 DOI 10.1002/pmic.200402011 *Proteomics* 2005, *5,* 4864–4884

REGULAR ARTICLE

Proteome mapping of mature pollen of *Arabidopsis thaliana*

Rachel Holmes-Davis[1], Charlene K. Tanaka[2], William H. Vensel[2], William J. Hurkman[2] and Sheila McCormick[1]

[1] Plant Gene Expression Center, USDA/ARS and UC Berkeley, Albany, CA, USA
[2] Crop Improvement and Utilization Research Unit, Western Regional Research Center, USDA/ARS, Albany, CA, USA

The male gametophyte of *Arabidopsis* is a three-celled pollen grain that is thought to contain almost all the mRNAs needed for germination and rapid pollen tube growth. We generated a reference map of the *Arabidopsis* mature pollen proteome by using multiple protein extraction techniques followed by 2-DE and ESI-MS/MS. We identified 135 distinct proteins from a total of 179 protein spots. We found that half of the identified proteins are involved in metabolism (20%), energy generation (17%), or cell structure (12%); these percentages are similar to those determined for the pollen transcriptome and this similarity is consistent with the idea that in addition to the mRNAs, the mature pollen grain contains proteins necessary for germination and rapid pollen tube growth. We identified ten proteins of unknown function, three of which are flower- or pollen-specific, and we identified nine proteins whose RNAs were absent from the transcriptome, seven of which are involved in metabolism, energy generation, or cell wall structure. Our work complements and extends recent analyses of the pollen transcriptome.

Received: November 24, 2004
Revised: March 18, 2005
Accepted: April 6, 2005

Keywords:
Electrospray ionization tandem mass spectrometry (ESI-MS/MS) / Male gametophyte / Plant / Two-dimensional gel electrophoresis (2-DE)

1 Introduction

In plants male gametogenesis occurs within the anthers of the flower and begins with the division of a diploid sporophytic cell. The daughter cells of this division, the tapetal initial and the pollen mother cell, have different fates. The tapetal initials form the tapetum, the tissue that supplies nutrients to the developing pollen grain and secretes compounds onto the surface of the developing pollen grain. The pollen mother cells undergo meiosis to yield a tetrad of haploid cells called microspores. The tapetum releases an enzyme called callase to degrade the callose wall that holds the microspores together. Once freed from the callose wall, these uninucleate microspores undergo an asymmetric mitotic division to form the bicellular pollen grain, in which the smaller generative cell is fully enclosed within the larger vegetative cell. In approximately 70% of plant families (*e.g.*, Solanaceae and Liliaceae) bicellular pollen grains are released from the anther before the generative cell divides to form two sperm cells. In these species the second mitotic division occurs while the pollen tube grows through the pistil, the female organ of the flower. In other plant families, such as the Cruciferae, to which *Arabidopsis* belongs, the second mitotic division occurs before the pollen is shed from the plant.

Upon pollination, the vegetative cell grows a pollen tube that delivers the sperm cells to the embryo sac for fertilization of the egg and the central cell. Genes involved in pollen germination have mostly been identified in lily and in plants in the Solanaceae [1–9], because large quantities of pollen can easily be collected and *in vitro* pollen germination is relatively simple to perform in these species [10, 11]. However, genomic information about these species is incomplete and a proteomics approach to identify proteins in these spe-

Correspondence: Dr. Sheila McCormick, Plant Gene Expression Center, USDA/ARS and UC Berkeley, 800 Buchanan St., Albany, CA 94710, USA
E-mail: sheilamc@nature.berkeley.edu
Fax: +1-510-559-5678

cies would be difficult because of limited protein database information. The completion of the sequence of the *Arabidopsis* genome [12] makes this plant ideal for proteomics studies, but collecting large quantities of *Arabidopsis* pollen can be difficult. The development of new vacuuming techniques for collecting milligram amounts of pollen [13] now makes it possible to study the mature pollen proteome.

Few detailed proteomic studies have been carried out on the plant male gametophyte. The most comprehensive of these studies are the proteome reference maps that have been generated for rice anther proteins during male gametophyte development [14, 15]. These studies focused on identifying a variety of sporophytic and gametophytic proteins important for microspore development. With this baseline knowledge, Imin *et al*. [16] detected 70 proteins that were differentially accumulated during early cold stress in anthers and identified 18 of those 70 proteins by MALDI-TOF MS. In addition to the rice anther studies, a proteomic study of the *Arabidopsis* pollen coat, following cyclohexane extraction [17], identified 12 proteins, 9 of which were oleosins and lipases. These proteins, along with lipids, are thought to support the initial signaling interactions between the pollen and pistil [17–19].

Recent publications of the *Arabidopsis* pollen transcriptome [20, 21] presented a genome-wide examination of the RNAs present within in the mature pollen grain and identified those RNAs that are enriched or selectively expressed in pollen. Of the approximately 26 000 genes in the *Arabidopsis* genome [22] approximately 6600 are expressed in pollen [23, 24]. These studies determined that approximately one-third of the genes that are expressed in vegetative tissues are not expressed in pollen and that 10–20% of the genes expressed in pollen are specific to pollen [20, 21]. Nearly half of the pollen-specific mRNAs encode proteins needed for cell wall metabolism, for signal transduction, or for the generation of the actin-rich cytoskeleton [20], supporting the idea that the mature pollen grain contains mRNAs needed for rapid germination and pollen tube growth [25]. The presence in mature pollen of both mRNAs and the corresponding proteins are documented for several genes that are expressed in pollen [3, 10, 26–29], although there is one report where the corresponding protein was not detected until after germination [30].

We wanted to address on a larger scale the question of whether the proteins encoded by RNAs are present in pollen or are translated later. To this end, we used a proteomics approach to identify proteins that are present in the mature pollen grain. We used 2-DE followed by ESI-MS/MS to generate a reference map of the *Arabidopsis* mature pollen proteome. Our findings complement the recent transcriptome studies of pollen and significantly extend the proteomics studies. Of the 135 distinct proteins we identified, half are involved in metabolism (20%), energy generation (17%), or cell structure (12%), percentages close to those determined for the pollen transcriptome. Additionally, we identified nine proteins whose RNAs were absent from the transcriptome [23, 24]; seven of these nine proteins are also involved in metabolism, energy generation, or cell wall structure.

2 Materials and methods

2.1 Isolation of proteins from mature *Arabidopsis* pollen

Mature pollen from *Arabidopsis thaliana* (Columbia ecotype) was harvested using a modified hand-held vacuum cleaner outfitted with plumbing fittings and nylon meshes [13]. In order to collect sufficient amounts of pollen to isolate enough protein for this study, pollen was collected from approximately 10 000 plants that were grown at three separate times for a total of ten collections (pollen can be collected every 2–3 days from plants that are flowering). After collection, the pollen was weighed and stored at −80°C. Proteins were extracted from 35 mg of pollen pooled from several collections using the method of Giavalisco *et al*. [31]. Because they were isolated from pooled pollen samples, the protein samples should be considered "averaged samples" and not biological replicates. Mature pollen was homogenized using a ground-glass tissue grinder (Kontes, Vineland, NJ) in approximately 400 μL of a buffer containing 2.5 parts of mixture 1 and 1 part of mixture 2 [31] such that the final extraction buffer contained 71.4 mM KCl, 14.3% v/v glycerol, 37.7 mM Tris pH 7.4, 0.29 mM pepstatin A, 0.4 mM PMSF, 28.6% ethanol, and one Complete™ tablet (Roche Applied Science, Indianapolis, IN) *per* 2 mL of solution. The homogenate was centrifuged at 202 000 × *g* at 4°C for 1 h. The proteins in the supernatant (salt-soluble (SS) fraction) were precipitated with 5 vol of 100% acetone and incubated at -20°C for 1 h. After centrifugation, the pellet was washed three times with 80% acetone and then resuspended in urea buffer (9 M urea, 4% NP 40, 1% DTT, and 2% ampholytes). The pellet from the SS fraction was homogenized in a ground-glass tissue grinder with approximately 100 μL R2D2 buffer (5 M urea, 2 M thiourea, 2% CHAPS, 2% 3-(decyldimethylamino)propanesulfate inner salt (SB3–10), 1 M DTT, 100 mM tris 2-carboxyethyl phosphine (TCEP), 0.5% ampholytes pH 4–6.5, and 0.25% ampholytes pH 3–10) [32]. The homogenate was centrifuged at 202 000 × *g* at 4°C for 1 h; the proteins in the supernatant (salt-insoluble (SI) fraction) were precipitated with 5 vol of 100% acetone and incubated at −20°C for 1 h. After centrifugation, the pellet was washed three times with 80% acetone and then resuspended in urea buffer. For SDS extraction, 25 mg of mature pollen was homogenized in 200 μL SDS extraction buffer (200 mM Tris pH 8.9, 4% SDS, 20% glycerol, and 80 mM DTT) using a ground-glass tissue grinder (Kontes). The homogenate was centrifuged at 202 000 × *g* at 4°C for 1 h and the proteins in the supernatant were precipitated with 5 vol of 100% acetone and incubated at −20°C for 1 h. After centrifugation, the pellet was washed three times with 80% acetone and then resuspended in urea buffer.

All protein samples were incubated for 1 h at room temperature and centrifuged at 16 000 × *g* for 15 min (Eppendorf 5415C, Brinkman Instruments, Westbury, NY) to remove insoluble material before electrophoresis. The protein samples were quantitated using the BioRad DC protein Assay Kit (BioRad, Hercules, CA).

2.2 2-DE

Proteins were separated in the first dimension by IEF (Mini Protean II Tube Cell; BioRad Laboratories) according to the manufacturer instructions with the modifications outlined here. The gels (8.7 cm long and 1 mm in diameter) contained 9.2 M urea, 4% (total monomer) acrylamide-Bis, 2% NP 40, 2% 3–10 Iso-Dalt Grade Servalyts (Crescent Chemical, Islandia, NY), 0.015% ammonium persulfate, and 0.125% TEMED. The upper electrode (anode) buffer was 0.2% v/v sulfuric acid and the lower electrode buffer (cathode) was 0.5% v/v ethanolamine. The gels were prefocused at 200 V for 10 min, 300 V for 15 min, and 400 V for 15 min. Samples containing 18 μg of protein were loaded at the acidic end of the IEF gel and overlaid with 5 M urea. IEF gels were run at 500 V for 10 min and then at 750 V for 1 h. The gels were extruded into microcentrifuge tubes using a tube gel ejector attached to a 10 mL syringe without buffer. Equilibration buffer (2.3% SDS, 10% glycerol, 0.05% DTT, 62.5 mM Tris-Cl pH 6.8) was added and the gels were frozen immediately by placing the tubes in dry ice; tubes were stored at −70°C. Proteins were separated in the second dimension by SDS gel electrophoresis (X-Cell Sure-Lock Mini Cell electrophoresis system; Invitrogen, Carlsbad, CA). IEF gels were thawed, placed on top of Novex NuPAGE Bis-Tris 4–12% acrylamide minigels (8 cm by 8 cm) with 2-DE and reference wells (Invitrogen), and overlaid with 45 μL of equilibration buffer. Mark12™ (Invitrogen) unstained molecular weight standard (3 μL) was loaded into the reference well. The SDS gels were run with NuPAGE MES-SDS running buffer (Invitrogen) for 48 min at 200 V. The gels were stained with CBB G-250 (Sigma, St. Louis, MO) according to Kasarda *et al.* [33], destained in water for 3–5 h, and stored at 4°C in 20% ammonium sulfate. We ran at least four gels for each protein fraction. With the exception of a few of the exceptionally low-abundance proteins, which were not visible on all gels, the protein patterns were identical. The gels were digitized with a calibrated scanner (PowerLook III, UMAX Technologies, Dallas, TX) at 300 dpi; the same settings were used for all gels. Computer software (Progenesis Workstation ver. 2002.1; Non-linear Dynamics, Newcastle upon Tyne, UK) was used to calculate normalized spot volumes (individual spot volume/total spot volume × 100) for the three fractions. For each protein, the normalized protein volume was compared with the corresponding mRNA signal in the pollen transcriptome [23, 24]. Proteins identified from spots with multiple protein identifications were excluded from the analysis, as were the proteins whose corresponding mRNAs are not represented in the pollen transcriptome.

2.3 Protein identification by MS

Protein spots were excised from 2-DE gels and digested in-gel as described previously [34], except that during the final extraction step ACN was replaced with 10% formic acid. We pooled spots from four gels of each protein extraction. ESI-MS/MS was performed with an API QSTAR Pulsar *i* hybrid quadrupole-TOF instrument (Applied Biosystems/MDS Sciex, Toronto, Canada). Sample introduction was *via* a Proxeon Biosystems (Odense, Denmark) nano-electrospray source. In-gel digests (20 μL) were loaded automatically onto a C-18 trap cartridge and chromatographed on an RP column (Vydac 238EV5.07515, 75 μm × 150 mm; Hesperia, CA) fitted at the effluent end with a coated spray tip (FS360–50–5-CE; New Objective Woburn, MA). An LC Packings nano-flow LC system (Dionex, Sunnyvale, CA) with an autosampler, column-switching device, loading pump, and nano-flow solvent delivery system was used to elute the column. The elution solvents were: A (0.5% acetic acid) and B (80% ACN, 0.5% acetic acid). Samples were eluted at a column flow rate of 220 nL/min. The column was equilibrated with 8% B and 92% A before each injection. At injection, a linear gradient was initiated, reaching 80% B and 20% A by 12 min. At 12 min, the column was flushed with 80% B and 20% A for 1 min then returned to the starting condition within 1 min. Before loading the next sample, the column was equilibrated with 8% B and 92% A for 14 min. The mass spectrometer was calibrated daily and operated above a resolution of 8000 with a mass accuracy of 10–50 ppm with external calibration. The acquisition cycle time of 4 s consisted of a single 1 s MS "survey" scan followed by a 3 s MS/MS scan. The dynamic exclusion window was set to always exclude previously fragmented masses. Doubly or triply charged ions with intensities greater than 40 counts in the survey scan were selected for fragmentation. Collision energy optimized for charge state and m/z was determined by the Analyst QS software. Nitrogen was used for the collision gas and the pressure in the collision cell ranged from 3×10^{-6} to 6×10^{-6} torr.

Spectra were converted to a text format and analyzed using the KNEXUS data automation client (Genomic Solutions, Ann Arbor, MI) with the Sonar search engine selected. Sonar incorporates a, b, and y ions and uses vector alignment and spectrum dot products to calculate correlations between mass spectra and database peptide sequences [35]. The following constraints were used to search the databases: tryptic peptides with no more than one missed cleavage, iodoacetamide modification of sulfhydryl groups, partial oxidation of methionine residues, mass range of 2–120 kDa, and a p*I* of 5–9 for the intact protein. The NCBI nonredundant protein database (TIGR AGI annotation 11.0, release date January 12, 2004) was searched and mass spectrometer identification scores are reported as expectation values. Sonar reports protein identifications and expectation values, *i.e.*, the number of matches expected if the matches were completely random. An expectation value of 1 means that at least one similar match would be expected when searching a database that did

not include the protein sequence that truly matches the data. Thus a lower expectation value makes it more likely that the match is correct and not random. An expectation value of 1×10^{-3} can be interpreted as one chance in 1000 that the match was random. Thus, as expectation values approach zero, confidence in the identification increases. *A. thaliana* proteins identified with expectation values of 1×10^{-3} or less are reported.

2.4 Determination of functional categories and predictions of subcellular localization for the identified proteins

The proteins were categorized by function according to data from MAtDB (http://mips.gsf.de/proj/thal/db/search/search_frame.html), the Goldberg Laboratory EST list (http://www.mcdb.ucla.edu/Research/Goldberg/ests/functional_categories-index.htm), and using the Classification SuperViewer on the *Arabidopsis* Functional Genomics Tools of the BBC website (http://bbc.botany.utoronto.ca/welcome.htm). Most proteins were similarly categorized by all three methods. In the cases of discrepancies or for proteins to which no classification had been assigned, annotation was completed manually.

For proteins with unknown function, protein localization was predicted with PSORT (http://psort.ims.u-tokyo.ac.jp/) after selecting "plant" as the source of the input sequence and protein alignments were performed with ClustalW (http://www.ebi.ac.uk/clustalw/).

2.5 RT-PCR analysis

RNA was extracted from 3-wk-old leaves, 3-wk-old roots, open flower buds, closed flower buds wherein pollen was not dehisced, mature pollen, and from pistils removed from closed flower buds, using the RNeasy kit (Qiagen, Valencia, CA). Three micrograms of total RNA was used in each RT reaction. The RT reactions were completed using random hexamers and the SuperScript II kit (Invitrogen). Primers for RT-PCR were as follows: for At5g61720, forward 5'-CGGGTCAAAAGACAACCAGT-3' and reverse 5'-CCCATTTTGGGGGTATTCTT-3' (yields a 514 bp genomic fragment or a 409 bp cDNA amplified fragment); for At5g62750 forward 5'-GAGGAACACAACAAGGCTGAG-3' and reverse 5'-acattcagcagcatttgttattg-3' (yields a 401 bp fragment); and for At3g01250 forward 5'-CTCATGCAAGCAGAGGAGAA-3' and reverse 5'-CTTTCTCCTTCTCCGGAACC-3' (yields a 492 bp fragment). Each 50 μL PCR reaction mix contained 1 μL of the RT reaction, 0.2 mM dNTPs, 0.2 μM forward primer, 0.2 μM reverse primer, 1 × Taq buffer (NEB, Beverly, MA), and 2.5 U Taq polymerase (NEB). The PCR cycling conditions were: 5 min at 95°C; 30 cycles at 30 s at 95°C, 30 s at 60°C, 75 s at 72°C; 10 min at 72°C. The control used for RT-PCR was a putative vacuolar ATPase (At4g38510); ATPase forward 5'-ATCATCCTCCTGATGCTCG-3' and reverse 5'-CGTGGTCAAGTTCTCGAAG-3' (yields a 424 bp cDNA amplified fragment). Photographs of the gels were digitized using the Epson 4870 (Epson America, Long Beach, CA) and the images were inverted using Adobe Photoshop Elements version 2.0 (Adobe Systems, San Jose, CA).

3 Results and discussion

Mature *Arabidopsis* pollen was harvested using a modified hand-held vacuum cleaner outfitted with nylon meshes with decreasing pore size. Flower parts are trapped on the 80 and 35 μm meshes and pollen (<30 μm diameter) collects on the 6-μm mesh. We assessed the purity of the pollen by light microscopy. Figure 1A shows a typical field of vacuum-collected pollen; increased magnification (Fig. 1B) shows that no other plant parts were present in the preparation.

A multistep procedure was used to extract proteins from mature pollen of *A. thaliana*, in order to obtain proteins with different solubility properties and to generate a large pool of proteins to identify. Mature pollen proteins were first separated into SS and SI fractions using a buffer containing 71.4 mM KCl. The SI fraction was then treated with a buffer to obtain proteins soluble in detergents and chaotropes (R2D2 [32]). Because SDS solubilization works well for analysis of plant proteins, including membrane proteins [36], mature pollen proteins were separately extracted using a 4% SDS buffer. As shown in Fig. 2, 2-DE resolved 455 protein spots in the SS fraction (Fig. 2A), 610 protein spots in the SI

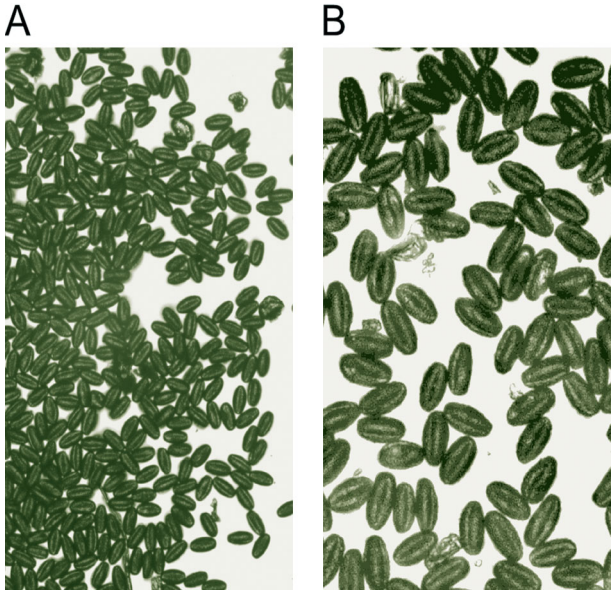
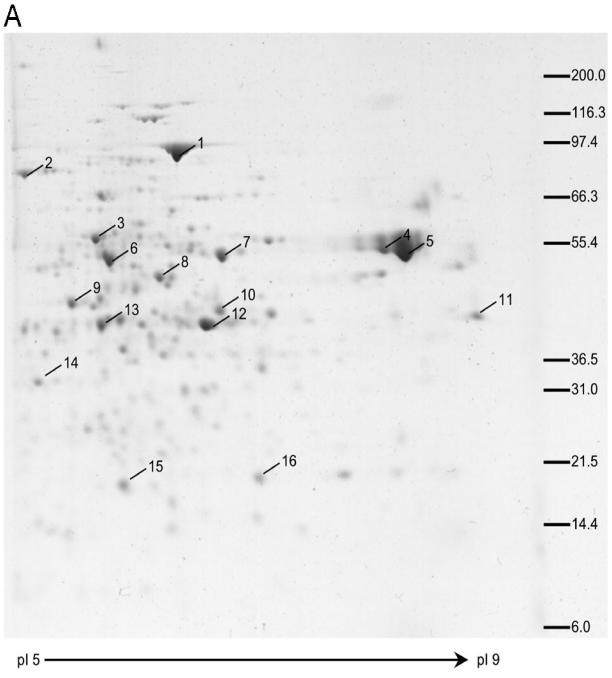


Figure 1. Pollen grains from a representative vacuum-collected sample. (A) Approximately 5 mg of pollen collected on a 6-μm nylon mesh. (B) Higher magnification of the pollen grains showing the purity of the sample, specifically the absence of plant debris. Pollen grains appear dark because the sample is illuminated from below. White bar represents 10 μm.

fraction (Fig. 2B), and 688 protein spots in the SDS fraction (Fig. 2C). The 2-DE patterns for the SI and SS fractions were qualitatively and quantitatively different. The SI fraction had more proteins with MW of 55–200 kDa and basic p*I* than did the SS fraction. The pattern for the SDS fraction was similar to the pattern for the SS fraction. This is not surprising because the SDS fraction is a total pollen extract and includes



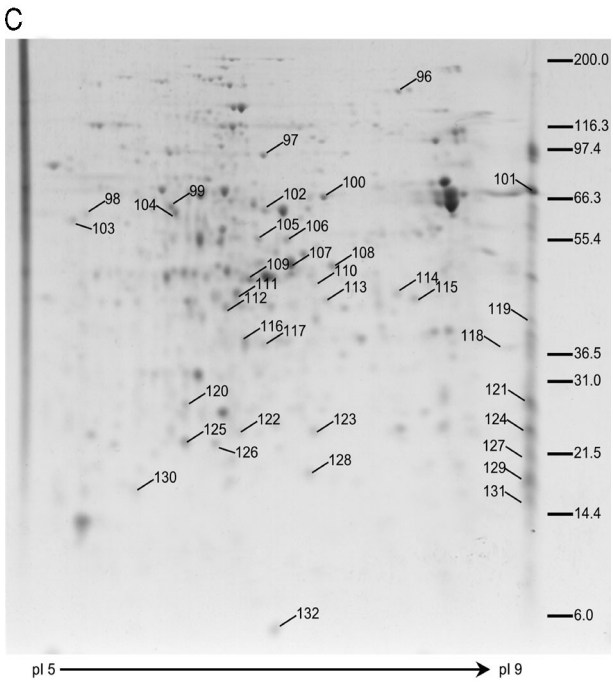


Figure 2. 2-DE gels of the SS, SI, and SDS extracted proteins from mature pollen. (A) Proteins soluble in 71.4 mM KCl. (B) Proteins insoluble in 71.4 mM KCl but soluble in R2D2 buffer. (C) Proteins extracted with 4% SDS. All spots with protein identifications are numbered and are referenced in Table 1 and Table S1.

proteins present in both the SS and SI fractions. In addition, the buffers used for solubilizing the SI and SDS fraction proteins contained nonionic detergents and would be expected to extract similar protein populations [36–39]. Because SDS is highly effective in solubilizing many different types of proteins, including those with many transmembrane domains, we thought it was worthwhile to extract proteins with SDS, even though not all would be resolved by 2-DE.

We excised a combined total of 225 randomly chosen spots from the gels of the three extractions. To ensure a distribution of abundant and lower-abundance proteins and to include proteins from a wide range of molecular weights and p*I*s, we picked spots from all sections of the gels. We identified 179 proteins (Table 1, Table S1) representing 135 distinct genes (*i.e.*, proteins with distinct gene numbers or At numbers, http://www.*Arabidopsis*.org/). Of these 135 proteins, 114 are represented as RNAs in the *Arabidopsis* pollen transcriptome [23, 24], 9 are not expressed in the transcriptome, and 12 are not represented on the Affymetrix ATH1 Genome Array that was used for the transcriptome analysis. About 10% of the proteins we identified are encoded by genes identified as pollen-specific from microarray analysis [20, 21] or as estimated from EST representation.

Of the 135 proteins identified (Table 1, Table S1), 27 (20%) are predicted to function in metabolism, 23 (17%) are predicted to function in energy generation, and 17 (12%) are predicted to function in cell structure (Table 2). Our results

Table 1. Proteins identified in mature *Arabidopsis* pollen

Spot number	Extraction method[a)]	At number	Protein similarity	p*I*	kDa	Expect[b)]	Pollen expr.[c)]	Pollen-specific[d)]
Cell growth and division (3)[e)]								
95	SI	At4g29340	Profilin 3	5.0	14.6	3.4×10^{-18}	Yes	Yes
116	SDS	At5g40770	Prohibitin	7.0	30.4	1.3×10^{-5}	Yes	No
81	SI	At3g16640	Translationally controlled tumor protein-like protein, related to guanine-nucleotide free chaperones	4.5	18.9	1.3×10^{-23}	Yes	No
Cell structure (17)								
9	SS	At2g37620	Actin 1	5.3	42.0	3.1×10^{-48}	No	No
51	SI	At3g18780	Actin 12	5.4	42.0	3.2×10^{-46}	Yes	No
92	SI	At4g25590	Actin-depolymerizing factor-related protein	5.1	15.4	1.8×10^{-15}	Yes	Yes
130	SDS	At5g52360	Actin-depolymerizing factor, putative	5.6	16.0	4.9×10^{-23}	Yes	Yes
93	SI	At5g52360	Actin-depolymerizing factor, putative	5.6	16.0	2.2×10^{-8}	Yes	Yes
45	SI	At3g14040	Exopolygalacturonase	8.6	46.4	2.0×10^{-8}	NR[d)]	No
5	SS	At3g14040	Exopolygalacturonase	8.6	46.4	7.0×10^{-42}	NR	No
76	SI	At3g05930[f)]	Germin-like protein, GLP8, possible oxalate oxidase activity	8.9	22.7	2.7×10^{-11}	Yes	Yes
91	SI	At3g05930[f)]	Germin-like protein, GLP8, possible oxalate oxidase activity	8.9	22.7	4.4×10^{-5}	Yes	Yes
96	SDS	At2g16730	Glycosyl hydrolase family 35 protein	8.8	96.6	1.4×10^{-23}	Yes	Yes
17	SI	At2g16730	Glycosyl hydrolase family 35 protein	8.8	96.6	1.0×10^{-28}	Yes	Yes
125	SDS	At4g24640	Invertase/pectin methylesterase inhibitor family	5.5	20.1	1.6×10^{-12}	Yes	Yes
16	SS	At4g24640	Invertase/pectin methylesterase inhibitor family	5.5	20.1	2.2×10^{-20}	Yes	Yes
110	SDS	At3g13390	Pectinesterase (pectin methylesterase) family	9.1	62.3	4.2×10^{-33}	Yes	No
23	SI	At3g13390	Pectinesterase (pectin methylesterase) family	9.1	62.3	1.9×10^{-9}	Yes	No
54	SI	At3g62170	Pollen-specific pectin esterase, putative	8.7	63.4	2.2×10^{-18}	Yes	No
11	SS	At3g62170	Pollen-specific pectin esterase, putative	8.7	63.4	9.2×10^{-30}	Yes	No
30	SI	At3g07850[f)]	Polygalacturonase (EC 3.2.1.15)	8.6	46.3	5.5×10^{-8}	No	No
40	SI	At3g07850[f)]	Polygalacturonase (EC 3.2.1.15)	8.6	46.3	9.6×10^{-12}	No	No
4	SS	At3g07850	Polygalacturonase (EC 3.2.1.15)	8.6	46.3	2.1×10^{-47}	No	No
44	SI	At1g02790	Polygalacturonase, putative	8.5	45.1	4.9×10^{-19}	Yes	No
59	SI	At3g02230	Reversibly glycosylated polypeptide-1 (RGP-1), alpha-1,4-glucan-protein synthase	5.6	41.1	2.0×10^{-56}	NR	No
13	SS	At3g02230	Reversibly glycosylated polypeptide-1 (RGP-1), alpha-1,4-glucan-prot synthase	5.6	41.1	4.8×10^{-43}	NR	No
56	SI	At5g15650	Reversibly glycosylated polypeptide-2, (RGP-2) alpha-1,4-glucan-protein synthase	5.8	41.4	6.0×10^{-50}	No	No
66	SI	At1g74010	Strictosidine synthase family protein	8.3	34.3	5.4×10^{-34}	Yes	No
39	SI	At1g04820	Tubulin alpha-2/alpha-4 chain (TUA4)	4.7	49.5	1.1×10^{-4}	Yes	No
41	SI	At5g44340	Tubulin beta-4 chain (TUB4)	4.8	50.3	9.0×10^{-18}	Yes	No

Table 1. Continued

Spot number	Extraction method[a)]	At number	Protein similarity	p*I*	kDa	Expect[b)]	Pollen expr.[c)]	Pollen-specific[d)]
Disease and defense (8)								
73	SI	At1g07890[f)]	Ascorbate peroxidase, putative (APX)	5.7	27.8	6.9×10^{-40}	Yes	No
77	SI	At1g07890[f)]	Ascorbate peroxidase, putative (APX)	5.7	27.8	1.5×10^{-33}	Yes	No
100	SDS	At1g20630	Catalase 1	6.9	57.3	7.9×10^{-31}	Yes	No
130	SDS	At1g08830	Copper/zinc superoxide dismutase (CSD1)	5.4	14.8	1.8×10^{-4}	Yes	No
125	SDS	At4g11600[f)]	Glutathione peroxidase, putative	9.4	25.7	5.1×10^{-3}	Yes	No
126	SDS	At4g11600[f)]	Glutathione peroxidase, putative	9.4	25.7	2.9×10^{-5}	Yes	No
87	SI	At4g11600	Glutathione peroxidase, putative	9.4	25.7	1.8×10^{-8}	Yes	No
79	SI	At2g47730	Glutathione S-transferase 6 (GST class phi)	6.1	24.1	9.0×10^{-59}	Yes	No
122	SDS	At3g06050	Peroxiredoxin, alkyl hydroperoxide reductase, thiol-specific antioxidant	9.0	21.3	2.3×10^{-25}	Yes	No
86	SI	At3g06050	Peroxiredoxin, alkyl hydroperoxide reductase, thiol-specific antioxidant	9.0	21.3	6.9×10^{-29}	Yes	No
79	SI	At3g10920	Putative (Mn) superoxide dismutase	8.5	25.4	4.3×10^{-26}	Yes	No
102	SDS	At3g02360	6-Phosphogluconate dehydrogenase family protein	7.0	53.8	1.8×10^{-26}	Yes	No
Energy (23)								
85	SI	At3g52300	ATP synthase D chain-related	5.1	19.6	6.7×10^{-28}	Yes	No
82	SI	At5g13450	ATP synthase delta chain subunit, putative/H(+)-transporting two-sector ATPase	9.1	26.3	7.9×10^{-23}	Yes	No
88	SI	At5g47030	ATP synthase delta chain, mitochondrial precursor	6.2	21.5	2.7×10^{-4}	Yes	No
74	SI	At2g21870	ATP synthase fad subunit	9.0	25.1	1.2×10^{-47}	Yes	No
25	SI	At1g78900	ATPase 70 kDa subunit-related	5.1	69.1	1.2×10^{-18}	Yes	No
98	SDS	At2g36530[f)]	Enolase (2-phospho-D-glycerate hydroylase)	5.5	48.0	7.8×10^{-16}	Yes	No
99	SDS	At2g36530[f)]	Enolase (2-phospho-D-glycerate hydroylase)	5.5	48.0	1.6×10^{-5}	Yes	No
104	SDS	At2g36530[f)]	Enolase (2-phospho-D-glycerate hydroylase)	5.5	48.0	2.9×10^{-60}	Yes	No
6	SS	At2g36530	Enolase (2-phospho-D-glycerate hydroylase)	5.5	48.0	1.5×10^{-84}	Yes	No
112	SDS	At2g05990	Enoyl-ACP reductase, chloroplast, putative	9.2	41.5	3.0×10^{-13}	Yes	No
81	SI	At1g47420	Ferripyochelin binding protein	6.2	28.1	4.0×10^{-23}	Yes	No
57	SI	At3g52930	Fructose-bisphosphate aldolase, putative	6.0	38.8	3.2×10^{-55}	Yes	No
108	SDS	At5g03690	Fructose-bisphosphate aldolase-like protein	8.1	38.9	5.2×10^{-61}	Yes	No
102	SDS	At2g47510	Fumarate hydratase, putative/fumarase, putative	7.6	53.5	4.0×10^{-13}	No	No
109	SDS	At3g04120	Glyceraldehyde-3-phosphate dehydrogenase C subunit (GapC)	6.6	37.1	5.0×10^{-12}	Yes	No
58	SI	At3g04120	Glyceraldehyde-3-phosphate dehydrogenase C subunit (GapC)	6.3	37.1	2.9×10^{-56}	Yes	No

Table 1. Continued

Spot number	Extraction method[a)]	At number	Protein similarity	p*I*	kDa	Expect[b)]	Pollen expr.[c)]	Pollen-specific[d)]
15	SS	At3g04120	Glyceraldehyde-3-phosphate dehydrogenase C subunit (GapC)	6.3	37.1	1.1×10^{-20}	Yes	No
12	SS	At1g13440	Glyceraldehyde-3-phosphate dehydrogenase-related	6.7	37.0	1.5×10^{-52}	Yes	No
31	SI	At1g20260	Hypothetical protein, vacuolar ATP synthase subunit B, putative	5.1	54.4	1.7×10^{-26}	NR	No
8	SS	At1g65930	Isocitrate dehydrogenase [$NADP^+$], putative	6.1	46.0	2.1×10^{-76}	Yes	No
43	SI	At1g65930	Isocitrate dehydrogenase [$NADP^+$], putative	6.1	46.0	1.7×10^{-26}	Yes	No
112	SDS	At3g47520	Malate dehydrogenase (EC 1.1.1.37), chloroplast	8.5	42.6	4.0×10^{-22}	No	No
111	SDS	At1g04410	Malate dehydrogenase, cytosolic, putative	6.1	35.9	5.9×10^{-36}	Yes	No
64	SI	At5g43330	Malate dehydrogenase, cytosolic, putative	6.3	36.0	4.7×10^{-37}	Yes	No
97	SDS	At5g25880	Malate oxidoreductase, putative	6.5	65.0	2.2×10^{-43}	Yes	Yes
20	SI	At5g37510	NADH dehydrogenase (ubiquinone), mitochondrial, putative	6.2	82.2	9.1×10^{-56}	Yes	No
60	SI	At2g20360	NADH-ubiquinone oxireductase-related	9.3	44.0	1.7×10^{-22}	Yes	No
55	SI	At2g47470	Protein disulfide isomerase family	5.8	39.8	1.2×10^{-41}	Yes	No
111	SDS	At2g17420	Putative thioredoxin reductase	5.4	31.5	3.7×10^{-24}	No	No
Intracellular traffic (5)								
90	SI	At1g10630	ADP-ribosylation factor 1	6.4	20.6	1.6×10^{-25}	Yes	No
36	SI	At3g59920	Rab GDP dissociation inhibitor (GDI2)	5.4	49.9	1.4×10^{-9}	Yes	No
80	SI	At5g47200	Ras-related GTP-binding protein, putative, RAB1C	5.3	22.5	5.0×10^{-29}	Yes	No
80	SI	At4g17530	Ras-related GTP-binding protein, putative, very strong similarity to RAB1C	5.9	24.8	2.8×10^{-30}	Yes	No
19	SI	At5g17920[f)]	5-Methyltetrahydropteroyltriglutamate-homocysteine *S*-methyltransferase	6.1	84.6	8.1×10^{-88}	NR	No
48	SI	At5g17920[f)]	5-Methyltetrahydropteroyltriglutamate-homocysteine *S*-methyltransferase	6.1	84.6	7.6×10^{-19}	NR	No
Metabolism (27)								
42	SI	At2g33150[f)]	Acetyl-CoA C-acyltransferase (3-ketoacyl-CoA thiolase), putative	8.6	49.0	5.3×10^{-42}	Yes	No
45	SI	At2g33150[f)]	Acetyl-CoA C-acyltransferase (3-ketoacyl-CoA thiolase), putative	8.6	49.0	2.2×10^{-17}	Yes	No
110	SDS	At4g25900	Aldose 1-epimerase family protein	9.6	36.1	3.5×10^{-5}	Yes	No
64	SI	At1g26320	Allyl alcohol dehydrogenase, putative	5.5	39.0	2.8×10^{-12}	Yes	Yes
50	SI	At5g19550	Aspartate aminotransferase	6.8	44.5	8.4×10^{-14}	Yes	No
47	SI	At5g11520	Aspartate aminotransferase, chloroplast (transaminase A/Asp3)	9.3	49.1	5.2×10^{-11}	Yes	No
113	SDS	At5g04180	Carbonic anhydrase family protein	7.1	31.5	1.8×10^{-15}	Yes	No
116	SDS	At1g23730	Carbonic anhydrase, putative/carbonate dehydratase, putative	6.5	29.2	7.9×10^{-41}	Yes	No

Table 1. Continued

Spot number	Extraction method[a)]	At number	Protein similarity	p*I*	kDa	Expect[b)]	Pollen expr.[c)]	Pollen-specific[d)]
43	SI	At2g44350	Citrate synthase	6.4	52.9	2.8×10^{-28}	Yes	No
111	SDS	At2g45600	Esterase, lipase-putative	6.0	36.8	5.0×10^{-21}	Yes	No
34	SI	At1g80460	Glycerol kinase-related	5.9	56.9	6.5×10^{-22}	Yes	No
100	SDS	At5g26780	Glycine8.9	59.5		7.0×10^{-12}	Yes	No
124	SDS	At1g67290	Glyoxal oxidase-related	9.0	68.1	9.1×10^{-4}	Yes	No
22	SI	At1g67290	Glyoxal oxidase-related	9.0	68.1	9.5×10^{-31}	Yes	No
102	SDS	At4g13930	Hydroxymethyltransferase	7.3	51.7	1.5×10^{-36}	No	No
7	SS	At4g13930	Hydroxymethyltransferase	7.3	51.7	6.7×10^{-62}	No	No
75	SI	At4g01480	Inorganic phosphatase-related, putative pyrophosphatase	6.1	24.8	4.6×10^{-3}	Yes	No
19	SI	At3g03780	Methionine synthase-related	6.1	84.9	3.4×10^{-72}	Yes	No
1	SS	At3g03780	Methionine synthase-related	6.1	84.9	4.2×10^{-105}	Yes	No
112	SDS	At1g63000	NAD-dependent5.7	33.8		9.4×10^{-25}	Yes	No
114	SDS	At2g47650	NAD-dependent epimerase/ dehydratase family protein	8.9	50.1	5.9×10^{-29}	No	No
128	SDS	At4g11010	Nucleoside diphosphate kinase 3 (ndpk3)	9.3	25.8	4.5×10^{-14}	Yes	No
91	SI	At4g11010	Nucleoside diphosphate kinase 3 (ndpk3)	9.3	25.8	1.9×10^{-13}	Yes	No
63	SI	At4g10260	PfkB type carbohydrate kinase protein family	6.9	35.0	1.5×10^{-37}	Yes	No
52	SI	At2g17630	Phosphoserine aminotransferase-related	8.3	47.0	7.4×10^{-5}	Yes	No
27	SI	At1g09780	Putative 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	5.2	60.7	7.8×10^{-22}	Yes	No
14	SS	At1g11840	Putative lactoylglutathione lyase	5.0	32.0	1.2×10^{-51}	Yes	No
34	SI	At3g03250[f)]	Putative UDP-glucose pyrophosphorylase	5.8	51.9	5.0×10^{-82}	Yes	No
35	SI	At3g03250[f)]	Putative UDP-glucose pyrophosphorylase	5.9	51.8	4.6×10^{-12}	Yes	No
37	SI	At3g03250[f)]	Putative UDP-glucose pyrophosphorylase	5.9	51.8	2.1×10^{-11}	Yes	No
38	SI	At3g03250[f)]	Putative UDP-glucose pyrophosphorylase	5.9	51.8	3.3×10^{-5}	Yes	No
46	SI	At2g36880[f)]	*S*-adenosylmethionine synthetase-related	5.8	42.9	9.8×10^{-57}	Yes	No
47	SI	At2g36880[f)]	*S*-adenosylmethionine synthetase-related	5.8	42.9	2.3×10^{-31}	Yes	No
48	SI	At2g36880[f)]	*S*-adenosylmethionine synthetase-related	5.8	42.9	5.4×10^{-7}	Yes	No
94	SI	At5g42890	Sterol carrier protein 2 (SCP-2) family protein	9.2	13.5	9.1×10^{-13}	Yes	No
120	SDS	At5g26667	Uridylate kinase, uridine monophosphate kinase (PYR6)	5.8	22.6	2.0×10^{-30}	NR	No
99	SDS	At5g57655	Xylose isomerase family protein	5.6	54.0	5.7×10^{-65}	Yes	No
36	SI	At5g57655	Xylose isomerase family protein	5.6	54.0	1.4×10^{-27}	Yes	No

Table 1. Continued

Spot number	Extraction method[a)]	At number	Protein similarity	p*I*	kDa	Expect[b)]	Pollen expr.[c)]	Pollen-specific[d)]
Protein destination and storage (12)								
28	SI	At1g21750	Protein disulfide isomerase, putative	4.8	55.8	5.7×10^{-47}	Yes	No
103	SDS	At1g21750	Protein disulfide isomerase, putative	4.8	55.8	1.4×10^{-50}	Yes	No
83	SI	At2g16600	Cytosolic cyclophilin (ROC3)	8.7	18.7	6.6×10^{-13}	Yes	No
2	SS	At5g02500	DnaK-type molecular chaperone hsc70.1	5.0	71.7	1.5×10^{-110}	Yes	No
61	SI	At1g07750	Globulin-like protein	5.8	38.5	5.1×10^{-18}	Yes	No
70	SI	At3g09440	Heat shock cognate 70 kDa protein 3 (HSC70–3)	4.9	71.5	1.8×10^{-6}	Yes	No
105	SDS	At5g42020	Luminal binding protein (BiP)	5.1	73.7	1.8×10^{-11}	NR	No
21	SI	At5g42020	Luminal binding protein (BiP)	5.1	73.7	1.5×10^{-80}	NR	No
22	SI	At5g28540	Luminal binding protein 1 precursor (BiP-1) (AtBP1)	5.1	73.9	1.6×10^{-8}	NR	No
83	SI	At4g04200	Microsomal signal peptidase 25 kDa subunit	6.7	13.9	3.5×10^{-5}	NR	No
83	SI	At2g39960	Microsomal signal peptidase 25 kDa subunit, putative	9.0	21.7	3.5×10^{-5}	No	No
123	SDS	At4g38740	Peptidyl-prolyl *cis-trans* isomerase (ROC1)	7.7	18.6	1.6×10^{-9}	Yes	No
84	SI	At2g21130	Peptidylprolyl isomerase (EC 5.2.1.8) similar to CYP2	8.3	18.6	2.5×10^{-10}	Yes	No
132	SDS	At4g05320	Polyubiquitin UBQ10, putative	9.1	28.1	1.6×10^{-9}	Yes	No
Protein synthesis (13)								
29	SI	At1g07920	Elongation factor 1-alpha	9.1	108.1	1.1×10^{-29}	NR	No
115	SDS	At1g07930	Elongation factor 1-alpha	9.2	49.7	2.4×10^{-20}	NR	No
101	SDS	At5g60390	Elongation factor 1-alpha	9.1	108.1	1.7×10^{-20}	NR	No
48	SI	At4g02930	Elongation factor Tu (EF-Tu), putative	6.2	49.6	4.3×10^{-35}	Yes	No
124	SDS	At5g45775	Ribosomal protein L11 (RPL11D), 60S	9.9	19.9	6.1×10^{-15}	NR	No
124	SDS	At2g37190	Ribosomal protein L12 (RPL12A), 60S	9.1	18.0	8.9×10^{-15}	Yes	No
129	SDS	At2g20450	Ribosomal protein L14 (RPL14A), 60S	10.1	15.5	1.9×10^{-19}	Yes	No
121	SDS	At1g33140	Ribosomal protein L9, putative	9.4	22.1	8.8×10^{-26}	No	No
115	SDS	At3g09200	Ribosomal protein P0, 60S acidic, putative	4.9	34.1	8.6×10^{-12}	Yes	No
127	SDS	At2g36160	Ribosomal protein S14 (RPS14A), 40S	10.6	16.3	1.3×10^{-17}	Yes	No
131	SDS	At1g07770	Ribosomal protein S15A (RPS15aA), 40S	10.5	14.8	2.2×10^{-7}	Yes	No
119	SDS	At2g41840	Ribosomal protein S2 (RPS2C), 40S	10.3	31.1	1.1×10^{-29}	Yes	No
71	SI	At3g05500	Rubber elongation factor (REF) family protein	7.6	27.3	2.8×10^{-25}	Yes	No
Signal transduction (9)								
117	SDS	At5g63400	Adenylate kinase	6.9	27.1	2.3×10^{-28}	Yes	No
26	SI	At5g61790	Calnexin 1 (CNX1)	4.8	60.8	2.3×10^{-59}	Yes	No
18	SI	At3g09840	Cell division cycle protein 48 (CDC48A)	5.1	90.1	1.1×10^{-5}	Yes	No
99	SDS	At5g09550	GDP dissociation inhibitor	5.1	50.0	5.0×10^{-3}	Yes	No
86	SI	At1g56330	GTP-binding protein SAR1B	6.5	22.0	5.2×10^{-7}	Yes	No
111	SDS	At5g16760	Inositol 1,3,4-trisphosphate 5/6-kinase	5.9	36.4	2.4×10^{-9}	Yes	No

Table 1. Continued

Spot number	Extraction method[a)]	At number	Protein similarity	p*I*	kDa	Expect[b)]	Pollen expr.[c)]	Pollen-specific[d)]
69	SI	At4g20670	Pollen coat receptor kinase, putative	9.2	30.8	2.0×10^{-6}	NR	No
122	SDS	At4g02080	SAR1A/GTP-binding secretory factor	7.7	22.0	6.4×10^{-3}	Yes	No
82	SI	At4g02080	SAR1A/GTP-binding secretory factor	7.0	22.0	7.3×10^{-7}	Yes	No
62	SI	At1g18080	WD-40 repeat protein, AtArcA gene	7.6	36.1	6.8×10^{-19}	Yes	No
Transport (9)								
32	SI	At5g08670	H^+-transporting ATP synthase beta chain-related	6.2	59.8	9.0×10^{-105}	NR	No
3	SS	At5g08670	H^+-transporting ATP synthase beta chain-related	6.2	59.8	9.3×10^{-94}	NR	No
33	SI	At5g08680	H^+-transporting ATP synthase beta chain-related	6.1	60.0	6.7×10^{-22}	NR	No
89	SI	At4g16160	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/ Tim23 family protein	7.0	18.6	3.5×10^{-16}	Yes	No
72	SI	At5g20230	Plastocyanin-like domain containing protein, blue copper-binding protein, GPI anchored protein	4.7	20.2	6.0×10^{-12}	Yes	No
65	SI	At5g67500[f)]	Porin-related protein	8.9	29.6	4.5×10^{-53}	Yes	No
67	SI	At5g67500[f)]	Porin-related protein	8.9	29.6	5.1×10^{-17}	Yes	No
67	SI	At3g01280	Porin-related protein, similar to voltage-dependent anion-selective channel protein, VDAC	8.8	29.4	1.4×10^{-37}	Yes	No
131	SDS	At4g32470	Ubiquinol-cytochrome C reductase complex 14 kDa protein, putative	9.7	14.5	3.2×10^{-9}	Yes	No
118	SDS	At3g08560	Vacuolar ATP synthase subunit E, putative	9.2	27.0	7.2×10^{-9}	Yes	Yes
128	SDS	At5g15090	Voltage-dependent anion-selective channel protein hsr2 (high sugar response)	7.8	29.2	2.4×10^{-4}	Yes	No
68	SI	At5g15090[f)]	Voltage-dependent anion-selective channel protein hsr2 (high sugar response)	7.8	29.2	1.6×10^{-54}	Yes	No
69	SI	At5g15090[f)]	Voltage-dependent anion-selective channel protein hsr2 (high sugar response)	7.8	29.2	1.1×10^{-17}	Yes	No
Unknown function (9)								
130	SDS	At1g29250	Expressed protein, predicted membrane bound	5.3	14.5	4.1×10^{-10}	Yes	No
106	SDS	At1g30580	Expressed protein, predicted GTP-binding protein, contains DUF933	6.4	44.7	5.6×10^{-3}	Yes	No
129	SDS	At3g01250[f)]	Expressed protein, predicted secreted protein	9.4	18.6	4.9×10^{-6}	Yes	Yes
131	SDS	At3g01250[f)]	Expressed protein, predicted secreted protein	9.4	18.6	2.1×10^{-8}	Yes	Yes
24	SI	At3g28830	Expressed protein, predicted to be secreted	9.7	55.4	2.6×10^{-27}	Yes	No
129	SDS	At3g28980	Expressed protein, predicted to be secreted	8.9	49.1	9.7×10^{-9}	Yes	No

Table 1. Continued

Spot number	Extraction method[a)]	At number	Protein similarity	p*I*	kDa	Expect[b)]	Pollen expr.[c)]	Pollen-specific[d)]
49	SI	At3g28980	Expressed protein, predicted to be secreted	8.9	49.1	7.0×10^{-11}	Yes	No
119	SDS	At3g49720	Expressed protein, predicted mitochondria-targeted protein	9.4	28.8	7.8×10^{-4}	Yes	No
107	SDS	At5g61720	Expressed protein, predicted to be secreted, predicted ATP-binding protein	7.5	43.4	5.7×10^{-35}	Yes	Yes
52	SI	At5g61720[f)]	Expressed protein, predicted to be secreted, predicted ATP-binding protein	7.5	43.4	1.4×10^{-15}	Yes	Yes
53	SI	At5g61720[f)]	Expressed protein, predicted to be secreted, predicted ATP-binding protein	7.5	43.4	3.2×10^{-64}	Yes	Yes
10	SS	At5g61720	Expressed protein, predicted to be secreted, predicted ATP-binding protein	7.5	43.4	1.1×10^{-68}	Yes	Yes
78	SI	At5g62750	Expressed protein, predicted nucleotide binding protein	8.5	14.2	4.5×10^{-5}	Yes	Yes
119	SDS	At5g65810	Expressed protein, predicted mitochondria-targeted protein	8.9	24.0	5.7×10^{-4}	Yes	No

a) SS, salt-soluble fraction; SI, salt-insoluble fraction; SDS, SDS fraction.
b) Expect–expectation value, the number of matches expected if the matches were completely random (see Section 2 for details).
c) Pollen expressed as determined by expression on Affymetrix ATH1 microarray.
d) Pollen specificity determined by expression on Affymetrix ATH1 microarray and EST representation.
e) Number in parenthesis is the number of unique proteins in each functional category.
f) These proteins were identified from multiple protein spots on the same gel.

agree with the observation that the pollen transcriptome exhibits preferential expression of genes encoding cell wall and cytoskeleton-related proteins [20]. Because the role of the pollen grain is to grow a pollen tube and deliver the sperm cells to the embryo sac, it is reasonable that the pollen grain stores proteins needed for early pollen tube growth and proteins for energy generation [25]. We identified several ribosomal proteins (At2g36160, At1g07770, At2g41840, At1g33140, At5g45775, At2g37190, At2g20450, At3g09200) that are likely required for further protein synthesis throughout germination and pollen tube growth. Pollen tubes grow *via* tip growth, which requires the continuous deposition of new cell wall and plasma membrane at the end of the pollen tube [40]. Cytoskeletal proteins such as actins (At2g37620, At3g18780, At5g59370), profilins (At2g19770, At4g29340), and tubulins (At1g04820, At5g44340) are required for the transport of cell wall components. Enzymes such as polygalacturonases (At1g02790, At3g07850, At3g14060) [41], exopolygalacturonase (At3g14040) [42], pectin esterases (At3g62170, At3g13390) [43], glycosyl hydrolase (At2g16730), and 1,4-β-glucan synthase (cellulose synthase, At3g02230, At5g15650) [44] are also needed for pollen tube growth. In addition, we identified two pollen-specific actin-depolymerizing factors (At4g25590, At5g52360) and one profilin (At2g19770). Actin-depolymerizing factor [8, 45] and profilin [46] regulate the dynamics of the actin cytoskeleton in the growing pollen tube [8], presumably by maintaining an optimal level of actin cycling necessary for pollen tube growth and reorientation.

The main goal of this study was to identify many different types of proteins in mature pollen. However, we were particularly interested in identifying plasma membrane-bound proteins because of their importance in pollen tube growth [2, 3, 47] and presumed importance for fertilization [48]. Although we identified proteins that are membrane-bound, most of those proteins are predicted to be in the mitochondrial membrane, *e.g.*, ATPases, porin-related proteins, and voltage-dependent anion-selective channel proteins. We identified only two proteins that are predicted to be plasma membrane-bound, At1g29250 and At5g20230. At1g29250 is annotated as an expressed protein and is predicted to have one N-terminal transmembrane-spanning domain. A PSI-BLAST search of this protein returned only proteins from *Arabidopsis* and rice. Based on microarray data [23, 24] and EST database searches (results not shown), this protein is not pollen-specific, but it appears to be plant-specific. On the other hand, At5g20230 is a known glycosylphosphatidylinositol-anchored protein (GAP) [49, 50] with

a blue copper-binding domain [51]. Blue copper-binding proteins play roles in both pollen germination [52] and in pollen tube guidance [53], and in general, GAPs are required for pollen germination and pollen tube growth in *Arabidopsis* [50].

3.1 Comparing protein and mRNA levels

One limitation to using 2-DE gels for a proteomics study is that only higher-abundance proteins are identified. We were interested in determining whether the more abundant proteins we identified also had higher mRNA levels in the transcriptome. To address this question, we quantitated the protein spots in all three fractions and compared the normalized protein volumes to their corresponding mRNA signals in the transcriptome [23, 24]. We included 77 proteins in this analysis. We excluded those proteins that were identified from spots with multiple identifications and those whose corresponding mRNAs were absent from the mature pollen transcriptome.

In order to make comparisons with protein levels, we similarly represented mRNA values as a percentage of the total, using the approximately 6600 genes that are represented in the mature pollen transcriptome [23, 24]. Figure 3 compares the normalized protein volumes to mRNA signal for 77 proteins. In general, we found an inverse relationship: mRNAs that were abundant corresponded to lower-abundance proteins while mRNAs that were less abundant corresponded to some of the more abundant proteins (Fig. 3 and Table 2). The abundant proteins that had an inverse relationship with their corresponding mRNAs were primarily involved in energy generation (At1g13440, At2g21870, At2g36530, At2g47470, At3g52930), while the low-abundance proteins whose mRNAs were abundant were mainly involved in cell structure (At1g74010, At2g16730, At3g13390, At4g24640, At4g 25590, At5g44340). It seems reasonable that the mature pollen grain stores proteins it will need for the immediate generation of energy to initiate germination and pollen tube growth and that mRNAs encoding proteins necessary for cell structure are ready for immediate translation upon germination of the pollen grain.

There were, however, exceptions to the inverse relationship: some abundant proteins had corresponding mRNAs that were also high, and some low-abundance proteins had corresponding mRNAs that were also low. To determine if there were a trend within this category, we compared the 25 most abundant proteins with the 25 most abundant mRNAs (of those that correspond to proteins we identified) and found that 11 of the 25 mRNAs represented genes important for cell structure, whereas 10 of the 25 proteins are involved in energy production and metabolism (Fig. S1 and Table S2). We found only eight genes whose proteins and mRNAs both fell into the most abundant categories (Table S2); all (At1g04820, At2g36680, At3g03780, At3g04120, At3g05930, At3g28830, At3g62170, At4g29340, and At5g61720) are involved in cell structure or metabolism. Among the 25 least abundant proteins and 25 least abundant mRNAs there were 10 genes (At1g21750, At1g30580, At1g65930, At2g20360, At2g36530, At3g09840, At4g16160, At5g37510, At5g42890, and At5g63400) whose proteins and mRNAs both fell into the least abundant categories (Fig. S3 and Table S3). Most of these proteins are necessary for energy generation and signal transduction. Interestingly, the low-abundance proteins and mRNAs contained relatively few (three) proteins or mRNAs (one) necessary for cell structure. Overall, these comparisons suggest that the mature pollen grain puts the majority of its resources into making proteins and mRNAs that it will need immediately upon germination.

Not all of the proteins we identified are encoded by mRNAs represented in the pollen transcriptome. We identified nine proteins (Table 3) whose mRNAs are missing from the transcriptome [23, 24]. EST representation (NCBI, http://www.ncbi.nlm.nih.gov/) of these proteins indicates that none are pollen-specific. For the proteins that are predicted to be secreted (At2g17420, At2g39960, and At3g02230), one explanation for their presence might be that they were deposited on the pollen grain by the tapetum, the sporophytic tissue of the anther that surrounds the male gametophytes during their early development. Alternately, for any of these nine proteins, the mRNA encoding them might be short-lived or be of extremely low-abundance and thus appear absent from the transcriptome.

Of the nine proteins whose mRNAs are absent from the transcriptome, seven are important in cell structure, energy generation, or metabolism, further supporting the idea that the mature pollen grain contains the proteins necessary for germination and rapid pollen tube growth [25]. One secreted protein important for cell structure, At3g07850, is a putative polygalacturonase. Polygalacturonases are responsible for the hydrolysis of 1,4-α-D-galactosiduronic linkages in pectate and other galacturonans. Polygalacturonases are present inside the pollen grain and are important for pollen development and pollen tube growth [54–56]. However, some polygalacturonases are deposited on the outside of the pollen grain. During *Arabidopsis* pollen development, a tetrad of microspores is formed after meiosis II. In wild-type plants, the pollen mother cell wall around the tetrad is degraded and the four microspores separate [57]. In *Arabidopsis* plants with mutations in the QUARTET genes, the dissolution of the wall does not occur and the pollen grains remain as a quartet. One of the QUARTET genes, *QRT3 (At4g20050)*, is transiently expressed in the tapetum of *Arabidopsis* during microspore development and encodes an endopolygalacturonase with polygalacturonase activity [58]. Therefore, it is possible that At3g07850 or At3g14060, another secreted protein, are deposited on the pollen grain by the tapetum, as is QUARTET 3.

3.2 Comparisons of our results with the results of other studies

In a previous study of the *Arabidopsis* pollen coat proteome, the 12 proteins identified were mainly oleosins and lipases [17]. Of these proteins, only one, At4g20670, was identified in

Table 2. Comparison of normalized protein volumes to mRNA levels as a percentage of the pollen transcriptome

Gene number	Function	Identity	Normalized protein volume	mRNA levels as% of pollen transcriptome	Fraction	Spot number
At5g04180	Metabolism	Carbonic anhydrase family protein	0.3280	0.3076	SDS	113
At1g02790	Cell structure	Polygalacturonase, putative	0.2820	0.2625	SI	44
At3g13390	Cell structure	Pectinesterase (pectin methylesterase) family	0.1800	0.2492	SI	23
At3g28830	Unknown	Expressed protein, predicted to be secreted	1.3700	0.2464	SI	24
At3g62170	Cell structure	Pollen-specific pectin esterase, putative	0.4150	0.2452	SS	11
At3g62170	Cell structure	Pollen-specific pectin esterase, putative	0.3550	0.2452	SI	54
At3g05930	Disease and defense	Germin-like protein, GLP8, possible oxalate oxidase activity	0.3820	0.2329	SI	76
At4g24640	Cell structure	Invertase/pectin methylesterase inhibitor family	0.1980	0.2272	SS	16
At5g61720	Unknown	Expressed protein, predicted to be secreted, predicted ATP-binding protein	0.2850	0.2176	SS	10
At5g61720	Unknown	Expressed protein, predicted to be secreted, predicted ATP-binding protein	0.7360	0.2176	SI	53
At5g61720	Unknown	Expressed protein, predicted to be secreted, predicted ATP-binding protein	0.3270	0.2176	SDS	107
At5g44340	Cell structure	Tubulin beta-4 chain (TUB4)	0.2400	0.1912	SI	41
At2g16730	Cell structure	Glycosyl hydrolase family 35 protein	0.2350	0.1752	SI	17
At2g16730	Cell structure	Glycosyl hydrolase family 35 protein	0.1420	0.1752	SDS	96
At5g20230	Transport	Plastocyanin-like domain containing protein, blue copper-binding protein, GPI anchored protein	0.3610	0.1649	SI	72
At3g28980	Unknown	Expressed protein, predicted to be secreted	0.3290	0.1622	SI	49
At1g74010	Cell structure	Strictosidine synthase family protein	0.0930	0.1370	SI	66
At4g05320	Protein destination and storage	Polyubiquitin UBQ10, putative	0.1020	0.1158	SDS	132
At2g36880	Metabolism	*S*-adenosylmethionine synthetase-related	2.1740	0.0921	SI	46
At5g52360	Cell structure	Actin-depolymerizing factor, putative	0.2860	0.0846	SI	93
At4g25590	Cell structure	Actin-depolymerizing factor-related protein	0.2190	0.0789	SI	92
At4g29340	Cell growth and division	Profilin 3	1.0280	0.0787	SI	95
At3g05500	Protein synthesis	Rubber elongation factor (REF) family protein	0.0550	0.0675	SI	71
At3g03780	Metabolism	Methionine synthase-related	3.6840	0.0536	SS	1
At5g03690	Energy	Fructose-bisphosphate aldolase-like protein	0.1270	0.0531	SDS	108
At5g62750	Unknown	Expressed protein, predicted nucleotide binding protein	0.1370	0.0521	SI	78
At1g04820	Cell structure	Tubulin alpha-2/alpha-4 chain (TUA4)	0.4040	0.0518	SI	39
At2g21130	Protein destination and storage	Peptidylprolyl isomerase (EC 5.2.1.8) similar to CYP2	0.1160	0.0324	SI	84
At3g04120	Energy	Glyceraldehyde-3-phosphate dehydrogenase C subunit (GapC)	0.3300	0.0299	SS	15
At3g04120	Energy	Glyceraldehyde-3-phosphate dehydrogenase C subunit (GapC)	0.9750	0.0299	SI	58
At3g04120	Energy	Glyceraldehyde-3-phosphate dehydrogenase C subunit (GapC)	0.0850	0.0299	SDS	109

Table 2. Continued

Gene number	Function	Identity	Normalized protein volume	mRNA levels as% of pollen transcriptome	Fraction	Spot number
At2g33150	Metabolism	Acetyl-CoA C-acyltransferase (3-ketoacyl-CoA thiolase), putative	0.1860	0.0265	SI	42
At3g03250	Metabolism	Putative UDP-glucose pyrophosphorylase	0.9390	0.0256	SI	35
At3g03250	Metabolism	Putative UDP-glucose pyrophosphorylase	0.2570	0.0256	SI	37
At3g03250	Metabolism	Putative UDP-glucose pyrophosphorylase	0.3340	0.0256	SI	38
At4g01480	Metabolism	Inorganic phosphatase-related, putative pyrophosphatase	0.3960	0.0217	SI	75
At1g07890	Disease and defense	Ascorbate peroxidase, putative (APX)	0.8570	0.0214	SI	73
At1g07890	Disease and defense	Ascorbate peroxidase, putative (APX)	0.2390	0.0214	SI	77
At4g11600	Disease and defense	Glutathione peroxidase, putative	0.2570	0.0190	SI	87
At4g11600	Disease and defense	Glutathione peroxidase, putative	0.0730	0.0190	SDS	126
At4g38740	Protein destination and storage	Peptidyl-prolyl *cis-trans* isomerase (ROC1)	0.0250	0.0186	SDS	123
At3g08560	Transport	Vacuolar ATP synthase subunit E, putative	0.2150	0.0164	SDS	118
At1g13440	Energy	Glyceraldehyde-3-phosphate dehydrogenase-related	0.3310	0.0151	SS	12
At3g52300	Energy	ATP synthase D chain-related	0.3170	0.0150	SI	85
At5g25880	Energy	Malate oxidoreductase, putative	0.0230	0.0150	SDS	97
At5g47030	Energy	ATP synthase delta chain, mitochondrial precursor	0.2450	0.0137	SI	88
At2g21870	Energy	ATP synthase fad subunit	0.4380	0.0133	SI	74
At1g11840	Metabolism	Putative lactoylglutathione lyase	0.2430	0.0133	SS	14
At3g52930	Energy	Fructose-bisphosphate aldolase, putative	0.7370	0.0130	SI	57
At1g65930	Energy	Isocitrate dehydrogenase [$NADP^+$], putative	0.1240	0.0127	SS	8
At5g02500	Protein destination and storage	DnaK-type molecular chaperone hsc70.1	1.0790	0.0122	SS	2
At2g36530	Energy	Enolase (2-phospho-D-glycerate hydroylase)	0.1080	0.0117	SS	6
At2g36530	Energy	Enolase (2-phospho-D-glycerate hydroylase)	0.1290	0.0117	SDS	98
At2g36530	Energy	Enolase (2-phospho-D-glycerate hydroylase)	0.3620	0.0117	SDS	104
At1g78900	Energy	ATPase 70 kDa subunit –related	0.2690	0.0115	SI	25
At1g07750	Protein destination and storage	Globulin-like protein	0.9970	0.0111	SI	61
At4g10260	Metabolism	PfkB type carbohydrate kinase protein family	0.6700	0.0107	SI	63
At3g09840	Signal transduction	Cell division cycle protein 48 (CDC48A)	0.1730	0.0090	SI	18
At2g20360	Energy	NADH-ubiquinone oxireductase-related	0.1400	0.0086	SI	60
At5g61790	Signal transduction	Calnexin 1 (CNX1)	0.6850	0.0084	SI	26
At1g21750	Energy	Protein disulfide isomerase, putative	0.9140	0.0078	SI	28
At1g21750	Protein synthesis	Protein disulfide isomerase, putative	0.0310	0.0078	SDS	103
At3g18780	Cell structure	Actin 12	0.8070	0.0078	SI	51
At5g15090	Transport	Voltage-dependent anion-selective channel protein hsr2 (high sugar response)	0.6260	0.0075	SI	68
At1g10630	Intracellular traffic	ADP-ribosylation factor 1	0.5060	0.0069	SI	90
At5g37510	Energy	NADH dehydrogenase (ubiquinone), mitochondrial, putative	0.0840	0.0066	SI	20
At1g18080	Signal transduction	WD-40 repeat protein, AtArcA gene	0.3120	0.0065	SI	62

Table 2. Continued

Gene number	Function	Identity	Normalized protein volume	mRNA levels as% of pollen transcriptome	Fraction	Spot number
At2g47470	Energy	Protein disulfide isomerase family	0.6240	0.0059	SI	55
At5g67500	Transport	Porin-related protein	0.5040	0.0058	SI	65
At3g10920	Disease and defense	Putative (Mn) superoxide dismutase	0.9680	0.0052	SI	79
At5g42890	Metabolism	Sterol carrier protein 2 (SCP-2) family protein	0.2020	0.0051	SI	94
At4g16160	Transport	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	0.0960	0.0048	SI	89
At5g63400	Signal transduction	Adenylate kinase	0.1480	0.0037	SDS	117
At3g09440	Protein destination and storage	Heat shock cognate 70 kDa protein 3 (HSC70–3)	0.3500	0.0032	SI	70
At1g09780	Metabolism	Putative 2,3-bisphosphoglycerate-independent phosphoglycerate mutase	0.2580	0.0025	SI	27
At1g30580	Unknown	Expressed protein, predicted GTP-binding protein, contains DUF933	0.0670	0.0021	SDS	106
At5g19550	Metabolism	Aspartate aminotransferase	0.2650	0.0019	SI	50

Table 3. Identified proteins with no detectable RNA in the transcriptome

Spot number	Extraction method[a)]	At number	Protein similarity	Theoretical p*I*	Theoretical kDa	Expect value
Cell structure (3)						
9	SS	At2g37620	Actin 1	5.3	42.0	3.1×10^{-48}
4	SS	At3g07850	Polygalacturonase	8.6	46.3	2.1×10^{-47}
30	SI	At3g07850	Polygalacturonase	8.6	46.3	5.5×10^{-8}
56	SI	At5g15650	Reversibly glycosylated polypeptide-2 (RGP-2), alpha-1,4-glucan-protein synthase	5.8	41.4	6.0×10^{-50}
Energy (2)						
102	SDS	At2g47510	Fumarate hydratase, putative	7.6	53.5	4.0×10^{-13}
111	SDS	At2g17420	Putative thioredoxin reductase	5.4	31.5	3.7×10^{-24}
Metabolism (2)						
114	SDS	At2g47650	NAD-dependent epimerase/ dehydratase family protein	8.9	50.1	5.9×10^{-29}
7	SI	At4g13930	Hydroxymethyltransferase	7.3	51.7	1.5×10^{-36}
102	SDS	At4g13930	Hydroxymethyltransferase	7.3	51.7	6.7×10^{-62}
Protein destination and storage (1)						
83	SI	At2g39960	Microsomal signal peptidase 25 kDa subunit, putative	9.0	21.7	3.5×10^{-5}
Protein synthesis (1)						
121	SDS	At1g33140	Ribosomal protein L9, putative	9.4	22.1	8.8×10^{-26}

a) SS, salt-soluble fraction; SI, salt-insoluble fraction; SDS, SDS fraction.

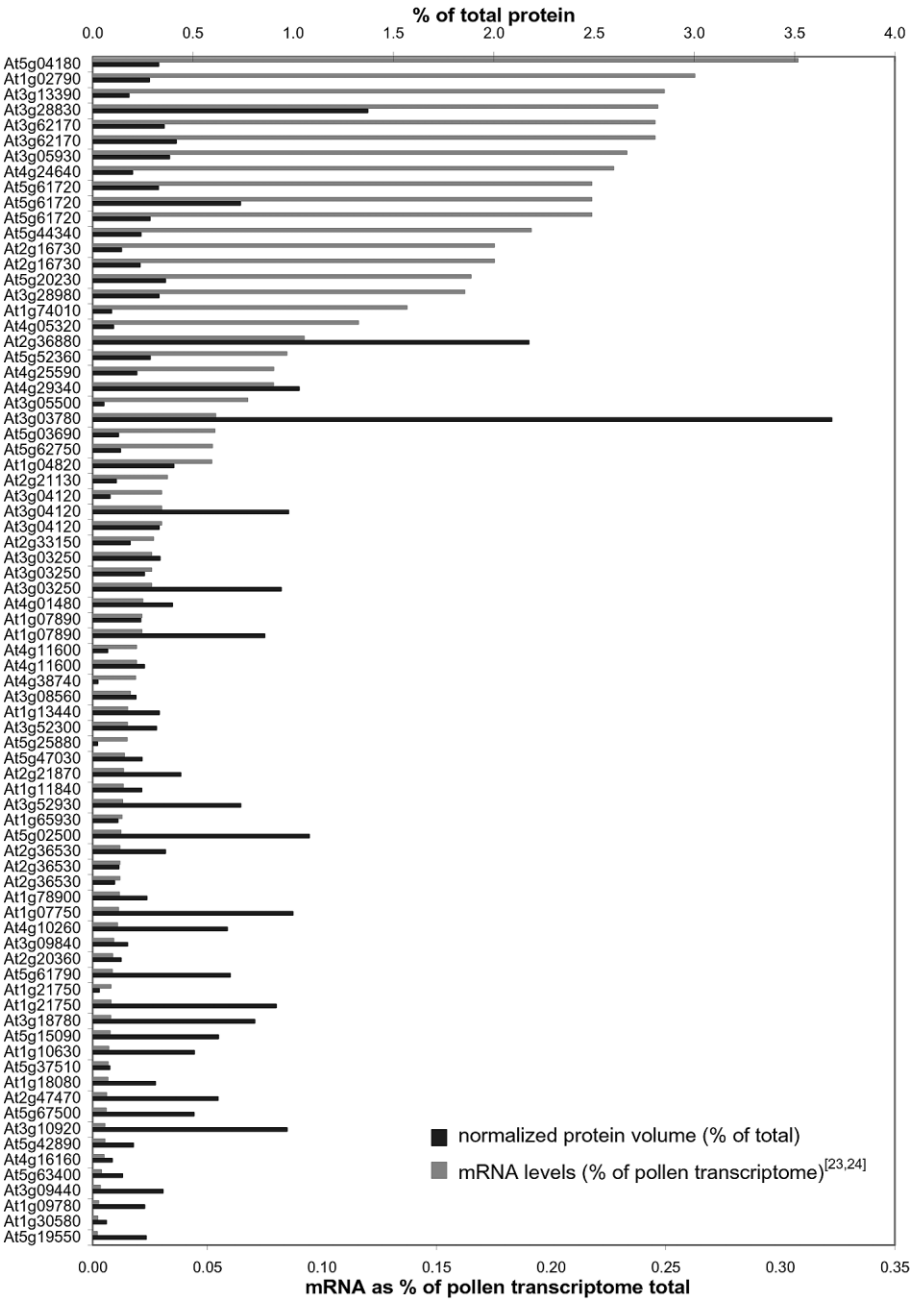


Figure 3. Comparison of normalized protein volumes and mRNA expression levels. Both protein volumes and mRNA expression levels are represented as a percentage of the total. mRNA signal information is based on the AtGenExpress data for the mature pollen proteome [23, 24]. Proteins that were identified from spots with multiple protein identifications were excluded from the analysis, as were proteins whose corresponding mRNAs are not represented in the mature pollen transcriptome.

our analysis. This protein was annotated in the earlier study [17] as a putative pollen coat receptor kinase; however, At4g20670 is predicted to be a secreted protein and does not have a kinase domain, although it is similar to the extracellular domains of some receptor kinases. The differences in protein composition between our study and the pollen coat proteome [17] are likely due to the solvents that were used to extract the proteins. Mayfield *et al.* [17] extracted the proteins using cyclohexane, a solvent known to strip proteins off the pollen coat. Because we used salt and detergents to extract the proteins from the entire pollen grain, it is not surprising that we identified very different proteins. In a second proteomics study that identified rice anther proteins during different stages of anther development, including the heading stage, the stage in which mature tricellular pollen grains are encased within the anther, Imin *et al.* [14] extracted proteins using a solution of acetone containing TCA and DTT. This extraction method is also different from our salt and detergent methods; however, TCA, like SDS, extracts a wide variety of proteins with different solubilities. Several proteins identified from heading stage anthers are important in sugar metabolism (vacuolar acid invertase, fructokinase),

or in cell elongation and cell expansion (β-expansin, profilin, and actin); cellular functions that are essential for pollen tube growth and two of which also were identified in our study.

3.3 Pollen-specific proteins

Most of the proteins identified as pollen-specific are necessary for cell structure: glycosyl hydrolase family protein (At2g16730), germin-like protein (GLP8, putative oxalate oxidase, At3g05930), pectin methylesterase inhibitor family protein (At4g24640), actin 4 (At5g59370), and actin-depolymerizing factors (At4g25590, At5g52360). In addition to pollen-specific cell structure proteins, proteins important for cell growth and division (profilin 3, At4g29340), metabolism (putative allyl alcohol dehydrogenase, At1g26320), energy (putative malate oxidoreductase, At5g25880), and transport (putative vacuolar ATP synthase subunit E, At3g08560) are also represented. Of the proteins we identified, 6% (9 of 139; Table 1, Table S1) are annotated only as expressed proteins. For these, PSI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) searches in the public databases returned no significant matches to proteins of known function. EST representation and microarray experiments [23, 24] indicate that six of the nine proteins are not pollen-specific. The remaining three proteins (At3g01250, At5g61720, and At5g62750) were predicted to be either flower-specific or pollen-specific. We performed RT-PCR to verify the expression pattern of these three genes. Figure S3 shows that At3g01250 and At5g61720 are pollen-specific, and that At5g62750 is expressed in pollen and pistils.

To learn more about these three proteins, we performed a variety of *in silico* analyses. At5g62750 (SI spot 78) is a 124 amino acid lysine-rich protein predicted to be localized to the nucleus. Unlike the other protein identifications (Table 1, Table S1) where multiple peptides matched database sequences, we identified this protein from a single 26 amino acid peptide. However, when we performed BLAST searches with this peptide, we found that it was unique in the predicted proteome; thus, we believe the identity is accurate. At5g62750 is 37% identical to a 121 amino acid protein of unknown function from rice (accession number BAC9937). This rice protein is annotated as a KED-like protein, most likely because 52% of its amino acids are lysine (K), glutamic acid (E), and aspartic acid (D). KED proteins [59] represent a family of wound-response proteins about which nothing else is known. Alternatively, because At5g62750 has many lysines, it may be a histone-like protein.

At3g01250 (SDS, spots 128 and 131) is a 164 amino acid protein and is predicted to be secreted. The level of the At3g01250 mRNA in pollen was extremely low. At3g01250 was identified from two distinct spots on the gel, suggesting that it is post-translationally modified. Indeed, several predicted phosphorylation sites are present in this protein as are potential myristoylation and glycosylation sites.

At5g61720 was identified in all three fractions (SS spot 10; SI spots 52 and 53; SDS spot 107). This 384 amino acid protein is predicted to be secreted and it shares approximately 30% sequence identity to At3g28830 (SI spot 24) and 21% sequence identity to At3g28980 (SI spot 49, SDS spot 128), both of which are present in pollen (Table 1, Table S1), but neither of which is pollen-specific (data not shown). Interestingly, all three of these proteins have significant similarity to a protein encoded by another gene on chromosome 5 (At5g39870) and to proteins of unknown function (At3g28870, At3g28780, At3g28790, At3g28810, At3g28820, and At3g28840) that are encoded by a gene cluster on chromosome 3. We used MEME 3.0 (http://meme.sdsc.edu/meme/website/meme.html) to determine whether any highly conserved regions exist in these proteins. Figure 4 shows two highly conserved motifs: the first begins immediately after the predicted signal sequence at residues 31–38. The second motif is found in seven of the nine proteins and is located about 250 residues C-terminal of the first motif. Both motifs are similar to each other with regard to conserved residues and to the spacing between the conserved residues. For example, both motifs contain the consensus [LF]-[IVL]-X(2)-L-E-X(11)-FF-X(2)-[LF]-[KRE]-X(2)-[MAL]. However, each motif has different conserved residues N-terminal to this consensus. The first motif has the consensus [P]-X(4)-[DE]-X(6)-[KDR]-X(4) directly N-terminal to the shared consensus, whereas the second motif has the consensus E-X(6)-[VM]-X(2). PHI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) was used to identify other proteins containing the shared consensus sequence. By searching with less stringent variations of the consensus sequence, P-X(4)-D-X(6)-K-X(4)-F-[ML]-X(2)-L-E-X(3)-P-X(5)-K-X-[FL]-F, we found two translated ESTs from *Brassica* anthers (GenBank accession numbers CD839918 and CD841398) that were more similar to At5g61720 than they were to each other. The alignment of the proteins deduced from the translation of these ESTs with At5g61720 is shown in Fig. 5. We found no matches in other species, suggesting that this protein family is specific to *Arabidopsis* and *Brassica*.

3.4 Proteins identified multiple times

Of the 179 identified proteins, 31 were identified multiple times and 13 of these proteins were identified from multiple spots on the same gel (Table 1, Table S1). Proteins can migrate as multiple spots on 2-DE gels for several reasons. Naturally occurring protein isoforms can result from *in vivo* PTMs such as phosphorylation, glycosylation, or acetylation. These modifications do not significantly change the molecular weight of a protein but they will likely cause a shift in p*I* in either the acidic or basic direction. For example, At2g36880, an *S*-adenosylmethionine synthetase-related protein, migrates as a string of three spots (SI spots 46, 47, and 48). This spot pattern could be the result of phosphorylation, as *S*-adenosylmethionine synthetase is believed to be phosphorylated [60]. The same might be true for At5g15090,

Motif 1

Gene	start site residue	p value[b]	motif sequence
At3g28810	38	$4.10e^{-56}$	PPQTLKDVEPYTVKVVIVFVSDLEKECPKTNKFKAFFEKLRAFAKYVCPI
At3g28820	38	$8.37e^{-55}$	FPQTLKDVEPYTVKVVMVFVSDLEKECPKTNKFKTFFEKLRAVAKYVCPI
At3g28830[a]	36	$1.18e^{-37}$	PPKKCNEFTPYANKGMITLVTDLEGSSPATTEFKTFFTQFKSYMSFIETT
At3g28840	35	$8.63e^{-45}$	PPRRSRDFRPFACRGMLKFVDVLELKCPLKPQYKSFFGNLRSYMNFINSA
At3g28770	31	$5.65e^{-36}$	PPNTIKDIEPYISDRALGFVLKLENNCPIREQLRSFFEKLKDLLKLESSV
At3g28780[c]	35	$3.83e^{-37}$	FPKNSQDFEPFAYKGMLSFVDNLENMAPEKGEYKDFFSKLKAFMSFINTA
At3g28790	36	$9.71e^{-45}$	FPRRCIDFAPYAGKGMLMLVSNLEGGCPATREFKQFFSTFKSYMSFISSA
At3g28980[a]	38	$3.39e^{-52}$	DPKTVKDVEPYTVKVVMVFVADLEKECPKTNKFKAFFEKLRGFAKYVCPI
At5g39870	34	$9.56e^{-45}$	QPLTVKDVESYTIKVVTTFLVELEKECPKTEKFKVFFEKLKAYSKYLCPV
At5g61720[a]	35	$4.31e^{-42}$	FPKLGKDYEAFANKGISDFLGDLEGMCPKTAEFKDFFENLKDYMAFFNSA

Motif 2

Gene	start site residue	p value[b]	motif sequence
At3g28810	269	$1.02e^{-58}$	TVTQVEEETSKDVSTFIMNLEKKCPQKEEYKVFFEQLKGTMIAPPKER
At3g28820	269	$1.02e^{-58}$	TVTQVEEETSKDVSTFIMNLEKKCPQKEEYKVFFEQLKGTMIAPPKER
At3g28830[a]	365	$1.49e^{-37}$	TVKQVESETSKEVMSFIMQLEKKYAAKAELKVFFESLKSSMQASASVG
At3g28770	1875	$8.25e^{-35}$	TTTDTESNTSKEVTSFISNLEEKSPGTQEFQSFFQKLKDYMKYLCPVS
At3g28790	441	$3.74e^{-34}$	SVKEVETQTSSEVNSFISNLEKKYTGNSELKVFFEKLKTSMSASAKLS
At3g28980[a]	277	$7.34e^{-57}$	TVTQVEEETSKDVSTFIMNLEKKCPQKEEFKVFFEQLKGTMIAPRKER
At5g61720[a]	220	$5.98e^{-31}$	MGKAIEPFAYKGMSDFLGSLESKCPATPEFKDFFVKLEDYMACFKLVS

[a]Proteins identified in this study.
[b]P value was calculated by the MEME program and represents the certainty of the match to the consensus sequence.
[c]GenBank accession number BAD43471.1was used in this alignment. At3g28780 has two in frame ATG sites. GenBank accession number NP_189520.1 differs from BAD43471.1 because the second ATG is used as the translational start site; thus, the first 60 amino acids of the protein are missing.

Figure 4. Conserved motifs of unknown protein family. Alignment of MEME 3.0-generated conserved motifs from three proteins of unknown function (At3g28830, At3g28980, and At5g61720) identified in this study with similar proteins found in the *Arabidopsis* genome. Motif 1 is located at the N-terminus immediately following the predicted signal peptide. Position of Motif 2 is not as conserved as that of Motif 1.

```
At5g61720        MGKISLAICLTLLVTLSTVYETQGTFSLPLYLKNFPKLGKDYEAFANKGISDFLGDLEGM 60
gi|32521858      MAKTSLTIYLSLLVALSTVYETQGTFSLPLYLKNFPKVGHDFESFAYKGMMDFMGDLEGK 60
gi|32523338      ------------------------TFSLPHYLGDFPKMSKDFEPFAYKGMSAFLGALESK 36
                                         ***** ** :***:.:*:*.** **:  *:* **.

At5g61720        CPKTAEFKDFFENLKDYMAFFNSAAPGSKDNQFEMFVKSEKLFKALSAFNSSKGGTSEDS 120
gi|32521858      CPQTTEFKDFFLKLKDYTACYSSTAPGSKDLQVELSIKSETLFRAMSDFSGTKGGTSEDS 120
gi|32523338      CPATAEFKDLFVKVVDYMACFKS------GIKIEMQEKSVKLFRAISVLDGTNGETSVDS 90
                 ** *:****:* :: ** * :.*      . :.*:  ** .**:*:* :..::* ** **

At5g61720        WKLVDGLLSMGKGLVEMKKSGSKEITFEERRDLISSMVKWARAIGLFVKAASEDKGQSID 180
gi|32521858      WTLVDGLLSMGKSLVEMKKSGSKEITFEQRKETIQSMVKWTRGIGLFVKKVSESKGKSID 180
gi|32523338      WRMVDGMLSMGILVTEMKKNVSQEITFEQRKELIGAMVKWARAIGLLVKTASEKKGQSID 150
                 * :***:****  :.****. *:*****:*:: * :****:*.***:** .**.**:***

At5g61720        LASFGIDYDNNVESPFSKRAMYETQGTFSLPHYVKNTPKMGKAIEPFAYKGMSDFLGSLE 240
gi|32521858      LSSFGIDYDNXVSSP-SERALYETXGTFSLPHYV------------------------- 213
gi|32523338      LASFGVDYSPHVASP-IKGANGETIK---------------------------------- 175
                 *:***:**.  * **  : *  **
```

Figure 5. Alignment of At5g61720 with two translated ESTs from *Brassica*. Two ESTs from *Brassica* were identified that contain Motif 1 of the protein family. Highlighted residues designate Motif 1. Only the first 240 amino acids of At5g61720 are shown.

the voltage-dependent anion-selective channel protein HSR2. It appears as two spots with the same molecular weight but with different p*I*s. This protein is known to be present in plants in multiple isoforms that are differentially phosphorylated [61].

Proteins that were present in multiple spots that differed considerably in molecular weight and/or p*I* were also found. The large differences in molecular weight could result from proteins being translated from alternatively spliced mRNAs. At1g07890, a putative ascorbate peroxidase, was identified from two spots that varied in M_r and p*I*. It is possible that the two proteins are translated from different mRNA splice variants, as is known to occur with ascorbate peroxidase in other plants [62, 63]. Alternatively, differences in M_r and p*I* could result from proteolytic cleavage during protein extraction or sample handling, or from normal degradation of the protein.

4 Concluding remarks

We used different protein extraction techniques, 2-DE, and ESI-MS/MS to generate a partial proteome reference map of mature pollen of *Arabidopsis*. By establishing this map, we

have contributed to the understanding of the biochemistry of pollen. As techniques for *Arabidopsis* pollen germination *in vitro* improve [13], comparative proteomic studies between mature pollen and germinated or mutant pollen will become possible. With the ability to isolate relatively large quantities of *Arabidopsis* pollen, proteomic analyses of sperm [64] should be feasible.

Supplementary Material

Table S1. Proteins identified from mature *Arabidopsis* pollen. This table is identical to Table 1 but it can be downloaded as a Microsoft Word file that is searchable and has links to the GenBank protein sequences.
Table S2. The 25 most abundant proteins and the 25 most highly expressed mRNAs.
Table S3. The 25 least abundant proteins and the 25 lowest expressed mRNAs.

Figure S1. Comparison of the functional classes of the 25 most abundant proteins and the 25 most highly expressed mRNAs that correspond to proteins identified.
Figure S2. Comparison of the functional classes of the 25 least abundant proteins and the 25 least expressed mRNAs that correspond to proteins identified.
Figure S3. RT-PCR of flower-specific proteins of unknown function. Relative mRNA abundance levels corresponding to three of the unknown proteins identified in this study. RT-PCR is shown for the six tissue types tested. For At3g01250, a three-fold more PCR product was loaded onto the gel.

We thank Sheila Johnson-Brousseau, Heping Yang, and Michelle Engel for comments on this manuscript. This work was supported by the US Department of Agriculture under Grant number CRIS 5335–21000–011–00D.

5 References

[1] Cheung, A. Y., May, B., Kawata, E. E., Gu, Q., Wu, H. M., *Plant J.* 1993, *3*, 151–160.

[2] Muschietti, J., Eyal, Y., McCormick, S., *Plant Cell* 1998, *10*, 319–330.

[3] Tang, W., Ezcurra, I., Muschietti, J., McCormick, S., *Plant Cell* 2002, *14*, 2277–2287.

[4] Tang, W., Kelley, D., Ezcurra, I., Cotter, R., McCormick, S., *Plant J.* 2004, *39*, 343–353.

[5] Mousavi, A., Hiratsuka, R., Takase, H., Hiratsuka, K., Hotta, Y., *Plant Cell Physiol.* 1999, *40*, 406–416.

[6] Pertl, H., Himly, M., Gehwolf, R., Kriechbaumer, R. *et al.*, *Planta* 2001, *213*, 132–141.

[7] Allwood, E. G., Anthony, R. G., Smertenko, A. P., Reichelt, S. *et al.*, *Plant Cell* 2002, *14*, 2915–2927.

[8] Chen, C. Y., Wong, E. I., Vidali, L., Estavillo, A. *et al.*, *Plant Cell* 2002, *14*, 2175–2190.

[9] Cheung, A. Y., Chen, C. Y., Glaven, R. H., de Graaf, B. H. *et al.*, *Plant Cell* 2002, *14*, 945–962.

[10] Muschietti, J., Dircks, L., Vancanneyt, G., McCormick, S., *Plant J.* 1994, *6*, 321–338.

[11] Obermeyer, G., Klaushofer, H., Nagl, M., Hoeftberger, M., Bentrup, F.-W., *Planta* 1998, *207*, 303–312.

[12] The Arabidopsis Genome Initiative, *Nature* 2000, *408*, 796–815.

[13] Johnson-Brousseau, S. A., McCormick, S., *Plant J.* 2004, *39*, 761–775.

[14] Imin, N., Kerim, T., Weinman, J. J., Rolfe, B. G., *Proteomics* 2001, *1*, 1149–1161.

[15] Kerim, T., Imin, N., Weinman, J. J., Rolfe, B. G., *Proteomics* 2003, *3*, 738–751.

[16] Imin, N., Kerim, T., Rolfe, B. G., Weinman, J. J., *Proteomics* 2004, *4*, 1873–1882.

[17] Mayfield, J. A., Fiebig, A., Johnstone, S. E., Preuss, D., *Science* 2001, *292*, 2482–2485.

[18] Preuss, D., Lemieux, B., Yen, G., Davis, R. W., *Genes Dev.* 1993, *7*, 974–985.

[19] Mayfield, J. A., Preuss, D., *Nat. Cell Biol.* 2000, *2*, 128–130.

[20] Honys, D., Twell, D., *Plant Physiol.* 2003, *132*, 640–652.

[21] Becker, J. D., Boavida, L. C., Carneiro, J., Haury, M., Feijo, J. A., *Plant Physiol.* 2003, *133*, 713–725.

[22] Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J. *et al.*, *Science* 2003, *301*, 653–657.

[23] *NASCArrays Experiment: AtGenExpress: Developmental Series (flowers and pollen)*, date accessed 7–9-2004, http://128.243.111.177/narrays/experimentpage.pl?experimentid=152

[24] Craigon, D. J., James, N., Okyere, J., Higgins, J. *et al.*, *Nucleic Acids Res.* 2004, *32 Database issue*, D575–D577.

[25] Mascarenhas, J. P., *Bot. Rev.* 1975, *41*, 259–314.

[26] Curie, C., McCormick, S., *Plant Cell* 1997, *9*, 2025–2036.

[27] de Dios Alche, J., Castro, A. J., Olmedilla, A., Fernandez, M. C. *et al.*, *J. Cell Sci.* 1999, *112(Pt 15)*, 2501–2509.

[28] Huang, J. C., Lin, S. M., Wang, C. S., *Plant Cell Physiol.* 2000, *41*, 477–485.

[29] Wengier, D., Valsecchi, I., Cabanas, M. L., Tang, W. H. *et al.*, *Proc. Natl. Acad. Sci. USA* 2003, *100*, 6860–6865.

[30] Honys, D., Combe, J. P., Twell, D., Capkova, V., *Sex Plant. Reprod.* 2000, *13*, 135–144.

[31] Giavalisco, P., Nordhoff, E., Lehrach, H., Gobom, J., Klose, J., *Electrophoresis* 2003, *24*, 207–216.

[32] Mechin, V., Consoli, L., Le Guilloux, M., Damerval, C., *Proteomics* 2003, *3*, 1299–1302.

[33] Kasarda, D. D., Woodard, K. M., Adalsteins, A. E., *Cereal Chem.* 1998, *75*, 70–71.

[34] Vensel, W. H., Harden, L., Tanaka, C. K., Hurkman, W. J., Haddon, W. F., *J. Biomol. Technol.* 2002, *13*, 95–100.

[35] Fenyo, D., Beavis, R. C., *Anal. Chem.* 2003, *75*, 768–774.

[36] Hurkman, W. J., Tanaka, C. K., DuPont, F. M., *Plant Physiol.* 1988, *88*, 1263–1273.

[37] Santoni, V., Rabilloud, T., Doumas, P., Rouquie, D. *et al.*, *Electrophoresis* 1999, *20*, 705–711.

[38] Luche, S., Santoni, V., Rabilloud, T., *Proteomics* 2003, *3*, 249–253.

[39] Santoni, V., Kieffer, S., Desclaux, D., Masson, F., Rabilloud, T., *Electrophoresis* 2000, *21*, 3329–3344.

[40] Yang, Z., *Curr. Opin. Plant Biol.* 1998, *1*, 525–530.

[41] Hadfield, K. A., Bennett, A. B., *Plant Physiol.* 1998, *117*, 337–343.

[42] Brown, S. M., Crouch, M. L., *Plant Cell* 1990, *2*, 263–274.

[43] Li, Y. Q., Mareck, A., Faleri, C., Moscatelli, A. *et al.*, *Planta* 2002, *214*, 734–740.

[44] Ferguson, C., Teeri, T. T., Siika-aho, M., Read, S. M., Bacic, A., *Plant J.* 1998, *206*, 452–460.

[45] Chen, C. Y., Cheung, A. Y., Wu, H. M., *Plant Cell* 2003, *15*, 237–249.

[46] Vidali, L., McKenna, S. T., Hepler, P. K., *Mol. Biol. Cell* 2001, *12*, 2534–2545.

[47] Scholz-Starke, J., Buttner, M., Sauer, N., *Plant Physiol.* 2003, *131*, 70–77.

[48] Talbot, P., Shur, B. D., Myles, D. G., *Biol. Reprod.* 2003, *68*, 1–9.

[49] Borner, G. H., Lilley, K. S., Stevens, T. J., Dupree, P., *Plant Physiol.* 2003, *132*, 568–577.

[50] Lalanne, E., Honys, D., Johnson, A., Borner, G. H. *et al.*, *Plant Cell* 2004, *16*, 229–240.

[51] Van Gysel, A., Van Montagu, M., Inze, D., *Gene* 1993, *136*, 79–85.

[52] Weterings, K., Reijnen, W., van Aarssen, R., Kortstee, A. *et al.*, *Plant Mol. Biol.* 1992, *18*, 1101–1111.

[53] Kim, S., Mollet, J. C., Dong, J., Zhang, K. *et al.*, *Proc. Natl. Acad. Sci. USA* 2003, *100*, 16125–16130.

[54] Allen, R. L., Lonsdale, D. M., *Plant Mol. Biol.* 1992, *20*, 343–345.

[55] Allen, R. L., Lonsdale, D. M., *Plant J.* 1993, *3*, 261–271.

[56] Tebbutt, S. J., Rogers, H. J., Lonsdale, D. M., *Plant Mol. Biol.* 1994, *25*, 283–297.

[57] McCormick, S., *Plant Cell* 2004, *16(Suppl)*, S142–S153.

[58] Rhee, S. Y., Osborne, E., Poindexter, P. D., Somerville, C. R., *Plant Physiol.* 2003, *133*, 1170–1180.

[59] Hara, K., Yagi, M., Koizumi, N., Kusano, T., Sano, H., *Plant Cell Physiol.* 2000, *41*, 684–691.

[60] Pajares, M. A., Corrales, F., Duran, C., Mato, J. M., Alvarez, L., *FEBS Lett.* 1992, *309*, 1–4.

[61] Al Bitar, F., Roosens, N., Smeyers, M., Vauterin, M. *et al.*, *Biochim. Biophys. Acta* 2003, *1625*, 43–51.

[62] Yoshimura, K., Yabuta, Y., Ishikawa, T., Shigeoka, S., *J. Biol. Chem.* 2002, *277*, 40623–40632.

[63] Yoshimura, K., Yabuta, Y., Tamoi, M., Ishikawa, T., Shigeoka, S., *Biochem. J.* 1999, *338(Pt 1)*, 41–48.

[64] Engel, M. L., Chaboud, A., Dumas, C., McCormick, S., *Plant J.* 2003, *34*, 697–707.

