Plant Physiology Preview. Published on July 30, 2008, as DOI:10.1104/pp.108.125229

To be submitted to Plant Physiology

1 **Running head:** Sperm cell transcriptome in *Arabidopsis thaliana*

2

3 Prof. José A. Feijó

4

5 Depto. Biologia Vegetal, Faculdade de Ciências da Universidade de Lisboa; Lisboa,

6 Campo Grande, 1700, Portugal

7 Instituto Gulbenkian de Ciência; Centro de Biologia do Desenvolvimento, Oeiras, 2780-

8 901, Portugal

9

10 Phone: +351 214407941

11 Fax: +351 214407970

12 E-mail: jfeijo@fc.ul.pt

13

14

15

16

17

18 **Research Area:** Genome Analysis

19

20

21

22 **Keywords:** Arabidopsis, male gametes, FACS, transcriptomics

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.
Copyright 2008 by the American Society of Plant Biologists

1 # Comparative Transcriptomics of *Arabidopsis thaliana* Sperm Cells

2

3 Filipe Borges[1], Gabriela Gomes[1,*], Rui Gardner[1], Nuno Moreno[1], Sheila McCormick[2],

4 José A. Feijó[1,3,§], Jörg D. Becker[1]

5

6 [1] Instituto Gulbenkian de Ciência; Centro de Biologia do Desenvolvimento, Oeiras,

7 2780-901, Portugal

8 [2] Plant Gene Expression Center, USDA/ARS-UC-Berkeley, 800 Buchanan St., Albany,

9 CA, 94710

10 [3] Depto. Biologia Vegetal, Faculdade de Ciências da Universidade de Lisboa; Lisboa,

11 Campo Grande, 1700, Portugal

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 **Footnotes**

2 * Current address: Centro de Engenharia Biológica e Química, Instituto Superior
3 Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
4 [§] Corresponding author; Phone/Fax: +351 214407941/+351 214407970, jfeijo@fc.ul.pt
5

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 **ABSTRACT**

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

23

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 **INTRODUCTION**

2 In angiosperms, meiosis in the anthers yields haploid unicellular microspores.
3 Subsequently, pollen mitosis I (PM I) yields a larger vegetative cell and a smaller
4 generative cell (GC). The GC undergoes PM II, a symmetric division that yields two
5 sperm cells. In *Arabidopsis thaliana* PM II occurs before anthesis, so that three-celled
6 pollen grains (a vegetative cell and two sperm cells within the vegetative cell
7 cytoplasm) are later released from the anthers (Dumas et al., 1985; Yu et al., 1989;
8 Boavida et al., 2005). When the male gametophyte (pollen grain) meets the papillae of a
9 receptive stigma, a complex series of cell-cell signalling events will drive pollen tube
10 growth towards the embryo sac (female gametophyte). Upon arrival the pollen tube tip
11 bursts, discharging the two sperm cells. To achieve double fertilization, each sperm cell
12 fuses with an egg or a central cell, to yield the zygote and primary endosperm cell,
13 respectively (Boavida et al., 2005), but the mechanisms underlying double fusion
14 remain relatively unknown. It was recently shown that a mutation in the Arabidopsis
15 *CDC2A* gene has a paternal effect, whereby mutant pollen produces only one sperm cell
16 that exclusively fertilizes the egg cell (Nowack et al., 2006). Although this mutant
17 arrested embryo development early, its single fertilization event somehow triggered
18 autonomous proliferation of the endosperm. The loss of activity of the Chromatin
19 Assembly Factor 1 (CAF1) also prevents PMII, but in this case the resulting single
20 sperm cell could fertilize either female gamete (Chen et al., 2008).

21 To date a few genes have been described as specifically expressed in Arabidopsis
22 sperm cells; some of these appear to be important for pollen tube guidance (von Besser
23 et al., 2006) or for regulating their own transcriptional program (Okada et al., 2005).
24 Moreover, post-mitotic male gametes of Arabidopsis continue to progress through the S
25 phase of cell cycle as they move through the pollen tube, reaching the G2 phase just

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 before double fertilization (Friedman, 1999). These findings suggest that there is an
2 intricate control over molecular programs in the sperm cells that are coupled to cell
3 cycle transitions.

4 Although sperm cells have highly condensed chromatin, the original assumption that
5 they are transcriptionally quiescent was proven wrong by the activity of *AtGEX1*
6 promoter, whose expression is confined to sperm cells, and is not expressed in the
7 progenitor generative cell (Engel et al., 2005). A cDNA library was constructed from
8 FACS-purified sperm cells of *Zea mays*, which revealed a diverse complement of
9 mRNAs representing at least 2,560 genes (Engel et al., 2003). These findings were
10 significantly more than the restricted number of transcripts previously described in plant
11 sperm cells (Xu et al., 1999, 1999; Singh et al., 2002; Okada et al., 2005). In addition
12 they led to the identification of *AtGEX1* and *AtGEX2* (Engel et al., 2005), whose
13 promoters drive expression in Arabidopsis sperm cells. Additional sperm-specific genes
14 have been reported more recently (Mori et al., 2006; von Besser et al., 2006; Ingouff et
15 al., 2007), including *DUO1*, a MYB transcription factor in *A. thaliana* that is important
16 for the G2/M transition during pollen mitosis II (Rotman et al., 2005). Transcriptional
17 profiling of generative cells of *Lilium longiflorum* (Okada et al., 2006; Okada et al.,
18 2007) revealed a considerable overlap with the maize sperm ESTs (Engel et al., 2003)
19 and the Arabidopsis pollen microarray datasets (Becker et al., 2003; Honys and Twell,
20 2003). Despite this previous work, the male gamete ESTs from maize, Plumbago, and
21 lily were limited, so that testing the potential importance of sperm-expressed transcripts
22 by means of reverse genetics could not be comprehensive, even when an Arabidopsis
23 homolog could be identified for such ESTs. Several studies using the Affymetrix ATH1
24 GeneChip platform to characterize the transcriptomes of mature pollen grains and
25 earlier developmental stages demonstrated that this type of analysis can provide starting

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 points from which to dissect the genetic programs driving development and functions of
2 the male gametophyte (reviewed in Becker and Feijo, 2007). But technical difficulties in
3 obtaining sufficient amounts of pure biological material constituted a major hindrance
4 for transcriptional profiling of Arabidopsis sperm cells.

5 Here we used a newly developed protocol, based on Fluorescence-Activated Cell
6 Sorting (FACS), to isolate and purify sperm cells from transgenic *A. thaliana* plants that
7 were expressing enhanced green fluorescent protein under the control of a sperm-
8 specific promoter (*AtGEX2::eGFP*). Consequently, we could use Affymetrix ATH1
9 Genome Arrays to profile the mRNA complement of these male gametes. The direct
10 comparison of their transcriptome with those of pollen and seedlings, as well as with
11 additional ATH1 data sets from a variety of vegetative tissues, showed that the sperm
12 cell transcriptome was distinct. We identified transcripts that were enriched and/or
13 preferentially expressed in sperm cells. Functional classification of sperm-enriched
14 transcripts showed that DNA repair, ubiquitin-mediated proteolysis and proteins
15 required for progression through the cell cycle were over-represented categories. Our
16 analyses provide a number of hypotheses for testing and point out future challenges
17 towards understanding the role of genes expressed in the male gametes before, during,
18 and possibly after double fertilization.
19

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 **RESULTS AND DISCUSSION**

2

3 **Fluorescence-Activated Cell Sorting of Arabidopsis Sperm Cells**

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

22

23 **A Remarkable Diversity of mRNAs in Sperm**

24 To determine the sperm cell transcriptome, we used Arabidopsis GeneChip ATH1
25 Genome arrays, representing 22,392 different genes. Although microarray data sets for

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 seedlings and pollen exist (Pina et al., 2005), we repeated analyses for these samples
2 using identical starting amounts of total RNA (16 ng), since these amounts were less
3 than those used in previous studies in our lab. Correlation coefficients above 0.97
4 between the three replicates of each sample type underline the reproducibility of the
5 array data (see Supplemental Figure 4). Based on the MicroArray Suite (MAS) 5
6 detection algorithm, the mean percentages of detected Present calls were 27% for sperm
7 cells, 33% for pollen and 64% for seedlings, corresponding to 5,829, 7,177 and 14,464
8 genes, respectively. Summarized expression data of all samples are shown in
9 Supplemental Table I.
10 A cDNA library from FACS-purified sperm cells of *Zea mays* yielded around 5,000
11 sequenced ESTs that matched with ~1,385 annotated genes in the Arabidopsis genome
12 (e-values of 1e-8 or smaller) (Engel et al., 2003), although 138 of those were not
13 represented on ATH1. A Venn diagram (Fig. 3) was used to display the number of
14 genes common between our samples and the identifiers of Arabidopsis genes matching
15 the maize sperm cell ESTs (restricted to genes represented on the array). Although
16 differences in the transcriptomes of sperm cells from monocot and dicot plants are
17 expected, it was surprising that only 594 genes were shared between our dataset of
18 Arabidopsis sperm cell-expressed genes and the Arabidopsis genes listed as
19 corresponding to maize sperm ESTs. One explanation of this might be that the maize
20 EST comparisons identified a different family member as the best Arabidopsis match.
21 Another explanation is that cDNA sequencing could have identified mRNAs that are not
22 detected in the microarray analyses. To test these hypotheses, we analyzed 395 maize
23 sperm ESTs for which their putative Arabidopsis homologs were called Absent by the
24 microarray experiment. This analysis confirmed that most (238) were singletons, while
25 the others were mostly represented by 2, 3 or 4 ESTs. These findings support the idea

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 that rare messages might have been sequenced from the maize sperm library but would
2 not have been detected by the microarray. A few genes were represented by multiple
3 ESTs (e.g. ATPase encoding gene represented by 23 ESTs) and these belong to gene
4 families, so that the gene selected as the best match could have been a different member
5 than the one detected in Arabidopsis sperm cells. The overlap between Arabidopsis
6 sperm cells and seedling encompassed 4,757 genes, representing almost the entire
7 sperm transcriptome, with approximately 2,400 of those genes showing enriched
8 expression in sperm. As expected, the vast majority of genes called Present in
9 Arabidopsis sperm cells were also detected in Arabidopsis pollen (3,813). To address
10 the possibility of contamination from RNA derived from the vegetative cell in our
11 FACS-purified sperm cell samples, we screened the dataset for genes reported as
12 pollen-specific (expression confined only to the vegetative cell). A good example can be
13 found in Strompen *et al*. (2005), where isoforms of the Vacuolar H^+-ATPase subunit E
14 were analyzed by means of comparison of VHA-E1-GFP, E2-GFP and E3-GFP fusions
15 expression in mature pollen. E1, the major isoform, was expressed in sperm cells but
16 not in the vegetative cell, E2 was expressed only in the vegetative cell, and E3 was
17 expressed in both the vegetative cell and sperm cells. The genes encoding all three
18 isoforms are represented on ATH1 (*At4g11150*, *At3g08560* and *At1g64200*,
19 respectively). Both E1 and E3 transcripts were detected in sperm cells, whereas E2 was
20 not. All three transcripts were detected in pollen. Another example is
21 *AtVEX1/At5g62850*, specifically expressed in the vegetative cell cytoplasm and nucleus
22 (Engel et al., 2005), and called Absent in the sperm cell sample, but Present in pollen.
23 For further validation of our microarray data, RT-PCR was performed. One chosen
24 target was one of the highest expressed genes in pollen, encoding a carbonic anhydrase
25 protein (*At5g04180*), and whose expression was not detected in sperm cells. In

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 agreement with our microarray data, this transcript could not be detected in sperm cells
2 or in seedling by RT-PCR, but high levels were detected in pollen (Fig. 4A). As a
3 further proof, two additional independent samples of total RNA from sperm cells were
4 used to confirm enrichment for a sperm-specific transcript *AtMGH3|At1g19890* (Okada
5 et al., 2005), and absence of transcripts derived from the vegetative cell,
6 *AtVEX1|At5g62850* (Engel et al., 2005), by RT-PCR (Fig. 4C).
7

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

10 We carried out comparative analyses of the sperm cell transcriptome with
11 transcriptional datasets from representative vegetative tissues of *A. thaliana*. Principal
12 component analysis (PCA) is a robust method to project high-dimensional, global
13 expression data onto the three principal components. The closer two points are in the
14 plot, the more similar the samples are in terms of their global gene expression profile.
15 As shown in Fig. 5B, the first principal component separates sperm cells and pollen
16 from the vegetative tissues, while the second principal component shows a further
17 separation between pollen and sperm cells. The first principal component demonstrates
18 the uniqueness of the expression profile in sperm cells. This was already demonstrated
19 for pollen in previous studies (Pina et al., 2005), but here we can additionally observe
20 that pollen and sperm cells separate along the second principal component, thus
21 emphasizing the differences in their transcriptional profiles. A similar separation of
22 samples can be obtained with hierarchical clustering (Fig. 5A). The DAVID Gene
23 Functional Classification Tool (Huang et al., 2007) was used to condense the list of
24 genes showing enriched expression in sperm cells. This tool agglomerates the genes into
25 functional clusters according to their biological significance and enrichment score, in

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 relation to all genes represented on the array (Table I and Supplemental Table 2). The
2 distribution through the first clusters highlight over-represented GO categories, i.e.
3 DNA metabolism (DNA replication and repair), Ubiquitin-Mediated Proteolysis
4 (ubiquitin ligase activity) and Cell Cycle. Although most of the genes expressed in
5 sperm cells were also detected in pollen, it was previously shown that the over-
6 represented GO terms for genes with enriched expression in pollen are signaling, vesicle
7 trafficking and membrane transport (Pina et al., 2005). Previous studies on sperm cell
8 cycle activity (Friedman, 1999) showed that sperm spend most of their development
9 (from the generative cell undergoing mitosis until entry into the embryo sac) in the S
10 phase of the cell cycle. Plant sperm also have a complex set of transcripts for control of
11 protein fate and degradation through ubiquitination. Interestingly, these processes are
12 also essential during spermatogenesis of mammals, for histone-to-protamine
13 replacement (Baarends et al., 1999), as well as for many other key events in
14 gametogenesis and fertilization (reviewed in (Sakai et al., 2004). In concordance with
15 our results, genes involved in DNA repair, activation of cyclins and ubiquitination were
16 also highly represented in ESTs from generative cells of *Lilium longiflorum* (Xu et al.,
17 1998; Singh et al., 2002; Okada et al., 2006).
18
19 **Identification of Novel mRNAs in Male Gametes: Towards Understanding Their**
20 **Role in Fertilization and Early Development**
21 Genes showing preferential or enriched expression in sperm cells are primary
22 candidates for roles in male gamete development and fertilization. To identify sperm-
23 preferentially expressed transcripts we performed a comparative analysis with a number
24 of previously reported microarray datasets: leaves and siliques (Pina et al., 2005),
25 ovules and unpollinated pistils (Boavida et al., submitted) and AtGenExpress datasets

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 (Schmid et al., 2005) available through the web-based application Genevestigator
2 (Zimmermann et al., 2004), selecting samples that did not contain pollen. Seventy-four
3 genes were identified (Table II) and distributed according to their molecular function (as
4 annotated in TAIR as of May 2007). Notably, this set includes two genes previously
5 reported as sperm-specific: histone H3, *AtMGH3|At1g19890* (Okada et al., 2005) and
6 *HAP2(GCS1)|At4g11720* (von Besser et al., 2006). We performed RT-PCR analysis for
7 several genes within this list, including *AtMGH3* and *HAP2*, in order to confirm the
8 microarray data reported in this study (sperm cells, pollen and seedling), and
9 additionally with ovule and silique samples (Fig. 4D). Although we could not amplify
10 any of the transcripts from the seedling sample, we detected expression for several of
11 these genes in both ovule and silique samples. We can exclude contamination of
12 genomic DNA in the cDNA templates as source of these amplification products,
13 because we used primers amplifying intron-spanning products whenever possible.
14 Surprisingly, one gene whose expression was detected in ovules and siliques is
15 *HAP2(GCS1)|At4g11720*. Previous studies in Arabidopsis showed only a male-specific
16 role for HAP2 (Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006), but on
17 the other hand, RT-PCR analysis indicated that a *HAP2* homolog in *Chlamydomonas*
18 was expressed in both *plus* and *minus* gametes (Mori et al., 2006). Interestingly,
19 although *HAP2* transcripts were detected in *Chlamydomonas plus* gametes, a recent
20 study has shown that HAP2 protein is essential for fusion in *minus* gametes only (Liu et
21 al., 2008). We do not know if the genes we detected are transcribed in the whole ovule
22 or only in a fraction of the cells comprising the ovule, e.g. the egg cell and/or central
23 cell. It was demonstrated that *GEX2*, previously described as specifically expressed in
24 sperm cells (Engel et al., 2005), is expressed also within the female gametophyte
25 (Alandete-Saez et al., 2008).

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

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

24 Two genes (*At3g49450*, *At3g19890*), which according to our RT-PCR analyses are
25 expressed only in sperm cells, encode proteins with an F-Box domain in their N-

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 terminus, and it is notable that the most highly expressed gene in Arabidopsis sperm
2 cells is annotated as an F-box protein (*At3g62230*). We used MAPMAN (Thimm et al.,
3 2004) to display the expression profiles of genes involved in the ubiquitin/26S
4 proteasome pathway. Among the F-Box proteins represented on the array and expressed
5 in both seedlings and sperm cells, most had higher expression values in sperm than in
6 seedlings (Supplemental Fig. 3). Only about 20 of the more than 450 F-Box proteins
7 have had their biological functions elucidated: they are involved in the regulation of
8 diverse cellular processes including cell cycle transitions, transcription, signal
9 transduction, circadian rhythms, floral development (Patton et al., 1998; del Pozo and
10 Estelle, 2000; Xiao and Jang, 2000; Schwager et al., 2007) and gametophytic self-
11 incompatibility (Sijacic et al., 2004; Hua et al., 2007). The C-termini of F-Box proteins
12 usually contain domains conferring substrate-binding specificity, being later targeted for
13 degradation by the ubiquitin-ligase complex SCF (SKP-Cullin-F-Box). Why this
14 diversity and enrichment of F-Box proteins in sperm cells? One possible hypothesis is
15 that proteins needed during earlier stages of microgametogenesis might need to be
16 targeted for degradation in later stages, e.g. cell cycle regulators. A recent study
17 supports it, since the activity of UBP3/UBP4 deubiquitinating enzymes (essential for the
18 Ub/26S proteasome system) is crucial for PMII in Arabidopsis (Doelling et al., 2007).
19 On the other hand these and other transcripts or proteins stored in sperm cells might be
20 delivered to both egg and central cells upon fertilization and play a role in early
21 embryonic development. Comprehensive studies highlighting the presence of RNAs in
22 human spermatozoa focused on its potential to profile past events of spermatogenesis
23 and suggested a role for certain sperm RNAs in early zygotic development (Ostermeier
24 et al., 2002). Those particular transcripts were later shown to be delivered to the oocyte
25 at fertilization (Ostermeier et al., 2004).

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1

2 **Transcription Factors**

3 The *AtGEX1* and *AtGEX2* promoters, as well as other sperm-expressed promoters
4 discussed in Engel et al. (2005) have binding sites for Dof-type transcription factors. Of
5 the 32 Dof-type transcription factors in Arabidopsis, three (*At3g47500*, *At5g39660* and
6 *At5g62430*) were detected in sperm cells; one or more of these probably regulates
7 sperm-specific promoters. There are several MYB-type transcription factors with
8 relatively high expression in sperm cells, while other transcription factor families have
9 restricted representation. For example, there are many homeobox transcription factors in
10 Arabidopsis, but only one of this group (*At3g61150*) has substantial expression in sperm
11 cells. As a group, genes encoding scarecrow transcription factors are mostly not
12 expressed in pollen or in sperm cells, but 3 (*At2g45160*, *At2g29060*, and *At1g63100*) are
13 detected in sperm cells and not in pollen, implying that these 3 might play sperm-
14 specific roles. The bHLH transcription factor family is one of the largest in Arabidopsis
15 (Toledo-Ortiz et al., 2003), but only a few were detected in sperm cells, and one of these
16 (*At2g42300*) is not detected in pollen, again suggesting that it may play a specific role in
17 male gametes. Notably, the zinc finger (C2H2 type) transcription factor *At4g35700*
18 shows the highest expression level of all sperm-preferentially expressed genes identified
19 in this study. This transcription factor might be expressed only in sperm cells although it
20 was also detected in siliques by RT-PCR. Given that the detection level for this
21 transcript is very high in pollen, and higher in sperm cells than in pollen (see
22 Supplemental Fig. 2), it is likely that its weak detection in siliques is pollen-derived.

23

24 **Signal Transduction Pathways**

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 Sperm cells might have unique signalling pathways. For example, the small GTPases of
2 the Rop family are important for many aspects of cytoplasmic signalling, including
3 actin cytoskeleton reorganization (Nibau et al., 2006). The GTP/GDP status of Rops is
4 controlled by the Rop-GEF family, as well as by RopGAPs and ROP-GDIs (Klahre et
5 al., 2006; Klahre and Kost, 2006; Zhang and McCormick, 2007). There are 3 Rops
6 expressed in sperm cells (*At2g17800*, *At3g51300* and *At4g35950*), as well as one of the
7 3 RopGDIs in Arabidopsis (*At3g07880*), and a RopGAP (*At5g61530*). There are 14
8 genes encoding PRONE-type RopGEFs in Arabidopsis, and 5 of these are pollen-
9 specific or selective (Zhang and McCormick, 2007). None of the 14 RopGEFs were
10 detected in sperm cells. Another protein, Spike1 (*At4g16340*) also has RopGEF activity
11 (Basu et al., 2005), but its transcript was not detected in pollen or in sperm cells. It is
12 unclear how the GTP/GDP status of Rops in sperm cells can be controlled if RopGEFs
13 are absent. We should not exclude the hypothesis that some known RopGEFs may be in
14 fact expressed in sperm cells, but their transcripts are below the detection threshold.
15 In sperm cells, some complicated signalling cascades might be simplified. For
16 example, MAP kinase cascades are central to many signaling pathways in plants, and
17 there is often crosstalk between different members in different signaling pathways
18 (Mishra et al., 2006). Of the 20 MAPKKKs in Arabidopsis, several were detected in
19 sperm cells, and two (*MAPKKK19|At3g50310* and *MAPKKK20|At5g67080*) have
20 extremely high expression; these two were not detected in pollen, implying that sperm-
21 specific signaling pathways might exist. Indeed, these MAPKKKs must act on
22 MAPKK3 (*At5g40440*), the only one of the 10 MAPKKs in Arabidopsis that could be
23 detected in sperm cells. Several MAPKs are also expressed, with *At2g01450* showing
24 the highest expression level. Other examples of gene families with restricted expression
25 in sperm cells include the large lectin receptor kinase family, only a few of which were

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 detected, and the auxin efflux carriers, only one (*At5g01990*) of which is expressed in
2 sperm, while a different one (*At2g17500*) is expressed in pollen.
3

4 **Small RNA and DNA Methylation Pathways**

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

17 Since our last study (Pina et al., 2005) the list of genes known or thought to be
18 involved in small RNA pathways has increased significantly. When we include genes
19 implicated in RNA-directed DNA methylation (RdDM), maintenance of DNA-
20 methylation and active demethylation (reviewed in Vazquez, 2006; Matzke et al., 2007),
21 18 of the 53 genes are expressed in sperm cells (Supplemental Table I). Five of these
22 transcripts (*AGO9|At5g21150*, *DDM1|At5g66750*, *DRB4|At3g62800*, *MET1|At5g49160*
23 and *SUVH5|At2g35160*) are so highly enriched in sperm cells that their detection in
24 pollen is very likely to be sperm cell-derived. Besides the cytosine methyltransferase
25 (MTase) *MET1*, a number of other transcripts expressed in sperm cells are implicated in

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 maintenance of DNA-methylation: MET1 works in conjunction with HDA6|*At5g63110*
2 to maintain CG methylation during DNA replication. Notably, the SWI2/SNF2
3 chromatin-remodeling factor DDM1, which is also involved in maintenance of CG
4 methylation, is highly enriched in sperm cells, The histone-H3-lysine-9 MTases
5 SUVH5, SUVH4|*At5g13960* and SUVH6|*At2g22740* are thought to maintain non-CG
6 methylation via control of the DNA MTase CMT3|*At1g69770* (Ebbs and Bender, 2006).
7 Interestingly *CMT3* expression could not be detected in sperm cells as we confirmed by
8 RT-PCR (Fig. 4B), but the MTase *DRM2* was. This MTase is also involved in the
9 maintenance of non-CG methylation, but more importantly in RdDM by catalyzing de
10 novo methylation of cytosines in all sequence contexts, in conjunction with DRD1 (not
11 represented on the ATH1 array), the Polymerase IVb complex and AGO4 or AGO6.
12 While the largest subunit of Pol IVb (*NRPD1b|At2g40030*) is expressed in sperm cells,
13 expression of its second largest subunit *NRPD2a|At3g23780* could not be detected (Fig.
14 4B). *AGO6* is expressed, but *AGO9*, a member of the AGO4/AGO6 subfamily (Zheng
15 et al., 2007), shows much higher expression levels, indicating a possible role for AGO9
16 in the context of de novo methylation in sperm cells. siRNAs serve as triggers for de
17 novo methylation, and their genesis involves RDR2|*At4g11130*, HEN1|*At4g20910* and
18 DCL3|*At3g43920*, with *DCL3* apparently not expressed in sperm cells (Fig. 4B). Instead
19 *DCL1|At1g01040* expression was detected, as well as very high expression levels for
20 the *AGO1*-homolog *AGO5|At2g27880.* Interestingly, it was shown recently that
21 expression of *OsMEL1*, the rice ortholog of *AGO5*, is germ-cell specific and
22 indispensable for premeiotic mitosis and meiosis during sporogenesis (Nonomura et al.,
23 2007). Taken together these data suggest that sperm cells in mature pollen grains are
24 actively regulating the epigenetic state of their genome through RdDM and maintenance
25 of DNA methylation. This is in concordance with sperm cells of Arabidopsis being in S-

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 phase at anthesis (Friedman, 1999; Durbarry et al., 2005). Perhaps these pathways are
2 necessary to subdue selfish DNA elements after DNA replication. The high expression
3 levels of AGO 5 and 9 as well as the apparent absence of CMT3 and DCL3 transcripts
4 suggest that novel small RNA pathways might act in sperm cells. The enriched
5 expression of the dsRNA-binding protein *DRB4*|*At3g62800*, which functions in the
6 trans-acting siRNA pathway (Nakazawa et al., 2007) adds a further interesting piece to
7 this puzzle, as does the enriched expression of Morpheus' molecule 1
8 (*MOM1*|*At1g08060*), which is involved in a DNA-methylation-independent epigenetic
9 silencing pathway (Mittelsten Scheid et al., 2002; Vaillant et al., 2006). Notably most of
10 the transcripts mentioned here can be found in clusters one and six of our functional
11 classification (Table I), highlighting the importance of DNA repair and epigenetic
12 processes in sperm cells.
13

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 **CONCLUSIONS**

2 Here we developed a method to isolate *A. thaliana* sperm cells by fluorescence-
3 activated cell sorting, thus allowing the first whole-genome transcriptional analysis of
4 plant sperm. Because so little is known about sperm cell biology, we could raise a
5 number of questions by bio-informatics inference which eventually will open the way to
6 test the respective underlying hypotheses by direct experimental methods. Furthermore,
7 we identified a number of biological processes associated with sperm and identified
8 several targets for reverse genetics analyses as these proteins are likely to play roles in
9 double fertilization.

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Transgenic *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 21ºC-23ºC) and then transferred to long-day conditions (16 h light) to induce flowering. Twelve 4 week-old seedlings of Arabidopsis ecotype Col-0 were grown on solid medium (B-5 with 0.8% phytoagar; Duchefa, Haarlem, The Netherlands) 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 300mL of a glass beaker) and placed in a humid chamber at room temperature for 1 hour 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 sucrose, 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 Incorporated, New York, USA) 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®, USA), then the pollen was concentrated by passing the filtrate through a 10µm pore size mesh filter (SEFAR AG, Heiden, Switzerland) using a vacuum pump (KNF Aero Mat, USA). The

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 material collected on the filter was subsequently washed with 1-2mL of buffer, to
2 release pollen and impurities, and then filtered through a 28µm mesh filter (SEFAR AG,
3 Heiden, Switzerland), to dispose of debris larger than pollen, directly into a glass
4 homogenizer (Kontes Glass Company, Tissue Grind Pestle SC and Tissue Grind Tube
5 S2, Vineland, New Jersey, USA). Pollen grains were subsequently disrupted by
6 applying 3 circular up and down movements. The resulting solution, containing free
7 sperm cells as well as both burst and intact pollen grains, was filtered through a 15µm
8 pore size mesh filter (SEFAR AG, Heiden, Switzerland) in order to eliminate most of
9 the debris and the remaining intact pollen grains. To the filtrate enriched in sperm cells,
10 1µM DRAQ5™ (Biostatus Limited, Alexis Corporation, Lausen, Switzerland) was
11 added in order to stain genomic DNA, and maintained at 4ºC before FACS. A scheme
12 summarizing the procedure is available as Supplemental Fig. 1.
13
14 **Fluorescence Activated Cell Sorting**
15 Fluorescence-activated sperm cell sorting was performed in a Moflo High-Speed Cell
16 Sorter (Dako-Cytomation Inc. Fort Collins, CO, USA), using a 100 µm ceramic nozzle
17 with 30 psi sheath pressure, a 488 nm laser line from a Coherent Sapphire 488-200
18 CDRH laser for eGFP excitation, and a 632.8 nm laser line from a Spectra-Physics
19 107B 25mW HeNe laser to excite DRAQ5. GFP and DRAQ5 were detected using a
20 530/40 nm and a 700/75 nm HQ band pass filter, respectively. SE buffer was used as
21 the sheath solution, both for hydrodynamic stability and subsequent analysis of sorted
22 sperm cells. Sperm cells viability after sorting was evaluated by fluorescein diacetate
23 staining according to a procedure previously described for pollen (Heslop-Harrison and
24 Heslop-Harrison, 1970). For subsequent RNA extraction, sperm cells were sorted
25 directly into RNeasy extraction buffer RLT (Qiagen, Germany).

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 **RNA extraction, Biotin-Labelling of cRNA and Hybridization**

2 Total RNA was isolated using the RNeasy Micro Kit (Qiagen, Germany), from three
3 biological replicates of sperm cells, pollen and seedlings. RNA integrity was assessed
4 using an Agilent 2100 Bioanalyser with a RNA 6000 Nano Assay (Agilent
5 Technologies, Palo Alto, CA), and processed for use on Affymetrix Arabidopsis ATH1
6 Genome Arrays (Santa Clara, CA, USA) according to the manufacturer's Two-Cycle
7 Target Labeling Assay. Briefly, 16 ng of total RNA containing spiked-in Poly-A RNA
8 controls (GeneChip Expression GeneChip Eukaryotic Poly-A RNA Control Kit;
9 Affymetrix) was used in a reverse transcription reaction (Two-Cycle DNA synthesis kit;
10 Affymetrix) to generate first-strand cDNA. After second-strand synthesis, double-
11 stranded cDNA was used in an *in vitro* transcription (IVT) reaction to generate cRNA
12 (MEGAscript T7 kit; Ambion, Austin, TX). The cRNA obtained was used for a second
13 round of cDNA and cRNA synthesis, resulting in biotinylated cRNA (GeneChip
14 Expression 3'-Amplification Reagents for IVT-Labeling; Affymetrix). The size
15 distribution of the cRNA and fragmented cRNA, respectively, was assessed using an
16 Agilent 2100 Bioanalyzer with a RNA 6000 Nano Assay.

17 Fifteen micrograms of fragmented cRNA was used in a 300-µL hybridization
18 containing added hybridization controls. 200 µL of the mixture was hybridized on
19 arrays for 16 h at 45°C. Standard post-hybridization washes and double-stain protocols
20 (FS450_0004) were used on an Affymetrix GeneChip Fluidics Station 450, in
21 conjunction with the GeneChip Hybridization Wash and Stain kit (Affymetrix). Arrays
22 were scanned on an Affymetrix GeneChip scanner 3000 7G. All quality parameters for
23 the arrays were confirmed to be in the recommended range. GeneChip datasets for the 9
24 arrays used in this study are available in a MIAME-compliant format through
25 ArrayExpress (accession no. E-ATMX-35).

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 **Data Analysis and Gene Functional Classification**

2 Absent and Present calls were generated using Affymetrix GCOS 1.4 software and all
3 subsequent analyses were conducted using dChip software as of April 2007
4 (http://www.dchip.org, Wong Lab, Harvard) (see detailed protocol on Supplemental
5 Material). For downstream analyses, only genes called Present in at least two replicates
6 of each sample were considered.

7 Hierarchical clustering and Principal Component Analysis (PCA) were computed using
8 Partek Genomics Suite (Partek, St. Louis, Missourii, USA). We performed comparative
9 analyses of our samples with datasets of leaves, flowers and siliques (Pina et al., 2005),
10 and with datasets from ovules and unpollinated pistils (Boavida et al., submitted). The
11 graphical representation SnailView (Becker et al., 2003) was applied to visualize the
12 mean pattern of expression for the 5,829 genes detected in sperm cells, co-plotted with
13 correspondent pollen data.

14 To identify genes for which expression has been only detected in sperm cells, we
15 performed comparisons with microarray data of vegetative tissues from previous studies
16 (Pina et al., 2005) and using the web-based application Genevestigator (Zimmermann et
17 al., 2004). First, we selected genes called Present in sperm cells but Absent in pollen
18 and vegetative tissues. It was also necessary to set a fold-change cut-off for genes
19 enriched in sperm cells (relative to pollen) to infer which genes detected in pollen might
20 be sperm-selective. We found that increasing the fold-change cut-off between the
21 logarithmic expression values of pollen and sperm cells yielded a better correlation and
22 for that reason we used a fold change of 3, without any direct statistical significance. A
23 list of probe IDs representing 254 genes was uploaded on Genevestigator software
24 (Zimmermann et al., 2004), selecting for analysis only high quality arrays of the
25 AtGenExpress Database (Schmid et al., 2005) from experiments of developmental

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 stages of *Arabidopsis thaliana*. Samples containing mature pollen or pollen undergoing
2 microgametogenesis were excluded. To compare the expression of genes across the
3 selected arrays we used the Digital Northern tool, and genes called Present with a p-
4 value $\leq$ 0.05 in at least two of the three replicates were excluded.
5 The DAVID gene functional classification tool (Huang et al., 2007) was used to
6 condense the list of genes detected in our sample set into functionally related groups.
7 We used the novel agglomeration method to cluster the three main gene ontology charts
8 (Biological Process, Molecular Function and Cellular Component) in a meaningful
9 network context. The following parameters were used: Classification Stringency
10 (Medium), Similarity Term Overlap (3), Similarity Threshold (0.50), Initial Group
11 Membership (3), Final Group Membership (3), Multiple linkage threshold (0.50).
12

13 **RT-PCR Analysis**

14 Non-hybridized cRNA from one replicate of each sample (sperm cells, pollen, seedling,
15 ovule and silique) was used to prepare double-stranded cDNA. For expression analysis,
16 5 ng of each template cDNA was used in reactions of 35 PCR cycles. First Strand
17 cDNA synthesis by SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA) was
18 performed on two additional samples of total RNA (~50ng each) from sperm cells and
19 used, together with cDNA from pollen and leaf samples, in reactions of 40 PCR cycles.
20 The sequences of all the primers used are available as Supplemental Table III.

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 **SUPPLEMENTAL MATERIAL**

2

3 **Supplemental Table I -** Expression summary of all genes represented on the ATH1
4 array and sRNAs and Methylation Analysis. Average of signal values and
5 Present/Absent calls in sperm cells, pollen and seedling samples, highlighting sperm-
6 enriched, -depleted, -selective and preferentially expressed genes. The comparison of
7 gene expression between all three samples is represented by the values of fold change.
8 Gene families involved in the Small RNA and DNA methylation pathways were
9 analysed separately in terms of enrichment in sperm cells relative to seedlings.

10

11 **Supplemental Table II** - Gene functional Analysis by DAVID tools. Functional
12 Clustering of genes detected in sperm cells, pollen, seedlings and highlighting those
13 with enriched expression in sperm, for which the genes representing the first fifteen
14 clusters with a higher score are presented.

15

16 **Supplemental Table III** - Primers used for RT-PCR analysis.

17

18 **Supplemental Figure 1** - Schematic representation of the procedure for isolating
19 Arabidopsis sperm cells.

20

21 **Supplemental Figure 2** - Snail View representation of gene expression in sperm cells
22 and pollen. The expression levels of 5,829 expressed genes in Arabidopsis sperm cells
23 (dots) were co-plotted with the expression levels in pollen (line). Gene expression
24 values are represented in angular coordinates, in which the angle defines the gene rank
25 (clockwise from middle) and the radius defines the logarithm of gene expression values.

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

Genes were ranked according to increasing expression in pollen. A clear cut line of dots on the left side of the Snail View (dashed line) indicates that highly-expressed sperm cell transcripts contribute to the corresponding expression level of that given gene in pollen

Supplemental Figure 3 - Overview of the SCF E3 complex. Fold-change (comparing the logarithmic values of gene expression in sperm cells relative to seedlings) is represented using an adapted version of the 26S/ubiquitin-dependent proteasome pathway for the MAPMAN tool (Thimm et al., 2004). Colored squares symbolize genes represented on ATH1 that are associated with each component of the SCF E3 complex. Blue: up-regulated in sperm; Red: down-regulated in sperm; White: no change; Gray: Absent in sperm.

Supplemental Figure 4 - Quality parameters analyzed by Expression Console™ software (Affymetrix) for RMA (Robust Multichip Analysis) summarized data. (a) Pearson's Correlation of signal intensities. (b) MvA plots show the overall reproducibility of a single replicate within each different sample (sperm cells, pollen and seedling), according to the data variability 'M', as a function of the mean 'A'.

Supplemental Protocol – Genechip data analysis

ACKNOWLEDGMENTS

We thank Rob Martienssen and Keith Slotkin (Cold Spring Harbor Laboratory) for fruitful discussions. This work was supported by grant POCTI/BIA-BCM/60046/2004; PPCDT/BIA-BCM/61270/2004 from Fundacão para a Ciência e a Tecnologia (FCT), Portugal. FB and GG acknowledge a BIC within this project. JDB and RG were

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

1 supported by FCT Post-Doc fellowships SFRH/BPD/31047/2006 and

2 SFRH/BPD/14930/ 2004, respectively.

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

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**:** ssn015

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. *In* Abstracts of the 16th International Conference on Arabidopsis Research, Madison, WI, USA

Becker D, Geiger D, Dunkel M, Roller A, Bertl A, Latz A, Carpaneto A, Dietrich P, Roelfsema MRG, Voelker C, Schmidt D, Mueller-Roeber B, Czempinski K, Hedrich R (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 (2005) The making of gametes in higher plants. Int J Dev Biol **49:** 595-614

Boavida LC, Vieira AM, Becker JD, Feijo JA (2005) 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

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

Doelling JH, Phillips AR, Soyler-Ogretim G, Wise J, Chandler J, Callis J, Otegui MS, Vierstra RD (2007) The Ubiquitin-Specific Protease Subfamily 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

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

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, Billker O (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

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

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

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

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 Rep **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

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

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 (1999) 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 (1999) 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 **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

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

FIGURE LEGENDS

Figure 1. Fluorescence-activated sperm cell sorting based on cell size (forward scatter, FSC), intracellular complexity (side scatter, 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.

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.

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

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

common genes between *Arabidopsis* sperm cells and pollen, maize sperm cells, and Arabidopsis seedlings is 3813, 594, 4757, respectively.

Figure 4. RT-PCR analysis. Gel figures presenting confirmatory RT-PCR analysis for (a) a gene highly expressed in pollen but not detected in sperm cells in our microarray analysis, encoding a Carbonic Anhydrase family protein (*At5g44340*); and (b) genes involved in the sRNAs and RNA-directed DNA Methylation pathways, which are not detectable in sperm cells, *DCL3* (*At3g43920*), *CMT3* (*At1g69770*), *NRPD2a* (*At3g23780*); (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.

Figure 5. Hierarchical clustering and Principal Component Analysis 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; Boavida et al., submitted); (a) Hierarchical clustering dendrogram, using Pearson's Dissimilarity to calculate row dissimilarity and Ward's method for row clustering. (b) Principal Component Analysis.

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

Table I - Functional classification of sperm-enriched genes

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	Cysteine-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

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.

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

Table II - Genes preferentially expressed in sperm cells

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)	11305
256313_s_at	*At1g35850 / At5g59280*	APUM17 (ARABIDOPSIS PUMILIO 17); APUM16 (ARABIDOPSIS PUMILIO 16)	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*	Transducin family protein / WD-40 repeat family protein	647
247016_at	*At5g66970*	GTP binding	54.5
255124_at	*At4g08560*	APUM15 (ARABIDOPSIS PUMILIO 15)	60
Hydrolase Activity			
264896_at	*At1g23210*	Glycosyl hydrolase family 9 protein	10415
261217_at	*At1g32850*	Ubiquitin carboxyl-terminal hydrolase family protein	1865
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	3470
255892_at	*At1g17910*	Wall-associated kinase, putative	117
254009_at	*At4g26390*	Pyruvate kinase, putative	1179
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	1579
257372_at	*At2g43220*	DC1 domain-containing protein	1300
247737_at	*At5g59200*	Pentatricopeptide (PPR) repeat-containing protein	1212
246388_at	*At1g77405*	Unknown protein	208
267047_at	*At2g34370*	Pentatricopeptide (PPR) repeat-containing protein	122
248898_at	*At5g46370*	KCO2 (CA2+ ACTIVATED OUTWARD RECTIFYING K+ CHANNEL 2)	914
Other Enzyme Activity			
263217_at	*At1g30740*	FAD-binding domain-containing protein	1104
263041_at	*At1g23320*	Alliinase family protein	103
256760_at	*At3g25650*	ASK15 (ARABIDOPSIS SKP1-LIKE 15)	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	16885
264269_at	*At1g60240*	Apical meristem formation protein-related	11364
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-mannose pyrophosphorylase, putative	51
Transporter Activity			
260850_at	*At1g21870*	Glucose-6-phosphate/phosphate translocator-related	705
Unknown			
251384_at	*At3g60760**	Unknown protein	4936
249468_at	*At5g39650*	Unknown protein	4326
252258_at	*At3g49450**	F-box family protein	3779
248072_at	*At5g55680*	Glycine-rich protein	1903
249981_at	*At5g18510*	Unknown protein	1830
256047_at	*At1g07060*	Unknown protein	1578
257603_at	*At3g13820*	F-box family protein	1400
247341_at	*At5g63720*	Unknown protein; similar to IMP dehydrogenase/GMP reductase [Medicago truncatula]	962
257443_at	*At2g22050*	Unknown; similar to Cyclin-like F-box; Serine/threonine protein phosphatase, BSU1 [Medicago truncatula]	873
254183_at	*At4g23960*	F-box family protein	730
254883_at	*At4g11720*	GCS1/HAP2 (GENERATIVE CELL-SPECIFIC 1) (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	1179
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 LIKE 5)	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

* Genes analyzed by RT-PCR whose expression has been only detected in sperm cells and in some cases in pollen, but probably being sperm-derived.

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

Seventy four 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; Boavida et al., submitted) and the AtGenExpress database (Schmid et al., 2005), distributed according to their molecular function as annotated in TAIR.

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

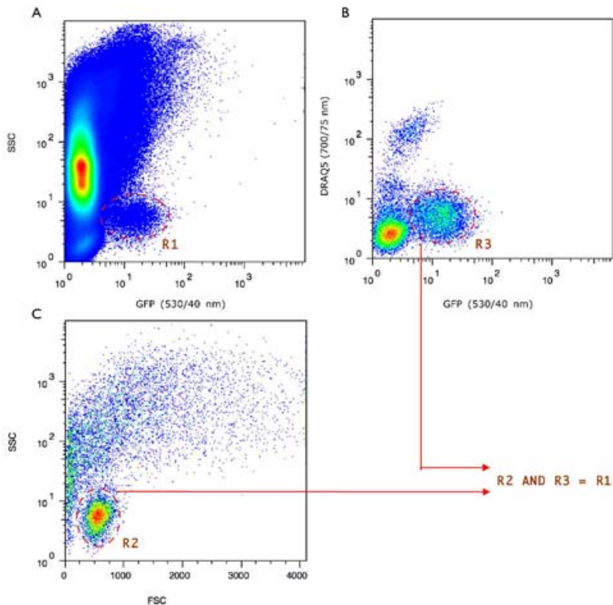

Figure 1. Fluorescence-activated sperm cell sorting based on cell size (forward scatter, FSC), intracellular complexity (side scatter, SSC), GFP signal, and presence of intracellular DNA, via DRAQ5 staining. Low granulosity (low SSC) and GFP positive signals were used to identify the sperm cell population (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.

Downloaded from
Copyright © 2008

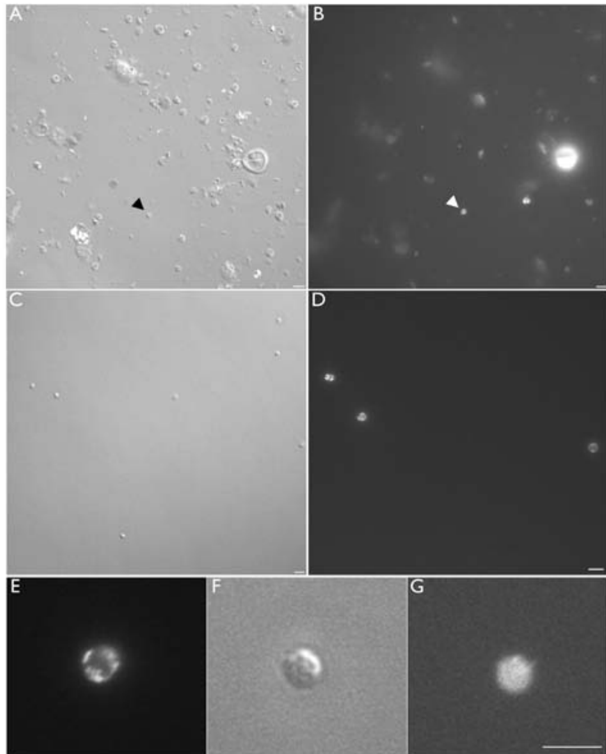

Figure 2. Visualization of FACS-purified sperm cells. Wide-field fluorescence microscopy was used to visualize Arabidopsis sperm cells expressing *AtGEX2::eGFP* (Engel et al., 2005) before (b) and after FACS-purification (d). DIC microscopy confirmed that the debris in the filtrate before FACS (a) was 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.

Downloaded from
Copyright © 2008 A

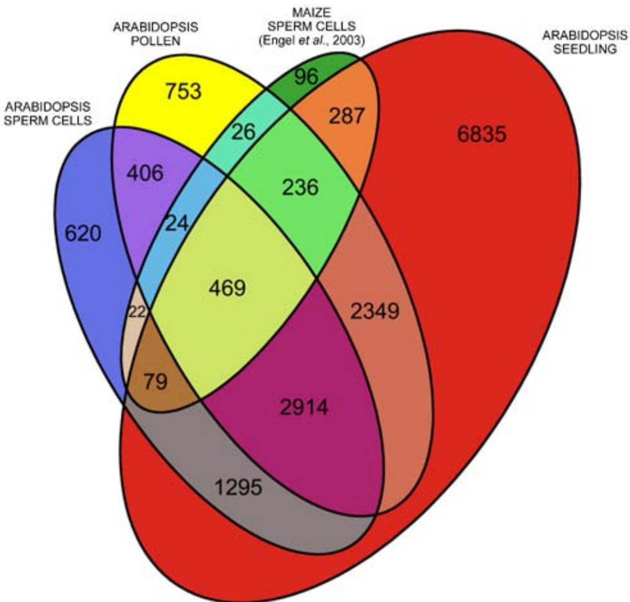

Figure 3. Venn diagram, depicting the overlap between genes whose expression was called present in sperm cells (5,829), pollen (7,177), seedlings (14,464), and additionally the Arabidopsis genes that are represented on the ATH1 array (1,239) and that match those of maize sperm cells ESTs (Engel et al., 2003). Intersection of common genes between *Arabidopsis* sperm cells and pollen, maize sperm cells, and Arabidopsis seedlings is 3813, 594, 4757, respectively.

Downloaded from
Copyright © 2008 A

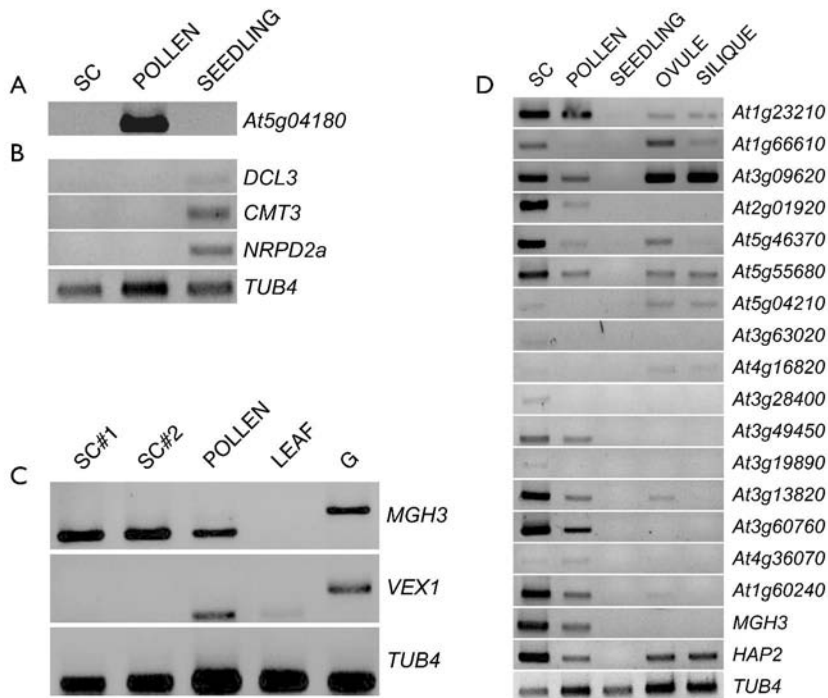

Figure 4. RT-PCR analysis. Gel figures presenting confirmatory RT-PCR analysis for (a) a gene highly expressed in pollen but not detected in sperm cells in our microarray analysis, encoding a Carbonic Anhydrase family protein (*At5g44340*); and (b) genes involved in the sRNAs and RNA-directed DNA Methylation pathways, which are not detectable in sperm cells, *DCL3* (*At3g43920*), *CMT3* (*At1g69770*), *NRPD2a* (*At3g23780*); (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.

Downloaded from on July 24, 2017 - Published by www.plantphysiol.org
Copyright © 2008 American Society of Plant Biologists. All rights reserved.

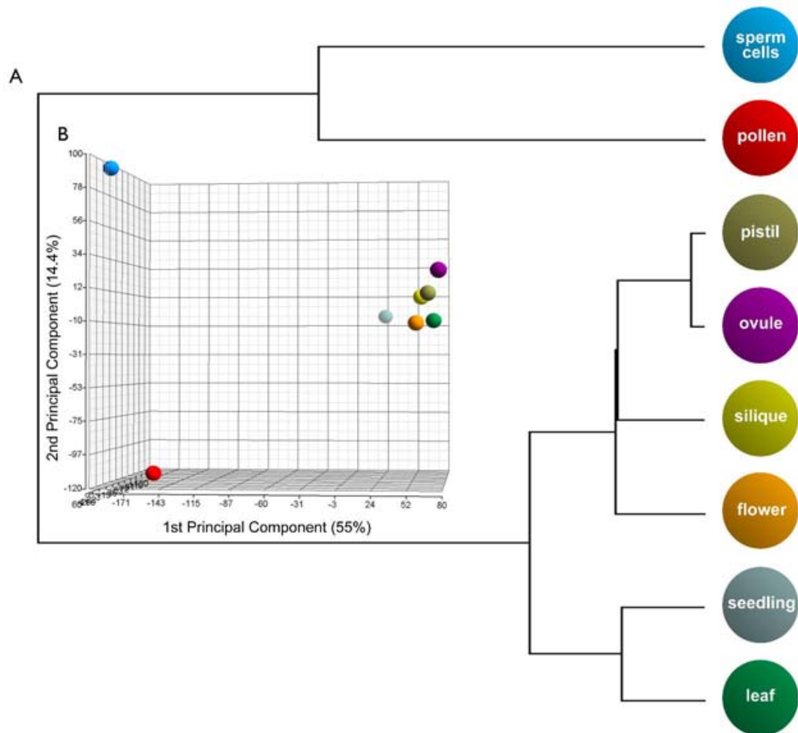

Figure 5. Hierarchical clustering and Principal Component Analysis of tissue-dependent gene expression patterns. Transcriptome data of seedlings, pollen and sperm cells were compared with those of leaf, flower, silique, ovule and unpollinated pistil from previous studies (Pina et al., 2005; Boavida et al., submitted). (a) Hierarchical clustering dendrogram, using Pearson's Dissimilarity to calculate row dissimilarity and Ward's method for row clustering. (b) Principal Component Analysis.

Downloaded from on July 24, 2017 - Published by w
Copyright © 2008 American Society of Plant Biologis