Comparative Transcriptomics of Arabidopsis Sperm Cells[1][C][W]

Filipe Borges, Gabriela Gomes[2], Rui Gardner, Nuno Moreno, Sheila McCormick, José A. Feijó*, and Jörg D. Becker

Instituto Gulbenkian de Ciência, Centro de Biologia do Desenvolvimento, Oeiras, 2780–901, Portugal (F.B., G.G., R.G., N.M., J.A.F., J.D.B.); Plant Gene Expression Center, United States Department of Agriculture/ Agricultural Research Service, University of California, Berkeley, Albany, California 94710 (S.M.); and Depto. Biologia Vegetal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, Lisbon, 1700, Portugal (J.A.F.)

In flowering plants, the two sperm cells are embedded within the cytoplasm of the growing pollen tube and as such are passively transported to the embryo sac, wherein double fertilization occurs upon their release. Understanding the mechanisms and conditions by which male gametes mature and take part in fertilization are crucial goals in the study of plant reproduction. Studies of gene expression in male gametes of maize (*Zea mays*) and *Plumbago* and in lily (*Lilium longiflorum*) generative cells already showed that the previously held view of transcriptionally inert male gametes was not true, but genome-wide studies were lacking. Analyses in the model plant Arabidopsis (*Arabidopsis thaliana*) were hindered, because no method to isolate sperm cells was available. Here, we used fluorescence-activated cell sorting to isolate sperm cells from Arabidopsis, allowing GeneChip analysis of their transcriptome at a genome-wide level. Comparative analysis of the sperm cell transcriptome with those of representative sporophytic tissues and of pollen showed that sperm has a distinct and diverse transcriptional profile. Functional classifications of genes with enriched expression in sperm cells showed that DNA repair, ubiquitin-mediated proteolysis, and cell cycle progression are overrepresented Gene Ontology categories. Moreover, analysis of the small RNA and DNA methylation pathways suggests that distinct mechanisms might be involved in regulating the epigenetic state of the paternal genome. We identified numerous candidate genes whose involvement in sperm cell development and fertilization can now be directly tested in Arabidopsis. These results provide a roadmap to decipher the role of sperm-expressed proteins.

In angiosperms, meiosis in the anthers yields haploid unicellular microspores. Subsequently, pollen mitosis I (PM I) yields a larger vegetative cell and a smaller generative cell (GC). The GC undergoes PM II, a symmetric division that yields two sperm cells. In Arabidopsis (*Arabidopsis thaliana*), PM II occurs before anthesis, so that three-celled pollen grains (a vegetative cell and two sperm cells within the vegetative cell cytoplasm) are later released from the anthers (Dumas et al., 1985; Yu et al., 1989; Boavida et al., 2005a). When the male gametophyte (pollen grain) meets the papillae of a receptive stigma, a complex series of cell-cell signaling events will drive pollen tube growth toward the embryo sac (female gametophyte). Upon arrival, the pollen tube tip bursts, discharging the two sperm cells. To achieve double fertilization, each sperm cell fuses with an egg or a central cell to yield the zygote and primary endosperm cell, respectively (Boavida et al., 2005b), but the mechanisms underlying double fusion remain relatively unknown. It was recently shown that a mutation in the Arabidopsis *CDC2A* gene has a paternal effect, whereby mutant pollen produces only one sperm cell that exclusively fertilizes the egg cell (Nowack et al., 2006). Although this mutant arrested embryo development early, its single fertilization event somehow triggered autonomous proliferation of the endosperm. The loss of activity of the Chromatin Assembly Factor 1 also prevents PM II, but in this case, the resulting single sperm cell could fertilize either female gamete (Chen et al., 2008).

To date, a few genes have been described as specifically expressed in Arabidopsis sperm cells; some of these appear to be important for pollen tube guidance (von Besser et al., 2006) or for regulating their own transcriptional program (Okada et al., 2005). More-

[1] This work was supported by Fundacão para a Ciência e a Tecnologia, Portugal (grant nos. POCTI/BIA–BCM/60046/2004 and PPCDT/BIA–BCM/61270/2004; a BIC within this project to F.B. and G.G.; and postdoc fellowships SFRH/BPD/31047/2006 and SFRH/BPD/14930/2004, respectively, to J.D.B. and R.G.).

[2] Present address: Centro de Engenharia Biológica e Química, Instituto Superior Técnico, Av. Rovisco Pais, 1049–001 Lisboa, Portugal.

* Corresponding author; e-mail jfeijo@igc.gulbenkian.pt.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: José A. Feijó (jfeijo@igc.gulbenkian.pt).

[C] Some figures in this article are displayed in color online but in black and white in the print edition.

[W] The online version of this article contains Web-only data.

www.plantphysiol.org/cgi/doi/10.1104/pp.108.125229

over, postmitotic male gametes of Arabidopsis continue to progress through the S phase of the cell cycle as they move through the pollen tube, reaching the G2 phase just before double fertilization (Friedman, 1999). These findings suggest that there is an intricate control over molecular programs in the sperm cells that are coupled to cell cycle transitions.

Although sperm cells have highly condensed chromatin, the original assumption that they are transcriptionally quiescent was proven wrong by the activity of *AtGEX1* promoter, whose expression is confined to sperm cells and is not expressed in the progenitor GC (Engel et al., 2005). A cDNA library was constructed from fluorescence-activated cell sorting (FACS)-purified sperm cells of maize (*Zea mays*), which revealed a diverse complement of mRNAs representing at least 2,560 genes (Engel et al., 2003). These findings were significantly more than the restricted number of transcripts previously described in plant sperm cells (Xu et al., 1999a, 1999b; Singh et al., 2002; Okada et al., 2005). In addition, they led to the identification of *AtGEX1* and *AtGEX2* (Engel et al., 2005), whose promoters drive expression in Arabidopsis sperm cells. Additional sperm-specific genes have been reported more recently (Mori et al., 2006; von Besser et al., 2006; Ingouff et al., 2007), including *DUO1*, a MYB transcription factor in Arabidopsis that is important for the G2/M transition during PM II (Rotman et al., 2005). Transcriptional profiling of GCs of lily (*Lilium longiflorum*; Okada et al., 2006, 2007) revealed a considerable overlap with the maize sperm ESTs (Engel et al., 2003) and the Arabidopsis pollen microarray datasets (Becker et al., 2003; Honys and Twell, 2003). Despite this previous work, the male gamete ESTs from maize, *Plumbago*, and lily were limited, so that testing the potential importance of sperm-expressed transcripts by means of reverse genetics could not be comprehensive, even when an Arabidopsis homolog could be identified for such ESTs. Several studies using the Affymetrix ATH1 GeneChip platform to characterize the transcriptomes of mature pollen grains and earlier developmental stages demonstrated that this type of analysis can provide starting points from which to dissect the genetic programs driving development and functions of the male gametophyte (for review, see Becker and Feijo, 2007). But technical difficulties in obtaining sufficient amounts of pure biological material constituted a major hindrance for transcriptional profiling of Arabidopsis sperm cells.

Here, we used a newly developed protocol, based on FACS, to isolate and purify sperm cells from transgenic Arabidopsis plants that were expressing enhanced GFP under the control of a sperm-specific promoter (*AtGEX2::eGFP*). Consequently, we could use Affymetrix ATH1 Genome Arrays to profile the mRNA complement of these male gametes. The direct comparison of their transcriptome with those of pollen and seedlings, as well as with additional ATH1 data sets from a variety of vegetative tissues, showed that the sperm cell transcriptome was distinct. We identified transcripts that were enriched and/or preferentially expressed in sperm cells. Functional classification of sperm-enriched transcripts showed that DNA repair, ubiquitin-mediated proteolysis, and proteins required for progression through the cell cycle were overrepresented categories. Our analyses provide a number of hypotheses for testing and point out future challenges toward understanding the role of genes expressed in the male gametes before, during, and possibly after double fertilization.

RESULTS AND DISCUSSION

FACS of Arabidopsis Sperm Cells

Arabidopsis sperm cells expressing eGFP (Engel et al., 2005) were isolated using a high-speed cell sorter from crude fractions resulting from a complex purification protocol (see Supplemental Fig. S1). They were selected based on their size and intracellular complexity (as assessed by the angle of their light scatter properties), their GFP signal, and presence of DNA. Low granulosity (low side scatter [SSC]) and GFP-positive signals were used to identify the sperm cell population (R1 in Fig. 1A). A low forward scatter (FSC) signal (small particles) within the GFP positive population further discriminated between sperm cells and other small entities within the sample (R2 in Fig. 1B). To assure exclusion of debris within the small-sized population containing GFP-positive sperm cells, we used DRAQ5, a live cell DNA marker (R3 in Fig. 1C). To guarantee purity, sperm cells were sorted using a logical combination of regions R2 and R3, and we verified that cells sorted from region R1 only yielded the same population (results not shown). Each session yielded around 100,000 cells, from which total RNA was isolated immediately. A high degree of purity of the sorted sperm cell fractions was verified routinely before and during sorting, by wide-field fluorescence microscopy (Fig. 2), and later by reverse transcription (RT)-PCR analysis on known vegetative nucleus- and sperm cell-specific genes (Fig. 4C), as described below. The viability of isolated sperm cells was confirmed by fluorescein diacetate staining (Fig. 2G).

A Remarkable Diversity of mRNAs in Sperm

To determine the sperm cell transcriptome, we used Arabidopsis GeneChip ATH1 Genome arrays, representing 22,392 different genes. Although microarray data sets for seedlings and pollen exist (Pina et al., 2005), we repeated analyses for these samples using identical starting amounts of total RNA (16 ng), because these amounts were less than those used in previous studies in our lab. Correlation coefficients above 0.97 between the three replicates of each sample type underline the reproducibility of the array data (see Supplemental Fig. S4). Based on the MicroArray Suite 5 detection algorithm, the mean percentages of detected present calls were 27% for sperm cells, 33%

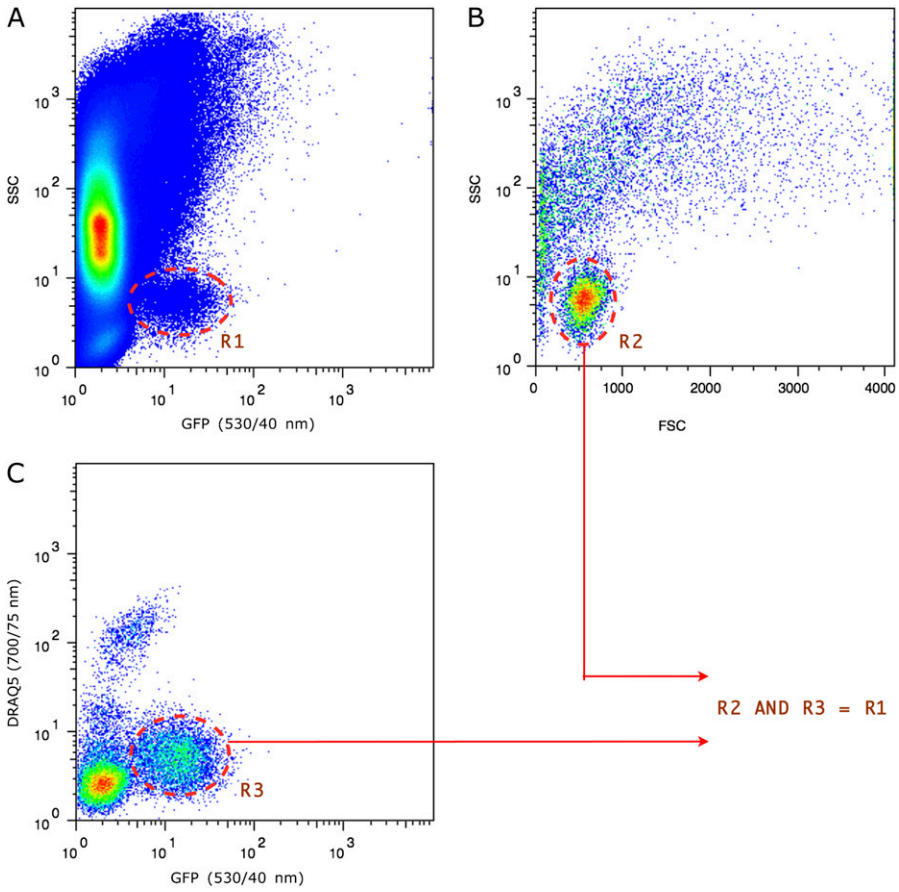

Figure 1. Fluorescence-activated sperm cell sorting based on cell size (FSC), intracellular complexity (SSC), GFP signal, and presence of intracellular DNA, via DRAQ5 staining. Low granulosity (low SSC) and GFP positive signals were used to identify the sperm cell population (R1) from the total population. To guarantee purity, a low FSC signal (small particles; R2) within the GFP/DRAQ5 double positive population (R3) were used to exclude other small particles. A displays total population, B shows cells within region R2, and C shows cells within region R3. [See online article for color version of this figure.]

for pollen, and 64% for seedlings, corresponding to 5,829, 7,177, and 14,464 genes, respectively. Summarized expression data of all samples are shown in Supplemental Table S1.

A cDNA library from FACS-purified sperm cells of maize yielded around 5,000 sequenced ESTs that matched with approximately 1,385 annotated genes in the Arabidopsis genome (e-values of 1e-8 or smaller; Engel et al., 2003), although 138 of those were not represented on ATH1. A Venn diagram (Fig. 3) was used to display the number of genes common between our samples and the identifiers of Arabidopsis genes matching the maize sperm cell ESTs (restricted to genes represented on the array). Although differences in the transcriptomes of sperm cells from monocot and dicot plants are expected, it was surprising that only 594 genes were shared between our dataset of Arabidopsis sperm cell-expressed genes and the Arabidopsis genes listed as corresponding to maize sperm ESTs. One explanation for this might be that the maize EST comparisons identified a different family member as the best Arabidopsis match. Another explanation is that cDNA sequencing could have identified mRNAs that are not detected in the microarray analyses. To test these hypotheses, we analyzed 395 maize sperm ESTs for which their putative Arabidopsis homologs were called absent by the microarray experiment. This analysis confirmed that most (238) were singletons, while the others were mostly represented by 2, 3, or 4 ESTs. These findings support the idea that rare messages might have been sequenced from the maize sperm library but would not have been detected by the microarray. A few genes were represented by multiple ESTs (e.g. ATPase encoding gene represented by 23 ESTs) and these belong to gene families, so that the gene selected as the best match could have been a different member than the one detected in Arabidopsis sperm cells. The overlap between Arabidopsis sperm cells and seedling encompassed 4,757 genes, representing almost the entire sperm transcriptome, with approximately 2,400 of those genes showing enriched expression in sperm. As expected, the vast majority of genes called present in Arabidopsis sperm cells were also detected in Arabidopsis pollen (3,813). To address the possibility of contamination from RNA derived from the vegetative cell in our FACS-purified sperm cell samples, we screened the dataset for genes reported as pollen-specific (expression confined only to the vegetative cell). A good example can be found in Strompen et al. (2005), where isoforms of the vacuolar H^+-ATPase subunit E were analyzed by means of comparison of VHA-E1-GFP, E2-GFP, and E3-GFP fusion expression in mature pollen. E1, the major isoform, was expressed in sperm cells but not in the vegetative cell, E2 was expressed only in the vegetative cell, and E3 was

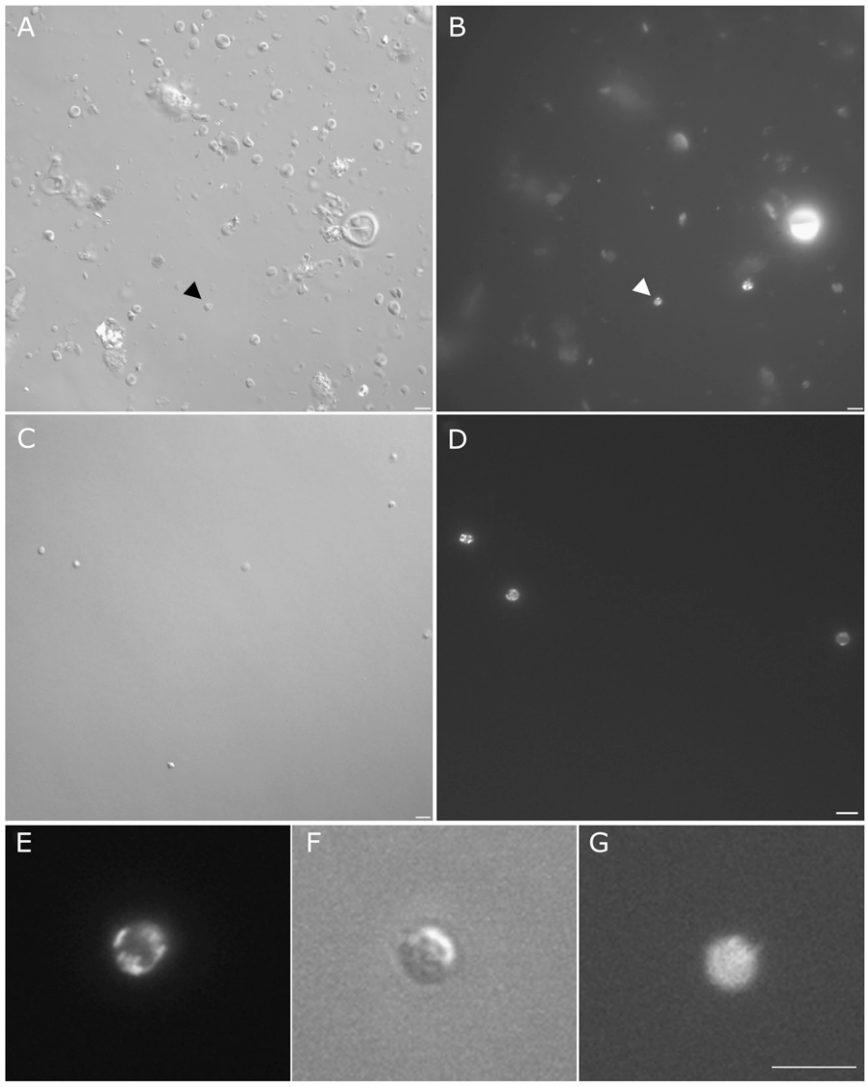

Figure 2. Visualization of FACS-purified sperm cells. Wide-field fluorescence microscopy was used to visualize Arabidopsis sperm cells expressing *AtGEX2*::*eGFP* (Engel et al., 2005) before (B) and after FACS purification (D). Differential interference microscopy (DIC) microscopy confirmed that the debris in the filtrate before FACS (A) was removed after sorting (C). For DIC imaging of FACS-purified sperm cells (C), we captured and merged several images along the optical axis. A higher magnification of a sorted sperm cell shows GFP fluorescence (E), cell-shape integrity by DIC microscopy (F), and cell viability using fluorescein diacetate staining (G). The bars represent 5 microns and the arrowheads are pointing to sperm cells.

expressed in both the vegetative cell and sperm cells. The genes encoding all three isoforms are represented on ATH1 (*At4g11150*, *At3g08560*, and *At1g64200*, respectively). Both E1 and E3 transcripts were detected in sperm cells, whereas E2 was not. All three transcripts were detected in pollen. Another example is *AtVEX1*|*At5g62850*, specifically expressed in the vegetative cell cytoplasm and nucleus (Engel et al., 2005) and called absent in the sperm cell sample, but present in pollen. For further validation of our microarray data, RT-PCR was performed. One chosen target was one of the highest expressed genes in pollen, encoding a carbonic anhydrase protein (*At5g04180*), and whose expression was not detected in sperm cells. In agreement with our microarray data, this transcript could not be detected in sperm cells or in seedling by RT-PCR, but high levels were detected in pollen (Fig. 4A). As further proof, two additional independent samples of total RNA from sperm cells were used to confirm enrichment for a sperm-specific transcript, *AtMGH3*|*At1g19890* (Okada et al., 2005), and absence of transcripts derived from the vegetative cell, *AtVEX1*|*At5g62850* (Engel et al., 2005), by RT-PCR (Fig. 4C).

A Unique Transcriptional Profile to Control DNA Repair, Ubiquitination, and Cell Cycle Progression

We carried out comparative analyses of the sperm cell transcriptome with transcriptional datasets from representative vegetative tissues of Arabidopsis. Principal component analysis (PCA) is a robust method to project high-dimensional, global expression data onto the three principal components. The closer two points are in the plot, the more similar the samples are in terms of their global gene expression profile. As shown in Figure 5B, the first principal component separates sperm cells and pollen from the vegetative tissues, while the second principal component shows a further separation between pollen and sperm cells. The first principal component demonstrates the uniqueness of the expression profile in sperm cells. This was already demonstrated for pollen in previous studies (Pina

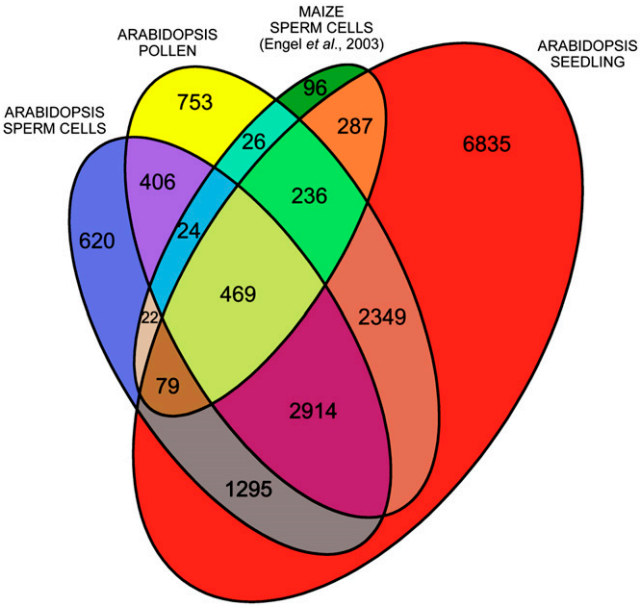

Figure 3. Venn diagram, depicting the overlap between genes whose expression was called present in sperm cells (5,829), pollen (7,177), seedlings (14,464), and additionally the Arabidopsis genes that are represented on the ATH1 array (1,239) and that match those of maize sperm cells ESTs (Engel et al., 2003). Intersection of common genes between Arabidopsis sperm cells and pollen, maize sperm cells, and Arabidopsis seedlings is 3,813, 594, and 4,757, respectively. [See online article for color version of this figure.]

et al., 2005), but here we can additionally observe that pollen and sperm cells separate along the second principal component, thus emphasizing the differences in their transcriptional profiles. A similar separation of samples can be obtained with hierarchical clustering (Fig. 5A). The DAVID Gene Functional Classification Tool (Huang et al., 2007) was used to condense the list of genes showing enriched expression in sperm cells. This tool agglomerates the genes into functional clusters according to their biological significance and enrichment score in relation to all genes represented on the array (Table I; Supplemental Table S2). The distribution through the first clusters highlight overrepresented Gene Ontology (GO) categories, i.e. DNA metabolism (DNA replication and repair), ubiquitin-mediated proteolysis (ubiquitin ligase activity), and cell cycle. Although most of the genes expressed in sperm cells were also detected in pollen, it was previously shown that the overrepresented GO terms for genes with enriched expression in pollen are signaling, vesicle trafficking, and membrane transport (Pina et al., 2005). Previous studies on sperm cell cycle activity (Friedman, 1999) showed that sperm spend most of their development (from the GC undergoing mitosis until entry into the embryo sac) in the S phase of the cell cycle. Plant sperm also have a complex set of transcripts for control of protein fate and degradation through ubiquitination. Interestingly, these processes are also essential during spermatogenesis of mammals, for histone-to-protamine replacement (Baarends et al., 1999), as well as for many other key events in gametogenesis and fertilization (for review, see Sakai et al., 2004). In concordance with our results, genes involved in DNA repair, activation of cyclins, and ubiquitination were also highly represented in ESTs from GCs of lily (Xu et al., 1998; Singh et al., 2002; Okada et al., 2006).

Identification of Novel mRNAs in Male Gametes: Toward Understanding Their Role in Fertilization and Early Development

Genes showing preferential or enriched expression in sperm cells are primary candidates for roles in male gamete development and fertilization. To identify sperm-preferentially expressed transcripts, we performed a comparative analysis with a number of previously reported microarray datasets: leaves and siliques (Pina et al., 2005), ovules, and unpollinated pistils (L.C. Boavida, F. Borges, J.D. Becker, and J.A. Feijó, unpublished data) and AtGenExpress datasets (Schmid et al., 2005) available through the Web-based application Genevestigator (Zimmermann et al., 2004), selecting samples that did not contain pollen. Seventy-four genes were identified (Table II) and distributed according to their molecular function (as annotated in The Arabidopsis Information Resource [TAIR] as of May 2007). Notably, this set includes two genes previously reported as sperm specific: histone H3, *AtMGH3* | *At1g19890* (Okada et al., 2005), and *HAP2 (GCS1)* | *At4g11720* (von Besser et al., 2006). We performed RT-PCR analysis for several genes within this list, including *AtMGH3* and *HAP2*, to confirm the microarray data reported in this study (sperm cells, pollen, and seedling), and additionally with ovule and silique samples (Fig. 4D). Although we could not amplify any of the transcripts from the seedling sample, we detected expression for several of these genes in both ovule and silique samples. We can exclude contamination of genomic DNA in the cDNA templates as source of these amplification products, because we used primers amplifying intron-spanning products whenever possible. Surprisingly, one gene whose expression was detected in ovules and siliques is *HAP2(GCS1)* | *At4g11720*. Previous studies in Arabidopsis showed only a male-specific role for HAP2 (Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006), but on the other hand, RT-PCR analysis indicated that a *HAP2* homolog in *Chlamydomonas* was expressed in both *plus* and *minus* gametes (Mori et al., 2006). Interestingly, although *HAP2* transcripts were detected in *Chlamydomonas plus* gametes, a recent study has shown that HAP2 protein is essential for fusion in *minus* gametes only (Liu et al., 2008). We do not know if the genes we detected are transcribed in the whole ovule or only in a fraction of the cells comprising the ovule, e.g. the egg cell and/or central cell. It was demonstrated that *GEX2*, previously described as specifically expressed in sperm cells (Engel

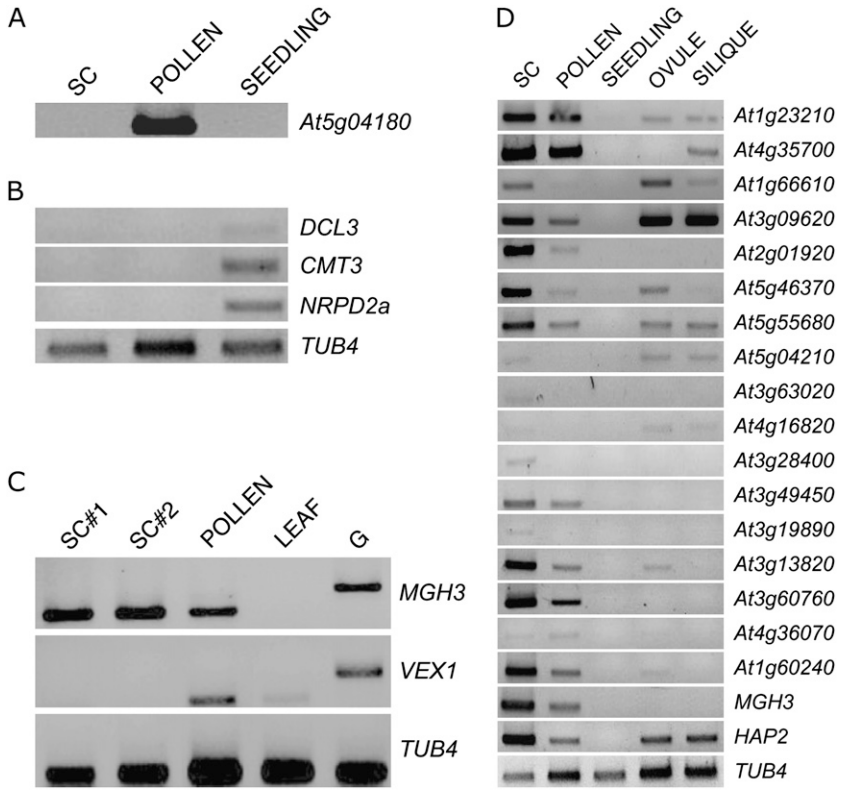

Figure 4. RT-PCR analysis. A and B, Gel figures presenting confirmatory RT-PCR analysis for a gene highly expressed in pollen but not detected in sperm cells in our microarray analysis, encoding a carbonic anhydrase family protein (*At5g44340*; A); and genes involved in the sRNAs and RdDM pathways, which are not detectable in sperm cells, *DCL3* (*At3g43920*), *CMT3* (*At1g69770*), and *NRPD2a* (*At3g23780*; B). C, RT-PCR on total RNA from two samples of FACS-isolated sperm cells (SC#1, SC#2), pollen, and leaf, showing enrichment of *MGH3* (*At1g19890*) transcripts and absence of *VEX1* (*At5g62850*) transcripts in both sperm cell replicates. D, Expression of several genes presented in Table II was tested by RT-PCR in sperm cells, pollen, seedling, ovule, and silique cDNA samples. *TUB4* (*At5g04180*) was used as positive control. SC, Sperm cells; G, genomic DNA.

et al., 2005), is expressed also within the female gametophyte (Alandete-Saez et al., 2008).

Even though some of the genes reported in Table II were detected in ovules and/or siliques, most of them show enriched expression in sperm cells. A highly expressed gene within this list encodes a protein that belongs to the glycosyl hydrolase family (*At1g23210*). Interestingly, in mammals a glycosyl hydrolase (hyaluronidase) protein, PH-20, located at both the plasma and acrosomal membranes in spermatozoa, is involved in adhesion to the zona pellucida (Lathrop et al., 1990). Another gene of note is *KCO2*|*At5g46370*, encoding a member of the Arabidopsis K^+ channel family AtTPK(KCO) (Voelker et al., 2006). Within the same family, *AtTPK4(KCO4)* is predominantly expressed in pollen tubes and is spontaneously activated and modulated by external Ca^{2+} and cytosolic pH (Becker et al., 2004). In mammalian sperm, capacitation refers to intracellular alkalinization and changes in membrane potential, conferring on sperm the capacity to reach and fertilize the egg. Such mechanisms were shown to be driven by outwardly rectifying K^+ currents (Navarro et al., 2007), whereby an inward Ca^{2+}-selective current is maximized by $I_{CatSper}$ (Kirichok et al., 2006). We suggest that K^+ homeostasis in Arabidopsis sperm cells might be at least partially controlled by AtTPK2(KCO2) and that plant male gametes might also undergo a process similar to spermatozoa capacitation. Other channel proteins might play important roles in sperm cells, such as maintaining turgor; for example, two members (MSL2|*At5g10490* and MSL3|*At1g58200*) of the mechanosensitive ion channel family (Haswell and Meyerowitz, 2006) were detected in sperm cells, and both are expressed more highly in sperm than in pollen. The predominant phenotype of mutations in these genes was in chloroplast size and shape (Haswell and Meyerowitz, 2006), but some fertility problems were reported as well.

Two genes (*At3g49450*, *At3g19890*), which, according to our RT-PCR analyses are expressed only in sperm cells, encode proteins with an F-box domain in their N terminus, and it is notable that the most highly expressed gene in Arabidopsis sperm cells is annotated as an F-box protein (*At3g62230*). We used MAPMAN (Thimm et al., 2004) to display the expression profiles of genes involved in the ubiquitin/26S proteasome pathway. Among the F-box proteins represented on the array and expressed in both seedlings and sperm cells, most had higher expression values in sperm than in seedlings (Supplemental Fig. S3). Only about 20 of the more than 450 F-box proteins have had their biological functions elucidated; they are involved in the regulation of diverse cellular processes, including cell cycle transitions, transcription, signal transduction, circadian rhythms, floral development (Patton et al., 1998; del Pozo and Estelle, 2000; Xiao and Jang, 2000; Schwager et al., 2007), and gametophytic self-incompatibility (Sijacic et al., 2004; Hua et al., 2007). The C termini of F-box proteins usually contain domains conferring substrate-binding specificity, being later targeted for degradation by the ubiquitin-ligase complex SKP-Cullin-F-box. Why this

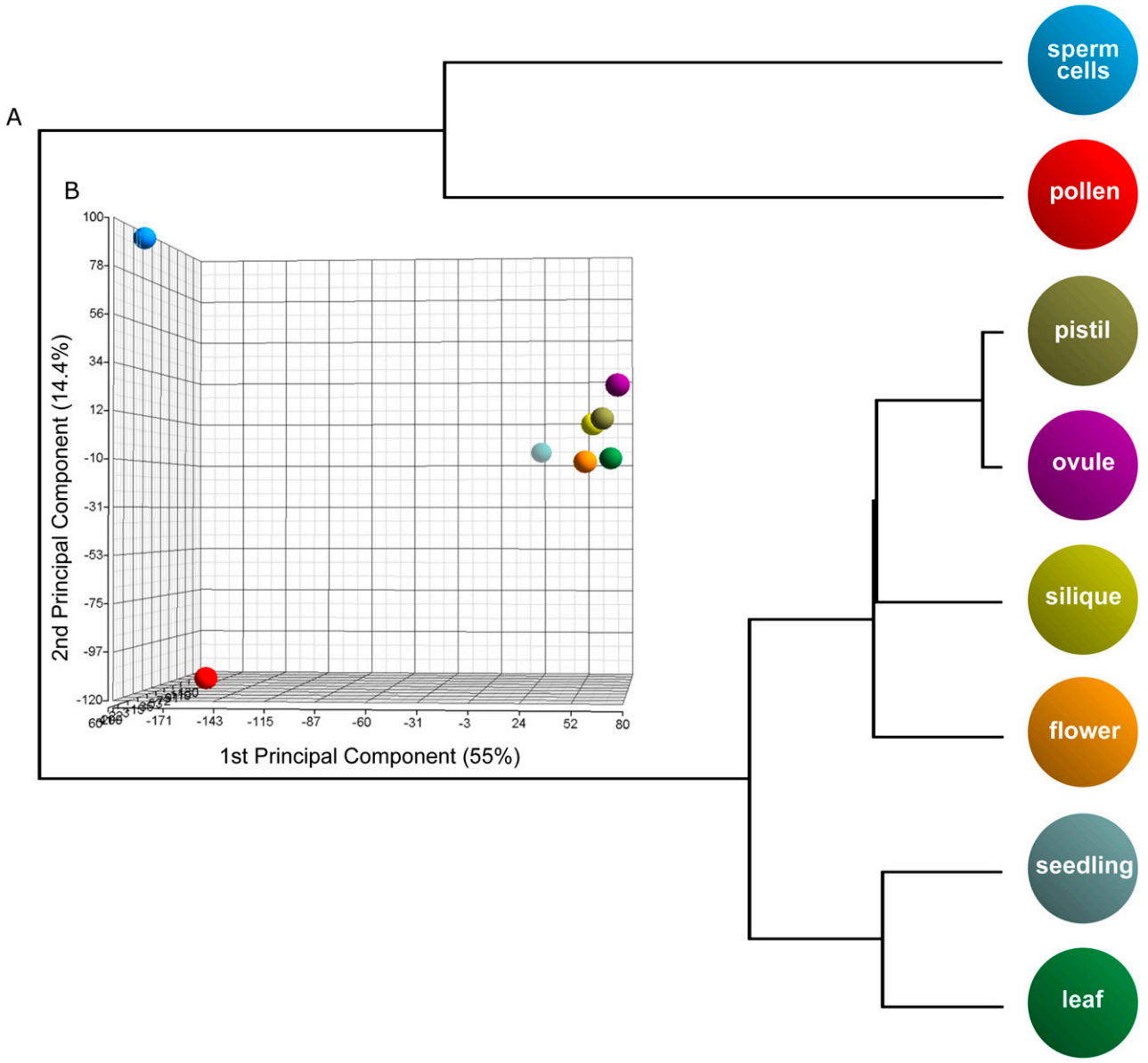

Figure 5. Hierarchical clustering and PCA of tissue-dependent gene expression patterns. Transcriptome data of seedlings, pollen, and sperm cells were compared with those of leaf, flower, silique, ovule, and unpollinated pistil from previous studies (Pina et al., 2005; L.C. Boavida, F. Borges, J.D. Becker, and J.A. Feijó, unpublished data). A, Hierarchical clustering dendrogram, using Pearson's dissimilarity to calculate row dissimilarity and Ward's method for row clustering. B, PCA. [See online article for color version of this figure.]

diversity and enrichment of F-box proteins in sperm cells? One possible hypothesis is that proteins needed during earlier stages of microgametogenesis might need to be targeted for degradation in later stages, e.g. cell cycle regulators. A recent study supports it, because the activity of UBP3/UBP4 deubiquitinating enzymes (essential for the Ub/26S proteasome system) is crucial for PM II in Arabidopsis (Doelling et al., 2007). On the other hand, these and other transcripts or proteins stored in sperm cells might be delivered to both egg and central cells upon fertilization and play a role in early embryonic development. Comprehensive studies highlighting the presence of RNAs in human spermatozoa focused on its potential to profile past events of spermatogenesis and suggested a role for certain sperm RNAs in early zygotic development (Ostermeier et al., 2002). Those particular transcripts were later shown to be delivered to the oocyte at fertilization (Ostermeier et al., 2004).

Transcription Factors

The *AtGEX1* and *AtGEX2* promoters, as well as other sperm-expressed promoters discussed in Engel et al. (2005), have binding sites for Dof-type transcription factors. Of the 32 Dof-type transcription factors in Arabi-

Table I. *Functional classification of sperm-enriched genes*

Gene ontology charts (biological process, molecular function, and cellular component) for sperm-enriched genes were functionally clustered by DAVID tools according to their enrichment score and against all genes represented on the ATH1 array. The representative annotations terms associated to each cluster were manually selected.

Cluster	Annotation Term	Enrichment Score
1	DNA metabolism/DNA repair	8.34
2	Ubiquitin-mediated proteolysis	6.51
3	Cell cycle	6.42
4	Mitosis	3.94
5	Biopolymer metabolism	3.45
6	Chromosome organization and biogenesis	3.23
7	Heterotrimeric G-protein complex	3.18
8	CDK regulator activity	2.43
9	ATPase activity	2.3
10	Cys-type peptidase activity	2.25
11	mRNA metabolism	2.17
12	Cytoskeleton	2.08
13	Intracellular transport	2.06
14	Unfolded protein binding	1.98
15	Damaged DNA binding	1.91

dopsis, three (*At3g47500*, *At5g39660*, and *At5g62430*) were detected in sperm cells; one or more of these probably regulates sperm-specific promoters. There are several MYB-type transcription factors with relatively high expression in sperm cells, while other transcription factor families have restricted representation. For example, there are many homeobox transcription factors in Arabidopsis, but only one of this group (*At3g61150*) has substantial expression in sperm cells. As a group, genes encoding scarecrow transcription factors are mostly not expressed in pollen or in sperm cells, but three (*At2g45160*, *At2g29060*, and *At1g63100*) are detected in sperm cells and not in pollen, implying that these three might play sperm-specific roles. The bHLH transcription factor family is one of the largest in Arabidopsis (Toledo-Ortiz et al., 2003), but only a few were detected in sperm cells, and one of these (*At2g42300*) is not detected in pollen, again suggesting that it may play a specific role in male gametes. Notably, the zinc finger (C2H2 type) transcription factor *At4g35700* shows the highest expression level of all sperm-preferentially expressed genes identified in this study. This transcription factor might be expressed only in sperm cells, although it was also detected in siliques by RT-PCR. Given that the detection level for this transcript is very high in pollen and higher in sperm cells than in pollen (Fig. 4D), it is likely that its weak detection in siliques is pollen derived.

Signal Transduction Pathways

Sperm cells might have unique signaling pathways. For example, the small GTPases of the Rop family are important for many aspects of cytoplasmic signaling, including actin cytoskeleton reorganization (Nibau et al., 2006). The GTP/GDP status of Rops is controlled by the Rop-GEF family, as well as by RopGAPs and ROP-GDIs (Klahre et al., 2006; Klahre and Kost, 2006; Zhang and McCormick, 2007). There are three Rops expressed in sperm cells (*At2g17800*, *At3g51300*, and *At4g35950*), as well as one of the three RopGDIs in Arabidopsis (*At3g07880*) and a RopGAP (*At5g61530*). There are 14 genes encoding PRONE-type RopGEFs in Arabidopsis, and five of these are pollen specific or selective (Zhang and McCormick, 2007). None of the 14 RopGEFs was detected in sperm cells. Another protein, SPIKE1 (*At4g16340*), also has RopGEF activity (Basu et al., 2005), but its transcript was not detected in pollen or in sperm cells. It is unclear how the GTP/GDP status of Rops in sperm cells can be controlled if RopGEFs are absent. We should not exclude the hypothesis that some known RopGEFs may be in fact expressed in sperm cells, but their transcripts are below the detection threshold.

In sperm cells, some complicated signaling cascades might be simplified. For example, mitogen-activated protein kinase (MAPK) cascades are central to many signaling pathways in plants, and there is often cross talk between different members in different signaling pathways (Mishra et al., 2006). Of the 20 MAPKKKs in Arabidopsis, several were detected in sperm cells and two (*MAPKKK19|At3g50310* and *MAPKKK20|At5g67080*) have extremely high expression; these two were not detected in pollen, implying that sperm-specific signaling pathways might exist. Indeed, these MAPKKKs must act on MAPKK3 (*At5g40440*), the only one of the 10 MAPKKs in Arabidopsis that could be detected in sperm cells. Several MAPKs are also expressed, with *At2g01450* showing the highest expression level. Other examples of gene families with restricted expression in sperm cells include the large lectin receptor kinase family, only a few of which were detected, and the auxin efflux carriers, only one (*At5g01990*) of which is expressed in sperm, while a different one (*At2g17500*) is expressed in pollen.

Table II. *Genes preferentially expressed in sperm cells*

*, Genes analyzed by RT-PCR whose expression has been detected only in sperm cells and in some cases in pollen, but probably being sperm-derived. Shown are 74 genes appearing to be preferentially expressed in sperm cells based on comparisons with microarray data of several vegetative tissues from previous studies (Pina et al., 2005; L.C. Boavida, F. Borges, J.D. Becker, and J.A. Feijó, unpublished data) and the AtGenExpress database (Schmid et al., 2005), distributed according to their molecular function as annotated in TAIR. AGI, Arabidopsis Genome Initiative.

Probe Set ID	AGI ID	Description	Signal
DNA or RNA binding			
255815_at	At1g19890*	ATMGH3/MGH3 (MALE-GAMETE-SPECIFIC HISTONE H3; Okada et al., 2005)	11,305
256313_s_at	At1g35850 | At5g59280	APUM17 (ARABIDOPSIS PUMILIO17); PUM16 (ARABIDOPSIS PUMILIO16)	335
245720_at	At5g04210	RNA recognition motif (RRM)-containing protein	129
251572_at	At3g58390	Eukaryotic release factor 1 family protein/eRF1 family protein	69
266933_at	At2g07760	Zinc knuckle (CCHC-type) family protein	50
253192_at	At4g35370	Transducing family protein/WD-40 repeat family protein	647
247016_at	At5g66970	GTP binding	55
255124_at	At4g08560	APUM15 (ARABIDOPSIS PUMILIO15)	60
Hydrolase activity			
264896_at	At1g23210	Glycosyl hydrolase family 9 protein	10,415
261217_at	At1g32850	Ubiquitin carboxyl-terminal hydrolase family protein	1,865
257378_s_at	At2g02290 | At5g23470	NLI interacting factor (NIF) family protein	389
258725_at	At3g09620	DEAD/DEAH-box helicase, putative	210
245447_at	At4g16820	Lipase class 3 family protein	157
262042_at	At1g80140	Glycoside hydrolase family 28 protein/polygalacturonase (pectinase) family protein	69
261278_at	At1g05800	Lipase class 3 family protein	46
258740_at	At3g05780	Lon protease, putative	661
Kinase activity			
263577_at	At2g17090	Protein kinase family protein	3,470
255892_at	At1g17910	Wall-associated kinase, putative	117
254009_at	At4g26390	Pyruvate kinase, putative	1,179
253128_at	At4g36070	CPK18 (calcium-dependent protein kinase 18)	48
Other binding			
263304_at	At2g01920*	Epsin N-terminal homology (ENTH) domain-containing protein/clathrin assembly protein-related	1,579
257372_at	At2g43220	DC1 domain-containing protein	1,300
247737_at	At5g59200	Pentatricopeptide (PPR) repeat-containing protein	1,212
246388_at	At1g77405	Unknown protein	208
267047_at	At2g34370	Pentatricopeptide (PPR) repeat-containing protein	122
248898_at	At5g46370	KCO2 (CA2+ ACTIVATED OUTWARD RECTIFYING K+ CHANNEL2)	914
Other enzyme activity			
263217_at	At1g30740	FAD-binding domain-containing protein	1,104
263041_at	At1g23320	Alliinase family protein	103
256760_at	At3g25650	ASK15 (ARABIDOPSIS SKP1-LIKE15)	34
Protein binding			
256408_at	At1g66610	Seven in absentia (SINA) protein, putative	76
263301_x_at	At2g04970 | At2g06440 | At2g14140 | At2g15200	Heat shock protein (At2g04970, At2g14140); Unknown protein (At2g06440, At2g15200)	56
Transcription factor activity			
253153_at	At4g35700	Zinc finger (C2H2 type) family protein	16,885
264269_at	At1g60240	Apical meristem formation protein-related	11,364
263417_at	At2g17180	Zinc finger (C2H2 type) family protein	567
255514_s_at	At5g65330	MADS-box family protein	85
Transferase activity			
251794_at	At3g55590	GDP-Man pyrophosphorylase, putative	51
Transporter activity			
260850_at	At1g21870	Glc-6-P/phosphate translocator-related	705
Unknown			
251384_at	At3g60760*	Unknown protein	4,936
249468_at	At5g39650	Unknown protein	4,326
252258_at	At3g49450*	F-box family protein	3,779

(*Table continues on following page.*)

Table II. (*Continued from previous page.*)

Probe Set ID	AGI ID	Description	Signal
248072_at	At5g55680	Gly-rich protein	1,903
249981_at	At5g18510	Unknown protein	1,830
256047_at	At1g07060	Unknown protein	1,578
257603_at	At3g13820	F-box family protein	1,400
247341_at	At5g63720	Unknown protein; similar to IMP dehydrogenase/GMP reductase [*M. truncatula*]	962
257443_at	At2g22050	Unknown; similar to Cyclin-like F-box; Ser/Thr protein phosphatase, BSU1 [*M. truncatula*]	873
254183_at	At4g23960	F-box family protein	730
254883_at	At4g11720	GCS1/HAP2 (GENERATIVE CELL-SPECIFIC1; von Besser et al., 2006)	662
248196_at	At5g54150	Unknown protein	658
251201_at	At3g63020*	Unknown protein	443
258431_at	At3g16580	F-box family protein	290
257993_at	At3g19890*	F-box family protein	255
257980_at	At3g20760	Unknown protein	225
267049_at	At2g34210	KOW domain-containing transcription factor family protein	1,179
248214_at	At5g53670	Unknown protein	174
257585_at	At3g12420	Pseudogene of 3′-5′ exonuclease containing protein	163
267399_at	At2g44195	Unknown protein	151
254184_at	At4g23970	Unknown protein	131
246638_at	At5g34880	Gypsy-like retrotransposon family	107
245135_at	At2g45230	Non-LTR retrotransposon family (LINE)	87
258319_at	At3g22700	F-box family protein	83
255399_at	At4g03750	Unknown protein	82
265606_s_at	At3g09510 | At2g31520 | At2g25550	Unknown protein	70
249257_at	At5g41640	Unknown protein	67
254325_at	At4g22650	Unknown protein	61
265838_at	At2g14550	Pseudogene	61
259465_at	At1g19030	Unknown protein	60
247728_at	At5g59510	DVL18/RTFL5 (ROTUNDIFOLIA LIKE5)	51
257998_at	At3g27510	Unknown protein	50
264732_at	At1g62160	Pseudogene	49
259634_at	At1g56380	Mitochondrial transcription termination factor family protein/mTERF family protein	47
262472_at	At1g50160	Unknown protein	46
257843_at	At3g28400*	hAT-like transposase family (hobo/Ac/Tam3)	34
245754_at	At1g35183	Unknown protein	26

Small RNA and DNA Methylation Pathways

From analysis of maize sperm ESTs and their respective Arabidopsis homologs (Engel et al., 2003), we had previously hypothesized that small RNA pathways might be partially activated in Arabidopsis sperm cells (Pina et al., 2005), which stood in contrast to an apparent complete lack of expression in mature pollen grains, i.e. absent calls for all genes in these pathways (Pina et al., 2005). However, our new pollen data set shows present calls for three of the 15 transcripts that were below the detection threshold in our previous study (*AGO1*|*At1g48410*, *AGO4*|*At2g27040*, and *AGO9*|*At5g21150*; Supplemental Table S1). Given that the biological material was obtained in the same way in the two studies and that the expression values for the three AGO transcripts are relatively low, our best explanation for this discrepancy is that improved chemistry for sample processing, array hybridization, and staining resulted in a better signal to noise ratio and thus a higher sensitivity.

Since our last study (Pina et al., 2005), the list of genes known or thought to be involved in small RNA pathways has increased significantly. When we include genes implicated in RNA-directed DNA methylation (RdDM), maintenance of DNA methylation, and active demethylation (for review, see Vazquez, 2006; Matzke et al., 2007), 18 of the 53 genes are expressed in sperm cells (Supplemental Table S1). Five of these transcripts (*AGO9*|*At5g21150*, *DDM1*|*At5g66750*, *DRB4*|*At3g62800*, *MET1*|*At5g49160*, and *SUVH5*|*At2g35160*) are so highly enriched in sperm cells that their detection in pollen is very likely to be sperm cell derived. Besides the cytosine methyltransferase (MTase) *MET1*, a number of other transcripts expressed in sperm cells are implicated in maintenance of DNA methylation; MET1 works in conjunction with HDA6|*At5g63110* to maintain

CG methylation during DNA replication. Notably, the SWI2/SNF2 chromatin-remodeling factor DDM1, which is also involved in maintenance of CG methylation, is highly enriched in sperm cells. The histone-H3-Lys-9 MTases SUVH5, SUVH4 | *At5g13960*, and SUVH6 | *At2g22740* are thought to maintain non-CG methylation via control of the DNA MTase CMT3 | *At1g69770* (Ebbs and Bender, 2006). Interestingly, *CMT3* expression could not be detected in sperm cells, as we confirmed by RT-PCR (Fig. 4B), but the MTase *DRM2* was. This MTase is also involved in the maintenance of non-CG methylation, but more importantly, in RdDM by catalyzing de novo methylation of cytosines in all sequence contexts in conjunction with DRD1 (not represented on the ATH1 array), the Polymerase IVb complex, and AGO4 or AGO6. While the largest subunit of Pol IVb (*NRPD1b* | *At2g40030*) is expressed in sperm cells, expression of its second largest subunit, *NRPD2a* | *At3g23780*, could not be detected (Fig. 4B). *AGO6* is expressed, but *AGO9*, a member of the AGO4/AGO6 subfamily (Zheng et al., 2007), shows much higher expression levels, indicating a possible role for AGO9 in the context of de novo methylation in sperm cells. Small interfering RNAs serve as triggers for de novo methylation, and their genesis involves RDR2 | *At4g11130*, HEN1 | *At4g20910*, and DCL3 | *At3g43920*, with *DCL3* apparently not expressed in sperm cells (Fig. 4B). Instead, *DCL1* | *At1g01040* expression was detected, as well as very high expression levels for the *AGO1*-homolog *AGO5* | *At2g27880*. Interestingly, it was shown recently that expression of *OsMEL1*, the rice (*Oryza sativa*) ortholog of *AGO5*, is germ-cell specific and indispensable for premeiotic mitosis and meiosis during sporogenesis (Nonomura et al., 2007). Taken together, these data suggest that sperm cells in mature pollen grains are actively regulating the epigenetic state of their genome through RdDM and maintenance of DNA methylation. This is in concordance with sperm cells of Arabidopsis being in S-phase at anthesis (Friedman, 1999; Durbarry et al., 2005). Perhaps these pathways are necessary to subdue selfish DNA elements after DNA replication. The high expression levels of *AGO5* and *AGO9* as well as the apparent absence of *CMT3* and *DCL3* transcripts suggest that novel small RNA pathways might act in sperm cells. The enriched expression of the dsRNA-binding protein *DRB4* | *At3g62800*, which functions in the trans-acting siRNA pathway (Nakazawa et al., 2007), adds a further interesting piece to this puzzle, as does the enriched expression of the Morpheus' molecule 1 gene (*At1g08060*), which is involved in a DNA methylation-independent epigenetic silencing pathway (Mittelsten Scheid et al., 2002; Vaillant et al., 2006). Notably, most of the transcripts mentioned here can be found in clusters one and six of our functional classification (Table I), highlighting the importance of DNA repair and epigenetic processes in sperm cells.

CONCLUSION

Here, we developed a method to isolate Arabidopsis sperm cells by FACS, thus allowing the first whole-genome transcriptional analysis of plant sperm. Because so little is known about sperm cell biology, we could raise a number of questions by bioinformatics inference, which eventually will open the way to test the respective underlying hypotheses by direct experimental methods. Furthermore, we identified a number of biological processes associated with sperm and identified several targets for reverse genetics analyses as these proteins are likely to play roles in double fertilization.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Transgenic Arabidopsis (*Arabidopsis thaliana*) *AtGEX2::eGFP* plants (Engel et al., 2005) were used for both sperm cell and pollen isolation. The plants were grown for 8 weeks in short-day conditions (8 h light at 21°C–23°C) and then transferred to long-day conditions (16 h light) to induce flowering. Twelve 4-week-old seedlings of Arabidopsis ecotype Columbia were grown on solid medium (B-5 with 0.8% phytoagar; Duchefa) and pooled for RNA extraction. Pollen isolation and FACS were performed as previously reported (Becker et al., 2003). All sample types were processed in triplicate from three individual RNA extractions.

Isolation of Arabidopsis Sperm Cells

To obtain sperm cells from mature pollen, inflorescences from transgenic plants expressing *AtGEX2::eGFP* were collected (filling approximately 300 mL of a glass beaker) and placed in a humid chamber at room temperature for 1 h to hydrate. The sperm extraction (SE) buffer (1.3 mM H_3BO_3, 3.6 mM $CaCl_2$, 0.74 mM KH_2PO_4, 438 mM Suc, 7 mM MOPS, 0.83 mM $MgSO_4$ in double distilled water) was adjusted to pH 6.0 with NaOH, filter-sterilized (0.22 μm filtration, 150 mL Bottle Top Filter with a 45-mm neck, Corning) and vacuum-degassed. Inflorescences were immersed and mixed in 500 mL of SE buffer until the pollen grains were released, yielding a yellowish solution. To remove flowers and other plant tissues, the solution was filtered through Miracloth (Calbiochem), then the pollen was concentrated by passing the filtrate through a 10-μm-pore size mesh filter (SEFAR AG) using a vacuum pump (KNF Aero Mat). The material collected on the filter was subsequently washed with 1 to 2 mL of buffer to release pollen and impurities and then filtered through a 28-μm mesh filter (SEFAR AG), to dispose of debris larger than pollen, directly into a glass homogenizer (Kontes Glass, Tissue Grind Pestle SC and Tissue Grind Tube S2). Pollen grains were subsequently disrupted by applying three circular up and down movements. The resulting solution, containing free sperm cells as well as both burst and intact pollen grains, was filtered through a 15-μm-pore size mesh filter (SEFAR AG) to eliminate most of the debris and the remaining intact pollen grains. To the filtrate enriched in sperm cells, 1 μM DRAQ5 (Biostatus Limited, Alexis) was added to stain genomic DNA and maintained at 4°C before FACS. A scheme summarizing the procedure is available as Supplemental Figure S1.

FACS

Fluorescence-activated sperm cell sorting was performed in a Moflo High-Speed Cell Sorter (Dako-Cytomation), using a 100-μm ceramic nozzle with 30 psi sheath pressure, a 488-nm laser line from a Coherent Sapphire 488 to 200 CDRH laser for eGFP excitation, and a 632.8-nm laser line from a Spectra-Physics 107B 25-mW HeNe laser to excite DRAQ5. GFP and DRAQ5 were detected using a 530-/40-nm and a 700-/75-nm HQ band pass filter, respectively. SE buffer was used as the sheath solution, both for hydrodynamic stability and subsequent analysis of sorted sperm cells. Sperm cells viability after sorting was evaluated by fluorescein diacetate staining according to a procedure previously described for pollen (Heslop-Harrison and Heslop-Harrison, 1970). For subsequent RNA extraction, sperm cells were sorted directly into RNeasy extraction buffer RLT (Qiagen).

RNA Extraction, Biotin-Labeling of cRNA, and Hybridization

Total RNA was isolated using the RNeasy Micro kit (Qiagen) from three biological replicates of sperm cells, pollen, and seedlings. RNA integrity was

assessed using an Agilent 2100 Bioanalyser with an RNA 6000 Nano Assay (Agilent Technologies) and processed for use on Affymetrix Arabidopsis ATH1 Genome Arrays (Santa Clara) according to the manufacturer's Two-Cycle Target Labeling Assay. Briefly, 16 ng of total RNA containing spiked-in Poly-A RNA controls (GeneChip Expression GeneChip Eukaryotic Poly-A RNA Control kit; Affymetrix) was used in an RT reaction (Two-Cycle DNA synthesis kit, Affymetrix) to generate first-strand cDNA. After second-strand synthesis, double-stranded cDNA was used in an in vitro transcription reaction to generate cRNA (MEGAscript T7 kit; Ambion). The cRNA obtained was used for a second round of cDNA and cRNA synthesis, resulting in biotinylated cRNA (GeneChip Expression 3′-Amplification Reagents for in vitro transcription labeling; Affymetrix). The size distribution of the cRNA and fragmented cRNA, respectively, was assessed using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Assay.

Fifteen micrograms of fragmented cRNA was used in a 300-μL hybridization containing added hybridization controls. A total of 200 μL of the mixture was hybridized on arrays for 16 h at 45°C. Standard posthybridization washes and double-stain protocols (FS450_0004) were used on an Affymetrix GeneChip Fluidics Station 450, in conjunction with the GeneChip Hybridization Wash and Stain kit (Affymetrix). Arrays were scanned on an Affymetrix GeneChip scanner 3000 7G. All quality parameters for the arrays were confirmed to be in the recommended range. GeneChip datasets for the nine arrays used in this study are available in a MIAME-compliant format through ArrayExpress (accession no. E-ATMX-35).

Data Analysis and Gene Functional Classification

Absent and present calls were generated using Affymetrix GCOS 1.4 software and all subsequent analyses were conducted using dChip software as of April 2007 (http://www.dchip.org; Wong Lab; see detailed protocol in Supplemental Protocol S1). For downstream analyses, only genes called present in at least two replicates of each sample were considered.

Hierarchical clustering and PCA were computed using Partek Genomics Suite (Partek). We performed comparative analyses of our samples with datasets of leaves, flowers, and siliques (Pina et al., 2005) and with datasets from ovules and unpollinated pistils (L.C. Boavida, F. Borges, J.D. Becker, and J.A. Feijó, unpublished data). The graphical representation SnailView (Becker et al., 2003) was applied to visualize the mean pattern of expression for the 5,829 genes detected in sperm cells coplotted with correspondent pollen data.

To identify genes for which expression has been only detected in sperm cells, we performed comparisons with microarray data of vegetative tissues from previous studies (Pina et al., 2005) and using the Web-based application Genevestigator (Zimmermann et al., 2004). First, we selected genes called present in sperm cells but absent in pollen and vegetative tissues. It was also necessary to set a fold-change cutoff for genes enriched in sperm cells (relative to pollen) to infer which genes detected in pollen might be sperm selective. We found that increasing the fold-change cutoff between the logarithmic expression values of pollen and sperm cells yielded a better correlation and for that reason we used a fold-change of 3, without any direct statistical significance. A list of probe IDs representing 254 genes was uploaded on Genevestigator software (Zimmermann et al., 2004), selecting for analysis only high quality arrays of the AtGenExpress Database (Schmid et al., 2005) from experiments of developmental stages of Arabidopsis. Samples containing mature pollen or pollen undergoing microgametogenesis were excluded. To compare the expression of genes across the selected arrays, we used the Digital northern tool, and genes called present with a P-value ≤ 0.05 in at least two of the three replicates were excluded.

The DAVID gene functional classification tool (Huang et al., 2007) was used to condense the list of genes detected in our sample set into functionally related groups. We used the novel agglomeration method to cluster the three main gene ontology charts (Biological Process, Molecular Function and Cellular Component) in a meaningful network context. The following parameters were used: classification stringency (medium), similarity term overlap (3), similarity threshold (0.50), initial group membership (3), final group membership (3), and multiple linkage threshold (0.50).

RT-PCR Analysis

Nonhybridized cRNA from one replicate of each sample (sperm cells, pollen, seedling, ovule, and silique) was used to prepare double-stranded cDNA. For expression analysis, 5 ng of each template cDNA was used in reactions of 35 PCR cycles. First-strand cDNA synthesis by SuperScript III reverse transcriptase (Invitrogen) was performed on two additional samples of total RNA (approximately 50 ng each) from sperm cells and used, together with cDNA from pollen and leaf samples, in reactions of 40 PCR cycles. The sequences of all the primers used are available as Supplemental Table S3.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Schematic representation of the procedure for isolating Arabidopsis sperm cells.

Supplemental Figure S2. Snail View representation of gene expression in sperm cells and pollen.

Supplemental Figure S3. Overview of the SKP-Cullin-F-box E3 complex.

Supplemental Figure S4. Quality parameters analyzed by Expression Console software (Affymetrix) for RMA (Robust Multichip Analysis) summarized data.

Supplemental Table S1. Expression summary of all genes represented on the ATH1 array and sRNAs and methylation analysis.

Supplemental Table S2. Gene functional analysis by DAVID tools.

Supplemental Table S3. Primers used for RT-PCR analysis.

Supplemental Protocol S1. GeneChip data analysis.

ACKNOWLEDGMENTS

We thank Rob Martienssen and Keith Slotkin (Cold Spring Harbor Laboratory) for fruitful discussions.

Received June 23, 2008; accepted July 27, 2008; published July 30, 2008.

LITERATURE CITED

Alandete-Saez M, Ron M, McCormick S (2008) GEX3, expressed in the male gametophyte and in the egg cell of *Arabidopsis thaliana*, is essential for micropylar pollen tube guidance and plays a role during early embryogenesis. Mol Plant **1:** 586–598

Baarends WM, Hoogerbrugge JW, Roest HP, Ooms M, Vreeburg J, Hoeijmakers JH, Grootegoed JA (1999) Histone ubiquitination and chromatin remodeling in mouse spermatogenesis. Dev Biol **207:** 322–333

Basu D, Mallery E, Szymanski DB (2005) SPIKE1, a DOCK-family protein, is a guanine nucleotide exchange factor for Rho of plants (ROP) and positively regulates the WAVE/ARP2/3 pathway (abstract no. 137). Abstracts of the 16th International Conference on Arabidopsis Research. http://www.arabidopsis.org/news/abstracts.jsp

Becker D, Geiger D, Dunkel M, Roller A, Bertl A, Latz A, Carpaneto A, Dietrich P, Roelfsema MRG, Voelker C, et al (2004) AtTPK4, an Arabidopsis tandem-pore K+ channel, poised to control the pollen membrane voltage in a pH- and Ca2+-dependent manner. Proc Natl Acad Sci USA **101:** 15621–15626

Becker JD, Boavida LC, Carneiro J, Haury M, Feijo JA (2003) Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant Physiol **133:** 713–725

Becker JD, Feijo JA (2007) How many genes are needed to make a pollen tube? Lessons from transcriptomics. Ann Bot (Lond) **100:** 1117–1123

Boavida LC, Becker JD, Feijo JA (2005a) The making of gametes in higher plants. Int J Dev Biol **49:** 595–614

Boavida LC, Vieira AM, Becker JD, Feijo JA (2005b) Gametophyte interaction and sexual reproduction: how plants make a zygote. Int J Dev Biol **49:** 615–632

Chen Z, Tan JL, Ingouff M, Sundaresan V, Berger F (2008) Chromatin assembly factor 1 regulates the cell cycle but not cell fate during male gametogenesis in *Arabidopsis thaliana*. Development **135:** 65–73

del Pozo JC, Estelle M (2000) F-box proteins and protein degradation: an emerging theme in cellular regulation. Plant Mol Biol **44:** 123–128

Doelling JH, Phillips AR, Soyler-Ogretim G, Wise J, Chandler J, Callis J, Otegui MS, Vierstra RD (2007) The ubiquitin-specific protease subfam-

ily UBP3/UBP4 is essential for pollen development and transmission in Arabidopsis. Plant Physiol **145:** 801–813

Dumas C, Knox RB, Gaude T (1985) The spatial association of the sperm cells and vegetative nucleus in the pollen grain of Brassica. Protoplasma **124:** 168–174

Durbarry A, Vizir I, Twell D (2005) Male germ line development in Arabidopsis. *duo* pollen mutants reveal gametophytic regulators of generative cell cycle progression. Plant Physiol **137:** 297–307

Ebbs ML, Bender J (2006) Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. Plant Cell **18:** 1166–1176

Engel ML, Chaboud A, Dumas C, McCormick S (2003) Sperm cells of *Zea mays* have a complex complement of mRNAs. Plant J **34:** 697–707

Engel ML, Davis RH, McCormick S (2005) Green sperm. Identification of male gamete promoters in Arabidopsis. Plant Physiol **138:** 2124–2133

Friedman WE (1999) Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. Development **126:** 1065–1075

Haswell ES, Meyerowitz EM (2006) MscS-like proteins control plastid size and shape in *Arabidopsis thaliana*. Curr Biol **16:** 1–11

Heslop-Harrison J, Heslop-Harrison Y (1970) Evaluation of pollen viability by enzymatically induced fluorescence; intracellular hydrolysis of fluorescein diacetate. Stain Technol **45:** 115–120

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

Hua Z, Meng X, Kao TH (2007) Comparison of *Petunia inflata* S-locus F-box protein (Pi SLF) with Pi SLF like proteins reveals its unique function in S-RNase based self-incompatibility. Plant Cell **19:** 3593–3609

Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol **8:** R183

Ingouff M, Hamamura Y, Gourgues M, Higashiyama T, Berger F (2007) Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. Curr Biol **17:** 1032–1037

Johnson MA, von Besser K, Zhou Q, Smith E, Aux G, Patton D, Levin JZ, Preuss D (2004) Arabidopsis *hapless* mutations define essential gametophytic functions. Genetics **168:** 971–982

Kirichok Y, Navarro B, Clapham DE (2006) Whole-cell patch-clamp measurements of spermatozoa reveal an alkaline-activated Ca2+ channel. Nature **439:** 737–740

Klahre U, Becker C, Schmitt AC, Kost B (2006) Nt-RhoGDI2 regulates Rac/Rop signaling and polar cell growth in tobacco pollen tubes. Plant J **46:** 1018–1031

Klahre U, Kost B (2006) Tobacco RhoGTPase ACTIVATING PROTEIN1 spatially restricts signaling of RAC/Rop to the apex of pollen tubes. Plant Cell **18:** 3033–3046

Lathrop WF, Carmichael EP, Myles DG, Primakoff P (1990) cDNA cloning reveals the molecular structure of a sperm surface protein, PH-20, involved in sperm-egg adhesion and the wide distribution of its gene among mammals. J Cell Biol **111:** 2939–2949

Liu Y, Tewari R, Ning J, Blagborough AM, Garbom S, Pei J, Grishin NV, Steele RE, Sinden RE, Snell WJ, et al (2008) The conserved plant sterility gene *HAP2* functions after attachment of fusogenic membranes in Chlamydomonas and Plasmodium gametes. Genes Dev **22:** 1051–1068

Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJ (2007) Targets of RNA-directed DNA methylation. Curr Opin Plant Biol **10:** 512–519

Mishra G, Zhang WH, Deng F, Zhao J, Wang XM (2006) A bifurcating pathway directs abscisic acid effects on stomatal closure and opening in Arabidopsis. Science **312:** 264–266

Mittelsten Scheid O, Probst AV, Afsar K, Paszkowski J (2002) Two regulatory levels of transcriptional gene silencing in Arabidopsis. Proc Natl Acad Sci USA **99:** 13659–13662

Mori T, Kuroiwa H, Higashiyama T, Kuroiwa T (2006) Generative Cell Specific 1 is essential for angiosperm fertilization. Nat Cell Biol **8:** 64–71

Nakazawa Y, Hiraguri A, Moriyama H, Fukuhara T (2007) The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the trans-acting siRNA pathway. Plant Mol Biol **63:** 777–785

Navarro B, Kirichok Y, Clapham DE (2007) KSper, a pH-sensitive K+ current that controls sperm membrane potential. Proc Natl Acad Sci USA **104:** 7688–7692

Nibau C, Wu HM, Cheung AY (2006) RAC/ROP GTPases: 'hubs' for signal integration and diversification in plants. Trends Plant Sci **11:** 309–315

Nonomura K, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, Kurata N (2007) A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell **19:** 2583–2594

Nowack MK, Grini PE, Jakoby MJ, Lafos M, Koncz C, Schnittger A (2006) A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. Nat Genet **38:** 63–67

Okada T, Bhalla PL, Singh MB (2006) Expressed sequence tag analysis of *Lilium longiflorum* generative cells. Plant Cell Physiol **47:** 698–705

Okada T, Endo M, Singh MB, Bhalla PL (2005) Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant *AtMGH3*. Plant J **44:** 557–568

Okada T, Singh MB, Bhalla PL (2007) Transcriptome profiling of *Lilium longiflorum* generative cells by cDNA microarray. Plant Cell Rep **26:** 1045–1052

Ostermeier GC, Dix DJ, Miller D, Khatri P, Krawetz SA (2002) Spermatozoal RNA profiles of normal fertile men. Lancet **360:** 772–777

Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA (2004) Reproductive biology: delivering spermatozoan RNA to the oocyte. Nature **429:** 154

Patton EE, Willems AR, Tyers M (1998) Combinatorial control in ubiquitin-dependent proteolysis: Don't Skp the F-box hypothesis. Trends Genet **14:** 236–243

Pina C, Pinto F, Feijo JA, Becker JD (2005) Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol **138:** 744–756

Rotman N, Durbarry A, Wardle A, Yang WC, Chaboud A, Faure JE, Berger F, Twell D (2005) A novel class of MYB factors controls sperm-cell formation in plants. Curr Biol **15:** 244–248

Sakai N, Sawada MT, Sawada H (2004) Non-traditional roles of ubiquitin-proteasome system in fertilization and gametogenesis. Int J Biochem Cell Biol **36:** 776–784

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet **37:** 501–506

Schwager KM, Calderon-Villalobos LI, Dohmann EM, Willige BC, Knierer S, Nill C, Schwechheimer C (2007) Characterization of the VIER F-BOX PROTEINE genes from *Arabidopsis* reveals their importance for plant growth and development. Plant Cell **19:** 1163–1178

Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AG, Huang S, Kao TH (2004) Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature **429:** 302–305

Singh MB, Xu HL, Bhalla PL, Zhang ZJ, Swoboda I, Russell SD (2002) Developmental expression of polyubiquitin genes and distribution of ubiquitinated proteins in generative and sperm cells. Sex Plant Reprod **14:** 325–329

Strompen G, Dettmer J, Stierhof YD, Schumacher K, Jurgens G, Mayer U (2005) Arabidopsis vacuolar H+-ATPase subunit E isoform 1 is required for Golgi organization and vacuole function in embryogenesis. Plant J **41:** 125–132

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J **37:** 914–939

Toledo-Ortiz G, Huq E, Quail PH (2003) The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell **15:** 1749–1770

Vaillant I, Schubert I, Tourmente S, Mathieu O (2006) MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. EMBO Rep **7:** 1273–1278

Vazquez F (2006) Arabidopsis endogenous small RNAs: highways and byways. Trends Plant Sci **11:** 460–468

Voelker C, Schmidt D, Mueller-Roeber B, Czempinski K (2006) Members of the Arabidopsis AtTPK/KCO family form homomeric vacuolar channels in planta. Plant J **48:** 296–306

von Besser K, Frank AC, Johnson MA, Preuss D (2006) Arabidopsis *HAP2 (GCS1)* is a sperm-specific gene required for pollen tube guidance and fertilization. Development **133:** 4761–4769

Xiao W, Jang J (2000) F-box proteins in Arabidopsis. Trends Plant Sci **5:** 454–457

Xu HL, Swoboda I, Bhalla PL, Sijbers AM, Zhao CX, Ong EK, Hoeijmakers

JHJ, Singh MB (1998) Plant homologue of human excision repair gene *ERCC1* points to conservation of DNA repair mechanisms. Plant J **13:** 823–829

Xu HL, Swoboda I, Bhalla PL, Singh MB (1999a) Male gametic cell-specific expression of *H2A* and *H3* histone genes. Plant Mol Biol **39:** 607–614

Xu HL, Swoboda I, Bhalla PL, Singh MB (1999b) Male gametic cell-specific gene expression in flowering plants. Proc Natl Acad Sci USA **96:** 2554–2558

Yu HS, Hu SY, Zhu C (1989) Ultrastructure of sperm cells and the male germ unit in pollen tubes of *Nicotiana tabacum.* Protoplasma **152:** 29–36

Zhang Y, McCormick S (2007) A distinct mechanism regulating a pollen-specific guanine nucleotide exchange factor for the small GTPase Rop in *Arabidopsis thaliana.* Proc Natl Acad Sci USA **104:** 18830–18835

Zheng X, Zhu J, Kapoor A, Zhu JK (2007) Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. EMBO J **26:** 1691–1701

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol **136:** 2621–2632