Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochemical and Biophysical Research Communications 337 (2005) 1257–1266
BBRC
www.elsevier.com/locate/ybbrc

A reference map of the *Arabidopsis thaliana* mature pollen proteome ☆

Sandra Noir, Anne Bräutigam, Thomas Colby, Jürgen Schmidt, Ralph Panstruga *

Max-Planck-Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, D-50829 Köln, Germany

Received 15 September 2005
Available online 7 October 2005

Abstract

The male gametophyte (or pollen) plays an obligatory role during sexual reproduction of higher plants. The extremely reduced complexity of this organ renders pollen a valuable experimental system for studying fundamental aspects of plant biology such as cell fate determination, cell–cell interactions, cell polarity, and tip-growth. Here, we present the first reference map of the mature pollen proteome of the dicotyledonous model plant species, *Arabidopsis thaliana*. Based on two-dimensional gel electrophoresis, matrix-assisted laser desorption/ionization time-of-flight, and electrospray quadrupole time-of-flight mass spectrometry, we reproducibly identified 121 different proteins in 145 individual spots. The presence, subcellular localization, and functional classification of the identified proteins are discussed in relation to the pollen transcriptome and the full protein complement encoded by the nuclear *Arabidopsis* genome.
© 2005 Elsevier Inc. All rights reserved.

Keywords: *Arabidopsis thaliana*; Male gametophyte; Mass spectrometry; Mature pollen; Proteome; Two-dimensional gel electrophoresis

In higher plants, development of the male gametophyte is a well-programmed and elaborate process. Male sporogenesis begins with the division of a diploid sporophytic cell giving rise to the anther wall of the stamen and the sporogenic cells. The latter cells then undergo several mitoses to differentiate into pollen mother cells. Subsequently, each diploid pollen mother cell forms a tetrad of haploid microspores via a round of DNA replication followed by two consecutive meiotic divisions. Each uninucleate microspore then undergoes an asymmetric mitotic division forming a large vegetative cell and a smaller generative cell (i.e., the bicellular pollen stage). In *Arabidopsis*, the generative cell undergoes another mitotic division giving rise to two sperm cells (i.e., the tricellular pollen stage). Following pollination, the vegetative cell controls the further development of the mature pollen grain and growth of the pollen tube into the style until both sperm cell nuclei are delivered to the embryo sac in the ovule, where they participate in double fertilization [1,2].

The last two decades have been marked by increasing efforts to decipher the genetic and molecular basis of pollen development and functions [reviewed in 1,3,4]. In the model plant *Arabidopsis thaliana*, the extremely reduced, tricellular male gametophyte constitutes an ideal experimental system for analyses of important biological processes in higher plant reproduction. In addition, it represents a very useful model for studying fundamental aspects of plant biology such as cell fate determination, cell–cell interactions, cell polarity, and tip-growth [5–7].

The availability of the full genome sequence of *Arabidopsis* [8] has made genome-wide, microarray-based analyses of the male gametophyte transcriptome of this model plant possible [9–12]. 13,977 male gametophyte-expressed mRNAs were recently identified using the ATH1 Genome Array, which covers ~81% of the currently estimated 28,000 protein-coding genes in *Arabidopsis*. It seems that 9.7% of these transcripts are specific for the male gametophyte [11]. The majority of the respective genes appear to be expressed during the two earliest developmental stages

☆ *Abbreviations:* 2-DE, two-dimensional gel electrophoresis; Chaps, 3-(3-cholamidopropyl)dimethylammonio-1-propane sulfonate; DAPI, 4′,6-diamino-phenylindole; FDA, fluorescein diacetate; IEF, isoelectric focusing; LC–MS/MS, liquid chromatography coupled to tandem mass spectrometry; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; MS, mass spectrometry; p*I*, isoelectric point; PMF, peptide mass fingerprint.

* Corresponding author. Fax: +49 221 5062 353.
E-mail address: panstrug@mpiz-koeln.mpg.de (R. Panstruga).

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.09.185

of sporogenesis (i.e., in microspores and bicellular pollen). Indeed, 8788 transcripts were found to be expressed in tricellular pollen [11] and either 7235 [11] or 6587 [12] in mature pollen. Compared to the uninucleate microspore stage, male-gametophyte-specific genes identified from the mature pollen grain stage exhibit an increase in the collective proportion of cell wall-, cytoskeleton-, signaling-, and transport-related genes [11]. This rise is thought to reflect the functional specialization of mature pollen in preparation for a dramatic change in its cell growth pattern during pollen germination and pollen tube growth.

Although global gene expression analysis is useful for selecting candidates for functional studies, the presence or absence of a given transcript does not necessarily have predictive power for the presence or absence of the protein it encodes. For example, minute levels of a transcript may be sufficient for adequate accumulation of the respective protein, while post-transcriptional and/or post-translational regulation might negatively affect accumulation of particular proteins encoded by well-represented transcripts. Thus, in addition to transcriptome analyses, it is equally important to determine the protein complement of a given tissue. To this end, a two-dimensional gel electrophoresis (2-DE)-based proteomic approach has already been applied to study the differential protein synthesis and protein phosphorylation patterns of *Brassica napus* microspores during microspore embryogenesis [13] and to compare pollen coat protein profiles of self-incompatible and self-fertile *Festuca pratensis* [14]. In addition, to develop information about the proteins present at different stages of pollen development, 2-DE and mass spectrometry have been used to analyze the protein complement of rice anthers [15,16]. In this study, the authors were able to identify ~150 protein spots that varied among different stages of anther development. So far, the only reported proteome analysis in *A. thaliana* described the constituents of the pollen coat [17]. The work presented here reports the application of proteome analysis techniques to characterize the *Arabidopsis* mature pollen stage. Using 2-DE in combination with MALDI-TOF MS and LC–MS/MS analyses, we generated the first reference map of the *Arabidopsis* mature pollen proteome.

Materials and methods

Plant material and pollen isolation. Plants of *A. thaliana* Columbia (Col-0) ecotype were grown in the greenhouse at 20 °C-day and 18 °C-night and 50% relative humidity, under 8 h light/16 h dark for the 3 first weeks, and then 16 h light/8 h dark regimes until flowering. In addition to wild-type plants, three homozygous double knockout (T-DNA or transposon insertion) lines in genes *AtMlo5* and *AtMlo9* were employed in the experiments. These lines are homozygous F_2 progeny from crosses of insertion lines SLAT_15-21 (*Atmlo5*-2) and GARLIC_348_B04 (*Atmlo9*-2), SALK_118934 (*Atmlo5*-5) and SALK_073198 (*Atmlo9*-3), as well as SLAT_24-28 (*Atmlo5*-1) and GARLIC_348_B04 (*Atmlo9*-2), respectively. Sequence-indexed SALK [18] and GARLIC (now referred to as SAIL) [19] T-DNA insertion lines were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC; http://arabidopsis.info/) and Syngenta Biotechnology, respectively. SLAT transposon insertion lines [20] were selected by PCR using pooled DNA samples and appropriate oligonucleotides. Mature pollen grains from ~400 plants per genotype were harvested using a vacuum system adapted from Johnson-Brousseau and McCormick [21], and stocked at −20 °C until protein extraction (see below).

Pollen purity and viability were assessed using representative samples of the frozen pollen batches. The purity of isolated pollen was determined by light microscopy and epifluorescence microscopy after 4′,6-diaminophenylindole (DAPI, Sigma) staining (2.5 μg/ml DAPI in 50 mM PBS, pH 7.2, with 0.01% Tween 20 and 5% DMSO). Samples were observed via epifluorescence 1 h after incubation in DAPI solution at room temperature. Pollen viability was analyzed 2 h after rehydration at 4 °C in a high humidity chamber and staining with fluorescein diacetate (FDA, Fluka). A stock solution of 2 mg/ml FDA was made in acetone and diluted drop-by-drop in 0.5 M of sucrose until the solution remained milky. Samples were observed by epifluorescence 5 min after incubation at room temperature.

Protein extraction. Pollen grains were suspended in 600 μl of extraction buffer (1 protease inhibitor cocktail tablet (Roche); 10 mM DTT; 50 mM, pH 8.0, Tris-Base; 10 mM EDTA; and 0.5% Chaps) and a small amount of glass beads (diameters of 425–600 μm, Sigma) were added. Pollen grains were broken by five repetitions of vortexing (30 s) interrupted by short pauses on ice (30 s each). After centrifugation at 16,000*g* for 5 min at 4 °C, the supernatant was removed and stored on ice. From the remaining pellet, the extraction procedure was repeated and supernatants were pooled. Protein concentration of the supernatant was determined by the Bradford assay with bovine serum albumin as the standard. Supernatants were dried frozen and stored at −80 °C until loading for isoelectric focusing.

Two-dimensional gel electrophoresis. Two-dimensional polyacrylamide gel electrophoresis (2-DE) was performed using the NuPAGE ZOOM Benchtop Proteomics system (Invitrogen). Proteins (100 μg) were solubilized in 165 μl sample rehydration buffer (7 M urea, 2 M thiourea, 2% Chaps, 0.5% ZOOM Carrier Ampholytes, pH 3–10 (Invitrogen), 20 mM DTT, and 0.1% bromophenol blue). Prior to isoelectric focusing (IEF) ZOOM strips, pH 3–10, NL (Invitrogen) were incubated in the rehydration solution containing the sample for 16 h, and IEF was conducted using the following step gradient: 0–175 V (1 min), 175 V (15 min), 175–2000 V (45 min), and 2000 V (25 min). After IEF, the strips used for gel electrophoresis were first equilibrated in 4.5 ml lithium dodecyl sulfate (LDS) sample buffer together with 0.5 ml of 10× sample reducing agent (Invitrogen), and subsequently in the same solution containing 125 mM iodoacetamide without reducing agent (15 min each). Samples were separated in the second dimension on NuPAGE Novex 4–12% Bis–Tris ZOOM gels in Mes–SDS running buffer (Invitrogen), and proteins were stained with colloidal Coomassie using Imperial Protein Stain (Pierce).

In-gel digestion and mass spectrometry. After 2-DE, spots of various intensities were automatically picked (PROTEINEERsp, Bruker) and tryptically digested using the DP Chemical 96 Kits for fully automated in-gel digestion (PROTEINEERdp, Bruker). Aliquots of the digests were automatically prepared (PROTEINEERdp, Bruker) for subsequent MALDI-TOF analysis on AnchorChip targets (Bruker) according to Gobom et al. [22]. Mass spectra of tryptic peptides were taken with a Bruker Reflex IV MALDI-TOF MS. The obtained peptide mass fingerprints (PMFs) were processed in Xmass 5.1.16 (Bruker) and used to identify the corresponding proteins in the ProteinScape 1.2 database system (Protagen AG) which triggered Mascot (Matrix Science) and ProFound (Genomic Solutions) searches. Sample aliquots of spots containing several putative proteins were submitted to in-depth LC–MS/MS analyses.

LC–MS/MS analysis. Residual organic solvent was removed from sample digests via vacuum centrifugation prior to injection into a Waters CapLC coupled to a Q-ToFII mass spectrometer (Micromass/Waters). Samples were loaded onto a Waters Symmetry 300 c18 trapping column for on-column focusing and eluted over a Waters Atlantis c18 analytical column (75 μm × 150 mm) with a 5–50% ACN gradient containing 0.1% formic acid. The elution was analyzed in survey mode under the control of Masslynx4.0 (Waters). Multiply charged peaks of sufficient intensity were automatically fragmented in the argon collision cell and the resulting MS/MS spectra were processed in Masslynx4.0 before submission to Mascot (Matrix Science) for searches against the NCBI non-redundant database.

Pollen transcriptome analysis. The pollen transcriptome analysis has been performed using Affymetrix GeneChip technology. Two independent

experiments have been carried out using Col-0 wild-type plants and the three *Atmlo5/Atmlo9* double mutant lines. Mature pollen grains from ~400 plants each per genotype were harvested using the above-mentioned vacuum system. Total RNAs were extracted using the RNeasy Plant Mini Kit (Qiagen) and used for *Arabidospsis* GeneChip ATH1 array hybridizations (University of Cologne—Germany) under standard conditions [23]. MicroArray Suite 5.0 software was used for data analysis.

Database searching. The PMF data were used to screen the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein). When AGI names could not be identified, data were completed by using the TAIR BLAST 2.2.8 tool. The TAIR database (http://www.arabidopsis.org/index.jsp) was also used to annotate identified proteins and to assign their predicted subcellular localization. The latter were corroborated by using the pSORT Wolf software for Protein Subcellular Localization Prediction (http://wolfpsort.seq.cbrc.jp/). Potential transmembrane domains were investigated by using the TAIR Bulk Protein Search Tool. Functional categories were assigned according to the Functional Catalogue Database of MIPS (http://mips.gsf.de/projects/funcat). When multiple functional categories were allotted to an individual protein, the most probable (i.e., lowest *p* value) was chosen. *Arabidopsis* organs where identified genes are preferentially expressed were deduced by using the Gene Atlas tool of Genevestigator (https://www.genevestigator.ethz.ch/) [24].

Results and discussion

The results of this study were obtained in the context of a comparative proteome analysis between *A. thaliana* Col-0 wild-type plants and *Atmlo5/Atmlo9* double knockout mutants in the Col-0 genetic background (see Materials and methods for details). Genes *AtMlo5* and *AtMlo9* encode two members of the family of heptahelical MLO membrane proteins [25] that are preferentially expressed during pollen development [Z.Y. Chen and co-workers, personal communication; S. Noir and R. Panstruga, unpublished]. In the context of this report, we take advantage of the fact that no differences between the proteomes of mature pollen from wild-type plants and *Atmlo5/Atmlo9* knockout mutants could be observed and employ the respective dataset for the establishment of the first reference map of the *Arabidopsis* mature pollen proteome.

Mining of pollen source material

Approximately 20 mg of mature pollen each was collected from ~400 plants per genotype (wild-type plus three distinct *Atmlo5/Atmlo9* double mutants). Microscopic examination of pollen populations after DAPI staining confirmed the homogeneity of these samples (i.e., no other developmental stage besides mature tricellular pollen was observed). Little cellular debris was detected and FDA staining revealed that the proportion of viable pollen grains was ~75% (data not shown). From each pollen sample, ~1800–2000 μg of total protein could be extracted.

Establishment of a proteome map of Arabidopsis mature pollen

Aliquots (~100 μg each) of the protein samples were subjected to 2-DE as described in Materials and methods. Analysis of at least two replicate gels of the four distinct *A. thaliana* genotypes revealed no reproducible differences between samples derived from Col-0 wild-type plants and *Atmlo5/Atmlo9* double mutants (data not shown). In total, up to 145 mature pollen protein spots could be reproducibly resolved over a p*I* range of 3–10 on Coomassie blue-stained gels (Fig. 1).

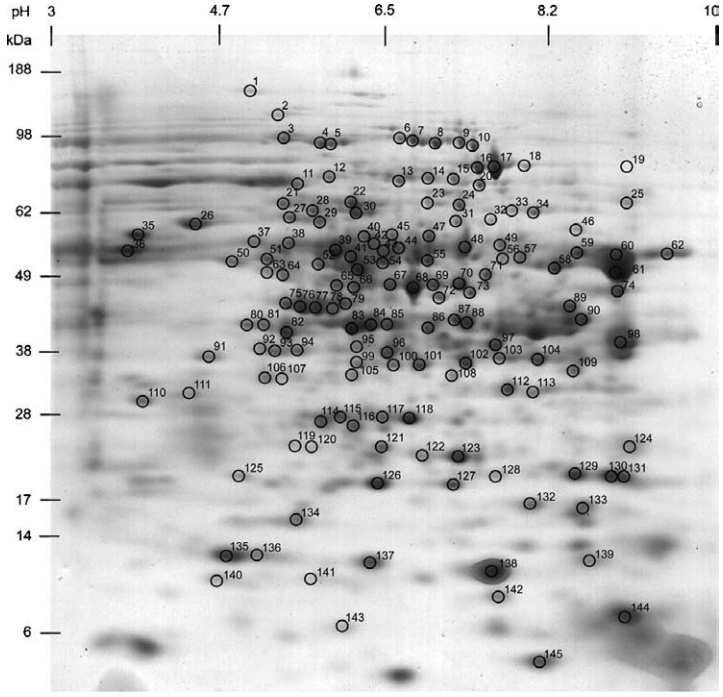

Fig. 1. Coomassie blue-stained two-dimensional gel of a total protein extract (100 μg) from mature pollen of *Arabidopsis*. Molecular masses are given in kDa and approximate isoelectric points are shown. The identification of each labeled spot can be found in Table 1.

Table 1
Results of protein identification in *Arabidopsis* mature pollen

Spot No.[a]	AGI name	Gene index	Protein identity (TAIR description)	MM/p*I*[b]	Sequence coverage[c]	ProFound-score[d]	Mascot-score[e]	Subcellular localization[f]	Functional category[g]	Genevestigator[h]	Micro array detection[i]
1	At1g49490	gi|5430752	Leucine-rich repeat family protein/extensin family protein	90.9/4.8	17.7	2.4	131	vacuole	unc.	1	P
2	At2g41740	gi|22136974	Villin 2	107.8/5.1	24.1	1.5	146	nucleus	cell wall	2	P
3	At3g09840	gi|6681343	Cell division cycle protein 48	89.3/5.0	50.7	2	329	endomb.	prot. proc.	3	P
4	At4g15530	gi|7268300	Pyruvate phosphate dikinase family protein	104.9/5.8	31.3	2.3	273	cytp.	energy	2	P
5	At4g15530	gi|7268300	Pyruvate phosphate dikinase family protein	104.9/5.8	34.5	2.2	302	cytp.	energy	2	P
6	At4g26970	gi|23308183	Putative aconitate hydratase	108.3/7.0	36.0	2.2	293	endomb.	energy	2	P
7	At4g26970	gi|23308183	Putative aconitate hydratase	108.3/7.0	43.4	1.7	307	endomb.	energy	2	P
8	At2g05710	gi|4586021	Aconitate hydratase	98.1/5.7	59.0	2.4	472	cytp.	energy	3	P
9	At4g35830	gi|7270535	Aconitate hydratase	98.1/6.0	45.1	2	305	cytp.	energy	3	P
10	At1g56075	gi|23397162	Putative elongation factor 2	93.8/5.9	53.7	2.4	318	cytp.	unc.	no data	no data
11	At5g09590	gi|7671430	Heat shock protein 70	72.9/5.5	53.7	2.2	230	endomb.	defense	3	undeter.
12	At3g60750	gi|22136900	Putative transketolase	79.9/5.9	39.7	2	228	endomb.	energy	4	undeter.
13	At2g45290	gi|20196914	Putative transketolase	68.8/5.6	43.4	1.8	214	endomb.	energy	2	undeter.
14**	At2g07698	gi|14916970	Putative ATP synthase alpha chain	55.3/6.2	[j]	[j]	94	endomb.	energy	3	A
14**	At3g24170	gi|14532810	Putative glutathione reductase	54.3/6.4	[j]	[j]	64	cytp.	unc.	4	undeter.
15	At5g65690	gi|10177326	Putative phosphoenolpyruvate carboxykinase	68.7/6.0	44.3	1.9	222	unknown	energy	no data	no data
16	At5g17920	gi|20147149	Homocysteine methylase	84.3/6.1	45.4	2.4	295	cytp.	metabolism	no data	no data
17	At5g17920	gi|20147149	Homocysteine methylase	84.3/6.1	65.8	2.4	404	cytp.	metabolism	no data	no data
18	At5g17920	gi|20147149	Homocysteine methylase	84.3/6.1	55.0	2.4	328	cytp.	metabolism	no data	no data
19	At5g17920	gi|20147149	Homocysteine methylase	84.3/6.1	38.2	2.3	214	cytp.	metabolism	no data	no data
20	At1g53500	gi|33090264	NAD-dependent epimerase/dehydratase family protein	75.2/6.0	19.8	2.4	121	unknown	unc.	1	P
21	At1g78900	gi|30725440	Vacuolar ATP synthase catalytic subunit A	68.8/5.0	57.0	2.3	329	cytp.	transport	3	P
22	At5g66760	gi|21700795	Succinate dehydrogenase flavoprotein subunit	69.6/5.8	68.0	2.4	392	endomb.	energy	2	P
23	At1g23190	gi|20148521	Phosphoglucomutase	63.1/5.9	39.8	2.4	210	cytp.	energy	2	P
24	At1g23190	gi|20148521	Phosphoglucomutase	63.1/5.9	56.3	2.2	353	cytp.	energy	2	P
25	At1g67290	gi|28394059	Glyoxal oxidase-related	67.7/9.6	59.5	2.4	354	endomb.	metabolism	1	P
26	At1g21750	gi|17104689	Disulfide isomerase-like protein	55.6/4.7	66.7	2.4	458	endomb.	unc.	3	undeter.
27	At3g23990	gi|34098917	Chaperonin	61.2/5.5	60.0	2.4	344	endomb.	prot. proc.	3	A
28	At3g54440	gi|30523396	Glycoside hydrolase family 2 protein	68.4/5.1	34.3	2.3	282	nucleus	metaC	no data	no data
29	At2g33210	gi|2924773	Putative chaperonin	55.2/5.2	46.9	1.5	174	endomb.	prot. proc.	3	A
30	At3g08590	gi|21280833	Putative 2,3-biphosphoglycerate-independent phosphoglycerate mutase	60.7/5.5	65.0	1.6	286	cytp.	unc.	3	undeter.
31	At3g11830	gi|6671939	Putative chaperonin	59.7/6.0	40.2	2.4	209	cytp.	cell cycle	3	undeter.
32	At3g48990	gi|20799715	AMP-dependent synthetase and ligase family protein	55.5/6.1	18.3	1.5	87	unknown	metabolism	3	P
32	At5g42740	gi|11094242	Glucose-6-phosphate isomerase	61.5/6.2	44.7	2.4	204	unknown	energy	3	A
33	At5g25880	gi|5107826	Putative malate oxidoreductase	64.6/6.6	27.9	2.4	170	cytp.	energy	1	P
34	At5g25880	gi|5107826	Putative malate oxidoreductase	64.6/6.6	59.4	2.4	264	cytp.	energy	1	P
35	At1g56340	gi|30725696	Calreticulin 1	48.5/4.3	35.8	2.4	104	endomb.	cell fate	3	undeter.
36	At1g09210	gi|16974341	Calreticulin 2	48.1/4.2	31.6	2.4	114	endomb.	unc.	2	undeter.
37	At4g38510	gi|7270834	Putative vacuolar ATP synthase subunit B	54.3/4.9	51.1	2.2	210	vacuole	transport	1	P
38	At3g13930	gi|23397124	Putative dihydrolipoamide *S*-acetyltransferase	58.4/8.6	52.1	2.1	258	endomb.	energy	2	undeter.
39	At5g08690	gi|17939849	ATP synthase β-chain 2	63.3/6.6	66.6	2.4	413	endomb.	energy	3	P
40	At5g08690	gi|17939849	ATP synthase β-chain 2	63.3/6.6	27.2	0.9	67	endomb.	energy	3	P
41	At5g08690	gi|17939849	ATP synthase β-chain 2	63.3/6.6	42.1	2.3	145	endomb.	energy	2	P
42	At4g34200	gi|23297595	Putative D-3-phosphoglycerate dehydrogenase	63.3/6.2	44.3	2.3	179	endomb.	energy	3	undeter.
43	At4g13940	gi|20148279	Adenosylhomocysteinase	53.3/5.6	49.7	2.1	256	cytp.	metabolism	3	P
44	At5g17310	gi|14532836	Putative UTP–glucose-1-phosphate uridylyltransferase	51.9/5.6	70.0	2.1	279	endomb.	metaC	2	P

45	At5g20890	gi|21537317	Putative chaperonin	57.2/5.5	68.1	2	263	cytp.	transport	3	undeter.
46	At3g18190	gi|20466822	Putative chaperonin	57.7/8.7	39.0	2.4	188	nucleus	prot. proc.	3	A
47	At3g02090	gi|6513923	Putative peptidase beta-subunit	59.1/6.3	55.9	1.4	212	endomb.	unc.	3	P
48	At2g07698	gi|14916970	Putative ATP synthase alpha-chain	55.0/6.2	45.6	2.4	195	endomb.	energy	3	A
49	At5g65690	gi|10177326	Putative phosphoenolpyruvate carboxykinase	68.7/6.0	44.3	1.9	222	unknown	energy	no data	no data
50	At5g44340	gi|16323374	Tubulin β-4 chain	49.8/4.6	65.3	2.4	318	cytoskel	cell fate	1	P
51	At1g50010	gi|23506129	Tubulin α-2/α-4 chain	49.5/4.8	71.3	2.4	329	cytoskel	cytosk.	2	P
52	At1g51980	gi|20258957	Putative peptidase alpha subunit	54.4/5.9	29.0	1.9	103	endomb.	unc.	3	P
53	At2g36530	gi|23297411	Enolase	47.7/5.5	77.0	2	338	cytp.	metaC	3	undeter.
54	At5g15490	gi|53749198	Putative UDP-glucose 6-dehydrogenase	53.1/5.7	64.6	2.2	272	endomb.	metaC	1	P
55	At5g15490	gi|53749198	Putative UDP-glucose 6-dehydrogenase	53.1/5.7	65.0	1.9	302	endomb.	metaC	1	P
56**	At5g34850	gi|20259673	Calcineurin-like phosphoesterase family protein	55.0/6.9	[j]	[j]	64	endomb.	unc.	2	P
57	At4g13930	gi|20334774	Putative glycine hydroxymethyltransferase	51.7/7.0	32.1	1.9	92	cytp.	metabolism	no data	no data
58	At2g47510	gi|2529676	putative fumarate hydratase	53.0/9.1	46.5	1.3	177	endomb.	energy	no data	no data
58	At4g13930	gi|20334774	Putative glycine hydroxymethyltransferase	51.7/7.0	32.1	1.9	92	cytp.	metabolism	no data	no data
59	At1g20630	gi|2511725	Catalase 1	56.8/7.4	60.4	2.2	355	cytp.	defense	2	P
60	At3g07850	gi|6648197	Exopolygalacturonase	45.6/9.7	65.3	2.4	267	endomb.	unc.	1	P
61	At3g07850	gi|6648197	Exopolygalacturonase	45.6/9.7	66.7	2.4	277	endomb.	unc.	1	P
62	At1g07920	gi|295789	Elongation factor 1-α	49.5/9.8	35.2	2.4	130	cytp.	prot. synth.	no data	no data
63	At5g03630	gi|22655344	Putative monodehydroascorbate reductase	47.5/5.1	50.3	2.4	209	cytp.	metabolism	4	A
64	At5g03630	gi|22655344	Putative monodehydroascorbate reductase	47.5/5.1	55.2	2	199	cytp.	metabolism	4	A
65	At3g13920	gi|14594802	Translation initiation factor 4A-1	41.8/8.8	45.8	1.8	142	unknown	cell fate	1	P
66	At1g02500	gi|81647	*S*-Adenosylmethionine synthetase 1	43.1/5.4	39.7	2.3	146	cytp.	metaC	3	P
67	At2g36880	gi|22137172	Putative *S*-adenosylmethionine synthetase	42.5/5.7	31.0	1.2	140	cytp.	metaC	3	P
67	At4g01850	gi|23308349	*S*-Adenosylmethionine synthetase 2	43.2/5.6	36.4	1.3	142	cytp.	metaC	3	P
68	At2g36880	gi|22137172	Putative *S*-adenosylmethionine synthetase	42.5/5.7	75.4	2.4	382	cytp.	metaC	3	P
69	At1g65930	gi|20453235	Putative isocitrate dehydrogenase	45.7/6.1	41.5	2	161	endomb.	energy	3	P
70	At1g65930	gi|20453235	Putative isocitrate dehydrogenase	45.7/6.1	55.1	2.1	203	endomb.	energy	3	P
71	At2g35840	gi|11127757	Sucrose-phosphatase 1	47.6/6.2	53.3	2.3	206	endomb.	energy	3	undeter.
72	At4g02930	gi|31376381	Putative elongation factor Tu	49.4/6.3	69.6	2.4	341	endomb.	prot. synth.	3	undeter.
73	At5g07440	gi|7576182	Glutamate dehydrogenase 2	44.7/6.1	51.1	2.2	181	endomb.	metabolism	2	P
74	At2g33150	gi|13194830	Putative acetyl-CoA C-acyltransferase	48.5/9.5	71.4	2.1	245	cytp.	energy	3	P
75	At2g37620	gi|71633	Actin 1	41.7/5.2	62.9	2.2	295	cytoskel	cytosk.	1	P
75	At3g53750	gi|21554576	Actin 3	41.8/5.2	62.9	2.2	295	cytoskel	cytosk.	1	P
76	At2g37620	gi|71633	Actin 1	41.7/5.2	61.0	2.3	274	cytoskel	cytosk.	1	P
76	At3g53750	gi|21554576	Actin 3	41.8/5.2	62.9	2.2	295	cytoskel	cytosk.	1	P
77	At5g59370	gi|8885543	Actin 4	41.8/5.3	38.2	1.9	164	cytoskel	cytosk.	1	P
77	At3g46520	gi|28827524	Actin 12	41.8/5.3	38.2	1.9	164	cytoskel	cytosk.	1	P
78	At1g79550	gi|30725646	Putative phosphoglycerate kinase	42.1/5.4	74.3	1.6	263	cytp.	metaC	3	undeter.
79	At2g20420	gi|22136422	Putative succinyl-CoA ligase β-chain	45.3/6.3	53.7	1.7	148	endomb.	unc.	3	P
80	At5g16510	gi|21280929	Putative reversibly glycosylated polypeptide	38.6/4.9	48.9	2.4	200	endomb.	metabolism	1	P
81	At1g35720	gi|12083278	Annexin 1	36.2/5.1	79.2	1.6	222	cytp.	unc.	3	P
82	At3g09820	gi|6681336	Adenosine kinase 1	37.8/5.2	76.7	2.2	297	cytp.	metabolism	3	P
83	At3g02230	gi|31711848	Reversibly glycosylated polypeptide-1	40.6/5.5	78.7	2.4	303	endomb.	cell wall	1	P
84	At5g15650	gi|21464559	Reversibly glycosylated polypeptide-2	40.9/5.7	79.7	2.4	272	endomb.	cell wall	no data	no data
85	At2g47470	gi|24417274	Disulfide isomerase-like	39.5/5.7	56.5	2.3	153	endomb.	unc.	3	undeter.
86	At3g17940	gi|9294498	Aldose 1-epimerase family protein	37.2/5.9	67.2	1.3	162	cytp.	unc.	1	P
87	At3g52930	gi|21592946	Putative fructose-bisphosphate aldolase	38.5/6.0	43.6	2.1	102	cytp.	metaC	3	undeter.
88	At3g52930	gi|21592946	Putative fructose-bisphosphate aldolase	38.5/6.0	80.2	2.4	286	cytp.	metaC	3	undeter.
89	At2g30970	gi|22136256	Aspartate aminotransferase	47.7/9.1	62.3	2.3	296	endomb.	metabolism	2	undeter.
90	At5g61720	gi|30793953	Expressed protein, *A. thaliana*	43.1/8.6	51.8	2	228	endomb.	unc.	1	P
91	At1g26480	gi|30017291	14-3-3 protein GF14 iota	30.5/4.7	63.4	2.4	232	cytp.	prot. proc.	1	P
92	At3g59480	gi|6996284	pfkB-type carbohydrate kinase family protein	35.0/5.1	73.0	2.2	273	cytp.	energy	1	A

(*continued on next page*)

Table 1 (*continued*)

Spot No.[a]	AGI name	Gene index	Protein identity (TAIR description)	MM/p*I*[b]	Sequence coverage[c]	ProFound-score[d]	Mascot-score[e]	Subcellular localization[f]	Functional category[g]	Genevestigator[h]	Micro array detection[i]
93	At3g59480	gi|6996284	pfkB-type carbohydrate kinase family protein	35.0/5.1	65.0	2.4	294	cytp.	energy	1	A
94	At2g31390	gi|22136070	pfkB-type carbohydrate kinase family protein	35.2/5.2	46.5	2.2	176	cytp.	energy	3	P
95	At1g07750	gi|24111337	Cupin family protein	38.3/5.8	41.9	2.3	174	cytp.	unc.	1	undeter.
96	At1g07750	gi|24111337	Cupin family protein	38.3/5.8	48.0	2.1	170	cytp.	unc.	1	undeter.
97	At5g43330	gi|20259605	Putative malate dehydrogenase	35.7/6.4	68.4	2.4	297	cytp.	energy	1	P
98	At1g69940	gi|12325236	Pectinesterase family protein	36.6/9.3	51.5	2.2	207	cell wall	cell wall	1	P
99	At5g01410	gi|20453411	Putative stress-responsive protein	33.2/5.7	17.5	1.1	64	cytp.	metabolism	3	P
100	At5g01410	gi|20453411	Putative stress-responsive protein	33.2/5.7	41.4	1.9	158	cytp.	metabolism	3	P
101	At1g63000	gi|14423536	Expressed protein, *A. thaliana*	33.6/5.8	33.9	2.4	104	unknown	unc.	2	P
102	At2g28680	gi|4580389	Cupin family protein	38.4/6.3	65.7	2.4	266	cytp.	unc.	1	undeter.
103	At4g10260	gi|4538955	pfkB-type carbohydrate kinase family protein	34.7/7.7	58.3	2.3	169	cytp.	energy	1	A
104	At4g10260	gi|4538955	pfkB-type carbohydrate kinase family protein	34.7/7.7	65.4	2.1	270	cytp.	energy	1	A
105	At1g64980	gi|34146814	Expressed protein, *A. thaliana*	30.2/5.3	41.7	1.8	114	cytp.	unc.	1	P
106	At1g11840	gi|15810219	Putative lactoylglutathione lyase	31.9/4.9	52.3	2.4	189	unknown	unc.	3	P
107	At1g11840	gi|15810219	Putative lactoylglutathione lyase	31.9/4.9	30.4	2.0	82	unknown	unc.	3	P
108	At1g11840	gi|15810219	Putative lactoylglutathione lyase	31.9/4.9	52.3	2.0	154	unknown	unc.	3	P
109	At5g04180	gi|21618128	Carbonic anhydrase family protein	31.4/7.9	43.3	2.3	182	endomb.	unc.	1	P
110	At5g19510	gi|15810631	Elongation factor 1B α-subunit 2	24.2/4.3	45.5	2.3	149	ribosome	prot. synth.	3	P
111	At1g78300	gi|21618266	14-3-3 protein GF14 ω	29.1/4.6	83.0	1.5	222	cytp.	prot	1	P
112	At1g23730	gi|21386933	Putative carbonic anhydrase	28.8/6.7	67.4	2.4	231	cytp.	unc.	1	undeter.
113	At1g47260	gi|21280965	Bacterial transferase hexapeptide repeat-containing protein	30.0/6.9	58.3	1.4	157	endomb.	unc.	3	P
113	At5g20080	gi|21592883	Putative NADH-cytochrome *b*5 reductase	36.0/9.3	57.3	1.7	172	endomb.	energy	3	A
114	At2g21870	gi|19310625	Expressed protein, *A. thaliana*	27.6/6.3	60	2.4	150	endomb.	energy	3	P
115	At3g55440	gi|11270444	Putative triosephosphate isomerase	27.1/5.1	93.3	2.1	272	cytp.	metaC	3	undeter.
116	At2g46860	gi|20197320	Putative inorganic pyrophosphatase	24.9/5.5	40.7	0.8	75	cytp.	metabolism	1	P
117	At1g07890	gi|21554322	L-Ascorbate peroxidase 1	27.5/5.7	48.8	2.3	124	endomb.	defense	3	undeter.
118	At1g07890	gi|21554322	L-Ascorbate peroxidase 1	27.5/5.7	74.4	2.4	252	endomb.	defense	3	undeter.
119	At5g01600	gi|12642862	Ferritin 1	28.2/5.7	35.7	2.1	121	endomb.	transport	4	P
120	At5g20720	gi|17065646	20 kDa chaperonin	26.8/9.3	69.6	2.3	181	endomb.	transport	4	A
121	At3g60180	gi|2497486	Putative uridylate kinase	22.5/5.7	89.1	2.3	263	cytp.	metabolism	4	undeter.
122	At2g47730	gi|20197312	Glutathione *S*-transferase 6	24.1/6.1	26.0	1.2	69	endomb.	metabolism	2	P
123	At2g47730	gi|20197312	Glutathione *S*-transferase 6	24.1/6.1	82.8	2.4	232	endomb.	metabolism	2	P
124	At3g05930	gi|6714406	Germin-like protein	23.0/9.6	58.9	2.3	148	endomb.	defense	1	P
125	At3g52300	gi|21555349	ATP synthase D chain-related	19.6/4.9	80.4	2.3	210	cytp.	energy	3	P
126	At4g24640	gi|21593482	Invertase/pectin methylesterase inhibitor family protein	19.9/5.5	77.5	2.4	280	endomb.	unc.	1	P
127	At4g11600	gi|21617919	Putative glutathione peroxidase	18.6/7.4	66.9	2.4	162	endomb.	defense	2	P
128	At3g06050	gi|20466103	Alkyl hydroperoxide reductase	21.4/9.6	65.2	2.4	199	endomb.	unc.	2	P
129	At2g21130	gi|21593051	Peptidyl-prolyl *cis–trans* isomerase/cyclophilin	18.5/9.4	82.8	2.4	196	cytp.	prot. proc.	2	P
129	At4g38740	gi|21593682	Peptidyl-prolyl *cis–trans* isomerase/cyclophilin	18.4/9.0	77.9	2.1	180	cytp.	prot. proc.	3	P
130	At2g21130	gi|21593051	Peptidyl-prolyl *cis–trans* isomerase/cyclophilin	18.5/9.4	90.2	2.1	201	cytp.	prot. proc.	2	P
131	At2g21130	gi|21593051	Peptidyl-prolyl *cis–trans* isomerase/cyclophilin	18.5/9.4	90.2	1.9	198	cytp.	prot. proc.	2	P
132*	At3g07850	gi|11994377	Exopolygalacturonase	45.7/9.7	31.0	2.3	156	endomb.	unc.	1	P
133*	At3g07850	gi|6648197	Exopolygalacturonase	45.6/9.7	30.9	2.3	136	endomb.	unc.	1	P
134	At5g52360	gi|10177402	Putative actin-depolymerizing factor	15.3/5.4	78.5	2.3	125	cytp.	cell wall	1	P
135	At4g29340	gi|21537389	Profilin 3	14.4/4.9	57.5	2.0	114	cytp.	cytosk.	1	P
136*	At4g29340	gi|21537389	Profilin 3	14.4/4.9	44.0	1.9	82	cytp.	cytosk.	1	P
137	At1g15415	gi|5103841	Expressed protein (low similarity to LEA protein)	10.4/5.9	76.0	2.4	137	nucleus	unc.	no data	no data
138	At4g13560	gi|28827234	LEA domain-containing protein	11.6/7.7	88.1	2.4	178	nucleus	unc.	1	P
139	At4g13560	gi|28827234	LEA domain-containing protein	11.6/7.7	62.4	2.3	106	nucleus	unc.	1	P
140	At5g12140	gi|21593904	Putative cysteine protease inhibitor	11.2/4.9	58.4	0.7	64	cytp.	unc.	3	P
141	At3g51030	gi|21617958	Thioredoxin H-type 1	12.7/5.6	48.2	0.7	68	cytp.	unc.	1	P

142	At5g40370	gi|2192438	Putative glutaredoxin	11.7/7.7	90.1	2.4	187	endomb.	metabolism	3	P
143	At1g04670	gi|5978673	Expressed protein, *A. thaliana*	13.4/9.9	38.9	0.7	76	endomb.	unc.	1	P
144	At1g62820	gi|7529028	Putative calmodulin	16.6/4.7	43.9	[j]	66	unknown	unc.	3	P
144**	At4g13230	gi|2137131	LEA domain-containing protein	13.1/10.3	[j]	[j]	76	endomb.	unc.	1	P
145	At4g02890	gi|3158374	Polyubiquitin	15.4/7.6	68.6	1.7	133	cytp.	prot. proc.	3	P

[a] Proteins indicated with asterisk were identified (**) or validated (*) by LC–MS/MS analysis. All other proteins were identified by MALDI-TOF MS.

[b] Molecular mass (kDa) and p*I* of identified protein (according to the Proteinscape software; Bruker Daltonics).

[c] Sequence coverage (percentage of the complete protein sequence identified).

[d] ProFound score (http://prowl/profound_help.html).

[e] Mascot score (http://www.matrixscience.com/help/scoring_help.html).

[f] Subcellular localization: cytp., cytoplasm; endomb., endomembrane (i.e., secretory pathway, mitochondrion or chloroplast).

[g] Functional category: unc., unclassified protein; metabolism, general metabolism; energy, energy metabolism; transport, cellular transport; prot. proc., protein processing; prot. synth., protein synthesis; cytosk., cytoskeleton; metaC, C-compound and carbohydrate metabolism; defense, defense and stress response.

[h] Genevestigator data: (1) preferentially expressed in stamen; (2) preferentially expressed in stamen and some other organs; (3) expressed in all organs; (4) preferentially expressed in another organ than stamen; (no data) gene not on ATH1 GeneChip.

[i] Microarray detection according to the transcriptome analysis performed in our laboratory (cf. Materials and methods). (A) mRNAs absent (not detected in any genotype); (P) mRNAs present (detected in any genotype); (undeter.) undetermined (present only in some genotypes; value close to threshold); (no data) gene not on ATH1 GeneChip.

[j] These numbers are not available for the LC–MS/MS analysis.

Protein identification by PMF analysis

All protein spots were processed by automated in-gel tryptic digestion and MALDI-TOF MS analysis to generate PMFs. Spots with non-significant identification scores were further analyzed by LC–MS/MS. The 155 spectra obtained were used to screen the NCBI database, which enabled us to assign putative identity to 145 spots representing 121 different proteins (Table 1). For 10 of the 145 spots analyzed, two distinct polypeptides were identified either by MALDI-TOF MS or by LC–MS/MS. In addition, LC–MS/MS revealed the identity of 4 further proteins (Table 1). Identified proteins ranged in calculated molecular mass from 10.4 to 108 kDa and in calculated p*I* from 4.2 to 9.8 (Table 1). The average deviations of the measured peptide masses compared to the theoretical values of the respective peptides of the identified proteins were less than 0.1 Da. The identity of 105 spots was assigned in at least two replicates of different genotypes, 40 others were characterized in at least two replicates of the same genotype. The low complexity of the male gametophyte and the relative purity of the pollen samples appear to result in a beneficial signal-to-noise ratio which may explain the highly significant Mascot and ProFound scores obtained in our analysis (Table 1). All identified proteins were designated with arbitrary spot numbers as shown in Fig. 1. As indicated in Table 1, all Gene Index (GI) annotations could be assigned with the corresponding AGI name [8] and the designation of the corresponding genes is given according to the TAIR database (http://www.arabidopsis.org/index.jsp). The predicted molecular masses and isoelectric points for the majority of the identified proteins were consistent with the experimental data as judged from the location of the respective polypeptide spots within the 2D gels. This further corroborated the identity of the recognized polypeptides. Exceptions included spot 1 (At1g49490) which exhibits a higher molecular mass (~150 kDa) than calculated (90.9 kDa). Indeed, big proteins can display a strong tendency to aggregate, complicating migration. Conversely, products related to spots 132 and 133 (At3g07849) show a low molecular mass (~17 kDa) compared with the calculated value (~45 kDa). This could result from post-translational modifications or degradation of intact proteins during the extraction procedure.

Presence of multiple polypeptide isoforms

Generally, proteins can be assigned to two classes of isoforms. The first class comprises polypeptide variants encoded by the same gene (e.g., splice variants or different post-translational modifications). The second class encompasses highly sequence-related protein isoforms encoded by distinct genes. In our study, 27 candidates of the first class of isoforms were detected and 16 of the second. Approximately 44% of all isoforms likely represent different phosphorylation states since they are characterized by the same apparent molecular mass but different p*I* values. As

suggested by the previous analysis from rice anthers about proteins present at different stages of pollen development [16], occurrence of multiple isoforms originating from a single gene could indicate that post-translational modifications might play a significant role during pollen development.

Subcellular localization

The potential subcellular localization of identified polypeptides was assigned using web-based tools as described in Materials and methods. Despite a somewhat different classification, the observed proportions were compared to the subcellular localization prediction of the whole *Arabidopsis* protein set [26]. We noted that the largest fraction (~41%) of the identified proteins appears to be cytoplasmic which is consistent with the respective value for the whole *Arabidopsis* proteome (58%). A second significant proportion (~40%) is potentially targeted to the endomembrane system which is also similar to the respective fraction of the total *Arabidopsis* protein set (43%). Furthermore, approximately 4% of the pollen polypeptides are predicted to reside in the nucleus, ~5% are presumably associated with the cytoskeleton, ~1.6% reside in the vacuole, ~0.8% in the cell wall, ~0.8% in the ribosome, and the localization of ~7% of the proteins cannot currently be predicted. Notably, none of the *Arabidopsis* pollen coat proteins described in the study of Mayfield et al. [17] was identified in our analysis. The extreme hydrophobic properties of these proteins could mean that they were not extracted in our experimental conditions. Besides, only 6 proteins with predicted transmembrane domains (At2g07698, At1g09210, At2g47470, At1g69940, At1g56340, and At3g07850) were identified. This may relate to the fact that generally integral membrane proteins, in particular those with multiple membrane-spanning domains, do not resolve well during isoelectric focusing and are therefore inherently underrepresented in 2-DE [27]. This fact represents a well-known limitation of 2-DE-based proteomic approaches that also applies to our study.

Functional categories

We determined the predicted functional categories assigned to the identified pollen proteins. The majority of the polypeptides are involved in metabolism (~42%). This could potentially reflect a special requirement of the mature pollen, since only 22.5% of the total proteome of *Arabidopsis* is dedicated to this function [8]. Alternatively, since metabolic enzymes are generally highly abundant, soluble proteins, this functional category might be overrepresented in our analysis. The latter hypothesis appears more likely since only ~19% of the pollen transcriptome encodes metabolic proteins (see below; [10]). In detail, ~19% of the identified proteins implicated in metabolic functions are related to carbohydrate metabolism (e.g., triosephosphate isomerase -At3g55440-, glycoside hydrolase -At3g5440-, glucose dehydrogenase -At5g15490-, or fructose-bisphospate aldolase -At3g52930) and ~51% to energy metabolism. Similar to the *Arabidopsis* total proteome analysis (i.e., 11.7%) [8], ~9% of the *Arabidopsis* mature pollen proteome is implicated in biogenesis of cellular compounds. With respect to this function, enzymes required for the synthesis of cell wall constituents, such as pectinesterase -At1g69940- or exopolygalacturonase -At4g24640, At3g07850- have been identified. Cytoskeleton-associated proteins have also been found (e.g., actin -At2g37620, At3g53750, At3g46520, At5g59370-, profilin -At4g29340-, and tubulin -At1g50010, At5g44340). The latter polypeptides are presumably involved in the regulation of polarized tip-growth of pollen tubes upon germination and along the female reproductive tract [7]. In addition, more than 9% of the identified proteins exhibit a protein processing function (14% in the *Arabidopsis* total proteome [8]). Interestingly, the majority of these proteins (75%) are preferentially implicated in the determination of protein fate rather than in protein synthesis. This is consistent with the idea that *Arabidopsis* mature pollen is charged with a pre-formed translational apparatus enabling rapid activation upon hydration and germination [10]. A considerable number (27%, i.e., 33 proteins) of the experimentally identified polypeptides are of unknown function. This number is essentially in accordance with the respective transcriptome data (see below). A large set of unclassified proteins was also found in recent plant, human, and yeast mitochondrial proteome analyses [28–30] and interpreted as an indication for a putative wealth of as yet undiscovered mitochondrial functions. By analogy, presence of a major set of proteins with unknown functions in pollen may hint at as yet unidentified cellular processes, of which some might be specific for the male gametophyte. The latter hypothesis is supported by the fact that 16 of the respective genes are preferentially expressed in the stamen (Genevestigator data). The identification of these unclassified pollen proteins provides a basis for the use of reverse genetics to identify novel biological functions in plants by taking advantage of the reduced complexity of the pollen system.

Comparison with pollen transcriptome data

Expression data of genes corresponding to proteins identified in our study were investigated by using the microarray data deposited at the Genevestigator server [24]. This revealed that a relative majority of these genes (40%) are expressed in all organs. Approximately 48% of this group are implicated in metabolic functions, suggesting that they act in general plant-wide housekeeping. A good proportion (30.5%) of our candidates are preferentially expressed in the stamen organ and 16.5% are preferentially expressed in stamen and some other organs. Only a minority of the candidates (5%) are preferentially expressed in another organ than stamen.

Presence of proteins identified in our study was compared to data obtained by a previous transcriptome

analysis of mature pollen [10]. In terms of functional categories, general metabolism (20.6%), protein processing (9%), and stress-related (4%) functions assigned on the basis of our proteomic approach reflect similar proportions, namely 18%, 6%, and 6%, respectively, in the microarray-based transcriptome study. Polypeptides devoted to cellular transport (4%) and signaling (0%) functions appear underrepresented in our proteome analysis compared to gene expression data (8% and 13%, respectively). Complete absence of proteins implicated in cellular signaling may be due to the fact that many of these polypeptides are either of low abundance and/or integral membrane proteins that are difficult to detect in 2-DE-based approach. Low abundance may also account for a similar discrepancy with regard to transcriptional regulators (0% in proteome versus 7% in transcriptome). Surprisingly, products exhibiting a function in energy metabolism appear to be highly represented in the proteome (21%) compared to the transcriptome (1%). This may indicate that mature pollen is well prepared with pre-existing metabolic enzymes for rapid energy conversion upon germination and pollen tube expansion. Both our proteomic approach and the global analysis of the pollen transcriptome revealed a high proportion of as yet unclassified proteins (27% and 19%, respectively; see above).

Besides the above-described global comparison, we compared each identified protein with its respective transcriptome data. The latter either originate from experiments performed in our laboratory (cf. Materials and methods) or publicly available datasets ([12] and unpublished datasets available at the NASC website; http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl; slides Honys_MPG1_SLD; ATGE_73A and DT001_AG_pollen, respectively). In total, transcripts of proteins unequivocally identified in 14 spots (corresponding to 10 distinct AGI loci) have been reproducibly classified as absent (based on MicroArray Suite 5.0 software) in our own transcriptome analyses (At2g07698; At2g33210; At3g18190; At3g23990; At3g59480; At4g10260; At5g03630; At5g20080; At5g20720; At5g42740, see Table 1; please note that our transcriptome studies were carried out by applying the same plant growth conditions and techniques for pollen mining as for the proteome analysis). Likewise, nine of the ten transcripts were declared as absent in the study of Pina et al. [12], while the remaining datasets deposited on the NASC webpage provide ambiguous results for these proteins. The latter might be due to differences in plant growth and/or pollen mining as compared to our techniques. In summary, the respective transcripts might either be absent in mature pollen or present at very low levels. This observation implies that low transcript abundance may suffice to ensure accumulation of a significant proportion of protein in some instances (e.g., for proteins with a low turnover rate). Alternatively, the respective polypeptides might have been translated at an earlier stage of pollen development (e.g., in the bicellular stage) and still persist in mature pollen. In any case, this finding further underlines the importance of proteome studies as valuable supplements to transcriptome analyses.

Acknowledgments

We acknowledge the SALK Institute (La Jolla, USA), Syngenta Biotechnology Inc. (Research Triangle Park, USA), and the Sainsbury Laboratory (Norwich UK) for providing *Arabidopsis* knockout lines either directly to us or via the NASC repository. This work was supported by grants of the Max-Planck society to R.P. and J.S.

References

[1] S. McCormick, Male gametophyte development, Plant Cell 5 (1993) 1265–1275.
[2] D. Twell, S.K. Park, E. Lalanne, Asymmetric division and cell-fate determination in developing pollen, Trends Plant Sci. 3 (1998) 305–310.
[3] J.P. Mascarenhas, Molecular mechanisms of pollen tube growth and differentiation, Plant Cell 5 (1993) 1303–1314.
[4] S. McCormick, Control of male gametophyte development, Plant Cell 16 (2004) S142–S153.
[5] J.A. Feijó, J. Sainhas, T. Holdaway-Clarke, M.S. Cordeiro, J.G. Kunkel, P.K. Hepler, Cellular oscillations and the regulation of growth: the pollen tube paradigm, BioEssays 23 (2001) 86–94.
[6] J.A. Feijó, S.S. Costa, A.M. Prado, J.D. Becker, A.C. Certa, Signaling by tips, Curr. Opin. Plant Biol. 7 (2004) 589–598.
[7] K. Hepler, L. Vidali, A.Y. Cheung, Polarized cell growth in higher plants, Annu. Rev. Cell Dev. Biol. 17 (2001) 159–187.
[8] *Arabidopsis* Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, Nature 408 (2000) 796–815.
[9] J.D. Becker, L.C. Boavida, J. Carneiro, M. Haury, J.A. Feijó, Transcriptional profiling of *Arabidopsis* tissues reveals the unique characteristics of the pollen transcriptome, Plant Physiol. 133 (2003) 713–725.
[10] D. Honys, D. Twell, Comparative analysis of the *Arabidopsis* pollen transcriptome, Plant Physiol. 132 (2003) 640–652.
[11] D. Honys, D. Twell, Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*, Genome Biol. 5 (2004) R85.
[12] C. Pina, F. Pinto, J.A. Feijó, J.D. Becker, Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation, Plant Physiol. 138 (2005) 744–756.
[13] J. Cordewener, J. Bergervoet, C.M. Liu, Changes in protein synthesis and phosphorylation during microspore embryogenesis in *Brassica napus*, J. Plant Physiol. 156 (2000) 156–163.
[14] A. Kalinowski, K. Winiarczyk, B. Wojciechowska, Pollen proteins after two-dimensional gel electrophoresis and pollen morphology of the amphiploids *Aegilops kotschyi* and *Ae. variabilis* with *Secale cereale*, Sex, Plant Reprod. 14 (2001) 153–161.
[15] N. Imin, T. Kerim, J.J. Weinman, B.G. Rolfe, Characterization of rice anther proteins expressed at the young microspore stage, Proteomics 1 (2001) 1149–1161.
[16] T. Kerim, N. Imin, J.J. Weinman, B.G. Rolfe, Proteome analysis of male gametophyte development in rice anthers, Proteomics 3 (2003) 738–751.
[17] J.A. Mayfield, A. Fiebig, S.E. Johnstone, D. Preuss, Gene families from the *Arabidopsis thaliana* pollen coat proteome, Science 292 (2001) 2482–2485.
[18] J.M. Alonso, A.N. Stepanova, T.J. Leisse, C.J. Kim, H. Chen, P. Shinn, D.K. Stevenson, J. Zimmerman, P. Barajas, R. Cheuk, C. Gadrinab, C. Heller, A. Jeske, E. Koesema, C.C. Meyers, H. Parker, L. Prednis, Y. Ansari, N. Choy, H. Deen, M. Geralt, N. Hazari, E. Hom, M. Karnes, C. Mulholland, R. Ndubaku, I. Schmidt, P. Guzman, L. Aguilar-Henonin, M. Schmid, D. Weigel, D.E. Carter, T. Marchand, E. Risseeuw, D. Brogden, A. Zeko, W.L. Crosby, C.C.

Berry, J.R. Ecker, Genome-wide insertional mutagenesis of *Arabidopsis thaliana*, Science 301 (2003) 653–657.

[19] A. Sessions, E. Burke, G. Presting, G. Aux, J. McElver, D. Patton, B. Dietrich, P. Ho, J. Bacwaden, C. Ko, J. Clarke, D. Cotton, D. Bullis, J. Snell, T. Miguel, D. Hutchison, B. Kimmerly, T. Mitzel, F. Katagiri, J. Glazebrook, M. Law, S. Goff, A high-throughput *Arabidopsis* reverse genetics system, Plant Cell 14 (2002) 2985–2994.

[20] A. Tissier, S. Marillonnet, V. Klimyuk, K. Patel, M. Angel Torres, G. Murphy, J.D.G. Jones, Multiple independent defective suppressor–mutator transposon insertions in *Arabidopsis*: a tool for functional genomics, Plant Cell 11 (1999) 1841–1852.

[21] S.A. Johnson-Brousseau, S. McCormick, A compendium of methods useful for characterizing *Arabidopsis* pollen mutants and gametophytically expressed genes, Plant J. 39 (2004) 761–775.

[22] J. Gobom, M. Schuerenberg, M. Mueller, D. Theiss, H. Lehrach, E. Nordhoff, Cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics, Anal. Chem. 73 (2001) 434–438.

[23] J.C. Redman, B.J. Haas, G. Tanimoto, C.D. Town, Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array, Plant J. 38 (2004) 545–561.

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

[25] A. Devoto, H.A. Hartmann, P. Piffanelli, C. Elliott, C. Simmons, G. Taramino, C.-S. Goh, F.E. Cohen, B.C. Emerson, P. Schulze-Lefert, R. Panstruga, Molecular phylogeny and evolution of the plant-specific seven transmembrane MLO family, J. Mol. Evol. 56 (2003) 77–88.

[26] R.A. Gutiérrez, P.J. Green, K. Keegstra, J.B. Ohlrogge, Phylogenetic profiling of the *Arabidopsis thaliana* proteome: what proteins distinguish plants from other organisms? Genome Biol. 5 (2004) R53.

[27] N.A. Jones, Two-dimensional polyacrylamide gel electrophoresis for proteome analyses, in: P. Cutler (Ed.), Methods in Molecular Biology, vol. 244, Humana Press, Totowa, NJ (USA), 2004, pp. 353–359.

[28] J.L. Heazlewood, J.S. Tonti-Filippini, A.M. Gout, D.A. Day, J. Whelan, A.H. Millar, Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins, Plant Cell 16 (2004) 241–256.

[29] S.W. Taylor, E. Fahy, B. Zhang, G.M. Glenn, D.E. Warnock, S. Wiley, A.N. Murphy, S.P. Gaucher, R.A. Capaldi, B.W. Gibson, S.S. Ghosh, Characterization of the human heart mitochondrial proteome, Nat. Biotechnol. 21 (2003) 281–286.

[30] A. Sickmann, J. Reinders, Y. Wagner, C. Joppich, R. Zahedi, H. Meyer, B. Schonfisch, I. Perschil, A. Chacinska, B. Guiard, P. Rehling, N. Pfanner, C. Meisinger, The proteome of *Saccharomyces cerevisiae* mitochondria, Proc. Natl. Acad. Sci. USA 100 (2003) 13207–13212.