage, 395 genes only during the middle stage, and 61 genes only during the late stage. Sixty-six (66) genes are differentially expressed in early and middle stages, 28 genes in early and late stages and 47 genes in middle and late stages. Sixty (60) genes are differentially expressed in all embryo development stages

**Additional file 6: Differential expression of stage-specific genes during each time point of embryo development (from 2 to 22 DAP).** To characterize the transcriptional changes taking place between developmental stages, a Volcano Plot with Student's t-test analysis was used to identify genes that were specifically expressed at only one time point when compared to UO (2 DAP vs UO, 4 DAP vs UO, etc) during embryogenesis from 0 to 22 DAP, with a  $p$  value  $\leq 0.01$  and a  $\geq \pm 2$  fold cut-off.

### Acknowledgements

We thank Gabriel Téodorescu for plant care and maintenance. F. Tebbji was the recipient of a Ph. D. fellowships from Le Fonds québécois de la recherche sur la nature et les technologies (FQRNT). The Natural Sciences and Engineering Research Council of Canada (NSERC), FQRNT, and the Canada Research Chair program are also acknowledged for their financial support (D. P. Matton).

### Author details

<sup>1</sup>Institut de recherche en biologie végétale, Département de sciences biologiques, Université de Montréal, 4101 rue Sherbrooke est, Montréal, Québec, H1X 2B2, Canada. <sup>2</sup>Biotechnology Research Institute, National Research Council, 6100 Royalmount Avenue, Montreal, QC, H4P 2R2, Canada.

### Authors' contributions

FT carried out the experiments, interpreted the data, and drafted the manuscript. AN gave extensive advices on microarray data analysis and edited the paper. DPM conceived the research, directed the overall project, and played a major part in writing and results interpretation. All authors have read and approved the final manuscript.

Received: 21 December 2009 Accepted: 12 August 2010

Published: 12 August 2010

### References

- Drews GN, Yadegari R: **Development and function of the angiosperm female gametophyte.** *Annu Rev Genet* 2002, **36**:99-124.
- Russell SD: **Double fertilization.** *Int Rev Cytol* 1992, **140**:357-388.
- Russell SD: **Attraction and transport of male gametes for fertilization.** *Sex Plant Reprod* 1996, **9**:337-342.
- Guignard L: **Sur les anthérozoïdes et la double copulation sexuelle chez les végétaux angiospermes.** *C R Acad Sci Paris* 1899, **128**:864-871.
- Nawaschin SG: **Resultate einer Revision der Befruchtungsvorgänge bei Lilium Martagon und Fritillaria tenella.** *Bull Acad Sci St Petersburg* 1898, **9**:377-382.
- Mansfield SG, Briarty LG: **Early embryogenesis in Arabidopsis thaliana. II. The developing embryo.** *Canadian journal of botany* 1991, **69**:461-476.
- Goldberg RB, Barker SJ, Perez-Grau L: **Regulation of gene expression during plant embryogenesis.** *Cell* 1989, **56**:149-160.
- West M, Harada JJ: **Embryogenesis in Higher Plants: An Overview.** *Plant Cell* 1993, **5**:1361-1369.
- Thomas TL: **Gene expression during plant embryogenesis and germination: an overview.** *Plant Cell* 1993, **5**:1401-1410.
- Lindsey K, Topping JF: **Embryogenesis: a Question of Pattern.** *J Exp Bot* 1993, **44**:359-374.
- Dnyansagar VR, Cooper DC: **Development of the Seed of Solanum phureja.** *American Journal of Botany* 1960, **47**:176-186.
- Jenik PD, Gillmor CS, Lukowitz W: **Embryonic Patterning in Arabidopsis thaliana.** *Annual Review of Cell and Developmental Biology* 2007, **23**:207-236.
- Chandler JW, Cole M, Flier A, Grewe B, Werr W: **The AP2 transcription factors DORNROSCHE and DORNROSCHE-LIKE redundantly control Arabidopsis embryo patterning via interaction with PHAVOLUTA.** *Development* 2007, **134**:1653-1662.
- Griffith ME, Mayer U, Capron A, Ngo QA, Surendrarao A, McClinton R, Jurgens G, Sundaresan V: **The TORMOZ gene encodes a nucleolar protein required for regulated division planes and embryo development in Arabidopsis.** *Plant Cell* 2007, **19**:2246-2263.
- Lahmy S, Guilleminot J, Schmit AC, Pelletier G, Chaboute ME, Devic M: **QQT proteins colocalize with microtubules and are essential for early embryo development in Arabidopsis.** *Plant J* 2007, **50**:615-626.
- Vidaurre DP, Ploense S, Krogan NT, Berleth T: **AMP1 and MP antagonistically regulate embryo and meristem development in Arabidopsis.** *Development* 2007, **134**:2561-2567.
- Boisson-Demier A, Frietsch S, Kim TH, Dizon MB, Schroeder JI: **The peroxin loss-of-function mutation abstinence by mutual consent disrupts male-female gametophyte recognition.** *Curr Biol* 2008, **18**:63-68.
- Neuffer MG, Sheridan WF: **Defective Kernel Mutants of Maize. I. Genetic and Lethality Studies.** *Genetics* 1980, **95**:929-944.
- Sheridan WF, Neuffer MG: **Defective Kernel Mutants of Maize II. Morphological and Embryo Culture Studies.** *Genetics* 1980, **95**:945-960.
- Scanlon MJ, Stinard PS, James MG, Myers AM, Robertson DS: **Genetic analysis of 63 mutations affecting maize kernel development isolated from Mutator stocks.** *Genetics* 1994, **136**:281-294.
- Clark JK, Sheridan WF: **Isolation and Characterization of 51 embryo-specific Mutations of Maize.** *Plant Cell* 1991, **3**:935-951.
- Tzafirir I, Pena-Muralla R, Dickerman A, Berg M, Rogers R, Hutchens S, Sweeney TC, McElver J, Aux G, Patton D, Meinke D: **Identification of genes required for embryo development in Arabidopsis.** *Plant Physiol* 2004, **135**:1206-1220.
- Gray-Mitsumune M, O'Brien M, Bertrand C, Tebbji F, Nantel A, Matton DP: **Loss of ovule identity induced by overexpression of the fertilization-related kinase 2 (ScFRK2), a MAPKKK from Solanum chacoense.** *J Exp Bot* 2006, **57**:4171-4187.
- Gomez LD, Baud S, Gilday A, Li Y, Graham IA: **Delayed embryo development in the ARABIDOPSIS TREHALOSE-6-PHOSPHATE SYNTHASE 1 mutant is associated with altered cell wall structure, decreased cell division and starch accumulation.** *Plant J* 2006, **46**:69-84.
- Liu NY, Zhang ZF, Yang WC: **Isolation of embryo-specific mutants in Arabidopsis: genetic and phenotypic analysis.** *Methods Mol Biol* 2008, **427**:101-109.
- Colombo M, Masiero S, Vanzulli S, Lardelli P, Kater MM, Colombo L: **AGL23, a type I MAD5-box gene that controls female gametophyte and embryo development in Arabidopsis.** *Plant J* 2008, **54**:1037-1048.
- Nuccio ML, Thomas TL: **ATS1 and ATS3: two novel embryo-specific genes in Arabidopsis thaliana.** *Plant Mol Biol* 1999, **39**:1153-1163.
- Heck GR, Perry SE, Nichols KW, Fernandez DE: **AGL15, a MAD5 domain protein expressed in developing embryos.** *Plant Cell* 1995, **7**:1271-1282.
- Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization.** *Mol Biol Cell* 1998, **9**:3273-3297.
- Harmer SL, Kay SA: **Microarrays: determining the balance of cellular transcription.** *Plant Cell* 2000, **12**:613-616.
- Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM: **Coordinated plant defense responses in Arabidopsis revealed by microarray analysis.** *Proc Natl Acad Sci USA* 2000, **97**:11655-11660.
- Aharoni A, Keizer LC, Van Den Broeck HC, Blanco-Portales R, Munoz-Blanco J, Bois G, Smit P, De Vos RC, O'Connell AP: **Novel insight into vascular, stress, and auxin-dependent and -independent gene expression programs in strawberry, a non-climacteric fruit.** *Plant Physiol* 2002, **129**:1019-1031.
- Cheong YH, Chang HS, Gupta R, Wang X, Zhu T, Luan S: **Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis.** *Plant Physiol* 2002, **129**:661-677.
- Becker JD, Boavida LC, Carneiro J, Haury M, Feijo JA: **Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome.** *Plant Physiol* 2003, **133**:713-725.
- Hony D, Twell D: **Comparative analysis of the Arabidopsis pollen transcriptome.** *Plant Physiol* 2003, **132**:640-652.
- Hennig L, Grussem W, Grossniklaus U, Kohler C: **Transcriptional programs of early reproductive stages in Arabidopsis.** *Plant Physiol* 2004, **135**:1765-1775.
- Spencer MW, Casson SA, Lindsey K: **Transcriptional profiling of the Arabidopsis embryo.** *Plant Physiol* 2007, **143**:924-940.
- Lan L, Chen W, Lai Y, Suo J, Kong Z, Li C, Lu Y, Zhang Y, Zhao X, Zhang X, et al: **Monitoring of gene expression profiles and isolation of candidate genes involved in pollination and fertilization in rice (Oryza sativa L.) with a 10 K cDNA microarray.** *Plant Mol Biol* 2004, **54**:471-487.
- Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J: **Microarray analysis of developing Arabidopsis seeds.** *Plant Physiol* 2000, **124**:1570-1581.
- Zhang X, Feng B, Zhang Q, Zhang D, Altman N, Ma H: **Genome-wide expression profiling and identification of gene activities during early flower development in Arabidopsis.** *Plant Mol Biol* 2005, **58**:401-419.
- Wellmer F, Riechmann JL, Alves-Ferreira M, Meyerowitz EM: **Genome-wide analysis of spatial gene expression in Arabidopsis flowers.** *Plant Cell* 2004, **16**:1314-1326.

42. Lee JM, Williams ME, Tingey SV, Rafalski JA: **DNA array profiling of gene expression changes during maize embryo development.** *Funct Integr Genomics* 2002, **2**:13-27.
43. Casson S, Spencer M, Walker K, Lindsey K: **Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis.** *Plant J* 2005, **42**:111-123.
44. Wellmer F, Alves-Ferreira M, Dubois A, Riechmann JL, Meyerowitz EM: **Genome-wide analysis of gene expression during early Arabidopsis flower development.** *PLoS Genet* 2006, **2**:e117.
45. Desikan R, AH-M S, Hancock JT, Neill SJ: **Regulation of the Arabidopsis transcriptome by oxidative stress.** *Plant Physiol* 2001, **127**:159-172.
46. Perez-Amador MA, Lidder P, Johnson MA, Landgraf J, Wisman E, Green PJ: **New molecular phenotypes in the dst mutants of Arabidopsis revealed by DNA microarray analysis.** *Plant Cell* 2001, **13**:2703-2717.
47. Germain H, Rudd S, Zotti C, Caron S, O'Brien M, Chantha SC, Lagace M, Major F, Matton DP: **A 6374 unigenes set corresponding to low abundance transcripts expressed following fertilization in Solanum chacoense Bitt, and characterization of 30 receptor-like kinases.** *Plant Mol Biol* 2005, **59**:515-532.
48. Hawkes JG, Hjerting JP: *The Potatoes of Argentina, Brazil, Paraguay and Uruguay* Oxford, UK: Oxford University Press 1969.
49. Van Sint Jan V, Laublin G, Birhman RK, Cappadocia M: **Genetic analysis of leaf explant regenerability in Solanum chacoense.** *Plant Cell, Tissue and Organ Culture* 1996, **47**:9-13(15).
50. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**:207-210.
51. Schefe JH, Lehmann KE, Buschmann IR, Unger T, Funke-Kaiser H: **Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula.** *J Mol Med* 2006, **84**:901-910.
52. Churchill GA: **Fundamentals of experimental design for cDNA microarrays.** *Nat Genet* 2002, **32**(Suppl):490-495.
53. Chantha SC, Emerald BS, Matton DP: **Characterization of the plant Notchless homolog, a WD repeat protein involved in seed development.** *Plant Mol Biol* 2006, **62**:897-912.
54. Fowlkes EB, Mallows CL: **A Method for Comparing Two Hierarchical Clusterings.** *Journal of the American Statistical Association* 1983, **78**:553-584.
55. Park SY, Jauh GY, Mollet JC, Eckard KJ, Nothnagel EA, Walling LL, Lord EM: **A lipid transfer-like protein is necessary for lily pollen tube adhesion to an in vitro stylar matrix.** *Plant Cell* 2000, **12**:151-164.
56. Sarowar S, Kim YJ, Kim KD, Hwang BK, Ok SH, Shin JS: **Overexpression of lipid transfer protein (LTP) genes enhances resistance to plant pathogens and LTP functions in long-distance systemic signaling in tobacco.** *Plant Cell Rep* 2009, **28**:419-427.
57. Vroemen CW, Langeveld S, Mayer U, Ripper G, Jurgens G, Van Kammen A, De Vries SC: **Pattern Formation in the Arabidopsis Embryo Revealed by Position-Specific Lipid Transfer Protein Gene Expression.** *Plant Cell* 1996, **8**:783-791.
58. Wolters-Arts M, Lush WM, Mariani C: **Lipids are required for directional pollen-tube growth.** *Nature* 1998, **392**:818-821.
59. Nieuwland J, Feron R, Huisman BA, Fasolino A, Hilbers CW, Derksen J, Mariani C: **Lipid transfer proteins enhance cell wall extension in tobacco.** *Plant Cell* 2005, **17**:2009-2019.
60. Kader JC: **Lipid-transfer proteins: a puzzling family of plant proteins.** *Trends Plant Sci* 1997, **2**:66-70.
61. Johnson R, Narvaez J, An G, Ryan C: **Expression of proteinase inhibitors I and II in transgenic tobacco plants: effects on natural defense against Manduca sexta larvae.** *Proceedings of the National Academy of Sciences of the United States of America* 1989, **86**:9871-9875.
62. Peña-Cortes H, Sanchez-Serrano J, Rocha-Sosa M, Willmitzer L: **Systemic induction of proteinase-inhibitor-II gene expression in potato plants by wounding.** *Planta* 1988, **174**:84-89.
63. Martineau B, McBride KE, Houck CM: **Regulation of metalcarboxypeptidase inhibitor gene expression in tomato.** *Molecular and General Genetics MGG* 1991, **228**:281-286.
64. Villanueva J, Canals F, Prat S, Ludevid D, Querol E, Avilés FX: **Characterization of the wound-induced metalcarboxypeptidase inhibitor from potato1 The nucleotide sequence data reported here have been submitted to the GenBank Nucleotide Sequence Database and are available under accession number AF060551.1: cDNA sequence, induction of gene expression, subcellular immunolocalization and potential roles of the C-terminal propeptide.** *FEBS letters* 1998, **440**:175-182.
65. Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, *et al*: **Functional annotation of the Arabidopsis genome using controlled vocabularies.** *Plant Physiol* 2004, **135**:745-755.
66. Meinke DW: **Embryo-lethal mutants of Arabidopsis thaliana: analysis of mutants with a wide range of lethal phases.** *TAG Theoretical and Applied Genetics* 1985, **69**:543-552.
67. Errampalli D, Patton D, Castle L, Mickelson L, Hansen K, Schnall J, Feldmann K, Meinke D: **Embryonic Lethals and T-DNA Insertional Mutagenesis in Arabidopsis.** *Plant Cell* 1991, **3**:149-157.
68. Jurgens G, Mayer U, Torres Ruiz R, Berleth T, Misera S: **Genetic analysis of pattern formation in the Arabidopsis embryo.** *Dev Suppl* 1991, **1**:27-38.
69. Elster R, Bommer P, Sheridan WF, Werr W: **Analysis of four embryo-specific mutants in Zea mays reveals that incomplete radial organization of the proembryo interferes with subsequent development.** *Dev Genes Evol* 2000, **210**:300-310.
70. Devic M, Albert S, Delseny M: **Induction and expression of seed-specific promoters in Arabidopsis embryo-defective mutants.** *Plant J* 1996, **9**:205-215.
71. Altmann T, Felix G, Jessop A, Kauschmann A, Uwer U, Pena-Cortes H, Willmitzer L: **Ac/Ds transposon mutagenesis in Arabidopsis thaliana: mutant spectrum and frequency of Ds insertion mutants.** *Mol Gen Genet* 1995, **247**:646-652.
72. Mayer U, Ruiz RAT, Berleth T, Misera S, Juurgens G: **Mutations affecting body organization in the Arabidopsis embryo.** *Nature* 1991, **353**:402-407.
73. Goldberg RB, de Paiva G, Yadegari R: **Plant Embryogenesis: Zygote to Seed.** *Science* 1994, **266**:605-614.
74. Johri B, Ambegaokar K, Srivastava P: *Comparative embryology of angiosperms* Springer 1992.
75. Geeta R: **The origin and maintenance of nuclear endosperms: viewing development through a phylogenetic lens.** *Proc Biol Sci* 2003, **270**:29-35.
76. Rounsley SD, Ditta GS, Yanofsky MF: **Diverse Roles for MADS Box Genes in Arabidopsis Development.** *Plant Cell* 1995, **7**:1259-1269.
77. Golz JF, Hudson A: **Plant development: YABBYs claw to the fore.** *Curr Biol* 1999, **9**:R861-863.
78. Bowman JL: **The YABBY gene family and abaxial cell fate.** *Curr Opin Plant Biol* 2000, **3**:17-22.
79. Bowman JL, Smyth DR, Meyerowitz EM: **Genetic interactions among floral homeotic genes of Arabidopsis.** *Development* 1991, **112**:1-20.
80. Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Munster T, Winter KU, Saedler H: **A short history of MADS-box genes in plants.** *Plant Mol Biol* 2000, **42**:115-149.
81. Ng M, Yanofsky MF: **Function and evolution of the plant MADS-box gene family.** *Nat Rev Genet* 2001, **2**:186-195.
82. Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM: **The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors.** *Nature* 1990, **346**:35-39.
83. Jack T, Brockman LL, Meyerowitz EM: **The homeotic gene APETALA3 of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens.** *Cell* 1992, **68**:683-697.
84. Azumi Y, Liu D, Zhao D, Li W, Wang G, Hu Y, Ma H: **Homolog interaction during meiotic prophase I in Arabidopsis requires the SOLO DANCERS gene encoding a novel cyclin-like protein.** *Embo J* 2002, **21**:3081-3095.
85. Vroemen CW, Mordhorst AP, Albrecht C, Kwaaitaal MA, de Vries SC: **The CUP-SHAPED COTYLEDON3 gene is required for boundary and shoot meristem formation in Arabidopsis.** *Plant Cell* 2003, **15**:1563-1577.
86. Walling L, Drews GN, Goldberg RB: **Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels.** *Proc Natl Acad Sci USA* 1986, **83**:2123-2127.
87. Gatehouse JA, Evans IM, Croy RRD, Boulter D: **Differential Expression of Genes During Legume Seed Development.** *Philosophical Transactions of the Royal Society of London B, Biological Sciences* 1986, **314**:367-384.
88. Sprunck S, Baumann U, Edwards K, Langridge P, Dresselhaus T: **The transcript composition of egg cells changes significantly following fertilization in wheat (Triticum aestivum L.).** *Plant J* 2005, **41**:660-672.
89. Porceddu A, Stals H, Reichheld JP, Segers G, De Veylder L, Barroco RP, Casteels P, Van Montagu M, Inze D, Mironov V: **A plant-specific cyclin-**

- dependent kinase is involved in the control of G2/M progression in plants. *J Biol Chem* 2001, **276**:36354-36360.
90. Gown AM, Jiang JJ, Matles H, Skelly M, Goodpaster T, Cass L, Reshatof M, Spaulding D, Coltrera MD: **Validation of the S-phase specificity of histone (H3) in situ hybridization in normal and malignant cells.** *J Histochem Cytochem* 1996, **44**:221-226.
91. Tanimoto EY, Rost TL, Comai L: **DNA Replication-Dependent Histone H2A mRNA Expression in Pea Root Tips.** *Plant Physiol* 1993, **103**:1291-1297.
92. Breyne P, Dreesen R, Vandepoele K, De Veylder L, Van Breusegem F, Callewaert L, Rombauts S, Raes J, Cannoot B, Engler G, et al: **Transcriptome analysis during cell division in plants.** *Proc Natl Acad Sci USA* 2002, **99**:14825-14830.
93. Cormier P: **Translation factors: from protein synthesis to cell cycle regulation and tumorigenesis.** *MS - Medecine sciences* 2000, **16**:378-385.
94. Brukhin V, Gheyselinck J, Gagliardini V, Genschik P, Grossniklaus U: **The RPN1 Subunit of the 26S Proteasome in Arabidopsis Is Essential for Embryogenesis.** *Plant Cell* 2005, **17**:2723-2737.
95. Kushwaha R, Singh A, Chattopadhyay S: **Calmodulin7 plays an important role as transcriptional regulator in Arabidopsis seedling development.** *Plant Cell* 2008, **20**:1747-1759.
96. Ma W, Smigel A, Tsai YC, Braam J, Berkowitz GA: **Innate immunity signaling: cytosolic Ca<sup>2+</sup> elevation is linked to downstream nitric oxide generation through the action of calmodulin or a calmodulin-like protein.** *Plant Physiol* 2008, **148**:818-828.
97. Navarro L, Zipfel C, Rowland O, Keller I, Robatzek S, Boller T, Jones JD: **The transcriptional innate immune response to flg22. Interplay and overlap with Avr gene-dependent defense responses and bacterial pathogenesis.** *Plant Physiol* 2004, **135**:1113-1128.
98. Durrant WE, Rowland O, Piedras P, Hammond-Kosack KE, Jones JD: **cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles.** *Plant Cell* 2000, **12**:963-977.
99. Benschop JJ, Mohammed S, O'Flaherty M, Heck AJ, Slijper M, Menke FL: **Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis.** *Mol Cell Proteomics* 2007, **6**:1198-1214.
100. Pislariu CI, Dickstein R: **An IRE-like AGC kinase gene, MtIRE, has unique expression in the invasion zone of developing root nodules in *Medicago truncatula*.** *Plant Physiol* 2007, **144**:682-694.
101. Moore I, Diefenthal T, Zarsky V, Schell J, Palme K: **A homolog of the mammalian GTPase Rab2 is present in Arabidopsis and is expressed predominantly in pollen grains and seedlings.** *Proc Natl Acad Sci USA* 1997, **94**:762-767.
102. Abe M, Takahashi T, Komeda Y: **Cloning and characterization of an L1 layer-specific gene in Arabidopsis thaliana.** *Plant Cell Physiol* 1999, **40**:571-580.
103. Long JA, Ohno C, Smith ZR, Meyerowitz EM: **TOPLESS regulates apical embryonic fate in Arabidopsis.** *Science* 2006, **312**:1520-1523.
104. Lehti-Shiu MD, Adamczyk BJ, Fernandez DE: **Expression of MADS-box genes during the embryonic phase in Arabidopsis.** *Plant Mol Biol* 2005, **58**:89-107.
105. Angenent GC, Franken J, Busscher M, van Dijken A, van Went JL, Dons HJ, van Tunen AJ: **A novel class of MADS box genes is involved in ovule development in petunia.** *Plant Cell* 1995, **7**:1569-1582.
106. Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF: **Assessing the redundancy of MADS-box genes during carpel and ovule development.** *Nature* 2003, **424**:85-88.
107. Sieber P, Gheyselinck J, Gross-Hardt R, Laux T, Grossniklaus U, Schneitz K: **Pattern formation during early ovule development in Arabidopsis thaliana.** *Dev Biol* 2004, **273**:321-334.
108. Becerra C, Puigdomenech P, Vicent CM: **Computational and experimental analysis identifies Arabidopsis genes specifically expressed during early seed development.** *BMC Genomics* 2006, **7**:38.
109. Yu HJ, Hogan P, Sundaresan V: **Analysis of the female gametophyte transcriptome of Arabidopsis by comparative expression profiling.** *Plant Physiol* 2005, **139**:1853-1869.
110. Tung CW, Dwyer KG, Nasrallah ME, Nasrallah JB: **Genome-wide identification of genes expressed in Arabidopsis pistils specifically along the path of pollen tube growth.** *Plant Physiol* 2005, **138**:977-989.
111. Nielsen ME, Lok F, Nielsen HB: **Distinct developmental defense activations in barley embryos identified by transcriptome profiling.** *Plant Mol Biol* 2006, **61**:589-601.
112. Gollack D, Vera P, Dietz KJ: **Expression of subtilisin-like serine proteases in Arabidopsis thaliana is cell-specific and responds to jasmonic acid and heavy metals with developmental differences.** *Physiol Plant* 2003, **118**:64-73.
113. Blein JP, Cotos-Thevenot P, Marion D, Ponchet M: **From elicitors to lipid-transfer proteins: a new insight in cell signalling involved in plant defence mechanisms.** *Trends Plant Sci* 2002, **7**:293-296.
114. Maldonado AM, Doerner P, Dixon RA, Lamb CJ, Cameron RK: **A putative lipid transfer protein involved in systemic resistance signalling in Arabidopsis.** *Nature* 2002, **419**:399-403.
115. Trevino MB, MA OC: **Three drought-responsive members of the nonspecific lipid-transfer protein gene family in *Lycopersicon pennellii* show different developmental patterns of expression.** *Plant Physiol* 1998, **116**:1461-1468.
116. Sheoran IS, Olson DJ, Ross AR, Sawhney VK: **Proteome analysis of embryo and endosperm from germinating tomato seeds.** *Proteomics* 2005, **5**:3752-3764.
117. Palanivelu R, Brass L, Edlund AF, Preuss D: **Pollen tube growth and guidance is regulated by POP2, an Arabidopsis gene that controls GABA levels.** *Cell* 2003, **114**:47-59.
118. Wilhelm LK, Preuss D: **Self-sterility in Arabidopsis due to defective pollen tube guidance.** *Science* 1996, **274**:1535-1537.
119. Sweetlove LJ, Heazlewood JL, Herald V, Holtzapffel R, Day DA, Leaver CJ, Millar AH: **The impact of oxidative stress on Arabidopsis mitochondria.** *The Plant Journal* 2002, **32**:891-904.
120. Cazale AC, Clement M, Chiarenza S, Roncato MA, Pochon N, Creff A, Marin E, Leonhardt N, Noel LD: **Altered expression of cytosolic/nuclear HSC70-1 molecular chaperone affects development and abiotic stress tolerance in Arabidopsis thaliana.** *J Exp Bot* 2009, **60**:2653-2664.
121. Lan L, Li M, Lai Y, Xu W, Kong Z, Ying K, Han B, Xue Y: **Microarray analysis reveals similarities and variations in genetic programs controlling pollination/fertilization and stress responses in rice (*Oryza sativa* L.).** *Plant Mol Biol* 2005, **59**:151-164.
122. Chantha SC, Tebbji F, Matton DP: **From the notch signaling pathway to ribosome biogenesis.** *Plant Signal Behav* 2007, **2**:168-170.
123. O'Brien M, Chantha SC, Rahier A, Matton DP: **Lipid signaling in plants. Cloning and expression analysis of the obtusifolios 14alpha-demethylase from *Solanum chacoense* Bitt., a pollination- and fertilization-induced gene with both obtusifolios and lanosterol demethylase activity.** *Plant Physiol* 2005, **139**:734-749.
124. Yamaoka S, Leaver CJ: **EMB2473/MIRO1, an Arabidopsis Miro GTPase, is required for embryogenesis and influences mitochondrial morphology in pollen.** *Plant Cell* 2008, **20**:589-601.
125. Balk J, Aguilar Netz DJ, Tepper K, Pierik AJ, Lill R: **The essential WD40 protein Cia1 is involved in a late step of cytosolic and nuclear iron-sulfur protein assembly.** *Mol Cell Biol* 2005, **25**:10833-10841.
126. Johnstone RW, Wang J, Tommerup N, Vissing H, Roberts T, Shi Y: **Ciao 1 is a novel WD40 protein that interacts with the tumor suppressor protein WT1.** *J Biol Chem* 1998, **273**:10880-10887.
127. Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ: **Sorting signals, N-terminal modifications and abundance of the chloroplast proteome.** *PLoS One* 2008, **3**:e1994.
128. Weijers D, Benkova E, Jager KE, Schlereth A, Hamann T, Kientz M, Wilmoth JC, Reed JW, Jurgens G: **Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators.** *EMBO J* 2005, **24**:1874-1885.
129. Friml J, Vieten A, Sauer M, Weijers D, Schwarz H, Hamann T, Offringa R, Jurgens G: **Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis.** *Nature* 2003, **426**:147-153.
130. Weijers D, Schlereth A, Ehrismann JS, Schwank G, Kientz M, Jurgens G: **Auxin triggers transient local signaling for cell specification in Arabidopsis embryogenesis.** *Dev Cell* 2006, **10**:265-270.
131. Jürgens G: **Growing up green: cellular basis of plant development.** *Mech Dev* 2003, **120**:1395-1406.
132. Jurgens G: **Apical-basal pattern formation in Arabidopsis embryogenesis.** *EMBO J* 2001, **20**:3609-3616.
133. Zimmerman JL: **Somatic Embryogenesis: A Model for Early Development in Higher Plants.** *Plant Cell* 1993, **5**:1411-1423.
134. Chen JG, Ullah H, Young JC, Sussman MR, Jones AM: **ABP1 is required for organized cell elongation and division in Arabidopsis embryogenesis.** *Genes & Development* 2001, **15**:902-911.

135. Hadfi K, Speth V, Neuhaus G: **Auxin-induced developmental patterns in *Brassica juncea* embryos.** *Development* 1998, **125**:879-887.
136. Tanaka H, Dhonukshe P, Brewer P, Friml J: **Spatiotemporal asymmetric auxin distribution: a means to coordinate plant development.** *Cellular and Molecular Life Sciences* 2006, **63**:2738-2754.
137. Eeuwens CJ, Schwabe WW: **Seed and podwall development in *P. sativum* L. in relation to extracted and applied hormones.** *Journal of Experimental Botany* 1975, **26**.
138. Takase T, Nakazawa M, Ishikawa A, Kawashima M, Ichikawa T, Takahashi N, Shimada H, Manabe K, Matsui M: **ydk1-D, an auxin-responsive GH3 mutant that is involved in hypocotyl and root elongation.** *Plant J* 2004, **37**:471-483.
139. Jurgens G, Torres Ruiz RA, Berleth T: **Embryonic pattern formation in flowering plants.** *Annu Rev Genet* 1994, **28**:351-371.
140. Przemecck GK, Mattsson J, Hardtke CS, Sung ZR, Berleth T: **Studies on the role of the Arabidopsis gene MONOPTEROS in vascular development and plant cell axialization.** *Planta* 1996, **200**:229-237.
141. Hardtke CS, Berleth T: **The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development.** *Embo J* 1998, **17**:1405-1411.
142. Hamann T, Mayer U, Jurgens G: **The auxin-insensitive bodenlos mutation affects primary root formation and apical-basal patterning in the Arabidopsis embryo.** *Development* 1999, **126**:1387-1395.
143. Hamann T, Benkova E, Baurle I, Kientz M, Jurgens G: **The Arabidopsis BODENLOS gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning.** *Genes Dev* 2002, **16**:1610-1615.
144. Barton MK, Poethig RS: **Formation of the shoot apical meristem in *Arabidopsis thaliana*: an analysis of development in the wild type and in the shoot meristemless mutant.** *Development* 1993, **119**:823-831.

doi:10.1186/1471-2229-10-174

**Cite this article as:** Tebbji et al.: Transcription profiling of fertilization and early seed development events in a solanaceous species using a 7.7 K cDNA microarray from *Solanum chacoense* ovules. *BMC Plant Biology* 2010 **10**:174.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

