Planta (2006) 224: 1004–1014
DOI 10.1007/s00425-006-0283-3

ORIGINAL ARTICLE

Heping Yang · Navpreet Kaur · Stephanie Kiriakopolos
Sheila McCormick

EST generation and analyses towards identifying female gametophyte-specific genes in *Zea mays* L.

Received: 23 January 2006 / Accepted: 2 April 2006 / Published online: 23 May 2006

Abstract The embryo sac (female gametophyte) plays an important role in double fertilization. The female gametophyte is composed of four specific cell types: the synergids that attract pollen tubes, the egg cell and central cell which are fusion partners for the two sperm cells, and the antipodal cells whose function is unknown. As a resource for gene discovery and to help identify genes exhibiting cell-specific expression patterns, we constructed cDNA libraries from female gametophytes and from egg cells of maize and sequenced more than 8,500 ESTs. These libraries represent diverse transcripts, potentially corresponding to 3,850 genes (contigs and singletons) from the female gametophyte and 963 genes (contigs and singletons) from the egg cell. In each collection, 16% of the contigs/singletons have no matches in databases and 3–5% encode hypothetical proteins; novel hypothetical proteins (not found within the female gametophyte contigs) were identified among the egg cell contigs. We examined 65 contigs by RT-PCR and 19 genes that were potentially female gametophyte-specific were identified. We used in situ hybridization to determine expression specificity for seven genes: one transcript was expressed both in the egg cell and in the central cell, one was expressed in the egg cell and synergids, two were expressed in the central cell, two were expressed in the synergids, and one was expressed in the central cell and the synergids. Four of these encode small, potentially secreted peptides that are dissimilar except for a conserved triple cysteine motif near their C-terminus. These EST resources should prove useful for identifying female gametophyte or cell-specific genes.

Keywords Central cell · Double fertilization · Egg cell · Gamete · Synergid · Triple cysteine motif protein

Abbreviations Cdk: Cyclin-dependent kinase · DIG: Digoxigenin · EA: Egg apparatus · EAL: EA1-like · EBE: Embryo sac/basal endosperm · eIF-5A: Eukarytic translation initiation factor 5A · ES: Embryo sac · ESR1g1: Embryo surrounding region 1g1 · EST: Expressed sequence tag · GPI-Aps: Glycosylphosphatidyl inositol-anchored proteins · *NP1*: Nucellain precursor · TLA1: Transparent leaf area1

Introduction

The life cycle of higher plants alternates between the spore-producing sporophyte, a pre-dominant diploid generation, and gamete-producing gametophytes, a reduced haploid generation. The pollen grain (male gametophyte) is composed of a large vegetative cell and two sperm cells. The female gametophyte typically consists of seven cells: the egg cell, two synergid cells, a binucleate central cell and three antipodal cells. In angiosperms, double fertilization occurs. The pollen tube enters a synergid to deliver the two sperm cells to the female gametophyte. One sperm cell fuses with the egg cell to form the zygote that develops into the embryo, whereas the second sperm cell fuses with the central cell to yield the primary endosperm cell that eventually develops into the endosperm. Although double fertilization was discovered over 100 years ago, molecules that control this unique feature of angiosperms are largely unexplored because of the inaccessibility of plant gametes, which are embedded in sporophytic tissues (reviewed in Weterings and Russell 2004).

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00425-006-0283-3 and is accessible for authorized users.

H. Yang · N. Kaur · S. Kiriakopolos · S. McCormick (✉)
Plant Gene Expression Center, United States Department of Agriculture, Agricultural Research Service, and Department of Plant and Microbial Biology, University of California at Berkeley, 800 Buchanan Street, Albany, CA 94710, USA
e-mail: sheilamc@nature.berkeley.edu

Present address: H. Yang
Monsanto Company, 800 N. Lindbergh Blvd., St. Louis, MO 63167, USA

Successful fertilization in angiosperms depends on communication between the male gametophyte and female floral organs. Sperm cells are within the vegetative cell of the pollen grain and, unlike animal sperm cells, are not motile but are transported to the female gametophyte by the pollen tube. The pollen tube responds to multiple signals on its journey to the female gametophyte (reviewed in Swanson et al. 2004) For example, genetic evidence indicated the existence of a female gametophyte-derived, long-range activity controlling pollen tube attraction (Hülskamp et al. 1995; Ray et al. 1997). Analyses of the *Arabidopsis* mutants *magatama* (Shimizu and Okada 2000), *gfa2* (Christensen et al. 2002), *feronia* (Huck et al. 2003) and *sirene* (Rotman et al. 2003) suggested that signals from the female gametophyte control pollen tube attraction, penetration, reception and subsequent sperm cell delivery. Disruption of these signals resulted in sterility although pollen tubes could still arrive at the female gametophyte (Shimizu and Okada 2000; Huck et al. 2003; Rotman et al. 2003). Furthermore, cell ablation experiments demonstrated that the synergids are the source of attraction signal(s) whose molecular nature is still unknown (Higashiyama et al. 2001). In maize, a small protein present in the synergids and egg cell, ZmEA1, is thought to play a role in pollen tube attraction (Márton et al. 2005).

Large-scale single-pass sequencing of randomly selected cDNA clones is useful for discovering novel gene functions, identifying novel gene family members, and defining gene expression profiles. Comparison of ESTs from different tissues within an organism helps to identify tissue-specific genes. Some EST resources exist for the female gametophyte. Young ear cDNA libraries were sequenced by the maize EST project (http://www.mutransposon.org/project/RescueMu/zmdb/est/library.php), but mRNAs from the female gametophyte are likely not well-represented in such libraries, because most of a young ear is composed of sporophytic tissues. A cDNA library was constructed from isolated egg cells of maize (Dresselhaus et al. 1994). Transcripts from that library include those encoding calreticulin (Dresselhaus et al. 1996), ribosomal proteins (Dresselhaus et al. 1999a), eukaryotic translation initiation factor 5A (eIF-5A) (Dresselhaus et al. 1999b), defensin-like proteins (Cordts et al. 2001), egg apparatus 1 (EA1) (Márton et al. 2005), and TRANSPARENT LEAF AREA1 (TLA1; Dresselhaus et al. 2005). Recently, the same group reported (Sprunck et al. 2005) 735 expressed sequence tags (ESTs) from a wheat egg cell cDNA library, representing 404 genes, most of which were metabolism related. Okamoto et al. (2004) identified eight abundant proteins in the maize egg cell, and Okamoto et al. (2005) identified six ESTs that were expressed in the egg cell. Le et al. (2005) prepared cDNA libraries from maize egg cells and central cells and performed a suppression-subtractive hybridization, followed by microarray screening of ~~1,000 cDNAs, to identify clones with enhanced expression in one of these cell types. They identified and report on ~~30 ESTs from each library. They used in situ hybridization for five of these, to determine cellular expression patterns, and identified one apparently egg-specific EST and one apparently central cell-specific EST. The synergids are responsible for pollen tube attraction but no synergid cDNA library exists. A function for the antipodal cells is unknown—in many species they degenerate in the mature female gametophyte, but in maize they proliferate—no antipodal cell cDNA library exists. Yu et al. (2005) recently identified female gametophyte transcripts in *Arabidopsis* by comparative transcriptomics between ovules containing female gametophytes and ovules lacking them, due to a mutation at the *sporocyteless/nozzle* gene.

We isolated female gametophytes from maize and constructed a cDNA library, so that genes expressed in any cell type of the maize female gametophyte would be represented. Single-pass sequencing of 7165 ESTS followed by bioinformatic analyses indicated that this female gametophyte library is useful for gene discovery; the contigs represent 3,850 genes and 183 correspond to genes annotated as hypothetical proteins. To enrich for egg cell-specific transcripts and to determine the egg cell transcriptome, we also constructed a cDNA library from isolated egg cells. Analysis of 1,415 egg cell ESTs revealed that egg cells have diverse transcripts. Notably, many ESTs from the egg cell library were not identified within the female gametophyte ESTs, demonstrating that enrichment for a single cell type is effective at identifying novel ESTs. We performed RT-PCR on 65 selected ESTs from the female gametophyte library and identified 19 putatively female gametophyte-specific genes. Five of these encode a novel family of small cysteine-rich proteins with a triple cysteine motif at or near the C-terminus; these have different cellular expression patterns within the embryo sac. By significantly increasing publically available ESTs from the female gametophyte and egg cell, we have provided the research community with a resource for further studies on gametophyte development and double fertilization.

Materials and methods

Isolation of female gametophytes and egg cells

Zea mays (inbred line A188) plants were grown in a greenhouse for a 16 h day, 25/19°C (day/night) and 70% relative humidity. An irradiance of 50–120 W m^{-2} was provided by 1,000 W lamps. The first ears appearing on 8-week-old plants were covered with paper bags before silk emergence and female gametophytes were isolated from these unpollinated ears that had been covered for 2 weeks.

Female gametophytes were isolated by enzymatic maceration of ovule slices followed by manual microdissection, essentially as described in Kranz et al. (1991). Briefly, unfertilized ears (on average, 15 cm in length with 10 cm long silks, including both visible and those

covered by husks) were harvested, the husks were surface-sterilized with ethanol (70%), then husk leaves were removed and ovaries collected from the middle of the ears. Ovaries were longitudinally cut, while observing under a dissecting scope, to obtain pieces containing female gametophytes. The pieces were kept in 0.53 M mannitol solution, pH 5.0 before enzymatic maceration. Approximately, 20 ovule pieces were incubated (25°C, ~~1 h) in a 3.5-cm diameter plastic dish with 1.5 ml of enzyme mixture containing 0.75% pectinase (Sigma, St Louis, MO), 0.25% pectolyase Y23 (Seishin, Tokyo, Japan), 0.5% hemicellulase (Sigma), and 0.5% cellulase Onozuka RS (Yakult Honsha, Tokyo, Japan). Female gametophytes were isolated with microdissection needles. To isolate egg cells, the female gametophyte wall was cut with microdissection needles. Female gametophytes and egg cells were collected into 1.5 ml Eppendorf tubes with disposable transfer pipettes (1.5 × 30 mm) (BIO-RAD, Richmond, CA) and rinsed with 0.53 M mannitol solution three times. The extra mannitol solution was removed after centrifugation at 3,000 rpm for 1 min. Isolated female gametophytes and egg cells were frozen in liquid nitrogen and stored at −80°C.

RNA extraction and library construction

Total RNA was extracted from 300 embryo sacs and 270 egg cells using 0.5 ml TRIzol Reagent (GIBCO BRL, Gaithersburg, MD) and quantified with Ribogreen (Molecular Probes, Eugene, OR). Total RNA was used to synthesize cDNA; cDNA libraries of female gametophytes or egg cells were constructed by oligo dT priming using a SMART PCR cDNA synthesis kit (Clontech, Palo Alto, CA). The resulting PCR-amplified cDNA was digested with *SfiI* and ligated to *SfiI*-cut lambda arms; the cDNA was oriented in the vector. The recombinants were packaged with Gigapack III Gold packaging reagents (Stratagene, La Jolla, CA). The primary cDNA libraries were used as the EST source.

Plasmid isolation and sequencing

Individual phage clones were converted into plasmids in *Escherichia coli* (strain BM25.8) using the Clontech Cre loxP system. Plasmid DNA was extracted from BM25.8 cells using DirectPrep 96 MiniPrep Kits (QIAGEN, Valencia, CA). cDNA inserts of 300 bp or longer were amplified from the 5′ end using the Clontech 5′ end sequencing primer that flanks the cDNA inserts and a BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster, CA). PCR products were sequenced with an ABI PRISM™ 3100 Genetic Analyzer (Applied Biosystems). If there were more than three Ns in a row, the sequence thereafter was removed to control sequence quality. The female gametophyte EST sequences are in GenBank; accession numbers are listed in Table S1. Sequencing is ongoing: updated information can be viewed at the McCormick lab web site, http://www.pgec.usda.gov/McCormick/mclab.html, or at NCBI (http://www.ncbi.nlm.nih.gov) using the query "*Zea mays* embryo sac" and "*Zea mays* egg cell".

Bioinformatics

Ambiguous, linker, and vector sequences were removed manually prior to bioinformatic analysis. Sequences were assembled with Sequencer™ 3.1.2 using the default settings. For annotation, manually processed sequences were compared against current public databases, first using Blastx (http://www.ncbi.nlm.nih.gov). If a sequence was too short to obtain matches in the protein databases, we used nucleotide-nucleotide blast (Blastn) to attempt to obtain a longer EST sequence and then re-ran Blastx with the longer query. All parameters used default values. The ESTs were manually annotated using the information returned by Blastx. A transcript was assigned a match to a protein of known identity if the probability of similarity score in Blastx was at the threshold *e* value of 10^{-8} or smaller. The match with the lowest score was chosen when an EST sequence matched more than one gene, as in the case of gene families. Amino acid sequences of the triple cysteine motif family were compared and an alignment was generated using Multalin (Corpet 1988). Contig assembly was carried out with Sequencer™ 3.1.2 using the default settings. Contigs were assigned functional categories based on Gene Ontology (Camon et al. 2003), after sequence comparisons of their predicted gene products with the current SWISS-PROT/TrEMBL (SWALL) databases.

Reverse transcription (RT)-PCR analysis

To test the purity of the female gametophyte RNA, gene-specific primers (Table S3) were used to amplify transcripts of *embryo sac* (*ES*) *1* and *ES2/3/4* (positive controls), *Nucellain precursor−1* (*NP1*) (Linnestad et al. 1998) and *embryo surrounding region 1g1* (*ESR1g1*) (negative controls). The *NP1* primers amplified the predicted size product from maize ovule slices (not shown). For analysis of multiple tissue expression patterns, total RNA was isolated from ovule slices with female gametophytes, ovule slices with nucellar cells only, silks, leaves, roots and pollen of 8-week-old maize plants. The RNA was treated with DNase I and then individually used to synthesize first strand cDNA using the Moloney Murine Leukemia Virus (MMLV) reverse transcriptase supplied in the First-strand cDNA synthesis kit (Invitrogen, Carlsbad, CA). First strand cDNA (50 ng) from each tissue was amplified with transcript-specific primers. PCR was run for 30 cycles; denaturation at 95°C for 30 s, primer annealing at 55°C for 30 s, and polymerization at 72°C for 3 min, in a Thermal Controller (MJ Research, Watertown, MA). RT-PCR products were separated by electrophoresis on a 1% agarose gel in 0.5× BE buffer.

In situ hybridization

Female gametophyte-specific cDNA inserts were cloned into pGEM 3Zf vector (Promega, Madison WI). Linear-

ized DNA templates (2 μg) were used to synthesize sense and antisense probes with SP6 or T7 RNA polymerase from the Digoxigenin (DIG) RNA labeling kit (Roche Applied Science, Indianapolis, IN). RNA probes longer than 500 bp were hydrolyzed to ~~500 bp fragments in fresh 0.1 M carbonate buffer (pH 10.2). DIG-labeled RNA probes were resuspended in 50% formamide and stored at −20°C.

Ovule pieces containing female gametophytes were fixed in 4% paraformaldehyde in 1× PBS at 4°C for 36 h, and then dehydrated at 4°C through a graded ethanol: H_2O series. The tissues were then placed in 75% EtOH and 25% Histoclear (National Diagnostics, Atlanta, GA) for 1.5 h, 50% EtOH and 50% Histoclear for 1.5 h, 25% EtOH and 75% Histoclear for 1.5 h, and 100% Histoclear for 3 × 1.5 h, all at room temperature. After Histoclear was gradually replaced with molten paraplast at 50°C, tissues were transferred into molds and stored at 4°C or used directly. Paraffin-embedded tissues were sectioned on a microtome to 9 μm thickness. Pretreatment, hybridization and washing (at 50°C), and detection were performed as described (Torres et al. 1995), except that the antibody (anti-digoxigenin-alkaline phosphatase fab fragments) was purchased from Roche Applied Science. Images were captured from a Zeiss Axiophot microscope (Thornwood, NY) with a SPOT camera (Diagnostic Instruments, Sterling Heights, MI).

Results

Quality of female gametophyte and egg cell cDNA libraries

We collected unfertilized maize ears and isolated 300 female gametophytes (Fig. S1). Dynamic cytoplasmic streaming in all the cells was visible, indicating that the isolated female gametophytes were viable. Each isolated female gametophyte still had a few nucellar cells attached; nucellar cells are from the sporophyte and RNA from them was therefore not wanted. Because it was too tedious to manually remove these attached nucellar cells, we tested the purity of the RNA from the isolated female gametophytes to determine if RNA from the few remaining nucellar cells could be detected. We performed RT-PCR with primers specific for the female gametophyte-expressed genes *ES1* and *ES2/3/4* (Cordts et al. 2001), a nucellar cell-specific gene, *NP1* (Linnestad et al. 1998) and an endosperm-specific gene *ESR1g1* (Opsahl-Ferstad et al. 1997). Figure 1 shows that only transcripts from the *ES1* and *ES2/3/4* genes were detected. Although we cannot completely exclude the presence of transcripts from nucellar-expressed genes, we concluded that the female gametophyte RNA was pure enough for cDNA library construction. We then synthesized cDNA using total RNA from female gametophytes and constructed a library using a PCR-based method. From one ligation and packaging we obtained about 225,000 pfu. To assess the quality of this library, we tested 116 randomly selected plaques: about 90% had cDNA inserts; insert sizes ranged from 0.3 to 2.9 kb, with the majority falling between 0.5 and 0.9 kb.

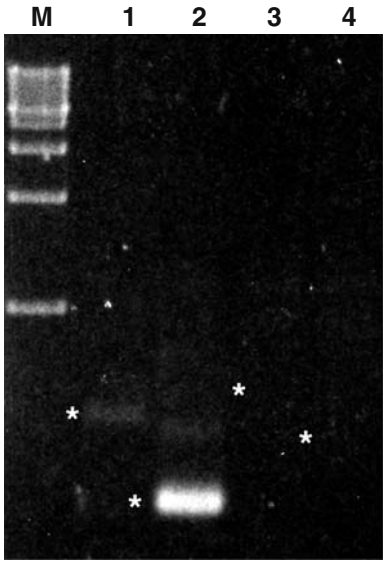

Fig. 1 Purity of isolated female gametophytes. Total RNA from isolated female gametophytes was used to synthesize cDNA. RT-PCR was performed with primers specific for *ES1*(1), *ES2/3/4* (2), *NP1* (3), and *ERS1g1*(4). The expected sizes are indicated with *asterisks*. Molecular weight marker is λ/*Hin*dIII (M)

Egg cells were isolated mechanically from isolated female gametophytes without using additional cell wall-degrading enzymes, to minimize stress responses (Fig. S2). We collected 270 egg cells and constructed a cDNA library from total RNA. From one ligation and packaging we obtained 325,000 pfu. We randomly selected 192 clones to assess the quality of this library: about 95% had cDNA inserts; insert sizes ranged from 0.3 to 3.2 kb, with the majority falling between 0.5 and 1.3 kb.

Diversity of transcripts

EST clones from the unamplified libraries were randomly picked and those with inserts larger than 300 bp were sequenced from their 5′ ends in a single run. EST sequencing from these libraries is ongoing; for updated information go to http://www.pgec.usda.gov/McCormick/mclab.html, or search Genbank (http://www.ncbi.nlm.nih.gov) with the query "*Zea mays* embryo sac" or "*Zea mays* egg cell".

About 30% of 7,165 female gametophyte ESTs were full length, or judged to be so based on size of insert and/or alignment with the closest match in the databases. A majority of the EST clones encoded full open reading frames (ORFs) or C-terminal coding regions with 3′ UTRs; less than 10% were 3′ UTRs of known or predicted genes. In order to assess the level of redundancy and to estimate the number of unique genes potentially represented in the ESTs, we assembled them into contigs.

Table S1 lists the contigs assembled from 7,165 ESTs and includes accession numbers, functional category and best BLASTP match. About 40% of the ESTs were singletons. The rest assembled into contigs of 2–188 clones. Figure 2 shows the 20 largest contigs from the female gametophyte. Each contig is composed of more than 20 overlapping or redundant ESTs. The majority of these clusters correspond to housekeeping genes. Overall, the current contigs and singletons correspond to potentially 3850 unique genes, indicating that the female gametophyte has diverse transcript populations.

The contigs were grouped according to functional categories as summarized in Fig. 3; over 70% encode known proteins and their functional categories are typical of the representations seen in other cDNA libraries or transcriptomes from various tissues (Hennig et al. 2004; Sprunck et al. 2005). Metabolism-related genes are the most frequently represented while genes involved in cell growth/division were least abundant. About 12% of the contigs were classified as "unknown protein" or "expressed protein" by comparison to sequences from rice, *Arabidopsis* or other species. The best matches for 4.7% of the contigs are to genes that are annotated as encoding hypothetical proteins in the *Arabidopsis*, rice or other genomes. Those with significant matches to *Arabidopsis* hypothetical proteins are indicated in Table S1; some others were similar to hypothetical proteins from rice, including 13 cysteine-rich proteins analyzed further below. Of the 989 ESTs with no significant matches, 386 assembled into 107 contigs while the others were singletons. Most of these contain an ORF; some of the rest might correspond to noncoding RNAs. When combined, these three categories (encoding unknown, hypothetical or novel proteins) constitute 30% of the contigs.

Diverse transcripts were also represented in the egg cell cDNA library (Table S2). Of the 1,415 ESTs, 806 were singletons and the other 609 ESTs assembled into 156 contigs, thus potentially representing 962 genes. Many of the most abundant contigs show no significant similarity to the databases, or correspond to unknown proteins (Fig. 2b). Overall, the functional distributions of the egg cell ESTs were similar to those of the female gametophyte (Fig. 3). Within the no significant similarity, hypothetical protein and unknown protein groups, there was limited overlap between the female gametophyte and egg contigs.

Identification of genes potentially specific to the female gametophyte

We reasoned that cDNAs that were female gametophyte specific might be absent or underrepresented in current EST databases. Because there are extensive ESTs from libraries made from diverse tissues of maize (http://www.tigr.org/tigr-scripts/tgi/libtc.pl?db=maizest) and most of our ESTs fell into a contig with these already sequenced ESTs, we did not randomly select ESTs and perform PCR to determine their expression patterns.

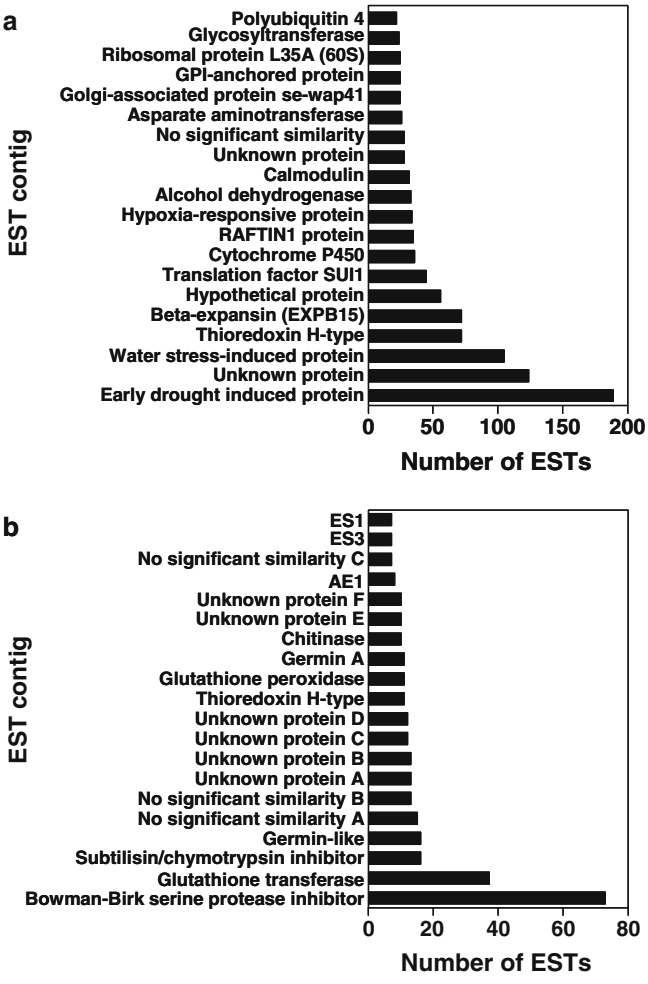

Fig. 2 Summary of the 20 largest contigs (> 20 ESTs) assembled from female gametophyte ESTs (**a**) or egg cell ESTs (**b**). The Genbank accession number listed after each name is of the best match in the database. Polyubiquitin 4 = S28426; Glycosyltransferase = XP_475798.1; Ribosomal protein L35A (60S) = AAL59231.1; GPI-anchored protein = CAE05527.1; Golgi-associated protein se-wap41 = T04331; Asparate aminotransferase = NP_916216.1; No significant similarity = ES12381 (DN828382); Unknown protein = XP_468711.1; Calmodulin = CAA46150.1; Alcohol dehydrogenase = AAA33434.1; Hypoxia-responsive protein = NP_198128.1; RAFTIN1 protein = XP_465009.1; Cytochrome P450 = NP_922325.1; Translation factor SUI1 = P56330; Hypothetical protein = Q8LHP0; Beta-expansin (EXPB15) = AAM73779.1; Thioredoxin, H-type = AAL67139.1; Water stress-induced protein = T07613; Unknown protein = AAF98579.1; Early drought-induced protein = AAM46894.1; ES1 = AAK08132.1; ES3 = AAK08134.1; No significant similarity C (triple cysteine motif containing) = E1872 (DT535575); AE1 = CAB56552.1; Unknown protein F = XP_479340.1; Unknown protein E = XP_479340.1; Chitinase = T03405; Germin A = XP_480464.1; Glutathione peroxidase = NP_192897.2; Thioredoxin H-type = T04090; Unknown protein D = CAE04145.1; Unknown protein C = NP_568713.1; Unknown protein B = XP_473116.1; Unknown protein A = NP_568088.1; No significant similarity B = E0002(DR451926); No significant similarity A = E0017 (DR451941); Germin-like = XP_480451.1; Subtilisin/chymotrypsin inhibitor = S61830; Glutathione transferase = NP_909709.1; Bowman–Birk serine protease inhibitor = NP_910046.1

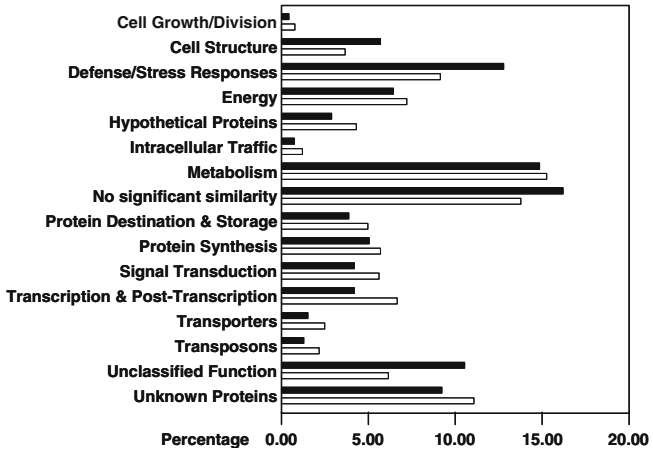

Fig. 3 Functional classification of maize egg cell (*open bars*) and female gametophyte (*filled bars*) contigs. The contigs listed in Tables S1 and S2 were grouped into different categories based on their biological functions

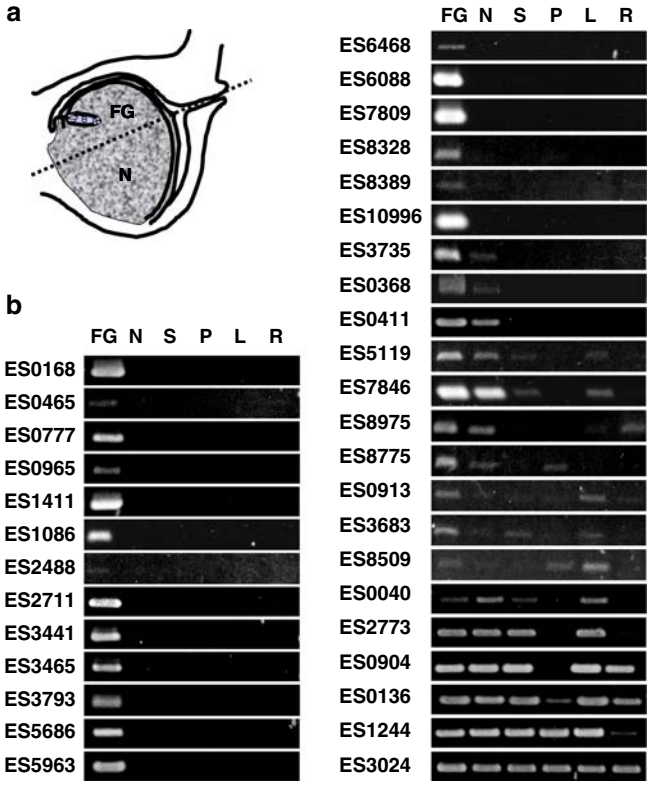

Fig. 4 Tissue-specific expression of selected genes. Total RNA from ovule slices with female gametophytes (*ES*), ovule slices lacking female gametophytes (*N*), silks (*S*), Anther (*A*), leaves (*L*) and roots (*R*) was used to synthesize cDNA. PCR was carried out and the products were separated by electrophoresis. For genes expressed in several tissues, only one representative from each expression pattern is shown. **a** A diagram showing how the cuts were made. **b** RT-PCR products

Indeed, when we checked expression patterns for ES0916 (eIF-5A) and ES0904 (blue copper-binding protein), we found that our RT-PCR results closely paralleled EST database representation. Instead we selected candidates for RT-PCR from those ESTs that had no database matches, or that encoded hypothetical proteins, or that encoded unknown proteins whose EST matches in public EST databases were only from reproductive tissues (flowers, ovaries, ovules, embryos and endosperm). Using primers specific for the *ES1* and *ES2/3/4* genes, we first confirmed that these female gametophyte-specific messages (Cordts et al. 2001) could be detected in cDNAs derived from ovule slices with female gametophytes, and were not detected in cDNAs derived from ovule slices lacking female gametophytes (data not shown).

Of 65 genes tested by RT-PCR, 19 appeared to be female gametophyte specific (Fig. 4), although genes expressed in nucellar cells surrounding the female gametophytes might also be represented in this group. Ten of these ESTs (ES0465, ES1086, ES2711, ES3441, ES3465, ES3793, ES5686, ES8328, ES8389, and ES10996) did not have any match in the current databases, five (ES1411, ES5963, ES6088, ES6468 and ES7809) encode hypothetical proteins, and two (ES0965 and ES2488) encode unknown proteins. The last two (ES0777 and ES0168) were similar to previously described proteins. ES0777 shares similarity with the maize ae1 protein. The *ae1* gene was expressed at a very early stage of androgenic embryogenesis in vitro and in early endosperm in vivo (Magnard et al. 2000); our results show that a similar protein is also expressed in the unfertilized female gametophyte. ES0168 encodes a protein similar to embryo sac/basal endosperm (EBE)-2 (CAD24798.1). *ZmEBE-2* is expressed in the female gametophyte and in the basal endosperm transfer layer of maize (Magnard et al. 2003).

Nine of the 19 putatively female gametophyte-specific ESTs (ES0777, ES1411, ES2711, ES3441, ES3465, ES3793, ES5686, ES5963 and ES10996) encode small cysteine-rich proteins. Five of these (ES2711, ES3441, ES3465, ES5686 and ES10996) display very low identity with each other, but their cysteine residues, including a triple cysteine motif at or near their C-termini (Fig. 5), are conserved among these proteins. All of these triple cysteine motif-containing proteins have a predicted N-terminal signal peptide, suggesting that they are secreted. The deduced amino acid sequences of ES1411, ES3793 and ES5963 are also cysteine-rich; however, they do not have the triple cysteine motif and do not share any similarity to each other. The deduced amino acid sequences of the other 10 putatively female gametophyte-specific ESTs are not cysteine-rich: ES0465 and ES1086 contain small ORFs while ES8328 and ES8389 have long ORFs. ES0965 encodes a protein similar to an *Arabidopsis* unknown protein (At1g21065). ES2488 is similar to an unknown protein (NP_913003.1) in rice. ES6088 is similar to a hypothetical protein (susceptibility homeodomain transcription factor) conserved in rice and *Arabidopsis* (At1g56580). ES6468 and ES7809 are similar to hypothetical proteins found only in rice.

Other tested clones showed different expression patterns (Fig. 4). Four ESTS (ES0368, ES0411, ES3735 and

Fig. 5 Alignment of amino acid sequences of the triple cysteine motif family. The amino acid sequences deduced from ES2711, ES3441, ES3465 and ES5686 were used to search TIGR (http://www.tigr-blast.tigr.org/tgi/) EST databases; retrieved EST sequences that contained a region encoding the triple cysteine motif were translated and aligned using Multalin (Corpet 1988). Single-letter abbreviations for the amino acid residues are used. Gaps are shown as *dashes*. Clone numbers from *Arabidopsis*, maize, rice and wheat are in *bold italic, bold, italic,* and *underlined italic*, respectively. Amino acids that are conserved in a majority of the sequences across species are indicated in the consensus. Clone numbers from the female gametophyte and the egg cell libraries are in underlined bold; those in normal typeface are from barley

ES5119) were expressed in ovule slices containing female gametophytes and in ovule slices lacking female gametophytes—presumably these genes are expressed in the nucellar cells. ES0368 encodes a protein similar to an *Arabidopsis* hypothetical protein (At5g66550). ES0411 shares similarity with a rice hypothetical protein (CAE05527.1) and with an *Arabidopsis* GPI-anchored protein (At5g63500). ES3735 and ES5119 could encode small proteins (80 and 49 amino acids, respectively) that do not have matches in current databases. Some genes, such as ES0040, ES0904 and ES2773, were expressed in most tissues but not in pollen. Others were detected in all tissues tested. ES3024 was expressed at nearly identical levels in all tissues.

Cell specificities of potentially female gametophyte-specific genes

The RT-PCR results shown in Fig. 4 did not distinguish cell specificity. We therefore used in situ hybridization to determine in which cells selected genes were expressed. For each probe, we examined 96 serial slices from six ovule pieces containing a female gametophyte. Since slices were 9 μm thick, not every slice went through a female reproductive cell, i.e. only one or two slices included the egg cell or synergid cells, and only 3–6 slices included the central cell. Therefore, there was no signal in most of the 96 serial slices. Figure 6 shows representative slices where there was a signal. Strikingly, different members of the triple cysteine family showed different expression patterns. ES2711 was expressed in both the egg cell and central cell (Fig. 6a). ES3441 was expressed in the central cell (Fig. 6c). ES3465 was expressed in the egg cell and synergids (Fig. 6e), and the transcripts of ES5686 were detected in synergids (Fig. 6g). The fifth triple cysteine gene was not tested. Three other tested genes also exhibited distinct expression patterns: ES0777 transcripts were detected in the central cell, transcripts for ES0965 were detected in the synergids, and ES0168 was expressed in both the central cell and the synergids (data not shown).

Triple cysteine motif-containing proteins are found in many plants

Because several members of the triple cysteine family showed intriguing expression patterns within the embryo sac, we searched databases to identify other triple cysteine members. TBLASTN searches were carried out against EST databases (http://www.tigrblast.tigr.org/tgi). No matching sequences were obtained when the full-length amino acid sequences were used as queries. However, when we used the triple cysteine motif and 3–5 amino acids N-terminal to it and set the expect threshold at 1,000, we identified a relatively large number of ESTs, from a variety of species, that could encode triple cysteine motif-containing proteins with predicted N-terminal signal sequences. Figure 5 shows an alignment of some triple cysteine motif-containing proteins. A similar query was used to search TAIR AGI proteins (http://www.Arabidopsis.org/Blast) and additional members were found among hypothetical proteins. There are at least 10 members of the triple

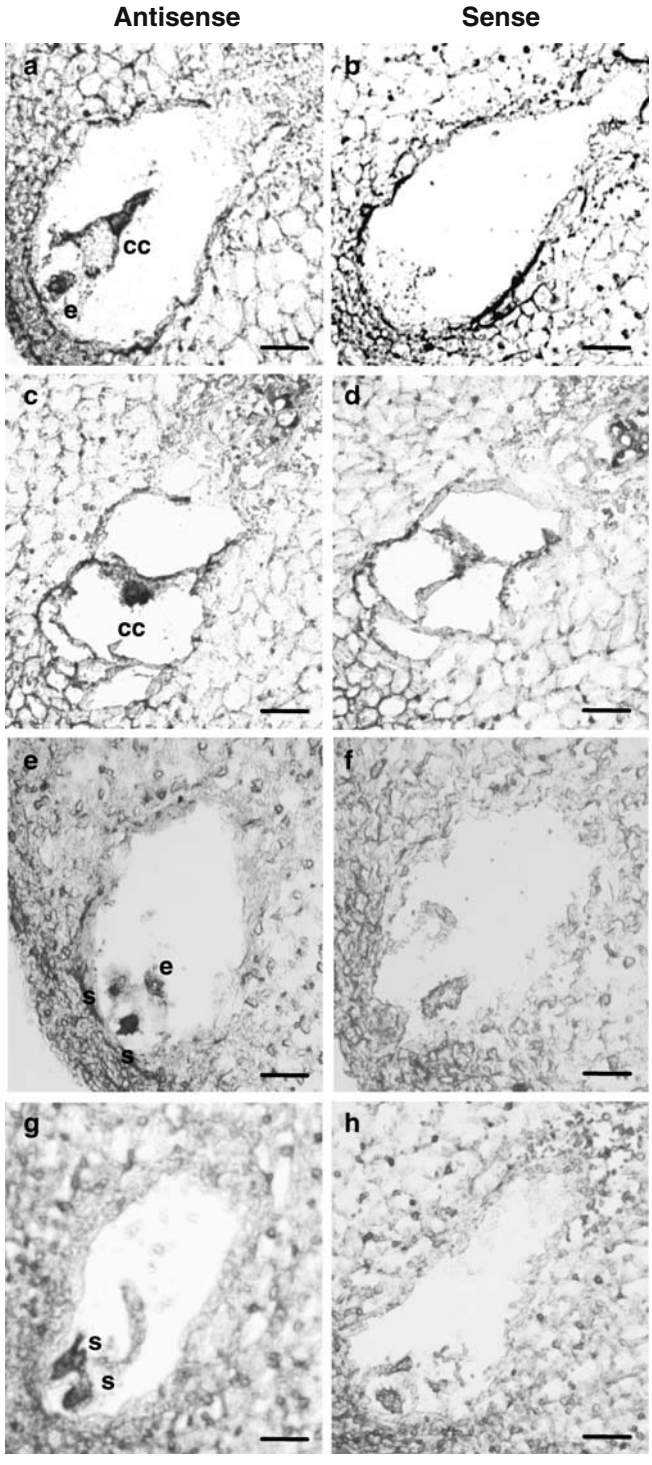

Fig. 6 Cell-specific expression of triple cysteine motif genes. DIG-labeled antisense probes (**a**, **c**, **e**, **g**) and sense probes (**b**, **d**, **f**, **h**) of ES2711 (**a**, **b**), ES3441 (**c**, **d**), ES3465 (**e**, **f**), and ES5686 (**g**, **h**) were used to detect transcripts on ovule slices. Each probe was hybridized with 96 serial slices from six ovule pieces containing a female gametophyte. *Bars* equal 100 μm. *a* antipodal cells; *cc* central cell; *e* egg cell; *s* synergid

cysteine motif family in the *Arabidopsis* genome (*At2g20070*, *At2g20465*, *At2g24693*, *At2g25305*, *At2g42885*, *At3g04540*, *At3g16895*, *At5g42235*, *At5g54220* and *At5g55132*).

Discussion

Although there are extensive ESTs from libraries made from diverse tissues (http://www.tigr.org/tdb/tgi/plant.shtml), transcripts specific to reproductive cells are underrepresented. Recently, transcriptional profiling (Yu et al. 2005) and mutant analyses (Pagnussat et al. 2005) were used to identify *Arabidopsis* genes that are involved in female gametophyte development and function. Here we described sequencing and bioinformatics analyses of ESTs from maize female gametophyte and egg cell cDNA libraries. The maize female gametophyte and egg cell have diverse transcripts. Enzymes related to metabolism were highly represented, an observation also made from analysis of ~~750 wheat egg cell ESTs (Sprunck et al. 2005). Genes involved in cell growth/division were least abundant in the female gametophyte, not surprising because cells in the mature female gametophyte do not divide before fertilization. ESTs encoding transcription and translation factors were overrepresented in the female gametophyte library; these results are consistent with recent expression profiling of reproductive tissues and comparisons to other tissues (Hennig et al. 2004). Defense-related transcripts were abundant in the female gametophyte (Table S1), egg cell (Table S2) and the wheat egg cell libraries (Sprunck et al. 2005), although it is possible that some stress-related transcripts might have been induced during isolation of the female gametophytes; a similar interpretation was used to explain the prevalence of a cold-induced transcript in the sperm cell library (Engel et al. 2003). However, it is also possible that defense systems are active in the female gametophyte in vivo (Heslop-Harrison et al. 1999).

A few female gametophyte-specific genes have been reported in maize (Cordts et al. 2001; Magnard et al. 2003; Márton et al. 2005). We obtained several ESTs identical or similar to these previously known female gametophyte-expressed genes, which validates our library. However, one of our objectives was to identify new female gametophyte-expressed genes potentially involved in pollen tube attraction and gamete interactions. Therefore, when we selected candidates for RT-PCR, we particularly focused on small and secreted proteins that might be involved in cell-to-cell interactions. Because this pre-selection strategy was successful (30% of the genes tested, Fig. 4), extending the strategy to other ESTs should prove fruitful.

Some transcripts already known to be egg cell-specific were enriched in the egg cell ESTs, relative to their abundance in the female gametophyte ESTs. For example, only 1/7165 of the female gametophyte ESTs encoded an ES2 protein (Cordts et al. 2001). By contrast, more than 1% of our egg cell ESTs encode ES proteins (Cordts et al. 2001). We found no *EA1* transcripts (Márton et al. 2005) in 7,165 female gametophyte ESTs but found four ESTs encoding EA1 in 1,415 egg cell ESTs. There were 7/1,415 ESTs encoding one member of the triple cysteine protein gene family in the egg library, but only ES6993 (1/7,165) corresponded to this member, suggesting that it may be egg cell-specific. About 3% of the egg cell ESTs encode germin-like (oxalate oxidase) proteins. By contrast, there were only 2/7,165 female gametophyte ESTs encoding germin-like proteins. The most abundant egg cell contig (Fig. 2b) encodes a protein similar to Bowman-Birk serine protease inhibitor, while only eight of these ESTs were sequenced from the female gametophyte library. On the other hand, the most abundant contig from the female gametophyte (Fig. 2a) encodes a protein annotated as early drought-induced, while only four of these ESTs were sequenced from the egg cell library. Most of the egg cell contigs that encode hypothetical proteins or unknown proteins, or that have no matches in current databases are not yet represented in the female gametophyte contigs. These findings suggest that transcripts from the egg cell were underrepresented even in the female gametophyte cDNA library. There is no overlap between the best *Arabidopsis* matches for hypothetical proteins in our datasets (Tables S1, S2) and the 22 hypothetical proteins indentifed by Yu et al (2005), as pointed out by Yu et al. (2005) in their Table S4. There is no overlap between the genes identified as having defects in fertilization (see Table S2 in Pagnussat et al. 2005) and the best *Arabidopsis* matches to our contigs (Tables S1, S2). Similarly, the subtraction procedure used by Le et al. (2005) identified a few ESTs not yet sequenced from our libraries.

There were several instances where different members of a particular gene family were differentially expressed in the female gametophyte or in sperm cells (Engel et al. 2003). For example, within the multi-gene family that encodes expansins (Lee et al. 2001) the beta expansins EXP14, EXP15, and EXP10 had no EST representation. ESTs for EXP14 and EXP15 were in the female gametophyte library while EXP10 was represented in the sperm cell ESTs; these results suggest that EXP14/EXP15 and EXP10 are gametic cell-expressed. Another example is glycosylphosphatidyl inositol-anchored proteins (GPI-APs); GPI-APs are represented in the female gametophyte ESTs and in the egg cell ESTs but not in the sperm cell ESTs (Engel et al. 2003). In mammals (GPI-APs) are egg cell surface proteins that are important in egg cell-sperm cell interactions (Alfieri et al. 2003).

It is expected that a given cell would have all components for a biochemical pathway or process. We surprisingly found that that was not always the case. For example, ESTs encoding seven large subunit ribosomal proteins and seven small subunit ribosomal proteins were represented in both the female gametophyte and sperm cell ESTs. However, ESTs encoding 17 large subunit and 9 small subunit ribosomal proteins were

represented only in the female gametophyte ESTs, while those for one large subunit ribosomal protein (L20) and one small subunit ribosomal protein (S14) were represented only in the sperm cell ESTs (Engel et al. 2003). We cannot exclude that depth of sequencing might influence the EST representation, but given that we found 28 ESTs encoding L35A in the female gametophyte ESTs, if L20 and S14 were expressed in the female gametophyte, it seems reasonable that we should have found at least one EST for L20 and S14. Perhaps the male and female gametes bring different ribosomal protein transcripts to the zygote.

Communication between the pollen tube and the female gametophyte controls the growth of the pollen tube towards its target, the female gametophyte. The signaling mechanisms are largely unknown but both surface-localized receptors and small molecules are likely involved. Although there are more than 600 receptor-like kinases in *Arabidopsis* genome, only a few ligands have been identified (reviewed in Torii 2004). Among the presumed female gametophyte-specific genes, we identified a novel gene family that is predicted to encode small, secreted and cysteine-rich proteins. Given their expression patterns (Fig. 6), these proteins are potential candidates for signaling roles, either as pollen tube attractants or for gamete interactions, as a small cysteine-rich protein is the ligand for the receptor kinase that mediates self-incompatibility (Schopfer et al. 1999), and small cysteine-rich proteins bind the extracellular domains of pollen receptor kinases (Tang et al. 2002, 2004). Eight of the triple cysteine motif-containing proteins in *Arabidopsis* (At2g24693, At2g25305, At2g42885, At3g04540, At3g16895, At5g42235, At5g54220, and At5g55132) were not predicted as individual genes in the first annotation (The *Arabidopsis* Genome Initiative 2000). Additional members may be missed in the current annotation.

This female gametophyte cDNA library, this egg cell cDNA library and these two EST databases are useful resources. In addition, the female gametophyte and egg cell ESTs provide expression evidence that over 350 proteins previously annotated as hypothetical in *Arabidopsis* and/or rice are in fact unknown proteins, whose functions in the female gametophyte can now be explored. The ESTs reported here, as well as other resources (Le et al. 2005; Pagnussat et al. 2005; Yu et al. 2005) can serve as cellular markers (Fig. 6) to study female gametophyte development and the promoters of such genes can be used to manipulate gene expression. In the future, these ESTs can be used to develop female reproductive cell-oriented gene chips to help identify the genes responsible for female reproductive mutants and to monitor changes in gene expression during gametophyte development, double fertilization and early embryogenesis.

Acknowledgments We thank Priti Patel, Michelle Meador, Teresa Mok, Jessica Kim and Jungsun Lee for technical assistance, and all members of our laboratory for useful discussions. This work was supported by the National Science Foundation (Plant Genome grant no. 0211742).

References

Alfieri JA, Martin AD, Takeda J, Kondoh G, Myles DG, Primakoff P (2003) Infertility in female mice with an oocyte-specific knockout of GPI-anchored proteins. J Cell Sci 116:2149–2155

Christensen CA, Gorsich SW, Brown RH, Jones LG, Brown J, Shaw JM, Drews GN (2002) Mitochondrial GFA2 is required for synergid cell death in *Arabidopsis*. Plant Cell 14:2215–2232

Camon E, Magrane M, Barrell D, Binns D, Fleischmann W, Kersey P, Mulder N, Oinn T, Maslen J, Cox A, Apweiler R (2003) The gene ontology annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro. Genome Res 13:662–672

Cordts S, Bantin J, Wittich PE, Kranz E, Lörz H, Dresselhaus T (2001) *ZmES* genes encode peptides with structural homology to defensins and are specifically expressed in the female gametophyte of maize. Plant J 25:103–114

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16:10881–10890

Dresselhaus T, Lörz H, Kranz E (1994) Representative cDNA libraries from few plant cells. Plant J 5:605–610

Dresselhaus T, Hagel C, Lörz H, Kranz E (1996) Isolation of a full-length cDNA encoding calreticulin from a PCR library of in vitro zygotes of maize. Plant Mol Biol 31:23–34

Dresselhaus T, Cordts S, Heuer S, Sauter M, Lörz H, Kranz E (1999a) Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles. Mol Gen Genet 261:416–427

Dresselhaus T, Cordts S, Lörz H (1999b) A transcript encoding translation initiation factor eIF-5A is stored in unfertilized egg cells of maize. Plant Mol Biol 39:1063–1071

Dresselhaus T, Amien S, Marton M, Strecke A, Brettschneider R, Cordts S (2005) *TRANSPARENT LEAF AREA1* encodes a secreted proteolipid required for anther maturation, morphogenesis, and differentiation during leaf development in maize. Plant Cell 17:730–745

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

Hennig L, Gruissem W, Grossniklaus U, Kohler C (2004) Transcriptional programs of early reproductive stages in *Arabidopsis*. Plant Physiol 135:1–11

Heslop-Harrison J, Heslop-Harrison JS, Heslop-Harrison Y (1999) The structure and prophylactic role of the angiosperm embryo sac and its associated tissues: *Zea mays* as a model. Protoplasma 209:256–272

Higashiyama T, Yabe S, Sasaki N, Nishimura Y, Miyagishima S, Kuroiwa H, Kuroiwa T (2001) Pollen tube attraction by the synergid cell. Science 293:1480–1483

Huck N, Moore JM, Federer M, Grossniklaus U (2003) The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception. Development 130:2149–2159

Hülskamp M, Schneitz K, Pruitt RE (1995) Genetic evidence for a long-range activity that directs pollen tube guidance in *Arabidopsis*. Plant Cell 7:57–64

Kranz E, Bautor J, Lörz H (1991) In vitro fertilization of single, isolated gametes of maize mediated by electrofusion. Sex Plant Reprod 4:12–16

Le Q, Gutiérrez-Marcos JF, Costa LM, Meyer S, Dickinson HG, Lörz H, Kranz E, Scholten S (2005) Construction and screening of subtracted cDNA libraries from limited populations of plant cells: a comparative analysis of gene expression between maize egg cells and central cells. Plant J 44:167–178

Lee Y, Choi D, Kende H (2001) Expansins: ever-expanding numbers and functions. Curr Opin Plant Biol 4:527–532

Linnestad C, Doan DN, Brown RC, Lemmon BE, Meyer DJ, Jung R, Olsen OA (1998) Nucellain, a barley homolog of the dicot

vacuolar-processing protease, is localized in nucellar cell walls. Plant Physiol 118:1169–1180

Magnard JL, Le Deunff E, Domenech J, Rogowsky PM, Testillano PS, Rougier M, Risueno MC, Vergne P, Dumas C (2000) Genes normally expressed in the endosperm are expressed at early stages of microspore embryogenesis in maize. Plant Mol Biol 44:559–574

Magnard JL, Lehouque G, Massonneau A, Frangne N, Heckel T, Gutierrez-Marcos J, Perez P, Dumas C, Rogowsky PM (2003) *ZmEBE* genes show a novel, continuous expression pattern in the central cell before fertilization and in specific domains of the resulting endosperm after fertilization. Plant Mol Biol 53:821–836

Márton ML, Cordt S, Broadhvest J, Dresselhaus T (2005) Micropylar pollen tube guidance by egg apparatus 1 of maize. Science 307:573–576

Okamoto T, Higuchi K, Shinkawa T, Isobe T, Lörz H, Koshiba T, Kranz E (2004) Identification of major proteins in maize egg cells. Plant Cell Physiol 45:1406–1412

Okamoto T, Scholten S, Lörz H, Kranz E (2005) Identification of genes that are up- or down-regulated in the apical or basal cell of maize two-celled embryos and monitoring their expression during zygote development by a cell manipulation- and PCR-based approach. Plant Cell Physiol 46:332–338

Opsahl-Ferstad HG, Le Deunff E, Dumas C, Rogowsky PM (1997) *ZmEsr*, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. Plant J 12:235–246

Pagnussat GC, Yu HJ, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie LF, Ye D, Sundaresan V (2005) Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. Development 132:603–14

Ray SM, Park SS, Ray A (1997) Pollen tube guidance by the female gametophyte. Development 124:2489–2498

Rotman N, Rozier F, Boavida L, Dumas C, Berger F, Faure JE (2003) Female control of male gamete delivery during fertilization in *Arabidopsis thaliana*. Curr Biol 13:432–436

Schopfer CR, Nasrallah ME, Nasrallah JB (1999) The male determinant of self-incompatibility in *Brassica*. Science 286:1697–1700

Shimizu KK, Okada K (2000) Attractive and repulsive interactions between female and male gametophytes in *Arabidopsis* pollen tube guidance. Development 127:4511–4518

Sprunck S, Baumann U, Edwards K, Langridge P, Dresselhaus T (2005) The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.) Plant J 41:660–672

Swanson R, Edlund AF, Preuss D (2004) Species specificity in pollen-pistil interactions. Annu Rev Genet 38:793–818

Tang W, Ezcurra I, Muschietti J, McCormick S (2002) A cysteine-rich extracellular protein, LAT52, interacts with the extracellular domain of the pollen receptor kinase LePRK2. Plant Cell 14:2277–2287

Tang W, Kelley D, Ezcurra I, Cotter R, McCormick S (2004) LeSTIG1, an extracellular binding partner for the pollen receptor kinases LePRK1 and LePRK2, promotes pollen tube growth in vitro. Plant J 39:343–353

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Torii KU (2004) Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. Int Rev Cytol 234:1–46

Torres MA, Rigau J, Puigdomenech P, Stiefel V (1995) Specific distribution of mRNAs in maize growing pollen tubes observed by whole-mount in situ hybridization with non-radioactive probes. Plant J 8:317–321

Weterings K, Russell SD (2004) Experimental analysis of the fertilization process. Plant Cell 16(Suppl):S107–S118

Yu HJ, Hogan P, Sundaresan V (2005) Analysis of the female gametophyte transcriptome of *Arabidopsis* by comparative expression profiling. Plant Physiol 139:1853–1869