SUPPLEMENTAL MATERIAL

NOTE

Due to the fact that we are using a novel deterministic classification scheme to classify our proteomics dataset, PRIDE cannot display the data in exactly the same manner as provided in Table S1. This is especially true for class 3a peptides and their respective spectral evidence. PRIDE is structured in such a way that only one protein accession is given for a certain peptide; as such it cannot properly display our class 3a peptides, which imply several distinct protein accessions.

Furthermore, in order to reflect which proteins are observed in one particular biochemical fraction, rather than in a technical experiment, we combined different biological replicates of the same fractionation protocol into one export file. Therefore, PRIDE displays 8 instead of the 15 experiments reported in the manuscript.

INDEX

SUPPLEMENTAL METHODS

SUPPLEMENTAL FIGURES

SUPPLEMENTAL TABLES

SUPPLEMENTAL METHODS

Pollen collection and protein extraction

Pollen collection. Mature pollen from *Arabidopsis thaliana* (Columbia accession) was collected using the "vacuum cleaner" method (Johnson-Brousseau and McCormick 2004). An in-house made wand, equipped with 3 nylon meshes of decreasing sizes 80 (filter for parts of the flower), 38 (filter for other debris), and 6µm (accumulates only pollen grains) was attached to a professional vacuum cleaner. Using the "vacuum method" (open inflorescences were vacuumed every second day for 2-3 weeks) we could to collect comfortable amounts of mature pollen grains, which were checked for purity under light microscope (Fig 1C).

Protein extraction. Pollen proteins were extracted using a sequential protein extraction method. The first protein fraction was obtained by performing a cyclohexane extraction of pollen coat proteins as described (Mayfield and Preuss 2000). The cyclohexane layer was collected by centrifugation and reduced under a stream of nitrogen (Der Fen Suen et al. 2003). After the cyclohexane evaporated, the pellet was resuspended in 50µl urea buffer. Pollen pellet recovered after the removal of the cyclohexane layer was frozen in liquid nitrogen and homogenized in a Mixer-Mill using glass beads. Pollen homogenate was resuspended in the 150µL buffer containing 100mM KCl, 50mM Tris-HCl pH8, 20% Glycerol (v/v), 1mM PMSF and 2x protease inhibitor cocktail (Complete Mini, Roche Diagnostics GmbH Mannheim, Germany) (Giavalisco et al. 2005). The homogenate was centrifuged for 30 min, at 20,000g and 4^{o}C, and the supernatant containing the salt soluble protein fraction was collected and stored at -20^{o}C. The remaining pellet was homogenized using an electric pestle in 150µL urea buffer containing 10mM Tris-HCl pH8, 5mM EDTA, 0.05% SDS, 6M Urea, 1M Thiourea, 2x protease inhibitor cocktail. The homogenate was centrifuged, the supernatant collected and stored as described above. The last protein extraction step consisted in the homogenization of the recovered pellet in 150µL detergent based buffer: 125mM Tris-HCl pH8**,** 2% SDS, 10% Glycerol, 50mM $Na_2S_2O_5$, 2x protease inhibitor cocktail (Martinez-Garcia et al. 1999). The resulting supernatant was collected and stored at -20^{o}C.

Protein separation and in gel digestion. Approximately 600µg protein per fraction was loaded individually onto a 0.75mm, 12% SDS-PAGE mini-gel (Mini-PROTEAN® 3 Cell, Bio-Rad) and separated by one-dimensional SDS-PAGE. Gels were cut into 0.4cm slices using a custom-made gel cutter, fragmented further into smaller pieces used for in gel digestion with trypsin, at 37 ^{o}C, over night (Baerenfaller et al. 2008).

Isotope-Coded Affinity Tagging (ICAT)

ICAT procedures were carried out according to the manufacturer's protocol for Cleavable ICAT®Reagent Kit for Protein Labeling (Applied Biosystems). Pollen proteins (1mg) were split and labeled on their Cysteine residues with a chemical tag consisting in an isotope-coded tag ($C_{10}H_{17}N_3O_3$; ^{13}C-heavy and ^{12}C-light) with specificity toward sulfhydryl groups, a cleavable eightfold deuterated linker and a biotin affinity tag (Gygi et al. 1999). Labeled proteins were trypsin digested and the resulting peptides were separated and analyzed by cation exchange chromatography, using an Agilent 1100 high-performance liquid chromatography (HPLC) system and a Polysulfoethyl A™ high-resolution cation-exchange column from PolyLC Inc., 5micron silica packing, pore-width of 300Å, column dimension 4.6 x 200mm. The collected HPLC fractions were pooled and the new set of samples (5 and respectively 7 samples for two ICAT experiments) was subjected to affinity chromatography (Avidin Cartridge provided by manufacturer) to isolate Cysteine labeled peptides. We analyzed both the cysteine-labeled and the non-labeled, flow-through fraction.

Mass spectrometry, data processing and analysis

The MS/MS spectra were searched using the TurboSEQUEST search algorithm version 27 (rev.12) against the *Arabidopsis thaliana* reference protein database TAIR7, concatenated with the reversed TAIR7 database as decoy database in order to estimate the overall false discovery rate (FDR) (Elias and Gygi 2007) plus roughly 260 additional known contaminants (i.e. porcine trypsin, various keratins, etc). The searches were performed allowing for one trypsin miss-cleavage, a mass tolerance of 3Da, and Methionine oxidation and Cysteine carbamidomethylation as variable modifications. For the case of ICAT-labeled peptides, the searches were done with a Cysteine fixed modification (ICAT-light 227.2603), a variable modification of 8.9339 at Cysteine (8.9339 represents the average mass difference between ICAT-heavy and light tags) and a variable Methionine oxidation. Database search results were further processed with the transproteomic pipeline (version 3.2.1_SQUALL) and PeptideProphet (Keller et al. 2002) using a probability cut-off of greater than or equal to 0.9 for the final dataset. The identifiable proteome (Brunner et al. 2007) was determined for TAIR7: An *in silico* database analysis of the 27,029 protein-coding genes (not considering pseudogenes and transposons), which encode 31,921 proteins revealed that a total of 29,988 distinct protein sequences (differing by at least one amino acid, and as such - at least in theory - distinguishable by proteomics methods) can be differentiated (Figure 2B, Figure S3).

Algorithm to extract the minimal set of proteins explaining the ambiguous peptide identifications (class 3B)

According to our classification scheme (Figure 2C), peptides of classes 1a, 1b and 3a are not substrings of any other protein sequences, implying as such a minimal set of proteins that explains the peptides of these three classes. Class 2 peptides on the other hand can imply proteins that could also contain peptides of class 1a, 1b and 3a, and as such these peptides can be explainable by these proteins. For example, if we have seen the proteins AT1G01100.1, AT1G01100.2, AT1G01100.4 with the peptide of class 1b NVTDLIMNVGAGGGGGAPVAAAAPAAGGGAAAAPAAEEK, these proteins can also explain the peptide AAGVSIESYWPMLFAK (Figure 2C). Thus in case we observe these two peptides, the minimal set of proteins explaining these peptides is AT1G01100.1, AT1G01100.2, AT1G01100.4 and the maximal set of proteins explaining these peptides is AT1G01100.1, AT1G01100.2, AT1G01100.3, AT1G01100.4.

Peptides of class 3b on the other hand, that imply proteins with different protein sequences and different gene models, can imply any number of proteins between MIN_EXPLAINABLE and MAX_EXPLAINABLE, where MIN_EXPLAINABLE is the minimal number of proteins that explains all peptides of class 3b after removing all the peptides explainable by class 1a-3a, and MAX_EXPLAINABLE is the maximal number of proteins explaining these peptides, i.e. all proteins seen with at least one class 3b peptide.

Following the Occam's Razor principle, one would want to find the minimal set of proteins explaining all the peptides called present in the sample(s). The problem of finding the minimal set of proteins can be formulated as an integer programming problem in the following way. Let y_i be a binary variable representing whether the protein *i* is in the sample ($y_i = 1$) or not ($y_i = 0$). Then the purpose would be to find the N dimensional configuration vector *Y* that solves the following integer programming problem:

Minimize: $\sum_{i=1}^{MAX\_EXPLAINABLE} y_i$

Subject to: $\sum_{j \in J} y_j$ >=1 for each class 3b peptide *z* and *J* is set of proteins containing the peptide z

This optimization problem can be solved using libraries for Linear Programming e.g. using the GNU Linear Programming Kit (http://www.gnu.org/software/glpk/). By applying this algorithm on the peptides of class 3b for the pollen dataset, all these peptides can be explained with as few as 132 proteins. The maximal set of proteins explaining these peptides on the other side has 2,557 proteins.

Re-mapping Affymetrix oligonucleotide probesets

Using the custom CDF libraries (version 10) downloaded from the Brainarray microarray laboratory we re-mapped the Affymetrix probesets against the TAIR7 database release and eliminated probesets that could be mapped to multiple positions in the genome. A total of 21,022 oligonucleotide probesets passed these criteria (3 gene models for non-coding RNA or non-coding retrotransposons were also removed), Figure S6. Due to statistical considerations (i.e. hybridization signals of genes with only few oligonucleotides in the probe set are less reliable) we flagged probesets with less than 6 oligonucleotides per probeset as lower quality. Since we did not assess differential gene expression but were mainly interested in present or absent calls, we left them in the final probeset dataset.
We re-analyzed the gene expression values of all 3 expression datasets and classified genes as present if the detection p-value was below 0.05, and further considered only those gene models that had a present call in both replicates of the developmental stages uninucleate microspores (UNM), bicellular pollen (BCP), and immature tricellular pollen (TCP) of the pollen development study (Honys and Twell 2004). Since this study had only one sample for mature pollen grain (MP), we established the total number of expressed genes in mature pollen by requiring genes to be expressed in 3 or more of the overall 6 mature pollen samples (Honys and Twell 2004; Pina et al. 2005; Schmid et al. 2005), Table S3.

Calculation of APEX values

We calculated the absolute protein expression measurement (APEX) for the pollen dataset as described (Lu et al. 2007). The APEX method estimates the absolute protein concentration from the proportionality between the number of peptides observed in the MS/MS experiments and the number of peptides expected from this protein based on the detectabilities of all *in silico* tryptic digested peptides (with up to 2 missed cleavages). First, a classifier for peptide detectability was built using Weka (version 3.4.4, http://www.cs.waikato.ac.nz/ml/weka/). The 40 most abundant pollen proteins (based on total spectral counts) were used to create the training set. All possible tryptic peptides generated by *in silico* digestion, which have molecular weights between 250-7,500 Da with at most 2 missed cleavages and at least 3 amino acids are labeled either as observed (peptide probability>0.9) or not observed according to PeptideProphet results. For each peptide, a feature vector was constructed from the molecular weight, the peptide length, and the frequency of each amino acid. A cost sensitive classifier based on bagging with a forest of random decision trees is used for training. Using the trained model, the classifier is applied to all *in silico* generated peptides from the mature pollen proteins, and probabilities of observing each peptide in the MS/MS experiment are obtained as an output. The sum of probabilities of observing each peptide from each protein is calculated, which is subsequently used in the APEX calculation by the equation from Lu et al. 2007. The estimate of the total number of protein molecules per cell is set as 2,5 x10^8 in this analysis.

Manual validation of mass spectra

To lower the FDR and to achieve very high data quality, we manually validated all protein identifications by a single spectrum of class 1a, 1b, 2, 3a (a total of 1,407 cases). 538 of these spectra

implied novel insights since the corresponding protein or gene model that had not been implied before or was inconsistently expressed in transcriptomics datasets; they were validated independently by two researchers. Conflicting judgments (< 5% of all cases) were additionally assessed by a third mass spectrometrist and checked against the search results of an independent search engine (Mascot). The single hit spectra, for which transcriptomics evidence had been observed (869), were validated by one researcher, and any unclear judgment assessed by an additional mass spectrometrist. The observation that 5 of the pollen mutants implied by independent genetic evidence (Table S5) were eliminated implies that this validation was quite stringent.

SUPPLEMENTAL FIGURES

Figure S1. Schematic representation of the plant life cycle

The plant life cycle alternates between a diploid (sporophytic generation) and a haploid generation (gametophytic generation). Cells in the sporophytic tissue in the sexual organs of the flower (anthers and carpels) undergo meiosis and the resulting haploid spores undergo several mitotic divisions to form multicellular gametophytes, which in turn form the gametes. Grey boxes refer to developmental processes.

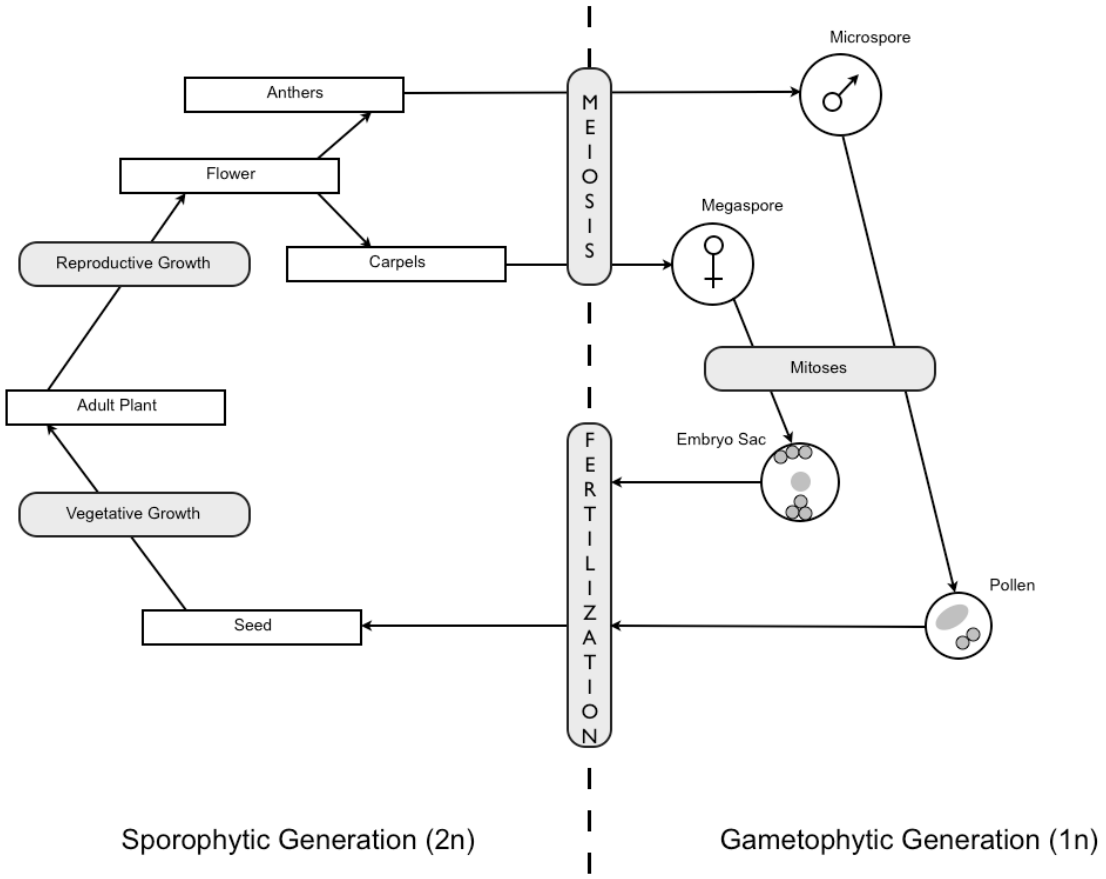


Figure S2. Overview of all mass spectrometry runs

We performed 15 independent experiments accounting for a total of 250 mass spectrometry runs. For the samples obtained after 1-D SDS PAGE separation and in-gel trypsin digestion, we performed a total 216 runs. One experiment represents one gel and the number of runs in parenthesis includes technical replicates. For the samples obtained by the ICAT method, we performed a total of 34 mass spectrometry runs.

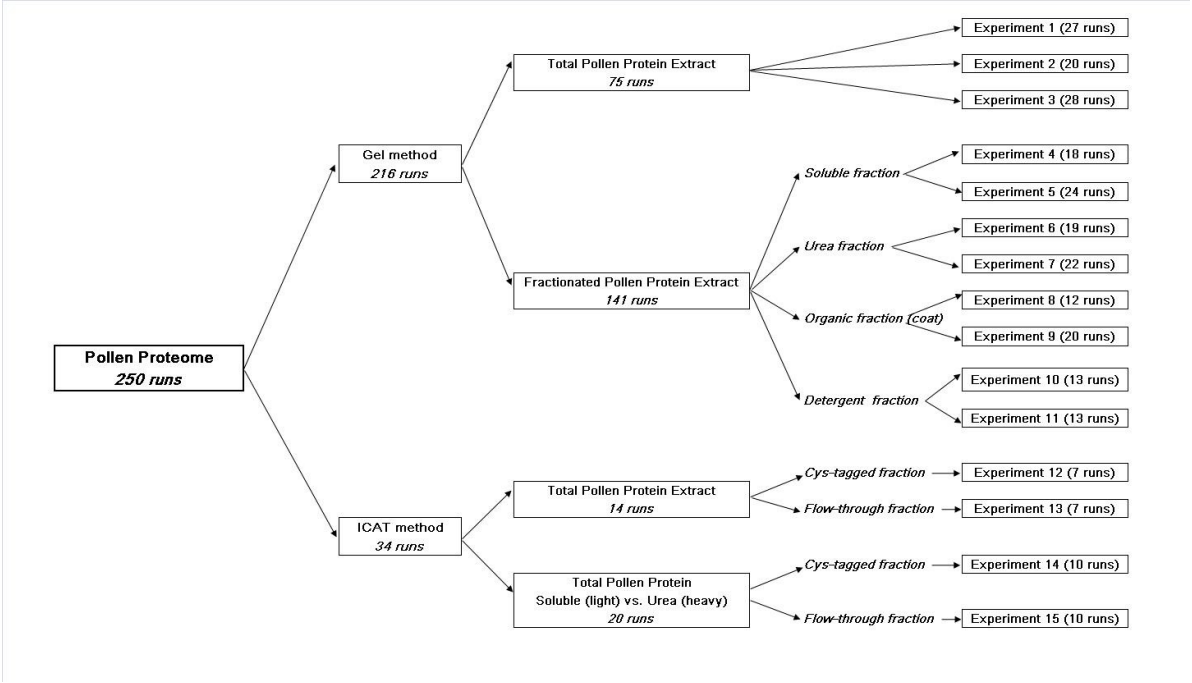


Figure S3. Results of an *in silico* analysis of the *Arabidopsis* reference database TAIR7

The *in silico* database analysis allows us to determine the identifiable proteome and establish the gene model – protein sequence – protein accession relationships (which are required for our deterministic peptide evidence classification scheme, see Figure 2B). Among 31,921 protein sequences encoded by 27,029 protein-coding gene models in TAIR7 (not considering pseudogenes and transposons), a total of 29,988 protein sequences differ by at least one amino acid and are thus – at least in theory – distinguishable by proteomics. Many protein sequences encoded by genes that differ only in the 5' or 3' UTR are not distinguishable by proteomics methods.

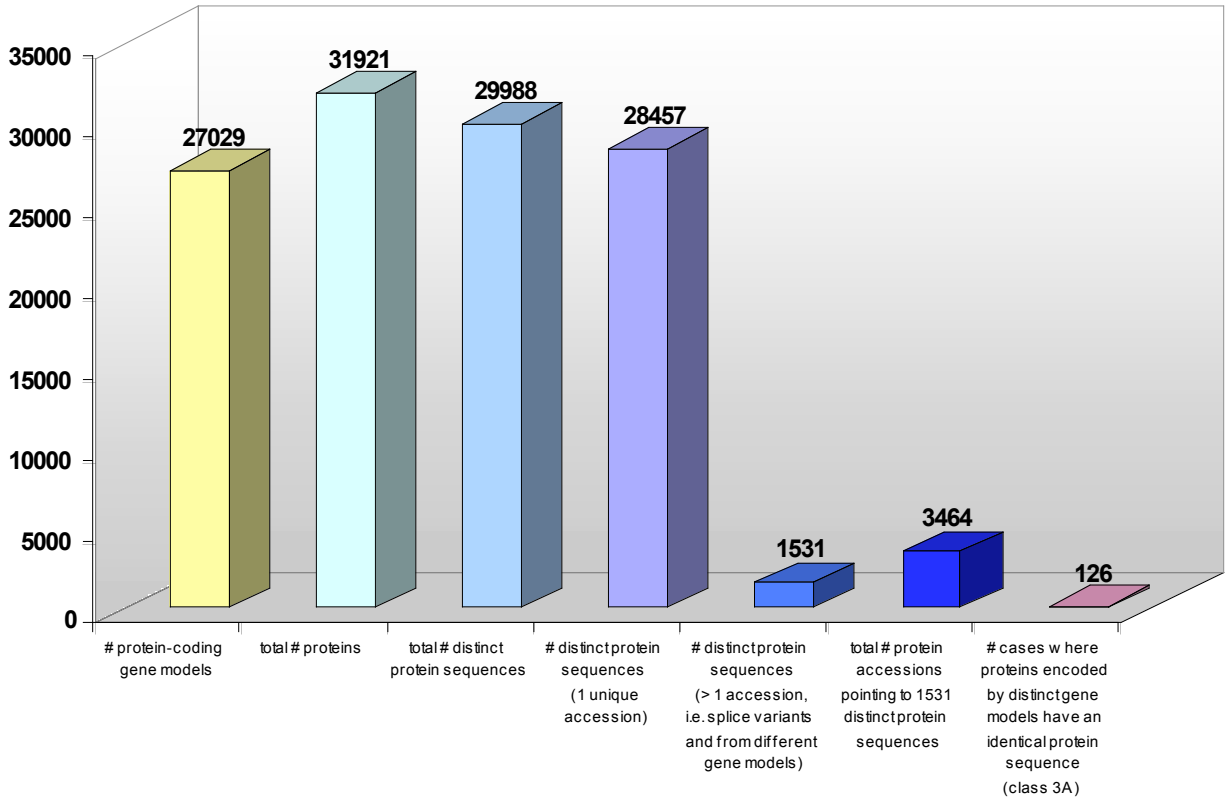
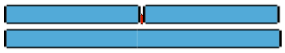


Figure S4. Peptide evidence for identified splice variants

A) Gene model *AT4G16160*

Sequence 1: AT4G16160.1
Length = 176 (1 .. 176)

Sequence 2: AT4G16160.2
Length = 178 (1 .. 178)



```
Score = 309 bits (791), Expect = 9e-83

Identities = 176/178 (98%), Positives = 176/178 (98%), Gaps = 2/178 (1%)

Query  1    MEKSGGRIVMDEIRSFEKAHLFDLGHPLLNRIADSFVKAAGVGALQAVSREAYFTVVDG-  59

            MEKSGGRIVMDEIRSFEKAHLFDLGHPLLNRIADSFVKAAGVGALQAVSREAYFTVVDG

Sbjct  1    MEKSGGRIVMDEIRSFEKAHLFDLGHPLLNRIADSFVKAAGVGALQAVSREAYFTVVDGA  60

Query  60   -FDSNNVGPPSEITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGM  118

             FDSNNVGPPSEITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGM

Sbjct  61   GFDSNNVGPPSEITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGM  120

Query  119  TEVRGGAHDWRNSAVAGALTGAAMAMTTSERTSHEQVVQSALTGAAISTAANLLSSVF  176

            TEVRGGAHDWRNSAVAGALTGAAMAMTTSERTSHEQVVQSALTGAAISTAANLLSSVF

Sbjct  121  TEVRGGAHDWRNSAVAGALTGAAMAMTTSERTSHEQVVQSALTGAAISTAANLLSSVF  178

                                                    (Source: Blast2Seqs, NCBI)
```

Experimental evidence (Table S1):

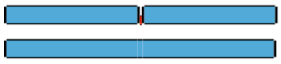
Protein/gene model | spectral count | transcriptomics evidence | 2-D gel evidence | peptide sequence | evidence class

```
AT4G16160.2|006|1|1|EAYFTVVDGAGFDSNNVGPPSEITGNKK|1a

AT4G16160.1|002|1|1|EAYFTVVDGFDSNNVGPPSEITGNKK|1a

AT4G16160|016|1|1|TSHEQVVQSALTGAAISTAANLLSSVF|2

AT4G16160|010|1|1|NSAVAGALTGAAMAMTTSER|2

AT4G16160|004|1|1|AHLFDLGHPLLNR|2

AT4G16160|002|1|1|NTGKESLQWGLAAGLYSGITYGMTEVR|2

AT4G16160|002|1|1|ESLQWGLAAGLYSGITYGMTEVR|2

AT4G16160|002|1|1|AAGVGALQAVSR|2
```

B) Gene model *AT3G05420*

Sequence 1: AT3G05420.1
Length = 668 (1 .. 668)

Sequence 2: AT3G05420.2
Length = 669 (1 .. 669)



```
Score = 1262 bits (3265), Expect = 0.0

Identities = 668/669 (99%), Positives = 668/669 (99%), Gaps = 1/669 (0%)
```

```
Query  1    MAMPRATSGPAYPERFYAAASYVGLDGSDSSAKNVISKFPDDTALLLYALYQQATVGPCN  60
            MAMPRATSGPAYPERFYAAASYVGLDGSDSSAKNVISKFPDDTALLLYALYQQATVGPCN
Sbjct  1    MAMPRATSGPAYPERFYAAASYVGLDGSDSSAKNVISKFPDDTALLLYALYQQATVGPCN  60
Query  61   TPKPSAWRPVEQSKWKSWQGLGTMPSIEAMRLFVKILEEDDPGWYSRASNDIPDPVVDVQ  120
            TPKPSAWRPVEQSKWKSWQGLGTMPSIEAMRLFVKILEEDDPGWYSRASNDIPDPVVDVQ
Sbjct  61   TPKPSAWRPVEQSKWKSWQGLGTMPSIEAMRLFVKILEEDDPGWYSRASNDIPDPVVDVQ  120
Query  121  IN-RAKDEPVVENGSTFSETKTISTENGRLAETQDKDVVSEDSNTVSVYNQWTAPQTSGQ  179
            IN RAKDEPVVENGSTFSETKTISTENGRLAETQDKDVVSEDSNTVSVYNQWTAPQTSGQ
Sbjct  121  INQRAKDEPVVENGSTFSETKTISTENGRLAETQDKDVVSEDSNTVSVYNQWTAPQTSGQ  180
Query  180  RPKARYEHGAAVIQDKMYIYGGNHNGRYLGDLHVLDLKSWTWSRVETKVATESQETSTPT  239
            RPKARYEHGAAVIQDKMYIYGGNHNGRYLGDLHVLDLKSWTWSRVETKVATESQETSTPT
Sbjct  181  RPKARYEHGAAVIQDKMYIYGGNHNGRYLGDLHVLDLKSWTWSRVETKVATESQETSTPT  240
Query  240  LLAPCAGHSLIAWDNKLLSIGGHTKDPSESMQVKVFDPHTITWSMLKTYGKPPVSRGGQS  299
            LLAPCAGHSLIAWDNKLLSIGGHTKDPSESMQVKVFDPHTITWSMLKTYGKPPVSRGGQS
Sbjct  241  LLAPCAGHSLIAWDNKLLSIGGHTKDPSESMQVKVFDPHTITWSMLKTYGKPPVSRGGQS  300
Query  300  VTMVGKTLVIFGGQDAKRSLLNDLHILDLDTMTWDEIDAVGVSPSPRSDHAAAVHAERFL  359
            VTMVGKTLVIFGGQDAKRSLLNDLHILDLDTMTWDEIDAVGVSPSPRSDHAAAVHAERFL
Sbjct  301  VTMVGKTLVIFGGQDAKRSLLNDLHILDLDTMTWDEIDAVGVSPSPRSDHAAAVHAERFL  360
Query  360  LIFGGGSHATCFDDLHVLDLQTMEWSRPAQQGDAPTPRAGHAGVTIGENWFIVGGGDNKS  419
            LIFGGGSHATCFDDLHVLDLQTMEWSRPAQQGDAPTPRAGHAGVTIGENWFIVGGGDNKS
Sbjct  361  LIFGGGSHATCFDDLHVLDLQTMEWSRPAQQGDAPTPRAGHAGVTIGENWFIVGGGDNKS  420
Query  420  GASESVVLNMSTLAWSVVASVQGRVPLASEGLSLVVSSYNGEDVLVAFGGYNGRYNNEIN  479
            GASESVVLNMSTLAWSVVASVQGRVPLASEGLSLVVSSYNGEDVLVAFGGYNGRYNNEIN
Sbjct  421  GASESVVLNMSTLAWSVVASVQGRVPLASEGLSLVVSSYNGEDVLVAFGGYNGRYNNEIN  480
Query  480  LLKPSHKSTLQTKTLEAPLPGSLSAVNNATTRDIESEVEVSQEGRVREIVMDNVNPGSKV  539
            LLKPSHKSTLQTKTLEAPLPGSLSAVNNATTRDIESEVEVSQEGRVREIVMDNVNPGSKV
Sbjct  481  LLKPSHKSTLQTKTLEAPLPGSLSAVNNATTRDIESEVEVSQEGRVREIVMDNVNPGSKV  540
Query  540  EGNSERIIATIKSEKEELEASLNKERMQTLQLRQELGEAELRNTDLYKELQSVRGQLAAE  599
            EGNSERIIATIKSEKEELEASLNKERMQTLQLRQELGEAELRNTDLYKELQSVRGQLAAE
Sbjct  541  EGNSERIIATIKSEKEELEASLNKERMQTLQLRQELGEAELRNTDLYKELQSVRGQLAAE  600
Query  600  QSRCFKLEVDVAELRQKLQTLETLQKELELLQRQKAASEQAAMNAKRQGSGGVWGWLAGS  659
            QSRCFKLEVDVAELRQKLQTLETLQKELELLQRQKAASEQAAMNAKRQGSGGVWGWLAGS
Sbjct  601  QSRCFKLEVDVAELRQKLQTLETLQKELELLQRQKAASEQAAMNAKRQGSGGVWGWLAGS  660
Query  660  PQEKDDDSP  668
            PQEKDDDSP
Sbjct  661  PQEKDDDSP  669
```

(Source: Blast2Seqs, NCBI)

Experimental evidence (Table S1):

Protein/gene model | spectral count | transcriptomics evidence | 2-D gel evidence | peptide sequence | evidence class

AT3G05420.2|001|1|0|**ASNDIPDPVVDVQINQR**|1a

AT3G05420.1|006|1|0|**ASNDIPDPVVDVQINR**|1a

AT3G05420|007|1|0|TLEAPLPGSLSAVNNATTR|2

AT3G05420|005|1|0|VFDPHTITWSMLK|2

AT3G05420|005|1|0|AGHAGVTIGENWFIVGGGDNK|2

AT3G05420|003|1|0|SGASESVVLNMSTLAWSVVASVQGR|2

AT3G05420|003|1|0|FYAAASYVGLDGSDSSAK|2

AT3G05420|002|1|0|ILEEDDPGWYSR|2

AT3G05420|002|1|0|DIESEVEVSQEGR|2

AT3G05420|002|1|0|AKDEPVVENGSTFSETK|2

AT3G05420|001|1|0|YEHGAAVIQDK|2

AT3G05420|001|1|0|VATESQETSTPTLLAPCAGHSLIAWDNK|2

AT3G05420|001|1|0|SEKEELEASLNK|2

AT3G05420|001|1|0|LQTLETLQKELELLQR|2

AT3G05420|001|1|0|EIVMDNVNPGSKVEGNSER|2

AT3G05420|001|1|0|DVVSEDSNTVSVYNQWTAPQTSGQRPK|2

AT3G05420|001|1|0|ARYEHGAAVIQDK|2

AT3G05420|001|1|0|AASEQAAMNAK|2

Figure S5. Spectral and distinct peptide coverage of the 3,467 pollen proteins

The two bar plots show the distribution of the 3,467 proteins with respect to the number of observed spectra as well as their respective distinct peptides. The majority of the 3,467 proteins are implied by two or more distinct (and information-rich) peptides (75%). The remaining 868 proteins, seen with only one peptide, are either identified by several spectra or are manually validated single hits.

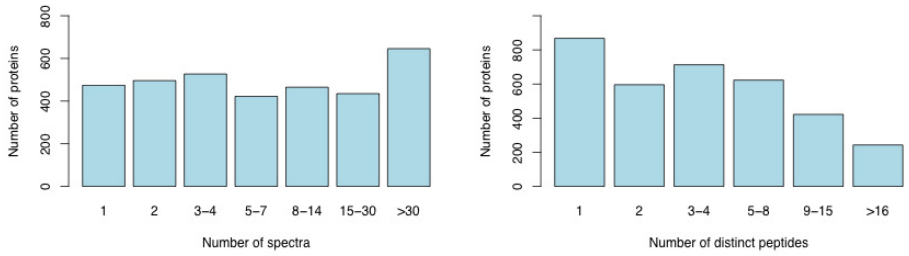


Figure S6. Re-mapping of Affymetrix probesets

Absolute count of the number of oligonucleotides per probeset after re-mapping of the Affymetrix probesets using the Brainarray custom CDF file (version 10) for TAIR7. All oligonucleotides that point to more than one genomic location were eliminated (Dai et al. 2005). 21,022 of the 22,500 ATH1 probesets (93%) passed the Brainarray re-mapping criteria, including 292 gene models that are represented by 3,4,5 oligos/probeset, respectively (shown in light blue in the figure). Since we did not assess differential gene expression but whether a transcript has been called present in previous pollen transcriptomics studies, we did leave these gene models whose gene expression signals would be of "lower quality" in the dataset. This allowed us to integrate unambiguous gene expression evidence of 21,022 gene models with proteomics data from 27,029 protein-coding gene models in TAIR7 (excluding pseudogenes, transposons, etc.), i.e. 78%. We had to acknowledge the fact that we might lose a few truly expressed gene models that are represented by Affymetrix probesets designed to analyze closely related genes.

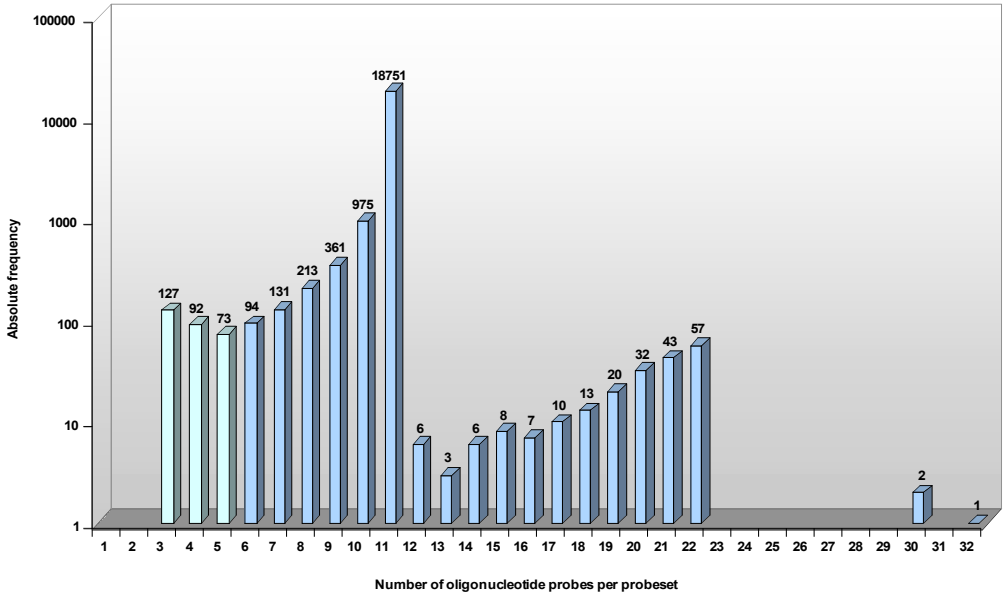
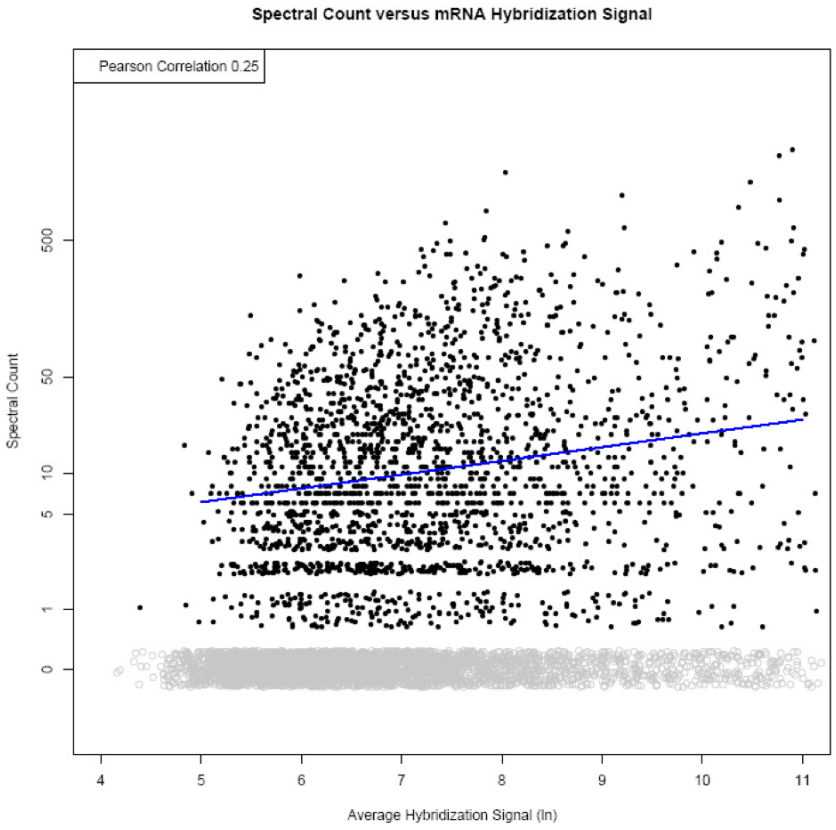
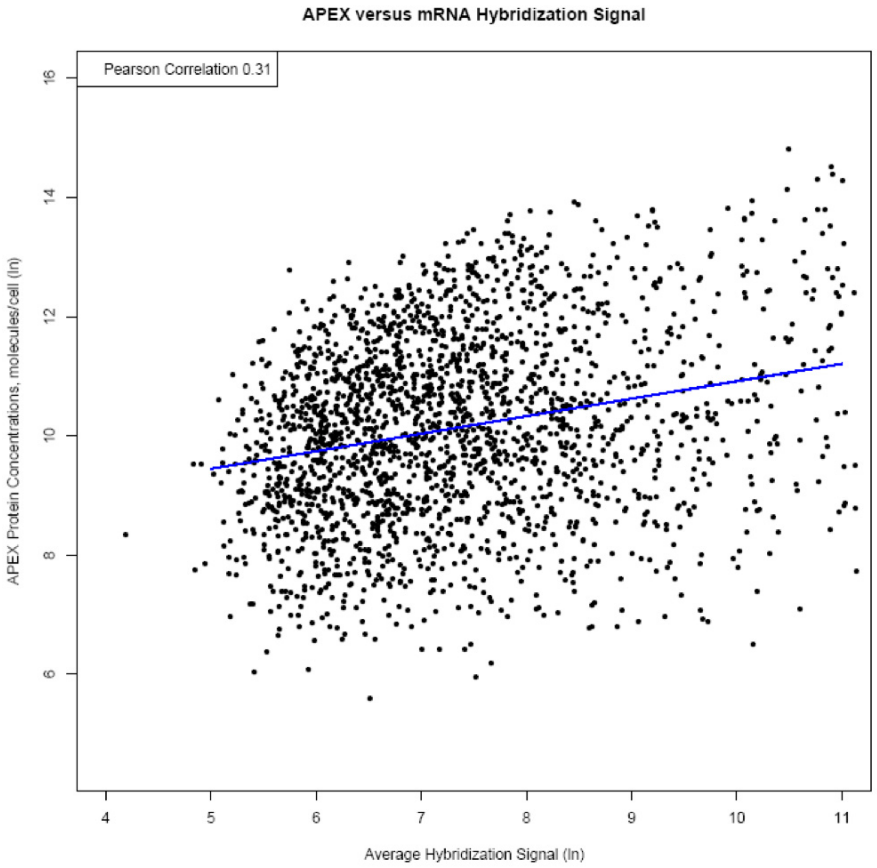


Figure S7: Correlation of transcriptomics and proteomics data

(A) Correlation of the average mRNA hybridization signals of the mature pollen study with three replicates (Schmid et al. 2005) and the spectral count summed up from proteins identified by peptides of class 1a, 1b, and 2 for the respective gene models. Gene models that were expressed but whose encoded proteins were not detected in our proteomics study (spectral count 0) are shown by grey, open circles. For proteins identified by 0, 1, 2, 3, 4 and 5 spectra, some noise was added to better visualize the data points. A linear regression line is fitted to the data (blue line).



(B) Correlation of the average mRNA hybridization signals with the APEX score (protein concentration given as molecules/cell), calculated as described (Lu et al. 2007).



(C) Density of the mean predicted peptide detectability for (i) proteins identified in our study (red line) and (ii) proteins encoded by gene models for which a gene expression signal has been detected by Schmid and colleagues (2005) (black line) and that we did not observe (p-value < 2.2 x10^{-16}).

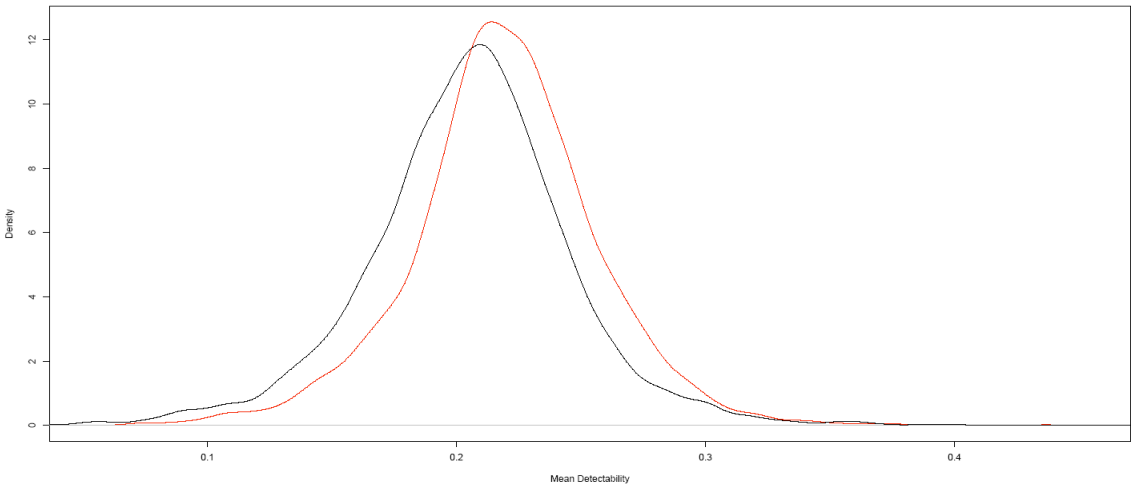


Figure S8. Abundance difference for proteins providing novel insights versus proteins identified by 2-D gel approaches

A scatterplot of the log protein length (x axis) versus log spectral count (y axis) and four boxplots are shown for 266 2-D gel identified proteins (red circles), and the 537 proteins newly identified in our study (blue circles) (red and blue bars represent the median). The 266 proteins observed in the 2D gel studies account for ~7.7% of the mature pollen proteome and have higher spectral counts and hence higher abundances (notice the shift on the x axis), they account for 37% of all spectra. In contrast, the 537 proteins that we imply for the first time in mature pollen account for ~15.5% of the 3,465 proteins; on the spectral count level they reflect only 13.5% of the overall observed spectra.

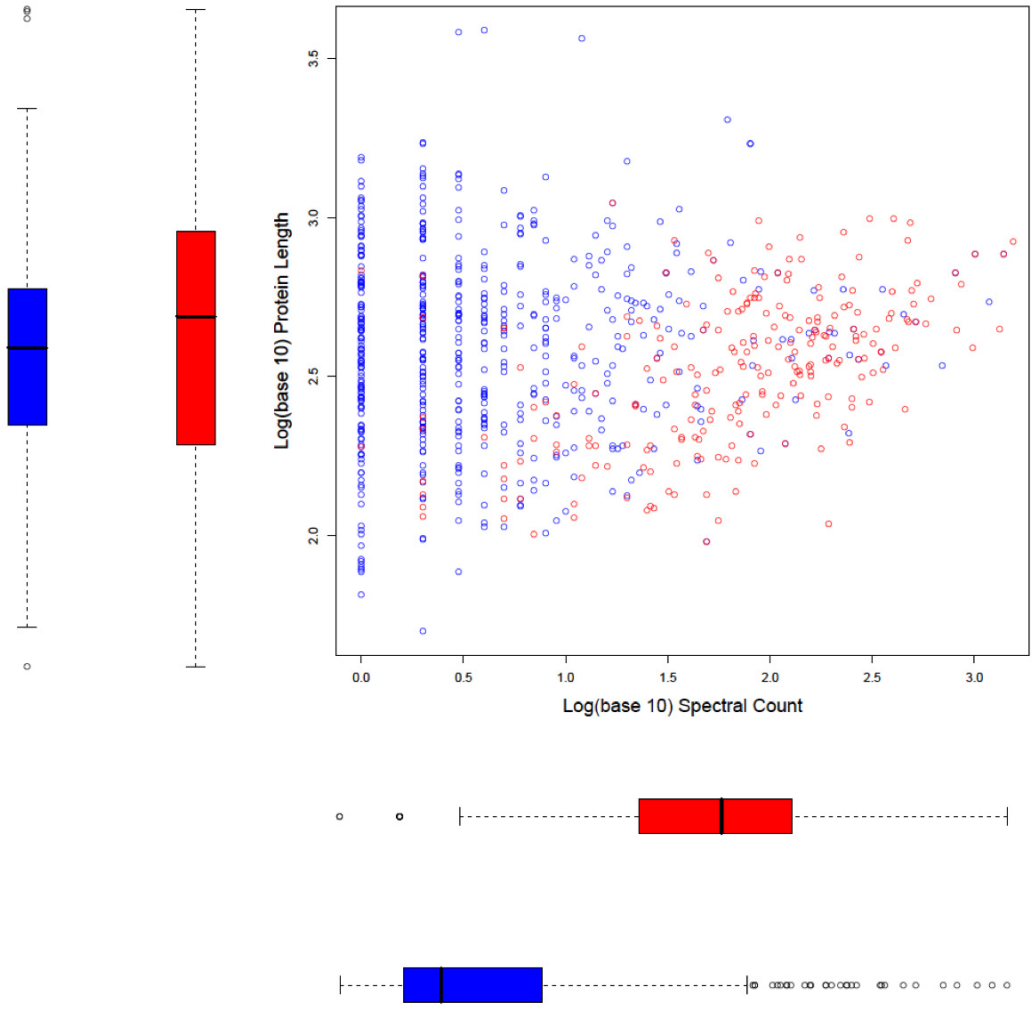


Figure S9. Heatmap representation of the clustering analysis of various *Arabidopsis* tissues

Heatmap representation of the respective distance matrix built upon Jaccard similarities. The darker the color, the less similar two tissues are. The hierarchical clustering proceeds in an agglomerative way and pollen and seed, which are clustered together, are the last to join the dendrogram.

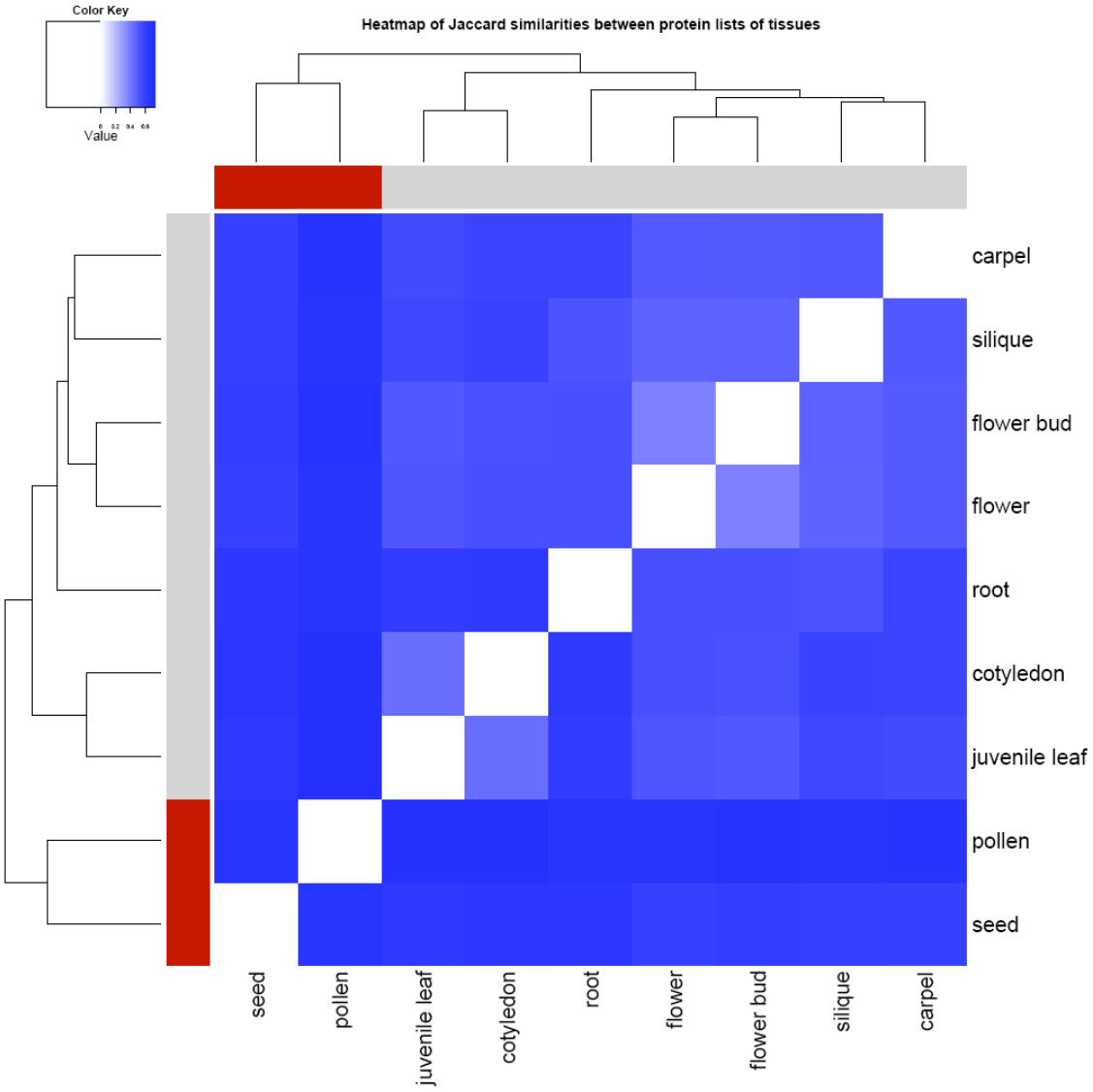
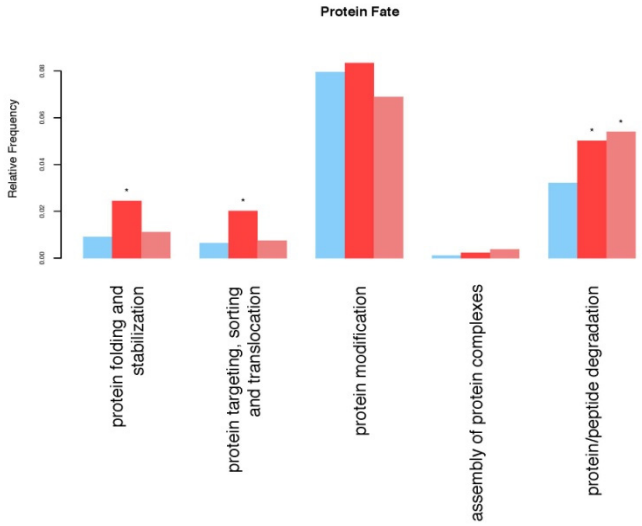


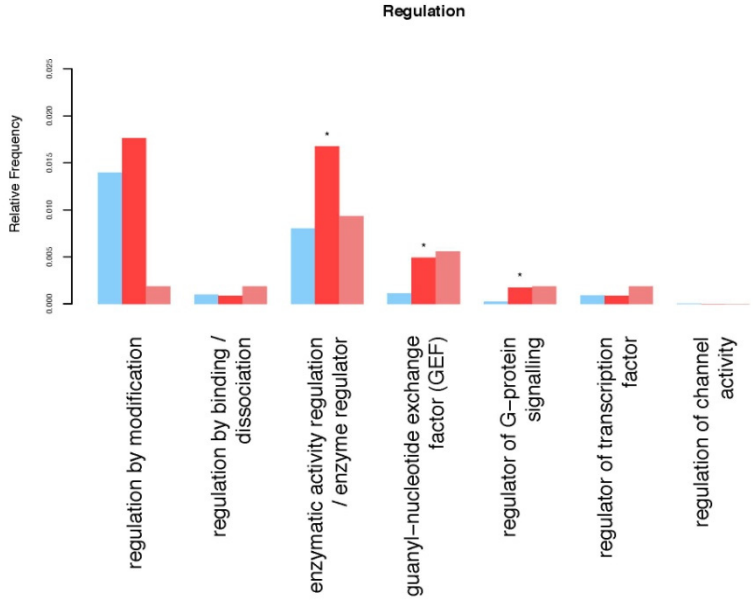
Figure S10. detailed representation of selected FunCat categories.

The subcategories of selected FunCat categories that are overrepresented in the pollen proteome are shown: protein fate (A), regulation of metabolism and protein function (B), cellular transport (C), transcription (D), and development (E).

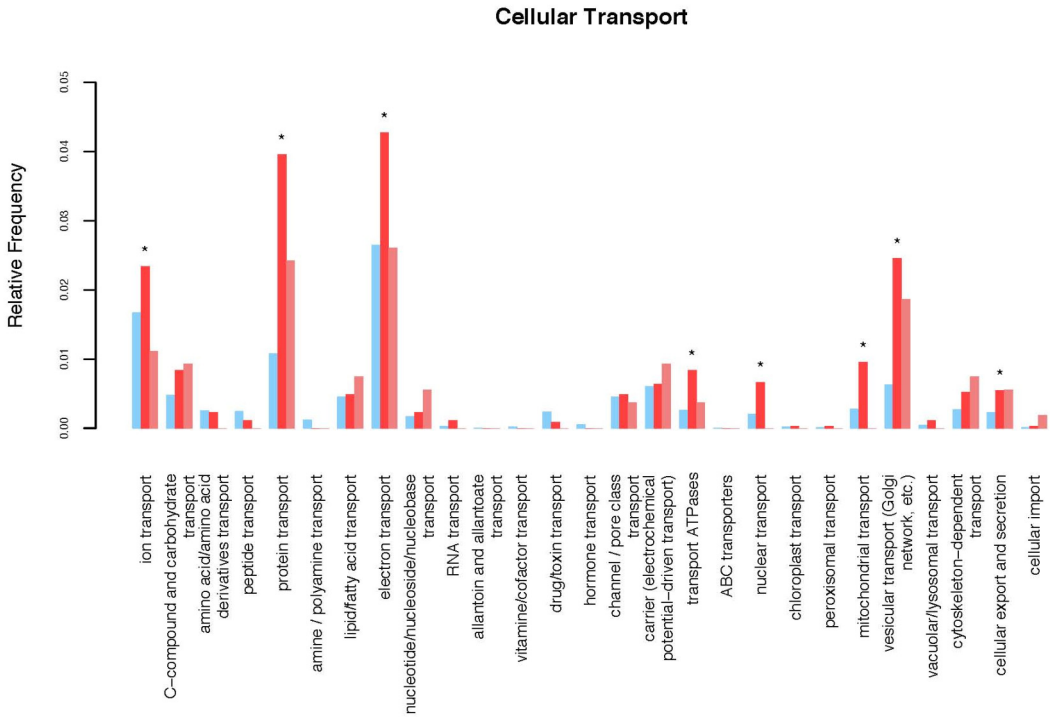
A. Protein Fate



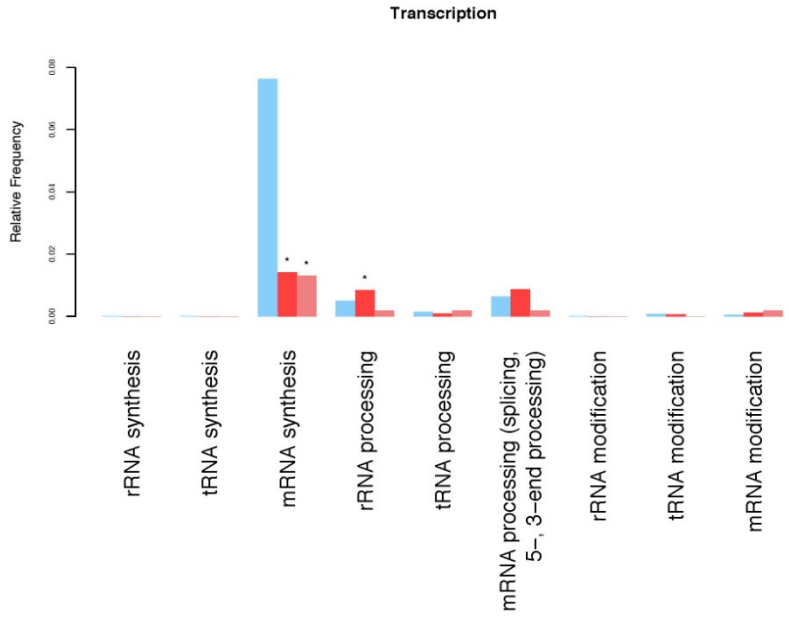
B. Regulation of metabolism and protein function



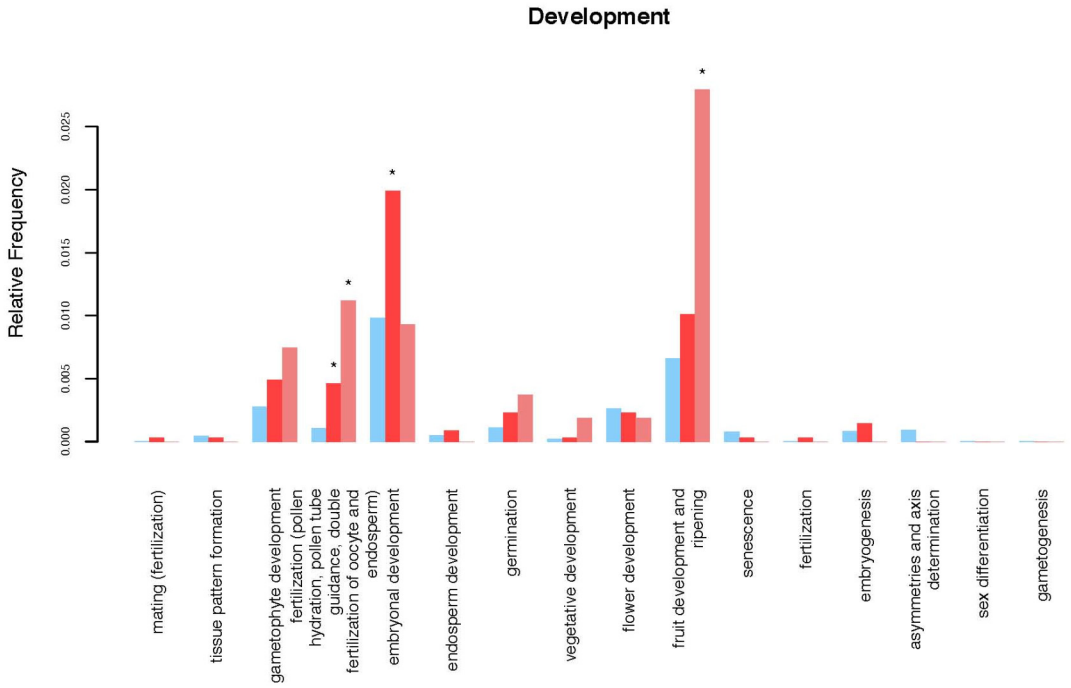
C. Cellular Transport



D. Transcription



E. Development



SUPPLEMENTAL TABLES

Table S1. Overview of the identified proteins and peptides in the mature pollen proteome (provided as a separate Excel file)

For each identified protein we provide the *Arabidopsis* gene model identifier AGI or, for class 2 identifications, the gene model name(s) (column 1), the peptide sequence (column 2) and its respective spectral count (column 3). In addition, we list whether previous transcriptomics data (11,150 genes expressed during pollen development, see Figure 3C) or 2-D gel proteomics evidence has been reported (columns 4, 5), and finally in column 6 we show the evidence class of the respective peptide. Furthermore, we show a minimal list of 132 proteins explaining the class 3b peptides that are not explained by proteins (at least one) of the 3,467 reference data set.

Table S2. Overview of the 28 "2-D Gel approach -exclusive" protein identifications (provided as a separate pdf document)

Our identifiable proteome index indicates that for several of the 28 2-D exclusive protein identifications more than one protein sequence can be encoded by the respective gene model (e.g. *AT1G07930*). For several cases, we have ambiguous peptide evidence (class 3b), e.g. for the gene model *AT1G07920* we recorded 996 class 3b spectra.

Furthermore, for 12 identifications we have recorded no evidence; all of these are contributed by the study by Sheoran et al., 2007. We show the AGI of the gene model identified in the respective 2-D study/studies (columns 1, 2); the number of protein sequences encoded by the gene model and how many protein accessions point to those protein sequences (columns 3, 4); the protein parameters length and isoelectric point ((columns 5, 6); the total number of class 3b spectra we observed (column 7), and candidate protein identifications based on the experimental peptide evidence.

Table S3. Re-analyzed transcriptomics datasets

The data from three transcriptomics studies on mature pollen (Pina et al. 2005; Schmid et al. 2005) or pollen development (Honys and Twell 2004) were assessed using the updated probeset mapping based on TAIR7, and the results are listed in the table below. In each cell, the number of expressed gene models is given. The sum of the gene models that are expressed in 2/2 replicates for the UNM, BCP, TCP pollen developmental stages, as well as in at least 3/6 replicates for mature pollen are shown.

	Honys and Twell, 2004							Pina et al., 2005		Schmid et al., 2005		
Identifier	UNM 1	UNM 2	BCP 1	BCP 2	TCP 1	TCP 2	MPG 1	Pollen A	Pollen B	Pollen 1	Pollen 2	Pollen 3
UNM1	10,511	8,916	9,981	8,831	6,386	5,954	5,445	4,435	4,207	5,665	5,725	5,357
UNM2	8,916	9,225	8,912	8,348	5,948	5,593	5,119	4,062	3,851	5,260	5,288	4,967
BCP1	9,981	8,912	11,243	9,142	6,855	6,420	5,837	4,832	4,617	6,111	6,180	5,787
BCP2	8,831	8,348	9,142	9,352	6,281	5,915	5,387	4,293	4,104	5,530	5,582	5,240
TCP1	6,386	5,948	6,855	6,281	7,450	6,486	5,815	4,716	4,540	5,935	5,877	5,612
TCP2	5,954	5,593	6,420	5,915	6,486	6,959	5,691	4,619	4,439	5,791	5,694	5,486
MPG1	5,445	5,119	5,837	5,387	5,815	5,691	6,434	4,572	4,428	5,576	5,557	5,411
Pollen A	4,435	4,062	4,832	4,293	4,716	4,619	4,572	5,650	4,808	4,978	5,012	4,900
Pollen B	4,207	3,851	4,617	4,104	4,540	4,439	4,428	4,808	5,391	4,801	4,842	4,728
Pollen 1	5,665	5,260	6,111	5,530	5,935	5,791	5,576	4,978	4,801	6,913	6,129	5,950
Pollen 2	5,725	5,288	6,180	5,582	5,877	5,694	5,557	5,012	4,842	6,129	7,035	6,030
Pollen 3	5,357	4,967	5,787	5,240	5,612	5,486	5,411	4,900	4,728	5,950	6,030	6,576
Sum	**8,915**		**9,141**		**6,485**		**6,708 (present in at least 3/6)**					
Total 1	**10,637**						**6,708**					
Total 2	**11,150**											

Table S4. List of unambiguously identified gene models expressed in mature pollen and during all stages of pollen development
(provided as a separate Excel file)

Based on the integration of our proteomics data with previous transcriptomics data using TAIR7 as common reference point, we provide a list of 7,901 gene models expressed in mature pollen (column 1; covered by the green and red data sets in Figure 3C), and a list of 11,687 gene models expressed throughout pollen development (column 2; covered by the green, red and blue data sets in Figure 3C).

Table S5. Overview of genes that, when mutated, have an observed phenotype during pollen development

We show genes that when mutated are known to affect late and early stages of pollen development. We show the gene name, the *Arabidopsis* gene identifier (AGI), the respective literature reference, the assigned evidence class based on the experimentally observed peptides, and the spectral count for the unambiguously identified proteins (or gene model, as in the case of ATAPT1, AT1G27450, which encodes 2 different protein sequences) is shown. The 5 AGIs highlighted in bold with a grey background are those observed as interactors of the DUF1216 family in the AraNet interactor analysis.

1. Genes affecting late stages of pollen development				
Gene Name	**AGI**	**Reference**	**Peptide Evidence Class**	**Spectral Count**
ACA9	*AT3G21180*	(Schiøtt et al. 2004)	1a	77
ACT12	*AT3G46520*	(Huang et al. 1996)	1a	132
ACT4	*AT5G59370*	(Huang et al. 1996)	1b	67
ADF10	*AT5G52360*	(Ruzicka et al. 2007)	1a	32
ADF11	*AT1G01750*	(Ruzicka et al. 2007)	-	0
ADF7	*AT4G25590*	(Ruzicka et al. 2007)	1a	68
ADF8	*AT4G00680*	(Ruzicka et al. 2007)	-	0
AGL18	*AT3G57390*	(Verelst et al. 2007)	-	0
AGL29	*AT2G34440*	(Verelst et al. 2007)	-	0
AHA3	*AT5G57350*	(Robertson et al. 2004)	-	0
AMC/PEX13	*AT3G07560*	(Boisson-Dernier et al. 2008)	-	0
ATAPT1	*AT1G27450*	(Zhang et al. 2002)	2	30
ATAPY1	*AT3G04080*	(Wolf et al. 2007)	1a	14
ATAPY2	*AT5G18280*	(Wolf et al. 2007)	1a	7
AtBECLIN1	*AT3G61710*	(Qin et al. 2007)	-	0
AtIPK2a	*AT5G07370*	(Xu et al. 2005)	1b	2
AtMGH3	*AT1G19890*	(Okada et al. 2005)	-	0
AtMYB32	*AT4G34990*	(Preston et al. 2004)	-	0
AtMYB4	*AT4G38620*	(Preston et al. 2004)	-	0
AtPMEI1	*AT1G48020*	(Wolf et al. 2003)	1a	9

AtPPME1	*AT1G69940*	(Tian et al. 2006)	1a	28
AtPTEN1	*AT5G39400*	(Gupta et al. 2002)	-	0
AtSTP11	*AT5G23270*	(Schneidereit et al. 2005)	-	0
AtSTP2	*AT1G07340*	(Truernit et al. 1999)	1a	6
AtSTP4	*AT3G19930*	(Truernit et al. 1996)	1a	232
AtSTP6	*AT3G05960*	(Scholz-Starke et al. 2003)	1a	70
AtSTP9	*AT1G50310*	(Schneidereit et al. 2003)	1a	282
AtSUC1	*AT1G71880*	(Stadler et al. 1999)	1a	69
AtTAF6	*AT1G04950*	(Lagoa et al. 2005)	-	0
AtTAF6b	*AT1G54360*	(Lagoa et al. 2005)	-	0
AtUSP	*AT5G52560*	(Schnurr et al. 2006)	1a	27
CalS5	*AT2G13680*	(Dong et al. 2005; Nishikawa et al. 2005)	1a	140
CAP1	*AT4G34490*	(Deeks et al. 2007)	1a	50
CER3	*AT5G02310*	(Hulskamp et al. 1995)	1a	1
CER6/POP1/CUT1	*AT1G68530*	(Fiebig et al. 2000 ; Hulskamp et al. 1995; Millar et al. 1999)	-	0
CESA1	*AT4G32410*	(Persson et al. 2007)	-	0
CESA3	*AT5G05170*	(Persson et al. 2007)	-	0
CESA6	*AT5G64740*	(Persson et al. 2007)	-	0
CNGC18	*AT5G14870*	(Frietsch et al. 2007)	-	0
COMATOSE	*AT4G39850*	(Footitt et al. 2007)	1a	42
GPT1	*AT5G54800*	(Niewiadomski et al. 2005)	1a	16
GRP17	*AT5G07530*	(Mayfield and Preuss 2000)	1a	1,181
HAP2/GCS1	*AT4G11720*	(von Besser et al. 2006 ; Mori et al. 2006)	-	0
IDH	*AT4G35650*	(Lemaitre and Hodges 2006)	1a	39
KIP	*AT5G49680*	(Procissi et al. 2003)	1a	50
LCB1	*AT4G36480*	(Teng et al. 2008)	1b	9
LCB2a	*AT3G48780*	(Dietrich et al. 2008; Teng et al. 2008)	1a	4
LCB2c	*AT5G23670*	(Dietrich et al. 2008; Teng et al. 2008)	1b	3

MIA	*AT5G23630*	(Jakobsen et al. 2005)	1a	18
MOR1/GEM1	*AT2G35630*	(Lalanne and Twell 2002)	1a	7
NMNAT	*AT5G55810*	(Hashida et al. 2007)	-	0
NPG1	*AT2G43040*	(Golovkin and Reddy 2003)	1a	20
OLEO1	*AT2G25890*	(Twell et al. 1990)	1a	10
OSB1	*AT1G47720*	(Zaegel et al. 2006)	-	0
POP2	*AT3G22200*	(Palanivelu et al. 2003; Wilhelmi and Preuss 1999)	1a	307
RGP1	*AT3G02230*	(Drakakaki et al. 2006)	1a	271
RGP2	*AT5G15650*	(Drakakaki et al. 2006)	1a	194
RGP5	*AT5G16510*	(Drakakaki et al. 2006)	1b	213
ROPGEF1	*AT4G38430*	(Gu et al. 2006)	-	0
ROPGEF2	*AT1G01700*	(Gu et al. 2006)	-	0
ROPGEF4	*AT2G45890*	(Gu et al. 2006)	-	0
ROPGEF5	*AT5G05940*	(Gu et al. 2006)	1a	1
SDH1-1	*AT5G66760*	(León et al. 2007)	1a	121
SEC8	*AT3G10380*	(Cole et al. 2005)	1a	11
SPIK	*AT2G25600*	(Mouline et al. 2002)	1a	19
SYP21	*AT5G16830*	(Sanderfoot et al. 2001)	1a	1
SYP22	*AT5G46860*	(Sanderfoot et al. 2001)	1a	15
SYP41	*AT5G26980*	(Sanderfoot et al. 2001)	-	0
SYP42	*AT4G02195*	(Sanderfoot et al. 2001)	1a	2
TPLATE	*AT3G01780*	(Van Damme et al. 2006)	1a	14
TUA1	*AT1G64740*	(Carpenter et al. 1992)	1a	37
UGE2	*AT4G23920*	(Rosti et al. 2007)	1a	4
UGE3	*AT1G63180*	(Rosti et al. 2007)	1a	12
VACUOLELESS 1	*AT2G38020*	(Hicks et al. 2004)	1a	4
VANGUARD1	*AT2G47040*	(Jiang et al. 2005)	1a	355
VHA-A	*AT1G78900*	(Dettmer et al. 2005)	1b	525
VHA-E2	*AT3G08560*	(Strompen et al. 2005)	1a	64
AtCDC48A	*AT3G09840*	(Park et al. 2008)	1a	99

AtCDC48B	*AT3G53230*	(Park et al. 2008)	1a	5
AtCDC48C	*AT5G03340*	(Park et al. 2008)	1a	6
AGP6	*AT5G14380*	(Levitin et al. 2008)	-	0
AGP11	*AT3G01700*	(Levitin et al. 2008)	-	0
POK	*AT1G71270*	(Guermonprez et al. 2008; Lobstein et al. 2004)	1a	4
AtKAS2	*AT1G74960*	(Hakozaki et al. 2008)	1b	1
AtKDSA1	*AT1G79500*	(Delmas et al. 2003; Delmas et al. 2008; Matsuura et al. 2003)	1b	9
AtKDSA2	*AT1G16340*	(Delmas et al. 2003; Delmas et al. 2008; Matsuura et al. 2003)	1b	2
AtKDSB	*AT1G53000*	Royo et al. 2000	1a	16
AtKDTA	*AT5G03770*	(Delmas et al. 2003; Delmas et al. 2008)	-	0
PUR4	*AT1G74260*	(Berthomé et al. 2008)	1a	22
ARGAH1	*AT4G08900*	(Brownfield et al. 2008)	1a	12
SGC	*AT3G53810*	(Wan et al. 2008)	-	0
ADLC1	*AT1G14830*	(Kang et al. 2003)	1a	56
RPA	*AT2G35210*	(Song et al. 2006)	1a	13
SDG4	*AT4G30860*	(Cartagena et al. 2008)	-	0
MIRO1/EMB2473	*AT5G27540*	(Yamaoka and Leaver 2008)	1b	34
MIRO2	*AT3G63150*	(Yamaoka and Leaver 2008)	1a	15
MIRO3	*AT3G05310*	(Yamaoka and Leaver 2008)	-	0
AtGSL8	*AT2G36850*	(Toller et al. 2008)	1a	6
AtGSL10	*AT3G07160*	(Toller et al. 2008)	1a	4
ER-ANT1	*AT5G17400*	(Leroch et al. 2008)	-	0
2. Genes affecting early stages of pollen development				
AtMYB103	*AT5G56110*	(Zhang et al. 2007)	-	0
BRAHMA	*AT2G46020*	(Hurtado et al. 2006)	-	0
CDKA1	*AT3G48750*	(Iwakawa et al. 2006)	-	0
MS1	*AT5G22260*	(Ito et al. 2007)	-	0

MS5	*AT4G20900*	(Glover et al. 1998; Taylor et al. 1998)	-	0
DUET/MMD1	*AT1G66170*	(Reddy et al. 2003; Yang et al. 2003)	-	0
DUO1	*AT3G60460*	(Rotman et al. 2005)	-	0
RHF1a	*AT4G14220*	(Liu et al. 2008)	-	0
RHF2a	*AT5G22000*	(Liu et al. 2008)	-	0
ASK1	*AT5G42190*	(Wang and Yang 2006)	-	0
UBP3	*AT4G39910*	(Doelling et al. 2007)	-	0
UBP4	*AT2G22310*	(Doelling et al. 2007)	-	0
KINESIN-12A/PAKRP1	*AT4G14150*	(Lee et al. 2007)	-	0
KINESIN-12B/PAKRP1L	*AT3G23670*	(Lee et al. 2007)	-	0
NEF1	*AT5G13390*	(Ariizumi et al. 2004)	-	0
GSL1	*AT4G04970*	(Enns et al. 2005)	-	0
GSL5	*AT4G03550*	(Enns et al. 2005)	-	0
SERK1	*AT1G71830*	(Colcombet et al. 2005)	-	0
SERK2	*AT1G34210*	(Colcombet et al. 2005)	-	0
SOL1/TSO-like	*AT3G22760*	(Andersen et al. 2007)	-	0
TETRASPORE	*AT3G43210*	(Spielman et al. 1997; Yang et al. 2003)	-	0
QRT1	*AT5G55590*	(Francis et al. 2006; Preuss et al. 1994)	-	0
QRT3	*AT4G20050*	(Preuss et al. 1994; Rhee et al. 2003; Rhee and Somerville 1998)	1b	1
TDF1/AtMYB35	*AT3G28470*	(Zhu et al. 2008)	-	0
TSO1	*AT3G22780*	(Andersen et al. 2007)	-	0
AtTSM1/CCoAOMT	*AT1G67990*	(Fellenberg et al. 2008)	-	0
RPG1	*AT5G40260*	(Guan et al. 2008)	-	0

Table S6. Percentage of selectively expressed gene models identified in various *Arabidopsis* tissues

Percentage of tissue-restricted protein expression of the mature pollen proteome compared to gene model identifications from 8 different tissues previously published in an *Arabidopsis* tissue-index (Baerenfaller et al. 2008); the identifications made in cell suspension culture were not considered. The percentages listed are based on the comparison of these two large scale *Arabidopsis* shotgun proteomics datasets.

Arabidopsis **Tissue**	**Number of experimentally identified gene models**	**Number of gene models with restricted expression**	**Percentage of gene models with restricted expression**
Root	6,125	1,085	17.7
Cotyledon	3,665	217	5.9
Juvenile Leaf	3,892	231	5.9
Carpel	3,946	200	5.1
Flower	5,215	401	7.7
Flower Bud	5,104	323	6.3
Silique	5,779	630	10.9
Seed	3,789	516	13.6
Pollen	3,465	588	17.0

Table S7: Number of gene models underlying the 1st level of FunCat Functional Categories represented in Figure 6.

Functional Category	Gene Models TAIR7	Gene Models Pollen	Gene Models Pollen Exclusive
METABOLISM	4,830	1,028	127
ENERGY	450	194	21
STORAGE PROTEIN	65	11	5
CELL CYCLE AND DNA PROCESSING	1,449	72	8
TRANSCRIPTION	2,493	107	12
PROTEIN SYNTHESIS	628	241	25
PROTEIN FATE	3,074	558	66
REGULATION OF METABOLISM AND PROTEIN FUNCTION	603	126	11
CELLULAR TRANSPORT	2,405	542	59
CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION MECHANISM	1,238	154	15
CELL RESCUE, DEFENCE AND VIRULENCE	1,350	219	18
INTERACTION WITH THE ENVIRONMENT	1,453	211	19
SYSTEMIC INTERACTION WITH THE ENVIRONMENT	663	55	2
CELL FATE	378	55	12
DEVELOPMENT	835	160	32
BIOGENESIS OF CELLULAR COMPONENTS	1,515	129	23
UNCLASSIFIED PROTEINS	7,952	584	121

Table S8. Summary of selected over-represented PFAM domains in pollen

We show the PFAM domain, the respective domain description and accession, the hits observed in the pollen proteome versus those observed in all distinct TAIR7 proteins, and finally the p-value for the Fisher's Exact test of the over-represented domains before multiple testing correction and after multiple testing correction.

PFAM Domain Name	PFAM Domain Description	PFAM Accession	Hits Pollen	Hits TAIR7	p-value before correction	p-value after correction
Proteasome	Proteasome A-type and B-type	PF00227.17	22	29	1.58×10^{-15}	1.03×10^{-13}
Ras	Ras family	PF00071.13	41	100	5.83×10^{-14}	3.73×10^{-12}
Miro	Miro-like protein	PF08477.4	42	107	1.72×10^{-13}	1.09×10^{-11}
Epimerase	NAD dependent epimerase/dehydratase family	PF01370.12	41	103	1.91×10^{-13}	1.19×10^{-11}
RmlD_sub_bind	RmlD substrate binding domain	PF04321.8	21	31	2.77×10^{-13}	1.69×10^{-11}
HEAT	HEAT repeat	PF02985.13	38	102	1.57×10^{-11}	9.43×10^{-10}
HSP70	Hsp70 protein	PF00012.11	16	22	3.71×10^{-11}	2.19×10^{-9}
NAD_binding_4	Male sterility protein	PF07993.3	31	83	1.06×10^{-9}	6.12×10^{-8}
EMP70	Endomembrane protein 70	PF02990.7	11	13	3.03×10^{-9}	1.73×10^{-7}
Cpn60_TCP1	TCP-1/cpn60 chaperonin family	PF00118.15	14	21	3.91×10^{-9}	2.19×10^{-7}
DUF1216	Protein of unknown function (DUF1216)	PF06746.2	10	11	4.15×10^{-9}	2.28×10^{-7}
Hydrolase	haloacid dehalogenase-like hydrolase	PF00702.17	28	74	4.76×10^{-09}	2.57×10^{-07}
Mito_carr	Mitochondrial carrier protein	PF00153.18	25	62	6.85×10^{-9}	3.63×10^{-7}
Sec7	Sec7 domain	PF01369.11	8	8	3.17×10^{-8}	1.65×10^{-6}
UBA	UBA/TS-N domain	PF00627.22	19	43	7.58×10^{-8}	3.86×10^{-6}
Adaptin_N	Adaptin N terminal region	PF01602.11	10	14	2.71×10^{-7}	1.36×10^{-5}
Cation_ATPase_N	Cation transporter/ATPase,	PF00690.17	13	26	1.55×10^{-6}	7.61×10^{-5}

	N-terminus					
E1-E2_ATPase	E1-E2 ATPase	PF00122.11	15	37	6.59×10^{-6}	0.0003
SNARE	SNARE domain	PF05739.10	14	34	1.08×10^{-5}	0.0005
Porin_3	Eukaryotic porin	PF01459.13	6	7	1.50×10^{-5}	0.0007
PABP	Poly-adenylate binding protein, unique domain	PF00658.9	6	7	1.50×10^{-5}	0.0007
Sec23_helical	Sec23/Sec24 helical domain	PF04815.6	8	13	2.38×10^{-5}	0.0010
Sec23_trunk	Sec23/Sec24 trunk domain	PF04811.6	8	13	2.38×10^{-5}	0.0010
zf-Sec23_Sec24	Sec23/Sec24 zinc finger	PF04810.6	8	13	2.38×10^{-5}	0.0010
ArfGap	Putative GTPase activating protein for Arf	PF01412.9	10	21	4.43×10^{-5}	0.0018
Tubulin	Tubulin/FtsZ family, GTPase domain	PF00091.16	9	18	6.66×10^{-5}	0.0027
Tubulin_C	Tubulin/FtsZ family, C-terminal domain	PF03953.8	9	18	6.66×10^{-5}	0.0027
Cation_ATPase_C	Cation transporting ATPase, C-terminus	PF00689.12	8	15	9.60×10^{-5}	0.0036
IBN_N	Importin-beta N-terminal domain	PF03810.10	8	15	9.60×10^{-5}	0.0036
Coatomer_WDAD	Coatomer WD associated region	PF04053.5	5	6	0.0001	0.0040
Sec23_BS	Sec23/Sec24 beta-sandwich domain	PF08033.3	7	12	0.0001	0.0045
C2	C2 domain	PF00168.21	25	103	0.0002	0.0078
Methyltransf_11	Methyltransferase domain	PF08241.3	23	93	0.0003	0.0096
Oleosin	Oleosin	PF01277.8	9	21	0.0003	0.0096
Metallophos	Calcineurin-like phosphoesterase	PF00149.19	21	82	0.0003	0.0100
Dynamin_M	Dynamin central region	PF01031.11	7	14	0.0004	0.0135
Profilin	Profilin	PF00235.10	4	5	0.0008	0.0235
Cpn10	Chaperonin 10 Kd	PF00166.12	4	5	0.0008	0.0235

	subunit					
Arf	ADP-ribosylation factor family	PF00025.12	9	24	0.0009	0.0235
EF1_GNE	EF-1 guanine nucleotide exchange domain	PF00736.10	4	5	0.0008	0.0235
Glyco_hydro_42	Beta-galactosidase	PF02449.6	5	8	0.0009	0.0235
WD40	WD domain, G-beta repeat	PF00400.23	48	265	0.0011	0.0260
TPR_2	Tetratricopeptide repeat	PF07719.8	32	159	0.0012	0.0281
GTP_EFTU_D2	Elongation factor Tu domain 2	PF03144.16	9	25	0.0013	0.0289
GTP_EFTU	Elongation factor Tu GTP binding domain	PF00009.18	10	30	0.0014	0.0297
Syntaxin	Syntaxin	PF00804.16	8	21	0.0016	0.0309
Rho_GDI	RHO protein GDP dissociation inhibitor	PF02115.8	3	3	0.0015	0.0309
HSP90	Hsp90 protein	PF00183.9	5	9	0.0017	0.0312
Sec1	Sec1 family	PF00995.14	4	6	0.0022	0.0336
Alpha_adaptinC2	Adaptin C-terminal domain	PF02883.11	4	6	0.0022	0.0336
PMEI	Plant invertase/pectin methylesterase inhibitor	PF04043.6	24	112	0.0020	0.0336
Myosin_N	Myosin N-terminal SH3-like domain	PF02736.10	6	13	0.0020	0.0336
Myosin_head	Myosin head (motor domain)	PF00063.12	7	18	0.0027	0.0356
Dynamin_N	Dynamin family	PF00350.14	8	23	0.0031	0.0375
Sec63	Sec63 Brl domain	PF02889.7	4	7	0.0047	0.0514
Calreticulin	Calreticulin family	PF00262.9	4	7	0.0047	0.0514
GDI	GDP dissociation inhibitor	PF00996.9	3	5	0.0129	0.0666
ubiquitin	Ubiquitin family	PF00240.14	15	66	0.0074	0.0666
efhand	EF hand	PF00036.23	31	171	0.0074	0.0666
Clat_adaptor_	Clathrin adaptor	PF01217.11	5	12	0.0081	0.0666

s	complex small chain					
Pectinesterase	Pectinesterase	PF01095.10	14	67	0.0196	0.0666
Synaptobrevin	Synaptobrevin	PF00957.12	6	17	0.0095	0.0666
DnaJ_C	DnaJ C terminal region	PF01556.9	7	22	0.0097	0.0666
Glyco_hydro_32C	Glycosyl hydrolases family 32 C terminal	PF08244.3	4	9	0.0139	0.0666
PRONE	PRONE (Plant-specific Rop nucleotide exchanger)	PF03759.4	5	14	0.0168	0.0666

Table S9. Selected protein families in pollen (provided as a separate Excel file)

Table S10: Overview of the protein family containing DUF1216 domains

The domain of unknown function, DUF1216, is a conserved protein domain within *Arabidopsis thaliana*. The family of DUF1216 proteins has members that contain one or two copies of the domain. In the table, we show the protein accession number, the peptide evidence class, the spectral count, the rank in the list of identified proteins based on spectral count, and whether transcriptomics evidence has been observed in early (UNM, BCP, TCP) or mature pollen stages (MPG).

For AT5G48575.1, no probeset is present on the Affymetrix ATH1 array. On the other hand, one common probeset (256580_s_at) represents both AT3G28810 and AT3G28820. During re-mapping this ambiguous probeset was eliminated. Our proteomics data, however, provide unambiguous peptide evidence that both proteins encoded by these two distinct gene models are expressed in mature pollen.

Protein Accession	Protein Class	Spectral Count	Rank by Spectral Count	Transcriptomics Data (early)	Transcriptomics Data (late)
AT3G28770.1	1a	106	227	1	1
AT3G28780.1	1a	300	56	1	1
AT3G28790.1	1a	2,075	2	1	1
AT3G28810.1	1a	207	99	0	0
AT3G28820.1	1a	155	142	0	0
AT3G28830.1	1a	2,284	1	1	1
AT3G28840.1	1a	364	46	1	1
AT3G28980.1	1a	1,331	5	1	1
AT5G39870.1	1a	27	691	1	1
AT5G48575.1		0		0	0
AT5G61720.1	1a	981	9	1	1

References

Andersen, S.U., Algreen-Petersen, R.G., Hoedl, M., Jurkiewicz, A., Cvitanich, C., Braunschweig, U., Schauser, L., Oh, S.A., Twell, D., and Jensen, E.O. 2007. The conserved cysteine-rich domain of a tesmin/TSO1-like protein binds zinc *in vitro* and TSO1 is required for both male and female fertility in *Arabidopsis thaliana*. *J Exp Bot* **58:** 3657-3670.

Ariizumi, T., Hatakeyama, K., Hinata, K., Inatsugi, R., Nishida, I., Sato, S., Kato, T., Tabata, S., and Toriyama, K. 2004. Disruption of the novel plant protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility in *Arabidopsis thaliana*. *The Plant Journal* **39** 170-181.

Baerenfaller, K., Grossmann, J., Grobei, M.A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., Zimmermann, P., Grossniklaus, U., Gruissem, W., and Baginsky, S. 2008. Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science* **320:** 938-941.

Berthomé, R., Thomasset, M., Maene, M., Bourgeois, N., Froger, N., and Budar, F. 2008. *pur4* mutations are lethal to the male, but not the female, gametophyte and affect sporophyte development in *Arabidopsis*. *Plant Physiology* **147:** 650-660.

Boisson-Dernier, A., Frietsch, S., Kim, T.H., Dizon, M.B., and Schroeder, J.I. 2008. The peroxin loss-of-function mutation *abstinence by mutual consent* disrupts male-female gametophyte recognition. *Curr Biol* **18:** 63-68.

Brownfield, D.L., Todd, C.D., and Deyholos, M.K. 2008. Analysis of *Arabidopsis* arginase gene transcription patterns indicates specific biological functions for recently diverged paralogs *Plant Mol Biol* **67:** 429-440.

Brunner, E., Ahrens, C.H., Mohanty, S., Baetschmann, H., Loevenich, S., Potthast, F., Deutsch, E.W., Panse, C., de Lichtenberg, U., Rinner, O. et al. 2007. A high-quality catalog of the *Drosophila melanogaster* proteome. *Nat Biotechnol* **25:** 576-583.

Carpenter, J.L., Ploense, S.E., Snustad, D.P., and Silfloway, C.D. 1992. Preferential expression of an alpha-tubulin gene of *Arabidopsis* in pollen. *The Plant Cell* **4:** 557-571.

Cartagena, J.A., Matsunaga, S., Seki, M., Kurihara, D., Yokoyama, M., Shinozaki, K., Fujimoto, S., Azumi, Y., Uchiyama, S., and Fukui, K. 2008. The *Arabidopsis* SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen. *Developmental Biology* **315:** 355–368.

Colcombet, J., Boisson-Dernier, A., Ros-Palau, R., Vera, C.E., and Schroeder, J.I. 2005. *Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASES1* and *2* are essential for tapetum development and microspore maturation. *The Plant Cell* **17:** 3350-3361.

Cole, R.A., Synek, L., Zarsky, V., and Fowler, J.E. 2005. SEC8, a subunit of the putative *Arabidopsis* Exocyst complex, facilitates pollen germination and competitive pollen tube growth. *Plant Physiology* **138:** 2005-2018.

Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H. et al. 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* **33:** e175.

Deeks, M.J., Rodrigues, C., Dimmock, S., Ketelaar, T., Maciver, S.K., Malhó, R., and Hussey, P.J. 2007. *Arabidopsis* CAP1 – a key regulator of actin organisation and development. *Journal of Cell Science* **120:** 2609-2618.

Delmas, F., Petit, J., Joubes, J., Seveno, M., Paccalet, T., Hernould, M., Lerouge, P., Mouras, A., and Chevalier, C. 2003. The gene expression and enzyme activity of plant 3-deoxy-D-manno-2-octulosonic acid-8-phosphate synthase are preferentially associated with cell division in a cell cycle-dependent manner. *Plant Physiology* **133:** 348-360.

Delmas, F., Séveno, M., Northey, J.G.B., Hernould, M., Lerouge, P., McCourt, P., and Chevalier, C. 2008. The synthesis of the rhamnogalacturonan II component 3-deoxy-D-manno-2-octulosonic acid (Kdo) is required for pollen tube growth and elongation. *Journal of Experimental Botany* **59:** 2639-2647.

Der Fen Suen, Wu, S.S.H., Chang, H.C., Dhugga, K.S., and Huang, A.H.C. 2003. Cell wall reactive proteins in the coat and wall of maize pollen - potential role in pollen tube growth on the stigma and through the style. *The Journal of Biological Chemistry* **278:** 43672–43681.

Dettmer, J., Schubert, D., Calvo-Weimar, O., Stierhof, Y.D., Schmidt, R., and Schumacher, K. 2005. Essential role of the V-ATPase in male gametophyte development. *The Plant Journal* **41:** 117-124.

Dietrich, C.R., Han, G., Chen, M., Berg, R.H., Dunn, T.M., and Cahoon, E.B. 2008. Loss-of-function mutations and inducible RNAi suppression of *Arabidopsis LCB2* genes reveal the critical role of sphingolipids in gametophytic and sporophytic cell viability. *The Plant Journal* **54:** 284-298.

Doelling, J.H., Phillips, A.R., Soyler-Ogretim, G., Wise, J., Chandler, J., Callis, J., Otegui, M.S., and Vierstra, R.D. 2007. The ubiquitin-specific protease subfamily UBP3/UBP4 is essential for pollen development and transmission in *Arabidopsis*. *Plant Physiology* **145:** 801-813.

Dong, X., Hong, Z., Sivaramakrishnan, M., Mahfouz, M., and Verma, D. 2005. Callose synthase (CalS5) is required for exine formation during microgametogenesis and for pollen viability in *Arabidopsis*. *The Plant Journal* **42:** 315–328.

Drakakaki, G., Zabotina, O., Delgado, I., Robert, S., Keegstra, K., and Raikhel, N. 2006. *Arabidopsis RGP1* and *RGP2* are essential for pollen development. *Plant Physiology* **142:** 1480-1492.

Elias, J.E. and Gygi, S.P. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4:** 207-214.

Enns, L.C., Kanaoka, M.M., Torii, K.U., Comai, L., Okada, K., and Cleland, R.E. 2005. Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility. *Plant Mol Biol* **58:** 333-349.

Fellenberg, C., Milkowski, C., Hause, B., Lange, P.-R., Böttcher, C., Schmidt, J., and Vogt, T. 2008. Tapetum-specific location of a cation-dependent O-methyltransferase in *Arabidopsis thaliana*. *The Plant Journal* **56:** 132-145.

Fiebig, A., Mayfield, J.A., Miley, N.L., Chau, S., Fischer, R.L., and Preuss, D. 2000. Alterations in *CER6*, a gene identical to *CUT1*, differentially affect long-chain lipid content on the surface of pollen and stems. *The Plant Cell* **12:** 2001-2008.

Footitt, S., Dietrich, D., Fait, A., Fernie, A.R., Holdsworth, M.J., Baker, A., and Theodoulou, F.L. 2007. The COMATOSE ATP-Binding cassette transporter is required for full fertility in *Arabidopsis*. *Plant Physiology* **144:** 1467-1480.

Francis, K.E., Lam, S.Y., and Copenhaver, G.P. 2006. Separation of *Arabidopsis* pollen tetrads is regulated by *QRT1*, a pectin methylesterase gene. *Plant Physiology* **142:** 1004-1013.

Frietsch, S., Wang, Y.F., Sladek, C., Poulsen, L.R., Romanowsky, S.M., Schroeder, J.I., and Harper, J.F. 2007. A cyclic nucleotide-gated channel is essential for polarized tip growth of pollen. *Proc Natl Acad Sci U S A* **104:** 14531-14536.

Giavalisco, P., Nordhoff, E., Kreitler, T., Klöppel, K.-D., Lehrach, H., Klose, J., and Gobom, J. 2005. Proteome analysis of *Arabidopsis thaliana* by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionisation-time of flight mass spectrometry. *Proteomics* **5:** 1902 - 1913.

Glover, J., Grelon, M., Craig, S., Chaudhury, A., and Dennis, E. 1998. Cloning and characterization of *MS5* from *Arabidopsis*: a gene critical in male meiosis. *The Plant Journal* **15:** 345-356.

Golovkin, M. and Reddy, A.S.N. 2003. A calmodulin-binding protein from *Arabidopsis* has an essential role in pollen germination. *PNAS* **100:** 10558-10563.

Gu, Y., Li, S., Lord, E.M., and Yang, Z. 2006. Members of a novel class of *Arabidopsis* Rho guanine nucleotide exchange factors control Rho GTPase-dependent polar growth. *The Plant Cell* **18:** 366-381.

Guan, Y.-F., Huang, X.-Y., Zhu, J., Gao, J.-F., Zhang, H.-X., and Yang, Z.-N. 2008. *RUPTURED POLLEN GRAIN1*, a member of the MtN3/saliva gene family, is crucial for exine pattern formation and cell integrity of microspores in *Arabidopsis*. *Plant Physiology* **147:** 852-863.

Guermonprez, H., Smertenko, A., Crosnier, M.-T., Durandet, M., Vrielynck, N., Guerche, P., Hussey, P.J., Satiat-Jeunemaitre, B., and Bonhomme, S. 2008. The POK/AtVPS52 protein localizes to several distinct post-Golgi compartments in sporophytic and gametophytic cells. *Journal of Experimental Botany* **59:** 3087-3098.

Gupta, R., Ting, J.T.L., Sokolov, L.N., Johnson, S.A., and Luan, S. 2002. A Tumor Suppressor Homolog, AtPTEN1, Is Essential for Pollen Development in *Arabidopsis*. *The Plant Cell* **14:** 2495-2507.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17:** 994-999.

Hakozaki, H., Park, J.I., Endo, M., Takada, Y., Kazama, T., Takeda, Y., Suzuki, G., Kawagishi-Kobayashi, M., and Watanabe, M. 2008. Expression and developmental function of the 3-ketoacyl-ACP synthase2 gene in *Arabidopsis thaliana*. *Genes Genet Syst* **83:** 143-152.

Hashida, S.-n., Takahashi, H., Kawai-Yamada, M., and Uchimiya, H. 2007. *Arabidopsis thaliana* nicotinate/nicotinamide mononucleotide adenyltransferase (AtNMNAT) is required for pollen tube growth. *The Plant Journal* **49:** 694-703.

Hicks, G.R., Rojo, E., Hong, S., Carter, D.G., and Raikhel, N.V. 2004. Geminating pollen has tubular vacuoles, displays highly dynamic vacuole biogenesis, and requires *VACUOLESS1* for proper function. *Plant Physiology* **134:** 1227-1239.

Honys, D. and Twell, D. 2004. Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biology* **5:** R85.

Huang, S., An, Y.-Q., McDowell, J.M., McKinney, E.C., and Meagher, R.B. 1996. The *Arabidopsis thaliana ACT4/ACT12* actin gene subclass is strongly expressed throughout pollen development. *The Plant Journal* **10:** 189-202.

Hulskamp, M., Kopczak, S.D., Horejsi, T.F., Kihl, B.K., and Pruitt, R.E. 1995. Identification of genes required for pollen-stigma recognition in *Arabidopsis thaliana*. *The Plant Journal* **8:** 703-715.

Hurtado, L., Farrona, S., and Reyes, J.C. 2006. The putative SWI/SNF complex subunit BRAHMA activates flower homeotic genes in *Arabidopsis thaliana*. *Plant Mol Biol* **62:** 291-304.

Ito, T., Nagata, N., Yoshiba, Y., Ohme-Takagi, M., Ma, H., and Shinozaki, K. 2007. *Arabidopsis MALE STERILITY1* encodes a PHD-type transcription factor and regulates pollen and tapetum development. *The Plant Cell* **19:** 3549-3562.

Iwakawa, H., Shinmyo, A., and Sekine, M. 2006. *Arabidopsis* CDKA;1, a cdc2 homologue, controls proliferation of generative cells in male gametogenesis. *The Plant Journal* **45:** 819-831.

Jakobsen, M.K., Poulsen, L.R., Schulz, A., Fleurat-Lessard, P., Annette Møller, Husted, S., Schiøtt, M., Amtmann, A., and Palmgren, M.G. 2005. Pollen development and fertilization in *Arabidopsis* is dependent on the *MALE GAMETOGENESIS IMPAIRED ANTHERS* gene encoding a Type V P-type ATPase. *Genes & Development* **19:** 2757-2769.

Jiang, L., Yang, S.L., Xie, L., Puah, C.S., Zhang, X.Q., Yang, W.C., Sundaresan, V., and Ye, D. 2005 *VANGUARD1* encodes a pectin methylesterase that enhances pollen tube growth in the *Arabidopsis* style and transmitting tract. *The Plant Cell* **17:** 584-596.

Johnson-Brousseau, S.A. and McCormick, S. 2004. A compendium of methods useful for characterizing *Arabidopsis* pollen mutants and gametophytically-expressed genes. *The Plant Journal* **39:** 761-775.

Kang, B.-H., Rancour, D.M., and Bednarek, S.Y. 2003. The dynamin-like protein ADL1C is essential for plasma membrane maintenance during pollen maturation. *The Plant Journal* **35:** 1-15.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74:** 5383-5392.

Lagoa, C., Clericia, E., Drenia, L., Horlowc, C., Caporalib, E., Colombob, L., and Kater, M.M. 2005. The *Arabidopsis* TFIID factor AtTAF6 controls pollen tube growth. *Developmental Biology* **285:** 91-100.

Lalanne, E. and Twell, D. 2002. Genetic control of male germ unit organization in *Arabidopsis*. *Plant Physiology* **129:** 865-875.

Lee, Y.R., Li, Y., and Liu, B. 2007. Two *Arabidopsis* phragmoplast-associated kinesins play a critical role in cytokinesis during male gametogenesis. *The Plant Cell* **19:** 2595-2605.

Lemaitre, T. and Hodges, M. 2006 Expression analysis of *Arabidopsis thaliana* NAD-dependent isocitrate dehydrogenase genes shows the presence of a functional subunit that is mainly expressed in the pollen and absent from vegetative organs. *Plant Cell Physiol.* **47:** 634-643.

León, G., Holuigue, L., and Jordana, X. 2007. Mitochondrial Complex II is essential for gametophyte development in *Arabidopsis thaliana*. *Plant Physiology* **143:** 1534-1546.

Leroch, M., Neuhaus, H.E., Kirchberger, S., Zimmermann, S., Melzer, M., Gerhold, J., and Tjadena, J. 2008. Identification of a novel adenine nucleotide transporter in the endoplasmic reticulum of *Arabidopsis*. *The Plant Cell* **20:** 438-451.

Levitin, B., Richter, D., Markovich, I., and Zik, M. 2008. Arabinogalactan proteins 6 and 11 are required for stamen and pollen function in *Arabidopsis*. *The Plant Journal* **56:** 351-363.

Liu, J., Zhang, Y., Qin, G., Tsuge, T., Sakaguchi, N., Luo, G., Sun, K., Shi, D., Aki, S., Zheng, N. et al. 2008. Targeted degradation of the cyclin-dependent kinase inhibitor ICK4/KRP6 by RING-type E3 ligases is essential for mitotic cell cycle progression during *Arabidopsis* gametogenesis. *The Plant Cell* **20:** 1538-1554.

Lobstein, E., Guyon, A., Ferault, M., Twell, D., Pelletier, G., and Bonhomme, S. 2004. The putative *Arabidopsis* homolog of yeast Vps52p is required for pollen tube elongation, localizes to Golgi, and might be involved in vesicle trafficking. *Plant Physiology* **135:** 1480-1490.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E.M. 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25:** 117-124.

Martinez-Garcia, J.F., Monte, E., and Quail, P.H. 1999. A simple, rapid and quantitative method for preparing *Arabidopsis* protein extracts for immunoblot analysis. *The Plant Journal* **20:** 251-257.

Matsuura, K., Miyagawa, I., Kobayashi, M., Ohta, D., and Matoh, T. 2003. *Arabidopsis* 3-deoxy-D-manno-oct-2-ulosonate-8-phosphate synthase: cDNA cloning and expression analyses. *J Exp Bot* **54:** 1785-1787.

Mayfield, J.A. and Preuss, D. 2000. Rapid initiation of *Arabidopsis* pollination requires the oleosin-domain protein GRP17 *Nat Cell Biol.* **2:** 128-130.

Millar, A.A., Clemens, S., Zachgo, S., Giblin, E.M., Taylor, D.C., and Kunst, L. 1999. *CUT1*, an *Arabidopsis* gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *The Plant Cell* **11:** 825-838.

Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa T. 2006. *GENERATIVE CELL SPECIFIC1* is essential for angiosperm fertilization. *Nat Cell Biol.* **8**: 64-71.

Mouline, K., Véry, A.-A., Gaymard, F., Boucherez, J., Pilot, G., Devic, M., Bouchez, D., Thibaud, J.-B., and Sentenac, H. 2002. Pollen tube development and competitive ability are impaired by disruption of a Shaker K+ channel in *Arabidopsis*. *Genes & Dev.* **16:** 339-350.

Niewiadomski, P., Knappe, S., Geimer, S., Fischer, K., Schulz, B., Unte, U.S., Rosso, M.G., Ache, P., Flugge, U.-I., and Schneidera, A. 2005. The *Arabidopsis* plastidic glucose 6-phosphate/phosphate translocator GPT1 is essential for pollen maturation and embryo sac development. *The Plant Cell* **17:** 760-775.

Nishikawa, S.-I., Zinkl, G.M., Swanson, R.J., Maruyama, D., and Preuss, D. 2005. Callose (beta-1,3 glucan) is essential for *Arabidopsis* pollen wall patterning, but not tube growth. *BMC Plant Biology* **5:** 22.

Okada, T., Endo, M., Mohan, B.S., and Bhalla, P.L. 2005. Analysis of the histone H3 gene family in *Arabidopsis* and identification of the male-gamete-specific variant AtMGH3. *The Plant Journal* **44:** 557-568.

Palanivelu, R., Brass, L., Edlund, A.F., and Preuss, D. 2003. Pollen tube growth and guidance is regulated by *POP2*, an *Arabidopsis* gene that controls GABA levels. *Cell* **114:** 47-59.

Park, S., Rancour, D.M., and Bednarek, S.Y. 2008. In planta analysis of the cell cycle-dependent localization of AtCDC48A and its critical roles in cell division, expansion, and differentiation. *Plant Physiology* **148:** 246-258.

Persson, S., Paredez, A., Carroll, A., Palsdottir, H., Doblin, M., Poindexter, P., Khitrov, N., Auer, M., and Somerville, C.R. 2007. Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. *Proc Natl Acad Sci U S A* **104:** 15566-15571.

Pina, C., Pinto, F., Feijo, J.A., and Becker, J.D. 2005. Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control and gene expression regulation. *Plant Physiology* **138:** 744–756.

Preston, J., Wheeler, J., Heazlewood, J., Li, S.F., and Parish, R.W. 2004. AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. *The Plant Journal* **40:** 979-995.

Preuss, D., Rhee, S.Y., and Davis, R.W. 1994. Tetrad analysis possible in *Arabidopsis* with mutation of the *QUARTET* (*QRT*) genes. *Science* **264:** 1458-1460.

Procissi, A., Guyon, A., Pierson, E.S., Giritch, A., Knuiman, B., Grandjean, O., Tonelli, C., Derksen, J., Pelletier, G., and Bonhomme, S. 2003. *KINKY POLLEN* encodes a SABRE-like protein required for tip growth in *Arabidopsis* and conserved among eukaryotes. *The Plant Journal* **36:** 894-904.

Qin, G., Ma, Z., Zhang, L., Xing, S., Hou, X., Deng, J., Liu, J., Chen, Z., Qu, L.-J., and Gu, H. 2007. *Arabidopsis AtBECLIN1/AtAtg6/AtVps30* is essential for pollen germination and plant development. *Cell Research* **17:** 249-263.

Reddy, T.V., Kaur, J., Agashe, B., Sundaresan, V., and Siddiqi, I. 2003. The *DUET* gene is necessary for chromosome organization and progression during male meiosis in *Arabidopsis* and encodes a PHD finger protein. *Development* **130:** 5975-5987.

Rhee, S.Y., Osborne, E., Poindexter, P.D., and Somerville, C.R. 2003. Microspore separation in the *quartet3* mutants of *Arabidopsis* is impaired by a defect in a developmentally regulated polygalacturonase required for pollen mother cell wall degradation. *Plant Physiology* **133:** 1170-1180.

Rhee, S.Y. and Somerville, C.R. 1998. Tetrad pollen formation in *quartet* mutants of *Arabidopsis thaliana* is associated with persistence of pectic polysaccharides of the pollen mother cell wall. *The Plant Journal* **15:** 79-88.

Robertson, W.R., Clark, K., Young, J.C., and Sussman, M.R. 2004. An *Arabidopsis thaliana* plasma membrane proton pump is essential for pollen development. *Genetics* **168:** 1677-1687.

Rosti, J., Barton, C.J., Albrecht, S., Dupree, P., Pauly, M., Findlay, K., Roberts, K., and Seifert, G.J. 2007. UDP-glucose 4-epimerase isoforms UGE2 and UGE4 cooperate in providing UDP-galactose for cell wall biosynthesis and growth of *Arabidopsis thaliana*. *The Plant Cell* **19:** 1565-1579.

Rotman, N., Durbarry, A., Wardle, A., Yang, W.C., Chaboud, A., Faure, J.-E., Berger, F., and Twell, D. 2005. A novel class of MYB factors controls sperm-cell formation in plants. *Curr Biol* **15:** 244-248.

Ruzicka, D.R., Kandasamy, M.K., McKinney, E.C., Burgos-Rivera, B., and Meagher, R.B. 2007. The ancient subclasses of *Arabidopsis* Actin Depolymerizing Factor genes exhibit novel and differential expression. *The Plant Journal* **52:** 460-472.

Sanderfoot, A.A., Pilgrim, M., Adam, L., and Raikhel, N.V. 2001. Disruption of individual members of *Arabidopsis* syntaxin gene families indicates each has essential functions. *The Plant Cell* **13:** 659–666.

Schiøtt, M., Romanowsky, S.M., Bækgaard, L., Jakobsen, M.K., Palmgren, M.G., and Harper, J.F. 2004. A plant plasma membrane Ca2+ pump is required for normal pollen tube growth and fertilization. *PNAS* **101:** 9502-9507.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* **37:** 501-506.

Schneidereit, A., Scholz-Starke, J., and Buttner, M. 2003. Functional characterization and expression analyses of the glucose-specific AtSTP9 monosaccharide transporter in pollen of *Arabidopsis*. *Plant Physiology* **133:** 182-190.

Schneidereit, A., Scholz-Starke, J., Sauer, N., and Büttner, M. 2005. AtSTP11, a pollen tube-specific monosaccharide transporter in *Arabidopsis*. *Planta* **221:** 48-55.

Schnurr, J.A., Storey, K.K., Jung, H.-J.G., Somers, D.A., and Gronwald, J.W. 2006. UDP-sugar pyrophosphorylase is essential for pollen development in *Arabidopsis*. *Planta* **224:** 520–532.

Scholz-Starke, J., Buttner, M., and Sauer, N. 2003. AtSTP6, a new pollen-specific H+-monosaccharide symporter from *Arabidopsis*. *Plant Physiology* **131:** 70-77.

Song, X.-F., Yang, C.-Y., Liu, J., and Yang, W.-C. 2006. RPA, a class II ARFGAP protein, activates ARF1 and U5 and plays a role in root hair development in *Arabidopsis*. *Plant Physiology* **141:** 966-976.

Spielman, M., Preuss, D., Li, F.L., Browne, W.E., Scott, R.J., and Dickinson, H.G. 1997. *TETRASPORE* is required for male meiotic cytokinesis in *Arabidopsis thaliana*. *Development* **124** 2645-2657.

Stadler, R., Truernit, E., Gahrtz, M., and Sauer, N. 1999. The AtSUC1 sucrose carrier may represent the osmotic driving force for anther dehiscence and pollen tube growth in *Arabidopsis*. *The Plant Journal* **19:** 269-278.

Strompen, G., Dettmer, J., Stierhof, Y., Schumacher, K., Jurgens, G., and Mayer, U. 2005. *Arabidopsis* vacuolar H-ATPase subunit E isoform 1 is required for Golgi organization and vacuole function in embryogenesis. *The Plant Journal* **41:** 125-132.

Taylor, P.E., Glover, J.A., Lavithis, M., Craig, S., Singh, M.B., Knox, R.B., Dennis, E.S., and Chaudhury, A.M. 1998. Genetic control of male fertility in *Arabidopsis thaliana*: structural analyses of postmeiotic developmental mutants. *Planta* **205:** 492-505.

Teng, C., Dong, H., Shi, L., Deng, Y., Mu, J., Zhang, J., Yang, X., and Zuo, J. 2008. Serine palmitoyltransferase, a key enzyme for de novo synthesis of sphingolipids, is essential for male gametophyte development in *Arabidopsis*. *Plant Physiology* **146:** 1322-1332.

Tian, G.-W., Chen, M.-H., Zaltsman, A., and Citovsky, V. 2006. Pollen-specific pectin methylesterase involved in pollen tube growth. *Developmental Biology* **294:** 83-91.

Toller, A., Brownfield, L., Neu, C., Twell, D., and Schulze-Lefert, P. 2008. Dual function of *Arabidopsis* glucan synthase-like genes *GSL8* and *GSL10* in male gametophyte development and plant growth. *The Plant Journal* **54:** 911-923.

Truernit, E., Schmid, J., Epple, P., Illig, J., and Sauerai, N. 1996. The sink-specific and stress-regulated *Arabidopsis STP4* gene: enhanced expression of a gene encoding a monosaccharide transporter by wounding, elicitors, and pathogen challenge. *The Plant Cell* **8:** 2169-2182.

Truernit, E., Stadler, R., Baier, K., and Sauer, N. 1999. A male gametophyte-specific monosaccharide transporter in *Arabidopsis*. *The Plant Journal* **17:** 191-201.

Twell, D., Yamaguchi, J., and McCormick, S. 1990. Pollen-specific gene expressionin transgenic plants: coordinate regulation of two different tomato gene promoters during microsporogenesis. *Development* **109:** 705-713.

Van Damme, D., Coutuer, S., Rycke, R.D., Bouget, F.-Y., Inzé, D., and Geelen, D. 2006. Somatic cytokinesis and pollen maturation in *Arabidopsis* depend on TPLATE, which has domains similar to coat proteins. *The Plant Cell* **18:** 3502-3518.

Verelst, W., Twell, D., de Folter, S., Immink, R., Saedler, H., and Munster, T. 2007. MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology* **8:** R249.

von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. 2006. *Arabidopsis HAP2* (*GCS1*) is a sperm-specific gene required for pollen tube guidance and fertilization. *Development* **133:** 4761-4769.

Wan, J., Patel, A., Mathieu, M., Kim, S.-Y., Xu, D., and Stacey, G. 2008. A lectin receptor-like kinase is required for pollen development in *Arabidopsis*. *Plant Mol Biol* **67:** 469-482.

Wang, Y. and Yang, M. 2006. The *ARABIDOPSIS* SKP1-LIKE1 (ASK1) protein acts predominately from leptotene to pachytene and represses homologous recombination in male meiosis. *Planta* **223:** 613-617.

Wilhelmi, L.K. and Preuss, D. 1999. Self-sterility in *Arabidopsis* due to defective pollen tube guidance. *Science* **274:** 1535-1537.

Wolf, C., Hennig, M., Romanovicz, D., and Steinebrunner, I. 2007. Developmental defects and seedling lethality in apyrase *AtAPY1* and *AtAPY2* double knockout mutants. *Plant Mol Biol* **64:** 657-672.

Wolf, S., Grsic-Rausch, S., Rausch, T., and Greiner, S. 2003. Identification of pollen-expressed pectin methylesterase inhibitors in *Arabidopsis*. *FEBS Letters* **555:** 551-555.

Xu, J., Brearley, C.A., Lin, W.-H., Wang, Y., Ye, R., Mueller-Roeber, B., Xu, Z.-H., and Xue, H.-W. 2005. A role of *Arabidopsis* inositol polyphosphate kinase, AtIPK2a, in pollen germination and root growth. *Plant Physiology* **137:** 94–103.

Yamaoka, S. and Leaver, C.J. 2008. EMB2473/MIRO1, an *Arabidopsis* Miro GTPase, is required for embryogenesis and influences mitochondrial morphology in pollen. *The Plant Cell* **20:** 589-601.

Yang, C.Y., Spielman, M., Coles, J.P., Li, Y., Ghelani, S., Bourdon, V., Brown, R.C., Lemmon, B.E., Scott, R.J., and Dickinson, H.G. 2003 *TETRASPORE* encodes a kinesin required for male meiotic cytokinesis in *Arabidopsis*. *The Plant Journal* **34** 229-240.

Zaegel, V., Guermann, B., Le Ret, M., Andres, C., Meyer, D., Erhardt, M., Canaday, J., Gualberto, J.M., and Imbaulta, P. 2006. The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *The Plant Cell* **18:** 3548-3563.

Zhang, C., Guinel, F.C., and Moffatt, B.A. 2002. A comparative ultrastructural study of pollen development in *Arabidopsis thaliana* ecotype Columbia and male-sterile mutant *apt1-3*. *Protoplasma* **219:** 59-71.

Zhang, Z.B., Zhu, J., Gao, J.F., Wang, C., Li, H., Li, H., Zhang, H.Q., Zhang, S., Wang, D.M., Wang, Q.X. et al. 2007. Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. *The Plant Journal* **52:** 528-538.

Zhu, J., Chen, H., Li, H., Gao, J.-F., Jiang, H., Wang, C., Guan, Y.-F., and Yang, Z.-N. 2008. *DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION1* is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. *The Plant Journal* **55:** 266-277.