New Phytologist

Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage

Scott D. Russell[1], Xiaoping Gou[1], Chui E. Wong[2], Xinkun Wang[3], Tong Yuan[1], Xiaoping Wei[1], Prem L. Bhalla[2] and Mohan B. Singh[2]

[1]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA; [2]Plant Molecular Biology and Biotechnology Laboratory, Australian Research Council Centre of Excellence for Integrative Legume Research, Melbourne School of Land and Environment, University of Melbourne, Parkville, Victoria 3010, Australia; [3]Higuchi Biosciences Center, University of Kansas, Lawrence, KS 66047, USA

Author for correspondence:
Scott D. Russell
Tel: +1 405 325 4391
Email: srussell@ou.edu

Received: *19 February 2012*
Accepted: *29 April 2012*

New Phytologist (2012) **195**: 560–573
doi: 10.1111/j.1469-8137.2012.04199.x

Key words: angiosperm sperm cells, chromatin modeling, expression profiling, gamete transcriptome, male gamete expression, male germ unit, *Oryza sativa*.

Summary

- Genomic assay of sperm cell RNA provides insight into functional control, modes of regulation, and contributions of male gametes to double fertilization.
- Sperm cells of rice (*Oryza sativa*) were isolated from field-grown, disease-free plants and RNA was processed for use with the full-genome Affymetrix microarray. Comparison with Gene Expression Omnibus (GEO) reference arrays confirmed expressionally distinct gene profiles.
- A total of 10 732 distinct gene sequences were detected in sperm cells, of which 1668 were not expressed in pollen or seedlings. Pathways enriched in male germ cells included ubiquitin-mediated pathways, pathways involved in chromatin modeling including histones, histone modification and nonhistone epigenetic modification, and pathways related to RNAi and gene silencing.
- Genome-wide expression patterns in angiosperm sperm cells indicate common and divergent themes in the male germline that appear to be largely self-regulating through highly up-regulated chromatin modification pathways. A core of highly conserved genes appear common to all sperm cells, but evidence is still emerging that another class of genes have diverged in expression between monocots and dicots since their divergence. Sperm cell transcripts present at fusion may be transmitted through plasmogamy during double fertilization to effect immediate post-fertilization expression of early embryo and (or) endosperm development.

Introduction

Sperm cells represent the male partner that fuses with the egg cell during fertilization in all multicellular eukaryotic organisms. In flowering plants, the male germ lineage is established during pollen mitosis I when the microspore divides to form the large pollen vegetative cell and the much smaller generative cell. This asymmetric division triggers establishment of the generative cell and the developmental divergence of the male germ lineage (Eady *et al.*, 1995). The male germ lineage is contained within the pollen vegetative cell inside the pollen grain and later tube. An ensuing mitotic division results in two nonmotile sperm cells, which become progressively smaller than typical somatic cells during their maturation and passage, occupying < 0.1% of the pollen grain (Russell & Strout, 2005). Although outwardly appearing to be simple cells, these two sperm cells have the capacity to fuse with egg cells and central cells, triggering double fertilization and embryogenesis. Double fertilization initiates remodeling of the egg into a totipotent zygote giving rise to the embryo and remodels the central cell into the nutritive endosperm.

Despite the small size of sperm cells and their dependence on surrounding pollen cytoplasm for nutrition and transport, these cells are transcriptionally active, possessing translation, regulatory and control elements, and a diversity of transcripts (reviewed by Singh *et al.*, 2008). An increasing number of sperm genes have proved to be essential for fertilization and normal embryo establishment, including, for example, *HAPLESS2* (*HAP2*) and *SHORT SUSPENSOR* (*SSP*). *HAP2* is known to encode a sperm-specific, surface-linked protein required for fertilization and is also implicated in directing pollen tubes to their female targets (von Besser *et al.*, 2006; Frank & Johnson, 2009). *SSP* is transmitted as a sperm transcript into the egg cell during gamete fusion and encodes the protein SSP, which activates the developmentally critical asymmetrical division of the zygote, producing a polarized proembryo, which has been shown to establish embryogenesis (Bayer *et al.*, 2009). Other sperm-originating transcripts have also been reported in zygotes and proembryos of tobacco (*Nicotiana tabacum*) (Ning *et al.*, 2006). Thus, accumulating evidence indicates that transcribed sperm products may fulfill essential roles in the successful establishment of the next

generation through untraditional nongenetic mechanisms (Gou *et al.*, 2009; Russell *et al.*, 2010). Sperm cells are also known to form products that may directly communicate with female gametes (Tian *et al.*, 2005), even when outwardly these cells appear to be mere passengers within the elongating pollen tube.

Although the genes controlling fertilization are being identified (Berger, 2008; Russell & Dresselhaus, 2008), only a small proportion of those in the male germ lineage have been fully characterized. The attraction of genomic profiling of transcripts generated by sperm cells lies in the identification of gene candidates selectively upregulated in the male germ line and in understanding their evolutionary involvement in reproductive biology. In Arabidopsis, the first report of what may constitute a canonical sperm transcriptome revealed 5829 transcribed genes using an Affymetrix 24K microarray (Affymetrix, Santa Clara, CA, USA; http://www.affymetrix.com/) (Borges *et al.*, 2008). That study confirmed over-representation of genes associated with DNA repair, ubiquitin-mediated proteolysis, epigenetic labeling and cell cycle progression which were also reported in prior expressed sequenced tag (EST) studies of sperm cells (Gou *et al.*, 2001, 2009; Engel *et al.*, 2003; Okada *et al.*, 2006). However, the evolutionary context of transcription and expression in the male germline will require genomic level investigations in a range of plants (Paterson *et al.*, 2010). This study is the first to extend this range to any monocot or crop plant. Only with expansion of these data will we gain genomic level understanding of the unique contribution of sperm cells to sexual reproduction and their role in fertility and crop productivity as founder cells with direct input into the fusion products of fertilization.

Oryza sativa L. (rice) is a useful model for studies of fertilization because genomic tools are available, and rice has an ideally short progamic phase—from pollination to fertilization—of 15–30 min. In the rice pollen tube, which contains the sperm, essentially no transcriptional changes were detected as compared with mature pollen (Wei *et al.*, 2010). Unlike Arabidopsis, sperm cells of rice are mature; they do not require completion of DNA synthesis and protracted pollen tube growth to effect fertilization (Friedman, 1999). This study provides genomic evidence of highly up-regulated transcripts encoding genes involved in chromatin modification in the male germ lineage, up-regulated pathways regulating miRNA and siRNA processing, and distinct transcription factors and signaling molecules indicating a unique transcriptional profile in the male germline.

Materials and Methods

Sperm and pollen isolation

Sperm cells were isolated from mature anthers of field-grown rice (*Oryza sativa* L. ssp. *japonica*, cultivar 'Katy'), courtesy of the Dale Bumpers National Rice Research Center and University of Arkansas Extension Station near Stuttgart, AR, USA. A centrifugation-based separation method was used for isolating sperm cells (Gou *et al.*, 1999). Collected anthers were immersed in cold 45% (w/v) sucrose solution and ground gently with a small glass rod to release pollen grains. This pollen mixture containing the sperm cells was filtered through 100-μm nylon mesh into a 15-ml conical tube and centrifuged at 300 ***g*** for 3 min. The pellet was resuspended in cold 45% (w/v) sucrose and washed once more, as above. Subsequent purifications were performed using autoclaved solutions on ice. Collected pollen grains were suspended in a 10× volume of 15% (w/v) sucrose solution and gently agitated for 20 min to allow pollen grains to burst, which releases sperm cells and pollen cytoplasm. The mixture was then filtered through 30-μm nylon mesh and the filtrate was layered on a 15%/40% discontinuous Percoll density gradient and centrifuged at 4000***g*** for 40 min at 4°C. The sperm cell-enriched fraction forms a band at the interface between 15% and 40% from which *c.* 100 μl is collected using a pipette. This sample is then diluted with a 4× volume of 15% sucrose solution, and centrifuged at 900***g*** for 10 min. The bottom 2 ml of the centrifuged volume was collected. The pellet was gently suspended and filtered though a 10-μm filter, with the filtrate layered on top of another 40% Percoll density gradient and centrifuged at 4000***g*** for 20 min. The band between the sample solution and 40% Percoll was again collected. This mixture contained concentrated, purified sperm cells. Another 4× volume of 15% sucrose solution was added to the mixture and centrifuged at 900***g*** for 10 min. Approximately 50 μl of purified sperm cells was collected from the bottom of the centrifuge tube and stored frozen at or below −80°C until use. Three biological replicates were isolated and retained for all samples during processing.

RNA preparation

Control tissues included mature rice pollen as a microgametophytic control and young seedlings as a sporophytic control for verifying sperm-enriched or -depleted probe matches. Anthesis rice pollen was isolated according to Russell *et al.* (2008) and frozen in liquid nitrogen. Seedlings grown from rice seed of Katy were germinated in soil, collected at developmental stage V3 (collar forming on leaf 3 of main stem; Counce *et al.*, 2000) and frozen in liquid nitrogen. All samples were stored at or below −80°C until RNA isolation. Total RNA was purified using the RNeasy plant mini kit according to the manufacturer's instructions (Qiagen; http://www.qiagen.com/). The RNA concentration and quality of pollen and seedlings were determined using routine spectrophotometric measurement and agarose gels. For sperm cell RNA, we did not determine concentration because of limited materials from the start. All accumulated isolated sperm RNA and 100 ng of total RNA of seedlings and mature pollen (calculated using spectrophotometric measurements at 260 and 280 nm on a Nanodrop ND-1000 spectrophotometer; NanoDrop Technologies, Wilmington, DE, USA; http://www.nanodrop.com/) were used for probe preparation for each of the three biological replicates performed.

RT-PCR analysis

Total RNA isolated was amplified using the SMART PCR cDNA synthesis kit according to the manufacturer's instructions

(Clontech; http://www.clontech.com/). For each sample, *c.* 10 ng of cDNA was used as the template in a 10-μl reaction volume for PCR amplification of target genes. For the PCR process, 25–30 cycles were used to ensure that the amount of amplified product remained in linear proportion to the initial template present in the reaction. The entire PCR reaction was separated on a 1% agarose gel containing 0.1 μg μl^{-1} ethidium bromide and visualized under UV light. Constitutive *PROFILIN-2* (*LOC_Os06g05880*) was used as an internal control. Real-time PCR analysis for selected transcripts was carried out in triplicate using EXPRESS SYBR® GreenER™ qPCR Supermix Universal kit (Invitrogen) with 1 ng of cDNA template according to the manufacturer's instructions. The starting concentration (expressed in arbitrary fluorescence units) of each transcript in a sample was calculated using LinRegPCR (Ramakers *et al.*, 2003) using raw fluorescence data generated by a Stratagene MX3000P (Invitrogen, Melbourne, Australia). This was then expressed relative to the starting concentration of *PROFILIN-2*. Primers are listed in Supporting Information Table S5.

Oligonucleotide microarray hybridization and data collection

As the amount of starting total RNA was low (in the range of 10–100 ng per sperm cell sample), the Affymetrix GeneChip Two-Cycle cDNA Synthesis Kit was used for target preparation with signal amplification. The Affymetrix 57K Rice Genome GeneChip oligonucleotide microarray was hybridized with 15 μg of fragmented cRNA for 16 h at 45°C, washed, stained, scanned and processed strictly following the Affymetrix GeneChip Expression Analysis Technical Manual as in Russell *et al.* (2008). Microarray data generated from all chips met quality control criteria set by Affymetrix. All data are posted in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database, accession entry GSE17002 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17002) for sperm, pollen vegetative cells and seedlings.

Calls for transcripts being present ($P < 0.05$), absent ($P > 0.065$) and marginal ($0.05 < P < 0.065$) were calculated based on the Affymetrix PM/MM method (MAS 5.0 algorithm, using Affymetrix GCOS version 1.1.1). Pathway level differential expression was determined using the Wilcoxon–Mann–Whitney test; a weighted average (Tukey's biweight) was used to determine fold differences between tissues. Functional classes were calculated by mapping rice orthologs on Arabidopsis BINS of the MAPMAN data-mining program (Thimm *et al.*, 2004), with gene assignments and annotation based on Affymetrix data (Netaffx website http://www.affymetrix.com/analysis/index.affx), rice genomic annotations (Michigan State University (MSU) release 7 and the Rice Annotation Project (RAP) release 5), RiceCyc (http://www.gramene.org/pathway/), GeneBins (Goffard & Weiller, 2007) and DAVID to estimate GO categorical enrichment and depletion (Huang *et al.*, 2007, 2009).

The current MSU annotation of the rice nuclear coding genome (release 7) recognizes 56 081 individual gene loci and 66 433 gene models, 103 loci in the plastid genome and 54 loci in the mitochondrial genome. Probes on the Affymetrix rice 57K GeneChip™ are 25-mer 3'-oligomeric sequences grouped into probe sets that consist of from eight to 16 exact match probes. BLASTn analysis of perfect matches to cDNAs, introns, untranslated regions (UTRs), short models, and intergenic models was used to assign probe sets to loci and to identify exact probe sequence matches and overlaps.

The number of expressed genes was estimated from probe sets by normalizing for multiple sampling by removing duplicate probe sets that map redundantly to a given locus, yielding 34 830 unigene probe sets (of 57 272 noncontrol probe sets). The RAP, which omitted transposable element motifs, reported *c.* 32 000 genes annotated (Itoh *et al.*, 2007). If transposable elements were added to their estimate, the result would be nearly identical to ours. Unanimous present calls were used to provide the most conservative estimate of transcripts expressed. Majority counts (two of three calls from biological replicates) increased gene estimates in pollen from 8101 to 9531, those in sperm from 10 732 to 12 331 and those in seedlings from 15 449 to 16 821. Mature pollen grains were reported to have from 9372 to 12 340 represented probe sets (Wei *et al.*, 2010; their additional file 2a), which generally agrees with our average of 11 830, and is well within the standard error. Wei *et al.* (2010) reported 5939 unigenes in pollen, their count including only genes annotated as 'expressed'. This under estimated the number of genes as The Institute for Genomic Research (TIGR) release 5 and MSU release 6.1 annotations reported only 68.543% and 69.317% of non-transposable element (TE) genes as 'expressed'. In the current MSU release 7 annotation, however, the percentage of genes reported as 'expressed' increased by over 31.2% (90.945%); this would raise their number of estimated expressed genes to *c.* 7793, which is not significantly different from our estimate of 8101 pollen genes.

Results

Sperm and pollen collected from disease-free, field-grown rice plants in three separate fields provided biological replicates for this study. Sperm cells and pollen isolated for this study were examined using interference contrast microscopy and fluorescein diacetate tests for intact, viable cells (Fig. 1a). RT-PCR of selected marker transcripts (Fig. 1b) verified that sperm isolates were strongly enriched and only a few of the most highly expressed pollen genes (Russell *et al.*, 2008) are reported present. *GENERATIVE CELL SPECIFIC 1 (GCS1)* is a sperm cell marker and *PROFILIN-2* is a loading control (Fig. 1b). Pearson's coefficient of correlation confirmed the reproducibility of the data, with mean r values of 0.983 for sperm, 0.993 for pollen, and 0.992 for seedlings. Each of the samples displayed a low correlation between different tissues, including sperm and pollen at $r = 0.250$ (Fig. 1c), sperm and seedlings at $r = 0.239$, and pollen and seedlings at $r = 0.178$. Scatter plots of all data sets are available at http://bomi.ou.edu/russell/plots/. These data reflect high sample consistency, excellent growth conditions and uniformity in biological and technical preparation, as is reflected in correlation matrix results (Fig. 1d).

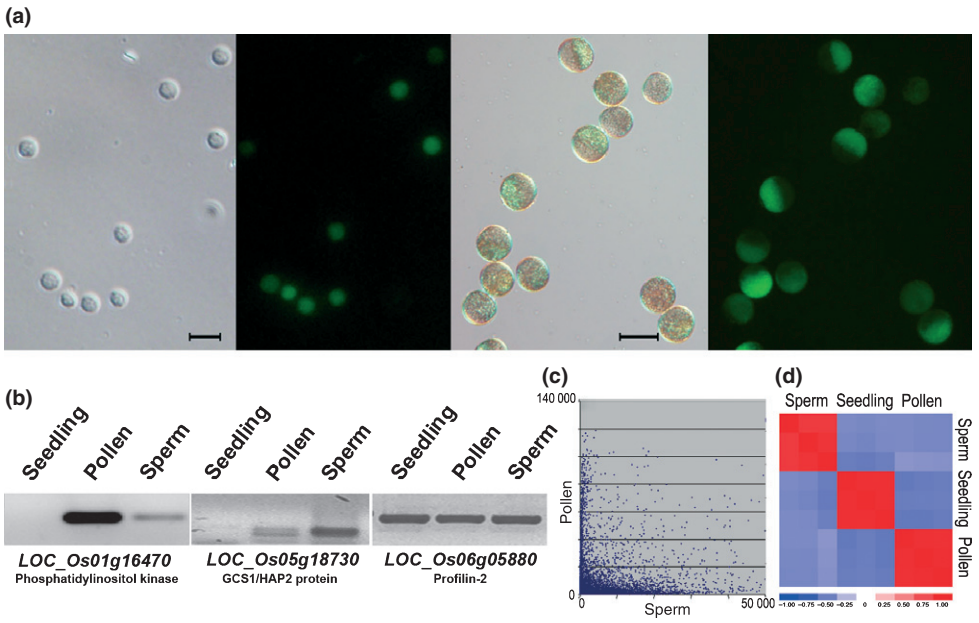

Fig. 1 (a) Differential interference contrast microscopy and fluorescein diacetate viability screen of isolated *Oryza sativa* sperm cells (left; bar, 5 μm) and pollen (right; bar, 50 μm). (b) RT-PCR validation for purity of cell isolations using marker genes for pollen (allergen *Ory s 1*), sperm *GENERATIVE CELL SPECIFIC 1 (GCS1)* and *HAPLESS 2 (HAP2)* and constitutive loading control (*PROFILIN-2*). (c) Scatter plot of pollen vs sperm cell probe set signal intensities. (d) Correlation matrix of respective probe set signal intensities in sperm, pollen and seedling.

Diversity of transcript profile in sperm cells

A Venn diagram shows relative gene expression in sperm, pollen and seedlings (Fig. 2a) based on unanimous present/absent (P/A) calls in triplicate samples. Seedlings expressed 15 449 genes using this method of estimation, followed by 10 732 for sperm cells and 8101 for pollen. Genes representing distinct, nonoverlapping sequences totaled 33 278 for the Affymetrix 57K rice GeneChip. Pollen-specific expression suggests unique transcription of at least 626 genes, representing 1.88% of genes represented on the *Oryza* 57K chip—similar to previous microarray-based reports for Arabidopsis pollen-specific genes, ranging from 737 (Pina *et al.*, 2005) to *c.* 800 in mature pollen (Honys & Twell, 2003). Seedling-specific expression indicates transcription of 5947 genes or *c.* 17.9% of genes represented on the 57K chip. Loci expressed in mature sperm indicate transcription of 1668 products, or an unexpectedly large 5.01% of genes represented on the 57K chip.

Sperm, pollen and seedlings transcribed 5537 gene sequences that appear to be represented in nearly all reference data accessed at NCBI GEO (platform GPL2025). We therefore believe that these may represent universal, potential 'housekeeping genes' of rice, in that they are expressed in all sporophyte and gametophyte cells observed to date. Approximately 20% of such transcripts are more highly represented in sperm cells than in other tissues and *c.* 5% are also highly represented in other reproductive tissues (Supporting Information Fig. S1). Although this might not reflect fundamental changes in function between this subset of up-regulated genes, there appear to be different levels of transcript abundance associated with the initiation of gametophyte development.

Functional categorization of sperm cell transcripts

The most highly represented functional categories in sperm cells involve metabolism, transcription and cell signaling (Fig. 2b).

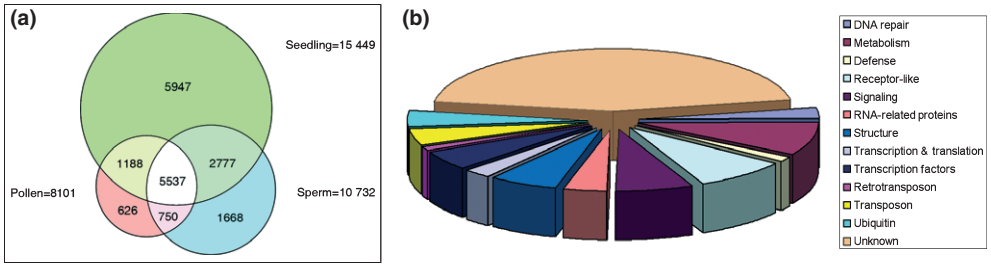

Fig. 2 (a) Venn diagram of genes expressed in different tissues of rice (*Oryza sativa*), including seedling (top), pollen (left) and sperm (right), in circles proportionate to the number of genes. (Probe sets corresponding to 14 785 unigene sequences were reported absent on the 57K chip.) (b) Pie chart showing functional categorization of transcripts expressed in sperm.

Functional categories up-regulated in sperm cells as compared with other tissues include transcription factors, cell signaling, protein modification, cellular identity and receptor-like molecules; these categories may each include some key players in functions unique to sperm cells.

Despite the small volume of male gametes, their short lifespan, and presumably small metabolic contribution to pollen, microarray results (this study and Borges *et al.*, 2008) suggest that sperm cells transcribe a diversity of genes. Table 1 lists a group of especially highly up-regulated sperm- or germline-selective transcripts that show limited expression in other tissues (Fig. 3, rows 1–38). Of the genes apparently restricted to sperm cells, *c.* 62% are unclassified, or 18% higher than the number of unclassified genes in seedlings. Some sperm transcripts share up-regulation with female germ lineages (Fig. 3, rows 39–45), supporting the theory that some genes may be conserved in the germline

Table 1 Highly sperm-enriched genes from the rice (*Oryza sativa*) genome rice microarray with $\log_2$ signal intensity based on dChip normalization

Probe set ID	Locus[1]	Annotation (MSU7, RAP5)	Intensity $\log_2$	Homology
Os.10737.1.S1_at	LOC_Os05g18730.1	Generative cell specific-1 (GCS1/HAP2)	14.20	At
Os.54874.1.S1_at	LOC_Os09g27040.1	GAMETE EXPRESSED 1 (GEX1)	12.47	Zm, At
OsAffx.17894.1.S1_at	LOC_Os09g25650.1	GAMETE EXPRESSED 2 (GEX2)	14.02	Zm, At
Os.41333.1.A1_at	LOC_Os01g42060.1	Expressed protein (similar to GAMETE EXPRESSED 3 (GEX3))	13.24	Zm, At
Os.53049.1.S1_at	LOC_Os04g46490.1	Aquaporin TIP5.1	14.61	Ll, Pz, Zm, At
Os.26448.1.A1_at	LOC_Os03g08070.1	Copper-transporting ATPase PAA1	10.59	Pz, Zm
Os.21018.1.S1_at	Os09g0525700	Generative cell specific-1; HAP2-GCS1 domain	14.00	
OsAffx.26224.1.S1_s_at	LOC_Os04g29090.1	FAD-binding and arabino-lactone oxidase protein	11.16	At
Os.18560.1.S1_at	LOC_Os05g01500.1	Tubulin-specific chaperone E	10.47	
OsAffx.3617.1.S1_at	LOC_Os03g55890.1	Ternary complex factor MIP1	10.34	
OsAffx.2553.1.S1_at	LOC_Os02g09580.1	OsFBX39 - F-box domain containing protein	10.26	Ll, Pz, Zm, At
Os.51974.1.S1_at	LOC_Os06g20860.1	Paramyosin	10.99	At
Os.50267.1.S1_at	LOC_Os08g34640.1	Receptor-like protein kinase precursor	13.12	Ll, Pz, Zm
Os.55267.1.S1_at	LOC_Os03g44630.1	Plastocyanin-like domain containing protein	13.75	Ll, Pz, Zm
Os.9559.1.S1_at	LOC_Os03g37570.1	Expressed protein	12.80	
OsAffx.2680.1.S1_at	LOC_Os02g19180.1	ZOS2-06 - C2H2 zinc finger protein	13.12	
Os.52171.1.S1_at	LOC_Os06g38950.1	ABC transporter, ATP-binding protein	14.10	
Os.32737.1.S1_at	LOC_Os11g08440.1	DnaK family protein	12.70	Ll, Pz, Zm, At
Os.38984.1.S1_at	LOC_Os01g23580.1	Inorganic H+ pyrophosphatase	12.65	Zm
Os.23286.1.A1_at	LOC_Os10g02920.1	Cytochrome b561	12.18	Zm, At
OsAffx.26489.1.S1_at	LOC_Os04g46760.1	Trehalose phosphatase	12.03	
Os.38283.1.S1_a_at	LOC_Os05g41550.1	Expressed protein	11.89	
Os.53437.1.S1_at	LOC_Os03g45980.1	Expressed protein	12.63	
Os.46544.1.A1_at	LOC_Os10g25060.1	Expressed protein	12.95	At
Os.9431.1.A1_a_at	LOC_Os08g16610.1	Rad21/Rec8-like protein	14.01	Zm, At
OsAffx.12136.1.S1_at	LOC_Os02g20530.1	Expressed protein	13.62	At
OsAffx.24724.1.S1_x_at	LOC_Os02g44600.1	Expressed protein	12.27	Pz, Zm, At
Os.4125.1.S1_at	UniGene Os.4125 Os12g0572800[2]	Similar to RSSG8 (RNA recognition/ankyrin motifs)	11.78	
Os.27370.1.S1_at	UniGene Os.54572 Os01g0353900[2]	Conserved hypothetical protein	12.51	
Os.54137.1.S1_at	LOC_Os02g08080.1	Expressed protein	12.84	Zm
OsAffx.31616.1.S1_at	LOC_Os12g06480.1	PHD-finger family protein	13.35	At
Os.52287.1.S1_at	LOC_Os02g02800.1	AGAP001222-PA protein (*Anopheles gambiae*-like methyl transferase)	13.02	
Os.54774.1.S1_at	Os07g0438300	Conserved hypothetical protein	10.68	
Os.52770.1.S1_at	Os03g0193300	Similar to nitrate transporter	12.70	
OsAffx.14496.1.S1_at	LOC_Os05g02030.1	OB-fold nucleic acid binding domain protein	10.18	
Os.18262.1.S1_at	Os11g/intergenic	-	11.81	
Os.50552.1.S1_at	LOC_Os08g35700.1	Leucine-rich repeat family protein	12.94	Pz, Zm, At
OsAffx.28262.1.S1_at	LOC_Os07g04520.1	Protein kinase	12.00	
Os.56612.1.A1_x_at	LOC_Os05g11980.1	Timeless protein	10.39	At
Os.34965.2.S1_s_at	LOC_Os06g07130.1	SHR5-receptor-like kinase	11.25	Zm, At
Os.54810.1.A1_at	Os08g0368000[2]	Coatomer delta subunit (Delta-coat protein)	12.54	Ll
Os.52821.1.S1_at	LOC_Os11g37200.1	Transmembrane BAX inhibitor motif-containing protein	13.62	Ll, Zm, At
Os.54486.1.S1_at	LOC_Os05g03320.1	Expressed protein	12.82	
Os.10491.1.S1_at	LOC_Os03g04690.1	Expressed protein	9.84	
Os.23535.1.A1_at	LOC_Os08g05820.1	Monocopper oxidase	14.07	

Homologies in *Lilium longiflorum* (Ll), *Plumbago zeylanica* (Pz), *Zea mays* (Zm) and *Arabidopsis thaliana* (At) are described in detail in Supporting Information Table S1.

[1]Five-numbered loci are those annotated at Michigan State University (MSU) ver. 7; seven-numbered loci are Rice Annotation Project (RAP) ver. 5.

[2]Also reported among Gou *et al.* (2001) rice sperm expressed sequence tags (ESTs).

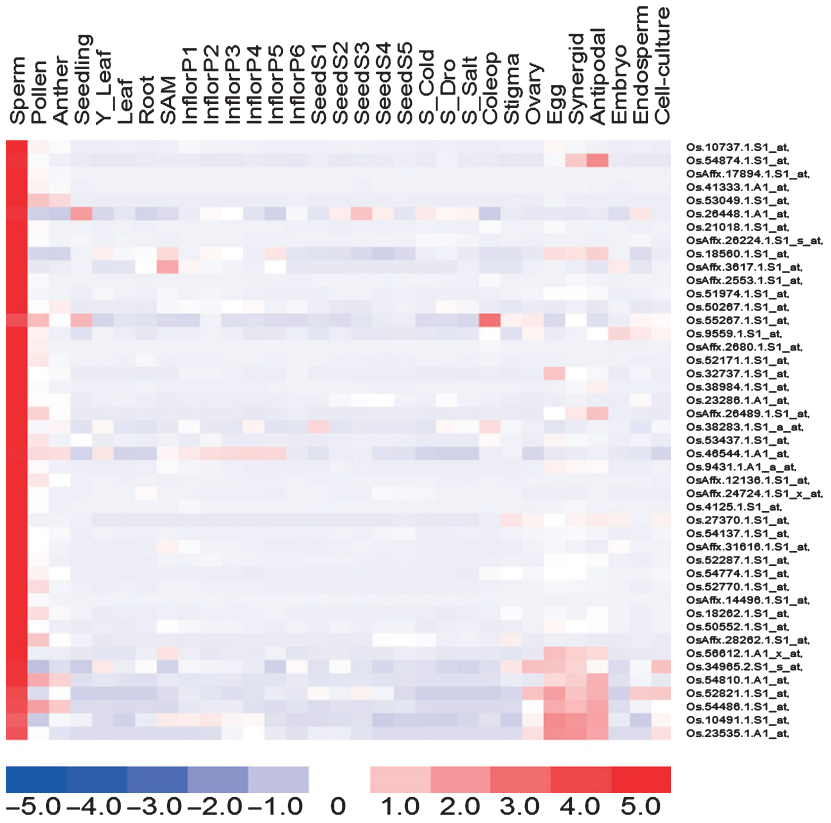

Fig. 3 Expression profiles of 45 highly up-regulated sperm genes in 31 different tissues. (See Table 1 and Supporting Information Table S1 for details). Y_, young; SAM, shoot apical meristem; S_, stress-related.

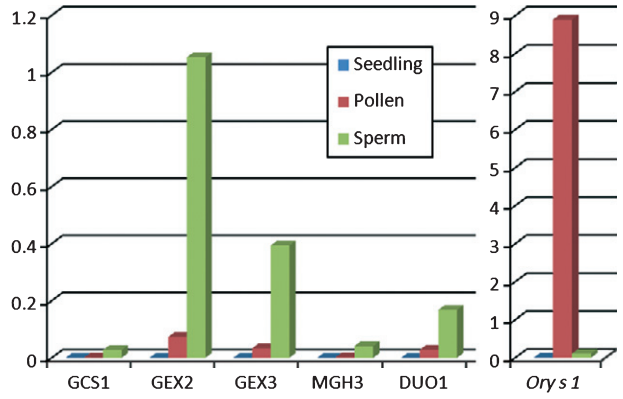

Fig. 4 qPCR of *Oryza sativa* sperm genes corresponding to GENERATIVE CELL SPECIFIC 1 (GCS1), GAMETE EXPRESSED 2 & 3 (GEX2, GEX3), MALE GERMLINE HISTONE H3 (MGH3) and R2R3 MYB transcription factor DUO1 POLLEN1 (DUO1), and pollen gene *Ory s 1*; all are scaled to *PROFILIN-2*.

between pollen and embryo sac lineages. Selected genes highly expressed in the microarray were examined using qRT-PCR (Fig. 4) and RT-PCR (Fig. 5). These results confirm microarray results for selected genes, including sperm markers GCS1, GAMETE EXPRESSED genes 1-3 (GEX1), GEX2 and GEX3 (Figs 1, 3–5). Additional highly transcribed sperm genes include sequences encoding aquaporin, F-box motif proteins, ubiquitin pathway-related proteins, DnaK Hsp70-related proteins, receptor-like kinases, cell signaling related proteins, apoptosis inhibitor BAX1 (BINDS ARCHAEL XPB helicase), and proteins involved in DNA repair pathways (Table 1); these have been found in multiple previous studies and putatively represent highly conserved functional themes of the male germline from both bicellular and tricellular pollen in monocot and dicot plants. The Arabidopsis transcriptome displayed sperm expression in homologs of 19 of the 45 transcripts shown in Table 1. Fourteen probe sets showed no close homolog, three homologs were called absent, whereas nine related loci were not represented on the Arabidopsis 24K genomic microarray chip (Table S1). Functional categories reported in sperm cell microarrays using the Arabidopsis and those annotated for the rice 57K Affymetrix genomic microarray chips reflected subtle differences in percentage representation, but overall gene numbers were much higher in *Oryza sativa* (Figs S3, S4), presumably largely because of historical expansion of the genome compared with that of Arabidopsis.

Divergent complements of expressed genes in sperm and other lineages

Principal components analysis (PCA) is a mathematically rigorous multivariate analysis method that reduces related variables to their component axes, allowing essentially *n*-dimensional data sets to be portrayed in typically three dimensions. Different rice tissues from GEO noncontrol probe sets depict the spatial

relationship of the first three principal components of normalized expression data (Fig. 6). According to these results, pollen vegetative cells and sperm cells define the two most divergent cell types, setting limits on three axes. The classical vegetative sporophyte tissues, by contrast, are aggregated distantly from pollen and sperm. In rice, sperm were clearly closer to the cluster of sporophytic cell types than they were to pollen. Female gametophytic cells clustered closely together near the sporophyte cluster, suggesting a much closer transcriptional relationship than displayed by the male germline (Fig. 6). Such differences in transcription presumably set the initial condition of the gametes at double fertilization and indicate significant distinctions between the male and female germ lineages.

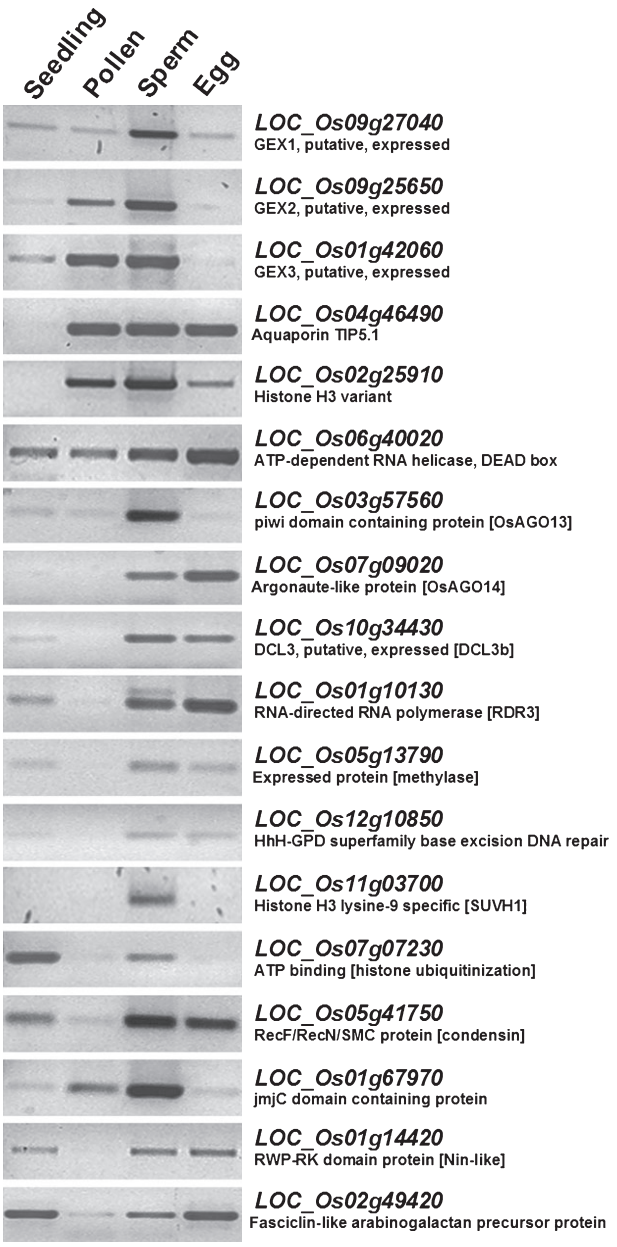

Fig. 5 RT-PCR analysis of selected transcripts from rice (*Oryza sativa*) tissues carried out under linear amplification conditions. The *PROFILIN-2* gene (see Fig. 1b) was used as an internal control.

Transcriptional profiles of sperm cells, pollen and seedlings suggest distinctly different patterns of gene expression

Signal intensities of sperm, pollen and seedling probe sets were compared to determine up- and down-regulated sequences that could be related to functional categories of genes (Table S2). The categories most markedly down-regulated between sperm cells and seedlings were those related to photosynthesis-associated pathways (e.g. redox-related genes, electron transport chain, nucleotide intermediates, and carbon-backbone synthetic pathways). By contrast, the most highly up-regulated sperm sequences compared with seedlings encoded proteins involved in ubiquitin pathways, DNA modification and repair, RNA transcription and regulation, modification of chromatin, protein degradation, signaling pathways, and a broad class of 'unknown proteins'. Down-regulated pathways in sperm cells compared with pollen included cell wall metabolism, transport, synthesis, degradation, electron transport, and secondary and primary metabolites, whereas the most enriched were RNA-related control, chromatin modeling, DNA repair and ubiquitin-mediated proteolysis (Table S2). Activation of the ubiquitin pathway is inferred by the abundance of transcriptional products in each of the essential components of the ubiquitin pathway. Gametophyte development depends on the activation of proteosome Regulatory Particle 5 (RPT5) (Gallois *et al.*, 2009), which is consistent with the importance of this pathway.

Among the most highly up-regulated sperm transcripts are those of genes that may facilitate or directly physically regulate sperm cell behavior (Tables 1, S1). One conspicuously abundant transcript encodes a Tonoplast Intrinsic Protein 5;1 (TIP5;1)-like aquaporin—which is the most abundant sperm-expressed transcript in rice (Table 1). Similar aquaporins are highly expressed in sperm of Arabidopsis (Borges *et al.*, 2008), *Zea mays* (Engel *et al.*, 2003) and *Plumbago zeylanica* S_{vn} (Gou *et al.*, 2009), suggesting a conserved and important role in sperm function. Such membrane-localized proteins as aquaporin, with the capacity to control cellular turgor, may regulate gametic membrane tensioning and contribute to sperm receptivity. Evidence of increased vacuolization of sperm cells during the final phases of pollen tube elongation *in vivo* has been reported in previous ultrastructure studies (Russell, 1992). Another highly transcribed gene, *LOC_Os05g18730*, is a presumed ortholog to GCS1 (Mori *et al.*, 2006) and HAP2—a membrane protein of Arabidopsis sperm cells required for gamete fusion and involved in pollen tube guidance (von Besser *et al.*, 2006). The HAP2 protein is also required for fusion in *Chlamydomonas*, indicating its highly conserved nature (Liu *et al.*, 2008). Signal intensity and fold changes in all sperm-responsive Affymetrix probe sets are shown in Table S3.

Sperm transcription factors

Seventy sperm-enhanced transcription factors (TFs) were found among detected probe sets in rice, including a number having no previous EST support, as may be expected given limited EST sampling of sperm cells in the past (Gou *et al.*, 1999). These

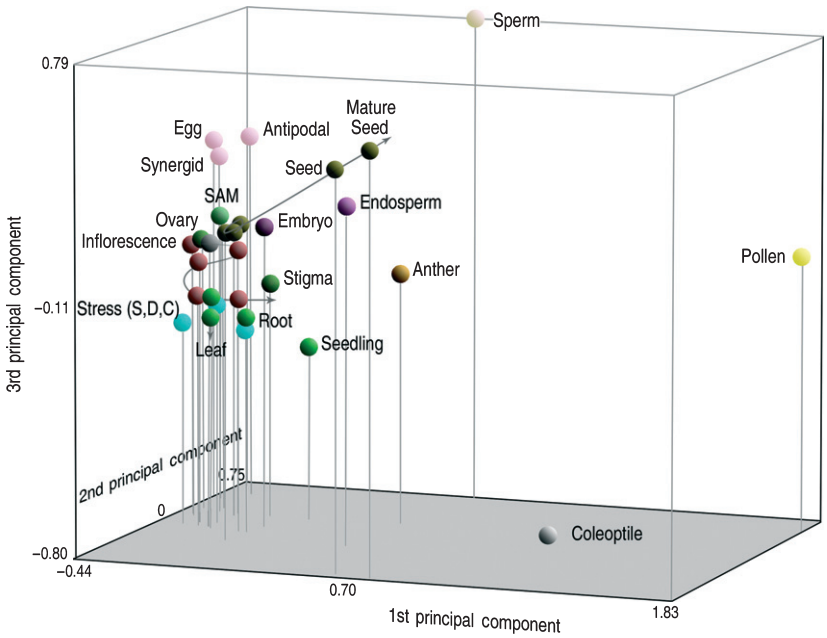

Fig. 6 Principal components analysis (PCA) of probe set signal intensity on the rice (*Oryza sativa*) GeneChip microarray reflecting gene expression patterns in 31 different tissues of rice. The relative distances of sperm and pollen profiles from those of other cell/tissue types (and each other) are indicative of particularly divergent patterns of overall gene expression in these two cell/tissue types.

sperm-enhanced TFs include Nin (Nodule Induction)-like and WRKY (WRKY motif zinc-finger-like) TFs (three of each), AP2/EREBP (APETALA2/ethylene-responsive element binding proteins), C2H2 (2 cysteine-2 histidine zinc-finger), CPP (cysteine-rich polycomb protein-like), MYB (myeloblastoma-like), PcG (Polycomb group), PHD (Plant HomeoDomain zinc finger) TFs (two each), and one each of bHLH (Helix-Loop-Helix), C3H (3 cysteine-1 histidine), ARR-B (type-B phospho-accepting response regulator), Homeobox, and NAC TFs. The transcription factor DUO1 (R2R3 MYB transcription factor DUO1 POLLEN1), which is not represented on the Affymetrix microarray, plays a key role in activating male germline genes of Arabidopsis (Borg *et al.*, 2011); in the current study, DUO1 is transcribed selectively in sperm cells (Fig. 4). An examination of upstream promoter regions of some of the most highly expressed sperm genes bear MYB binding motifs and are also enriched in other TF-binding domains consistent with TFs reported here as sperm selective (Sharma *et al.*, 2011). Some other highly sperm-enriched TFs are also implicated in modulating chromatin structure and controlling transcriptional repression of gene expression (Takeuchi *et al.*, 2006), notably including 14 of 20 Jumonji (Jmj) genes in this study, of which six are highly transcribed in sperm cells.

Interestingly, one TF transcribed in rice sperm cells appears to encode an AP2/EREBP TF homolog of BABY BOOM (*LOC_Os09g25600*)—which in Arabidopsis is associated with somatic embryogenesis (Boutilier *et al.*, 2002) and enhancement of plant regeneration (Srinivasan *et al.*, 2007). Although BABY BOOM has not been found in unfertilized egg cells of Arabidopsis (Curtis & Grossniklaus, 2008), it is highly transcribed in their sperm cells (Borges *et al.*, 2008).

Chromatin state and histone transcription in sperm cells

Modification of the chromatin state is an important requirement in the establishment of germ cells, in plants as well as in animals, and it is clear that the chromatin state of the gametes can precondition later patterns of imprinting and may strongly contribute to early expression (Luo *et al.*, 2011). Proteins contributing to chromatin state and chromatin-based gene activity include histones, DNA- and RNA-binding proteins, and enzymes that alter DNA and associated proteins through methylation and demethylation, acetylation and deacetylation, and control of their turnover, through synthesis and degradation (ubiquitination). These complex and interlinked pathways contribute to controlling expression on a local or regional genomic level and are highly conserved among eukaryotes (http://www.chromdb.org/).

Histone composition is a foundational element of chromatin state, and rice sperm cells display a unique combination of up-regulated transcripts in each major histone category (H1, H2A, H2B, H3 and H4), as is indicated among the 28 tissue types shown in Fig. 7. Among histone types, particular diversity is evident in histones H2B and especially H3. In Arabidopsis, a number of substitution H3 genes have been reported in sperm cells (Okada *et al.*, 2005) that are transcribed independently of replication and correspondingly lack an OCT promoter motif. Arabidopsis male-germline histone H3 (AtMGH3) is a sperm-specific histone H3 variant that has three rice histone H3 genes as close homologs in *Oryza sativa*, *HRT704*, *HRT11*, and *HRT12* (Chrom DB database, http://www.chromdb.org/), all of which are abundantly transcribed in the germline, as are the most highly transcribed H3 variants, *HRT707* and *HRT709* (Fig. 7). Each of these male germline-transcribed histones appears to

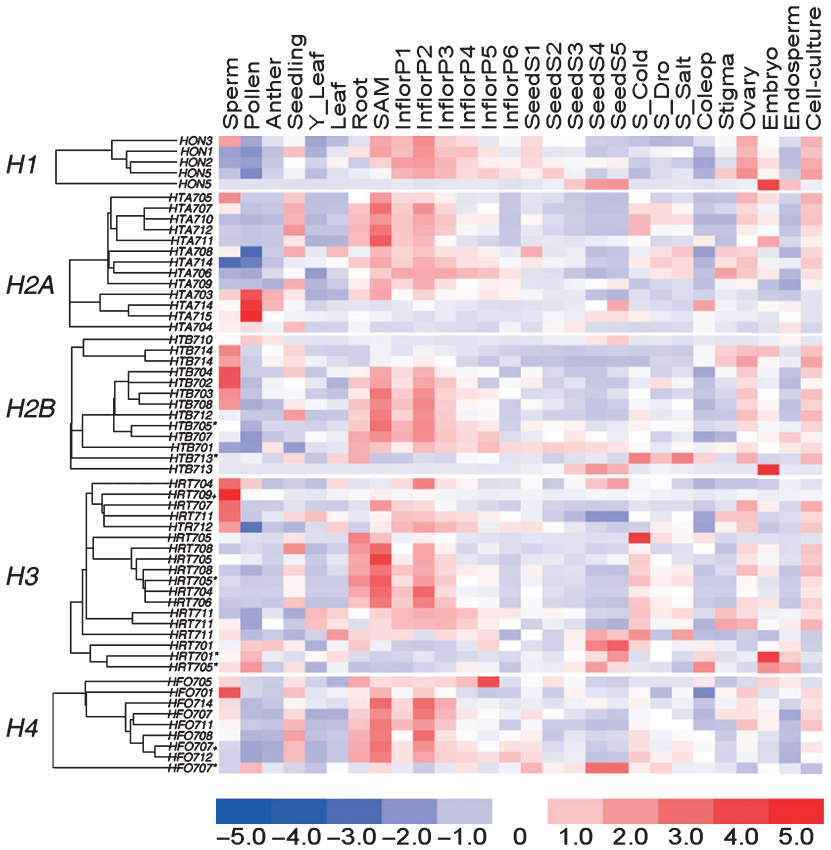

Fig. 7 Expression profiles of probe sets corresponding to histone genes represented on the rice (*Oryza sativa*) genome microarray chip and listed according to their ChromDB gene identifiers. Items marked with '*' have probe set sequences that overlap with more than one gene. Distinctive patterns of up- and down-regulation are particularly conspicuous in sperm and pollen. Y_, young; SAM, shoot apical meristem; S_, stress-related.

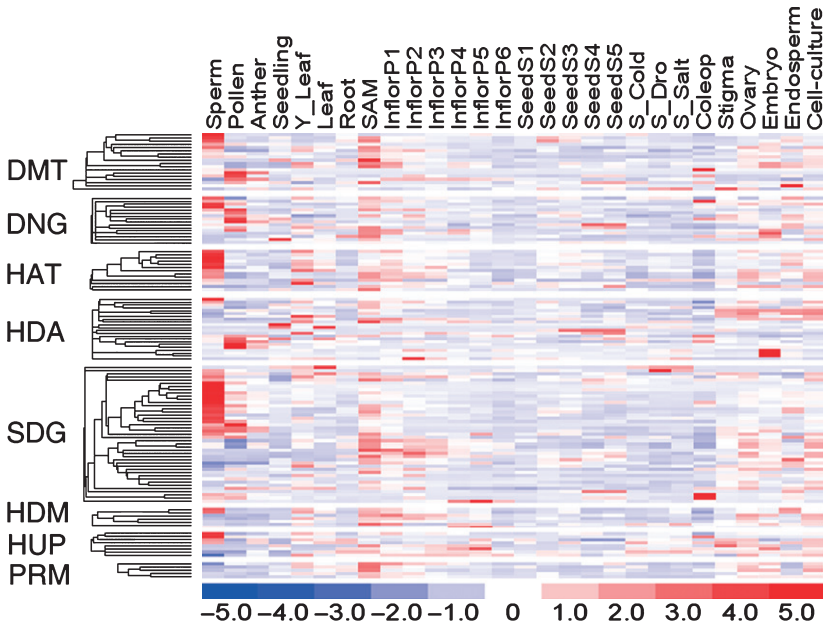

Fig. 8 Expression of probe sets corresponding to proteins involved in DNA and chromatin methylation, histone acetylation and histone ubiquitination in rice (*Oryza sativa*). DMT, DNA methyltransferase; DNG, DNA glycosylases; HAT, histone acetyltransferases; HDA, histone deacetylases; SDG, SET domain group proteins; HDM, histone demethylases; HUP, histone ubiquitination proteins; PRM, protein arginine methyltransferases. Y_, young; SAM, shoot apical meristem; S_, stress-related.

represent substitution histones, as they lack the OCT motif of histones transcribed during replication. Each of the histone subgroups has one or multiple sperm-enhanced representatives (Fig. 7) and each may be a target of further modification through chromatin modeling. Specialized histone proteins by themselves or in combination with other proteins may directly regulate transcription through changes in binding of chromatin to key gene regulatory elements.

Transcriptional activation of genes relating to chromatin state in sperm cells

Chromatin-modifying proteins known to modify DNA and DNA-associated proteins may in turn alter the activation, deactivation and longevity of DNA-associated proteins. Fig. 8 displays expression profiles for eight chromatin state modifiers. DNA methyltransferases (DMTs) mediate the addition of methyl groups at the DNA level (Pavlopoulou & Kossida, 2007), of which a unique group appears selectively up-regulated in sperm cells. Methylation, reversible by DNA glycosylases (DNGs), facilitates epigenetic reprogramming and DNA repair (Gehring *et al.*, 2009); a number of such genes are active in sperm cells and more types of DNG are active in pollen. Histone acetyltransferases (HATs) conjugate acetyl groups to histone proteins where these subgroups have the effect of loosening binding with DNA, thus facilitating transcription. By contrast, histone deacetylases (HDAs) tend to promote stronger bonding with DNA, thus inhibiting transcriptional activity. Sperm and pollen show distinct complements, as sperm cells reflect activation through enriched HAT transcript abundance, whereas the pollen reflects deactivation through enriched HDA transcript abundance.

Another group of histone-modifying proteins with complicated modes of controlling transcription are the SET domain group (SDG) proteins, which represent particularly conserved proteins that modify histone proteins through methylation of specific single or multiple lysine locations, with the consequence of activating or deactivating transcription. Expression profiles of SDG proteins involve H3K9 histone methylation, DNA-level CpG and CNG cytosine methylation, and other related chromatin modeling themes, indicating that SDG proteins may secondarily mediate a broad spectrum of protein- and DNA-level methylation (Ding *et al.*, 2007). Sperm show up-regulation of two histone demethylases (HMDs), which reverse the actions of SDG proteins, thus allowing a resetting of their chromatin state (Shinkai, 2007). Sperm also display up-regulated sequences encoding histone ubiquitination proteins (HUPs), which are believed to mediate DNA repair and accelerate histone turnover. Interestingly, protein arginine methyltransferases (PRMs), which are implicated in signal transduction, nuclear transport and transcription regulation controlling floral timing (Schmitz *et al.*, 2008), are down-regulated as a class in sperm cells.

RNAi pathway members in sperm cells

The expression of small RNAs acting through RNA-interference (RNAi) pathways is reflected in sperm cells through activation of an evolutionarily conserved family of proteins that include argonautes (AGO), dicer-like (DCL) and RNA-dependent RNA polymerases (RDRs).

Although early studies suggested transcriptional absence of RNAi pathways in mature pollen of Arabidopsis (Pina *et al.*, 2005), RNAi pathway genes appear to be present in sperm cells, including *AGO5*, *AGO6*, *AGO9*, and *DCL1* (Borges *et al.*, 2008), and have been more extensively characterized in pollen (Grant-Downton *et al.*, 2009). In rice sperm cells, transcripts encoding the PAZ (Piwi, Argonaut and Zwille) dsRNA recognition motif, characteristic of RNAi pathway genes, appear to be highly up-regulated compared with other tissues (Fig. S2). The seven argonautes represent each of the four major gene clades: *AGO1*, which functions in miRNA processing; *AGO4*, which cooperates with *DCL3* in chromatin silencing; *MEL1*, which functions in maintaining germ cell identity; and *ZIPPY*, which is implicated in heterochronic development (Nonomura *et al.*, 2007; Kapoor *et al.*, 2008). Most highly up-regulated in sperm cells are nontraditional *AGO* genes that encode a modified rather than the canonical *AGO1* catalytic site at the PIWI domain (Table 2). *DCL1* is implicated in biogenesis of miRNA, producing small RNAs from endogenous inverted repeats. *DCL3*, in turn, produces the 24nt siRNAs that mediate *de novo* DNA methylation, gene silencing and chromatin modification (Henderson *et al.*, 2006). Finding DCL3 transcripts absent from Arabidopsis sperm cells, Borges *et al.* (2008) suggested that novel small RNA pathways may be activated instead. In rice, however, DCL3 appears to be present in the sperm and egg, and thus siRNAs could be more conventionally involved in silencing transposable elements, a role that may involve pollen-produced elements being transported to sperm targets (Slotkin *et al.*, 2009). Sperm cells transcribe *RDR3*, which is one of four genes in rice that regulate and potentially amplify components of the RNAi machinery (Sijen *et al.*, 2001); these appear to be differentially enriched in the rice germline in general, as transcripts are present in both sperm and egg cells.

Differentially represented GO categories in sperm cell transcripts

The GeneBin analyses of *O. sativa* and Arabidopsis produced largely similar transcriptional profiles, with few exceptions with regard to percentage representation in categories (Fig. S3) but striking differences in the number of represented transcripts (Fig. S4). DAVID analysis found no categorical differences in the uniquely reported probe sets in sperm cells, as few have been functionally annotated. By contrast, significantly more probe sets were depleted in sperm (448) and in pollen (884) compared with sporophyte tissues. GO categories apparently depleted in sperm cells included metabolic cofactors, mitochondrial membrane proteins, redox-related pathways, auxin response pathways and protein binding motifs, whereas those depleted in the pollen vegetative cell included ribosome synthesis, hormone response, signal transduction-related histidine kinase pathways, tRNA acetylation processing, and helicase-related motifs (Table S4). In both cell types, these depletion motifs may reflect the short

Table 2 Differentially highly transcribed RNAi pathway genes in rice (*Oryza sativa*) sperm cells

Gene ID	Probe ID	Locus	Intensity (log_2)	Fold over seedling
Argonaute-related genes				
OsAGO1 group				
OsAGO1b	Os.8293.1.S1_at	LOC_Os09g27060	14.03	2.25
OsAGO17[1]	Os.54612.1.A1_at	LOC_Os02g07310	11.59	5.21
OsAGO4 group				
OsAGO16[2]	Os.19635.1.S1_at	LOC_Os07g16224	11.81	3.20
MEL1 group				
OsAGO12	Os.50897.1.S1_at	LOC_Os03g57560	12.71	3.66
OsAGO13[3]	OsAffx.25614.1.S1_at	LOC_Os03g47820	8.99	2.79
OsAGO14[2]	Os.53826.1.S1_at	LOC_Os07g09020	11.22	5.78
ZIPPY group				
OsAGO2[4]	Os.50333.1.S1_at	LOC_Os04g52540	13.83	3.14
OsAGO3[4]	Os.55120.1.S1_at	LOC_Os04g52550	8.17	3.05
Dicer-like-related genes				
DCL1 group				
DCL1a	Os.13488.1.S1_at	LOC_Os03g02970	12.06	2.71
DCL3 group				
DCL3b	OsAffx.8007.1.S1_at	LOC_Os10g34430	6.57	3.63
RNA-dependent RNase-related genes				
RDR3 group				
RDR3	Os.32212.1.A1_at	LOC_Os01g10130	7.96	3.48
RDR3	Os.32212.2.S1_x_at	LOC_Os01g10130	6.65	2.68

Microarray intensity was normalized using dChip. Motifs below represent divergences from the normal PIWI catalytic domain of D760, D845, and H986/H798 of AGO1. Gene IDs and motifs are according to Kapoor *et al.* (2008).
[1]HDR/C motif.
[2]DDH/P motif.
[3]-D-/H motif.
[4]DDD/H motif.

lifespan of the pollen tube and sperm cells such that metabolic needs can be met by surrounding cells.

Discussion

Compared with pollen, sperm cells contain numerous transcripts, consistent with their role as stem-cell-like founder cells. That such a distinct set of transcripts are present in sperm cells compared with pollen and other cells suggests that the regulatory role of the cells of the male germline is substantially autonomous from that of other tissues—this is all the more unexpected because the sperm cells are small, their volume diminishing progressively during maturation (Russell & Strout, 2005), and they are contained in a unique cell-within-a-cell relationship within the pollen. Sperm cells contain a large complement of transcripts that are apparently also expressed in other somatic tissues, presumably as consistently expressed transcripts with core metabolic functions. Although some transcripts in mature sperm cells are likely to persist from earlier developmental stages, others such as *DUO1*, *GCS1*, *GEX2*, *GEX3*, and *MGH3*, for example, have specific and potentially crucial roles in the biology of the male gamete.

The degree of transcriptomic complexity in sperm cells appears to exceed these diminutive cells' own metabolic requirements. In fact, some sperm transcripts appear not to be translated into protein in sperm cells, but may display delayed expression (Bayer *et al.*, 2009), although others are clearly transcribed and translated inside sperm cells (Ge *et al.*, 2011). Some sperm proteins could also persist from previous stages, but these may be very restricted given the highly up-regulated ubiquitin pathways that are a common theme in male germline studies (Singh *et al.*, 2008). Thus, both synthesis and degradation appear to have crucial roles in establishing the distinct transcriptional and expression profile of sperm cells, which reflects their own unique developmental niche.

Common and divergent themes in male germ lineage expression patterns

Among the most critically important conserved male germline genes are those that encode membrane proteins for fusion, such as GCS1, which apparently arose before the divergence of green plants (Liu *et al.*, 2008). In the next tier are common functional themes with evident homologs that meet the needs of regulating sperm expression, as for instance DUO1 (Borg *et al.*, 2011). By contrast, some expressed proteins appear to show relatively low conservation. For example, there are some highly transcribed sequences encoding proteins that have no obvious counterparts in other male germ lineages studied to date (Table S1), which suggests specialized functions and evolutionary divergence in the male germline potentially dating back to the divergence of monocots and dicots, some > 120 million years ago (Frohlich & Chase, 2007). Understanding both gene conservation and innovation in the context of germline evolution will require expanded

genomic studies (Paterson *et al.*, 2010). Rather than sperm transcripts merely serving their own specific metabolic needs, which are probably met largely by the pollen, this complex transcriptome appears to control its own unique expression pattern through extensive chromatin modeling both before and after fertilization (Grant-Downton & Dickinson, 2006; Ingouff *et al.*, 2007, 2010).

A significant proportion of the transcription of sperm cells appears to be related to programming the male germline determinants—a developmental event that begins with the asymmetric division of the microspore into pollen and generative cells (Eady *et al.*, 1995) and extends to sperm maturity (Twell, 2011). Sperm-expressed substitution histone H3 proteins and altered methylation state in the sperm nucleus represent examples of epigenetic chromatin modification (Okada *et al.*, 2005). The current study indicates that up-regulation of chromatin-modifying transcripts in the male germline reflects activation of multiple genes across nearly all classes of chromatin-modifying genes. Heterochromatin formation in the male germ lineage, a commonly known and historically described phenomenon (Maheshwari, 1950), appears to represent a most conspicuous self-regulating aspect of male germline epigenesis that has a significant molecular impact.

Cytoplasmic determinants unique to sperm cells may include abundant noncoding, small RNAs such as those involved in gene silencing (Slotkin *et al.*, 2009). Rice sperm cells are particularly enriched in transcripts encoding RNA-related processing proteins, and with highly up-regulated ubiquitin/proteosome pathways, are anticipated to result in dynamic shifts in proteome composition during maturation and the onset of receptivity. Numerous parallels are present between plants and animals in the establishment of their germlines (Dickinson & Grant-Downton, 2009).

In animals, germlines are established early in life and maintained by unique noncoding RNAs that are involved in germ identity through epigenetic marking and which play a crucial role in RNA silencing that prevents the expression of transpositional elements. In *Drosophila*, for example, piRNA and the protein Piwi are germline-specific subsets that are essential for spermatogenesis, and in mouse, similar orthologs, Miwi, Mili and their corresponding noncoding RNA, are also essential in the male germline (O'Donnell & Boeke, 2007). Regulation and suppression of transposable elements occur via up-regulation of an RNA-silencing mechanism as a frontline strategy for defending the germline genome from the uncontrolled insertion of transposable elements in plants, as well as animals. Pollen and sperm cells are known to contain 21nt miRNA sequences that are believed to be directed to the male germline (Slotkin *et al.*, 2009); small RNAs associated with AGO9 are specific for the female germline (Olmedo-Monfil *et al.*, 2010). Chromatin condensation driving the formation of heterochromatin in animals silences much of the male genome before fertilization and is augmented by polyamine binding which also inhibits transcription (Baulcombe, 2007). Although polyamines have not been found in nonmotile plant sperm cells, chromatin-silencing gene pathways are highly represented in rice sperm transcripts and could have a similar role in plants. Prevention of male germline transmission of viruses is critically required and is reflected in the diversity of highly up-regulated RNA silencing in animals (Ding & Voinnet, 2007). In rice, the up-regulation of a range of pathways for RNA silencing is noteworthy. Such defenses may aid in reducing the relatively rare occurrence of pollen-transmitted viral diseases (Mink, 2003).

Post-fertilization impact of paternally transcribed messages in development

An unexpectedly large complement of genes are enriched in sperm cells, compared with seedlings and the pollen vegetative cell; however, their role may not be evident before fertilization, nor are they necessarily translated, but they may influence post-fertilization development through transmission during plasmogamy. Recently, Bayer *et al.* (2009) demonstrated that, in Arabidopsis, *SHORT SUSPENSOR* (*SSP*) is an activator gene that is transcribed in the male germ lineage, transmitted into the egg cell and translated in the zygote, which initiates asymmetric division in the zygote. In *P. zeylanica*, which bears dimorphic sperm cells, similar transmission may occur; the dimorphic sperm cells of this plant are targeted to fuse specifically with either the egg or the central cell and display transcriptional profiles that appear to reflect the respective female cell with which the gamete will fuse (Gou *et al.*, 2009). For example, the sperm cell type that normally fuses with the central cell contains numerous copies of isopentenyl transferase, a control enzyme for cytokinin synthase, which drives endosperm development, whereas the sperm cell that fuses with the egg has an embryo-like profile (Russell *et al.*, 2010). Paternal transcripts have also been observed in tobacco zygotes using RT-PCR (Ning *et al.*, 2006) and are selectively persistent after fertilization (Xin *et al.*, 2011). Perhaps, as in animal systems, a complex repertoire of mRNAs may be delivered during fertilization (Ostermeier *et al.*, 2004; Krawetz, 2005). In plants, a similar failure to successfully perform *in vitro* fertilization using extracted sperm nuclei in maize (*Zea mays*) (Matthys-Rochon *et al.*, 1994) also suggests an essential role of the male cytoplasm in early post-fertilization development of plants.

Further insights into activated genes and pathways regulating flowering and male germline differentiation will advance not only our fundamental understanding of these reproductive cells but also cell–cell recognition, membrane fusion and fertilization and may aid in regulating these processes, which may be exploited in altering events from the earliest stages in seed development.

Acknowledgements

We thank Dr Yulin Jia (Dale Bumpers National Rice Research Center, Stuttgart, AR, USA) for providing field material and advice; Prof. Karen Moldenhauer (University of Arkansas Extension Station, Stuttgart, AR, USA) for seeds; and Cal Lemke (University of Oklahoma) for excellent technical assistance in growing greenhouse plants. We also thank Prof. Terry Speed and the software development staff of the Walter and Elisa Hall Institute of Medical Research (Melbourne, Australia) for advice on statistical data analysis, and Drs Peter Ades, Farzad Haerizadeh

and Harald Ottenhof (University of Melbourne) for additional help and encouragement. This work was supported by the Australian Research Council discovery grant DP 1097262 and US National Science Foundation award # IOS-1128145.

References

Baulcombe DC. 2007. Molecular biology: amplified silencing. *Science* **315**: 199–200.

Bayer M, Nawy T, Giglione C, Galli M, Meinnel T, Lukowitz W. 2009. Paternal control of embryonic patterning in *Arabidopsis thaliana*. *Science* **323**: 1485–1488.

Berger F. 2008. Double-fertilization, from myths to reality. *Sexual Plant Reproduction* **21**: 3–5.

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.

Borg M, Brownfield L, Khatab H, Sidorova A, Lingaya M, Twell D. 2011. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. *Plant Cell* **23**: 534–549.

Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijo JA, Becker JD. 2008. Comparative transcriptomics of *Arabidopsis thaliana* sperm cells. *Plant Physiology* **148**: 1168–1181.

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu CM, van Lammeren AA, Miki BL *et al.* 2002. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* **14**: 1737–1749.

Counce PA, Keisling TC, Mitchell AJ. 2000. A uniform, objective, and adaptive system for expressing rice development. *Crop Science* **40**: 436–443.

Curtis M, Grossniklaus U. 2008. Molecular control of autonomous embryo and endosperm development. *Sexual Plant Reproduction* **21**: 79–88.

Dickinson H, Grant-Downton R. 2009. Bridging the generation gap: flowering plant gametophytes and animal germlines reveal unexpected similarities. *Biological Reviews* **84**: 589–615.

Ding B, Zhu Y, Gao J, Yu Y, Cao K, Shen W-H, Dong A. 2007. Molecular characterization of three rice set-domain proteins. *Plant Science* **172**: 1072–1078.

Ding S-W, Voinnet O. 2007. Antiviral immunity directed by small RNAs. *Cell* **130**: 413–426.

Eady C, Lindsey K, Twell D. 1995. The significance of microspore division and division symmetry for vegetative cell-specific transcription and generative cell differentiation. *Plant Cell* **7**: 65–74.

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

Frank AC, Johnson MA. 2009. Expressing the diphtheria toxin A subunit from the *HAP2*(*GCS1*) promoter blocks sperm maturation and produces single sperm-like cells capable of fertilization. *Plant Physiology* **151**: 1390–1400.

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.

Frohlich MW, Chase MW. 2007. After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* **450**: 1184–1189.

Gallois J-L, Guyon-Debast A, Lecureuil A, Vezon D, Carpentier V, Bonhomme S, Guerche P. 2009. The *Arabidopsis* proteasome RPT5 subunits are essential for gametophyte development and show accession-dependent redundancy. *Plant Cell* **21**: 442–459.

Ge LL, Gou XP, Yuan T, Strout GW, Nakashima J, Blancaflor EB, Tian H, Russell SD. 2011. Migration of sperm cells during pollen tube elongation in *Arabidopsis thaliana* behavior during transport, maturation and upon dissociation of male germ unit associations. *Planta* **233**: 325–332.

Gehring M, Reik W, Henikoff S. 2009. DNA demethylation by DNA repair. *Trends in Genetics* **25**: 82–90.

Goffard N, Weiller G. 2007. GeneBins: a database for classifying gene expression data, with application to plant genome arrays. *BMC Bioinformatics* **8**: 87.

Gou XP, Wang SH, Chen F. 1999. Isolation and cytological observation of viable sperm cells of rice. *Acta Botanica Sinica* **41**: 669–674.

Gou XP, Xu Y, Tang L, Yan F, Chen F. 2001. Representative cDNA library from isolated rice sperm cells. *Acta Botanica Sinica* **43**: 1093–1096.

Gou XP, Yuan T, Wei XP, Russell SD. 2009. Gene expression in the dimorphic sperm cells of *Plumbago zeylanica*: transcript profiling, diversity, and relationship to cell type. *Plant Journal* **60**: 33–47.

Grant-Downton R, Le Trionnaire G, Schmid R, Rodriguez-Enriquez J, Hafidh S, Mehdi S, Twell D, Dickinson H. 2009. MicroRNA and tasiRNA diversity in mature pollen of *Arabidopsis thaliana*. *BMC Genomics* **10**: 643.

Grant-Downton RT, Dickinson HG. 2006. Epigenetics and its implications for plant biology 2. The 'epigenetic epiphany': epigenetics, evolution and beyond. *Annals of Botany* **97**: 11–27.

Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE. 2006. Dissecting *Arabidopsis thaliana* dicer function in small RNA processing, gene silencing and DNA methylation patterning. *Nature Genetics* **38**: 721–725.

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

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**: 44–57.

Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC *et al.* 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Research* **35**: W169–W175.

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

Ingouff M, Rademacher S, Holec S, Soljic L, Xin N, Readshaw A, Foo SH, Lahouze B, Sprunck S, Berger F. 2010. Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. *Current Biology* **20**: 2137–2143.

Itoh T, Tanaka T, Barrero RA, Yamasaki C, Fujii Y, Hilton PB, Antonio BA, Aono H, Apweiler R, Bruskiewich R *et al.* 2007. Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Research* **17**: 175–183.

Kapoor M, Arora R, Lama T, Nijhawan A, Khurana JP, Tyagi AK, Kapoor S. 2008. Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genomics* **9**: 451.

Krawetz SA. 2005. Paternal contribution: new insights and future challenges. *Nature Reviews Genetics* **6**: 633–642.

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 and Development* **22**: 1051–1068.

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A. 2011. A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genetics* **7**: e1002125.

Maheshwari P. 1950. *An Introduction to the Embryology of Angiosperms*. New York, NY, USA: McGraw-Hill.

Matthys-Rochon E, Mol R, Heizmann P, Dumas C. 1994. Isolation and microinjection of active sperm nuclei into egg cells and central cells of isolated maize embryo sacs. *Zygote* **2**: 152–157.

Mink GI. 2003. Pollen and seed-transmitted viruses and viroids. *Annual Review of Phytopathology* **31**: 375–402.

Mori T, Kuroiwa H, Higashiyama T, Kuroiwa T. 2006. GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nature Cell Biology* **8**: 64–71.

Ning J, Peng X-B, Qu L-H, Xin H-P, Yan T-T, Sun M. 2006. Differential gene expression in egg cells and zygotes suggests that the transcriptome is restructed before the first zygotic division in tobacco. *FEBS Letters* **580**: 1747–1752.

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.

O'Donnell KA, Boeke JD. 2007. Mighty piwis defend the germline against genome intruders. *Cell* **129**: 37–44.

Okada T, Bhalla PL, Singh MB. 2006. Expressed sequence tag analysis of *Lilium longiflorum* generative cells. *Plant Cell Physiology* **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 Journal* **44**: 557–568.

Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP. 2010. Control of female gamete formation by a small RNA pathway in *Arabidopsis*. *Nature* **464**: 628–632.

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

Paterson AH, Freeling M, Tang H, Wang X. 2010. Insights from the comparison of plant genome sequences. *Annual Review of Plant Biology* **61**: 349–372.

Pavlopoulou A, Kossida S. 2007. Plant cytosine-5 DNA methyltransferases: structure, function, and molecular evolution. *Genomics* **90**: 530–541.

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 Physiology* **138**: 744–756.

Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**: 62–66.

Russell SD. 1992. Double fertilization. *International Review of Cytology* **140**: 357–388.

Russell SD, Bhalla PL, Singh MB. 2008. Transcriptome-based examination of putative pollen allergens of rice (*Oryza sativa* ssp. *japonica*). *Molecular Plant* **1**: 751–759.

Russell SD, Dresselhaus T. 2008. Deciphering molecular mechanisms of fertilization in seed plants. *Sexual Plant Reproduction* **21**: 1.

Russell SD, Gou X, Wei X, Yuan T. 2010. Male gamete biology in flowering plants. *Biochemical Society Transactions* **38**: 598–603.

Russell SD, Strout GW. 2005. Microgametogenesis in *Plumbago zeylanica* (Plumbaginaceae). 2. Quantitative cell and organelle dynamics of the male reproductive cell lineage. *Sexual Plant Reproduction* **18**: 113–130.

Schmitz RJ, Sung S, Amasino RM. 2008. Histone arginine methylation is required for vernalization-induced epigenetic silencing of FLC in winter-annual *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **105**: 411–416.

Sharma N, Russell SD, Bhalla PL, Singh MB. 2011. Putative cis-regulatory elements in genes highly expressed in rice sperm cells. *BMC Research Notes* **4**: 319.

Shinkai Y. 2007. Regulation and function of H3K9 methylation. *Subcellular Biochemistry* **41**: 337–350.

Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S, Timmons L, Plasterk RHA, Fire A. 2001. On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**: 465–476.

Singh M, Bhalla P, Russell S. 2008. Molecular repertoire of flowering plant male germ cells. *Sexual Plant Reproduction* **21**: 27–36.

Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijó JA, Martienssen RA. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**: 461–472.

Srinivasan C, Liu Z, Heidmann I, Ence Darmo Jaya S, Fukuoka H, Joosen R, Lambalk J, Angenent G, Scorza R, Custers JBM *et al.* 2007. Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (*Nicotiana tabacum* L.). *Planta* **225**: 341–351.

Takeuchi T, Watanabe Y, Takano-Shimizu T, Kondo S. 2006. Roles of Jumonji and Jumonji family genes in chromatin regulation and development. *Developmental Dynamics* **235**: 2449–2459.

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 Journal* **37**: 914–939.

Tian HQ, Yuan T, Russell SD. 2005. Relationship between double fertilization and the cell cycle in male and female gametes of tobacco. *Sexual Plant Reproduction* **17**: 243–252.

Twell D. 2011. Male gametogenesis and germline specification in flowering plants. *Sexual Plant Reproduction* **24**: 149–160.

Wei LQ, Xu WY, Deng ZY, Su Z, Xue Y, Wang T. 2010. Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa*. *BMC Genomics* **11**: 338.

Xin HP, Peng XB, Ning J, Yan TT, Ma LG, Sun MX. 2011. Expressed sequence-tag analysis of tobacco sperm cells reveals a unique transcriptional profile and selective persistence of paternal transcripts after fertilization. *Sexual Plant Reproduction* **24**: 37–46.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Signal profiles of probe sets present in tissues of rice according to Affymetrix rice genome GeneChip and NCBI GEO data.

Fig. S2 Relative expression of probe sets encoding PAZ domain-containing proteins involved in RNA processing.

Fig. S3 Comparison of genes in each functional category by percentage according to GeneBins analysis of Arabidopsis (Borges *et al.*, 2008) and *Oryza sativa* sperm cells.

Fig. S4 Comparison of number of genes in each functional category according to GeneBins analysis of isolated Arabidopsis (Borges *et al.*, 2008) and *Oryza sativa* sperm cells.

Table S1 Highly up-regulated sperm transcripts, homologies, and occurrence in previous male germ lineage studies

Table S2 Comparison of functional category enrichment and depletion among the sperm cell, seedling and pollen

Table S3 Transcriptional profile of rice sperm, pollen and seedling (Excel spreadsheet)

Table S4 Depleted GO categories in sperm cells and pollen (Excel spreadsheet)

Table S5 List of primers used for RT-PCR and qPCR validations

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.