Charles University in Prague
Faculty of Science

Department of Plant Experimental Biology

Identification and characterization of transcription factors affecting late stages of male gametophyte development in *Arabidopsis thaliana*

Nikoleta Dupľáková

Ph.D. thesis

Prague 2010

Univerzita Karlova v Praze
Přírodovědecká fakulta

Katedra experimentální biologie rostlin

Identifikácia a charakterizácia transkripčných faktorov účastniacich sa regulácie neskorých štádií vývinu samčieho gametofytu *Arabidopsis thaliana*

Nikoleta Dupľáková

Dizertačná práca

Praha 2010

This is to certify that this Ph.D. thesis is not subject of any other defending procedure. It contains set of original results that have been or are being published in the international scientific journals

Prague, 10. 7. 2010, Nikoleta Dupľáková

Supervisor

RNDr. David Honys, Ph.D.

Laboratory of Pollen Biology
Institute of Experimental Botany AS CR
Rozvojová 263, Prague 6, 165 02
Czech Republic

Department of Plant Experimental Biology
Faculty of Science
Charles University in Prague
Viničná 5, Prague 2, 128 44
Czech republic

On behalf of the co-authors of the papers published, we hereby confirm the agreement with inclusion of the papers below into the dissertation thesis of Nikoleta Dupľáková. The papers were produced as a team work and the particular contribution of Nikoleta Dupľáková is specified at the beginning of relevant parts of the thesis.

Prague, 10. 7. 2010, David Honys

Poďakovanie

Na úvod by som rada poďakovala predovšetkým vedúcemu mojej dizertačnej práce RNDr. Davidovi Honysovi, Ph.D. za kvalitné a odborné vedenie mojej práce. Ďalej členom Laboratória biologie pylu na Ústave experimentálnej botaniky, hlavne Dr. Věře Čapkovej za cennú podporu a neustálu ochotu pomáhať. Patrikovi Hovancovi zase za obrovskú pomoc pri vzniku Arabidopsis Gene Profiler Family databázy. Davidovi Reňákovi, Věře Čapkovej a Davidovi Honysovi za pomoc pri preklade a revízii anglického textu dizertačnej práce.

Ďalej by som chcela poďakovať svojim blízkym a priateľom (ďakujem mamka, ocko, Patrik, Vierka) za obrovskú podporu nielen v časoch dobrých ale hlavne v tých zlých. Ale aj množstvu ďalších známych či neznámych ľudí, stretnutie s ktorými mi vždy vylúdi úsmev na tvári.

Som veľmi vďačná, že mi bolo umožnené pracovať na tak zaujímavej a inšpiratívnej téme.

Ďakujem.

Contents

1. Introduction

1.1. Male gametophyte of flowering plants

Life cycle of flowering plants consists of dominant diploid sporophytic phase and markedly reduced but vitally important haploid phase, male and female gametophyte. Male gametophyte is represented by pollen grains that are consisted by 2 or 3 highly specialized cells forming special "cell(s) within a cell" structure (Fig. 1.1) (Honys et al. 2006b). Vegetative and generative/sperm cell(s) show distinctive functional specialization, structure and cell fate underlaid by differential gene expression programs.

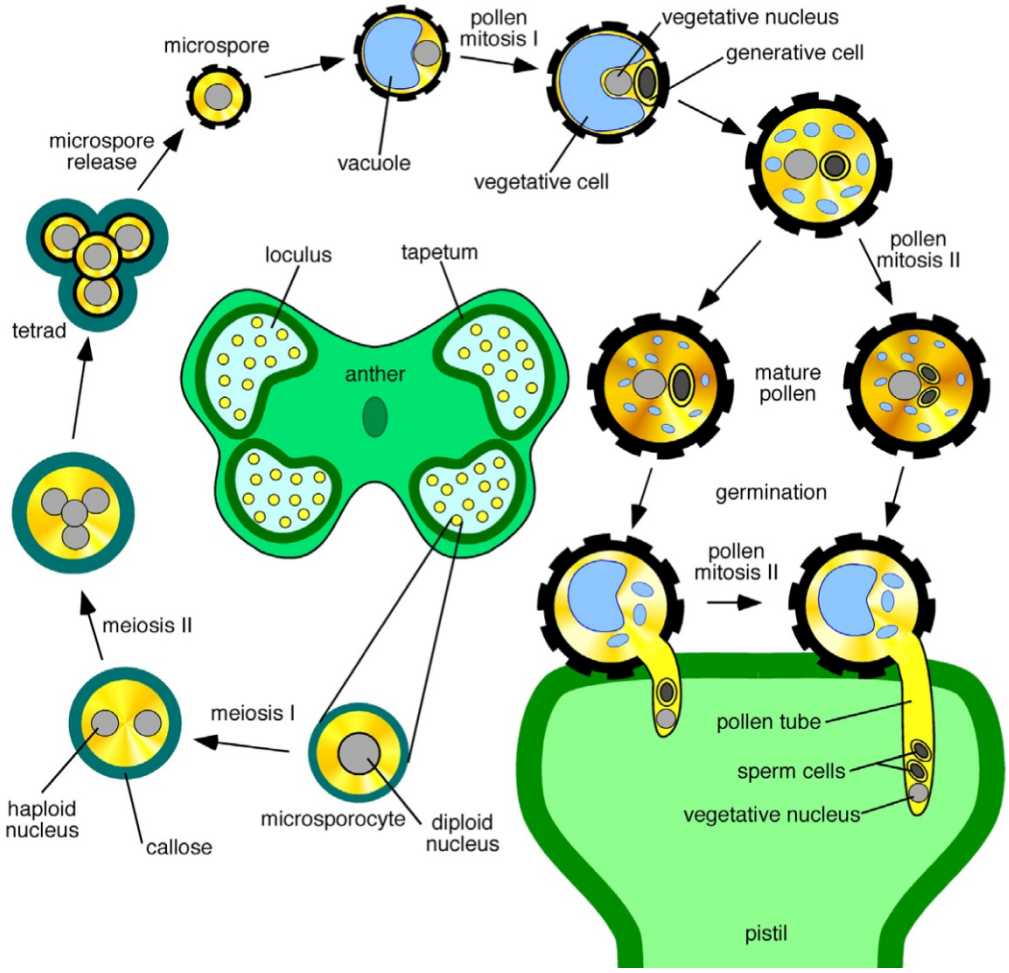


Fig. 1.1 Schematic diagram illustrating pollen development (Honys et al. 2006b)

1.2. *Pollen development*

The life of male gametophyte consists of two phases, concretely developmental and functional phase. First one is terminated by the formation of mature pollen shed from the anthers. The second, progamic or functional phase begins after pollination by pollen germination and subsequent pollen tube growth. It is ended by double fertilization. The whole process of male gametophyte development and function is precisely regulated by finely tuned gene expression and its dynamics in individual developmental stages (Borg et al. 2009).

There are two different pollen prototypes in relation to the timing of pollen mitosis II (PMII). In tricellular pollen this division occurs within the anther, whereas in bicellular pollen it occurs within the growing pollen tube. Although the majority of flowering plants (70% of species) including *Nicotiana tabacum* produce bicellular pollen, many important model and food crop plants such as *Arabidopsis thaliana, Oryza sativa, Triticum vulgare* and *Zea mays* produce advanced but often short-lived tricellular pollen grains (Twell et al. 2006; Singh and Bhalla 2007). Vegetative cell nurtures the developing gametes and, following successful pollination, forms to the pollen tube that carries the sperm cells to the female gametophyte. Sperm cells are delivered by growing pollen tube to the embryo sac where double fertilisation takes place (Borg et al. 2009).

Male gametophyte development takes place in specialized organs, stamens (Scott et al. 2004). Pollen grains are developed in anther loculi and are in close connection with metabolically active sporophytic cell layer, tapetum, that plays a critical role in many developmental processes such as the nutrition of microspores and the overall regulation of sporogenesis. The tapetum activity is necessary for microspore release by callase secretion causing the breakdown of callose wall surrounding the tetrads (Wilson and Yang 2004). Moreover, tapetum functions in pollen wall synthesis and patterning by secretion of pollen wall materials, primarily sporopollenin. Sporopollenin is composed of mixed polymers of fatty acid and phenylpropanoid derivatives and plays a important role in pollen dispersal and pollen–stigma interactions (Piffanelli et al. 1998; Zinkl et al. 1999).

1.3. High-throughput technologies for gene expression profiling

The current post-genomic era in biology can be characterized by enormous sequencing effort and incredible fast completion of genome sequences of many organisms. Up to date, 1945 organisms have been completely sequenced and still counting (http://www.genomesonline.org/, http://www.integratedgenomics.com/newgenomes.html). This effort as well as the general progress in molecular genetics and computer science stood in the beginning of the emerging high-throughput technological and biological approaches. Exploiting these approaches provided a significant opportunity to construct new genome-targeted questions concerning plant cellular functions. Accordingly, the picture that is being acquired about metabolic, developmental and physiological pathways becomes more and more complex.

Introduction of microarrays as an important high-throughput technology enabled global analyses of gene expression and as a consequence, made reverse genetics more applicable. The first reference of microarrays was as early as in 1994, when the first GeneChip® was developed by Stephen Fodor (McGall et al. 1997; Fodor et al. 1991; Pease et al. 1994). Since then, a qualitative development has increased the experimental accuracy and reliability and has brought wider coverage and many experimental setups. In broader sense, it is possible to characterize gene chips as ordered arrays of biomolecules on solid surface in two-dimensional pattern.

DNA arrays are currently the most commonly used and best-known microarray technology. DNA microarrays are composed of up to tens-of-thousands of DNA probes spotted or synthesized on a solid surface in an ordered high-density array of rows and columns. They permit the highly parallel measurement of labelled target nucleic acids within a complex mixture. DNA arrays work on the principle of the specific base-pairing of complementary nucleotides (A-T and C-G) and subsequent formation of probe-target interactions. Their main use lies in the detection rate of gene expression. However, number of novel applications is being introduced (Baginsky et al. 2010).

DNA microarrays have passed long way from “home-made” spotted cDNA arrays to commercially produced robust oligonucleotide gene chips. There have been several producers of DNA microarrays employing different probe-designing and array-printing technologies. Affymetrix, Agilent, Illumina, Roche Nimblegen, Applied Biosystems belong among the best

known. Due to historical reasons and despite of sometimes better performance of other systems, Affymetrix still holds the lead in number of published data. PubMed (http://www.ncbi.nlm.nih.gov/pubmed) search using individual microarray producers as keywords retrieved these numbers of citations: Affymetrix (2014), Agilent (196), Illumina (143), Applied Biosystems (52) and Nimblegen (23). That was one reason why we decided to use Affymetrix ATH1 Genome Arrays in our experiments.

Affymetrix ATH1 whole genome microarray was developed by Affymetrix in collaboration with Syngenta and became the first recognised "standard" in genome-wide transcriptomic studies in plants. According the last available annotation, over 22,500 probe sets harboured on the chip cover about 76% of the *Arabidopsis thaliana* genome [TAIR9, ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9_genome_release/].

The broader use of microarray technology raised the issue of data normalisation and comparability of different experiments with variable experimental set-ups. The introduction of the Minimum Information About Microarray experiments (MIAME) (Brazma 2009; Brazma et al. 2001) has increased the value and reproducibility of microarray experiments and has become a standard in documentation of array experiments and in the creation of databases of comparable transcriptomic experiments. Since the number of microarray experiments markedly increased in recent years, it is not surprising, that there are potential difficulties in the orientation in individual data sets. Microarray expression data are deposited on different servers, many of which are publicly accessible. Public plant microarray data are deposited in several databases including ArrayExpress (Brazma et al. 2003; Sarkans et al. 2005; Parkinson et al. 2009), GEO (Barrett et al. 2005), NASCarrays (Craigon et al. 2004) and the Stanford Microarray Database (Ball et al. 2005; Gollub et al. 2006; Gollub et al. 2003). These databases store thousands of individual datasets and some of them offer online tools for data normalization, filtering, statistical testing and pattern discovery. In parallel with the rapid accumulation of transcriptomic data, on-line analysis tools are being introduced to simplify their use. Global statistical data analysis methods contribute to the development of overall concepts about gene expression patterns and to query and compose working hypotheses. More recently, these applications are being supplemented with more specialized products offering visualization and specific data mining tools. Genevestigator (https://www.genevestigator.com/gv/index.jsp) (Zimmermann et al. 2004), Botany Array Resource (http://bar.utoronto.ca/) (Toufighi et al. 2005), Arabidopsis co-expression tool

(http://www.arabidopsis.leeds.ac.uk/act/) (Jen et al. 2006; Manfield et al. 2006), and ArrayExpress Expression Profiler (http://www.ebi.ac.uk/expressionprofiler/) (Kapushesky et al. 2004) represent some of the web-based tools used to analyse large microarray datasets.

1.4. Pollen transcriptomics

The whole process of the male gametophyte ontogenesis is finely tuned by modification of gene expression at all regulatory levels. This is accompanied by massive changes of gene expression profiles. A number of efficient and accurate gene expression analysis techniques have been widely exploited (Northern blot analysis, quantitative RT-PCR, cDNA differential screening) (Soares 1997; Hess et al. 1998). Most of these methods enabled the expression profiling at the level of single or relatively few selected genes. For the discovery of global functional or regulatory networks, the introduction of high-throughput technologies became essential. Several such methods have been developed including, cDNA fingerprinting (Money et al. 1996), serial analysis of gene expression – SAGE (Velculescu et al. 1995), massively parallel signature sequencing – MPSS (Brenner et al. 2000), high-density DNA oligonucleotide probe microarrays (Lockhart et al. 1996; Lipshutz et al. 1999) or cDNA arrays (Schena et al. 1995). Of these, particularly transcriptomic studies exploiting DNA microarrays provided a major input (Donson et al. 2002).

Before the dawning of the DNA microarray technology, gene-by-gene characterization led to the identification of approximately 150 pollen-expressed genes from different species, with strong evidence for pollen-specific expression in about 30 (reviewed in Twell 2002). Moreover, only 23 pollen-expressed genes were identified in Arabidopsis. The first study, exploiting 8K Affymetrix AG microarrays (Honys and Twell 2003), provided analysis for mature pollen based on approximately one-third of the Arabidopsis genome. This study led to the identification of 992 genes expressed in mature pollen of which 39% were considered pollen-specific. Accordingly, an observed significant 61% gametophytic/sporophytic overlap provided the ample potential to improve the fitness of the sporophytic generation through gametophytic competition and selection (Honys and Twell 2003). Based on such limited genome-wide studies, the estimation of pollen-expressed genes in *Arabidopsis* was between 3500 and 5500. The classification of pollen- expressed and pollen-specific genes into functional categories revealed several over-represented functional groups (cell wall, metabolism, cytoskeleton and signalling) among the pollen-specific genes. Moreover, pollen-specific genes were in general much more highly expressed than corresponding non-specific pollen-expressed genes (reviewed by (Twell et al. 2006).

Further elaboration was enabled by the availability of Affymetrix 23K Arabidopsis ATH1 arrays. Currently, there is a number of available datasets covering several stages of male gametophyte and functions. The most thoroughly studied stage is apparently mature pollen (6 experiments, 16 arrays (Borges et al. 2008; Honys and Twell 2004; Pina et al. 2005; Zimmermann et al. 2005; Qin et al. 2009; Wang et al. 2008)) serving as a baseline stage for most male gametophytic developmental studies. Three stages of immature pollen (uninucleate microspores (UNM), bicellular pollen (BCP), tricellular pollen (TCP), 1 experiment, 2 arrays per stage, (Honys and Twell 2004) completed the figure of pollen development. Recently, transcriptomic data covering the progamic phase have emerged – germinating pollen (2 experiments, 6 arrays, (Wang et al. 2008; Qin et al. 2009)) and pollen tubes (2 experiments, 9 arrays, (Qin et al. 2009; Wang et al. 2008)). Apart from revealing the number (Fig. 1.2) identity of the vast majority of male gametophyte-expressed genes, the major impact of these studies lies in the massively increased knowledge of the complexity and dynamics of haploid gene expression throughout single-cell development in plants (Honys and Twell 2004).

Although the availability of far more complete datasets shown that some statements published in the initial 8K study (Honys and Twell 2003) were unreasonably optimistic and required revision – namely the percentage of pollen-specific genes was reduced to cca 10% - most of the presented statements still remain valid. Affymetrix ATH1 studies confirmed the reduced complexity of male gametophytic transcriptome (reviewed by (Honys et al. 2006b; Twell et al. 2006)). Moreover, principal component analysis clearly demonstrated that the transcriptome of mature pollen differs markedly from all other tissues (Pina et al. 2005).

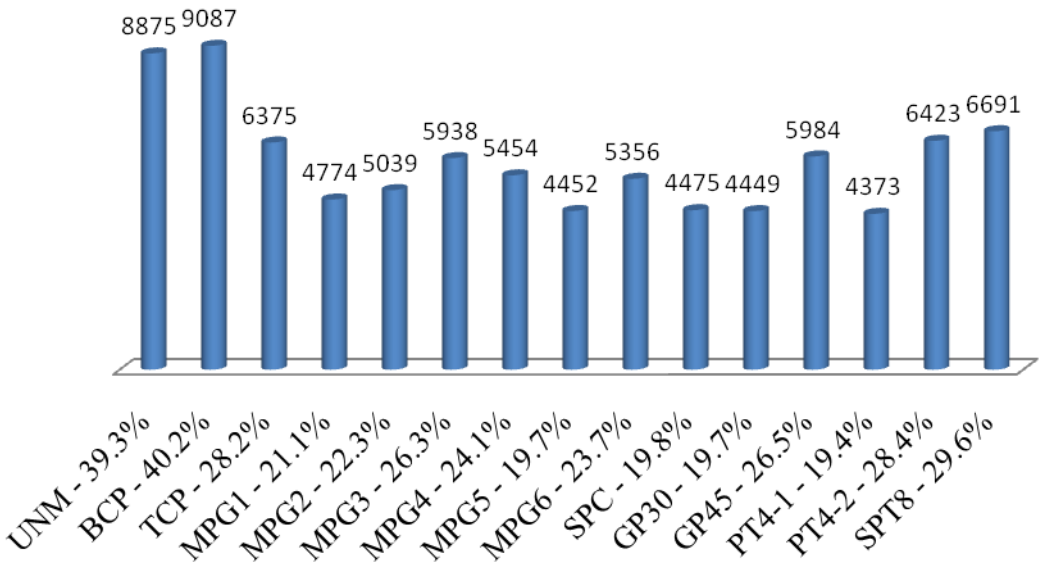


Fig. 1.2 The total number of expressed genes in individual developmental and progamic phase of pollen development based on Affymetrix ATH1 GeneChip 23K dataset; *UNM* uninucleate microspore, *BCP* bicellular pollen, *TCP* tricellular pollen, *MPG1* mature pollen grains (Honys and Twell 2004); *MPG2* mature pollen grains (Pina et al. 2005); *MPG3* mature pollen grains (Zimmermann et al. 2005); *MPG4* mature pollen grains (Borges et al. 2008); *MPG5* mature pollen grains (Qin et al. 2009); *MPG6* mature pollen grains (Wang et al. 2008); *SPC* sperm cells (Borges et al. 2008); *GP30* germinated pollen after 30 min. (Qin et al. 2009); *GP45* germinated pollen after 45 min. (Wang et al. 2008); *PT4-1* pollen tubes after 4 hours (Qin et al. 2009); *PT4-2* pollen tubes after 4 hours (Wang et al. 2008); *SPT8* pollen tubes semi in vitro, 3 h in poistil, 5 h in vitro (Qin et al. 2009)

1.5. Pollen transcription factors

Although the regulation of the development of the haploid plant generation is assured at all levels of gene expression, regulation of transcription still plays the key role. This is ensured by the activity of transcription factors functioning as transcription activators or repressors binding to cis-elements in promoters in the sequence-specific way and thus regulating the expression of downstream target genes. The plenitude of transcription factors, categorised into number of gene families, reflects the variability of target sequences as well as the variability of controlled processes. Many TF gene families are shared by plant, fungi and animal kingdoms, whereas others are specific just to one of them. Generally, the abundance and variability of transctiption factors is higher in plants than in the other kingdoms (Riechmann et al. 2000). In Arabidopsis genome, at least 1533 genes encoding putative transcription factors were identified (approximately 6% of the coding capacity of its genome) (Riechmann et al. 2000) with moderate variations published in other papers (Xiong et al. 2005; Riechmann 2002; Riechmann and Ratcliffe 2000; Gong et al. 2004; Guo et al. 2005). These differences are caused by inclusion of unclassified TFs and differences in bioinformatic search stringency (Qu and Zhu 2006). They were divided into more than 30 gene families (Riechmann et al. 2000; Riechmann 2002) Among the Arabidopsis TFs, about 45% are plant-specific whereas other TFs contain DNA-binding domains found also in other kingdoms (Riechmann et al. 2000; Riechmann 2002). . The size of individual TF gene families differs significantly – from single-copy or orphan genes (ie. *LEAFY* (*LFY*) or *SPOROCYTELESS* (*SPL*) (Riechmann 2006)) to gene families comprising over 100 members. Here, bHLH, AP2/ERF, MYB-(R1)R2R3 gene families are the largest in *Arabidopsis* genome and each of them contains approximately 6% of all transcription factors and/or 0.5% of total genes (Riechmann 2006).

The accessibility of high-throughput resources enabled the large-scale cloning of 1282 ORFs encoding *Arabidopsis* TFs (Gong et al. 2004) and the expression profiling of over 1400 TFs by real-time qRT-PCR (Czechowski et al. 2004). The introduction of microarrays enabled even more comprehensive expression studies of transcription factor genes. The Affymetrix ATH1 Genome Array harbours approximately 85% (1,350) of all transcription factors predicted in the Arabidopsis genome. Of them, 608 gave reliable expression signal in

the male gametophyte. That represented cca 45% of all transcription factors on the microarray. Moreover, as many as 54 TF genes (8.9% all male gametophytic TFs) were apparently male gametophyte-specific (Honys and Twell 2004). There was high variability in the representation of individual gene families in pollen. The most overexpressed TF families with more than 25 members expressed in the male gametophyte were CCAAT family, C2H2 zinc finger, WRKY, bZIP, TCP and GRAS gene families. On the contrary, AUX/IAA, HSF, bHLH, NAC, AP2-EREBP, HB, R2R3-MYB, MADS and C2C2 zinc finger gene families were all under-represented (Honys and Twell 2004). Expression profiles of male gametophytic TFs corresponded to general trends shown in total male gametophytic transcriptome. Besides a limited number of constitutively expressed genes, two major transcription factor genesets could be distinguished. One contains a major group of early-expressed genes and the second, a smaller group of genes, is more abundantly expressed later during pollen maturation. The reduction of expressed TFs after second pollen mitosis was in accord with the general tendency of reduction of mRNA complexity accompanied by increasing fraction of specific genes after PMII. Several gene families comprised predominantly early-expressed genes. These were the NAC, WRKY, TCP, ARF, Aux/IAA, HMG-box and Alfin-like transcription factors. On the other hand, C2H2 or TUBBY TFs showed predominately late expression profiles (Honys and Twell 2004). However, the biological relevance of these findings remains to be elucidated.

1.6. Transcription factors with demonstrated function in pollen development

Out of accord with high number of transcription factor genes expressed in male gametophyte, the specific function was assigned only to a limited number of them, predominantly by analyses of respective mutants. Of them, mutations affecting sporophyte development were much more common. Due to limited space, I mention only several key examples.

ABORTED MICROSPORES (AMS) is a member of MYC class of bHLH TFs. *ams* plants were completely devoid of pollen that was degenerated shortly after microspores release from the tetrad, prior to pollen mitosis I (Sorensen et al. 2003). *csa* is a male sterile mutant in rice (*Oryza sativa*) showing increased sugar contents in leaves and stems and reduced levels of sugars and starch in floral organs. *CSA* gene encodes a transcription factor belonging to R2R3 MYB transcription factor family that is expressed preferentially in the anther tapetal cells and in the sugar-transporting vascular tissues and probably regulates the sugar partitioning in male gametophyte development (Zhang et al. 2010). *MALE STERILITY (MS1)* encodes a nuclear protein with a PHD-finger domain accompanying RING finger and putative leucine zipper motifs. This factor has been proposed to regulate the expression of sporophytic genes necessary for pollen-exine formation since the mutant pollen showed aberrant exine deposition. Moreover, *MS1* was the first transcriptional regulator gene reported to affect the pollen development (Wilson et al. 2001; Ito et al. 2007; Ariizumi et al. 2005; Ito and Shinozaki 2002; Vizcay-Barrena and Wilson 2006; Yang et al. 2007). Gibberellins (GAs) play important roles in anther development. *GAMYB* is MYB transcription factor involved in GA-regulated gene expression in anthers (Kaneko et al. 2004; Aya et al. 2009; Tsuji et al. 2006). One of the mutants, *gamyb-4* shows abnormal enlarged tapetum and could not undergo normal meiosis and is involved in early anther development in rice (Liu et al. 2010). Jasmonate is another hormonal substance important for stamen development and pollen maturation. *MYB108*, *MYB21* and *MYB24* are transcription factors involved in the jasmonate response in *Arabidopsis* stamens and plants with insertion in these genes have reduced male fertility associated with delayed anther dehiscence, reduced pollen viability, and decreased fecundity (Mandaokar et al. 2003; Mandaokar and Browse 2009; Mandaokar et al. 2006).

MYB103 is regulated by *EMS1/EXS* and its downregulation results in premature tapetum degeneration and pollen abortion (Zhang et al. 2007). In addition to the above

regulatory network, mutation in *MYB32* tapetum-expressed gene led to the aberrant pollen development and partial male sterility (Preston et al. 2004). Similarly, mutation in the *MYB26* gene caused male sterility by defective cell wall fortification in endothecial cell layer in anthers (Steiner-Lange et al. 2003).

Among gametophytic mutations in TF genes, the list is significantly shorter as only five were studied in detail. Unusual R2R3 MYB gene, *DUO POLLEN1* (*DUO1*) was the first germ line-specific regulator to be identified. It is specifically expressed in the male germline and was shown to be a key regulator of germ cell division and sperm cell formation in *Arabidopsis* (Rotman et al. 2005; Brownfield et al. 2009a; Brownfield and Twell 2009; Durbarry et al. 2005). It is considered that DUO1 is specific regulator of genes required for G2-to-M phase transition in generative cells. *MALE MEIOCYTE DEATH1 (MMD1)* encoding PHD-finger protein is expressed later, preferentially during male meiosis. It participates in chromatin remodelling and/or transcriptional events for successful progression through meiosis as *mmd1* plants showed chromatin fragmentation followed by cell death before cytokinesis (Yang et al. 2003). TCP-family *TCP16* is expressed predominantly in developing microspores; *TCP16*–targeted RNAi caused rapid abortion of microspores and abnormal pollen structure (Takeda et al. 2006). *HAPLESS5* (*HAP5)* encodes AP2-domain protein and its mutation caused pollen morphological defects and abortion (Johnson et al. 2004). Finally, At*bZIP34* controls pollen wall patterning and affects several metabolic pathways (Gibalová et al. 2009).

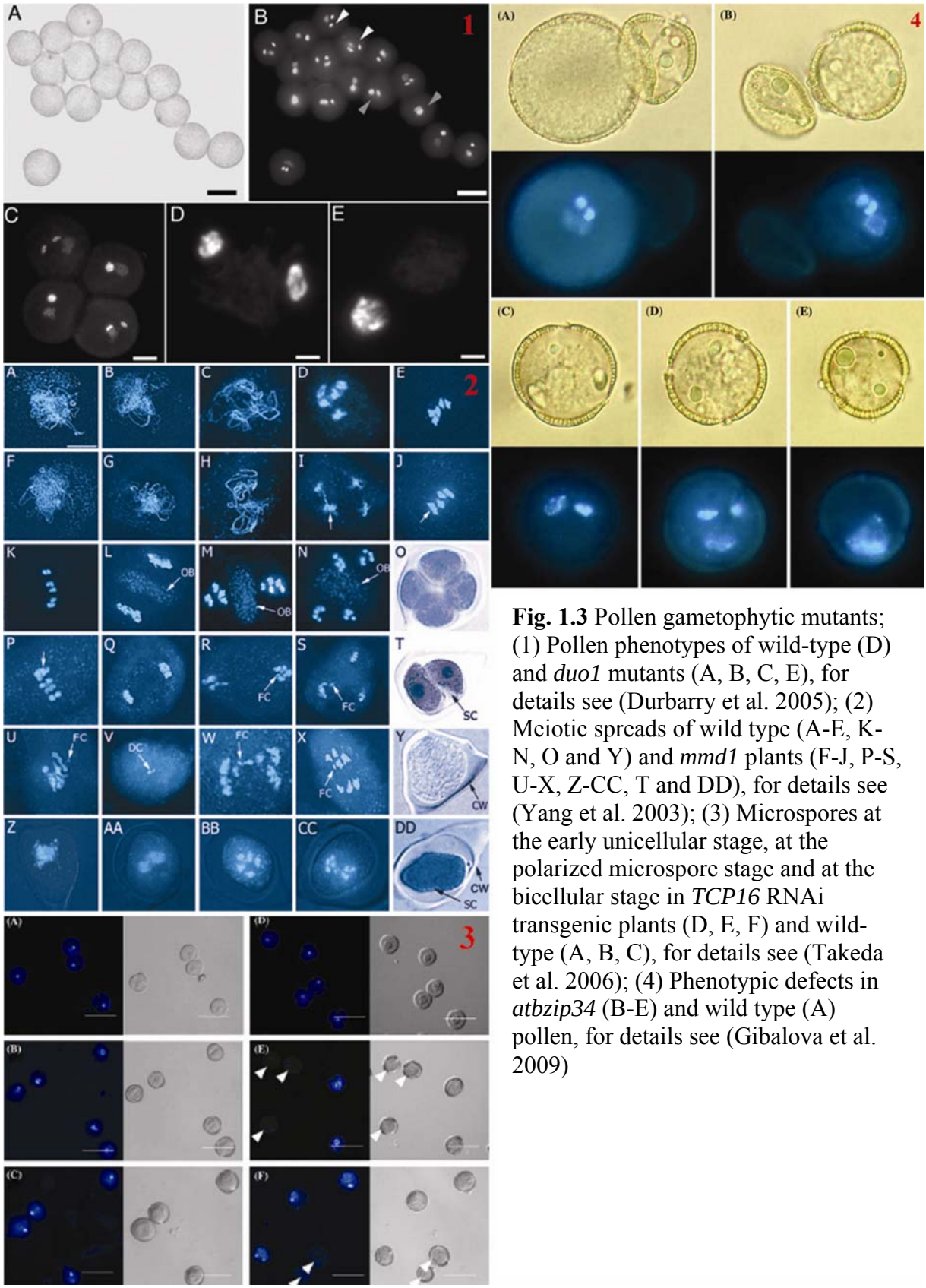


Fig. 1.3 Pollen gametophytic mutants; (1) Pollen phenotypes of wild-type (D) and *duo1* mutants (A, B, C, E), for details see (Durbarry et al. 2005); (2) Meiotic spreads of wild type (A-E, K-N, O and Y) and *mmd1* plants (F-J, P-S, U-X, Z-CC, T and DD), for details see (Yang et al. 2003); (3) Microspores at the early unicellular stage, at the polarized microspore stage and at the bicellular stage in *TCP16* RNAi transgenic plants (D, E, F) and wild-type (A, B, C), for details see (Takeda et al. 2006); (4) Phenotypic defects in *atbzip34* (B-E) and wild type (A) pollen, for details see (Gibalova et al. 2009)

2. Objectives

2.1. Objectives

The main objective of this thesis is the identification and characterisation of transcription factors involved in the regulation of late stages of *Arabidopsis* pollen development. To achieve this, available bioinformatics tools will be used in combination with the tools of reverse genetics.

The specific objectives of this work include:

1. Creation and maintenance of web-based database of *Arabidopsis thaliana* gene expression profiles
2. Selection of candidate genes encoding for transcription factors with the putative function in late stages of male gametophyte development
3. Demonstration of the regulatory function of selected candidate genes by phenotypic and genetic analyses of T-DNA insertion mutants in tested genes.

2.2. Strategies and work methods

The proposed project objectives will be met in four consecutive, however interdependent thematic units:

1. Development and maintenance of the gene expression database based on Affymetrix microarray data downloaded from various resources.
2. Identification of putative male gametophyte-expressed transcription factor genes.
3. Selection of candidate genes encoding for transcription factors with the likely function during late pollen development. Genes will be selected according to specific criteria including a) strict late expression profile, b) specific or at least significantly enriched (at least 50x) expression in the male gametophyte in comparison to sporophyte and c) the availability of appropriate T-DNA insertion lines.
4. Functional characterisation of selected genes demonstrating their importance for the flawless progress of late stages of pollen maturation.

3. Material and Methods

This chapter presents a simple survay of methods that I used for my Ph.D. thesis. Their detailed description and other information is provided in respective article.

3.1. Arabidopsis Gene Family Profiler (aGFP) – user-oriented transcriptomic database with easy-to-use graphic interface

- Data resources and definition of conditions for inclusion in Microarray experiments into Arabidopsis GFP database (Data resources)
- Creation of concept and Arabidopsis GFP database programming (Programming)
- The procedure for expression data normalization for Arabidopsis GFP database (Data normalization)
- The annotation of individual experiments and the loci involved in Arabidopsis GFP database (Annotation pages)
- The way how to define the gene families and superfamilies (Definition of gene families and superfamilies)

3.2. Wide-scale screening of T-DNA lines for transcription factor genes affecting male gametophyte development in Arabidopsis

- Processing of gene expression datasets coming from Affymetrix 23K ATH1 Arabidopsis GeneChips, including the use of publicly available databases of gene expression profiles to define a set of pollen specific or pollen significant genes (Microarray data)
- The use of publicly accessible databases to verify the expected pollen specific or pollen significant putative TF genes (Selection of candidate transcription factor genes)
- Using bioinformatics tools and mathematical techniques (eg. EPCLUST) for selection of specific pollen or pollen significant of putative transcription factors in the late stages of male gametophyte development (tricelullar pollen grain TCP, mature pollen grain MPG) (Selection of candidate transcription factor genes)

- Selection of appropriate and available T-DNA insertional lines (NASC SALK T-DNA lines, Syngenta - SAIL lines) to the disposal function of candidate genes and their subsequent genotypical verification (Selection of T-DNA insertion lines, Plant material and DNA extraction, Genotype analysis)
- Pollen grain phenotypic evaluation of T-DNA insertional mutant lines with knock down/out function of candidate genes (Microscopy analysis)

3.3. Functional characterization of phenotypically interesting mutants

3.3.1. AtbZIP34 is required for Arabidopsis pollen wall patterning and the control of several metabolic pathways in developing pollen

- Phenotypical characterization of *atbzip34* pollen *(*Microscopy in bright field and UV light)

3.3.2. C2-domain protein DEPOLL is required for intine development in Arabidopsis pollen

- Preparation of plant material (Plant material and growth conditions)
- Affymetrix ATH1 GeneChip 23K experiment (RNA extraction, probe preparation and DNA chip hybridization)
- Expression profile of *DEPOLL* gene (qRT-PCR, quantitative real-time PCR)
- Recovery of the normal phenotype of *depoll* pollen grains (Complementation analysis)
- Transmission of *depoll* mutant allele to the next plant generation (Transmission analysis)
- Phenotypical characterization of *depoll* pollen (Microscopy in bright field and UV light, Electron microscopy, Analysis of in vitro pollen tube growth)

4. Results

4.1. Creation and maintenance of web-based Arabidopsis thaliana gene expression database

We created an on-line Arabidopsis Gene Family Profiler database (arabidopsisGFP, aGFP, http://agfp.ueb.cas.cz/) that is composed as a relational MySQL database and Web server application programmed in PHP script language [PHP:Hypertext Preprocessor]. This database was developed to offer a new tool for the presentation of gene expression data, providing the user quick and intuitive way to display them using the innovative and unique concept of the virtual plant.

Arabidopsis GFP database was named after predefined gene families from which it consists. These gene families can be analysed immediately with several available data mining and visualisation tools. Next, we offered to the user a possibility of choice between normalization and gene detection algorithms. Last but not least the database employs an progressive idea "from simple to complex“ which is represented by the virtual plant. aGFP database provides the user also more traditional data presenting options (bar chart or tabulated display). All previously mentioned traits distinguish this database from others web-based tools for the work with microarray dataset.

Interactive virtual plant represents innovation and functional simplification of the preview of the transcription profile of single genes which comprises several growth stages defined according to Boyes et al. (Boyes et al. 2001). A white (low)-yellow-green (high) scale is used to depict the relative expression signals of individual genes, or gene families throughout the *Arabidopsis* life cycle. Mouse-over pointing to complex organs/tissues (i.e. flowers) causes opening of more detailed graphics showing individual organs (ie. sepals, petals, stamens, pistils and pollen). In the first approximation, an arithmetical mean expression signal from multiple experiments is displayed. In subsequent steps the user can choose to display expression data for individual plant organs or tissues at particular growth stages. This is accompanied by the option of progressive replacement of arithmetical means by individual expression values. So the user has the option to choose the different levels of visualization according his needs. This is in accordance with the adopted aGFP database concept "from simple to complex".

Finally, the user can switch from "virtual plant" visualization to a simple bar chart (standard or log-scaled) or tabulated display and can browse through individual experiments down to normalized or even raw data extracted from individual gene chips. Gene family data can also be visualized as a colorized spot chart. These attributes of aGFP contribute to a useful resource for the rapid bioinformatic analysis of *Arabidopsis* gene expression data through comparative expression profile analysis in a gene family based context.

The aGFP database can serve as a reference manual for the expression of individual genes and gene families in *Arabidopsis thaliana* under physiological conditions at different phases of the life cycle and in individual organs and tissues or cell types.

Survey of accessible transcriptomic data repositories and on-line data mining and visualisation tools to work are summarized in book chapter included in the PhD thesis. In this chapter we decided to compile a survey of microarray data normalization, filtering, statistical testing and pattern discovery servers with clear and brief description of their availabilities and differencies. For each database, basic information is available like URL, annotation of Arabidopsis genome, microarray data resource, used platforms, data normalisation, number of microarray datasets covered and access options. It is also distinguished between data repositories and/or data analysis tools.

We provide more detailed description of these portals - Arabidopsis Co-expression Tool (ACT), Arabidopsis Gene Family Profiler (aGFP), ArrayExpress, Bio-Array Resource (BAR), Gene Expression Omnibus (GEO), Genevestigator V3 (GV3), NASCArrays, The Stanford Microarray Databases (SMD). Microarray data analysis tools can be divided into two main categories. The first category comprises gene expression visualization tools, whereas tools forming the second category offer more thorough analyses of gene expression data, for example gene grouping and display of co-expressed genes. The most advanced software applications (Genevestigator V3 or BAR) offer both types of analyses. Other tools were designed specifically for intuitive visualization of gene expression at various morphological and developmental stages (aGFP).

The beginner in transcriptomic world will find in the chapter all necessary simplified direction how to be oriented in broad spectrum of expression data presented on-line and how to enter into this world with experimental data.

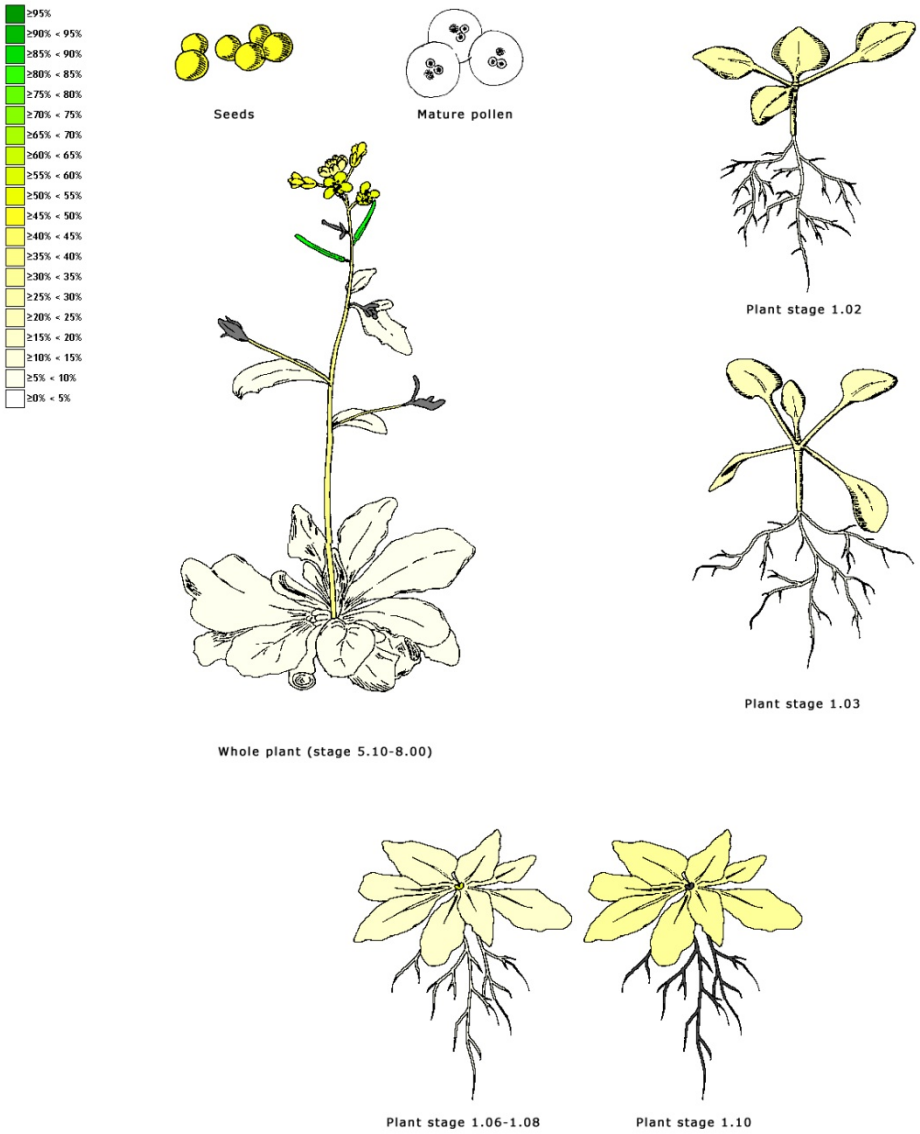


Fig. 4.1 Virtual plant with coloured expression scale for AtGenExpress experiments from Arabidopsis Gene Family profiler database (Dupľáková et al. 2007)

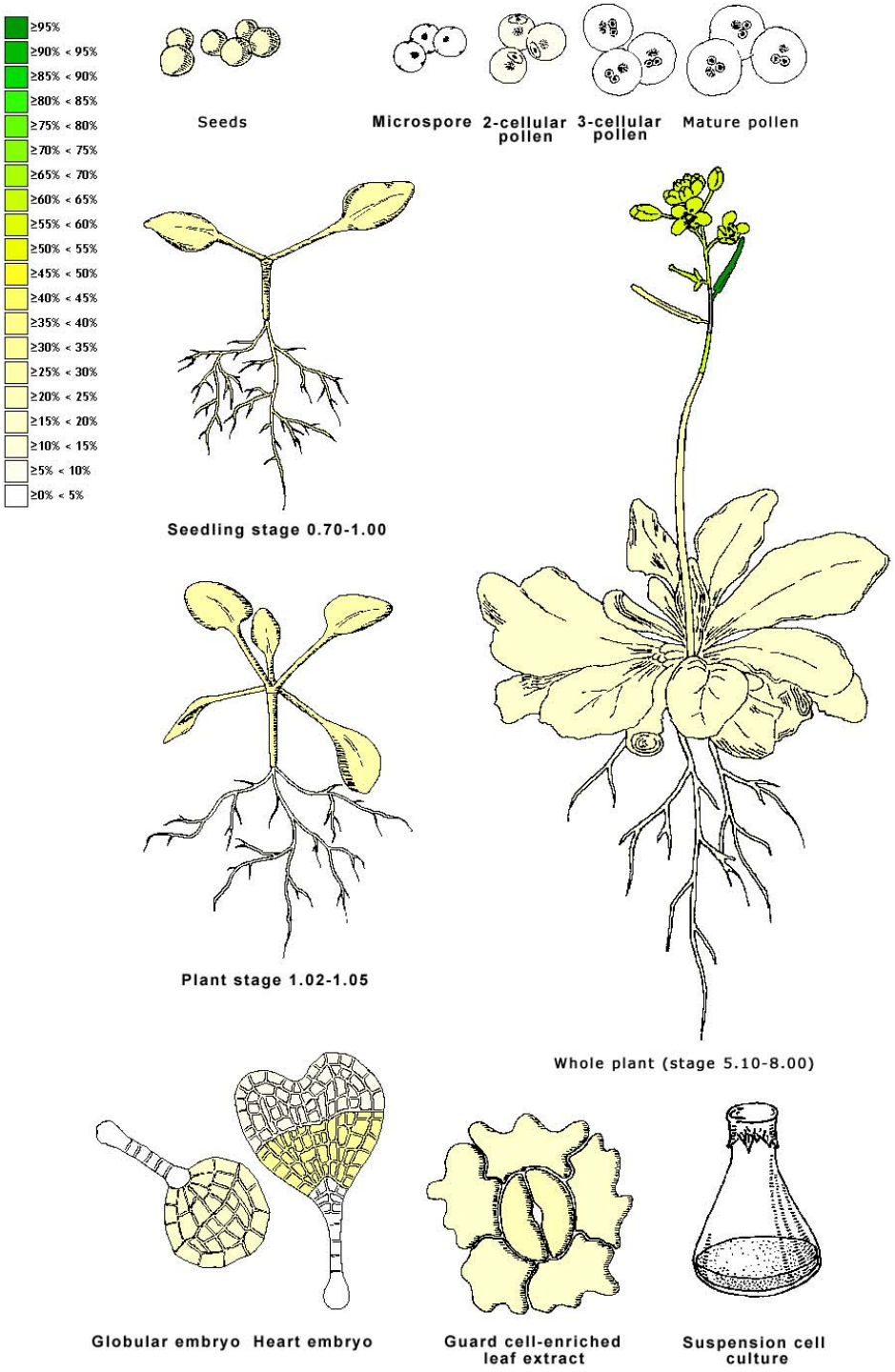


Fig. 4.2 Virtual plants with coloured expression scale for NASCArrays experiments from Arabidopsis Gene Family profiler database (Dupľáková et al. 2007)

Results summarised in this chapter were published in following papers:

1) **Dupl'áková N**, Reňák D, Hovanec P, Honysová B, Twell D, Honys D: **Arabidopsis Gene Family Profiler (aGFP) – user-oriented transcriptomic database with easy-to-use graphic interface.** *BMC Plant Biology* 2007, **7**:39.
   Impact factor 2009: 3.774

   My contribution: I participated in the concept definition of the arabidopsisGFP database and in the specification the data visualization concepts. I was fully responsible for the programming of all scripts and gene families and superfamilies definition. I was also involved in the manuscript preparation.

2) Honys D, **Dupľáková N**, Reňák D: **Online tools for presentation and analysis of plant microarray data.** In Moretti M.K., Rizzo L.J. (eds). Oligonucleotide Array Sequence Analysis, pp. 265-295, Nova Science Publishers, 2008.

   My contribution: I was involved in defining the overall concept of the paper and I was fully responsive for the detailed description of several databases and data repository servers.

4.2. Large-scare screening for transcription factor genes active during male gametophyte development

This chapter describes our pilot project for the discovery of the transcription factors regulatory network(s) important for male gametophyte development. Despite the long-term interest in understanding the metabolic and regulatory pathways in developing pollen, these processes are still far from being resolved. The actual work was divided into two separate sub-projects respecting the progression of pollen development. One part of the research was focused on early transcription factors responsible for the regulation of gene expression before the second pollen mitosis (PMII). The other part, which represented the subject of my study, was devoted to the late transcription factors. Late transcription factors are responsible for the control of the processes following PMII, including pollen germination or pollen tube growth.

The keynote of our study was that the knock-down of pollen-specific or significantly pollen-enriched transcripts is likely to cause the microscopically detectable disturbation of pollen phenotype. The analysis was preceded by the bioinformatic identification, selection and verification of candidate transcription factors. It was enabled by the availability of robust microarray data covering four stage of pollen development. These data were obtained within the scope of the long-term collaboration of our team with the laboratory of prof. Twell in the University of Leicester. In addition, we exploited the reverse genetic resources (AtGenExpress, NASCArrays). After the TF verification *in silico*, the appropriate T-DNA lines were ordered and phenotypically characterised using bright field and fluorescent microscopy.

In the screen, ten phenotype categories were distinguished affecting various structural or functional aspects including pollen abortion, presence of inclusions, variable pollen grain size, disrupted cell wall structure, cell cycle defects and male germ unit organization. Taken together, there were analysed 73 T-DNA insertion lines, representing 48 TF genes from 21 gene families expressed in either early or late pollen development. Thirteen lines were not confirmed to contain the T-DNA insertion. Among 60 confirmed lines, about a half (28 lines) showed strong phenotypic changes (i.e. ≥25% aberrant pollen) including four lines producing remarkable high proportion (70-100%) of disturbed pollen. However, remaining 32 lines exhibited mild defects or resembled wild-type appearance. There was no significant bias

towards any phenotype category amongst early and late TF genes and interestingly within individual TF families. The T-DNA line(s) exhibiting the most severe mutant phenotype disturbation(s) were selected for further functional characterization. In my case, the SAIL_1168_C11 line with the insertion in At1g70790 gene was selected for detailed research.

Results summarised in this chapter are being published in the following paper:

1) Reňák D*, **Dupľáková** N*, Honys D: **Wide-scale screening of T-DNA lines for transcription factor genes affecting male gametophyte development in Arabidopsis.** *Sexual Plant Reproduction* 2010. (in review)
   *should be regarded as equal contributors

   My contribution: As a joint first author I was responsive for the complete analysis of all late pollen transcription factors. It represented 37 lines. I was also involved in the manuscript preparation.

4.3. Functional characterisation of selected transcription factors

The first transcription factor identified in our laboratory was a factor involved in the early phases of male gametophyte development before the second pollen mitosis (PMII). This *AtbZIP34* transcription factor encoded by At2g42380 gene belongs to the ZIPs family. Both gametophytic and sporophytic role of *AtbZIP34* gene in proper formation and functional pollen grain had been verified. *AtbZIP34* was the first bZIP factor involved in male gametophyte development. The detailed phenotypic analyses of *atbzip34* mature pollen grains under the light and UV microscopy was my part in this publication.

atbzip34 pollen showed characteristic phenotypic defects affecting cell wall as well as pollen ultrastructural organization. Mutant pollen is distinguished by the characteristic wrinkled nexine and rare and deformed baculae and tecta, but does not lead to high levels of pollen abortion or male sterility. The exine patterning is affected but not the only phenotypic defects observed in *atbzip34* pollen. Characteristic inclusions observed in the cytoplasm of *atbZIP34* pollen suggested disturbance of metabolic pathways, possibly related to cellular transport and/or lipid metabolism. This assumption was supported by transmission electron microscopy showing differences in structure of endomembrane systems and lipid bodies. *atbZIP34* mutant pollen grains show morphological defects that result in reduced pollen germination efficiency and slower pollen tube growth both *in vitro* and *in vivo*. The deccessive role of *AtbZIP34* transcrition factor in formation of exine and in regulation of lipide metabolism or cell transport was suggested.

Our effort to select transcription factors involved in maturation of pollen grains (see paragraph 2) continues by the characterization of T-DNA line SAIL_1168_C11 with insertion in At1g70790 gene. The high proportion of dead pollen grains and many other phenotypic deviations are reflected in the name of gene *DEPOLL* (DEath POLLen grains), however, no transcription factor C2H2 domain but only C2 one was identified by means of *in-silico* analyses. So the function of this gene can not belong to a transcription regulation.

The last part of the Ph.D. thesis was devoted to a functional analyses of *depoll* mutant describing the key role of this gene in proper cell wall formation and function of intine and intine-plasmatic membrane connection. General ultrastructure of *depoll* pollen was similar to that of wild type. However, there were significant changes observed in the intine ultrastructure and organisation. In *depoll* pollen, the intine was greatly reduced and often

nearly missing, passing to granular structure. In pollen grains with reduced or missing intine, plasma membrane was detached from exine layer forming so called "ghost space". Interestingly, there was no plasma membrane detachment observed in aperture regions. This can provide an explaination why *depoll* pollen showed no germination defects. Germination of the pollen grains, formation and even the growth of the pollen tubes in *depoll* mutant are not affected.

Analyses of Affymetrix ATH1 gene expression data of *depoll* mutant identified several groups of genes with suggested or verified correlation toward metabolism of pollen grain cell wall (lipid transfer protein, polygalacturonase, invertase/pectin methylesterase inhibitor proteins, xyloglucan endotransglucosylase/hydrolase gene family).

Our task will be the determination of the involvement and function of DEPOLL proteins with C2 domain in particular metabolic and/or signal pathways. Future research on *DEPOLL* gene and its protein product represents a promising chance to come to understanding of so important phase of male gametophyte development.

Results summarised in this chapter were published in following papers:

1) Gibalová A, Reňák D, Matczuk K, **Dupľáková N**, Cháb D, Twell D, Honys D: **AtbZIP34 is required for *Arabidopsis* pollen wall patterning and the control of several metabolic pathways in developing pollen.** *Plant Molecular Biology* 2009, **70**:5.
   Impact factor 2009: 3.978

   My contribution: I was involbed in the phenotype evaluation of *atbzip34* pollen by bright field and fluorescent microscopy.

2) **Dupľáková N**, Novotný M, Fíla J, Levkaničová Z, Honys D: **C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen.** *Planta* 2010. (in preparation)

   My contribution: As a first author I was responsive for most of the work in this paper. I am being involved in the manuscript preparation.

5. Discussion

Knowledge of regulatory mechanisms employed on cellular level (Saez-Vasquez and Gadal 2010; Charon et al. 2010) allows us better understanding of processes controlling cell responses to internal as well as external stimuli. Male gametophyte is one of the models used in studies of regulatory and metabolic processes in contemporary biological sciences (Russell et al. 2010; Suzuki 2009; McCormick 2004). The reason why it is so extensively used as a model object is the haploid genome that provides option of immediate manifestation of recessive mutations, high grade of structural reduction and sophisticated method of genetic manipulations (Honys et al. 2006a; Mitsuda et al. 2006). Besides using a pollen grain as a model organism in order to understand heterogeneous biological mechanisms and general rules applied in plant kingdom (Dickinson and Grant-Downton 2009; Grant-Downton et al. 2009; Geitmann 2010; Rong et al. 2010; Kato et al. 2010), knowledge itself of the female gametophyte inevitably expands our basic knowledge and expertise (Geitmann 2010; Borg and Twell 2010; Stewman et al. 2010).

Individual phases of female gamethophyte development differ on the basis of significant morphological and physiological changes, taking place in them (Twell et al. 2006; McCormick 1993; Twell 2002) and are accompanied by strong changes at the transcriptome level. Development of female gametophytes may with respect to their nature be divided into two elementary phases – late and early (Honys and Twell 2004). Late phase of male gametophyte development is characterized by reduction of mRNA diversity however particularity of transcribed genes against early stages increases in contrast to that (uninucleate microspore UNM, bicellular pollen BCP) (Honys and Twell 2004). Tricellular pollen (TCP) and mature pollen grains (MPG) range among late phases of pollen development.

Above given fact is also accepted by our analysis of regulatory network of transcription factors within male gametophyte *Arabidopsis thaliana*, which has been divided into two separate but parallel running researches of early and late transcription factors. The initial reflective concept relating to the method of identification, selection and verification of gametophyte-specific or significant factors has in both cases been identical.

The principal subject of my Ph.D. thesis was the research of late transcription factors regulating pollen development in *Arabidopsis thaliana*. Functional, genetic and comparative trancriptomic studies indicate numerous representation of genes expressed in late stages of

pollen development within cell wall metabolism, signal transduction, pollen germination and pollen tube growth (Honys and Twell 2004; Pina et al. 2005; Becker et al. 2003; Honys and Twell 2003).

Application of modern technologies for gene expression analysis (Rensink and Buell 2005; Meyers et al. 2004a; Meyers et al. 2004b) including development of data mining tools represented enormous contribution for the plant biology (Stolc et al. 2005). Application of such approaches allows fast generating and verification of working hypotheses (Brady and Provart 2009; Aoki et al. 2007) and it has contributed e.g. to identification of 75 Mitogen-Activated Protein Kinase Kinase Kinases (MAPKKKs) in rice (Rao et al. 2010). Availability of transcription factor expression profiles from Affymetrix ATH1 Arabidopsis GeneChip 23K in individual stages of pollen development (UNM, BCP, TCP, MPG) (Honys and Twell 2004) allowed formulation of the basic method to identify pollen-specific or pollen-significant transcription factors. Data on sporophytic part of *Arabidopsis thaliana* were obtained from public available sources of NASCArrays (Craigon et al. 2004) and AtGenExpress (Schmid et al. 2005). Proper processing of extensive available transcriptomic data resulted in creation of Arabidopsis Gene Family Profiler (Arabidpsis GFP, aGFP, http://agfp.ueb.cas.cz/) database and a tool for their presentation and visualization. The effort has thus accomplished several tasks mutually interconnected in spirit and content.

Currently there are several databases and web-based tools for storage, imaging and analysis of "microarray expression data" from the plants (Arabidopsis Co-expression Tool (ACT) (Jen et al. 2006; Manfield et al. 2006), Arabidopsis Gene Family Profiler (aGFP) (Dupľáková et al. 2007), ArrayExpress (Brazma et al. 2003; Parkinson et al. 2009; Parkinson et al. 2005; Sarkans et al. 2005), Bio-Array Resource (BAR) , Gene Expression Omnibus (GEO) (Barrett et al. 2005; Edgar et al. 2002), Genevestigator V3 (GV3) (Zimmermann et al. 2005; Zimmermann et al. 2004), NASCArrays (Craigon et al. 2004), The Stanford Microarray Databases (SMD) (Gollub et al. 2006; Gollub et al. 2003; Ball et al. 2005)). Clear and brief description of their options and differences between them were drawn up in the publication on On-line tools for presentation and analysis of plant microarray data (Honys et al. 2008). Separate portals differ in numerous parameters. Part of them serves as data repositories (GEO), most of it however functions as a combination of data repository and analysis tool (GV3, BAR). Another differentiation parameter is their approach to imaging of expression data. We may encounter gene-centric and/or genome-centric approach. The first of the

approaches is able to select and image expression profile for genes meeting selected criteria (ACT), while most of the portals provide both options (GV3, BAR). This book chapter in addition provides basic information like URL, annotation of Arabidopsis genome, microarray data resource, used platforms, data normalisation, number of microarray datasets covered and access options.

We proposed the basic concept of our Arabidopsis GFP database upon options review, provided by individual gene expression data visualization and mining tools. None of the web-based tools has until then provided an option to visualize expression genes of the pre-defined gene families. Then we decided to provide users with the option to choose between two normalization and gene detection algorithms. Not the least we introduced a "virtual plant" that enables users to obtain a global expression profile of user specified or pre-defined gene families. The interactive "virtual plant" represents innovation and functional simplification of the insight into transcription profile of individual genes or gene families. This corresponds with the adopted aGFP database concept "from simple to complex". Finally, the user can switch from "virtual plant" visualization to a simple bar chart (standard or log-scaled) or tabulated display and can browse through individual experiments down to normalized or even raw data extracted from individual gene chips. Gene family data can also be visualized as a colorized spot chart (heat map). These attributes of aGFP contribute a useful resource for the rapid bioinformatic analysis of *Arabidopsis* gene expression data through comparative expression profile analysis in a gene familybased context.

Certain limitations of aGFP database result from the fact, that it applies resource data originated just from experiments with wild-type *Arabidopsis thaliana* grown under physiological conditions. The limitation may on the other hand be considered as one of its advantage. aGFP database may thus serve as a reference manual of expression genes under physiological conditions and provides user with access to vast collection of normalized Affymetrix ATH1 microarray datasets (120 experiments, 350 arrays) in various tissues in various stages of wild-type plant development. It thus provides a functional, simple and remarkable tool to create and verify overall concepts about gene expression patterns and to query and compose working hypotheses.

The concept of our search for specific and significant transcription factors alone combined the *in-silico* analysis with its own experimental efforts. At the time the Ph.D. thesis had been assigned as well as during the subsequent identification and selection of individual

factors only one microarray expression data for uninucleate microspore, bicellular pollen, tricellular pollen (Honys and Twell 2004) and two data resources for mature pollens were available (Honys and Twell 2004; Zimmermann et al. 2004). Though, in the course of time, other microarray datasets characterizing transcriptome of mature pollen grain became gradually available (Qin et al. 2009; Pina et al. 2005; Wang et al. 2008; Borges et al. 2008).

Besides bioinformatic tool EPCLUST (Kapushesky et al. 2004), databases of transcription factors within *Arabidopsis thaliana*, namely AGRIS and DATF (Davuluri et al. 2003b; Guo et al. 2005) our thesis employed reverse genetic approach. Reverse genetics is characterised by searching of changed phenotype manifestations in connection with particularly impaired metabolic or regulatory pathway (O'Malley and Ecker 2010). Implementation of mutants resulted e.g. in identification of plasma membrane zinc transporter in rice (Lee et al. 2010) or protein that regulates cellularization during endosperm development in *Arabidopsis* (Kang et al. 2008). We have applied large collections of T-DNA insertion mutants in our study (Krysan et al. 1999; Parinov and Sundaresan 2000; Krysan et al. 2002).

So far several tens of gametophytic and sporophytic mutations effecting shaping and functionality of male gametophyte have disposed functional characterisation (Kim et al. 2010; Chen et al. 2008; Iwakawa et al. 2006; Kim et al. 2008; Brownfield et al. 2009b). However only a smaller part represents mutations with proven relation to impaired functionality of transcription factors (Takeda et al. 2006; Brownfield et al. 2009a; Brownfield and Twell 2009; Durbarry et al. 2005; Rotman et al. 2005; Yang et al. 2003; Johnson et al. 2004; Gibalova et al. 2009; Mandaokar and Browse 2009; Sorensen et al. 2003; Liu et al. 2010; Zhang et al. 2010), though preliminary estimates based on chip analyses predict several hundreds of transcription factors expressed in pollen (Honys and Twell 2004).

The oldest identified genes of this mutation group include *MMD1* gene (At1g66170), which encodes plant homeo domain (PHD), finger protein MMD1 gene regulates chromatin-mediated gene expression during meiosis of *male meiocyte death1* (*mmd1*), gametophytic mutant shows numerous anomalies including defects in chromosome behavior, cytoplasmic shrinkage, and chromatin fragmentation, followed by cell death before cytokinesis (Yang et al. 2003).

Recent functionally characterized pollen mutants with impaired function of transcription factors include e.g. *csa* mutant, *gamyb-4* mutant and *myb108* mutant (Zhang et

al. 2010; Mandaokar and Browse 2009; Liu et al. 2010). *Carbon starved anther* (*csa*) sterile male mutant in rice shows reduced expression of monosaccharide transporter MST8 in anthers, which results in reduced levels of carbohydrates in late anthers. *CSA* gene encode an R2R3 MYB transcription factor, regarded for the key transcriptional regulator for sugar partitioning in rice during male reproductive development (Zhang et al. 2010).

The onset of our study of gametophyte-specific or significant late transcription factors was based on more than 1 500 putative transcription factors (Riechmann et al. 2000), which are expressed in gametophytic and/or various sporophytic tissues of *Arabidopsis thaliana*. Bio-information analysis ended up with 73 T-DNA insertion lines representing 48 genes of 21 TF families active in either early or late pollen development.

Two most numerous families of transcription factors in our study included C2H2 (8 members) and AP2/EREBP (5 members) and gene family. C2H2 zinc finger TF family belong to the major families in *Arabidopsis thaliana* and its number of genes is estimated from 66 to 211 according to different sources (e.g. RARTF (Iida et al. 2005), AGRIS (Davuluri et al. 2003b)). This type classical zinc finger contain protein domain with *c*onsensus pattern C - x(2,4) - C - x(3) - [LIVMFYWC] - x(8) - H - x(3,5) - H that contains two cysteines and two histidines involved in the tetrahedral coordination of a zinc atom (Takatsuji 1998; Englbrecht et al. 2004). Originally have been described in the *Xenopus* transcription factor TFIIIA (Ginsberg et al. 1984). The members of this TF family are utilized in flower, embryo, shoot and seed development, flowering time, root nodule development, abiotic stress response and light acclimatization response (FIS2 (Luo et al. 1999), SUP (Sakai et al. 1995)). So far several C2H2 zinc finger transcription factors expressed in anthers have been characterized (Takatsuji 1999; Kobayashi et al. 1998), while e.g. *TAZ1* gene (Kapoor et al. 2002) and *MEZ1* gene (Kapoor and Takatsuji 2006) cause pollen abortion in petunia. *taz1* as well as *mez1* mutations have sporophytic basis.

Another heavily studied family of transcription factors includes AP2/EREBP gene family, which is exclusively distributed in plants and named after its two founding members AP2 (APETALA2) (Okamuro et al. 1997; Gil-Humanes et al. 2009) and EREBPs (ethylene-responsive element binding proteins) (Leubner-Metzger et al. 1998). So far the function of this family members has been demonstrated in many aspects of the plant development with respect to stimulation through external and internal stimuli. These have been employed e.g. in floral organ identity determination, control of leaf epidermal cell identity, to forming part of

the mechanisms used by plants to respond to various types of biotic and environmental stress (Riechmann and Meyerowitz 1998; Dietz et al. 2010). The last functionally characterized representatives of the family of transcription factors include *WRI1* gene as a critical regulator of oil accumulation in maturing *Arabidopsis* seeds (Baud et al. 2009) or *EgAP2-1* gene expressed in meristematic and proliferating tissues of embryos in oil palm (Morcillo et al. 2007).

Our primary assumption for detection of gametophyte significant or specific transcription factors was, that insertion in respective transcription factor triggers strong phenotype and functional changes in the image of pollen grain in comparison with wild-type *Arabidopsis thaliana* with respect to their link to various metabolic, regulatory and/or signalling pathways in a cell (Twell et al. 2006; Honys et al. 2006b; Borg et al. 2009). This was the reason why we observed complete range of phenotype signs of mature pollen grains by means of the light and UV microscopy. The following signs such as: size, number of pollen grains, presence of pollen grains in formations (tetrads, triads, twins, single pollen grains), number of aborted pollen grains, nuclei position in male germ unit (MGU) and number of nuclei in MGU were evaluated. All typical aberrations in appearance of pollen grain, e.g. lipid inclusions were searched at the same time.

Surprising was the discovery of our large-scale study, that in spite of relatively high number of analysed transcription factors (48 in total) from different families (21 in total), in 1st place: number of observed aberrations had very low variability, in 2nd place: no significant aberrations were found between pollen defects of late and early transcription factors, in the 3rd place: no significant aberrations were found between pollen defects of individual transcription factor families, in the 4th place: remarkable mutant lines mostly demonstrated complete range of defects not just one defect type, in the 5th place: in most cases just the amount of all defects significantly differentiated interesting mutant lines from the uninteresting ones.

Only 28 out of 60 confirmed lines demonstrated more significant phenotype defects (i.e. ≥25% aberrant pollen) including four lines producing remarkable high proportion (70-100%) of disturbed pollen. Lower occurrence of defective pollen grains in interesting mutant lines compared to wild-type pollen grains does not correspond with 100% penetrance of defect(s) in homozygous plants. Observed situation fully complies with generally accepted idea on lower penetrance of pollen mutations. Literature describes several cases of incomplete penetrance, e.g. *hapless5* mutation with 25% morphologically aberrant pollen (Johnson et al.

2004), *limpet pollen* with 25% abnormal phenotype (Howden et al. 1998), *gemini pollen I* mutants showing 40% (Park et al. 1998; Park et al. 2004) and *sidecar pollen* showing 80% aberrant pollen (Chen and McCormick 1996). Reduced penetrance in pollen mutations is apparent especially in aberration study applying lines, bearing *quartet* mutation as a background for segregation analysis (Francis et al. 2006; Rhee and Somerville 1998).

Despite a low number of lines with more significant number of pollen grain phenotype aberrations, we detected two interesting lines, which particularly concerned SAIL_1168_C11 line with insert in the gene At1g70790 (*depoll* mutant) and SALK_18864 líniu with insert in the gene At2g42380 (*atbzip34* mutant). Both lines were functionally characterized. The first gene has been the subject as well as the output of my study of specific late transcription factors in male gametophyte. The second gene and its functional characterisation is the subject of Antónia Gibalová's Ph.D. study.

depoll mutant as well as *atbzip34* mutant show higher rate of aberrations in all observed phenotype signs of pollen grains in comparison with the wide-type pollen of *Arabidopsis thaliana*, but male sterility effects did not appear in any of the analysed mutants. In *depoll* mutant the number of aborted pollen grains 14.1%, 2-nuclear pollen grain 18%, pollen grain with eccentric MGU 4.8%, with misshaped MGU 19.7%, with linear MGU 16.9%, 1-nuclear pollen 4% and the total number of pollen grains without abortions in *depoll* mutant are within the range of 22.6%. Around a third of observed plants showed increased number of dead pollen grains in category over 10%, therefore the respective mutation was called *depoll* (DEath POLLen grains, *depoll*). *atbzip34* mutant shows significant aberrations in misshapen and misplaced nuclei, in addition large lipid inclutions in mature pollen grains were observed in this mutant. The subsequent electron microscopy revealed extensive disorders in cell wall micropatterning (Rose 2003; Bedinger 1992; Mohnen 2008; Sandhu et al. 2009; Willats et al. 2001). *atbzip34* mutant shows wrinkled nexine and rare and deformed baculae and tecta, *depoll* mutant shows disorders in regular shaping of entine and/or creation of functional connection of plasma membrane and a pollen cell wall. *atbzip34* mutant's disorders are most apparent in pollen ultra-structural organization of endomembrane system.

AtbZIP34 gene has been the first identified transcription factor significant in shaping and functionality of male gametophyte in our laboratory (Gibalova et al. 2009). *atbzip34* mutant falls within a relatively small family of gametophyte mutations, which cause significant morphological and functional aberrations of pollen grains (Takeda et al. 2006;

Brownfield et al. 2009a; Brownfield and Twell 2009; Durbarry et al. 2005; Rotman et al. 2005; Yang et al. 2003; Johnson et al. 2004). It should however be noted, that its sporophytic mode of action was confirmed at the same time on the basis of performed promotor analysis. (Gibalova et al. 2009). One of examples of dual functionality of the gene is *AtbZIP60* gene that was also expressed in tapetum and male gametophyte besides number of other tissues (Iwata et al. 2008).

AtbZIP34 falls within bZIP protein family, which together with C3H, C2H2 zinc finger, WRKY and TCP forms a group of families with high number of members expressed during male gametophyte development (Honys and Twell 2004). Basic leucine zipper (bZIP) proteins are exclusively represented in eucaryotic organisms (Correa et al. 2008), where they perform a wide range of physiological, biological and molecular functions, e.g. in abiotic and biotic stress responses (Hossain et al. 2010; Assuncao et al. 2010; Liu and Howell 2010), sugar signaling (Kang et al. 2010), seed maturation (Alonso et al. 2009) and others. AtbZIP34 was the first identified bZIP protein playing proved role in pollen development, whose mutation caused function disorders in pollen germination and pollen tube growth a transmission defect in *Arabidopsis* pollen development. Another bZIP transcription factor has currently been identified in tobacco (Iven et al. 2010), in particular the *BZI-1* gene acting in regulation of carbohydrate supply of the developing pollen. Like *AtbZIP34* gene it shows expression in tapetum and pollen. Very interesting fact is, that the respective gene acts either as a negative or positive transcriptional regulators during pollen development (Iven et al. 2010).

The second identified transcription factor in our study was the C2H2 zinc finger protein encoded by At1g70790 gene (AtTFDB Arabidopsis Gene Regulatory Information Server (AGRIS**)** (http://arabidopsis.med.ohio-state.edu/AtTFDB/AtTFBrowseResults.jsp?fam=C2H2) (Palaniswamy et al. 2006; Davuluri et al. 2003a). Its bio-informatic analysis of protein domains and prediction of 3-dimensional protein structure however strongly denied this idea (ExPASy PROSITE, http://www.expasy.ch/prosite/, (Gasteiger et al. 2003), I-TASSER, http://zhang.bioinformatics.ku.edu/I-TASSER/, (Zhang 2008)). Just the C2 domain calcium-binding motif in the N-tail (region 1 aa – 88 aa) has been identified in DEPOLL protein.

The C2 domain together with SH2, PTB, PH, SH3, WW and PDZ domains, are typical examples of intracellular protein modules (Rizo and Sudhof 1998). C2 domain was primarily

identified as a Ca^{2+} - binding motif and phospholipid binding to many C2 domains is regulated by Ca^{2+}. But subsequently they were diverged evolutionarily into Ca^{2+} -dependent and Ca^{2+} -independent forms that interact with multiple targets (Ca^{2+}, phospholipids, inositol polyphosphates, intracellular proteins) (Nalefski and Falke 1996). C2 domain was found in many cellular proteins involved in signal transduction or membrane trafficking (synaptotagmins (Perin et al. 1990), rabphilin-3(Shirataki et al. 1993), PLC (Rhee et al. 1989), Ras-GAP (Trahey et al. 1988), PKC (Coussens et al. 1986)), ubiquitin ligation (Nalefski and Falke 1996; Rizo and Sudhof 1998). One of the following tasks of our study will be to understand the role of DEPOLL protein with C2 domain within the male gametophyte development and make an ultimate confirmation or disconfirmation of its DNA-binding ability.

Year by year microarray experiments proved to be a suitable tool to detect genes involved in different physiological and pathogenic processes (Jiang et al. 2009; Jiang and Deyholos 2006, 2009) or contributed to detection of gene function with respect to particular regulatory and/or metabolic pathways they affect (Alves-Ferreira et al. 2007). Therefore microarray experiment employing Affymetrix ATH1 GeneChip 23K was included in the *atbzip34* and *depoll* mutant function analysis. The most affected metabolic pathways in *atbzip34* mutant were those related to cellular transport and/or lipid metabolism. Obtained information fully complied with the observed changes in structure of endomembrane system and lipid bodies. Several groups of significantly upregulated gene groups with proved or hypothetic relation to shaping of cell wall were discovered in *depoll* mutant, in particular the lipid transfer proteins (Arondel and Kader 1990; Thoma et al. 1993), polygalacturonase (Hadfield and Bennett 1998), invertase/pectin methylesterase inhibitor proteins (Hothorn et al. 2004), xyloglucan endotransglucosylase/hydrolase gene family (Campbell and Braam 1999).

DEPOLL protein unlike AtbZIP34 protein, which appears to be an activator of metabolic, and structural genes, shows signs of repressor of signalling pathways with relation to metabolism of pollen grain wall. Its *knockout* has brought significantly higher expression of genes related to shaping of pollen grain wall. It has become apparent in changes observable on pollen grain wall entine, however its cell wall structure in aperture area seems to be undisturbed. This complies with information, that neither sprouting of pollen grains, shaping nor the growth of pollen tube in *depoll* mutant is affected. Our transmission analysis has however revealed an interesting fact, that its pollen grains show increased level of vitality.

This seems to contrast with the increased number of observed defects e.g. dead pollen grains, which results in shortened length of *depoll* mutant siliques. *depoll* mutation thus causes defects of male gametophyte on one side (dead pollen grains), however viability increases in other surviving grainsWe suggested that not presence of DEPOLL protein lead to higher expression of many genes with relevance to cell wall metabolism. Their coordinated activities are distinguishable in elevated mutant pollen tube fitness.

The research revealed a few more interesting putative late transcription factors with respect to development aberrations observed in pollen. These concern genes At2g40620 (bZIP protein), At3g10470 (C2H2 zinc finger protein), At3g57390 (MADS-box protein), At4g05330 (C2H2 zinc finger protein), At5g54680 (bHLH protein) and At1g35490 (bZIP protein). Identification of such genes represents an interesting base for discovery of other significant components of male gametophyte regulatory network.

The presented results of our study demonstrate, that in spite of its extensiveness (48 putative transcription factors examined), we detected only several genes, whose complete or partial *knockout* and associated disturbance of their function affected the level of functionality and morphology of male gametophyte. Regarding the scope of the research running in the area of male gametophyte study, the number of hitherto described gametophyte mutations against the number of available mutants and modern technologies is very low. One of the primary causes may be the significant functional redundance of transcription factors with respect to their importance, they satisfy for the sake further continuation of vegetation life. The other striking problem is the fact, that most of *Arabidopsis* mutants show no visible, directly informative phenotype (Bouche and Bouchez 2001).Progress in methods and procedures for the identification of new pollen mutants can help us understand and uncover the developmental and regulatory mechanisms of male gametophyte lifecycle. Therefore, progressive study on male gametophyte represents a huge challenge in pollen biology.

6. Conclusions

My thesis was devoted to the investigation of a complex regulatory network of transcription factors in the *Arabidopsis thaliana* male gametophyte. In addition to the above-mentioned objective, the second aim of the thesis was the creation of a web-based transcriptome database and data mining tool characterized by an innovative concept of presentation and evaluation of the data from DNA chips.

The results of my thesis can be summarized as follows:

1. We created a database called Arabidopsis Gene Family Profiler (arabidopsisGFP) (http://agfp.ueb.cas.cz/) which was made available on-line. This database was developed in order to provide a new tool for presentation of gene expression data (transcriptome), allowing the user to display the data in a quick and intuitive way using the innovative and when launched also unique concept of a virtual plant. The database employs a progressive idea "from simple to complex“, but also provides the user with more traditional data presenting options. The aGFP database can serve as a reference manual for the expression of individual genes and gene families in *Arabidopsis thaliana* under physiological conditions at different phases of the life cycle and in individual organs and tissues or cell types.

2. A chapter reviewing available servers for deposition of the transcriptome data and on-line tools for their analysis is a part of the submitted PhD thesis. Substantial part of the text deals with the expression data originating from plants, primarily from *Arabidopsis thaliana*. Our review is intended to help the user to gain quick overview of the available tools and databases and to summarize the potential and advantages as well as disadvantages of their use. By this approach, we wish to make easier access for a common user to the complex transcripromic datasets in user friendly environment. The next result of my PhD thesis is the analysis of the regulatory network of late transcription factors which regulate development of the male gametophyte after the second pollen mitosis (PM II). *In-silico* identification and selection of the transcription factors *w*as followed by genotypic and phenotypic evaluation of mature pollen grains from individual T-DNA lines with knock-out function of several transcription factors. The total number

of the analyzed T-DNA lines is 37. They represent 21 early transcription factors from several gene families (C2H2 zinc finger, bZIP, MADS-box protein family, etc.). I have found and described several transcription factors in which the insertion mutation seriously affected the development and/or structure of the pollen grain (e.g. aborted pollen grains, changes in position and spatial arrangement of the male germ unit (MGU), increased number of pollen grains with one or two nuclei). The following genes from various metabolic and signaling pathways have been identified as potential candidates for pollen-specific developmental regulators: At2g40620 (bZIP protein), At3g10470 (C2H2 zinc finger protein), At3g57390 (MADS-box protein), At4g05330 (C2H2 zinc finger protein), At5g54680 (bHLH protein) and At1g35490 (bZIP protein).

3. The next result presented in the PhD thesis is a published paper on functional characterization of the *atbZIP34* mutant. My contribution to the project consisted in the phenotypic analysis of the mature pollen grains and subsequent evaluation of the analysis. *atbZIP34* mutant was discovered during analysis of the transcription factors network in the male gametophyte of *Arabidopsis thaliana*. Its mature pollen grains showed number of abnormalities observed in both fluorescence and electron microscopy, for example, defects in the shape and form of exine or reduced endomembrane system. Based on further results of my colleague, Ms Antonia Gibalova, the first author of the respective publication, and on results of other co-authors , the proposed role of the *AtbZIP34* factor is in the regulation of exine formation and its involvement in the regulation of the lipid metabolism and/or subcellular transport. Further analysis of the *AtbZIP34* transcription factor is the aim of the PhD thesis of Ms Antonia Gibalova.

4. My previous analysis of the transcription factors (see paragraph 3) discovered the interesting T-DNA line which showed high ratio of dead pollen grains and other phenotypic abnormalities when compared to the wild type pollen of *Arabidopsis thaliana*. The identified line was SAIL_1168_C11 which carries insertion in the At1g70790 gene. My initial idea that it was a C2H2 transcription factor, confirmed by available transcription factor databases, was not later confirmed by further *in-silico* analysis. With the help of bioinformatic tools I only discovered the C2 domain, which currently rules out the possibility that the gene functions as a transcription regulatory unit. My next experiments which are not the subject of this submitted PhD thesis will be focused on disproving the presence or finding a new DNA binding domain. The result of this part of

the PhD thesis is functional characterization of the At1g70790 gene which we call, based on the increased number of dead pollen grains, *DEPOLL* (DEath POLLen grains). The result of my research is a finding that the protein product of the *DEPOLL* gene is responsible for proper formation of the intine and/or functional coupling of the plasma membrane and the pollen grain wall. The consequent analysis of the expression data from the *depoll* pollen grains on the Affymetrix ATH1 array revealed several groups of genes with confirmed or expected relationship to the metabolism of the pollen grain wall (lipid transfer protein (LTP), polygalacturonase, invertase/pectin methylesterase inhibitor proteins, xyloglucan endotransglucosylase/hydrolase gene family). Examination of these genes represents further potential for the discovery of specific components which are essential for proper formation of the pollen grain intine including its functional coupling with the plasma membrane. Last but not least, my next step will be uncovering of the specific function of the DEPOLL protein which, with high probability, functions as a transcription inhibitor of many genes which are related to the formation of the pollen grain wall and its involvement in specific metabolic and/or signaling pathways.

7. Záver

Moja dizertačná práca bola venovaná zvýšeniu poznatkov o fungovaní komplexnej regulačnej siete transkripčných faktorov v samčom gametofyte *Arabidopsis thaliana*. Transkripčné faktory sú riadiacou jednotkou mnohých významných zmien nielen v rastlinnom organizme a sú zodpovedné za ich správny časový a priestorový priebeh. Ich odhalenie vrátane ich podrobnej funkčnej charakterizácie je dôležitou súčasťou nášho poznania. Okrem vyššie spomenutej snahy, ďalšou úlohou dizertačnej práce bolo vytvorenie inovatívneho konceptu prezentácie a hodnotenia dostupných expresných dát pochádzajúcich z DNA microarray čipov a prevedenie tohto prístupu do praxe vo forme on-line databázy a hodnotiacich nástrojov pre tieto expresné dáta.

Výsledky mojej dizertačnej práce je možné zhrnúť do následujúcich bodov:

1. Bola vytvorená on-line sprístupnená databáza nazvaná Arabidopsis Gene Family Profiler (arabidopsisGFP, http://agfp.ueb.cas.cz/). Táto databáza vznikla za účelom poskytnutia nového nástroja pre prezentáciu transkriptomických dát, ktorý umožní užívateľovi rýchly a intuitívny spôsob ich zobrazenia pomocou virtuálnej rastliny. Virtuálna rastlina využíva inovatívny koncept "od jednoduchému k zložitému", ale zároveň ponúka užívateľovi aj tradičné možnosti zobrazenia expresných dát. Naša databáza može slúžiť ako referenčná príručka expresie jednotlivých génov a génových rodín v rastline *Arabidopsis thaliana* za fyziologických podmienok v rôznych časových obdobiach života rastliny a s ohľadom na jej orgány, pletivá či jednotlivé bunky.

2. Ďalším výstupom mojej dizertačnej práce bola analýza siete neskorých transkripčných faktorov regulujúcich správny vývin samčieho gametofytu po druhej peľovej mitóze (PMII). *In-silico* identifikácia a výber týchto transkripčných faktorov bol zavŕšený genotypovým a fenotypovým hodnotením zrelých peľových zŕn jednotlivých T-DNA inzerčných línií s vyradenou funkciou rozličných transkripčných faktorov. Celkovo som analyzovala 37 T-DNA línií, ktoré prezentovali 21 neskorých transkripčných faktorov a zastupovali rôzne génové rodiny (C2H2 zinc finger, bZIP proteínová rodina, MADS-box proteínová rodina a ďalšie). Našla som a popísala niekoľko transkripčných faktorov, ktorých inzerčná mutácia spôsobila výrazné zmeny na vzhľade peľového zrna (napr. mŕtve peľové zrná, zvýšený počet peľových zŕn s dvoma alebo jedným jadrom a ďalšie

typy odchýlok). K potencionálnym kandidátom pre nájdenie ďalších peľovo špecifických regulátorov vývinu v rôznych metabolických či signálnych cestách patria At2g40620 gén (bZIP proteín), At3g10470 gén (C2H2 zinc finger proteín), At3g57390 gén (MADS-box proteín), At4g05330 gén (C2H2 zinc finger proteín), At5g54680 gén (bHLH proteín) alebo At1g35490 gén (bZIP proteín).

3. Ďalším výstupom vo forme článku, ktorý prezentujem v tejto dizertačnej práci je funkčná charakterizácia *atbZIP34* mutanta. Na tejto práci som sa podieľala fenotypovou analýzou zrelých peľových zŕn a jej vyhodnotením. *atbZIP34* mutant bol odhalený v priebehu výskumu siete transkripčných faktorov v samčom gametofyte *Arabidopsis thaliana*. Jeho zrelé peľové zrná vykazujú celý rad odchýlok pozorovaných na úrovni fluorescenčnej a elektrónovej mikroskopie, napríklad defekty v tvare a forme exiny či redukovaný endomembránový systém. Na základe dostupných výsledkov mojej kolegyne Antónie Gibalovej, prvej autorky príslušnej publikácie a ďalších členov autorského kolektívu je predpokladaná úlohu *AtbZIP34* faktora v správnom formovaní exiny a jeho zapojenie v regulácii lipidového metabolizmu a/alebo bunkového transportu. Ďalšia analýza transkripčného faktora *AtbZIP34* je predmetom dizertačnej práce Antónie Gibalovej.

4. Moja predchádzajúca analýza transkripčných faktorov (viď odstavec 2) odhalila zaujímavú T-DNA líniu, ktorá vykazovala vyšší podiel mŕtvych peľových zŕn a ďalších odchýlok vo fenotype zrelých peľových zŕn v porovnaní s divokými rastlinami *Arabidopsis thaliana*. Jednalo sa o líniu SAIL_1168_C11, ktorá obsahuje inzerciu v géne At1g70790. Moju pôvodnú predstavu, že ide o transkripčný faktor C2H2 potvrdenú dostupnými databázami transkripčných faktorov som následnými *in-silico* analýzami nepotvrdila. Pomocou bioinformatických nástrojov som v tomto géne odhalila len C2 doménu, čo aspoň zatiaľ jednoznačne vylučuje jeho funkciu ako transkripčnej regulačnej jednotky. Moje ďalšie testy, ktoré už nie sú predmetom predloženej dizertačnej práce budú zamerané na definitívne vylúčenie alebo nájdenie novej DNA väzbovej domény. Výstupom tejto časti dizertačnej práce je funkčná charakterizácia At1g70790 génu, ktorý sme vzhľadom na zvýšený počet mŕtvych peľových zŕn nazvali *DEPOLL* (DEath POLLen grains) gén. Výsledkom môjho skúmania bolo zistenie, že bielkovinový produkt *DEPOLL* génu je zodpovedný za správne utváranie intiny a/alebo funkčného prepojenia plazmatickej membrány a steny peľového zrna. Následná analýza expresných dát *depoll*

peľových zŕn pochádzajúcich z DNA čipu odhalila niekoľko skupín génov s potvrdeným alebo predpokladaným vzťahom k metabolizmu steny peľového zrna (lipid transferové proteíny, polygalakturonázy, invertázy/pektín metylesterázové inhibičné proteíny, xyloglukán endotransglukozylázy/hydrolázy). Výskum týchto génov predstavuje ďalší potenciál pre nájdenie špecifických komponentov nevyhnutných pre správne utváranie intiny peľového zrna vrátane jeho funkčného prepojenia s plazmatickou membránou. V neposlednom rade bude mojou ďalšou úlohou zistenie zapojenia a funkcie DEPOLL proteínu s C2 doménou v konkrétnych metabolických a/alebo signálnych cestách.

8. Chapters including papers

8.1. Arabidopsis Gene Family Profiler (aGFP) – user-oriented transcriptomic database with easy-to-use graphic interface

Dupl'áková Nikoleta, Reňák David, Hovanec Patrik, Honysová Barbora, Twell David, Honys David

BMC Plant Biology 2007, **7**:39

This paper presents the Arabidopsis Gene Family Profiler database and toolbox developed for the microarray data visualisation. It describes the general concept of the database and programming algorithm. Moreover, it gives the precise description of available graphical and gene definition tools offered by the aGFP database.

My contribution: I participated in the concept definition of the arabidopsisGFP database and in the specification the data visualization concepts. I was fully responsible for the programming of all scripts and gene families and superfamilies definition. I was also involved in the manuscript preparation.

Supporting online material on CD

BMC Plant Biology

BioMed Central

Database

Arabidopsis Gene Family Profiler (aGFP) – user-oriented transcriptomic database with easy-to-use graphic interface

Nikoleta Dupl'áková[1,2], David Reňák[1,2,3], Patrik Hovanec[4], Barbora Honysová[5], David Twell[6] and David Honys*[1,2]

Address: [1] Laboratory of Pollen Biology, Institute of Experimental Botany ASCR v.v.i., Rozvojová 263, 165 00 Prague 6, Czech Republic, [2]Department of Plant Physiology, Faculty of Science, Charles University in Prague, Viničná 5, 128 44, Prague 2, Czech Republic, [3]Department of Plant Physiology and Anatomy, Faculty of Biological Sciences, University of South Bohemia, Branišovská 31, 370 05 ŽKeské BudEjovice, Czech Republic, [4]Krymská 8/122, 101 00 Praha 10, Czech Republic, [5]Laboratory of Cell Biology, Institute of Experimental Botany ASCR v.v.i., Rozvojová 263, 165 00 Prague 6, Czech Republic and [6]Department of Biology, University of Leicester, Leicester, LE1 7RH, UK

Email: Nikoleta Dupl'áková - duplakova@ueb.cas.cz; David Reňák - renak@ueb.cas.cz; Patrik Hovanec - patrikhov@gmail.com; Barbora Honysová - honysova@ueb.cas.cz; David Twell - twe@leicester.ac.uk; David Honys* - honys@ueb.cas.cz

* Corresponding author

Published: 23 July 2007

BMC Plant Biology 2007, **7**:39 doi:10.1186/1471-2229-7-39

Received: 2 February 2007
Accepted: 23 July 2007

This article is available from: http://www.biomedcentral.com/1471-2229/7/39

Abstract

Background: Microarray technologies now belong to the standard functional genomics toolbox and have undergone massive development leading to increased genome coverage, accuracy and reliability. The number of experiments exploiting microarray technology has markedly increased in recent years. In parallel with the rapid accumulation of transcriptomic data, on-line analysis tools are being introduced to simplify their use. Global statistical data analysis methods contribute to the development of overall concepts about gene expression patterns and to query and compose working hypotheses. More recently, these applications are being supplemented with more specialized products offering visualization and specific data mining tools. We present a curated gene family-oriented gene expression database, Arabidopsis Gene Family Profiler (aGFP; http://agfp.ueb.cas.cz), which gives the user access to a large collection of normalised Affymetrix ATH1 microarray datasets. The database currently contains NASC Array and AtGenExpress transcriptomic datasets for various tissues at different developmental stages of wild type plants gathered from nearly 350 gene chips.

Results: The Arabidopsis GFP database has been designed as an easy-to-use tool for users needing an easily accessible resource for expression data of single genes, pre-defined gene families or custom gene sets, with the further possibility of keyword search. Arabidopsis Gene Family Profiler presents a user-friendly web interface using both graphic and text output. Data are stored at the MySQL server and individual queries are created in PHP script. The most distinguishable features of Arabidopsis Gene Family Profiler database are: 1) the presentation of normalized datasets (Affymetrix MAS algorithm and calculation of model-based gene-expression values based on the Perfect Match-only model); 2) the choice between two different normalization algorithms (Affymetrix MAS4 or MAS5 algorithms); 3) an intuitive interface; 4) an interactive "virtual plant" visualizing the spatial and developmental expression profiles of both gene families and individual genes.

Conclusion: Arabidopsis GFP gives users the possibility to analyze current Arabidopsis developmental transcriptomic data starting with simple global queries that can be expanded and further refined to visualize comparative and highly selective gene expression profiles.

Background

Completion and annotation of the *Arabidopsis thaliana* genome represented a major step in plant genetic research [1]. This knowledge enabled gene prediction, assignment of functional categories and gave an opportunity to study gene and chromosome organization including the distribution of transposable elements. Finally it has enabled the characterization of global gene expression patterns at the transcriptome level at different developmental stages and under various physiological and stress conditions. Efforts to reveal the biological functions of thousand of genes and their integration into proteome, metabolome and interactome networks has become the principal focus of many studies and represents key objective of the 2010 Project [2]

A number of efficient and accurate gene expression analysis technologies to determine the expression levels of individual genes have been widely exploited in recent decades (Northern blot analysis, quantitative reverse transcription-PCR, cDNA library screening). Most of these methods enable analysis of the expression of single or relatively few selected genes. For the discovery of partial or whole gene functional or regulatory networks, the development of high-throughput technologies is essential with genome-wide transcriptomic studies providing a major input [3]. Several such methods have been developed including, cDNA fingerprinting [4], serial analysis of gene expression – SAGE [5], massively parallel signature sequencing – MPSS [6], high-density DNA oligonucleotide probe microarrays [7,8] or cDNA arrays [9]. DNA microarray technologies are among the most frequently used methods for parallel global analysis of gene expression. These methods are based on the principle of selective and differential hybridization between sample target molecules and immobilized DNA probes. Hybridisation to probes arrayed on a solid surface report the relative abundance of DNA or RNA target molecules by fluorescent signal detection [10,11]. Microarray technologies now belong to the standard functional genomics toolbox [12,13] and have undergone massive development leading to increased genome coverage, accuracy and reliability. Whole Genome microarrays developed by Affymetrix (Santa Clara, CA, USA) in collaboration with Syngenta represented the first standard in genome wide transcriptomic studies in plants. Whole genome Affymetrix ATH1 GeneChips cover about 76% of the *Arabidopsis thaliana* genes [14]. Moreover, the introduction of the Minimum Information About Microarray experiments (MIAME) as standard documentation for array experiments and in transcriptomic databases, increasing the value and comparability of microarray data [15].

The number of experiments exploiting microarray technology has markedly increased in recent years. Not surprisingly, there are potential difficulties in navigating between different available data sets. Microarray expression data are deposited on servers, many of which are publicly accessible. Public plant microarray data are deposited in several databases including ArrayExpress [16], GEO [17], NASCarrays [18] and the Stanford Microarray Database [19-21]. Currently these databases store several thousands of individual datasets and some of these offer on-line tools for data normalization, filtering, statistical testing and pattern discovery [22-26].

In parallel with the rapid accumulation of transcriptomic data, on-line analysis tools are being introduced to simplify their use. Global statistical data analysis methods contribute to the identification of overall gene expression patterns and to query and compose new working hypotheses based on these findings [11,12,27,28]. More recently, these applications are being supplemented with more specialized products offering visualization and specific data mining tools. Genevestigator, Botany Array Resource, Arabidopsis co-expression tool, and Expression Profiler offer Web-based tools to analyse large microarray datasets. Genevestigator offers two types of queries: a gene-centric approach and a genome-centric approach, which are represented by several analysis tools; Gene Correlator, Gene Atlas, Gene Chronologer, Response Viewer and the Meta-Analyzer, that is among the most sophisticated complex amongst available microarray analysis toolboxes [29,30]. Botany Array Resource offers similar services supplemented with tools for discovery and analyses of cis-elements in promoters [31]. Expression Profiler (EP) provides tools for hierarchical and K-groups clustering, clustering comparison, similarity search or the signature algorithm [32,33]. The more specialized PathoPlant database on plant-pathogen interactions and components of signal transduction pathways related to plant pathogenesis also harbors gene expression data from *Arabidopsis thaliana* microarray experiments to enable searching for specific genes regulated upon pathogen infection or elicitor treatment [34,35]. Finally, Arabidopsis Co-Expression Tool (ACT) allows users to identify genes with expression patterns correlated across selected experiments or the complete data set and offers the novel clique finder tool [36-38].

In this article we introduce Arabidopsis Gene Family Profiler (arabidopsisGFP, aGFP), a web-based gene expression database with visualization tools. During programming, we took into account that for many microarray data users, extraction of global expression patterns of single genes or gene families can be time-consuming and its visualization difficult. Moreover, the use of various normalization algorithms in individual experiments makes direct comparison of genes of interest within various datasets uncertain. To solve these issues, we devel-

oped aGFP to provide the user with two normalization and gene detection algorithms and a "virtual plant" Arabidopsis Gene Family Profiler facility that enables users to obtain a global expression profile of user specified and/or pre-defined gene families. These attributes of aGFP contribute a useful resource for the rapid bioinformatic analysis of Arabidopsis gene expression data through comparative expression profile analysis in a gene family-based context.

Construction and content

The Arabidopsis Gene Family Profiler (aGFP) database was designed to give users the possibility to visualize expression patterns of individual genes, pre-defined gene families or user-defined gene sets in various tissues and at different developmental stages of wild type *Arabidopsis thaliana* plants. aGFP largely exploits microarray experiments obtained through the NASC AffyWatch transcriptomics service [39]. We adopted the general concept "from simple to complex". In the first approximation, an arithmetical mean expression signal from multiple experiments is displayed. In subsequent steps the user can choose to display expression data for individual plant organs or tissues at particular growth stages. This is accompanied by the option of progressive replacement of arithmetical means by individual expression values. So the user has the option to choose the different levels of visualization to suit needs. Finally, the user can switch from "virtual plant" visualization to a simple bar chart (standard or log-scaled) or tabulated display and can browse through individual experiments down to normalized or even raw data extracted from individual gene chips. Gene family data can also be visualized as a colorized spot chart.

Although the idea of web-based database tools is not novel, aGFP database offers a quick and interactive display of gene expression profiles using the virtual plant facility as well as alternate more conventional outputs. A novel feature of aGFP is that it enables the evaluation of the impact of normalization procedures on microarray expression data as well the possibility of rapid definition of user-defined families or gene groups. Simultaneously, aGFP serves as a facile and synoptic developmental reference guide for expression profiles of individual genes or gene families in wild-type *Arabidopsis thaliana* plants.

Data resources

The arabidopsisGFP database covers transcriptomic experiments accumulated from wild type *Arabidopsis thaliana* plants of various ecotypes grown under normal physiological conditions. Original raw microarray data were obtained from Nottingham Arabidopsis Stock Centre (NASC) through the AffyWatch service [39]. In order to ensure the quality and compatibility of expression data only microarray experiments using Affymetrix ATH1 whole genome arrays with at least two biological replicates were included. To date, arabidopsisGFP database covers transcriptomic data from 345 microarrays covering 120 experiments.

Programming

aGFP is composed as a relational MySQL database and Web server application, which is programmed in PHP script language [PHP:Hypertext Preprocessor]. Gene expression data are presented by dynamic HTML web pages with several types of graphic output. Graphs were generated using PHP module jpgraph [40]. HTML code was programmed to be correctly displayed in all commonly used internet browsers (Microsoft Internet Explorer/Mozilla Firefox/Opera). The user exploits a web-based interface for acquisition of custom-defined data. A user-friendly intuitive web-based interface is designed to enable simple and rapid navigation in aGFP. The aGFP database was created using general-to-specific strategy enabling the user to obtain a certain amount of information at every step with progressive targeted specification as the query develops.

Data normalization

All gametophytic and sporophytic datasets were normalized using freely available dChip 1.3 software [41]. The reliability and reproducibility of datasets was ensured by the use of duplicate or triplicate hybridization data in each experiment, normalization of all arrays to the median probe intensity and the use of normalized CEL intensities of all arrays for the calculation of model-based gene-expression values based on the Perfect Match-only model [42,43]. A given gene was scored as 'expressed' when it gave a reliable expression signal in all replicates. An expression signal value of '0' means that the detection call value was 'absent' or 'marginal' in at least one replicate provided. In arabidopsisGFP, the facility is provided to instantly switch between transcriptomic data normalized by two different algorithms – MAS 5.0 or MAS 4.0.

Annotation pages

Annotation of individual experiments is in accordance with MIAME standard [15]. *Arabidopsis thaliana* growth stages were according to Boyes et al. [44]. Affymetrix gene chips harbour several oligonucleotide probe types – prevalent unique probe sets (_at) accompanied by identical probe sets (_s set) and probes in a mixed probe set (_x set). Moreover, progressive Arabidopsis genome annotation has led to a reduction in the number of unique probe sets that has resulted in a reduction in the number of reliably 'present' genes. For these reasons, genes represented by these ambiguous probe sets were not included in the database [45]. This fact was taken into account and the aGFP database is regularly updated. Each locus in aGFP

database is associated with relevant annotation released by TAIR (currently version 6) [46], and direct links to other web resources are available for each gene – TIGR [47,48], MPSS [49], TAIR [50,51], MIPS [52,53].

Definition of gene families and superfamilies

arabidopsisGFP contains lists of pre-defined gene families and superfamilies enabling the rapid comparative visualization of expression profiles of their members. Genes in arabidopsisGFP are organized into two hierarchical levels consisting of gene families and superfamilies. All data were assembled from various relevant resources, the majority from TAIR – Arabidopsis Gene Family Information [54] and AGRIS [55,56]. Gene families were further organized in a different manner as gene family subclasses to different extent in each family and source. In order to simplify the different sub-divisions from different data sources, we rearranged them carefully and used only two levels, gene family and superfamily.

Utility and discussion

Data selection

At the aGFP home page, the user can select the search category (AGI number, BAC locus, Gene name, keyword) and two other input parameters; the gene detection algorithm (MAS4.0 or MAS5.0) and the source of expression data (AtGenExpress or NASCArrays). To make aGFP as comfortable as possible to use, at any stage of the query the user has the possibility to directly switch options between these pairs of parameters. This represents a distinct feature of the aGFP database that enables direct comparison of the influence of the detection algorithm or data resource on expression profiles.

aGFP database presents data normalised using two different algorithms, empirical MAS 4.0 and statistical MAS 5.0. Although MAS4.0 is believed to yields more false-positive calls [57], our analyses of four developmental stages of pollen development showed that the MAS5.0 detection algorithm tended to eliminate a number of genes originally detected as expressed by MAS4.0 and which were experimentally verified to be so [58]. This was often the case even for highly expressed genes (B. Honysová and D. Honys, unpublished results), highlighting the added value of the empirical MAS4.0 detection algorithm and comparative analysis.

Experiments included in the aGFP database are presented in two different subsets. The first subset contains data obtained within a scope of the AtGenExpress project [59], the second comprises all other datasets deposited at NASC and was labeled NASCArrays [26]. The reason for this separation was that AtGenExpress contains a structured set of experiments, involving Columbia-0 plants grown under comparable conditions to provide a gene expression atlas at several developmental stages. On the contrary, NASCArrays contains experiments carried out in various ecotypes grown under various conditions. Data in each subset are presented using several different graphical displays and, the user has an option to instantly switch between subsets in each environment (Fig. 1).

The other key feature of aGFP is the possibility to select pre-defined gene families and superfamilies. In subsequent steps, expression data for family members can be extracted down to the level of individual genes. Moreover, the user has also the possibility to work with custom-defined gene sets based on various search options (AGI number, Gene number, BAC locus, keyword search).

Data visualization

The aGFP database provides users with several different visualization formats. Apart from standard tables or bar charts (Fig. 2, 3, 4), an interactive virtual plant is used. The virtual plant comprises several growth stages defined according to Boyes et al. [44]. A white (low)-yellow-green (high) scale is used to depict the relative expression signals of individual genes, or gene families throughout the *Arabidopsis* life cycle. Mouse-over pointing to complex organs/tissues (i.e. flowers) causes opening of more detailed graphics showing individual organs (ie. sepals, petals, stamens, pistils and pollen; Fig. 5). This is in accordance with the adopted aGFP database concept "from simple to complex".

In addition to the virtual plant and bar chart graphics, expression profiles can also be visualized as interactive colorized spot charts or heat maps. The colorized spot chart uses a colour scale identical with that of the virtual

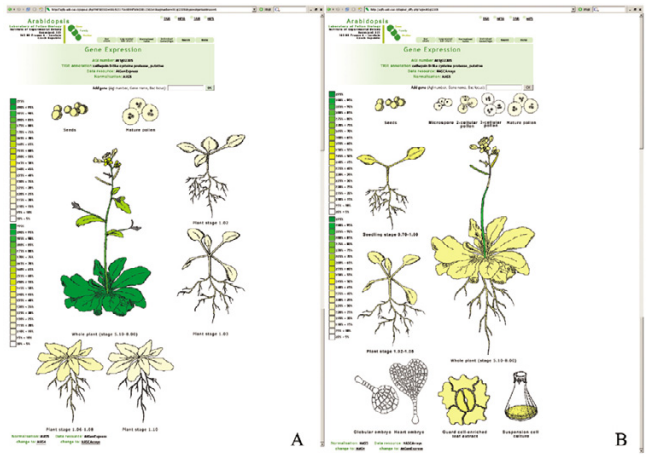


Figure 1
Visualisation of gene search results using interactive "virtual plant". AtGenExpress (A) and NASCArrays (B) expression data for At1g02305 are shown.

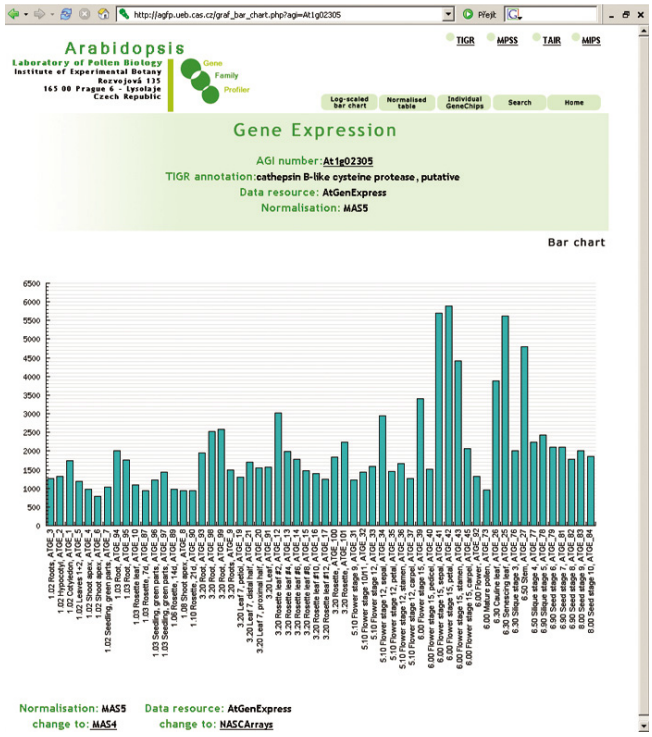


Figure 2
Expression data visualisation. Visualisation of At1g02305 gene expression using Bar chart option.

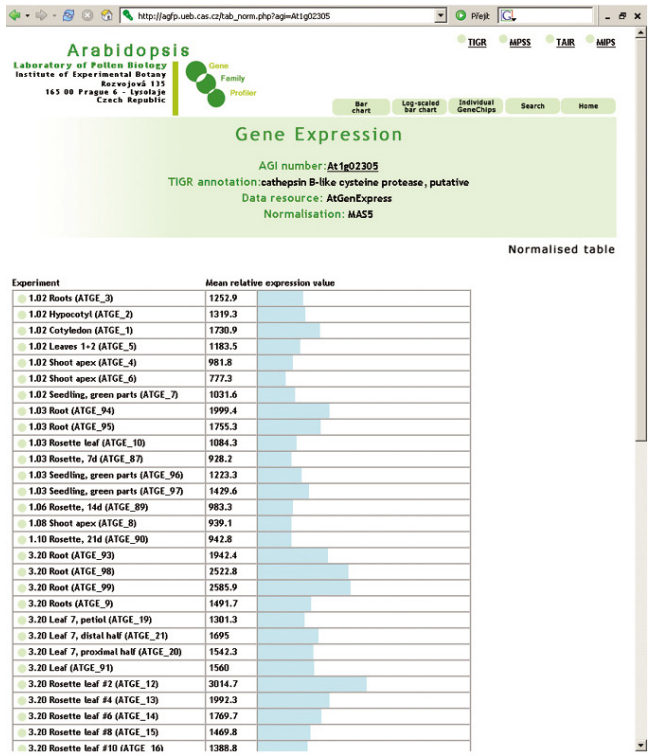


Figure 3
Expression data visualisation. Visualisation of At1g02305 gene expression using Normalised table option.

plant (Fig. 6). The interactivity of modeling is based on the possibility of ad hoc addition and removal of genes to and from a currently active set. Moreover, each spot contains information about expression signal value and experiment that is activated by mouse-over. Therefore visualized expression data are not merely summarized, but instead represent direct output from individual microarray experiments.

Data accessibility and legend

The complete datasets used in the aGFP database are described in the Legend available from the homepage and it is possible to trace the origin of all datasets. Moreover, data can be downloaded for individual and selected gene sets as a TAB-delimited text. This enables the direct import of downloaded data into spreadsheet editors such as Excel and database software such as Access and FileMaker. This text file contains a list of developmental and morphological stages, normalized expression data for the selected normalization algorithm and data source.

Conclusion

arabidopsisGFP is a microarray expression database of wild type *Arabidopsis thaliana* plants grown under physiological conditions. It gathers data from experiments using Affymetrix ATH1 whole genome arrays with two or more biological replicates. From the outset, it has been created as intuitive user-oriented web-tool employing a "general-to-specific" concept enabling the user to obtain certain amount of information at every step with progressive specification and refinement. The aGFP database contains several gene selection and grouping tools including predefined gene families. It also provides the user with different gene expression visualization options including a unique "virtual plant" graphic display. Easy switching of visualization options gives the user the possibility to rapidly select the most suitable form of data presentation. A novel advantage of the aGFP database is the provision of alternative normalization treatments of microarray data using statistical (MAS5.0) and empirical algorithms (MAS4.0). Together with the facile switch between these detection algorithms it provides the opportunity to instantly assess the reliability of gene expression data. Arabidopsis Gene Family Profiler represents a versatile tool for facile visualization of transcriptomic data that can be exploited in genome-led queries of gene and gene family functions and regulation.

Availability and requirements

The aGFP database is freely accessible and its concept offers the possibility to extract and visualize expression

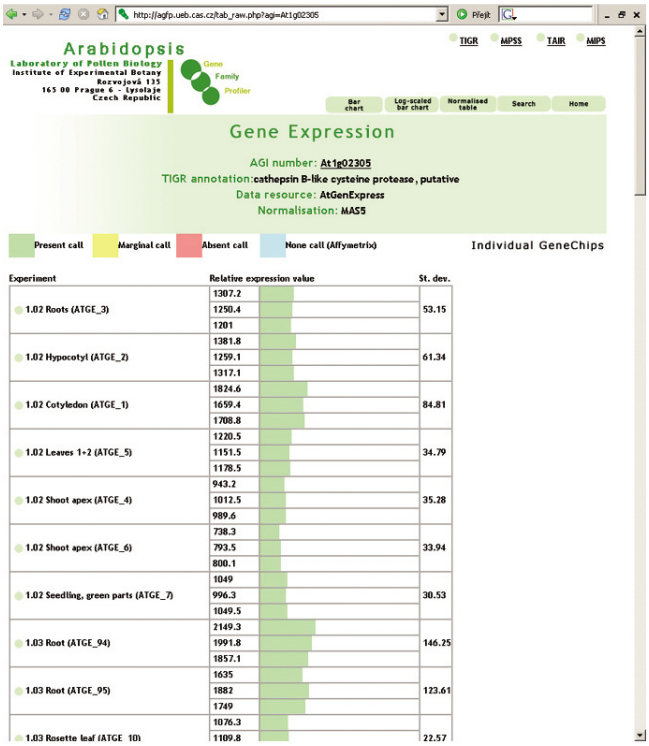


Figure 4
Expression data visualisation. Visualisation of At1g02305 gene expression using Individual GeneChips option.

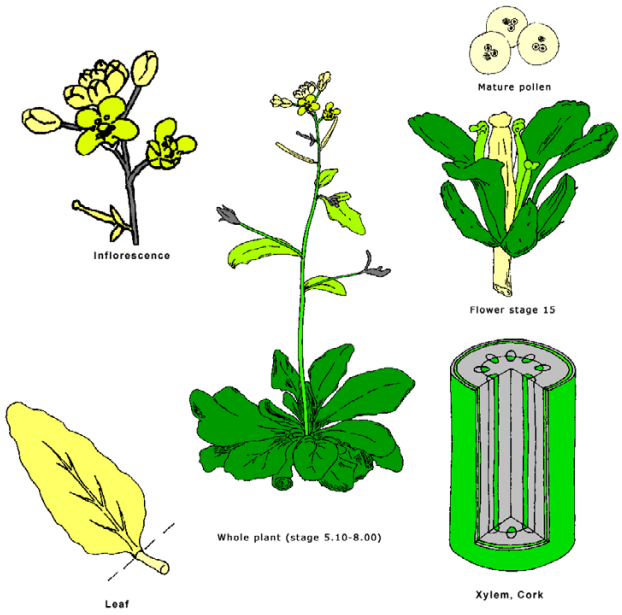


Figure 5
Individual organs visualised by "virtual plant". Detailed expression profile of At1g02305 visualised in selected organs.

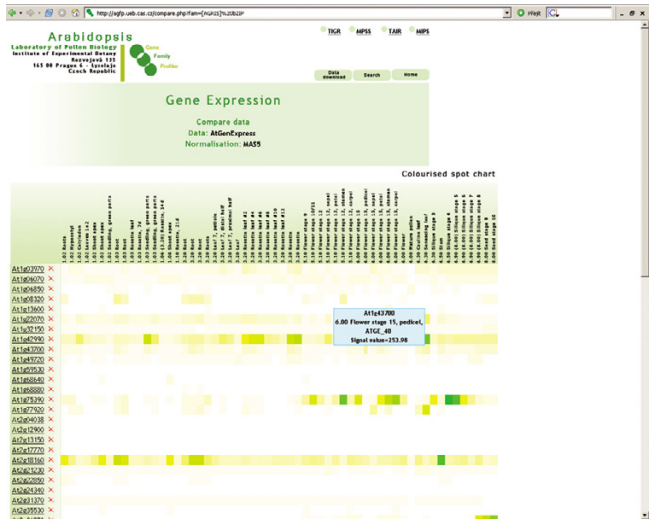


Figure 6
Colourised spot chart visualization. Visualization of expression profiles of bZIP gene family members (AtGenExpress).

profiles of individual genes, gene sets, gene families or gene superfamilies from a broad spectrum of microarray experiments covering various *Arabidopsis* organs, tissues and developmental stages. For these purposes, an innovative graphic concept of the "virtual plant" was introduced representing a clear and simple visualization of gene expression profiles in a morphological and developmental context. The arabidopsisGFP database is accessible at http://agfp.ueb.cas.cz/.

Authors' contributions

ND, DT and DH defined the concept of arabidopsisGFP database. ND and PH programmed all scripts. DH is responsible for data normalisation. DR worked on gene families and superfamilies definition. DH, DT, ND and BH specified the data visualization concepts and BH drew the "virtual plant". All authors read and approved the final manuscript.

Acknowledgements

Authors gratefully acknowledge the financial support from Grant Agency of The Czech Republic (522/06/0894), MSMT CR (LC06004) and The Royal Society (Joint Project 2004/R3-EU).

References

1. **Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.** *Nature* 2000, **408(6814):**796-815.
2. **Project 2010. To determine the functions of all genes in Arabidopsis thaliana by the year 2010** [http://www.nsf.gov/pubs/2006/nsf06612/nsf06612.htm]
3. Donson J, Fang Y, Espiritu-Santo G, Xing W, Salazar A, Miyamoto S, Armendarez V, Volkmuth W: **Comprehensive gene expression analysis by transcript profiling.** *Plant Mol Biol* 2002, **48(1-2):**75-97.
4. Money T, Reader S, Qu LJ, Dunford RP, Moore G: **AFLP-based mRNA fingerprinting.** *Nucleic Acids Res* 1996, **24(13):**2616-2617.

5. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270(5235):**484-487.
6. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays.** *Nat Biotechnol* 2000, **18(6):**630-634.
7. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: **Expression monitoring by hybridization to high-density oligonucleotide arrays.** *Nat Biotechnol* 1996, **14(13):**1675-1680.
8. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: **High density synthetic oligonucleotide arrays.** *Nat Genet* 1999, **21(1 Suppl):**20-24.
9. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270(5235):**467-470.
10. van Hal NL, Vorst O, van Houwelingen AM, Kok EJ, Peijnenburg A, Aharoni A, van Tunen AJ, Keijer J: **The application of DNA microarrays in gene expression analysis.** *J Biotechnol* 2000, **78(3):**271-280.
11. Clarke JD, Zhu T: **Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems: practical considerations and perspectives.** *Plant J* 2006, **45(4):**630-650.
12. Zhu T: **Global analysis of gene expression using GeneChip microarrays.** *Curr Opin Plant Biol* 2003, **6(5):**418-425.
13. Aharoni A, Vorst O: **DNA microarrays for functional plant genomics.** *Plant Mol Biol* 2002, **48(1-2):**99-118.
14. Hennig L, Menges M, Murray JA, Gruissem W: **Arabidopsis transcript profiling on Affymetrix GeneChip arrays.** *Plant Mol Biol* 2003, **53(4):**457-465.
15. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M: **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.** *Nat Genet* 2001, **29(4):**365-371.
16. **ArrayExpress** [http://www.ebi.ac.uk/arrayexpress/?]
17. **Gene Expression Omnibus** [http://www.ncbi.nlm.nih.gov/geo/]
18. **NASCarrays** [http://affymetrix.arabidopsis.info/]
19. Gollub J, Ball CA, Sherlock G: **The Stanford Microarray Database: a user's guide.** *Methods Mol Biol* 2006, **338:**191-208.
20. Ball CA, Awad IA, Demeter J, Gollub J, Hebert JM, Hernandez-Boussard T, Jin H, Matese JC, Nitzberg M, Wymore F, Zachariah ZK, Brown PO, Sherlock G: **The Stanford Microarray Database accommodates additional microarray platforms and data formats.** *Nucleic Acids Res* 2005, **33(Database issue):**D580-2.
21. **Stanford Microarray Database** [http://genome-www5.stanford.edu/]
22. Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara GG, Holloway E, Kapushesky M, Lilja P, Mukherjee G, Oezcimen A, Rayner T, Rocca-Serra P, Sharma A, Sansone S, Brazma A: **ArrayExpress--a public repository for microarray gene expression data at the EBI.** *Nucleic Acids Res* 2005, **33 (Database Issue):**D553-5.
23. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R: **NCBI GEO: mining millions of expression profiles--database and tools.** *Nucleic Acids Res* 2005, **33 (Database Issue):**D562-6.
24. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30(1):**207-210.
25. Gollub J, Ball CA, Binkley G, Demeter J, Finkelstein DB, Hebert JM, Hernandez-Boussard T, Jin H, Kaloper M, Matese JC, Schroeder M, Brown PO, Botstein D, Sherlock G: **The Stanford Microarray Database: data access and quality assessment tools.** *Nucleic Acids Res* 2003, **31(1):**94-96.
26. Craigon DJ, James N, Okyere J, Higgins J, Jotham J, May S: **NASCArrays: a repository for microarray data generated by NASC's transcriptomics service.** *Nucleic Acids Res* 2004, **32(Database issue):**D575-7.
27. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, Friend SH: **Functional discovery via a compendium of expression profiles.** *Cell* 2000, **102(1):**109-126.
28. Mandaokar A, Thines B, Shin B, Markus Lange B, Choi G, Koo YJ, Yoo YJ, Choi YD, Browse J: **Transcriptional regulators of stamen development in Arabidopsis identified by transcriptional profiling.** *Plant J* 2006, **46(6):**984-1008.
29. **Genevestigator** [https://www.genevestigator.ethz.ch/at/]
30. Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W: **GENEVESTIGATOR. Arabidopsis Microarray Database and Analysis Toolbox.** *Plant Physiol* 2004, **136(1):**2621-2632.
31. Toufighi K, Brady SM, Austin R, Ly E, Provart NJ: **The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses.** *Plant J* 2005, **43(1):**153-163.
32. Kapushesky M, Kemmeren P, Culhane AC, Durinck S, Ihmels J, Korner C, Kull M, Torrente A, Sarkans U, Vilo J, Brazma A: **Expression Profiler: next generation--an online platform for analysis of microarray data.** *Nucleic Acids Res* 2004, **32(Web Server issue):**W465-70.
33. **Expression Profiler** [http://ep.ebi.ac.uk/EP/]
34. Bulow L, Schindler M, Hehl R: **PathoPlant: a platform for microarray expression data to analyze co-regulated genes involved in plant defense responses.** *Nucleic Acids Res* 2007, **35(Database issue):**D841-5.
35. **PathoPlant** [http://www.pathoplant.de]
36. Manfield IW, Jen CH, Pinney JW, Michalopoulos I, Bradford JR, Gilmartin PM, Westhead DR: **Arabidopsis Co-expression Tool (ACT): web server tools for microarray-based gene expression analysis.** *Nucleic Acids Res* 2006, **34(Web Server issue):**W504-9.
37. Jen CH, Manfield IW, Michalopoulos I, Pinney JW, Willats WG, Gilmartin PM, Westhead DR: **The Arabidopsis co-expression tool (ACT): a WWW-based tool and database for microarray-based gene expression analysis.** *Plant J* 2006, **46(2):**336-348.
38. **Arabidopsis Co-Expression Tool** [http://www.arabidopsis.leeds.ac.uk/act/]
39. **NASC AffyWatch transcriptomics service** [http://affymetrix.arabidopsis.info/AffyWatch.html]
40. **JpGraph** [http://www.aditus.nu/jpgraph/]
41. **dChip Software** [http://biosun1.harvard.edu/complab/dchip/]
42. Li C, Wong WH: **Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection.** *Proc Natl Acad Sci U S A* 2001, **98(1):**31-36.
43. Li C, Hung Wong W: **Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application.** *Genome Biol* 2001, **2(8):**RESEARCH0032.
44. Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Gorlach J: **Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants.** *Plant Cell* 2001, **13(7):**1499-1510.
45. Wortman JR, Haas BJ, Hannick LI, Smith RK Jr., Maiti R, Ronning CM, Chan AP, Yu C, Ayele M, Whitelaw CA, White OR, Town CD: **Annotation of the Arabidopsis genome.** *Plant Physiol* 2003, **132(2):**461-468.
46. **TAIR Genome Annotation** [http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp]
47. **TIGR** [http://www.tigr.org/]
48. Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP: **The TIGR Plant Transcript Assemblies database.** *Nucleic Acids Res* 2007, **35(Database issue):**D846-51.
49. **MPSS** [http://mpss.udel.edu/at/]
50. Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P: **The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community.** *Nucleic Acids Res* 2003, **31(1):**224-228.
51. **TAIR** [http://www.arabidopsis.org/]
52. Schoof H, Ernst R, Nazarov V, Pfeifer L, Mewes HW, Mayer KF: **MIPS Arabidopsis thaliana Database (MAtDB): an integrated bio-**

logical knowledge resource for plant genomics. *Nucleic Acids Res* 2004, **32(Database issue):**D373-6.
53. **MIPS** [http://mips.gsf.de/proj/thal/db/]
54. **Arabidopsis Gene Family Information** [http://www.arabidopsis.org/browse/genefamily/index.jsp]
55. **AGRIS** [http://arabidopsis.med.ohio-state.edu/]
56. Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri RV, Grotewold E: **AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks.** *Plant Physiol* 2006, **140(3):**818-829.
57. Liu WM, Mei R, Di X, Ryder TB, Hubbell E, Dee S, Webster TA, Harrington CA, Ho MH, Baid J, Smeekens SP: **Analysis of high density expression microarrays with signed-rank call algorithms.** *Bioinformatics* 2002, **18(12):**1593-1599.
58. Twell D Oh S.-A., Honys, D.: **Pollen development, a genetic and transcriptomic view.** In *Plant Cell Monographs vol 3, The Pollen Tube* Edited by: Malhó R. Springer-Verlag Berlin, Heidelberg:pp 15-45.
59. Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU: **A gene expression map of Arabidopsis thaliana development.** *Nat Genet* 2005, **37(5):**501-506.

8.2. Online tools for presentation and analysis of plant microarray data

Honys David, **Dupľáková Nikoleta**, Reňák David

Moretti M.K., Rizzo L.J. (eds). Oligonucleotide Array Sequence Analysis, pp. 265-295, Nova Science Publishers, 2008

This book chapter represents an overview of available servers, which serve to store large transcriptomic datasets and tools for their analyses. This chapter was intended to help potential (in)experienced users with the orientation in increasingly complex network of microarray data and innovative approaches used in individual databases and toolboxes.

My contribution: I was involved in defining the overall concept of the paper and I was fully responsive for the detailed description of several databases and data repository servers.

Supporting online material on CD

In: Oligonucleotide Array Sequence Analysis ISBN: 978-1-60456-542-3
Editors: M.K. Moretti and L.J. Rizzo, pp. 265-295

Chapter 9

ONLINE TOOLS FOR PRESENTATION AND ANALYSIS OF PLANT MICROARRAY DATA

David Honys[1,2,*], Nikoleta Dupľáková[1,2] and David Reňák[1,2,3]
[1]Laboratory of Pollen Biology, Institute of Experimental Botany ASCR, v.v.i., Rozvojová 263, 165 02 Prague 6, Czech Republic
[2]Department of Plant Physiology, Faculty of Science, Charles University, Viničná 5, 128 44, Prague 2, Czech Republic
[3]Department of Plant Physiology and Anatomy, Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic

ABSTRACT

The complete sequencing and annotation of the *Arabidopsis thaliana* genome represented a major step in biological research. This knowledge enabled gene prediction, assignment of functional categories and gave an opportunity to characterize global gene expression patterns at the transcriptome level at different developmental stages and under various physiological and stress conditions. For the discovery of partial or whole regulatory or functional networks, the development of high-throughput technologies was inevitable with genome-wide transcriptomic studies providing an essential input. DNA microarray technologies are among the most frequently used methods for parallel global analysis of gene expression. Microarray technologies now belong to the standard functional genomics toolbox and have undergone massive development leading to increased genome coverage, accuracy and reliability. Whole genome microarrays developed by Affymetrix in collaboration with Syngenta represented the first standard in genome-wide transcriptomic studies in plants. Whole genome Affymetrix ATH1 GeneChips covers about 76% of the *Arabidopsis thaliana* genome. Moreover, the introduction of the Minimum Information About Microarray experiments (MIAME) has increased the value and reproducibility of microarray experiments and has become a standard in documentation of array experiments and in the creation of databases of comparable transcriptomic experiments. The number of experiments exploiting microarray technology has markedly increased in recent years. Not surprisingly, there are potential difficulties in the orientation in available data sets. Microarray expression data are deposited on servers, many of which are publicly accessible. Currently, these databases store several thousands of individual datasets and some of these offer online tools for data normalization, filtering,

* Author for correspondence

statistical testing and pattern discovery. In parallel with the rapid accumulation of transcriptomic data, on-line analysis tools are being introduced to simplify their use. Global statistical data analysis methods contribute to the development of overall concepts about gene expression patterns and to query and compose working hypotheses. More recently, these applications are being supplemented with more specialized products offering visualization and specific data mining tools. In this chapter, an overview of available on-line databases and web-based applications using microarray data is presented together with the information about their structure and elementary principles.

INTRODUCTION

Microarray technologies now belong to the standard functional genomics toolbox and have undergone massive development leading to increased genome coverage, accuracy and reliability. The number of experiments exploiting microarray technology has markedly increased in recent years. In parallel with the rapid accumulation of transcriptomic data, on-line analysis tools are being introduced to simplify their use. Global statistical data analysis methods contribute to the development of overall concepts about gene expression patterns and to query and compose working hypotheses. More recently, these applications are being supplemented with more specialized products offering visualization and specific data mining tools.

The knowledge of whole sequenced and annotated plant genomes that started with Arabidopsis thaliana [AGI 2000] enabled gene prediction, assignment of functional categories and gave an opportunity to study gene and chromosome organization including the distribution of transposable elements. Finally it has enabled the characterization of global gene expression patterns at the transcriptome level at different developmental stages and under various physiological and stress conditions. Efforts to reveal the biological functions of thousand of genes and their integration into proteome, metabolome and interactome networks has become the principal focus of many studies and represents a key objective of the 2010 Project (http://www.nsf.gov/pubs/2006/nsf06612/nsf06612.htm).

A number of efficient and accurate gene expression analysis technologies to determine the expression levels of individual genes have been widely exploited in recent decades (Northern blot analysis, quantitative reverse transcription-PCR, cDNA library screening). Most of these methods enable analysis of the expression of single or relatively few selected genes. For the discovery of partial or whole gene functional or regulatory networks, the development of high-throughput technologies is essential with genome-wide transcriptomic studies providing a major input [Donson et al. 2002]. Several such methods have been developed including, cDNA fingerprinting [Money et al. 1996], serial analysis of gene expression - SAGE [Velculescu et al. 1995], massively parallel signature sequencing - MPSS [Brenner et al. 2000], high-density DNA oligonucleotide probe microarrays [Lockhart et al. 1996, Lipshutz et al. 1999] or cDNA arrays [Schena et al. 1995]. DNA microarray technologies are among the most frequently used methods for parallel global analysis of gene expression. These methods are based on the principle of selective and differential hybridization between sample target molecules and immobilized DNA probes. Hybridisation to probes arrayed on a solid surface report the relative abundance of DNA or RNA target molecules by fluorescent signal detection [van Hall et al. 2000, Clarke and Zhu 2006]. Microarray technologies now belong to the standard functional genomics toolbox [Zhu 2003,

Aharoni and Vorst 2002] and have undergone massive development leading to increased genome coverage, accuracy and reliability. Whole Genome microarrays developed by Affymetrix (Santa Clara, CA, USA) in collaboration with Syngenta represented the first standard in genome wide transcriptomic studies in plants. Whole genome Affymetrix ATH1 GeneChips cover about 76% of the *Arabidopsis thaliana* genes [Hennig et al. 2003]. Moreover, the introduction of the Minimum Information About Microarray experiments (MIAME) as standard documentation for array experiments and in transcriptomic databases, increasing the value and comparability of microarray data [Brazma et al. 2001].

As microarray technology has become standard technique in large scale gene expression studies, the number of transcriptomic experiments has markedly increased in recent years. Not surprisingly, there are potential difficulties in navigating between different available data sets. Microarray expression data are deposited on servers, many of which are publicly accessible. Public plant microarray data are deposited in several databases including ArrayExpress, GEO, NASCArrays and the Stanford Microarray Database. Currently these databases store several thousands of individual datasets and some of these offer on-line tools for data normalization, filtering, statistical testing and pattern discovery [Parkinson et al. 2005, Barrett et al. 2005, Edgar et al. 2002, Gollub et al. 2003, Craigon et al. 2004]. In parallel with the rapid accumulation of transcriptomic data, on-line analysis tools are being introduced to simplify their use. Global statistical data analysis methods contribute to the identification of overall gene expression patterns and to query and compose new working hypotheses based on these findings [Clarke and Zhu 2006, Zhu 2003, Hughes et al. 2000, Mandaokar et al. 2006]. More recently, these applications are being supplemented with more specialized products offering visualization and specific data mining tools. Genevestigator, Botany Array Resource, Arabidopsis co-expression tool, and ArrayExpress Expression Profiler offer Web-based tools to analyse large microarray datasets.

In this chapter we describe and collate several available on-line portals devoted to storage and analyses of plant microarray data. Summary of individual portals is provided in Table 1. For each database, basic information is available like URL, annotation of Arabidopsis genome, microarray data resource, used platforms, data normalisation, number of microarray datasets covered and access options. It is also distinguished between data repositories and/or data analysis tools. Moreover, gene- or genome-centric database concepts are compared. Gene-centric approach means that the particular gene is in the centre of interest and relevant data can be acquired from individual experiments. On the contrary, genome-centric approach enables the identification of genes fulfilling given criteria. Most portals enable direct data download. However, they offer different formats that are listed in Table 1. In the following section we provide more detailed description of all portals. We always shortly introduce the overall database concept and general features with special attention paid to specific tools for data analysis.

Table 1. Overview of plant microarray databases

Portal	ACT	aGFP	ArrayExpress	BAR	GEO	Genevestigator V3	NASCArrays	SMD
Database type	Data repository and simple toolbox	Data visualisation toolbox	Data repository and toolbox	Data analysis and visualisation toolbox	Data repository	Data repository and toolbox	Data repository and simple toolbox	Data repository and toolbox
Concept	Genome-centric	Gene-centric	Gene-centric and genome-centric	Gene-centric and genome-centric	Gene-centric and genome-centric	Gene-centric and genome-centric	Gene-centric	Gene-centric and genome-centric
Arabidopsis genome annotation	TAIR7	TAIR7	Various	TAIR7	n/a	TAIR7	TIGR5	n/a
Microarray data resource	NASCArrays/GARNet	AtGenExpress, NASCArrays	EBI Array express or user-uploaded data	Bio-Array Resource Database, AtGenExpress, NASCArrays	User-uploaded data compatible with GEO structure	NASCArrays, FGCZ, GEO, ArrayExpress, AtGenExpress, TAIR	NASCArrays	Various including Affymetrix, Agilent, and spotted arrays
Platform(s)	Affymetrix ATH1 Whole Genome Array	Affymetrix ATH1 Whole Genome Array	Various	Affymetrix ATH1 Whole Genome Array	Various high-throughput gene expression data	Affymetrix AG and ATH1 Arrays	Affymetrix AG and ATH1 Arrays	Various
Data normalisation	MAS5.0	MAS4.0 and MAS5.0	Various	MAS5.0	Various	MAS5.0	MAS5.0	Various
No. of experiments	73 experiments, 544 arrays (Arabidopsis)	120 experiments and 350 arrays	3,051 experiments and over 50,000 arrays from over 200 organisms	BAR 150 arrays, NASCArrays incl. AtGenExpress 330/2954 (Arabidopsis)	4,134 platforms, 181,922 samples and 7,202 series from over 100 organisms	201 experiments and 3132 arrays (Arabidopsis)	330 experiments and 2954 arrays	over 70,000 arrays and 15,451 experiments from 53 organisms

Table 1. Continued

Portal	ACT	aGFP	ArrayExpress	BAR	GEO	Genevestigator V3	NASCArrays	SMD
Access	Free	Free	Free with optional password-protected account	Free	Free	Open (free), Classic (free for academic users) and Advanced (paid)	Free	Free with restricted access to non-public data
Data download	Yes, via NASCArrays	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Downloaded data format	Tab delimited, XLS files, Gnumeric, CSV files	TAB-delimited text	TXT or XLS format, MEGA-ML format	Depends on selected tool	Web-based interactive forms, XLS files, MINiML, SOFT text files	Depends on selected tool	CSV or TAB-delimited files	XLS and TXT files
URL	http://www.arabidopsis.leeds.ac.uk/act/index.php	http://aGFP.ueb.cas.cz	http://www.ebi.ac.uk/arrayexpress	http://bbc.botany.utoronto.ca	http://www.ncbi.nlm.nih.gov/geo/	https://www.genevestigator.ethz.ch/	http://affy.arabidopsis.info/narrays/experimentbrowse.pl	http://genome-www5.stanford.edu/
Key references	*Jen et al. 2006, Manfield et al. 2006*	*Duplakova et al. 2007*	*Torrente et al. 2005, Parkinson et al. 2006*	*Toufighi et al. 2005, Winter et al. 2007*	*Edgar et sl. 2002, Barrett et al. 2007*	*Zimmermann et al. 2004, Laule et al. 2006,*	*Craigon et al. 2004*	*Ball et al. 2005, Demeter et al. 2007*

PORTALS

Arabidopsis Co-expression Tool (ACT)

ACT is a web-based tool for gene expression analysis using large microarray data set from the Nottingham Arabidopsis Stock Centre. The database stores pre-calculated co-expression results for cca 21 800 genes based on data from over 400 arrays (Figure 1). ACT enables identification of gene co-expression patterns across user-selected single or multiple arrays. An additional Clique Finder tool provides quantitative method for the determination of correlation cut-offs leading to the generation of groups of genes that may share a common purpose [Manfield et al. 2006]. For further analysis of co-regulated genes, promoter element detection software is included for the identification of potential DNA upstream elements [Jen et al. 2006]. The bioinformatic analyses consist of three steps: (i) generation of a list of

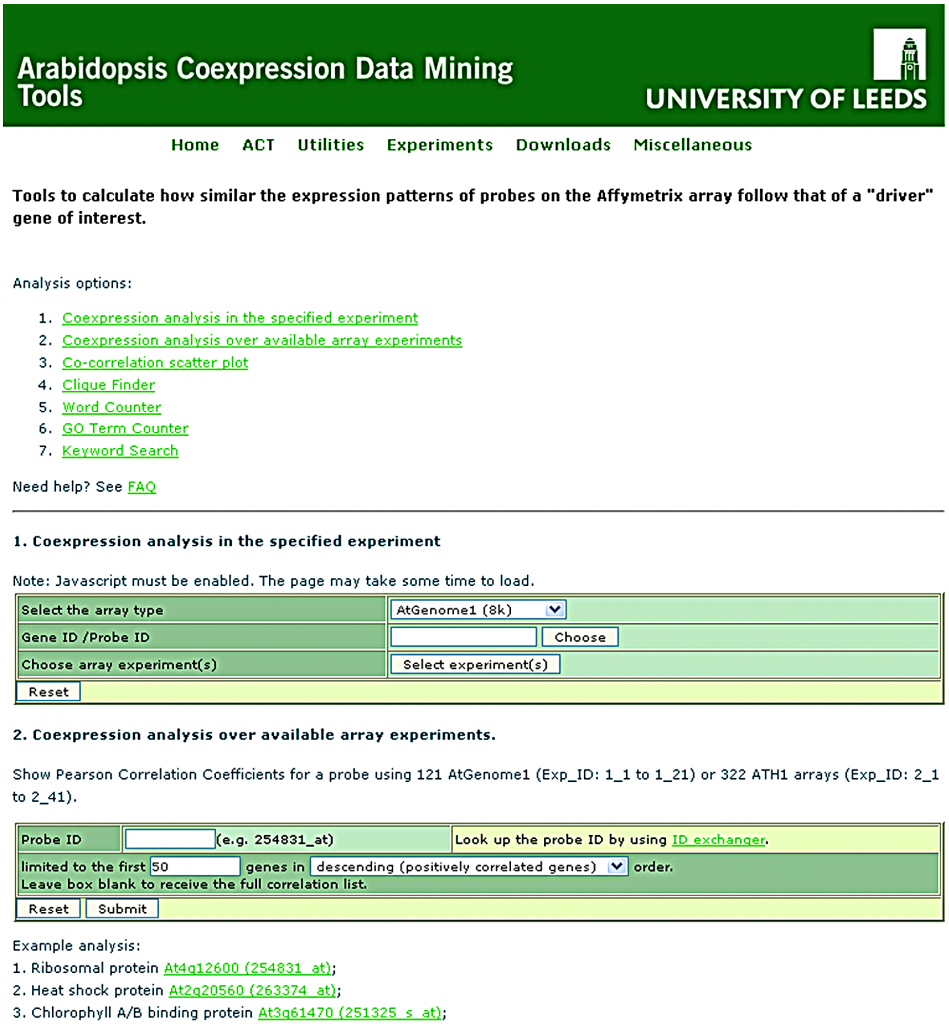


Figure 1. Homepage of the Arabidopsis Co-expression Tool available at http://www.arabidopsis.leeds.ac.uk/act/index.php.

negatively/positively co-expressed genes for a given gene of interest (driver); (ii) calculation of a co-correlation scatter plot of two selected genes; and (iii) identification of groups of genes that share statistical significant co-expression patterns. Co-correlation scatter plot analyses showing positively and negatively correlated genes may identify similar transcription regulatory mechanisms and can help to reveal the related biological functions of unannotated genes. Moreover, groups of genes with significant co-expression patterns often share the same biological theme and their members may overlap with other groups. It may then suggest their involvement in different biological processes. ACT provides a genome-centric approach allowing generation of a list of genes co-expressed within the driver gene throughout all experiments available. As an output, ACT offers a graphical view of gene expression patterns, list view of selected genes, co-correlation scatter plots; clusters of closely-associated probes and annotations of the most correlated genes to a given probe. Data can be downloaded as simple Tab-delimited TXT files, CSV, XLS files or in a Gnumeric format.

Specific Tools

Expression Pattern Displayer allows an exploration of gene expression patterns in a specified experiment or over all arrays available in the database.

Coexpression analysis was designed for the generation of lists of genes co-expressed with the probe (driver gene) according to their r-, p- and E-values.

Co-correlation scatter plot results in 2-D visualization of co-correlated data within two probes.

Clique Finder identifies genesets that are consistently co-expressed with each other throughout all microarray data.

Word count highlights annotations of the most correlated genes to a probe.

GO Terms offers Gene Ontology terms of the most correlated probes.

Sequence extractor extracts the specific region of cis-upstream or downstream sequences of single or multiple genes.

cis-Element Analyser analyses of cis-elements in the given sequence while **cis-Element_Locator** can be used for the localization of user-defined DNA motifs in promoters of single or multiple genes.

ID/Function Linker converts multiple Affymetrix probe IDs and AGI codes.

Distinguishable Features

Even though ACT uses the simplest measuring algorithm using the Pearson correlation coefficient, it is suggested to be an effective tool for the calculation of statistical significance [Manfield et al. 2006]. Researchers looking for positively and negatively correlated genes may (not) consider similar transcription regulatory mechanisms involved in related biological processes when revealing possible functions of unannotated genes. Moreover, the result of overlapping genes in each group may lead to a conclusion of being involved in different biological processes. In addition, the 2-D co-correlation plots can help in identifying sets of genes acting in particular metabolic, signaling or developmental processes. Finally, ACT tools for promoter analysis can be used to identify motifs which are often seen to underpin the observed expression correlation pattern. Other tools for identifying over-represented (but not

necessarily characterized) sequence motifs such as the Inclusive Motif Sampler are also useful for analyzing promoters of correlated genes.

Arabidopsis Gene Family Profiler (aGFP)

Currently, on-line analysis tools have been developed to simplify particular analyses of transcriptomic data from wide-ranging microarray analysis. Database Arabidopsis Gene Family Profiler (aGFP; http://aGFP.ueb.cas.cz) [Dupľáková et al. 2007] is a web-based tool offering visualization of microarray expression data of single genes, pre-defined gene families or even custom gene sets (Figure 2). aGFP gives users the possibility to analyze current *Arabidopsis* developmental transcriptomic data starting with simple global queries that can be expanded and further refined to visualize comparative and highly selective gene expression profiles. One of new features of aGFP is the possibility to display gene expression of pre-defined gene families and to easily adjust their composition by addition or reduction their members.

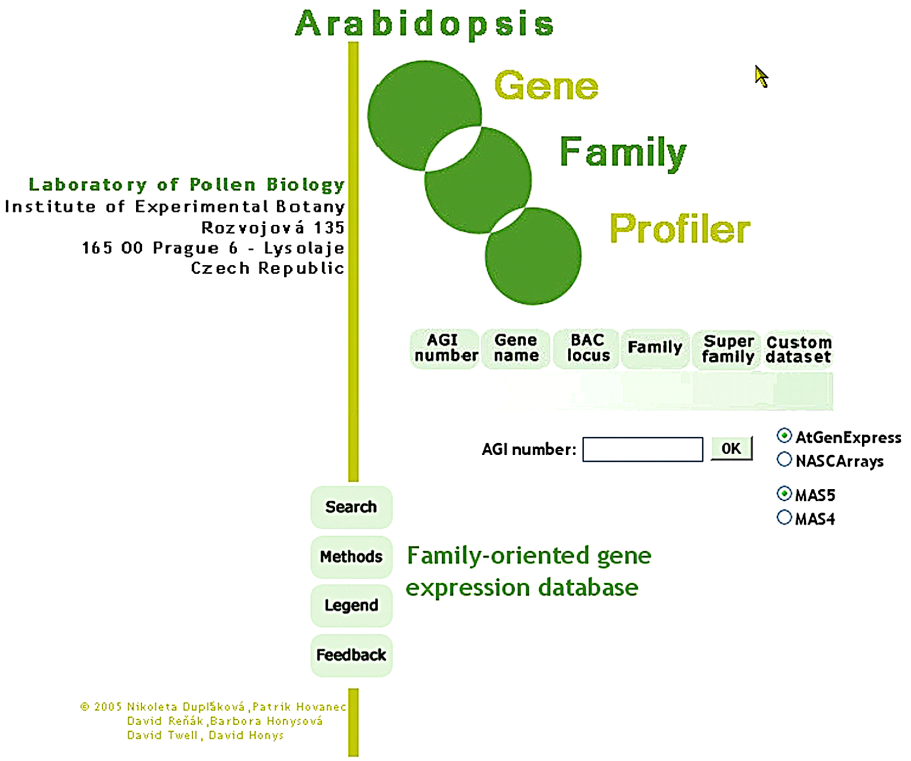


Figure 2. Homepage of the ArabidopsisGFP database available at http://aGFP.ueb.cas.cz.

aGFP database is based on a “gene-centric approach”. Authors adopted the general concept "from simple to complex". In the first approximation, an arithmetical mean expression signal from multiple experiments is displayed. In subsequent steps the user can choose to display expression data for individual plant organs or tissues at particular growth stages. This is accompanied by the option of progressive replacement of arithmetical means

by individual expression values. So users have the option to choose different levels of visualization to suit their needs.

aGFP offers several data search options together with two other input parameters; the gene detection algorithm (MAS4.0 or MAS5.0) and the source of expression data (AtGenExpress or NASCArrays). At any stage of the query the user has the possibility to directly switch options between these pairs of parameters.

aGFP uses two large freely available transcriptomic databases for various tissues at different developmental stages of wild type plants *Arabidopsis thaliana* grown at physiological conditions. The first subset contains data obtained within a scope of the AtGenExpress project [Schmid et al. 2005], the second comprises all other datasets deposited at NASC and was labeled NASCArrays [Craigon et al. 2004]. Data in each subset are presented using several different graphical displays and, the user has an option to instantly switch between subsets in each environment. Only experiments using Affymetrix ATH1 whole genome arrays with at least two biological replicates were included.

The complete datasets used in the aGFP database are described in the Legend available from the homepage and it is possible to trace the origin of all datasets. Gene expression data can be downloaded for individual and selected gene sets as a TAB-delimited TXT files. This enables the direct import of downloaded data into spreadsheet editors such as Excel and database software like Microsoft Access and FileMaker.

Moreover, aGFP allows a direct comparison of the influence of the detection algorithm or data resource on expression profiles. The database presents data normalised using two different algorithms, empirical Affymetrix MAS 4.0 and statistical MAS 5.0 with the possibility to switch between them at any point.

Genes in arabidopsisGFP are organized into two hierarchical levels consisting of gene families and superfamilies. All data were assembled from various relevant resources, the majority from TAIR – Arabidopsis Gene Family Information (http://www.arabidopsis.org/browse/genefamily/index.jsp) and AGRIS (http://arabidopsis.med.ohio-state.edu/) [Palaniswamy et al. 2006].

An interactive "virtual plant" represents the main tool for the visualization of the spatial and developmental gene expression profiles. Virtual plant uses a white-yellow-green scale to show relative expression signals of individual genes, or gene families throughout the *Arabidopsis* life cycle. In first step, the „virtual plant“ is displayed an mean expression signal from multiple experiments. In subsequent steps, the user can choose to display expression data for individual plant organs or tissues at particular growth stages. This is accompanied by the option of progressive replacement of arithmetical means by individual expression values. Finally, the user can switch from "virtual plant" visualization to a simple bar chart (standard or log-scaled) or tabulated display. The user can also browse through individual experiments down to normalized or even raw data extracted from individual gene chips. Gene family expression data can also be visualized as a colorized spot chart (heat map) with colour scale identical to that of the virtual plant. The interactivity of modeling is based on the possibility of ad hoc addition and removal of genes to and from currently active set. Moreover, each spot contains mouse-over-activated information about expression signal value and experiment. Annotation of individual experiments meets MIAME standard [Brazma et al. 2001].

Distinguishable Features

A novel feature of aGFP is that it enables the evaluation of the impact of normalization procedures on microarray gene expression data as well the possibility of rapid definition of user-defined gene families or other groups. Simultaneously, aGFP serves as a facile and synoptic developmental reference guide for expression profiles of individual genes or gene families in wild-type *Arabidopsis thaliana* plants. arabidopsisGFP contains lists of pre-defined gene families and superfamilies enabling the rapid comparative visualization of expression profiles of their members.

ArrayExpress

ArrayExpress (Figure 3) was developed as a public repository for gene expression data at the centre of a wider microarray informatics system at European Bioinformatics Institute (EBI) [Brazma et al. 2003]. It comprises well-annotated raw and processed microarray data originated from several platforms supporting MIAME requirements [Parkinson et al. 2005]. Currently, the database stores data of 1,500,000 gene expresion profiles from more than 50 000 hybridizations representing more than 200 organisms [Parkinson et al. 2007]. Microarray data are either part of EBI ArrayExpress or they can be user-uploaded from various resources including Affymetrix, Agilent, Illumina, Nimblegen, etc. [Brazma et al. 2003]. The majority of experimental data relates to gene expresion profiling studies, the remainder are array-based chromatin immunoprecipitation or comparative genomics studies [Parkinson et al. 2005].

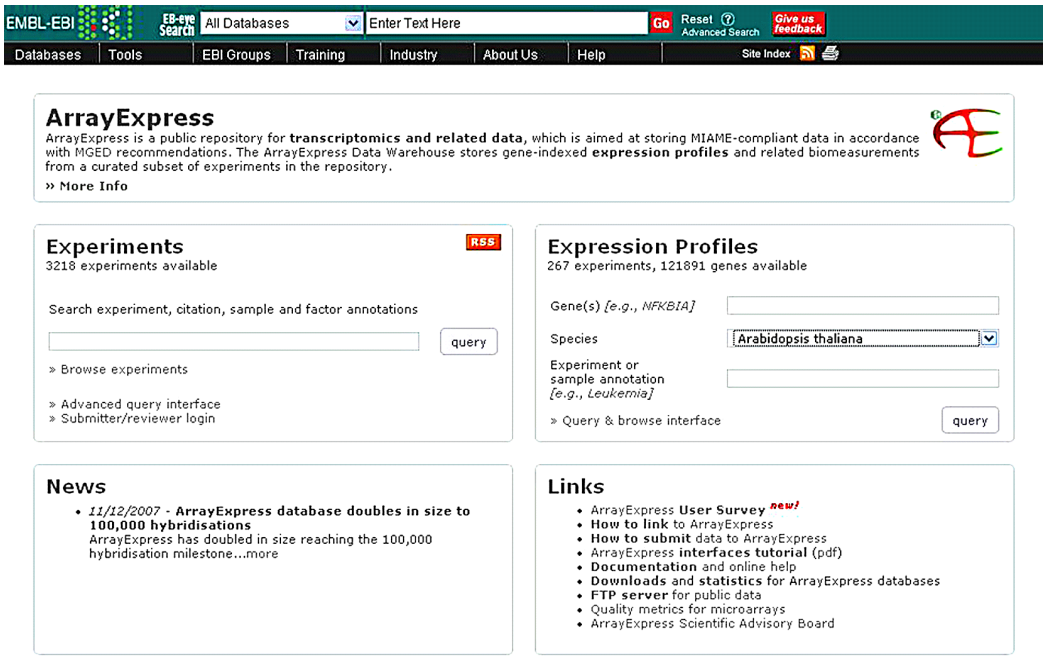


Figure 3. Homepage of the ArrayExpress database available at http://www.ebi.ac.uk/arrayexpress.

ArrayExpress was created to achieve three major goals, first, to serve as an archive for microarray data, second, to provide access to genes expression profiles, and third, to facilitate the sharing of microarry design and experimental protocols [Sarkans et al. 2005; Brazma et al. 2006].

ArrayExpress infrastructure consists of database itself and individual modules enabling user data submission, database query and data analyses [Brazma et al. 2003]. For data submission, three types of data are accepted, arrays, experiments and protocols. Altogether, they provide full information about the respective dataset but they can be uploaded separately under individual accession numbers. It simplifies the process of microarray data submission using standard platforms like Affymetrix. Data are submitted directly using three major submission routes: (i) online via the MIAMExpress data submittion tool [Brazma et al. 2003], (ii) via a spreadsheet suitable for all types and sizes based on technology type, species and experimental type, and (iii) via a MAGE-ML or MAGE-TAB (Microarray gene Expression Markup Language) [Spellman et al. 2002] pipelines set-up with external databases [Kapushensky et al. 2004]. As plant-based data represent about 25% of ArrayExpress datasets and vast majority of those based on Arabidopsis, specific open-source software application for the submission of Arabidopsis-based microarray data is available [Mukherjee et al. 2005]. MIAMExpress itself enables download of limited number of hybridisations with the requirement for filling out a web-form for each individual dataset. For this reason, BLoader software application was developed and implemented to generate the annotation information rapidly and submit the entire set in one batch [Schwager et al. 2005].

Since its introduction, ArrayExpress has undergone significant improvement as it has two major components. Original "ArrayExpress experiment repository" represents the main database comprising complete data while "ArrayExpress gene expression profile data warehouse" contains gene-indexed expression profiles from a curated subset of experiments from the repository [Brazma et al. 2006]. Both ArrayExpress and user-uploaded microarray data can be analysed and visualized exploiting integrated ExpressionProfiler toolbox that is described in detail below [Kapushesky et al. 2004].

Specific Tools

General tools:

Browse Experiments enables filters on species, array, experiment type, experimental factors, author, laboratory, publications, array design name and protocol type.

Browse Expression Profiles selects gene name or AGI, species, experiment or sample annotation.

Similarity Search is designed for arrangement of additional genes with similar expression profiles.

Expression Profiler — next generation:

Expression Profiler (EP) [Kapushesky et al. 2004] represents microarray data analysis and visualization toolbox implemented in the ArrayExpress database [Brazma et al. 2003]. EP was originally designed as a web-based platform for microarray data and other functional genomics-related data analysis [Kapushesky et al. 2004]. After the implementation of new architecture enabling the modularization of the original design and de-centralization of database development, the toolbox was renamed to Expression profiler: next generation

(EP:NG). Individual tools are described below. EP:NG can be accessed from ArrayExpress page or directly at http://www.ebi.ac.uk/expressionprofiler.

Individual EP:NG tools:

User-defined **Data Upload** of microarray data following specific data-organization and data description matrices including optional various metadata.

Data Selection provides brief basic statistical overview of microarray data and user-defined data selection and sub-selection.

Data Transformation represents optional procedures preceding or following data subselection. They include K-nearest neighbour imputation [Troyanskaya et al. 2001] to fill in missing values, LOWESS normalization [Yang et al. 2002], data conversion in two-channel experiments.

Similarity Search extracts and visualizes of a group of co-expressed genes in relation to given one(s).

Clustering is application of hierarchical and partitioning-based K-groups clustering algorithms [Torrente et al. 2005].

Clustering Comparison is a tool for the comparison and matching of two independent K-groups clustering results.

Signature Algorithm identifies co-expressed subset of genes in user-submitted geneset.

Ordination offers several tools for multivariate analysis methods; principal component analysis and correspondence analysis.

Between Group Analysis allows multiple discriminant approach used with expression data matrices.

Distinguishable Features

ArrayExpress database provides a free public repository of microarray gene expression data of various origins. In its modular architecture it links together gene expression data and number of data processing and analysis tools. Moreover, data submission in three separate steps resembling GEO enables future comparison and analyses of microarray data of different origin. Extra value was added by the subselection of specific gene expression data into growing curated ArrayExpress data warehouse. Data analysis and visualization tools are implemented in the integrated ExpressionProfiler package. This, together with unified data submission format, enables direct link to downloaded gene expression data, their efficient processing and analyses. Among most distinguishable features of EP:NG belong its gene clustering and statistical tools.

Bio-Array Resource (BAR)

The Bio-Array Resource (BAR) is an extensive microarray data repository accompanied by number of very helpful tools designed for data visualization and detailed analyses (Figure 4) [Toufighi et al. 2005]. BAR is a completely web-based database, exploiting MySQL with interfaces implemented in Perl and C. The site was originally named Botany Array Resource as it was designed for Arabidopsis-based data. Recently, a toolbox for the visualisation of gene expression in mouse tissues have been recently added [Winter et al.

2007] so further development and extension of the coverage is likely to be expected. Currently, BAR database comprises microarray data collected from University of Toronto microarray facility as well as data loaded from NASCArrays and AtGenExpress Consortium. The database has been designed as MIAME-compliant [Brazma et al. 2001] and all provided data and tools are publicly available through a web interface. BAR offers large number of various data analysis tools. They include not only gene expression visualization and analysis tools but also set of assorted on-line tools especially useful for genetic mapping and data formatting, visualization and clean up. An overview and brief characterization of individual tools is provided in the specific section.

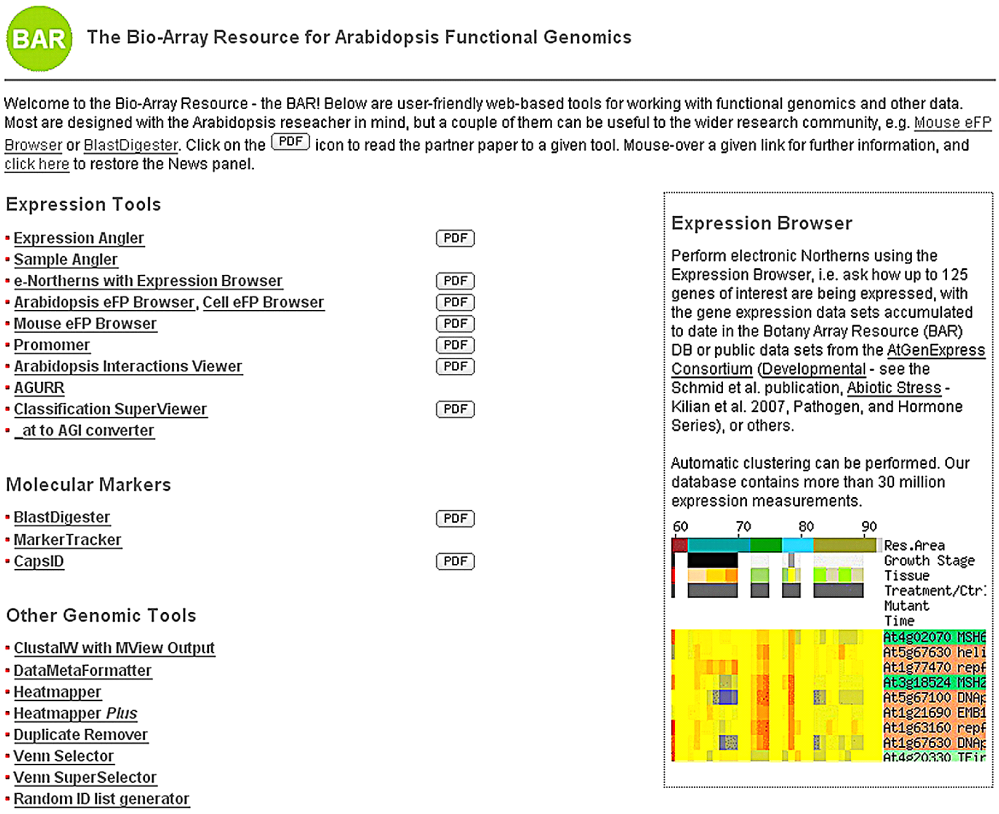


Figure 4. Homepage of the Bio-Array Resource database and toolbox available at http://bbc.botany.utoronto.ca.

Specific Tools

Gene expression tools:

Expression Angler identifies sets of co-regulated, anti-correlated, or condition/tissue-specific genes in various data sets.

Sample Angler identifies samples exhibiting similar expression profiles.

e-Northerns with Expression Browser is an electronic Northerns tool providing a graphical display of expression data of particular set of genes across all selected datasets.

Arabidopsis eFP Browser provides intuitive visualization of particular gene expression patterns.

Cell eFP Browser enables intuitive visualization of particular protein subcellular localization patterns based on documented and predicted subcellular localizations according to the SUBA database [Heazlewood et al. 2005, 2007].

Mouse eFP Browser offers intuitive visualization of particular gene expression patterns based on the Zhang et al. [2004] Functional Landscape of Mouse Gene Expression.

Promomer was designed for the identification of over-represented motifs in the promoter of a gene of interest, or in promoters of co-regulated genes.

Arabidopsis Interactions Viewer returns predicted or documented interaction partners with a gene of interest.

AGURR (Arabidopsis Genetic Uniqueness and Redundancy Revealer) is a program for the identification of samples in which one gene is uniquely expressed or in which several genes are potentially redundantly expressed.

Classification SuperViewer returns functional classification a of AGI IDs list based on the MIPS database.

at to AGI converter converts 22k Affymetrix GeneChip IDs into AGI IDs and vice versa.

Molecular Markers:

BlastDigester analyses a nucleotide BLAST output for restriction enzymes differentially cutting the aligned sequences [Ilic et al. 2005].

MarkerTracker is a repository for genetic markers.

CapsID serves as automated CAPS marker-based genotyping system [Taylor and Provart 2006].

Other general tools for basic genomic analyses:

ClustalW with MView Output is a web-based version of ClustalW.

DataMetaFormatter can be used for Arabidopsis gene expression data reformatting and adding pieces of meta-information (protein-protein interactions, functional classification).

Heatmapper and Heatmapper Plus applies a third dimension of information to a 2-D table via colour-coding.

Duplicate Remover removes duplicates from data lists.

Venn Selector shows identifiers in common and unique to two sets of sequences and fetching annotation.

Venn SuperSelector organizes multiple lists of genes or terms with associated values with an option of data export to Excel.

Random ID list generator is useful for the generation of *n* sets of *y* genes containing *z* number of randomly generated Arabidopsis AGI IDs.

Distinguishable Features

The BAR provides number of sophisticated and easy-to-use web-based tools for microarray data visualization and analyses. Moreover, these tools are accompanied by other "simple" but extremely helpful tools for data formatting, presentation and clean-up that certainly may be time-consuming. Microarray data are divided into several databases according to their origin and experiment type. In most tools, users can choose from Bio-Array

resource database, several AtGenExpress Consortium databases, NASC Arrays and sometimes even their own in-house databases.

Gene Expression Omnibus (GEO)

The Gene Expression Omnibus (GEO) was originally designed as a public data repository and retrieval system for high throughput gene expression data with an ambition to complement number of existing in house gene expression databases and to act as the central data distribution hub (Figure 5) [Edgar et al. 2002]. It represents one of large number of computational resources for the analysis of data stored in GenBank and other biological data provided by NCBI and available through NCBI's website http://www.ncbi.nlm.nih.gov [Wheeler et al. 2005, 2007]. Currently, the database has a MIAME-compliant infrastructure and contains both raw and processed data derived from over 100 organisms submitted by over

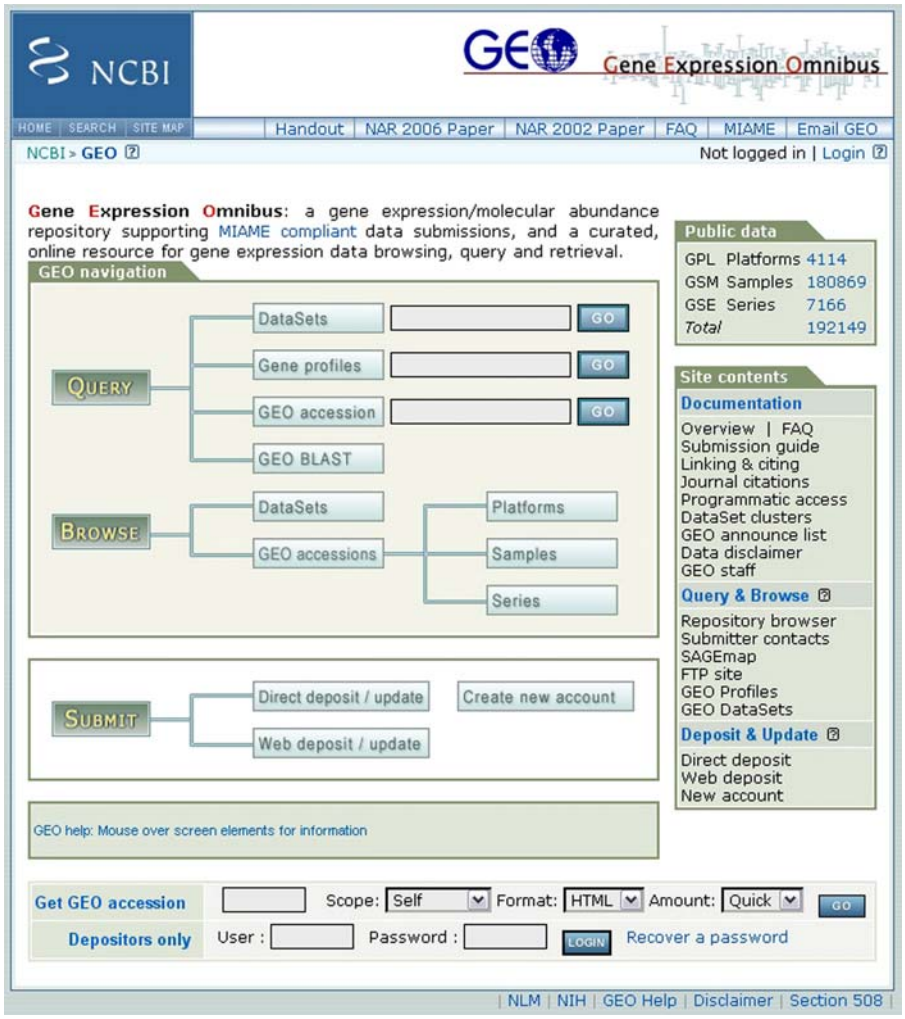


Figure 5. Homepage of the GEO database available at http://www.ncbi.nlm.nih.gov/geo/.

1,500 laboratories [Barrett and Edgar 2006a, Barrett et al. 2007]. GEO is not devoted to one technology but it contains microarray-based experiments measuring the abundance of mRNA, genomic DNA and protein molecules, as well as non-array-based technologies such as SAGE and mass spectrometry peptide profiling [Barrett et al. 2005].

The unifying feature of GEO database is that all data are segregated and stored as three principle components, platforms, samples and series. Platform provides a summary description of the array and data table by defining the array template. Sample describes the biological material, the experimental conditions under which the sample was handled and a respective data table containing hybridization signal levels for each feature on the corresponding previously defined platform. Finally, series defines a set of related samples that are considered as a part of an individual experiment [Barrett et al. 2007].

Specific Tools

There are two key options to query GEO using two NCBI Entrez databases [Barrett and Edgar 2006b]. **Entrez GEO DataSets** allows a genome-centric view on gene expression data in GEO. It returns a keyword-based identification of experiments of interest. **Entrez GEO Profiles** provides a gene-centric view on gene expression data in GEO. It returns expression profiles of genes of interest. Particular genes are identified using various search options. Both NCBI Entrez databases are supplemented with several additional tools. Alternatively, GEO database can be queried using GEO BLAST tool. It retrieves sequences and corresponding expression profiles similar to user-defined sequence.

Entrez GEO DataSets supplementary tools:

DataSet clusters is a hierarchical and K-means clustering tool for data visualization, selection and download.

Query group A versus B identifies genes with expression profiles meeting user-defined statistical criteria.

Subset effect flags was designed for the identification of potentially interesting genes by showing those with significantly different expression levels among experimental variables.

Entrez GEO Profiles supplementary tools:

Profile neighbors link allows connection and display of genes sharing similar expression profiles.

Sequence neighbors link enables connection and display of groups of genes related by sequence similarity.

Homolog neighbors link connects and displays groups of genes related by Homologene groups.

Links link links GEO data to related data in other NCBI resources (ie. PubMed, GenBank, Gene, UniGene, OMIM).

Distinguishable Features

GEO is a fast-growing data repository representing large compendium of various gene expression data. Its main impact lies in the complexity of microarray data and in their integration with other resources like sequence information, mapping and bibliographic data.

Moreover, GEO offers the possibility to download gene expression data to enable researchers the use of their own software. One example of such software package can be downloaded as a part of R/Bio/Conductor at http://bioconductor.org/packages/release/bioc/html/GEOquery.html. On the contrary, the diversity of area covered makes the development of robust and user-friendly statistical and data analysis tools difficult. Authors also plan further development of GEO and its extension in order to include non-gene-expression data types like chromatin-immunoprecipitation on arrays (ChIP-chip) studies, array comparative genomic hybridization (aCGH), SNP arrays and even proteomic data [Barrett et al. 2007].

Genevestigator V3 (GV3)

In 2007, new version of Genevestigator (V3 beta) was released (Figure 6). It serves as a bioinformatic service for microarray gene expression data from five organisms (*Arabidopsis thaliana, Mus musculus, Rattus norvegicus, Hordeum vulgare, Homo sapiens*) with total number 16,319 microarrays [Zimmermann et al. 2004, 2005, Laule et al. 2006]. For Arabidopsis, Genevestigator V3 covers 3222 experiments. Initial Genevestigator database was launched in 2004 as a web-based MySQL/PHP software application to query Arabidopsis microarray database with several analysis tools – Digital Northern, Gene Correlator, The, Gene Atlas, Gene Chronologer, Response Viewer and Meta-Analyser. Currently Genevestigator V3 operates as Java applet at a client's side. This client-server structure offers more sophisticated functions and allows more comfortable interface. Obviously, it is necessary to have Java runtime environment (JRE) version 1.4.2 or higher installed. The database is supplemented with several tools - Meta-Profile Analysis, Biomarker Search, Clustering Analysis, Pathway Projector. Genevestigator belongs among the most professional microarray analysis applications and provides many data mining tools allowing multilevel analyses. However, it does not serve as a data repository thus the user-controlled direct data upload is not possible. Currently, it hosts only expression data from Affymetrix GeneChips; for Arabidopsis, AG and ATH1 arrays are covered.

Transcriptomic data and their annotations were retrieved from many sources and repositories. Of these, NASC (http://affymetrix.arabidopsis.info/) [Craigon et al. 2004], FGCZ (http://www.fgcz.ethz.ch/), GEO (http://www.ncbi.nlm.nih.gov/geo/), [Edgar et al. 2002, 2006, Barrett et al. 2005, 2007, Barrett and Edgar 2006a], ArrayExpress (http://www.ebi.ac.uk/microarray-as/aer/?#ae-main[0]) [Parkinson et al. 2005, 2007, Brazma et al. 2003, 2006, Sarkans et al. 2005, Rocca-Serra et al. 2003], AtGenExpress [Schmid et al. 2005] and TAIR (http://www.arabidopsis.org/) [Rhee et al. 2003] belong among the public repositories. Genevestigator enables two main query approaches – a gene-centric and genome-centric. It also offers an extensive selection of data-searching options that are well described in the accompanying documentation (https://www.genevestigator.ethz.ch/index.php?option=com_content&task=view&id=34&Itemid=95).

Genevestigator V3 is generally publicly available. However, it offers three levels of access – Open, Classic and Advanced. Open access does not require the registration of user account but it offers only two tools (Northern and Selection tool). Classic access enables also the Meta-Profile Analysis toolbox and is free for academic institutions. Advanced access offers all tools available in GV3 (also Biomarker search, Clustering analysis and Pathway projector) but this access option is paid.

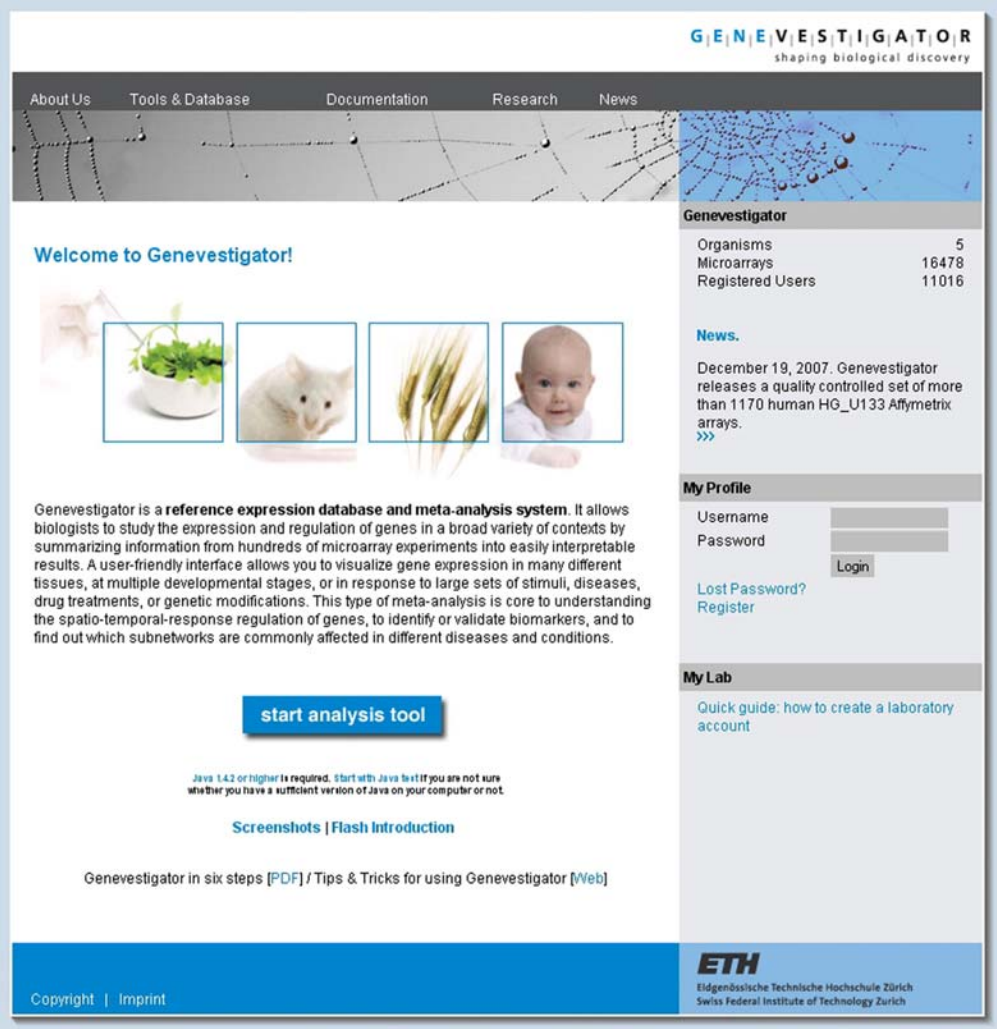


Figure 6. Homepage of the Genevestigator V3 database and toolbox available at https://www.genevestigator.ethz.ch/.

Data Mining and Analysis Tools

Meta-Profile Analysis comprise six tools - Selection, Northern, Anatomy, Development, Stimulus and Mutation that allow the discovery of gene expression pattern across various arrays (Selection, Northern) or categories (Anatomy, Development, Stimulus and Station). All Meta-profile Analysis tools are based on „gene-centric approach“ so they return signal intensity values of pre-defined genes and genesets. Further, these tools can be divided into two categories according to conditions and gene-array approach used. The gene-condition approach groups all arrays with same conditions for instance identical developmental or anatomical category („meta-profiles“).

Selection tool displays how selected gene or genes are expressed across all arrays in the database. The user can choose arrays and genes for display. Moreover, each experiment can

be selected and highlighted in context of all experiments. Results are visualised as a scatter plot with magnifying rectangle.

Northern displays how selected gene or genes are expressed across chosen arrays in the database. Selection and Northern tools may depict expression profiles of many genes through many arrays simultaneously. Results are shown as an editable scatter plot with possibility to switch the axes (horizontal or vertical).

Anatomy depicts how gene(s) of interest are expressed in different plant tissues, organs and cell cultures organised in a tree of anatomical categories. In each category, the average signal intensity value (mean value) from all selected arrays belonging to this category is plotted. Expression values from children nodes are always involved in the mean expression value in a parent node. Results are visualised as a scatter plot or heat map. Users shall be aware that the scatter plot displays only up to ten genes.

The **Development** tool is very similar to the Anatomy tool (identical output formats and expression data types) but it shows how strongly the selected gene(s) are expressed in different developmental stages.

The **Stimulus** tool visualises gene expression responses to external stimuli and reveals up- or down-regulated genes. The same stimuli from different experiments are processed separately. The user then can see one stimulus in one plot repeatedly. Results are shown as scatter plot or heath map with the indication of number of arrays under treatment (T) and control (C) conditions used for the calculation of ratio and log ratio. Again, the scatter plot displays up to ten genes.

Mutation displays the gene expression response to the mutations or other genetic modifications in the genome as a ratio or log ratio value. The Mutation tool features are identical as in the Stimulus tool. Both tools also have identical layout, output and options.

Biomarker Search helps user to identify or validate biomarker genes in given category. It consists of the same toolbox as Meta-Profile Analysis (Anatomy, Development, Stimulus, Mutation). Specific biomarker identification is based on target-base search. The score defined to identify specifically expressed genes is derived from the sum of averages from the selected target and base categories. It is possible to display up to 400 genes per search. All available tools (**Anatomy, Development Stimulus and Mutation**) return the list of probe sets with the highest scoring genes in given target group over against a base. Whereas Anatomy and Development work with absolute signals, Stimulus and Mutation use relative expression values. The Anatomy tool is visualized as the expandable tree menu of individual categories. The Development tool is displayed as a pictorial list of development categories. The Mutation tool simply shows a list of mutation categories. Results are shown as a heat map with blue-white (absolute expression values) or green-red color coding (relative expression values)

Clustering Analysis is used to identify local patterns in gene expression data in given categories (Arrays, Anatomy, Development, Stimulus, Mutation) and to discover the possible relationship among gene expression patterns. Huge advantage of GV3 clustering tool is that it enables clustering of genes within their biological context using meta-profiles in the analysis (Anatomy, Development, Stimulus, Mutation). After selection of particular experiments and genes the user chooses the clustering method. There are two different algorithms available – Hierarchical Clustering and Biclustering. Whereas the first results in a two-dimensional matrix consisting of genes and conditions by selected gene cluster and profile cluster, the other uses exact BiMax algorithm that identifies all biclusters in a matrix.

Pathway Projector represents a reaction network analysis system. It combines expression data from transcriptomic studies and literature knowledge about signaling and metabolic pathways. Currently, the Pathway Projector analysis tool is available only for two organisms – *Arabidopsis thaliana* and *Mus musculus*. In few subsequent steps, the user defines comparison set (array set) that contains treatment and control sets and then selects appropriate subnetworks. The pre-defined or user-defined datasets for four ontologies (anatomy, development, stimulus, mutation) represent great advantage of this tool. In addition to display a network of interest, Pathway Projector also enables the user to build a new network from predefined pathways. Moreover, it is possible to import and export pathway maps as a GVP files.

Distinguishable Features

The possibility to download graphical output in format and resolution ready for publication or web presentation represents very useful feature of newly released Genevestigator V3. The user can select between EPS, PNG or JPEG formats and set the figure parameters for export (width, height). Manual curation of microarray experiments represents great advantages of GV3. It results in Quality control (QC) that allows the detection of arrays with lower or poor quality and subsequently elimination of these arrays or experiment excluded from the microarray database. Identification of problematic arrays increases the output reliability of given data analysis. Genevestigator V3 categorizes experimental data that passed the QC check into separated subset called "High quality arrays only".

NASCArrays

NASCArrays [Craigon et al. 2004] is the Nottingham Arabidopsis Stock Centre's microarray database (http://affy.arabidopsis.info/narrays/experimentbrowse.pl) and it serves primarily as a repository for microarray data produced by NASC's transcriptomics service (Fig. 7). NASCArrays database harbours data from two types of Affymetrix GeneChips, AG Arrays and ATH1 Whole GenomeArrays. Microarray data are shared with other databases worldwide. In this database, individual arrays are referred to as "slides". Index page of NASCArrays allows obtaining expression data by Search tools, Treeview and Data mining tools. The easiest way of data selection is the keyword search in Experiment Search, NASCArrays Reference Number and Slide search with using many selection criteria e.g. growth conditions, development stage, etc. Alternatively, individual experiments can be selected in a Treeview (list of experiments) containing all experiments categorized according to various criteria [Whetzel et al. 2006]. Moreover, the Slide selection tool allows researcher to perform desired data analysis by choosing individual slides. Immediately after selection of appropriate slides, set of data mining tools can be applied with the possibility of data download.

The experiment page shows details for each experiment (abstract, contact details of the author, information about the experiment and list of all slides belonging to the respective experiment with precise experimental information.

For download, one or several slides can be selected using "Slide selection" function for the whole experiment or for up to 300 genes selected by bulk gene download. Data are supplied in one of two common data formats, CSV files or TAB-delimited text files. All data are well annotated with sample preparation details. Moreover, it is possible to download several data categories - Signal, StatPairsUsed, PresentCall, Detection P-value. For clustering purposes, only signal values may be downloaded.

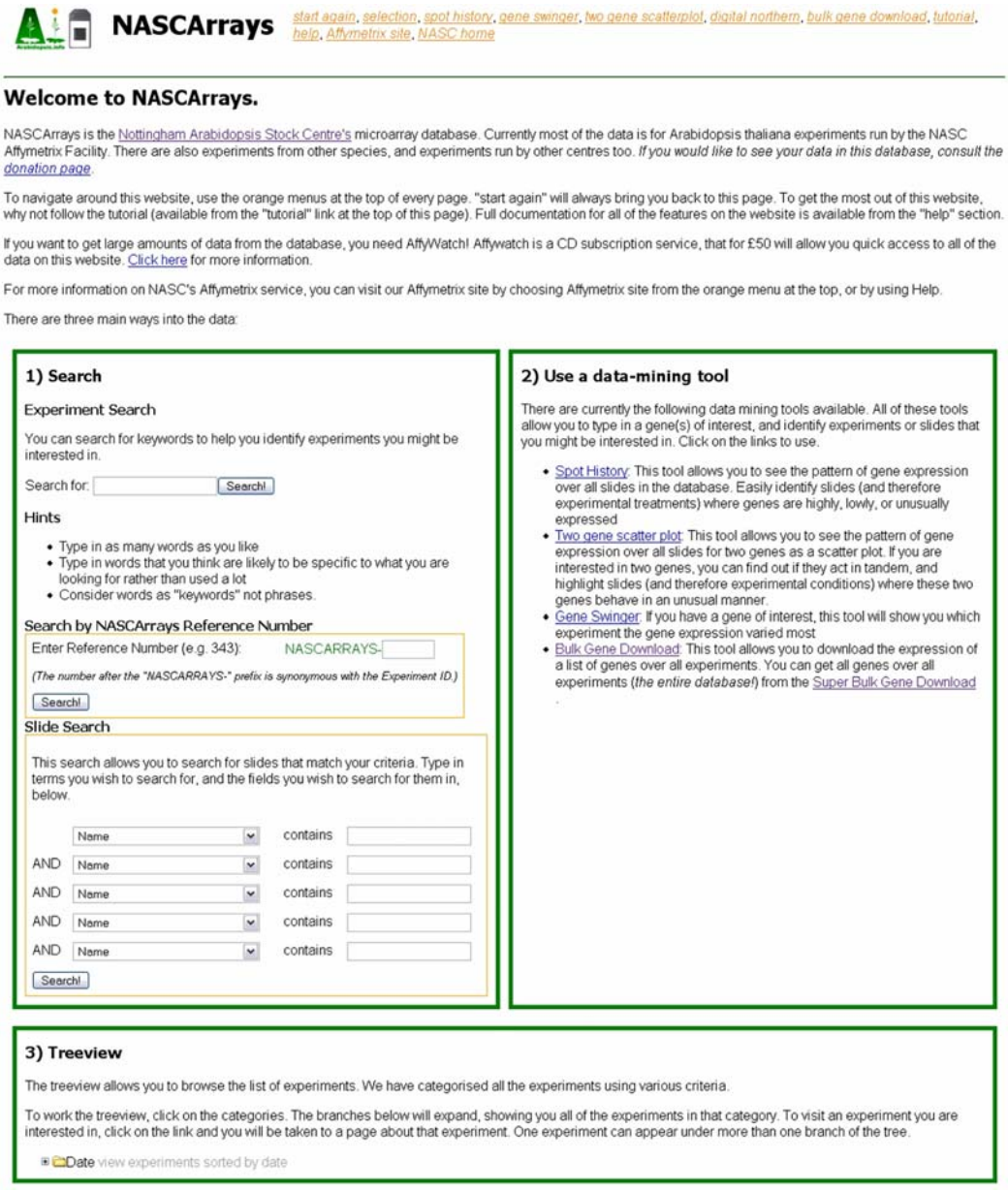


Figure 7. Homepage of the NASCArrays database available at http://affy.arabidopsis.info/narrays/experimentbrowse.pl.

Arabidopsis genome annotation supplied from Arabidopsis Ensembl (http://atensembl.arabidopsis.info/index.html) is based on TIGR3 version [Childs et al. 2007], (http://www.tigr.org/). For this reason, data must be interpreted carefully. However, user-supplied annotation for each sample is MIAME-compliant [Brazma et al. 2001] including information on how the original RNA sample preparation and treatment. The information about the each experiment and all used microarrays is very comprehensive. It can be accessed on the experiment page where each slide is represented by a slide box. Moreover, specially formatted data for clustering analysis using EBI-EPCLUST software [Kapushesky et al. 2004; http://ep.ebi.ac.uk/EP/EPCLUST/) can be acquired and all microarray data produced by NASC's Affymetrix service can be also mailed to users on CDs (AffyWatch, paid service).

Specific Tools

Spot history displays the distribution of expression of a gene of interest over all experiments in the database and gives the user the chance to find experiments or individual slides with distant expression values for given gene and therefore to identify "unusual experiments".

Two-gene scatter plot visualises expression profiles of two genes and reveals the relationships between these genes (co-expression) over all slides in a database.

Gene swinger is a web-based tool for experiment ranking according to the Coefficient of Variance value. It identifies experiments where the gene of interest shows the highest variability.

Digital Northern is an elementary tool for the visualization of relative expression signal and detection call (present, absent) of up to ten genes over all slides or over selected experiment or slides.

Simple Pairwise Analysis is a simple mining tool for identification of upregulated genes. After selection of just two slides it shows top 200 upregulated genes at each slide.

Bulk gene download enables download of up to 300 genes. After submitting list of AGI codes, related data from all the experiments in the database are downloaded.

Distinguishable Features

Extensive user-supplied annotation for each sample meets MIAME requirements [Brazma et al. 2001]. Data describing each experiment are very comprehensive including information on how the original RNA sample was prepared and how the RNA sample was subsequently hybridised. This information can be accessed on the experiment page. Among advantageous features belongs the availability of expression data specially formatted for clustering analysis by EBI-EPCLUST software [Kapushesky et al. 2004; http://ep.ebi.ac.uk/EP/EPCLUST/) and the availability of paid AffyWatch mailing service.

The Stanford Microarray Databases (SMD)

Originally, the database was established in 1999 for Stanford investigators but it is no longer restricted to local Stanford users. Now SMD serves as a data repository and microarray research database that is publicly available for all researches worldwide (Figure 8) [Sherlock

et al. 2001, Gollub et al. 2003]. Currently, the database contains the data for 53 organisms that represents almost 70,000 experiments (http://smd.stanford.edu/statictics.html). SMD supports wide collection of microarray platforms and software packages such as Affymetrix, Agilent, Combimatrix, GenePix, ScanAlyze, etc. Most data are public; however access to non-public data is limited to registered Stanford researchers and their collaborators. In principle, users can upload the data to the repository, subsequently retrieve and analyze their data using variety of data mining, analysing and displaying tools, and then, after publishing, made them available for the scientific community [Ball et al. 2005].

Figure 8. Homepage of the Stanford Microarray Database available at http://genome-www5.stanford.edu/.

After data retrieval, SMD provides various data processing and analysing tools. First, all data undergoes normalization and centering enabling cross-array comparison. Second, clustering algorithms reveal specific patterns within data using hierarchical clustering, self organising maps, singular value decomposition to identify missing data, alternatively, KNNImpute to estimate missing values in data matrices. Third, quality assessment tools offer other options including ANOVA analysis to detect spatial and print-tip bias on the array, HEEBO/MEEBO plots to view diagnostic and doping control, and graphing tools to produce histograms or scatter plots of user-selected fields [Demeter et al. 2007].

When searching for particular data, users can access variety of data mining tools in two different search forms – "Basic Search" and "Advanced Search". Options available in these forms give users great flexibility in their searches to specify the organism of interest or authors, and to further refine categories and subcategories by application of cellular, biochemical or physiological limits. Moreover, researchers can select from particular experiments, specify genes and follows other filtering criteria and are allowed to edit or delete the annotations associated with the data. Finally, obtained results can be displayed in different formats and links are provided to explore more details about genes, arrays and experiments of interest.

Specific Tools

Quality Assessment is an exploratory graphic tool producing histograms or scatter plots of user-selected fields to look at data points distribution, an ANOVA analysis to detect spatial and print tip bias on the array and a tool from BioConductor's ArrayQuality package to view diagnostic and doping control plots for HEEBO/MEEBO arrays.

Hierarchical Clustering and other Data Analysis Tools are intended for data clustering and visualizating of co-expressed genes.

Singular Value Decomposition provides a mathematical algorithm for the determination of unique orthogonal gene and corresponding array expression patterns.

KNNImpute enables computation of missing values in data matrices before applying singular value decomposition (SVD) displayed as raster image or bar graph; SVD is a mathematical approach allowing determination of unique orthogonal genes and corresponding array expression patterns.

GO TermFinder determines whether a list of genes produced by any number of microarray analysis software tools has a significant enrichement of GO terms using Gene Ontology annotation.

Array Color Tool provides a simplified view of the data ratio for a given microarray allowing the user to quickly examine the microarray for evidence of global effects such as printing biases using ANOVA calculation to measure the dependence of per-sot ratio on printing plate and on sector.

Experiment Selection Tool selects from variety of experiments by slide name, organism, category and subcategory of experiments.

Expresion History Tool allows users to find microarray expression data for a gene throughout all experiments available.

Data Filtering Options offers variety of statictical measurements to filter and select data.

Clustering and Image Generation provides choice of self organizing maps or clustering tools using Pearson Correlation or Euclidean Distance.

Gene Selection and Annotation Tool enables selection of one or all genes and their annotation by description, gene model, GenBank accession, clone ID, or BAC locus.

Distinguishable Features

SMD is a fast expanding database using extensive data mining tools for microarray data processing, analyzing, visualizing and sharing. Since employing wide spectra of microarray platforms, software, organisms and experimental conditions, the database is able to meet broad users' expectancy. SMD is currently the largest public repository for Arabidopsis Affymetrix microarray data. However, using complete data repository and all associated tools is possible only after paid registration for long-term data storage. SMD users can store dataset before and after filtering and centering in preclustering files. One can find useful to be able to precisely set up criteria for data mining and use the Expression History tool to explore the behavior of a gene of interest across all microarrays. SMD has also implemented pipe-line that converts sets of microarray data into MEGA-ML files and deposits them directly to ArrayExpress and GEO database.

CONCLUSION

Enormous increase of gene expression profiling studies in recent years have supported rapid formation of extensive data repositories and development of number of on-line software applications allowing researchers to exploit available gene expression data. This chapter provided brief description and comparison of microarray-based gene expression databases and web-based tools. Special attention was paid to Affymetrix GeneChip platform that is the most widely used and serves as a standard for the model species *Arabidopsis thaliana*. However, several included databases comprise data acquired from various platforms of microarray-based experiments (ArrayExpress, SMD) or even non-array based technologies (GEO). Moreover, these databases contain expression data from different organisms such as SMD storing data from 53 organisms, GEO from over 100 individual organisms and ArrayExpress carrying data from even more than 200 organisms. From Affymetrix-focused portals, number of databases contains data from both array types – the first generation GeneChip® 8K AG Genome Arrays covering approximately one third of Arabidopsis genome and later and more complete ATH1 Whole Genome Array (22K). Fewer databases are focused solely on the later type of Affymetrix ATH1 GeneChip (BAR, aGFP).

The largest data repositories and databases with associated analysis tools often stand as integrated components of extensive bioinformatic projects run by large institutions like European Bioinformatics Institute (EMBL-EBI) or National Center for Biotechnology Information (NCBI). Other databases and web-based tools represent extra value added to microarray service (NASCArrays) or were created by established laboratories working in transcriptomic or bioinformatic field (BAR, aGFP). A subset of microarray databases were originally designed as public data repositories and retrieval systems for high throughput expression data. They serve mainly as data storage and offer fewer analysis tools (GEO, ArrayExpress, SMD). On the other hand, other software applications were designed primarily for data analysis. They exploit microarray data from many gene expression databases and offer large number of various analysis tools (BAR, Genevestigator V3).

Microarray data analysis tools can be divided into two main categories. The first category comprises gene expression visualization tools, whereas tools forming the second one offer more thorough analyses of gene expression data, for example gene grouping and display of co-expressed genes. The most advanced software applications (Genevestigator V3 or BAR) offer both types of analyses. Other tools were designed specifically for intuitive visualization of gene expression at various morphological and developmental stages (aGFP).

Almost all databases and web-based tools presented in this chapter are freely publicly available. Moreover, some databases offer the possibility of password-protected account to increase the work comfort (ArrayExpress). As an exception, Genevestigator V3 offers three levels of access – Open, Classic and Advanced. Classic and Advanced access options supplement free Open access, they are paid (with the exception of Classic access for academic users) and offer more data analysis tools as described previously. In addition, paying registration is required in the Stanford MicroArray Database for long-term storage.

Individual databases exploit various microarray data resources. Most databases process data obtained from publicly available data resources and repositories (NASC, FGCZ, GEO, TAIR, etc). In specialized cases like NASCArrays, they represent a functional overlay of existing microarray service. As a supplement, several projects allow user-controlled data submission (GEO, ArrayExpress),

And how to recognize the quality of experimental data in the database? First, the use of curated experiments is important prerequisite for correct interpretation of expression data. Currently, the large number of databases is strictly MIAME-compliant. The minimum information about microarray experiment (MIAME) contributes to increased comparability and of microarray data and so increases the value of interpretation of obtained results [Brazma et al. 2001]. In some cases, the use of alternative normalization algorithms may prove useful (aGFP). So, the user can identify the quality database by means of well-annotated raw and processed microarray data too. Further criteria comprise the assessment of expression data quality by measurement and evaluation of several quality control metrics. Quality control (QC) allows the detection of arrays with lower or poor quality and subsequent elimination of these arrays or whole experiments from the microarray database. Identification of problematic arrays increase the output reliability of given data analysis. Genevestigator V3 categorizes experimental data that passed the QC check into separated subset called “High quality arrays only”. Moreover, several databases allow users not only to extract gene expression data, but also facilitate the sharing of microarray experiment design and experimental protocols. This possibility is very helpful as it markedly improves the design of new experiments. In addition to gene expression data characterized by expression signal (sometimes along with detection call, p-value) and data obtained as a result of database query and data analysis, selected databases also include other data labels (gene annotation, gene ontology). On the other hand, some databases do not clearly mention the version of Arabidopsis genome annotation and this imperfection can contribute to data misinterpretation.

In addition to the database query, data analyses and data submission, most databases offer also data download. Database commonly support various data format options, from well-known XLS, CSV or Tab-delimited TXT files to specialised formats like SOFT text file (GEO). When downloading defined datasets, the user can often acquire additional information e.g. StatPairsUsed, PresentCall, Detection P-value (Genevestigator V3).

To date, the vast amount of transcriptomic data is publicly available in number of data repositories together with various data analysis tools providing instruments for their exploitation as well as assigning putative functional information to analysed genes. The large number of software applications has a user-friendly web-interface and broad scale of analysis tool, which may help researchers to open up new perspective of gene expression data and to exploit them for the planning of specifically targeted experiments and verification of their hypotheses.

ACKNOWLEDGMENTS

Authors gratefully acknowledge the financial support from GACR (522/06/0896), GAAVCR (KJB600380701) and MSMT CR (LC06004, OC08011).

REFERENCES

Aharoni, A; Vorst, O. DNA microarrays for functional plant genomics. *Plant Mol Biol*, 2002, 48, 99-118.

Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, 2000, 408, 796-815.

Ball, CA; Awad, IAB; Demeter, J; Gollub, J; Hebert, JM; Hernandez-Boussard, T; Jin, H; Matese, JC; Nitzberg, M; Wymore, F; Zachariah, ZK; Brown, PO; Sherlock, G. *Nucleic Acids Research,* 2005, 33 D580-D582

Barrett, T; Suzek, TO; Troup, DB; Wilhite, SE; Ngau, WC; Ledoux, P; Rudnev, D; Lash, AE; Fujibuchi, W; Edgar, R. NCBI GEO: mining millions of expression profiles - database and tools. *Nucleic Acids Res,* 2005, 33 Database Issue, D562-566.

Barrett, T; Edgar, R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol,* 2006a, 411, 352-369.

Barrett, T; Edgar, R. Mining microarray data at NCBI's Gene Expression Omnibus (GEO). *Methods Mol Biol*, 2006b, 338, 175-190.

Barrett, T; Troup, DB; Wilhite, SE; Ledoux, P; Rudnev, D; Evangelista, C; Kim, IF; Soboleva, A; Tomashevsky, M; Edgar, R. NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res,* 2007, 35 Database issue, D760-765.

Brazma, A; Hingamp, P; Quackenbush, J; Sherlock, G; Spellman, P; Stoeckert, C; Aach, J; Ansorge, W; Ball, CA; Causton, HC; Gaasterland, T; Glenisson, P; Holstege, FC; Kim, IF; Markowitz, V; Matese, JC; Parkinson, H; Robinson, A; Sarkans, U; Schulze-Kremer, S; Stewart, J; Taylor, R; Vilo, J; Vingron, M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet,* 2001, 29(4), 365-371.

Brazma, A; Parkinson, H; Sarkans, U; Shojatalab, M; Vilo, J; Abeygunawardena, N; Holloway, E; Kapushesky, M; Kemmeren, P; Lara, GG; Oezcimen, A; Rocca-Serra, P; Sansone, SA. ArrayExpress - a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res,* 2003, 31(1), 68-71.

Brazma, A; Kapushesky, M; Parkinson, H; Sarkans, U; Shojatalab, M. Data storage and analysis in ArrayExpress. *Methods Enzymol,* 2006, 411, 370-386.

Brenner, S; Johnson, M; Bridgham, J; Golda, G; Lloyd, DH; Johnson, D; Luo, S; McCurdy, S; Foy, M; Ewan, M; Roth, R; George, D; Eletr, S; Albrecht, G; Vermaas, E; Williams, SR; Moon, K; Burcham, T; Pallas, M; DuBridge, RB; Kirchner, J; Fearon, K; Mao, J; Corcoran, K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol*, 2000, 18, 630-634.

Childs, KL; Hamilton, JP; Zhu, W; Ly, E; Cheung, F; Wu, H; Rabinowicz, PD; Town, CD; Buell, CR; Chan, AP. The TIGR Plant Transcript Assemblies database. *Nucleic Acids Res,* 2007, 35(Database issue):D846-851.

Clarke, JD; Zhu, T. Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems. practical considerations and perspectives. *Plant J*, 2006, 45, 630-650.

Craigon, DJ; James, N; Okyere, J; Higgins, J; Jotham, J; May, S. NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res*, 2004, 32(Database issue), D575-577.

Demeter, J; Beauheim, C; Gollub, J; Henandez- Bouccard, T; Jin, H; Maier, D; Matese, JC; Nitzberg, M; Wymore, F; Zachariah, ZK; Brown, PO; Sherlock, G; Ball, CA. *Nucleic Acids Research,* 2007, 35, D766-D770

Donson, J; Fang, Y; Espiritu-Santo, G; Xing, W; Salazar, A; Miyamoto, S; Armendarez, V; Volkmuth, W. Comprehensive gene expression analysis by transcript profiling. *Plant Mol Biol*, 2002, 48, 75-97.

Dupľáková, N; Reňák, D; Hovanec, P; Honysová, B; Twell, D; Honys D. Arabidopsis Gene Family Profiler (aGFP)--user-oriented transcriptomic database with easy-to-use graphic interface. BMC Plant Biol**,** 2007, 23, 7:39.

Edgar, R; Domrachev, M; Lash, AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res,* 2002, 30(1), 207-210.

Edgar, R; Barrett, T. NCBI GEO standards and services for microarray data. *Nat Biotechnol,* 2006, 24(12), 1471-1472.

Gollub, J; Ball, CA; Binklay, G; Demeter, J; Pinknkelstein, DB; Hebert, JM; Hernandez-Boussard, T; Heng, J; Kaloper, M; Matese, JC; Schroeder, M; Brown, PO; Botstein, D; Sherlock, D. *Nucleic Acids Research,* 2003, 31 94-96

Heazlewood, JL; Tonti-Filippini, J; Verboom, RE; Millar, AH. Combining experimental and predicted datasets for determination of the subcellular location of proteins in Arabidopsis. *Plant Physiol*, 2005, 139(2), 598-609.

Heazlewood, JL; Verboom, RE; Tonti-Filippini, J; Small, I; Millar, AH. SUBA: the Arabidopsis Subcellular Database. *Nucleic Acids Res*, 2007, 35(Database issue), D213-218.

Hennig, L; Menges, M; Murray, JA; Gruissem, W. Arabidopsis transcript profiling on Affymetrix GeneChip arrays. *Plant Mol Biol*, 2003, 53, 457-465.

Hughes, TR; Marton, MJ; Jones, AR; Roberts, CJ; Stoughton, R; Armour, CD; Bennett, HA; Coffey, E; Dai, H; He, YD; Kidd, MJ; King, AM; Meyer, MR; Slade, D; Lum, PY; Stepaniants, SB; Shoemaker, DD; Gachotte, D; Chakraburtty, K; Simon, J; Bard, M; Friend, SH. Functional discovery via a compendium of expression profiles. *Cell*, 2000, 102, 109-126.

Ilic, K; Berleth, T; Provart, NJ. BlastDigester - a web-based program for efficient CAPS marker design. *Trends Genet*, 2005, 21(1), 36.

Jen, CH; Manfield, IW; Michalopoulos, I; Piney, JW; Willats, WGT; Gilmartin, PM; Westhead, DR. *Plant Journal*, 2006, 46, 336-348.

Kapushesky, M; Kemmeren, P; Culhane, AC; Durinck, S; Ihmels, J; Körner, C; Kull, M; Torrente, A; Sarkans, U; Vilo, J; Brazma, A. Expression Profiler: next generation - an online platform for analysis of microarray data. *Nucleic Acids Res*, 2004, 32, W465-470.

Laule, O; Hirsch-Hoffmann, M; Hruz, T; Gruissem, W; Zimmermann, P. Web-based analysis of the mouse transcriptome using Genevestigator. *BMC Bioinformatics,* 2006, 7:311.

Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet*, 1999, 21, 20-24.

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol*, 1996, 14, 1675-1680.

Mandaokar A, Thines B, Shin B, Markus Lange B, Choi G, Koo YJ, Yoo YJ, Choi YD, Browse J. Transcriptional regulators of stamen development in Arabidopsis identified by transcriptional profiling. *Plant J*, 2006, 46, 984-1008.

Manfield, WI; Jen, CH; Pinney, JW; Michalopoulos, I; Bradford, JR; Gilmartin, PM; Westhead, DR. *Nucleic Acids Research,* 2006, 34, 504-509.

Money, T; Reader, S; Qu, LJ; Dunford, RP; Moore, G. AFLP-based mRNA fingerprinting. *Nucleic Acids Res*, 1996, 24, 2616-2617.

Mukherjee, G; Abeygunawardena, N; Parkinson, H; Contrino, S; Durinck, S; Farne, A; Holloway, E; Lilja, P; Moreau, Y; Oezcimen, A; Rayner, T; Sharma, A; Brazma, A; Sarkans, U; Shojatalab, M. Plant-based microarray data at the European Bioinformatics Institute. Introducing AtMIAMExpress, a submission tool for Arabidopsis gene expression data to ArrayExpress. *Plant Physiol,* 2005, 139(2), 632-636.

Palaniswamy, SK; James, S; Sun, H; Lamb, RS; Davuluri, RV; Grotewold, E. AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks. *Plant Physiol,* 2006, 140(3), 818-829.

Parkinson, H; Sarkans, U; Shojatalab, M; Abeygunawardena, N; Contrino, S; Coulson, R; Farne, A; Lara, GG; Holloway, E; Kapushesky, M; Lilja, P; Mukherjee, G; Oezcimen, A; Rayner, T; Rocca-Serra, P; Sharma, A; Sansone, S; Brazma, A. ArrayExpress - a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res,* 2005, 33 Database Issue, D553-555.

Parkinson, H; Kapushesky, M; Shojatalab, M; Abeygunawardena, N; Coulson, R; Farne, A; Holloway, E; Kolesnykov, N; Lilja, P; Lukk, M; Mani, R; Rayner, T; Sharma, A; William, E; Sarkans, U; Brazma, A. ArrayExpress - a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res,* 2007, 35 Database issue, D747-750.

Rhee, SY; Beavis, W; Berardini, TZ; Chen, G; Dixon, D; Doyle, A; Garcia-Hernandez, M; Huala, E; Lander, G; Montoya, M; Miller, N; Mueller, LA; Mundodi, S; Reiser, L; Tacklind, J; Weems, DC; Wu, Y; Xu, I; Yoo, D; Yoon, J; Zhang, P. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res, 2003, 31(1), 224-228.

Rocca-Serra, P; Brazma, A; Parkinson, H; Sarkans, U; Shojatalab, M; Contrino, S; Vilo, J; Abeygunawardena, N; Mukherjee, G; Holloway, E; Kapushesky, M; Kemmeren, P; Lara, GG; Oezcimen, A; Sansone, SA. ArrayExpress: a public database of gene expression data at EBI. *C R Biol,* 2003, 326(10-11), 1075-1078.

Sarkans, U; Parkinson, H; Lara, GG; Oezcimen, A; Sharma, A; Abeygunawardena, N; Contrino, S; Holloway, E; Rocca-Serra, P; Mukherjee, G; Shojatalab, M; Kapushesky, M; Sansone, SA; Farne, A; Rayner, T; Brazma, A. The ArrayExpress gene expression database: a software engineering and implementation perspective. *Bioinformatics,* 2005, 21(8), 1495-1501.

Schena, M; Shalon, D; Davis, RW; Brown, PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 1995, 270, 467-470.

Schmid, M; Davison, TS; Henz, SR; Pape, UJ; Demar, M; Vingron, M; Scholkopf, B; Weigel, D; Lohmann, JU. A gene expression map of Arabidopsis thaliana development. *Nat Genet*, 2005, 37(5), 501-506.

Schwager, C; Blake, J. Bloader - a batch loader application for MIAMExpress. *Bioinformatics*. 2005, 21(8):1727-1729.

Sherlock, G; Hernandez-Boussard, T; Kasarskis, A; Binkley, G; Matese, JC; Dwight, SS; Kaloper, M; Weng, S; Jin, H; Ball, CA; Eisen, MB; Spellman, PT; Brown, PO; Botstein, D; Cherry, JM. Nucleic Acids Research, 2001, 152-155.

Spellman, PT; Miller, M; Stewart, J; Troup, C; Sarkans, U; Chervitz, S; Bernhart, D; Sherlock, G; Ball, C; Lepage, M; Swiatek, M; Marks, WL; Goncalves, J; Markel, S; Iordan, D; Shojatalab, M; Pizarro, A; White, J; Hubley, R; Deutsch, E; Senger, M; Aronow, BJ; Robinson, A; Bassett, D; Stoeckert, CJ Jr; Brazma, A. Design and implementation of microarray gene expression markup langure (MAGE-ML). *Genome Biol*, 2002, 3, RESEARCH0046.

Taylor, J; Provart, NJ. CapsID: a web-based tool for developing parsimonious sets of CAPS molecular markers for genotyping. *BMC Genet*, 2006, 10, 7:27.

Torrente, A; Kapushesky, M; Brazma, A. A new algorithm for comparing and visualizing relationships between hierarchical and flat gene expression data clusterings. *Bioinformatics*, 2005, 21(21), 3993-3999.

Toufighi, K; Brady, SM; Austin, R; Ly, E; Provart, NJ. The Botany Array Resource: e-Northerns; Expression Angling; and promoter analyses. *Plant J*, 2005, 43(1), 153-63.

Troyanskaya, O; Cantor, M; Sherlock, G; Brown, P; Hastie, T; Tibshirani, R; Botstein, D; Altman, RB. Missing value estimation methods for DNA microarrays. *Bioinformatics*, 2001, 17(6), 520-525.

van Hal, NL; Vorst, O; van Houwelingen, AM; Kok, EJ; Peijnenburg, A; Aharoni, A; van Tunen, AJ; Keijer, J. The application of DNA microarrays in gene expression analysis. *J Biotechnol*, 2000, 78, 271-280.

Velculescu, VE; Zhang, L; Vogelstein, B; Kinzler, KW. Serial analysis of gene expression. *Science*, 1995, 270, 484-487.

Wheeler, DL; Barrett, T; Benson, DA; Bryant, SH; Canese, K; Church, DM; DiCuccio, M; Edgar, R; Federhen, S; Helmberg, W; Kenton, DL; Khovayko, O; Lipman, DJ; Madden, TL; Maglott, DR; Ostell, J; Pontius, JU; Pruitt, KD; Schuler, GD; Schriml, LM; Sequeira, E; Sherry, ST; Sirotkin, K; Starchenko, G; Suzek, TO; Tatusov, R; Tatusova, TA; Wagner, L; Yaschenko, E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, 2005, 33(Database issue), D39-45.

Wheeler, DL; Barrett, T; Benson, DA; Bryant, SH; Canese, K; Chetvernin, V; Church, DM; DiCuccio, M; Edgar, R; Federhen, S; Geer, LY; Kapustin, Y; Khovayko, O; Landsman, D; Lipman, DJ; Madden, TL; Maglott, DR; Ostell, J; Miller, V; Pruitt, KD; Schuler, GD; Sequeira, E; Sherry, ST; Sirotkin, K; Souvorov, A; Starchenko, G; Tatusov, RL; Tatusova, TA; Wagner ,L; Yaschenko, E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, 2007, 35(Database issue), D5-12.

Whetzel, PL; Parkinson, H; Causton, HC; Fan, L; Fostel, J; Fragoso, G; Game, L; Heiskanen, M; Morrison, N; Rocca-Serra, P; Sansone, SA; Taylor, C; White, J; Stoeckert, CJ Jr. The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics*, 2006, 22(7), 866-73.

Winter, D; Vinegar, B; Nahal, H; Ammar, R; Wilson, GV; Provart, NJ. An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS ONE*, 2007, 2(1), e718.

Yang, YH; Dudoit, S; Luu, P; Lin, DM; Peng, V; Ngai, J; Speed, TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*. 2002, 30(4), e15.

Zhang, W; Morris, QD; Chang, R; Shai, O; Bakowski, MA; Mitsakakis, N; Mohammad, N; Robinson, MD; Zirngibl, R; Somogyi, E; Laurin, N; Eftekharpour, E; Sat, E; Grigull, J; Pan, Q; Peng, WT; Krogan, N; Greenblatt, J; Fehlings, M; van der Kooy, D; Aubin, J; Bruneau, BG; Rossant, J; Blencowe, BJ; Frey, BJ; Hughes, TR. The functional landscape of mouse gene expression. *J Biol*. 2004, 3, 21.

Zhu, T. Global analysis of gene expression using GeneChip microarrays. *Curr Opin Plant Biol*, 2003, 6, 418-425.

Zimmermann, P; Hirsch-Hoffmann, M; Hennig, L; Gruissem, W. GENEVESTIGATOR. Arabidopsis Microarray Database and Analysis Toolbox. *Plant Physiol,* 2004, 136(1), 2621-2632.

Zimmermann, P; Hennig, L; Gruissem, W. Gene-expression analysis and network discovery using Genevestigator. Trends Plant Sci, 2005, 10(9), 407-409.

8.3. Wide-scale screening of T-DNA lines for transcription factor genes affecting male gametophyte development in Arabidopsis

Reňák David*, **Dupľáková** Nikoleta*, Honys David

*should be regarded as equal contributors

Sexual Plant Reproduction 2010 (in review)

This chapter describes our pilot project for the discovery of the transcription factors important for male gametophyte development. It presents the genetic and phenotypic analysis 73 T-DNA insertion representing 27 early and 21 late genes. The T-DNA line(s) exhibiting the most severe mutant phenotype disturbation(s) were selected for further functional characterization.

My contribution: As a joint first author I was responsive for the complete analysis of all late pollen transcription factors. It represented 37 lines. I was also involved in the manuscript preparation.

Supporting online material on CD

Wide-scale screening of T-DNA lines for transcription factor genes affecting male gametophyte development in *Arabidopsis*

David Reňák[1,2,*], Nikoleta Dupľáková[1*], David Honys[1,3]

[1]*Institute of Experimental Botany ASCR, Laboratory of Pollen Biology, Rozvojová 263, 165 02 Prague 6, Czech Republic*

[2]*University of South Bohemia, Faculty of Science, Department of Plant Physiology and Anatomy, Branišovská 31, 370 05 České Budějovice, Czech Republic*

[3]*Charles University in Prague, Faculty of Science, Department of Plant Experimental Biology, Viničná 5, 128 44 Prague 2, Czech Republic*

[*]should be regarded as equal contributors

Abstract

Male gametophyte development leading to the formation of a mature pollen grain is precisely controlled at various levels during its whole progression. Transcriptomic studies exploiting genome-wide microarray technologies revealed the uniqueness of pollen transcriptome and the dynamics of early and late successive global gene expression programs. However, the knowledge of transcription regulation is still very limited. In this study, we focused on the identification of pollen-expressed transcription factor (TF) genes involved in the regulation of male gametophyte development. To achieve this, reverse genetic approach was used. 73 T-DNA insertion lines were screened representing 48 genes of 21 TF families active in either early or late pollen development. In the screen, ten phenotype categories were distinguished affecting various structural or functional aspects including pollen abortion, presence of inclusions, variable pollen grain size, disrupted cell wall structure, cell cycle defects and male germ unit organization. Thirteen lines were not confirmed to contain the T-DNA insertion. Among 60 confirmed lines, about a half (28 lines) showed strong phenotypic changes (i.e. ≥25% aberrant pollen) including four lines producing remarkable high proportion (70-100%) of disturbed pollen. However, remaining 32 lines exhibited mild defects or resembled wild-type appearance. There was no significant bias towards any phenotype category amongst early and late TF genes and interestingly within individual TF families. Presented results have a potential to serve as a basal information resource for future research on the importance of respective TFs in male gametophyte development.

Keywords

Male gametophyte, transcription factor, T-DNA insertion line, phenotype screen, pollen phenotypic defects

Abbreviations

BCP: bicellular pollen; MGU: male germ unit; MPG: mature pollen grain; SAIL: (Syngenta *Arabidopsis* insertion library); TCP: tricellular pollen; TF: transcription factor; UNM: unicellular misrospore; ZFP: zinc-finger protein

Introduction

The haploid male gametophyte generation of flowering plants is represented by pollen grain consisting of two/three cells; one generative cell/two sperm cells and one vegetative cell later growing into a pollen tube. Pollen development is a complex process requiring the coordinated activity of various cells and tissues and their associated specific gene expression patterns. Hence, male gametophyte represents an excellent model for studies of fundamentally important developmental processes including cell specialization and differentiation, cell cycle regulation, cell polarity and growth, and the regulation of gene expression.

In a post-genomic era, Arabidopsis research has focused on the functional genomics aiming at the functional characterization of large collections of genes. Forward and reverse genetic screens have led to the identification of number of pollen mutant phenotypes caused by various mutagens (Chen and McCormic 1996; Grini et al. 1999; Park et al.1998), T-DNA insertions (Bonhomme et al. 1998; Howden et al. 1998; Johnson et al. 2004; Sessions et al. 2002) or transposons (Lalanne et al. 2004) and reviewed by Borg et al. (2009), Honys et al. (2006), and Twell et al. (2006). In particular, efficient reverse genetic approach, the essential component of functional genomics, has widely launched after a collection of characterized T-DNA insertion lines was made available for a broad scientific community. Moreover, high-throughput technologies enabling transcriptomic studies of global male gametophyte gene expression revealed high portion of male gametophytic transcriptome as well as significant overlap between gametophytic and sporophytic tissues. The total number of genes expressed during pollen development and function was estimated to be over 14,000. Surprisingly, only <5% represented genes showed strictly pollen-specific expression pattern (Honys and Twell 2004, Wang et al. 2008).

Despite a long-term research on the field of plant sexual reproduction, the developmental changes and regulatory mechanisms underlying the pollen grain formation are far from fully characterized. It is agreed that both transcription and translation play a key role in global and specific gene expression patterns during pollen maturation (Becker et al. 2003; Grobei et al. 2009; Honys and Twell 2003, 2004; Pina et al. 2005). On the contrary, the progamic phase of many species has been shown to be greatly independent on transcription but vitally dependent on translation instead (see Honys et al. 2006) that is widely facilitated by the activity of complex heavy multifunctional ribonucleoprotein particles (EPPs; Honys, Reňák et al. 2009). Despite of steadily growing knowledge-base of transcriptional regulation, the role of transcription factors (TFs) in male gametophyte development and function is still rather limited (reviewed by Borg et al. 2009; Honys et al. 2006; Twell et al. 2006).

In general, our present knowledge is concentrated mostly on sporophytic mutations. To date, over 45 mutations have been reported to cause cellular and developmental disorders, whose genes are involved in metabolism, protein synthesis, signal transduction, intracellular transport, cell wall synthesis, cytoskeleton and cytokinesis (reviewed by Borg et al. 2009; Honys et al. 2006; Twell et al. 2006). But only a few TF genes have been identified to play a role in sporophytic transcriptional regulation during anther formation followed by male gametophyte development. The gene regulatory network in *Arabidopsis* early anther development has been traced and partially disclosed. Leaving the ABCE model including *AGAMOUS* gene, downstream-regulated *SPOROCYTELESS/NOZZLE (SPL/NZZ)* MADS-box transcription factor determinates the sporogenous cell fate and anther wall development

(Ito at al. 2004; Schiefhaler et al. 1999; Yang et al. 1999). *SPL/NZZ* directly regulates bZIP-family protein *EXCESS MICROSPOROCYTES1/EXTRA SPOROGENOUS CELLS (EMS1/EXS)*. In *ems/exs* mutant, additional microsporocytes are formed instead of tapetal cells and thus viable pollen production is blocked (Canales et al. 2002; Zhao et al. 2002). *DYSFUNCTIONAL TAPETUM1 (DYT1),* a putative bHLH transcription factor, is likely to be involved in the regulation of many tapetal genes, either directly or indirectly, including *AMS* and *MS1* (Zhang et al. 2006). *ABORTED MICROSPORES (AMS)* is a member of MYC class of bHLH TFs. *ams* plants are completely devoid of pollen that is degenerated shortly after microspores release from the tetrad, prior to pollen mitosis I (Sorensen et al. 2003). *MALE STERILITY (MS1)* encodes a nuclear protein with a PHD-finger domain accompanying RING finger and putative leucine zipper motifs. This factor has been also proposed to regulate the expression of sporophytic genes necessary for pollen-exine formation since the mutant pollen shows aberrant exine deposition. Moreover, *MS1* was the first transcriptional regulator gene reported to affect the pollen development (Ariizumi et al. 2005; Ito and Shinozaki 2002; Ito et al. 2007; Vizcay-Barrena and Wilson 2006; Wilson et al. 2001; Yang et al. 2007). *MYB33/65,* regulated by *SPL/NZZ* either directly or via *DYT1,* are GYMYB-like MYB TFs whose mutations cause tapetum hypertrophy and premeiotic abortion of pollen development (Millar and Gubler 2005), *MYB103* is regulated by *EMS1/EXS* and its downregulation results in premature tapetum degeneration and pollen abortion (Zhang et al. 2007). In addition to the above regulatory network, *MYB32* is expressed in tapetum and its mutation leads to aberrant pollen development and partial male sterility (Preston et al. 2004). Similarly, mutation in*MYB26* gene causes male sterility by defective cell wall fortification in endothecial cell layer in anthers (Steiner-Lange et al. 2003). In *Petunia*, mutations in two tapetum-expressed genes encoding zinc-finger proteins TAZ1 (Kapoor et al. 2002) and MEZ1 (Kapoor and Takatsuji 2006) caused aberrant meiosis and pollen abortion. Besides specific TFs, *atTAF6* (member of TBP-associated transcription factors, subgroup of common TFIID) was identified to affect male gametophyte development causing reduced transmission of the mutant allele (Lago et al. 2005).

On the contrary, of male gametophytic transcription factors specifically expressed during pollen development, only five were studied in detail. Unusual R2R3 MYB gene, *DUO POLLEN1* (*DUO1*) is specifically expressed in the male germline and has been shown to be a key regulator of germ cell division and sperm cell formation Arabidopsis (Brownfield et al. 2009; Durbarry et al. 2005; Rotman et al. 2005). *MALE MEIOCYTE DEATH1 (MMD1)* encoding PHD-finger protein is expressed preferentially during male meiosis. It participates in chromatin remodelling and/or transcriptional events for successful progression through meiosis as *mmd1* plants show chromatin fragmentation followed by cell death before cytokinesis (Yang et al. 2003). TCP-family *TCP16* is expressed predominantly in developing microspores; *TCP16*–targeted RNAi caused rapid abortion of microspores and abnormal pollen structure (Takeda et al. 2006). *HAPLESS5* (*HAP5)* encodes AP2-domain protein and its mutation causes pollen morphological defects and abortion (Johnson et al. 2004). Finally, At*bZIP34* controls pollen wall patterning and affects several metabolic pathways (Gibalová et al. 2009). The interaction map of MADS-box genes (de Folter et al. 2005; Veron et al. 2007) and particularly the male gametophyte-specific regulatory network of their AtMIKC* subgroup (Verelst et al. 2007) belong among the best characterized. The authors presented the first view of a regulatory network controlling transcriptome dynamics during reproductive cell differentiation in plants. AtMIKC* proteins were shown to interact with each other forming five heterodimeric TF complexes that play a major role in the transcriptional switch during pollen maturation. Transcriptomic analysis of mutant plants revealed more than 1,300 regulated genes including downstream-acting TFs such as AGL18, AGL29, WRKY34, MYB97, MYB related CCA1, SCL8, EIL1 and bZIP1. It was the first analysis of a partial

regulatory network that controls transcriptome dynamics during reproductive cell differentiation in plants.

Yet there is apparent knowledge deficiency of transcriptional regulatory networks affecting the male gametophyte development. To address this question, we employed a large collection of T-DNA insertion lines for the wide-scale screen for transcription factors likely to play a regulatory role in male gametophyte development. Here we present results of the phenotype screen of 73 potential knock-down lines of genes encoding early and late male gametophytic transcription factors. We identified 28 lines with strong pollen phenotypic defects causing ≥25% abnormal pollen. Among 21 TF families included in the study, there was no preferences observed to any type of cellular disturbances, and moreover, no bias of early and late TFs towards any class of phenotypic defects. On the whole, revealing structural defects of respective TFs provides a broad information resource for their potential importance in regulatory networks of male gametophyte development.

Materials and methods

Microarray data

For candidate genes selection, previously published transcriptomic data from Affymetrix 23K Arabidopsis ATH1 arrays were used. These datasets included normalized gene expression levels for both gametophytic (uninucleate microspores (UNM), bicellular pollen (BCP), tricellular pollen (TCP) and mature pollen grain (MPG)) and number of sporophytic tissues (seedling, whole plant, stem, petiole, rosette leaf, root, root hair elongation zone and cell suspension culture). Publicly available transcriptomic datasets were downloaded from arabidopsisGFP database (http://aGFP.ueb.cas.cz; Duplakova et al. 2007). Microarray data were normalized using dChip 1.3 software (http://www.dchip.org). The reliability and reproducibility of analyses was ensured by the use of duplicates or triplicates in each experiment, the normalization of all arrays to the median probe intensity level and the use of normalized CEL intensities of all arrays for the calculation of model-based gene-expression values based on the Perfect Match-only model (Li and Wong 2001a,b).

Selection of candidate transcription factor genes

Genes encoding putative transcription factors were compiled from *Arabidopsis thaliana* transcription factors database AGRIS (Arabidopsis Gene Regulatory Information Server; http://arabidopsis.med.ohio-state.edu/ Davuluri et al. 2003; Palaniswamy et al. 2006) and DAFT (The Database of Arabidopsis Transcription Factors; http://datf.cbi.pku.edu.cn/; Guo et al. 2005).

Microarray datasets were analyzed using Microsoft Excel. Candidate genes were pre-selected and further analyzed prior to insertion lines ordering. For the pre-selection, only confirmed or putative transcription factor-coding genes exhibiting pollen-enriched or pollen-specific expression pattern were considered with preference to those with abundant expression levels in gametophytic tissues. The pre-selected set was further divided into two subsets representing genes with peak expression in early (UNM and BCP) or late (TCP and MPG) stages of male gametophyte development. Application of these criteria resulted in the selection of more than 220 genes in each subset.

Two pre-selected gene subsets were cluster-analyzed using publicly available EPCLUST software (Expression Profile data CLUSTering and analysis; http://www.bioinf.ebc.ee/EP/EP/EPCLUST/; Kapushesky et al. 2004). Two algorithms of

hierarchical and K-means type were applied resulting into two refined clusters of about 50 early and late pollen candidate genes with similar expression profiles.

Selection of T-DNA insertion lines

T-DNA collections were searched for insertions in selected candidate genes (T-DNA Express: Arabidopsis Gene Mapping Tool database; http://signal.salk.edu/cgi-bin/tdnaexpress; Alonso et al. 2003). The collection included both SAIL lines (pCSA110 and pDAP101 transformants) and SALK lines (pBIN-pROK2 transformants). When more insertion lines were available for particular gene, T-DNA lines with inserts in the promoter region or CDS rather than in 3'UTR were preferred.

Plant material and DNA extraction

Plants were grown in a greenhouse or a cultivation room under standard conditions of 16hr/8hr day/night regime, 20°C and illumination of 150 µmol $m^{-2}s^{-1}$. Approximately 30 mg of leaf tissue was collected from all plants at rosette stage of each independent mutant line. Samples were frozen in liquid nitrogen and ground with glass beads in Silamat S5 homogenizer for 12 sec. 250 µl of CTAB extraction buffer (1.4 M NaCl, 20 mM EDTA, 100 mM Tris pH 8.0, 3% CTAB) was added and the sample was extracted at room temperature (RT) for 20 min. Then, 250 µl of chloroform:isoamyl alcohol mixture (CH:IAA, 24:1) was added and the sample was mixed. Samples were centrifuged (13,000 rpm, 10 min, RT) and 200 µl of upper aqueous phase was transferred into a new tube containing 140 µl isopropanol. After vortexing, tubes were left 5 min for DNA denaturation and centrifuged (13,000 rpm, 2 min, RT). DNA pellet was washed with 1 ml 70% ethanol and, after a final centrifugation (13,000 rpm, 2 min, RT), carefully dried, resuspended in 50 µl H_2O and stored at -20°C.

Genotype analysis

PCR amplification was used for genotyping of each plant employing two pairs of primers, a gene specific primer designed individually for each gene and T-DNA left border-localized primer designed for SALK (LB22: 5'-GCGTGGACCGCTTGCTGCAACT-3') or SAIL (LB3_25: 5'-AATTTCATAACC-AATCTCGATACAC-3') insertion lines. Appropriate gene-specific primers designed by Primer3 software (http://www-genome.wi.mit.edu/ cgi-bin/primer/primer3_www.cgi) are listed in Table 1. PCR amplification was performed on Biometra T1 or Eppendorf Mastercycler ep-gradient thermocycler under following conditions: (1) 5 min at 94°C, (2) 35 cycles of 30 s at 94°C, 30 s at the optimal annealing temperature (45-65°C), and 1 min at 72°C, followed by (3) 10 min at 72°C and (4) pause at 4°C using Taq DNA polymerase (Fermentas). PCR products were separated on 1% TAE-agarose gels and visualised with ethidium bromide or SYBR Green.

Microscopy analysis

For phenotype analysis, three fully open flowers in developmental stage 14 (Smyth et al. 1990) were collected into each well of a 96-well plate containing 50 µl of DAPI (4',6-diamidino-2-phenylindole)/GUS staining buffer (400 ng/ml, Park et al. 1998; GUS solution: 50 mM Na_2HPO_4, 50 mM NaH_2PO_4, 10 mM EDTA, Gallagher 1992) and shortly shacken to release pollen grains from anthers. Pollen was observed in inverted microscope Nikon Eclipse TE2000-E in bright field and UV epifluorescence. The observation focused specifically on pollen size, shape and ultrastructure. The percentage of phenotypic defects was calculated

from observation of at least 300 pollen grains in each sample. For each pollen grain observed, only one most significant abnormality was counted (for details see Results and Discussion) in order to eliminate the artificial accumulation of phenotypic defects.

Results and Discussion

Data processing and selection of pollen specific TF genes

Previous male gametophytic developmental transcriptomic experiments revealed that of all 22,591 genes harboured on Affymetrix ATH1 Whole Genome Array, 1,595 putative TF genes were expressed in gametophytic and/or various sporophytic tissues (Honys and Twell 2004). In this experiment, datasets covering four gametophytic stages (UNM, BCP, TCP, MPG) and eight sporophytic tissues and cell types (cotyledon, whole sporophyte, stem, petiole, rosette leaf, root, root hair and cell suspension culture) were compared.

First, putative transcription factor genes with no or very low expression signal in any of four gametophytic stages were eliminated. Such treatment left 1,358 putative TF genes with reliable signal in at least one stage of male gametophyte development.

Second, genes preferentially expressed in male gametophyte were pre-selected. For this selection, a stringent, three-fold threshold value was used. Only genes with the strongest expression signal in male gametophyte at least 1.5-times (early genes) or 3-times (late genes) higher than the strongest expression signal in any sporophyte dataset were selected according to microarray data available at the time of selection. Such treatment trimmed the candidate group to 450 TF genes (225 early and 225 late TF genes) for subsequent clustering. As new male gametophytic datasets have emerged since the time of candidate genes selection (Borges et al., 2008, Wang et al., 2008, Qin et al., 2009), the actual threshold value does not always correspond to the original values.

Third, 450 pre-selected genes were cluster-analyzed using EPCLUST software. Two alternative algorithms were used, hierarchical (correlation measure-based distance) and K-means clustering (12 clusters used, Euclidean distance). It allowed the precise identification of two clusters of pollen-specific/enriched putative TF genes expressed in early (before PMII; UNM, BCP) and late stages (after PMII; TCP, MPG) of Arabidopsis male gametophyte development. Finally, 100 pollen-specific/enriched putative TF-encoding genes were manually selected from these two subsets (50 early and 50 late TF genes). Selected candidates were then individually searched for the T-DNA insertions available from either SALK or SAIL insertion line seed stocks. Wherever possible, insertion sites within gene coding regions or promoter sequences were preferred. The complete procedure resulted in a list of 27 early and 21 late genes for which 73 T-DNA insertion lines were ordered (Table 1). Most genes were represented by one insertion line; however, several genes matched to more than one insertion line either from SAIL or SALK seed stock. The updated table showing expression values and detection calls of all selected genes is presented as the Supplementary table 1.

Confirmation of the presence of T-DNA insertion

In all 73 lines, the presence of T-DNA insert at the correct position was verified. All gene-specific primers are listed in Table 1 as well as the length of expected PCR products and position of the insertion site within each gene. In most insertion lines, wild-type allele was recovered by PCR; the use of forward and reverse gene-specific primers gave expected products. The only exceptions were nine homozygous lines (Table 1). In 60 out of 73 lines (82%), the combination of T-DNA insert-specific LB primer and appropriate gene-specific

primer gave a PCR product with correct size. We identified nine homozygous lines in total (six lines harbouring the insertion in early TF genes and three in a late gene). Only two of them (SALK_114109 and SALK_130880, both in early TF genes) were annotated as homozygous in T-DNA Express (http://signal.salk.edu/cgi-bin/tdnaexpress), the rest was annotated as heterozygous. Further ten lines (four early TFs and six late TFs) annotated as homozygous elsewhere were found heterozygous or not confirmed in one case (Table 1). Moreover, the presence of expected T-DNA insert was not repeatedly confirmed in remaining 13 out of all 73 T-DNA insertion lines (18%). All unverified lines were excluded from the phenotype screen.

Phenotype categories

After genotyping of all insertion lines, mature pollen was screened for the putative phenotypic defects caused by the T-DNA insertion in the respective TF gene. For this, both bright field and epifluorescence microscopy was employed. We restricted the phenotype screen at mature pollen developmental stage with special attention paid to pollen abortion, cell wall structure, the presence of cytoplasmic inclusions and number, position and shape of nuclei. We presumed that the absence of early and late TF genes would lead to structural disorders throughout pollen development and thus would cause phenotypic defects persisting in mature pollen.

Several distinguishable categories of mutant phenotypic defects were observed and classified. Despite of 21 TF families represented by 73 insertion lines, pollen mutant phenotypes did not exhibit great variation. Such limited portfolio of pollen defects may seem unexpected at this wide-scale screening. However the role of transcription factors in complex networks should be regarded as different regulators can trigger various biochemical processes that finally affect similar cellular processes and functions like cytoskeleton organization, cell cycle, cellular transport, cell wall metabolism etc. Such pleiotropic activity has repeatedly been reported (Gibalová et al. 2009; Twell et al. 2006; Verelst et al. 2007). Five complex developmental and functional disorder classes were observed in comparison with standard wild-type phenotype. Those included (A) pollen abortion, (B) presence of cytoplasmic inclusion, (C) pollen grain size and cell wall structure, (D) cell cycle defects, and (E) MGU organization. However, the precise phenotype analyses have led to refinement of the classification of above dysfunctions into ten categories: A: (1) abortion; B: (2) inclusions; C: (3) small grain, (4) oval grain, (5) deformed cell wall; D: (6) one-celled pollen, (7) two-celled pollen; E: (8) eccentric MGU, (9) separated MGU, and (10) linear MGU (Fig. 1).

We mostly observed phenotypic abnormalities associated with cytoskeleton organization and function leading to the defective arrangement of MGU with their eccentric (Fig. 2J), separated (Fig. 2K) or linear position (Fig. 2L) in the pollen grain unlike their characteristic central arrangement in wild-type pollen (Fig. 2G). Because of the sole positional variation defects, we considered and discussed those phenotypes collectively as misarranged MGU from here onwards. Similar phenotypic defects were observed by Lalanne and Twell (2002) in *mgu-displaced (mud)* and *germ-unit-malformed (gum)* mutations. In this respect, eccentric MGU resembles *mud* and dispersed and/or linear MGU resembles *gum* mutation. However, *mud* and *gum* mutations are fully penetrant and cause remarkably reduced pollen fitness (Lalanne and Twell 2002). Other classes of mutations were likely to affect cell cycle and cell division. Namely, PMI defects resulted in one-celled pollen (Fig. 2H) and/or PMII defects resulted in two-celled pollen (Fig. 2I) with a vegetative nucleus and a generative cell. Cytoplasmic inclusions represented more specific phenotypic defects (Fig. 2C). Despite their unknown composition, they were likely formed as a consequence of impaired metabolical and/or secretory pathways as recently published for bZIP-family transcription

factor AtbZIP34 (Gibalová et al. 2009). The most serious and complex phenotypic defect(s) led to a complete cell abortion (Fig. 2B) with squashed pollen wall. Considering the pollen shape, we observed unusual pollen grain size and shape such as small (Fig. 2D) and oval pollen grains (Fig. 2E). However, these phenotypic defects were rare. Only one insertion line produced pollen with rough cell wall surface (Fig. 2F).

Phenotype screen of T-DNA insertion lines

Complete results of large phenotype screen are listed in Table 2. The table includes AGI code, gene family, insertion line and ten phenotype categories both in bright field and epifluorescence microscopy. While few T-DNA insertion lines showed strictly one type of phenotypic disorder, most showed more than one phenotypic defect. There was no significant bias towards any phenotype category when comparing early and late TFs. Moreover, there were no preferences among individual TF families in disrupting cellular events. The common cellular disruption affected MGU position resulting into three phenotypic defects, eccentric, separated or linear MGU (Fig. 1).

Among 60 confirmed lines, about a half (28 lines) showed strong phenotypic changes (i.e. ≥25% aberrant pollen) that are further discussed. But only few lines producing heavily disturbed pollen grains were observed, which reached high penetrance of 100% (SALK_140819), 80-85% (SALK_018864), 70-75% (SALK_010704) and 70-75% (SALK_084050). Generally, in the population of homozygous plants we mostly identified remarkable proportion of plants with wild-type phenotype counting the average penetrance in between 10-30%. Incomplete penetrance represents characteristic phenomenon of many pollen mutations as observed for *hapless5* mutation counting 25% morphologically aberrant pollen (Johnson et al. 2004), *limpet pollen* with 25% abnormal phenotype (Howden et al. 1998), *gemini pollen I* mutants showing 40% aberrant pollen (Park et al. 1998, 2004) and *sidecar pollen* producing 80% aberrant pollen (Chen and McCormick 1996). Other authors also suggested gametophytic mutations as pleiotropic, incompletely penetrant and displaying variable expressivity due to functional redundancy (Bonhomme et al. 1998; Drews and Yadegari 2002; Feldmann et al. 1997; Grini et al. 1999; Verelst et al. 2007). This phenomenon was especially appartent among transcription factors (Blanc et al. 2000; Verelst et al. 2007). Indeed, the vast majority of several hundred knockouts in *Arabidopsis* did not give rise to visible, directly informative phenotypes (Bouché and Bouchez 2001). Lower penetrance was also clearly demonstrated in several SAIL lines harbouring *quartet* mutation as a background for segregation analyses with all four pollen grains bound together in a tetrade (Francis et al. 2006; Rhee and Somerville 1998). Typically only one pollen grain in a tetrade showed developmental disorder instead of two in heterozygous plants or all four grains in homozygotes as expected for fully penetrant mutations.

Out of 16 heterozygous lines no homozygous plants were identified presumably due to limited population size. In the group of 60 T-DNA insertion lines analyzed, twelve were previously annotated as homozygous in T-DNA Express. However, only two of them were confirmed as homozygous. On the other hand, seven other lines previously annotated as heterozygous were identified as homozygous. The phenotypes of all homozygous plants were confirmed to co-segregate with the genotype after back-crossing to wild-type arabidopsis Col-0 plants.

The most prominent pollen phenotypic defects with respect to TF families

C2H2-type zinc finger proteins (ZFPs) are known to play a role in important pathways during plant vegetative growth, reproductive development and cold, salt and water stress

response (Milla et al. 2006; Sakamoto et al. 2000, 2004). A number of plant ZFPs from *Petunia* and *Arabidopsis* were also functionally characterised and are implicated in leaf and lateral shoot initiation, floral organogenesis, gametogenesis and stress response (Kapoor et al. 2006). *NUBBIN (NUB)* and *JAGGED (JAG)* promote the growth of the pollen-bearing microsporangia in anthers and are involved in the maintenance of the correct number of cell layers (Dinneny et al. 2006). In most C2H2 mutant lines abundant disturbances were noticed to affect the majority of pollen population. Four insertion lines of the same At4g05330 gene were tested (Table 2). The respective lines showed: 40% misarranged MGU, 10-15% two-celled pollen in homozygous SALK_090677; 20-25% misarranged MGU, 10-15% two-celled pollen in heterozygous plants of SALK_075976 line; 15-20% misarranged MGU, 10-15% two-celled pollen in heterozygous plants of SALK_075978; and finally, 20-25% misarranged MGU and 5-10% two-celled pollen in heterozygous plants of SALK_068135 line. Similarly, another C2H2 TF gene showed 40-45% misarranged MGU and 5% two-celled pollen in SALK_112731 line (At3g10470). Such large numbers implicated an important role of respective C2H2 factors in pollen-development regulatory network. Furthermore, it also highlighted their possible unique function and lack of functional redundancy that would be unusual in large C2H2 family. By contrast, members of size-comparable AtMIKC* MADS family showed large functional redundancy (Verelst et al. 2007)

CO-like C2C2 TFs are tightly involved in photoperiodic response pathway and photomorphogenesis conducted by *CONSTANS (CO)* and *CONSTANS-LIKE (COL)* genes. During floral transition, shoot apical meristem undergoes structural and functional changes including the cell cycle regulation, cytoskeletal and cell wall re-arrangement, energy metabolism, and hormonal response. Hence, the C2C2-CO-like factors play an important role in generative development. Accordingly, we observed strong structural disturbances in mutant pollen counting 40-45% misarranged MGU and 10-15% two-celled pollen in SALK_010704 line; 20% misarranged MGU and 10-15% two-celled pollen in heterozygous plants of SALK_138034 line; and 25% misarranged MGU and 15-20% two-celled pollen in SAIL_603_D10 line; all three lines hit At5g54470 gene as listed in Table 2.

The C3H family of zinc finger TFs is believed to play an effective role in stress tolerance since expression profile indicated that most members of this family were regulated by abiotic or biotic stresses (Wang et al. 2008). Pollen desiccation can represent the triggering event requiring their activity as mutations in two members of C3H TFs led to altered pollen phenotype: 40% misarranged MGU and 30% distinctively oval pollen grains in SALK_084050 line (At3g15740); 10% aborted grains in SAIL_764_B09 line (At3g15740); 15-20% misarranged MGU in SALK_041477 line (At2g37950) and 20% aborted grains, 10% two-celled pollen in SAIL_105_D03 line (At2g37950) (Tab. 2).

bHLHs form the second largest transcription factor family in plants and thus they govern a wide range of biological processes including growth and development (Riechmann et al. 2000; Toledo-Ortiz et al. 2003). bHLH transcription factors have been shown to dimerize with other bHLHs and with members of other transcription factor families such as MYBs or bZIPs and both partners may be required to control transcription of target genes (Abe et al. 2003; Grotewold et al. 2000). Considering male gametophyte development, *ABORTED MICROSPORES(AMS)* is up to date the only bHLH TF characterised; *ams* plants are completely devoid of mature pollen that degenerates shortly after microspores release from tetrads (Sorensen et al. 2003). However, in our screen, the most aberrant phenotype represented 10-15% aborted pollen in SALK_043690 plants (At5g54680)(Tab. 2).

The AP2 (APETALA2)/EREBP (Ethylene Responsive Element Binding Protein) multigene family includes numbers of developmentally and physiologically important transcription factors. AP2/EREBP genes are organized in two subfamilies: AP2 genes with two AP2 domains and EREBP genes with a single AP2/ERF (Ethylene Responsive Element

Binding Factor) domain. Most members of EREBP subfamily function in signal transduction pathways of biotic and environmental stress responses. Some transcription activators are involved in ABA signalling and draught stress responses such as *ABI4* (Finkelstein et al. 1998; Song et al. 2005), while others like DREB2 proteins are involved in ABA-insensitive draught stress signalling. Members of AP2 subfamily function as key developmental regulators in vegetative and reproductive organs including the floral homeotic gene *APETALA2* (Riechmann and Meyerowitz 1998; Shigyo et al. 2006). Considering pollen biology, *HAPLESS 5(HAP5)* encoding an AP2 domain protein affects pollen development. Its mutation leads to morphological defects of 25% pollen grains including aborted pollen (Johnson et al. 2004). In our screen, we observed morphological similarities to *hap5* phenotype including 25-30% aborted pollen in SAIL_405_F06 line (At1g25470), and 25-30% small pollen grains in SALK_102687 line (At3g11020). In addition, the later line produced 10% two-celled pollen and 5-10% eccentric MGU. Similarly, SALK_138253 (At5g53290) pollen also showed 20% eccentric MGU in heterozygous plants (Table 2).

R2R3-MYB family is one of the largest families of plant-specific TFs, thus predominantly involved in controlling plant-specific processes. The members control development, determination of cell fate and identity, hormonal signal transduction, cell division and phenylpropanoid metabolism. MYB proteins act as transcriptional activators as well as repressors and they often interact with members of bZIP family (Jin and Martin 1999; Stracke et al. 2001). *AtMYB125 (DUO1)* is involved in regulation of male gametophyte development by controlling cell cycle. *duo1* pollen fails to enter PMII as DUO1 is a specific regulator of genes required for G2 to M-phase transition (Durbarry et al. 2005). We identified two MYB factors with disturbed phenotypes: SAIL_557_G02 line (At2g39880; *AtMYB25*) producing 30% grains with deformed cell wall, 10% aborted pollen, 10% two-celled pollen and below 10% misarranged MGU; and heterozygous plants of SALK_039489 line (At2g32460; *AtMYB101*) producing 20% grains with misarranged MGU (Table 2). In a portion of mutant pollen grains a two-celled pollen phenotype was observed resembling that of *AtMYB125 (DUO1)*; however most other disturbances suggested that the factors were likely to play different roles in regulatory pathways.

In plants, bZIP factors were demonstrated to regulate organ and tissue differentiation, cell elongation, photomorphogenesis, energy metabolism, hormone and sugar signalling, flower maturation, seed development, and pathogen defence (Cluis et al. 2004; Weltmeier et al. 2009). According to their molecular organization, bZIP TFs were classified into ten groups, A to S (Jakoby et al. 2002). Group A bZIPs contain ABA-responsive elements (ABRE) and they are known to play a role in ABA or stress signalling to constitute an ABA-responsive complex and to induce drought, cold or high salinity gene expression (Guiltian et al. 1990; Jacoby et al. 2002; Shen et al. 1996). In rice, *OsbZIP72* (group A bZIP) is a positive regulator of ABA response and drought tolerance (Lu et al. 2009). Recently, group E bZIP TF *AtbZIP34* (At2g42380) was demonstrated to fulfil multiple roles in the pollen development; it controls cellular transport, lipid metabolism and cell wall patterning (Gibalová et al. 2009). Our observation confirmed the published results by analyses of two T-DNA insertion lines: SALK_018864 (At2g42380) plants produced 25-30% pollen contained cytoplasmic inclusions, 25-30% small pollen grains, 15% aborted pollen, 10% two-celled pollen, and SALK_128336 line (At2g42380) showed 20-25% linear MGU and 5% of two-celled pollen (Table 2). These data indicate an important role of bZIP transcription factors and their position in regulatory network of pollen development. Considering published and observed phenotypic defects, they are likely to be involved in cell cycle regulation, cytoskeletal arrangement, and cellular metabolism.

Tubby-like proteins and phospholipid scramblases (PLSCR) form a new superfamily of membrane tethered transcription factors with close evolutionary link. Both protein groups

are localized on the inner side of the cell membrane and are involved in enhancing receptor transactivation. Their localization appears to be reversible and controlled release from membranes enhances their nuclear localization and transcriptional activity (Boggon et al. 1999; Lai et al. 2004). Although no specific target gene has yet been identified for Tubby, *PLSCR1* is known to activate inositol 1,4,5-triphosphate receptor type 1 (IP3R1) and more recently, members of this family have been shown to have a role in cellular signalling (Bateman et al. 2008). Another Tubby-like protein, AtTLP9, was confirmed to interact with ASK1 (Arabidopsis Skp1-like 1) in yeast two-hybrid test. *ASK1*gene is essential for male meiosis; *ask1* mutation causes unequal segregation at anaphase I and abnormal chromosome separation (Yang et al. 1999). In yeast and humans, *SKP1* genes regulate the mitotic cell cycle too (Hoyt et al. 1997), so it is likely for them to play a similar role in *Arabidopsis*. The possible involvement of Tubby-like proteins in mitotic cell division can explain the observed 5-10% two-celled phenotype in SALK_092324 line (At1g53320), which also produced 20-25% pollen with misarranged MGU suggesting cytoskeleton dysfunction. Similarly, SAIL_618_F07 line (At1g47270) produced 5% two-celled pollen and 15% pollen with misarranged MGU and 5-10% aborted grains (Table 2).

JUMONJI domain proteins are generally involved in chromosomal remodelling and floral regulation (Noh et al. 2004, Lu et al. 2008). The role of JUMONJI in male gametophyte development is largely unknown but interestingly, *jhdm2a*, mutation in an animal counterpart in process of spermatogenesis, causes post-meiotic chromatin condensation defects. Jhdm2A directly binds to and controls the expression of *TRANSITION NUCLEAR PROTEIN 1 (TNP1)* and *PROTAMINE 1 (PRM1)* genes, whose products are required for packaging and condensation of sperm chromatin (Okada et al. 2007). Therefore, there may be some functional similarities in terms of cellular events during meiosis of microsporocytes or nuclei arrangement within the cell. It can explain the phenotypic defects of SALK_114109 (At1g63490) pollen with 15% eccentric nuclei. More severe phenotype was observed in the SAIL_513_D10 line (At2g34880): 35-40% pollen grains with inclusions implicated also other functional activities in addition to 10% misarranged MGU (Table 2). Besides functions in male gametophyte, *MEE27 (JMJ15)*, a member of JUMONJI family, is directly involved in female gametophyte development and causes the arrest at one-cell zygotic stage in *Arabidopsis* (Pagnussat et al. 2004).

Small G2-like family is defined by similarity to GOLDEN2 (G2) protein. Maize G2 protein is nuclear-localized (Hall et al. 1998) and forms both homodimers and heterodimers with ZmGLK1 (Rossini et al. 2001). In Arabidopsis, AtGLK1 and AtGLK2 proteins are expressed in partially overlapping domains in photosynthetic tissues and act as transcriptional regulators of chloroplast development. However, the putative role of G2-like family in male gametophyte development is unknown. Interestingly, SALK_140819 line (At2g20400) produced aberrant pollen with 95% misarranged MGU and 5% two-celled pollen and SAIL_888_A11 line (At2g20400) showed 25-30% misarranged MGU (Table 2). Despite the remarkable phenotype penetrance, it is impossible to speculate of any gametophytic regulatory function of the At2g20400 protein when there are no additional data available.

MADS-box genes encode large family of highly conserved transcription factors involved in diverse biological functions that are often key regulators in various vegetative and generative developmental processes (Alvarez-Buylla et al. 2000; Ferrario et al. 2006). In *Arabidopsis*, AtMIKC* is an over-represented class of MADS-box family and there is compelling evidence that AtMIKC* complexes control a MADS network during pollen development (Verelst et al. 2007). The authors revealed partial regulation network downstream of AtMIKC* genes and identified two non-MIKC MADS proteins (AGL18 and AGL29) that regulated a subset of AtMIKC* genes. In fact, agl65/66/104 triple mutant seemed to be largely unaffected except for pollen tube growth defects *in vitro* condition

providing not much information on pollen development itself (Verelst et al. 2007). In our screen, we identified At1g77980 *(AGL66)* whose mutation had impact in both tested lines, in SALK_072100: 25-30% misarranged MGU and 10-15% two-celled pollen, and SALK_072127: 15% misarranged MGU, 5% aborted pollen grains and 5% two-celled pollen (Table 2).

Compared to other kingdoms, CCAAT gene family encoding HAP proteins greatly expanded in the plant lineage. Thus, HAP proteins were likely to be recruited in plants to a wider range of functions than in animals and yeast, which contain just a single gene encoding each subunit (Edwards et al. 1998; Lee et al. 2003). Members of Arabidopsis CCAAT family participate in light response, photoperiodic flowering regulation, gene expression in response to sugar, and embryo development. HAP complex in plants enables gene activation by CCT domain proteins that play many important roles in environmental responses as suggested by Wenkel et al. (2006). In addition, AtHAP3 subunit was shown to interact with rice HAP3 and a MADS box transcription factor (Masiero et al. 2002). Yeast *hap* mutant showed pleiotropic phenotypes with a general reduction in cytochromes and reduced growth on nonfermentable carbohydrates suggesting that HAP complex controls expression of genes important for mitochondrial biogenesis (de Winde and Grivell 1993). Tested SALK_002235 line (At3g20910) produced 40% aborted pollen grains (Table 2) but we can only speculate if the HAP complex plays a similar role in pollen development.

The involvement of ARF factors in auxin signalling pathway and activating of early auxin-response genes is well documented (Wang et al. 2007; Weijers et al. 2005). The role of polar auxin flow in anther filaments was shown to be important for proper pollen mitosis during development of pollen grains (Cecchetti et al. 2008; Feng et al. 2006). Hirano et al. (2008) clearly demonstrated the gradual accumulation of auxin after PMI until pollen maturation as a result of higher expression of IAA synthetic genes during anther development. Subsequently IAA receptor genes were mostly expressed during the late stages of pollen development when various sets of *AUX/IAA* and *ARF* genes were expressed (Hirano et al. 2008). In our study we observed two-celled pollen phenotype in SAIL_170_F03 line (At2g28350) in 15-20% pollen indicating PMII defects similarly as in heterozygous plants of SAIL_451_F03 line (At1g77850) producing 5-10% two-celled pollen and 20% aborted pollen grains (Table 2). Detail experiments on excised root meristem of *Vicia faba* revealed similar effect of IAA that was able to markedly extend the transition from G1 to S phase and G2 to M phase of cell cycle and other studies on pollen further revealed that the meiosis process was normal while the mitosis at later stage was significantly affected (Feng et al. 2006; Polit et al. 2003).

Conclusion

In this study, we focused on the large-scale screen for transcription factor genes likely to be involved in the regulation of two pollen developmental gene expression programs. Thus we aimed to study genes active namely at early (UNM, BCP) and late (TCP, MPG) developmental stages. Because of the limited understanding of transcriptional regulatory networks in pollen biology, the reverse genetic approach employing wide-scale phenotype screens was chosen as a promising method for revealing putative transcriptional regulators. We present a survey of 73 lines harbouring T-DNA insertions in 48 transcription factor genes that were expressed during male gametophyte development. Numbers of insertion lines (28) were identified across different TF families producing pollen grains with remarkable structural disturbances counting ≥25%. On the whole, the global regulatory networks active in pollen development are largely unknown and the knowledge on roles of single transcription factors in pollen development is poor as well. Unravelling the complexity of gene expression

and description of the regulatory network represents an enormous challenge in contemporary pollen biology. Hence the presented results have a potential to serve as a baseline information resource for future research on the importance of the respective TF(s) in male gametophyte development.

Acknowledgements

This work was supported by the Grant Agency of the Academy of Sciences of the Czech Republic (grant KJB600380701), the Grant Agency of the Czech Republic (grant 522/09/0858) and the Ministry of Education, Youth and Sports of the Czech Republic (grants LC06004 and OC10054).

References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signalling. Plant Cell 15:63-78

Alonso JM, Stepanova AN, Weisse TJ, Kim CJ et al. (2003) Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. Science 301:653-657

Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF (2000) MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, root and trichomes. Plant J 24:457-466

Ariizumi T, Hatakeyama K, Hinata K, Sato S, Kato T, Tabata S, Toriyama K (2005) The *HKM* gene, which is identical to MS1 gene of *Arabidopsis thaliana*, is essential for primexine formation and exine pattern formation. Sex Plant Reprod 18:1-7

Bateman A, Finn RD, Sims PJ, Wiedmer T, Biegert A, Söding J (2008) Phospholipid scramblases and Tubby-like proteins belong to a new superfamily of membrane tethered transcription factors. Bioinformatics 2:159-162

Becker JD, Boavida LC, Carneiro J, Haury M, Feijo JA (2003) Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant Physiol 133:713-725

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000) Extensive duplication and reshuffling in *Arabidopsis thaliana* genome. Plant Cell 12:1093-1101

Boggon TJ, Shan WS, Santagata S, Myers SC, Shapiro L (1999) Implication of tubby proteins as transcription factors by a structure-based functional analysis. Science 2:2119-2125

Bonhomme S, Horlow C, Vezon D, de Laissardiere S, Guyon A, Ferault M, Marchand M, Bechtold N, Pelletier G (1998) T-DNA mediated disruption of essential gametophytic genes in Arabidopsis is unexpectedly rare and cannot be inferred from segregation distortion alone. Mol Gen Genet 260:444–452

Borg M, Brownfield L, Twell D (2009) Male gametophyte development: a molecular perspective. J Exp Bot 60:1465-1478

Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijo JA, Becker JD (2008) Comparative transcriptomics of Arabidopsis sperm cells. Plant Physiol 148: 1168-1181.

Bouché N, Bouchez D. Arabidopsis gene knockout: phenotypes wanted. Curr Op Plant Biol 4:111-117

Brownfield L, Hafidh S, Borg M, Sidorova A, Mori T, Twell D (2009) A plant germline-specific integrator of sperm specification and cell cycle progression. PLOS Genetics 5: e1000430

Canales C, Bhatt AM, Scott R, Dickinson H (2002) EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in Arabidopsis. Curr Biol 12:1718-1727

Cecchetti V, Altamura MM, Falasca G, Costantino P, Cardarelli M (2008) Auxin regulates Arabidopsis anther dehiscence, pollen maturation, and filament elongation. Plant Cell 20:1760-1774

Chen Y-CS, McCormick S (1996) Sidecar pollen, an Arabidopsis thaliana male gametophytic mutant with aberrant cell divisions during pollen development. Development 122: 3243–3253

Cluis CP, Mouchel CF, Hardtke CS (2004) The Arabidopsis transcription factor HY5 integrates light and hormone signaling pathways. Plant Journal 38:332-347

Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003) AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics 4:25

de Buck S, Jacobs A, van Montagu M, Depicker A (1999) The DNA sequences of T-DNA junctions suggest that complex T-DNA loci are formed by a recombination process resembling T-DNA integration. Plant J 20:295-304

de Folter S, Immink RGH, Kieffer M, Pařenicová L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, Davies B, Angenent GC (2005) Comprehensive interaction map of the Arabidopsis MADS box transcription factors. Plant Cell 17: 1424-1433

de Winde JH, Grivell LA (1993) Global regulation of mitochondrial biogenesis in *Saccharomyces cerevisiae.* Prog Nucleic Acid Res Mol Biol 46:51-91

Dinneny JR, Weigel D, Yanofsky MF (2006) NUBBIN and JAGGED define stamen and carpel shape in Arabidopsis. Development 133:1645-1655

Drews GN, Yadegari R (2002) Development and function of the angiosperm female gametophyte. Annu Rev Genet 36:99-124

Dupl'áková N, Reňák D, Hovanec P, Honysová B, Twell D, Honys D (2007) Arabidopsis Gene Family Profiler (aGFP)--user-oriented transcriptomic database with easy-to-use graphic interface. BMC Plant Biol 7:39

Durbarry A, Vizir I, Twell D (2005) Male germ line development in Arabidopsis duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. Plant Physiol 137: 297-307

Edwards D, Murray JAH, Smith AG (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis. Plant Physiol 117:1015-1022

Feldmann KA, Coury DA, Christianson ML (1997) Exceptional segregation of a selectable marker (KanR) in Arabidopsis identifies genes important for gametophytic growth and development. Genetics 47:1411-1422

Feng XL, Ni WM, Elge S, Mueller-Roeber B, Xu ZH, Xue HW (2006) Auxin flow in anther filaments is critical for pollen grain development through regulating pollen mitosis. Plant Mol Biol 61:215-226

Ferrario S, Shchennikova AV, Franken J, Immink RGH, Angenent GC (2006) Control of floral meristem determinacy in petunia by MADS-box transcription factors. Plant Physiol 140:890-898

Finkelstein RR, Wang ML, Lynch TJ, Rao S, Goodman HM (1998) The Arabidopsis abscisic acid response locus ABI4 encodes an APETALA2 domain protein. Plant Cell 10:1043-1054

Francis KE, Lam SY, Copenhaver GP (2006) Separation of Arabidopsis pollen tetrads is regulated by QUARTET1, a pectin methylesterase gene. Plant Physiol 142:1004-1013

Gibalová T, Reňák D, Matczuk K, Dupľáková N, Cháb D, Twell D, Honys D (2009) AtbZIP34 is required for Arabidopsis pollen wall patterning and the control of several metabolic pathways in developing pollen. Plant Mol Biol 70:581-601

Grini PE, Schnittger A, Schwarz H, Zimmermann I, Schwab B, Jurgens G, Hulskamp M (1999) Isolation of ethyl methanesulfonate-induced gametophytic mutants in Arabidopsis thaliana by a segregation distortion assay using the multimarker chromosome 1. Genetics 151:849–863

Grobei MA, Qeli E, Brunner E, Rehrauer H, Zhang R, Roschitzki B, Basler K, Ahrens CH, Grossniklaus U (2009) Deterministic protein inference for shotgun proteomics data provides new insights into Arabidopsis pollen development and function. Genome Res 19: 1786-1800

Grotewold E, Sainz MB, Tagliani L, Hernandez LM, Bowen B, Chandler VL (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. Proc Natl Acad Sci USA 97:13579-13584

Guiltinan MJ, Marcotte WR, Quatrano RS (1990) A plant leucine zipper protein that recognizes an abscisic response element. Science 250:267-271

Guo A, He K, Liu D, Bai S, Gu X, Wei L, Luo J (2005) DATF: a database of Arabidopsis transcription factors, Bioinformatics 2:2568-2569

Hall LN, Rossini L, Cribb L, Langdale JA (1998) GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf. Plant Cells 10:925-936

Hirano K, Aya K, Hobo T, Sakakibara H, Kojima M, Shim RA, Hasegawa Y, Ueguchi-Tanaka M, Matsuoka M (2008) Comprehensive transcriptome analysis of phytohormone biosynthesis and signaling genes in microspore/pollen and tapetum of rice. Plant Cell Physiol 49:1429-1450

Honys D, Reňák D, Feciková J, Jedelský PL, Nebesářová J, Dobrev P, Čapková V (2009) Cytoskeleton-associated large RNP complexes in tobacco male gametophyte (EPPs) are associated with ribosomes and are involved in protein synthesis, processing, and localization. J Proteom Res 8:2015-2031

Honys D, Reňák D, Twell D (2006) Male gametophyte development and function. In: Floriculture, Ornamental and Plant Biotechnology: Advances and Topical Issues I (eds). Global Science Books, UK, pp. 76-87

Honys D, Twell D (2003) Comparative analysis of the Arabidopsis pollen transcriptome. Plant Physiol 132:640-652

Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biol 5:R85

Howden R, Park SK, Moore JM, Orme J, Grossniklaus U, Twell D (1998) Selection of T-DNA-tagged male and female gametophytic mutants by segregation distortion in Arabidopsis. Genetics 14:621-631

Hoyt MA (1997) Eliminating all obstacles: Regulated proteolysis in the eukaryotic cell cycle. Cell 91:149-151.

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

Ito T, Shinozaki K (2002) The *MALE STERILITY1* gene of *Arabidopsis*, encoding a nuclear protein with a PHD-finger motif, is expressed in tapetal cells and required for pollen maturation. Plant Cell Physiol 3:1285-1292

Ito T, Wellmer F, Yu H, Das P, Ito N, Alves-Ferreira M, Riechmann JL, Mayerowitz EM (2004) The homeotic protein AGAMOUS control microsporogenesis by regulation of *SPOROCYTELESS*. Nature 430:356-360.

Jacoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann JJ, Kroj T, Parcy F (2002) bZIP transcription factors in *Arabidopsis*. Trends Plant Sci 7:106-111

Jin H, Martin C (1999) Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol 41:577-585

Johnson MA, von Besser K, Zhou Q, Smith E, Aux G, Patton D, Levin JZ, Preuss D (2004) Arabidopsis hapless mutations define essential gametophytic functions. Genetics 168:971-982

Kapoor S, Kobayashi A, Takatsuji H (2002) Silencing of the tapetum-specific zinc finger gene *TAZ1* causes premature degeneration of tapetum and pollen abortion in Petunia. Plant Cell 14:2353-2367

Kapoor S, Takatsuji H (2006) Silencing of ananther-specific zinc-finger gene, *MEZ1*, causes aberrant meiosis and pollen abortion in petunia. Plant Mol Biol 6:415-430

Kapushesky M, Kemmeren P, Culhane AC, Durinck S, Ihmels J, Krner C, Kull M, Torrente A, Sarkans U, Vilo J, Brazma A (2004) Expression Profiler: next generation-an online platform for analysis of microarray data. Nucleic Acids Res 32:W465-70 (Web Server issue)

Kononov ME, Bassuner B, Gelvin SB (1997) Integration of T-DNA binary vector "backbone" sequences into the tobacco genome: evidence for multiple complex patterns of integration. Plant J 11:945-957

Lago C, Clerici E, Dreni L, Horlow C, Caporali E, Colombo L, Kater MM (2005) The Arabidopsis TFIID factor AtTAF6 controls pollen tube growth. Develop Biol 285:91-100

Lai CP, Lee CL, Chen PH, Wu SH, Yang CC, Shaw JF (2004) Molecular analyses of the Arabidopsis TUBBY-like protein gene family. Plant Physiol 134:1586-1597

Lalanne E, Michaelidis C, Moore JM, Gagliano W, Johnson A, Patel R, Howden R, Vielle-Calzada J-P, Grossniklaus U, Twell D (2004) Analysis of transposon insertion mutants highlights the diversity of mechanisms underlying male progamic development in Arabidopsis. Genetics 167:1975-1986

Lalanne E, Twell D (2002) Genetic control of male germ unit organization in Arabidopsis. Plant Physiol 129:865-875

Laufs P, Autran D, Traas J (1999) A chromosomal paracentric inversion associated with T-DNA integration in *Arabidopsis*. Plant J 18:131-139

Lee HS, Fischer RL, Goldberg RB, Harada JJ (2003) Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. PNAS 100:2152-2156

Li C, Wong WH (2001a) Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. Proc Natl Acad Sci 98:31-36

Li C, Wong WH (2001b) Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol 2:research 0032.1-11

Lu F, Li B, Cui X, Liu C, Wang XJ, Cao X (2008) Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in Arabidopsis and rice. J Integr Plant Biol 50:886-896

Lu G, Gao C, Zheng X, Han B (2009) Identification of OsbZIP72 as a positive regulator of ABA response and drought tolerance in rice. Planta 229:605-615

Masiero S, Imbriano C, Ravasio F, Favaro R, Pelucchi N, Gorla MS, Mantovani R, Colombo L, Kater MM (2002) Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. J Biol Chem 277:26429-26435

Milla MAR, Townsend J, Chang IF, Cushman JC (2006) The Arabidopsis *AtDi19* gene family encodes a novel type of Cys2/His2 zinc-finger protein implicated in ABA-independent dehydration, high-salinity stress and light signaling pathways. Plant Mol Biol 61:13-30

Millar AA, Gubler F (2005) The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes redundantly facilitate anther development. Plant Cell 17:705-721

Nacry P, Camilleri C, Courtian B, Caboche M, Bouchez D (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis.* Genetics 149:641-650

Noh B, Lee SH, Kim HJ, Yi G, Shin EA, Lee M, Jung KJ, Doyle KJ, Amasino RM, Noh YS (2004) Divergent roles of a pair of homologous Jumonji/zinc-finger–class transcription factor proteins in the regulation of arabidopsis flowering time. Plant Cell 16:2601-2613

Okada Y, Scott G, Ray MK, Mishina Y, Zhang Y (2007) Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. Nature 450:119-123

Pagnussat GC, Yu HJ, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie LF, Ye D, Sundaresan V (2004) Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. Development 132:603-614

Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri RV, Grotewold E (2006) AGRIS and AtRegNet: a platform to link cis-regulatory elements and transcription factors into regulatory networks. Plant Physiol 140:818-829

Park SK, Howden R, Twell D (1998) The Arabidopsis thaliana gametophytic mutation gemini pollen 1 disrupts microspore polarity, division asymmetry and pollen cell fate. Development 125:3789-3799

Park SK, Rahman D, Oh SA, Twell D (2004) *Gemini pollen 2*, a male and female gametophytic cytokinesis defective mutation. Sex Plant Reprod 17:63-70

Pina C, Pinto F, Feijó JA, Becker JD (2005) Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol 138:744-756

Polit JT, Maszewski J, Kazmierczak A (2003) Effect of BAP and IAA on the expression of G1 and G2 control points and G1-S and G2-M transitions in root meristem cells of Vicia faba. Cell Biol Int 27:559-566

Preston J, Wheeler J, Heazlewood J, Li SF, Parish RW (2004) AtMYB32 is required for normal pollen development in *Arabidopsis thaliana.* Plant J 40:979-995

Qin Y, Leydon AR, Manziello A, Pandey R, Mount D, Denic S, Vasic B, Johnson MA, Palanivelu R (2009) Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. PLoS Genetics 5: e1000621.

Rhee SY, Somerville CR (2008) Tetrad pollen formation in quarted mutants of Arabidipsis thaliana is associated with persistence of pectic polysaccharides of pollen mother cell wall. Plant J 15:79-88

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105-2110

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biol Chem 379:633-646

Rossini L, Cribb L, Martin DJ, Langdale JA (2001) The Maize *Golden2* gene defines a novel class of transcriptional regulators in plants. Plant Cell 13:1231-1244

Rotman N, Durbarry A, Wardle A, Yang WC, Chaboud A, Faure JE, Berger F, Twell D (2005) A novel class of MYB factors controls sperm-cell formation in plants. Curr Biol 15 :244-248

Sakamoto H, Araki T, Meshi T, Iwabuchi M (2000) Expression of a subset of the *Arabidopsis* Cys_2/His_2-type zinc-finger protein gene family under water stress. Gene 248:23-32

Sakamoto H, Maruyama K, Sakuma Y, Meshi T, Iwabuchi M, Shinozaki K, Yamaguchi-Shinozaki K (2004) Arabidopsis Cys2His2-type zinc-finger proteins function as transcription repressors under drought, cold and high-salinity stress conditions. Plant Physiol 136:2734-2746

Schiefhaler U, Balasubramanian S, Sieber P, Chevalier D, Wisman E, Schneitz K (1999) Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organs development in *Arabidodpsis thaliana.* Proc Natl Acad Sci 96:11664-11669

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

Shen Q, Zhang P, Ho THD (1996) Modular nature of abscisic acid (ABA) response complexes: Composite promoter units that are necessary and sufficient for ABA induction of gene expression in barley. Plant Cell 8:1107-1119

Shigyo M, Hasebe M, Ito M (2006) Molecular evolution of the AP2 subfamily. Gene 366:256-265

Smyth DR, Bowman DL, Meyerowitz EM (1990) Early flower development in Arabidopsis. Plant Cell 2:755-767

Song CP, Agarwal M, Ohta M, Guo Y, Halfter U, Wang PC, Zhu JK (2005) Role of an Arabidopsis AP2/EREBP-type transcriptional repressor in abscisic acid drought stress responses. Plant Cell 17:2384-2396

Sorensen AM, Kröber S, Unte US, Huijser P, Dekker K, Saedler H (2003) The Arabidopsis *ABORTED MICROSPORES (AMS)* gene encodes a MYC class transcription factor. Plant J 33:413-423

Steiner-Lange S, Unte US, Eckstein L, Yang C, Wilson ZA, Schmelzer E, Dekker K, Saedler H (2003) Disruption of *Arabidopsis thaliana MYB26* results in male sterility due to non-dehiscent anthers. Plant J 34:519-528

Stracke R, Werber M, Weisshaar B (2001) The *R2R3-MYB* gene family in *Arabidopsis thaliana.* Curr Op Plant Biol 4:447-456

Takeda T, Amano K, Ohto MA, Nakamura K, Sato S, Kato T, Tabata S, Ueguchi C (2003) RNA interference of the *Arabidopsis* putative transcription factor *TCP16* gene results in abortion of early pollen development. Plant Mol Biol 61:165-177

Toledo-Ortiz G, Huq E, Quail PH (2003) The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell 15:1749-1770

Twell D, Oh S-A, Honys D (2006) Pollen development, a genetic and transcriptomic view. In: R. Malho (ed) The Pollen Tube: A Cellular and Molecular Perspective, Plant Cell Monographs Vol 3, Springer-Verlag, Berlin, pp. 15-45

Verelst W, Twell D, de Folter S, Immink R, Saedler H, Münster T (2007) MADS-complexes regulate transcriptome dynamics during pollen maturation. Genome Biol 8:R249

Veron AS, Kaufmann K, Bornberg-Bauer E (2007) Evidence of interaction network evolution by whole-genome duplication: a case study in MADS-box proteins. Mol Biol Evol 24:670-678

Vizcay-Barrena G, Wilson ZA (2006) Altered tapetal PCD and pollen wall development in the Arabidopsis *ms1* mutant. J Exp Bot 57:2709-2717

Wang D, Guo Y, Wu C, Yang G, Li Y, Zheng C (2008) Geneome-wide analysis of CCCH zinc finger family in Arabidopsis and rice. BMC Genomics 9:44

Wang D, Pei K, Fu Y, Sun Z, Li S, Liu H, Tang K, Han B, Tao Y (2007) Genome-wide analysis of the auxin response factors (ARF) gene family in rice (Oryza sativa). Gene 394:13-24

Wang Y, Zhang W-Z, Song L-F, Zou J-J, Su Z, Wu W-H (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in arabidopsis. Plant Physiol 148: 1201-1211.

Weijers D, Benkova E, Jäger KE, Schlereth A, Hamann T, Kientz M, Wilmoth JC, Reed JW, Jürgens G (2005) Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators. EMBO J 24:1874-1885

Weltmeier F, Rahmani F, Ehlert A, Dietrich K, Schutze K, Wang X, Chaban C, Hanson J, Teige M, Harter K, Vicente-Carbajosa J, Smeekens S, Droge-Laser S (2009) Expression pattern within the Arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. Plant Mol Biol 69:107-119

Wenkel S, Turck F, Singer K, Gissot L, Gourrierec JL, Samach A, Coupland G (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. Plant Cell 18:2971-2984

Wilson ZA, Marroll SM, Dawson J, Swarup R, Tighe PJ (2001) The Arabidopsis *MALE STERILITY (MS1*) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. Plant J 28:27-39

Yang C, Vizcay-Barrena G, Conner K, Wilson ZA (2007) *MALE STERILITY1* is required for tapetal development and pollen wall biosynthesis. Plant Cell 19:3530-3548

Yang M, Hu Y, Lodhi M, McCombie WR, Ma H (1999) The Arabidopsis SKP1-LIKE1 gene is essential for male meiosis and may control homologue separation. Proc Natl Acad Sci 96:11416-11421

Yang WC, Ye D, Xu J, Sundaresan V (1999) The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes Develop 13:2108-2117

Yang X, Makaroff C, Ma H (2003) The Arabidopsis *MALE MEIOCYTE DEATH1* gene encodes a PHD-finger protein that is required for male meiosis. Plant Cell 15:1281-1295

Zhang W, Sun Y, Timofejeva L, Chen C, Grossniklaus U, Ma H (2006) Regulation of *Arabidopsis* tapetum development and function by *DYSFUNCTIONAL TAPETUM(DYT1)* encoding a putative bHLH transcription factor. Development 133:3085-3095

Zhang ZB, Zhu J, Gao JF, Wang C, Li H, Li H, Zhang HQ, Zhang S, Wang DM, Wang QX, Huang H, Xia HJ, Yang ZN (2007) Transcription factor *AtMYB103* is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis. Plant J 52:528-538

Zhao DZ, Wang GF, Speal B, Ma H (2002) The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the Arabidopsis anther. Genes Dev 6:2021-2031

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

Table 1 The complete list of 49 early and late TF genes and corresponding 74 T-DNA insertion lines with marked position of insertion sites and sequence of the appropriate F and R PCR primers. A, homozygous lines confirmed by PCR; B, homozygous lines designated by T-DNA Express Database; C, heterozygous lines confirmed by PCR; D, lines without confirmed T-DNA insertion

AGI number	*insertion line*	*F primer*	*R primer*	*PCR product*	*insertion site*	*A*	*B*	*C*	*D*
Early TF genes									
At1g05290	SAIL_2_B09	ATCGGGCTGAAGCGTCAACA	ACAAGCTTGAGATCTCATATCTA	567	exon2				•
At1g16060	SALK_062861	GTACAAAGAGGGCCGTAGCA	CGCCATATGAAAAATGACCA	821	exon2			•	
At1g25470	SAIL_405_F06	GAAGAGAGAAAAGGCCAAACCT	TCGGGTCTAACGTAACAAGTGA	694	5'UTR			•	
At1g25540	SALK_059316	CTGCTGGATCTGCTGTTGTT	GCAAACTGGCCACCAACTAT	849	intron13			•	
At1g26590	SAIL_276_C10	CCAATGGAGATCTGCTGATGA	AGTGGGGATGAAATACTATCAG	527	exon1			•	
At1g26610	SAIL_271_A04	TTGGTGATTTTTGTGGAACTGAT	ATTGACAGAGACAAGCGAAATAA	507	promotor			•	
At1g44890	SAIL_423_H07	GTGGACAGTGACGGAGAAGT	TTTGTCTGTTTTTGCTACTATCAA	535	promotor			•	
At1g44890	SALK_026179	TTGGTTTTGGTTTCTAGACATGAG	ACCACAGATTGCAACAACCA	858	5'UTR			•	
At1g63490	SALK_014109	CTGAGGACTCCGTTTAGTTG	TTACGAGCTCGGAGTATGTT	586	intron24	•	•		
At1g77850	SAIL_451_F03	ACTAGCTAACGAGGGATAAGC	TTGCCGACGGCGGTGACAT	539	promotor			•	
At1g77850	SALK_062511	CTTTTTCGCCAAAACCTGAA	CACTGTTGCCCTCTCCTCTC	848	promoter			•	
At2g20400	SAIL_888_A11	TCTTTACGGAAATCTGTCATTTTT	CCACCGGTATGGCGGTTAAA	587	promotor			•	
At2g20400	SALK_140819	GGCTCTCGACATTGTTTGCT	GGCCTGAAGAAGACTGAAAGC	766	intron6	•			
At2g26940	SALK_112806	CTTCTGTGGAGATGAAAGAA	GAATGTGTGTTGCAAAGGTT	694	promoter				•
At2g28350	SAIL_170_F03	ACCCCAGTGTAGTACAAAAAAG	CGGACAAGACCATCGTAATAAA	557	promoter			•	
At2g28350	SALK_143232	TCTCTGGTTTGTTTGGTTCTTG	TAACGTGGAACGGAGAAACC	841	promoter		•	•	
At2g34820	SALK_118766	ACACTCAAAACCATTTCCAG	GAGAAATATTCGGGGTTCTG	594	intron1			•	
At2g34880	SAIL_513_D10	CACCTTCCAATGCTCTTACCA	CGGTTTATCGAGCTGTTCAGA	594	exon7			•	
At2g37950	SAIL_105_D03	TGGCTCACTCATTCTGTTTTG	GCTTGTATGGTTTTTGGGTTC	596	3'UTR			•	
At2g37950	SALK_041477	AATCTCGCAGGTTCTGTTT	GATCAAAGGCGACAAGTAAT	534	intron	•			
At2g39880	SAIL_557_G02	CTCCGGCGGACGAGAGATT	TGCTGGTTTTTTCTTTGCTAAATT	519	promotor			•	
At2g40670	SALK_142105	TCCCCATTGCTCTCTGTTTT	GGATATTGGTTCCAAGAAGCTG	878	exon1			•	
At3g11440	SAIL_808_B08	ACAAGTCCCTCGCAACTTAGT	AGCAGGAACAAGAAATGAGTATA	596	exon3				•
At3g11440	SALK_118611	CAACTGAAGACGGGATTTTA	GCAGCAAAATTTAGATCGTC	585	exon2		•	•	
At3g15740	SAIL_764_B09	CATGCGAACACGGCATTCGA	GATCGATGTCAAGCGGATCTA	560	5'UTR			•	
At3g15740	SALK_084050	GGGTTCATATAGCACAAAAC	CCGAAGAAGAAGATGATGAC	592	exon1	•			
At3g20910	SALK_002235	CAAGGCAGAGCTAGAGAAGA	CTGAAAGGGAAAAAGTCACA	605	exon5		•	•	
At3g23060	SALK_148143	ACCTAATGCATCACTCATGT	GCCAGGACTAGAAACATAACA	840	exon1				•
At4g04450	SALK_049063	TGCTGTCATGAGGCAACAAG	CAACGTTGCACCTGTTAGGTC	845	exon4				•
At4g26440	SAIL_1284_C01	AACTTGGGCTCAAGCTGAAAC	TTGGTGATTTTTGTGGAACTGAT	566	promotor	•			
At4g31060	SAIL_654_E02	AGCCGCAGCTTCCAGAAGG	GGCAGATCATGTGTGAACGAT	559	exon1			•	
At5g25830	SALK_052540	TCTCCCACCAATATCTCATC	ACGTTTAGTGCCGGAGTAT	693	3'UTR			•	
At5g41090	SAIL_1225_G03	CGGCGAATTAGTCGGTATCAA	GTCTCTCGAGCCACATAACC	538	exon2			•	
At5g41090	SALK_130880	TGGACTCTGTGTCATATCTT	AAGATATGACACAGAGTCCA	686	intron	•	•		
At5g53290	SAIL_659_D01	CTCCGGTGACGACGTTGTTA	GAGAAGGTCTCCACTTGCGA	522	exon1				•
At5g53290	SALK_138253	ACACCACGGAACTTCTTTAT	ACACGTTTTTCTCTCTCACC	600	exon1		•	•	
Late TF genes									
At1g35490	SAIL_687_B12	TAGAGATGGTATAGAATCTACAAA	AGACTCCACCATTGGTTTATTC	517	promotor				•
At1g47270	SAIL_618_F07	AGAAGGCGGCAAAATCCAGTC	ATTCGTACGGCACGAGATCCA	895	3'UTR			•	
At1g47270	SALK_056205	TCATCTCTGTATCATGGATGTTT	CGCCGTCGTGAGCTTCGAT	495	promotor				•
At1g53320	SALK_092324	TTGTTCCAAAGCAGTTTGCTC	TCCCTTCTTGCCTCAAATTGGT	906	intron		•	•	
At1g54830	SAIL_194_D02	GTCCGTACTTATCCCTACCC	GATTGGATAGCTTTGCTTTG	684	3'UTR			•	
At1g77980	SALK_072100	TGGCTACTGAGCTGAGGCATTC	CAACAACAACTTCTCATGGCA	846	exon			•	
At1g77980	SALK_072127	GACCTCGTGGTCTGCTTGTGC	TCCGACGTCGAGGTGAATCTT	920	exon			•	
At2g03060	SAIL_835_C11	GCAGTCAATTACCATCTTTCTC	GCTTGGAGAGGAAGATAAAGC	578	promotor				•
At2g13570	SALK_085886	ATTCGCAATGGGAAGAAACCG	CGAATTAATGTCGGCCGATTG	915	5'UTR		•	•	
At2g18280	SAIL_597_F08	GGACGTCAAAACAAATGTATGAA	CTTGCATCTAGCGTCTTTTTAC	554	promotor			•	
At2g32460	SALK_039489	GCACATCCCAAGTTGAAAACA	ACCAAAGACCGACCCATTCGT	881	exon		•	•	
At2g34440	SAIL_313_D03	GCTAAATCTGCACCCTTGTGA	CTCGTTCCTTCTCAGCTTCAA	535	exon			•	
At2g42380	SALK_018864	AAATTCGATGTGATCAAAGTGTTT	TGAGATTACTGTCCGACACACA	915	promotor			•	
At2g42380	SALK_128336	CCATAGTTCTTGTTTAATCTGTCTT	GCGGAAGTGTCAGTGTTATCGC	902	exon			•	
At3g04700	SAIL_285_E06	TCGTCAAATTTTAAGCAATGGAA	GGATCGTCCCCTGCCAAA	997	5'UTR			•	
At3g04700	SALK_027695	AAGGGCGCTTCAGACAATCTT	GAGCCACAATTTTCATCTCTTTT	577	intron			•	

At3g04700	SALK_073151	CTTCCCGAGGTCAGATCAGCA	CCCTGACTTCTCAAATCTCTTCA	944	exon			•	
At3g04700	SALK_143073	GCTCTCGCCTCTTCCTTCACC	GCCCTTCGTTCGTGAAACTTG	898	exon		•	•	
At3g10470	SALK_106077	CAAGTCGAGGGAATAAGTCGTCG	TTGATCAATCGGAAAAGATGAAA	910	promotor				•
At3g10470	SALK_112731	ACATCCGGGCAAGCATTAGGT	AGTTGGACCCGTCACGCTCTT	927	exon			•	
At3g11020	SAIL_1160_B02	CCCAACAACATAGGAGGATA	GACTCAAGAATCACAAAACC	837	promotor	•			
At3g11020	SALK_102687	AAGAAGGAGAGAAACCGAAA	CTGTTCCTGTTGCTGTTG	594	exon	•			
At3g11100	SAIL_1055_A04	TGTCGCCGTAGCGTCTTCG	CAACCTCATATTTACTGTCTTTC	570	5'UTR				•
At3g57390	SAIL_588_C06	AGGGATCAACTTTATTTGACTAT	CAGGGGTTGCTCTGTTTTCAT	864	exon/promotor			•	
At3g57390	SALK_083060	TGACGCCGAGGTTGCTCTTA	CCGTTAGTAAAACCCCGTAATT	563	5'UTR/intron				•
At4g05330	SALK_068135	TTGGAAATGAGAGGGGATTTGG	TGGAGAATTTAACACAACAATTT	872	promotor			•	
At4g05330	SALK_075976	GCGAATAAGTAACTTTCGGTTATGG	TGAGCACGTAGTTGTCATTTGCAG	946	5'UTR			•	
At4g05330	SALK_075978	GCGAATAAGTAACTTTCGGTTATGG	TGAGCACGTAGTTGTCATTTGCAG	946	5'UTR			•	
At4g05330	SALK_090677	CAAATGGGCCAATCACGAGAA	CAAGGAATTCAATGCCAAGAA	880	promotor			•	
At5g04390	SAIL_441_G10	ACGGGATAATGAGTGGAACAAA	GTTGTTTTCCGGGGTAGAAAC	581	5'UTR			•	
At5g16540	SALK_024253	GCCGATTCAATCATCCTCACG	CGTAAAGGAAGGCCAACCGAG	968	exon			•	
At5g54470	SAIL_603_D10	TTTTTATGGTGGATAATTTGTGGA	GCTTCTTAGCTATCTATCTATCCTG	911	promotor			•	
At5g54470	SALK_010704	TTGATTAAGGTTTCCGCGGCT	CAAAAATCGTTGTGATTTTCTTTT	929	promotor	•			
At5g54470	SALK_010712	TTGATTAAGGTTTCCGCGGCT	CAAAAATCGTTGTGATTTTCTTTT	929	promotor			•	
At5g54470	SALK_138034	CCTTGGCGATTTCAATAAAACGA	CGTGTCCGCAGAAGAAACGAG	882	exon			•	
At5g54680	SALK_004997	GAATTCACTAGGTTAATGCCCTGA	TAAATCCTCTCCTGAGGCCAA	936	intron		•		•
At5g54680	SALK_043690	TTTTGACCGTTTATGAAATGTTG	CCCAAATTTAAGCATAGAAACCA	936	5'UTR		•	•	

Table 2 Sumary of pollen phenotype screen for respective T-DNA lines harbouring insertions in early and late TF genes from different families. The numbers represent the percentual proportion of pollen phenotypic defects observed in each phenotype category

AGI number	*TF family*	*T-DNA line*	*abortion*	*inclusions*	*small grain*	*oval grain*	*deformed cell wall*	*one-celled*	*two-celled*	*eccentric MGU*	*separated MGU*	*linear MGU*	***SUM***	*analyzed pollen*	*line*	*insert verification*
Early TF genes																
At1g05290	C2C2-CO-like	SAIL_2_B09												no	no	no
At1g16060	AP2/EREBP	SALK_062861							5-10				**5-10**	HM	HZ	yes
At1g25470	AP2-EREBP	SAIL_405_F06	25-30						5	1-5	1-5		**35-40**	HM	HZ	yes
At1g25540	PFT1	SALK_059316	1-5						5-10				**10**	HZ	HZ	yes
At1g26590	C2H2	SAIL_276_C10	1-5							5	5		**10-15**	HM	HZ	yes
At1g26610	C2H2	SAIL_271_A04							10				**10**	HM	HZ	yes
At1g44890	unknow protein	SAIL_423_H07						5	20-25	1-5			**30**	HM	HZ	yes
At1g44890	unknow protein	SALK_026179						1-5	5-10				**10**	HM	HZ	yes
At1g63490	JUMONJI	SALK_114109	1-5						1-5	15			**20**	HM	HM	yes
At1g77850	ARF	SAIL_451_F03	20						5-10				**25-30**	HZ	HZ	yes
At1g77850	ARF	SALK_062511												HM	HZ	yes
At2g20400	G2-like	SAIL_888_A11				1-5				1-5	25		**30**	HM	HZ	yes
At2g20400	G2-like	SALK_140819							5	5	90		**100**	HM	HM	yes
At2g26940	C2H2	SALK_112806												no	no	no
At2g28350	ARF	SAIL_170_F03							15-20	1-5	1-5		**20-25**	HM	HZ	yes
At2g28350	ARF	SALK_143232	5						5-10				**10-15**	HM	HZ	yes
At2g34820	bHLH	SALK_118766												HZ	HZ	yes
At2g34880	JUMONJI	SAIL_513_D10	1-5	35-40						5	5		**50**	HM	HZ	yes
At2g37950	C3H	SAIL_105_D03	20						10	1-5	1-5		**35**	HM	HZ	yes
At2g37950	C3H	SALK_041477	5							5-10	10		**20-25**	HM	HM	yes
At2g39880	MYB	SAIL_557_G02	10				30		10	1-5	1-5		**55**	HM	HZ	yes
At2g40670	ARR	SALK_142105							10				**10**	HZ	HZ	yes
At3g11440	MYB	SAIL_808_B08												no	no	no
At3g11440	MYB	SALK_118611	1-5							1-5	1-5		**5-10**	HZ	HZ	yes
At3g15740	C3H	SAIL_764_B09	10										**10**	HZ	HZ	yes
At3g15740	C3H	SALK_084050				30	1-5			20	20		**70-75**	HM	HM	yes
At3g20910	CCAAT-HAP2	SALK_002235	40						1-5				**40-45**	HM	HZ	yes
At3g23060	C3H	SALK_148143												no	no	no
At4g04450	WRKY	SALK_049063												no	no	no
At4g26440	WRKY	SAIL_1284_C01							5-10	5			**10-15**	HM	HM	yes
At4g31060	AP2-EREBP	SAIL_654_E02	1-5						5	5-10			**15**	HM	HZ	yes
At5g25830	C2C2/GATA	SALK_052540	1-5								10		**10-15**	HM	HZ	yes
At5g41090	NAC	SAIL_1225_G03							5	5			**10**	HM	HZ	yes
At5g41090	NAC	SALK_130880			40				10				**50**	HM	HM	yes
At5g53290	AP2-EREBP	SAIL_659_D01												no	no	no
At5g53290	AP2/EREBP	SALK_138253	1-5							20			**20-25**	HZ	HZ	yes
Late TF genes																
At1g35490	bZIP	SAIL_687_B12												no	no	no
At1g47270	TUB	SAIL_618_F07	5-10						5	5-10	5-10		**25-30**	HM	HZ	yes
At1g47270	TUB	SALK_056205												no	no	no
At1g53320	TUB	SALK_092324							5-10	5	15-20		**30**	HM	HZ	yes
At1g54830	CCAAT-HAP5	SAIL_194_D02							1-5				**1-5**	HM	HZ	yes
At1g77980	MADS	SALK_072100							10-15	10-15	10	5	**40**	HZ	HZ	yes
At1g77980	MADS	SALK_072127	5						5			15	**25**	HZ	HZ	yes
At2g03060	MADS	SAIL_835_C11												no	no	no
At2g13570	CCAAT-HAP3	SALK_085886							5	1-5		1-5	**10**	HM	HZ	yes
At2g18280	TUB	SAIL_597_F08	1-5						5	5			**10-15**	HM	HZ	yes
At2g32460	MYB	SALK_039489	1-5					1-5		5-10		10-15	**25**	HZ	HZ	yes
At2g34440	MADS	SAIL_313_D03	1-5						10				**10-15**	HM	HZ	yes
At2g42380	bZIP	SALK_018864	15	25-30	25-30			1-5	10				**80-85**	HM	HZ	yes
At2g42380	bZIP	SALK_128336	1-5						5			20-25	**30**	HM	HZ	yes
At3g04700	unknow protein	SAIL_285_E06	1-5				1-5		5				**10**	HM	HZ	yes
At3g04700	unknow protein	SALK_027695								1-5			**1-5**	HM	HZ	yes
At3g04700	unknow protein	SALK_073151	5						5		10-15		**20-25**	HM	HZ	yes
At3g04700	unknow protein	SALK_143073	5-10						1-5				**10**	HM	HZ	yes

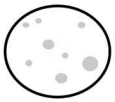
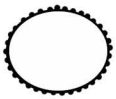
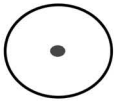
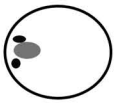
At3g10470	C2H2	SALK_106077									no	no	no
At3g10470	C2H2	SALK_112731				5	15-20	5-10	15-20	**45-50**	HM	HZ	yes
At3g11020	AP2-EREBP	SAIL_1160_B02	5			5				**10**	HM	HM	yes
At3g11020	AP2/EREBP	SALK_102687		25-30		10	5-10			**45**	HM	HM	yes
At3g11100	Trihelix	SAIL_1055_A04									no	no	no
At3g57390	MADS	SAIL_588_C06	5			10			15-20	**30-35**	HM	HZ	yes
At3g57390	MADS	SALK_083060									no	no	no
At4g05330	C2H2	SALK_068135	1-5			5-10	5	10-15	5	**30-35**	HZ	HZ	yes
At4g05330	C2H2	SALK_075976				10-15	5-10	10	5	**35**	HZ	HZ	yes
At4g05330	C2H2	SALK_075978	5			10-15	10	5-10		**35**	HZ	HZ	yes
At4g05330	C2H2	SALK_090677				10-15	10		30	**50-55**	HM	HZ	yes
At5g04390	C2H2	SAIL_441_G10			5			15		**20**	HM	HZ	yes
At5g16540	C2H2	SALK_024253	1-5					1-5		**5**	HZ	HZ	yes
At5g54470	C2C2-CO-like	SAIL_603_D10				15-20		5-10	15-20	**40-45**	HM	HZ	yes
At5g54470	C2C2-CO-like	SALK_010704	5	10-15		10-15	10-15	15-20	10-15	**70-75**	HM	HM	yes
At5g54470	C2C2-CO-like	SALK_010712	5			1-5			5-10	**15**	HZ	HZ	yes
At5g54470	C2C2-CO-like	SALK_138034				10-15	5-10		10-15	**30-35**	HZ	HZ	yes
At5g54680	bZIP	SALK_004997									no	no	no
At5g54680	bHLH	SALK_043690	10-15							**10-15**	HM	HZ	yes

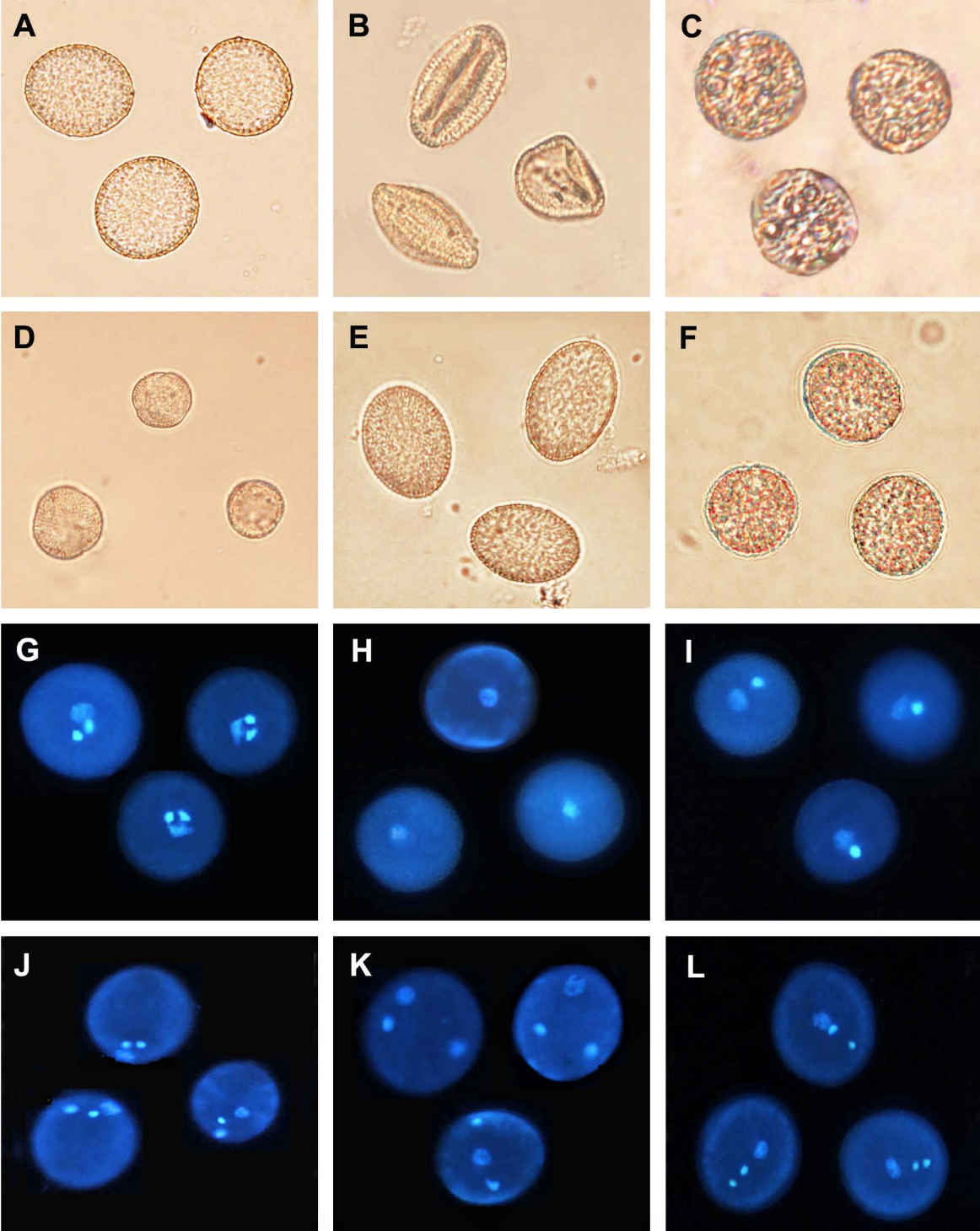
Figure legends

Fig. 1 Overview of ten mutant phenotype categories classified according to five functional disorders. A: cell abortion (1 abortion); B: presence of inclusions (2 inclusions); C: pollen grain size, cell wall structure (3 small grain, 4 oval grain, 5 deformed cell wall); D: cell cycle defects (6 one-celled pollen, 7 two-celled pollen); E: MGU organization (8 eccentric MGU, 9 separated MGU, 10 linear MGU)

Fig. 2 Portfolio of mutant phenotype categories and Arabidopsis pollen wild-type phenotype observed in bright field (a-f) and epifluorescent (g-l) microscopy. Phenotype categories: wild-type pollen (a, g); abortion (b); inclusions (c); small grain (d); oval grain (e); deformed cell wall (f); one-celled pollen (h); two-celled pollen (i); eccentric MGU (j); separated MGU (k) and linear MGU (l)

Supplementary table 1 Expression values and Detection calls of selected genesi n male gametophytic and sporophytic tissues. Sporophytic datasets were downloaded from the aGFP database (Duplakova et al., 2007). Gametophytic datasets were downloaded from public repositories and correspond to following publications: Honys (Honys and Twell, 2004), Pina (Pina et al., 2005), Zimmermann, (Zimmermann et al., 2004), Borges (Borges et al., 2008), Qin (Qin et al., 2009), Wang (Wang et al., 2008). Datasets are labeled as follows: UNM, uninucleate microspore; BCP, bicellular pollen; TCP, tricellular pollen; MPG, mature pollen grain; SPC, sperm cell; GP30, pollen germinated for 30 min; GP45, pollen germinated for 45 min; PT4, pollen tubes grown in vitro for 4 h; PT8, semi in vitro cultivated pollen tubes – 3 h in pistil, 5h in vitro; SL, seedlings; WP, whole plant; LF, leaves; PT, petioli; ST, stem; RT, roots; RH, root hair elongation zone; SU, suspension cell cultures

Functional disorder	Mutant phenotype		Description
A cell abortion	1 abortion		totaly colapsed pollen grain, no cytoplasm
B presence of inclusions	2 inclusions		number of inclusions in cytoplasm of unknown content
C pollen grain size, cell wall structure	3 small grain		small and round grain with a half diameter of wild-type pollen, normal MGU
	4 oval grain		distinctively oval grain
	5 deformed cell wall		deformed cell wall, rough cell surface
D cell cycle defects	6 one-celled pollen		only one nucleus with more condensation than regular vegetative nucleus
	7 two-celled pollen		one regular vegetative nucleus and one sperm cell-like nucleus, both in centre
E MGU organization	8 eccentric MGU		MGU shifted from the central position against cell wall, MGU itself may keep the same figure or sperm cells can be little separated from VN
	9 separated MGU		all nuclei separated from each other making a wide triangular figure of MGU
	10 linear MGU		all nuclei stand in linear arrangement with VN in the centre and sperm cells towards the cell wall


A
B
C
D
E
F
G
H
I
J
K
L

8.4. AtbZIP34 is required for Arabidopsis pollen wall patterning and the control of several metabolic pathways in developing pollen

Gibalová Antónia, Reňák David, Matczuk Katarzyna, **Dupľáková Nikoleta**, Cháb David, Twell David, Honys David

Plant Molecular Biology 2009, **70**:5

The article describes the functional characterization of AtbZIP34 transcription factor in Arabidopsis pollen. Several lines of evidence suggest its complex role in male reproductive development that involves a sporophytic role in exine patterning, and a sporophytic and/or gametophytic mode of action of AtbZIP34 in several metabolic pathways, namely regulation of lipid metabolism and/or cellular transport.

My contribution: I was involved in the phenotype evaluation of *atbzip34* pollen by bright field and fluorescent microscopy.

Supporting online material on CD

Plant Mol Biol (2009) 70:581–601
DOI 10.1007/s11103-009-9493-y

AtbZIP34 is required for Arabidopsis pollen wall patterning and the control of several metabolic pathways in developing pollen

Antónia Gibalová · David Reňák · Katarzyna Matczuk · Nikoleta Dupl'áková · David Cháb · David Twell · David Honys

Received: 21 January 2009 / Accepted: 15 April 2009 / Published online: 18 May 2009

Abstract Sexual plant reproduction depends on the production and differentiation of functional gametes by the haploid gametophyte generation. Currently, we have a limited understanding of the regulatory mechanisms that have evolved to specify the gametophytic developmental programs. To unravel such mechanisms, it is necessary to identify transcription factors (TF) that are part of such haploid regulatory networks. Here we focus on bZIP TFs that have critical roles in plants, animals and other kingdoms. We report the functional characterization of *Arabidopsis thaliana AtbZIP34* that is expressed in both gametophytic and surrounding sporophytic tissues during flower development. T-DNA insertion mutants in *AtbZIP34* show pollen morphological defects that result in reduced pollen germination efficiency and slower pollen tube growth both in vitro and in vivo. Light and fluorescence microscopy revealed misshapen and misplaced nuclei with large lipid inclusions in the cytoplasm of *atbzip34* pollen. Scanning and transmission electron microscopy revealed defects in exine shape and micropatterning and a reduced endomembrane system. Several lines of evidence, including the *AtbZIP34* expression pattern and the phenotypic defects observed, suggest a complex role in male reproductive development that involves a sporophytic role in exine patterning, and a sporophytic and/or gametophytic mode of action of AtbZIP34 in several metabolic pathways, namely regulation of lipid metabolism and/or cellular transport.

Keywords bZIP transcription factor · AtbZIP34 · Male gametophyte development · Lipid metabolism · Cellular transport · Cell wall formation · Transcriptomics

Electronic supplementary material The online version of this article (doi:10.1007/s11103-009-9493-y) contains supplementary material, which is available to authorized users.

A. Gibalová · D. Reňák · K. Matczuk · N. Dupl'áková · D. Honys (✉)
Laboratory of Pollen Biology, Institute of Experimental Botany ASCR, Rozvojová 263, 165 02 Praha 6, Czech Republic
e-mail: honys@ueb.cas.cz

A. Gibalová · D. Honys
Department of Plant Physiology, Faculty of Science, Charles University in Prague, Viničná 5, 128 44 Praha 2, Czech Republic

D. Reňák
Faculty of Biological Sciences, Department of Plant Physiology and Anatomy, University of South Bohemia, Branišovská 31, 370 05 Ceske Budejovice, Czech Republic

D. Cháb
Plant Reproduction Lab, Institute of Experimental Botany ASCR, Rozvojová 263, 165 02 Praha 6, Czech Republic

D. Twell
Department of Biology, University of Leicester, Leicester LE1 7RH, UK

Introduction

Male gametophyte development is a complex process requiring the coordinated participation of various cell and tissue types in the flower. The developmental phase leading to the formation of mature pollen grains is followed by the functional or progamic phase beginning with the impact of the grains on the stigma surface and terminated by double fertilization. Although its accessibility and highly reduced structure makes the male gametophyte an ideal model for developmental studies, we still have a limited knowledge of the regulatory mechanisms that specify gametophytic

development and function (McCormick 2004; Honys et al. 2006; Twell et al. 2006; Borg et al. 2009).

Previous genome-wide transcriptomic studies have led to the identification of 608 putative transcription factor (TF) genes active in at least one stage of male gametophyte development in *Arabidopsis* (Honys and Twell 2004). This represents ~45% of the 1,350 TF genes with corresponding probe sets on the Affymetrix ATH1 GeneChip. Of these 608 male gametophyte expressed TF transcripts, 54 (15.7%) were shown to be putatively pollen-specific. Several large protein families were overrepresented among male gametophyte expressed TFs, including C3H and C2H2 zinc finger proteins, WRKY, bZIP and TCP proteins. On the contrary, basic helix-loop-helix (bHLH) and APETALA2/ethylene response element binding protein-like (AP2/EREBP), MADS and R2R3-MYB gene families were underrepresented (Honys and Twell 2004). Interestingly, members of a few underrepresented gene families have been reported to function in male gametophyte and/or tapetum development (Ito et al. 2007; Takeda et al. 2006).

The most significant advances have originated from two studies of *Arabidopsis* MADS-box TFs (Verelst et al. 2007a, b). The authors identified pollen-specific MIKC* class of MADS-box proteins as major regulators of transcriptome dynamics during late stages of pollen development in *Arabidopsis*. MIKC* protein complexes were demonstrated to control a transcriptional switch directing pollen maturation that is essential for pollen competitive ability. The co-expression of five of the six AtMIKC* genes during late stages of pollen development suggests that they cooperate to establish a TF network active during the final stages of pollen development.

Several others TFs regulating male gametophyte development belong to the MYB family. An unusual R2R3 MYB gene, *DUO1* is specifically expressed in the male germline and has been shown to be a key regulator of germ cell division and sperm cell formation in *Arabidopsis* (Durbarry et al. 2005; Rotman et al. 2005; Brownfield et al. 2009). However, most of the known MYB-family factors act sporophytically. For example, knockout of two redundant genes, *MYB33* and *MYB65,* results in premeiotic abortion of pollen development (Millar and Gubler 2005). Moreover, expression of these genes is regulated at the post-transcriptional level by miRNAs miR159a and miR159b (Allen et al. 2007). Similarly, AtMYB103 is involved in the sporophytic control of microspore release and exine formation (Zhang et al. 2007) and forms part of regulatory network that acts downstream of another MYB gene, TDF1 (Zhu et al. 2008).

However, bZIP-family TFs have not yet been demonstrated to be directly involved in male gametophyte development. Compared to the largest TF gene families, the bZIP family is slightly smaller, consisting of 75 members in *Arabidopsis* (Jakoby et al. 2002), 89 or 92 in rice (Correa et al. 2008; Nijhawan et al. 2008) and 89 in *Populus trichocarpa* (Correa et al. 2008). The number of genes and distribution among subfamilies demonstrates the complexity and homogeneity of the bZIP gene family in angiosperms. Their chromosomal distribution and sequence similarities suggest that the bZIP TF family has diverged through multiple gene duplication events (Correa et al. 2008), contributing to their potential for regulating diverse gene networks. Putative AtbZIP proteins were clustered into 10 groups according to their domain structures and sequence similarities (Jakoby et al. 2002).

Basic leucine zipper (bZIP) proteins represent an exclusively eukaryotic class of enhancer-type TFs that are known to regulate many critical processes including histodifferentiation during embryogenesis (Darlington et al. 1998; Eferl et al. 1999; Wang et al. 1992). In adult animals, bZIP factors are involved in diverse processes such as metabolism, circadian rhythm, and learning and memory (Darlington et al. 1995, 1998; Sanyal et al. 2002; Yamaguchi et al. 2005). In yeast, bZIP proteins are necessary for sexual differentiation and entry into stationary phase (Takeda et al. 1995; Watanabe and Yamamoto 1996). In general, bZIP TFs appear to be mostly involved in regulatory processes of general metabolism and appear to act downstream in regulatory hierarchies.

In plants, bZIP factors have been shown to have important roles in organ and tissue differentiation, photomorphogenesis, cell elongation, nitrogen/carbon balance control, energy metabolism, hormone and sugar signalling, flower maturation, seed development and pathogen defence (Weltmeier et al. 2009, reviewed by Cluis et al. 2004; Correa et al. 2008; Jakoby et al. 2002). A group of bZIP TFs play important roles in the ABA signalling pathway in *Arabidopsis* and most ABA-responsive element-binding bZIPs belong to group A. Functional characterization of several group A bZIPs revealed that their expression is induced by ABA or abiotic stress (Choi et al. 2000; Finkelstein and Lynch 2000; Uno et al. 2000). Phylogenetic analysis showed that this group of bZIPs was evolutionarily conserved between *Arabidopsis* and rice. OsbZIP72, another member of group A, was recently shown to be a positive regulator of ABA response and drought tolerance in rice (Lu et al. 2008). Similarly another rice bZIP TF OsbZIP23 confers stress tolerance and ABA sensitivity (Xiang et al. 2008). Several bZIP TFs (*Arabidopsis thaliana* AtbZIP17, AtbZIP28, AtbZIP49 and AtbZIP60 with orthologues in *Nicotiana tabacum* NtbZIP60 and *N. benthamiana* NbbZIP60) were demonstrated to be membrane-bound in their cytoplasmic, inactive form (Iwata and Koizumi 2005; Liu et al. 2007a, b; Tajima et al. 2008; Tateda et al. 2008). These proteins are activated during the stress response by an intramembrane proteolysis

mechanism (RIP; reviewed by Seo et al. 2008) and have different sensitivities or responses to particular stimuli (reviewed in Chen et al. 2008). AtbZIP60 and AtbZIP28 are proteolysis-activated TFs directly involved in the endoplasmic reticulum stress response (Iwata et al. 2008; Liu et al. 2007a). The conserved presence of bZIP factors across all eukaryotic kingdoms, together with their roles in a myriad of cellular functions, underscores the importance of this class of enhancer-type TFs (Deppmann et al. 2006).

We carried out phenotypic screening of T-DNA insertion lines for candidate TFs potentially involved in regulation of male gametophyte development. A T-DNA insertion in *AtbZIP34*, encoded by At2g42380, resulted in obvious pollen morphological defects and was characterized further. Here we report the functional characterization of *AtbZIP34* and its expression in both gametophytic and surrounding sporophytic tissues during flower development. Our results demonstrate a role for AtbZIP34 in the sporophytic control of cell wall patterning and gametophytic control of pollen development. Transcriptomic analysis of *atbzip34* mutant pollen further identified altered patterns of gametophytic gene expression that highlight a role for AtbZIP34 in the control of pathways regulating cellular transport and lipid metabolism.

Materials and methods

Plant material and growth conditions

Arabidopsis T-DNA insertion line SALK 018864 (insertion in At2g42380 gene; kanamycin resistance) was used together with wild type ecotype Columbia-0 plants. Plants used for pollen isolation were grown in controlled-environment cabinets at 21°C under illumination of 150 µmol m^{-2} s^{-1} with a 16-h photoperiod. Pollen for microarray experiments was harvested from two independently grown populations according to Honys and Twell (2003). The purity of isolated pollen was determined by light microscopy and 4′,6-diamino-phenylindole-staining according to Park et al. (1998). Pollen viability was tested by fluorescence diacetate (FDA) staining according to Eady et al. (1995). In all tests, more than 1,000 grains were scored. Roots were grown from plants in liquid cultures as described previously (Honys and Twell 2003).

For genotyping of transgenic plants, gene-specific ZIP-F1, ZIP-R1 primers and insert-specific primer LB2 were used. Appropriate gene-specific primers were designed using Primer3 software (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) and are listed in Supplementary Table 1. Genomic DNA was isolated by a CTAB DNA extraction method modified from Weigel and Glazebrook (2002) in which initial grinding of leaf tissue frozen in liquid nitrogen was carried out ground for 12 s with glass beads in a dental amalgam mixer ESME Cap mix (3 M, Maplewood, MN). For segregation analyses, heterozygous plants *AtbZIP34/atbzip34* were allowed to self-fertilize and seeds were aseptically sown on ½ Murashige–Skoog media containing 10 µg $µl^{-1}$ kanamycin.

RNA extraction, probe preparation and DNA chip hybridization

Total RNA was extracted from 50 mg of isolated pollen using the RNeasy Plant Kit according to the manufacturer's instructions (Qiagen, Valencia, CA). For microarray experiments, RNA integrity was checked using an Agilent 2100 Bioanalyser (Agilent Technologies, Boblingen, Germany) at NASC. Biotinylated target RNA was prepared from 20 µg of total RNA as described in the Affymetrix GeneChip Technical Analysis Manual (Affymetrix, Santa Clara, CA). Preparation of cRNA probes and hybridization to ATH1 Genome Arrays and scanning were carried out as described (Honys and Twell 2003). Publicly available transcriptomic datasets were downloaded from arabidopsis GFP database (http://aGFP.ueb.cas.cz; Dupl'áková et al. 2007). All transcriptomics datasets were normalized using freely available dChip 1.3 software (http://www.dchip.org). The reliability and reproducibility of analyses was ensured by the use of duplicates or triplicates in each experiment, the normalization of all arrays to the median probe intensity level and the use of normalized CEL intensities of all arrays for the calculation of model-based gene-expression values based on the Perfect Match-only model (Li and Wong 2001a, b). As a reference, all four available mature pollen transcriptomic datasets were used and labelled MP1 (Ler, 2 repeats; Honys and Twell 2004), MP2 (Col-0, 3 repeats; Zimmermann et al. 2005), MP3 (Col-0, 2 repeats; Pina et al. 2005) and MP4 (2 repeats, wild type in this study). For each gene, the most deviant expression value was eliminated and the mean from the remaining three values was calculated (MPG).

RT-PCR

Pollen, stem, leaf and inflorescence RNA was isolated from Col-0 and/or *atbzip34/atbzip34* plants grown as described (Honys and Twell 2003). Pollen RNA used for RT-PCR analyses was obtained from plants that were grown independently from those used for microarray analysis. Samples of 1 µg total RNA were reverse transcribed in a 20-µL reaction using the ImProm-II Reverse Transcription System (Promega, Madison, WI) following the manufacturer's instructions. For PCR amplification, 1 µl of 50× diluted RT mix was used. The PCR reaction was carried out in 25 µL with 0.5 unit of *Taq* DNA polymerase (MBI

Fermentas, Vilnius, Lithuania), 1.2 mM $MgCl_2$, and 20 pmol of genotyping primers SALK_018864_F1 and SALK_018864_R1. The PCR program was as follows: 2 min at 95°C, 35 cycles of 15 s at 94°C, 15 s at the optimal annealing temperature 55°C, and 30 s at 72°C, followed by 10 min at 72°C.

The presence or absence of *AtbZIP34* transcripts in *atbzip34* mutant pollen was verified by RT-PCR of 5′ and 3′ end gene fragments separately. Wild type-pollen cDNA and genomic DNA was used as a control. PCR was performed with exon-localised primers: ZIP-F2, ZIP-R2 (exons 1, 2; upstream of insertion site), ZIP-F3, ZIP-R3 (exon 3–4; downstream of insertion site; Supplementary Table 1). The PCR program was as follows: 2 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at the annealing temperature (51°C), and 1 min at 72°C, followed by 10 min at 72°C.

qRT-PCR

Quantitative real-time PCR was carried out on a LightCycler 480, (Roche Applied Science, Mannheim, Germany) using LightCycler 480 SYBR Green I Master (according manufacturer′s instructions). The primers used were specific for genes encoding UDP-glucose epimerases (Supplementary Table 1) cDNA was produced with 1.8 μg of total RNA and 2 μl of 35 μM oligo-$(dT)_{23}$ in a 20 μl reaction. First strand cDNA was 20× diluted in a final volume of 10 μl with 500 nM of each of the HPLC purified primers. Reaction was performed in 96-well plastic plate (Roche, Mannheim, Germany). Real-time PCR data were collected on the light cycler with cycling conditions: 5 min of initial denaturation at 95°C, then 45 cycles of 10 s at 95°C, 10 s at 58°C, and 15 s at 72°C. PCR efficiencies were estimated from calibration curves generated from serial dilution of cDNAs. Real time PCR expression measurements are frequently normalized with the expression of reference gene. We used KAPP (kinase associated protein phosphatase, At5g19280) as a reference gene. The ratio of the relative amount of the target and reference gene was calculated as follows: E_R^{CpR}/E_T^{CpT} ($E_{T,}$ E_R: efficiency for target or reference gene qRT PCR assay; C_pT, C_pR: a crossing point for target or reference genes).

Promoter analysis

Developmental and tissue-specific expression profile of At2g42380 gene was evaluated using a promoter:eGFP:GUS construct. A 1,060 bp region upstream of *AtbZIP34* gene was PCR-amplified using pZIP-F and pZIP-R primers (Supplementary Table 1). An entry clone was prepared by cloning the promoter fragment into the pENTR2B vector (Invitrogen, Carlsbad, CA). From the entry clone, the *AtbZIP34* promoter fragment was further sub-cloned into the Gateway-destination vector pKGWFS7.0 (Karimi et al. 2005). Constructs were verified by restriction analysis and sequenced. *Arabidopsis* wt plants were transformed using the floral dip method (Clough and Bent 1998) and *Agrobacterium tumefaciens* strain GV3101. Transformants were selected on ½ MS medium-300 ml (0.66 g Murashige and Skoog basal medium, 3 g sucrose, 30 mg Myo-inositol, 150 mg MES (2-(*N*-morpholino)ethanesulfonic acid), 0.8% agar, pH 5.7 with KOH) containing 50 μg ml^{-1} kanamycin. Transformants were verified for T-DNA insertion by PCR. Flowers from T1 generation were collected to GUS buffer (0.1 M phosphate buffer, pH 7.0; 10 mM EDTA, pH 8.0; 0.1% triton X-100 supplemented with 1 mM X-glcA and 4 mM ferricyanide) After 48-h incubation at 37°C, samples were analyzed by bright field and fluorescence microscopy with Olympus DP50-CU microscope.

Complementation analysis

A 3,232 bp genomic fragment including the complete *AtbZIP34* gene and 720 bp of 5′ flanking DNA was PCR-amplified using ZIP-F and ZIP-R primers (Supplementary Table 1) and recombined into the pENTR2B vector (Invitrogen, Carlsbad, CA). This entry clone was further recombined into GATEWAY-compatible destination vector (VIB, Ghent, Belgium, Karimi et al. 2005). Constructs were verified by restriction analysis and sequenced. Homozygous *atbzip34* plants were transformed using the floral dip method (Clough and Bent 1998) and *Agrobacterium tumefaciens* strain GV3101. Transformants were selected on ½ MS medium-300 ml (0.66 g Murashige and Skoog basal medium, 3 g sucrose, 30 mg Myo-inositol, 150 mg MES, 0.8% agar, pH 5.7 with KOH) containing 50 μg ml^{-1} kanamycin. Transformants were verified for the presence of T-DNA by PCR with primers ZIP-F1 and ZIP-R1 (Supplementary Table 1). Phenotypic complementation was examined by bright field and fluorescence microscopy after DAPI staining as described (Park et al. 1998).

Electron microscopy

Freshly harvested material was fixed in a 2.5% (w/v) glutaraldehyde in a 0.1 M cacodylate buffer (pH 7.2) for 24 h at room temperature, washed with 4% glucose in 0.1 M PBS ($NaH_2PO_4 \times H_2O$, pH 7.0) for 15 min, post-fixed in 2% (w/v) osmium tetroxide in 0.1 M PBS buffer, washed with 4% glucose in 0.1 M PBS for 15 min, dehydrated through an ascending ethanol series (30–100% ethanol), and, via ethanol: acetone, to acetone. Samples were embedded in Poly/Bed® 812/Araldite 502 resins. Thin

sections (70 nm) were cut on a Reichert–Jung Ultracut E ultra-microtome and stained using uranyl acetate and lead citrate. Sections were analyzed and photographed using the JEM-1011 electron microscopes with Megaview III camera and analySIS 3.2 software (Soft Imaging System®).

For scanning electron microscopy, freshly harvested material was fixed in a 2.5% (w/v) glutaraldehyde in 0.1 M PBS for 24 h at room temperature, washed with 4% glucose in 0.1 M PBS for 15 min, dehydrated through an ascending ethanol series (30–100% ethanol), and, via ethanol:acetone, to acetone. Pollen samples for scanning electron microscopy was then critical point dried in CO_2, mounted on a stub, sputter coated with gold, and observed and photographed with a JEOL 6300 scanning microscope.

Analysis of in vitro pollen tube growth

Pollen was collected from just-open flowers of wild type and *atbzip34/atbzip34* plants. Pollen grains were germinated on a germination medium on microscope slide according to Boavida and McCormick (2007) with several modifications. Pollen germination medium (final volume 25 ml) was always prepared fresh from 0.5 M stock solutions of the main components (5 mM KCl, 0.01% H_3BO_3, 5 mM $CaCl_2$, 1 mM $MgSO_4$) using autoclaved water. Sucrose (10%) was added and dissolved and pH was then adjusted to 7.5 with NaOH. About 1.5% of low-melting agarose (Amresco, Solon, Ohio) was added and briefly heated in a microwave oven, just long enough for the agarose to melt. Glass slide was then filled with 500 µl melted germination media. Pollen from individual flowers was spread on the surface of germination pad by inverting the flower with the help of tweezers and gently bringing it onto agarose surface after its solidification. The whole flower could be used as a "brush" to spread pollen uniformly on the surface of the germination medium. Glass slides were immediately placed inside a moisture incubation chamber to avoid media dehydration and incubated for 10 h in the dark at 24°C. Samples were examined by bright field and fluorescence microscopy with an Olympus DP50-CU microscope.

Analysis of in vivo pollen tube growth

Flower buds from wild-type and *atbzip34* plants were emasculated and hand-pollinated on the following day. Wild type plants were pollinated with *atbzip34* pollen, and *atbzip34* plants by wt pollen. After 7 h, the styles were collected separately and fixed in ethanol/acetic acid (3:1) for 1 h at room temperature. After overnight softening in 8 M NaOH, the flowers were washed several times with distilled water and incubated with aniline blue solution (0.1% aniline blue in 0.1 M K_2HPO_4-KOH buffer, pH 11.0) for 3 h in the dark. The stained flowers were placed in a drop of 50% glycerol on a microscope slide and observed by epifluorescence microscopy.

Results

bZIP family TFs are widely expressed in *Arabidopsis*

AtbZIP genes form a large family of TFs comprising 75 annotated genes in *Arabidopsis* (Jakoby et al. 2002). Of these, 66 genes are represented on the Affymetrix ATH1 GeneChip and 24 showed reliable signals in the developing male gametophyte (Supplementary Table 2). Although most AtbZIP genes do not show strict or preferential expression according to transcriptomic data (Dupl'áková et al. 2007), At2g42380 encoding AtbZIP34 showed a pollen-enriched expression pattern suggesting its role in late male gametophyte development (Honys and Twell 2004). Further analyses of transcriptomic data including reproductive organs revealed that At2g42380 was active in the second and third whorls of stage 15 flowers (Smyth et al. 1990; Zimmermann et al. 2005). RT-PCR using RNA isolated from four stages of male gametophyte development, unicellular, bicellular, tricellular and mature pollen, and four sporophytic tissues revealed its cumulative expression and weak expression in whole flowers (Fig. 1A). This expression pattern suggested that AtbZIP34 represents a late pollen-enriched TF.

The expression pattern of *AtbZIP34* was further investigated in transgenic plants harbouring the *AtbZIP34* promoter fused to the eGFP:GUS reporter (Fig. 2). GUS assay confirmed previously investigated expression pattern by RT-PCR together with transcriptomic data that AtbZIP34 represents a late pollen-enriched TF. In stamens the GUS signal was first detectable throughout young anthers and later became concentrated in the tapetum (Fig. 2B). In young flower buds (stage 7–9), GUS signal was localized in anthers and pistils (Fig. 2A). In developed flowers (stage 14), GUS staining extended to whole anthers and filaments (Fig. 2D, G). In carpels, GUS staining was first detected in pistil vascular tissues and young female gametophytes before complete development of the integuments (Fig. 2C). After the developmental shift, the highest GUS activity was localized in funiculi connecting mature ovules with the placenta (Fig. 2L) and in papillar cells and adjacent stigmatic tissue (stage 14) (Fig. 2E, F). In ovules, GUS activity was detected only in earlier developmental stages (Fig. 2C). On the contrary, in the male gametophyte, GUS signal gradually accumulated from microspores to mature tricellular pollen grains (Fig. 2H–K). *AtbZIP34* promoter activity was also observed in vegetative organs and was always associated

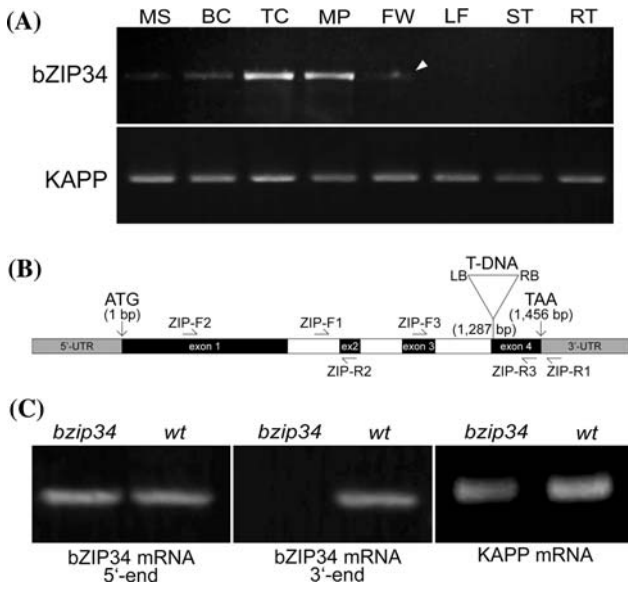


Fig. 1 Verification of At2g42380 expression profile and control KAPP gene expression by RT-PCR (**A**) in microspores (MS), bicellular (BC), tricellular (TC) and mature pollen (MP), whole flowers (FW), leaves (LF), stems (ST) and roots (RT). *White arrowhead* shows expression of At2g42380 in flowers. Diagram showing At2g42380 gene model (**B**) including T-DNA insertion site (*triangle*) and positions of respective primers—*arrows*, introns—*white boxes*, exons—*black boxes*, untranslated regions—*light grey boxes*, proximal promoter region—*dark grey box*, LB and RB—*left* and *right borders* of T-DNA. Expression analysis of both end regions of *AtbZIP34* transcript in wild type and *atbZIP34* pollen (**C**)—RT-PCR of *AtbZIP34* mRNA 5′-end (upstream of T-DNA insertion, primers ZIP-F2/ZIP-R2) and 3′-end regions (downstream of T-DNA insertion, primers ZIP-F3/ZIP-R3) as well as KAPP control transcript (primers KAPP-F/KAPP-R)

with vascular tissues in the distal regions of stems, leaves and siliques (data not shown).

Given its dynamic expression profile the regulatory function of AtbZIP34 TF is likely be complex. In this article, we focused our investigation on the role of AtbZIP34 in male gametophyte development and function.

Identification of an *AtbZIP34* T-DNA insertion mutant

We identified a T-DNA insertion in At2g42380 encoding AtbZIP34. In SALK_18864, the T-DNA insertion is located at the beginning of exon 4, after nucleotide 1,287 from the ATG initiation codon (Fig. 1B). The knock-down of *AtbZIP34* mRNA in pollen produced by homozygous SALK_18864 plants was verified by RT-PCR analysis of the transcripts upstream and downstream of the insertion site. The results confirm the absence of complete transcripts in *AtbZIP34* pollen using primer pair F1–R1, and partial transcripts downstream of the insertion site with primer pair F3–R3. However, 3′ truncated transcripts upstream of the insertion site were detected using primer pairs F2 and R2 (Fig. 1B), indicating that SALK_18864 represents a partial loss of function allele.

Cellular and pollen wall defects in *atbzip34* mutant pollen

The T-DNA insertion in the *AtbZIP34* gene is not lethal for gametophytic or sporophytic development as homozygous *atbzip34/atbzip34* plants were easily identified. However, pollen produced by homozygous plants showed characteristic phenotypic defects under bright field and fluorescence microscopy (Fig. 3). Five independent samples were observed ($n = 3{,}419$ pollen grains). In bright field observations, 56.2 ± 9.5% appeared similar to wild type pollen. After DAPI staining this percentage was lower (44.1 ± 5.5% of all pollen examined). The occurrence of collapsed pollen was 15.5 ± 3.9%. Despite the low percentage of unicellular microspores (2.8 ± 1.3%) and bicellular pollen (9.9 ± 2.3%), a fraction of tricellular pollen (26.7 ± 5.5%) contained malformed or displaced male germ units, often with unusual vegetative nuclei. These nuclei were larger and more diffuse than in wild type (Fig. 3B, D). Taken together, the majority of pollen grains exhibiting phenotypic abnormalities were tricellular, but these were smaller in diameter (*atbzip34*: $d = 12.05 \pm 1.54$ μm; $n = 30$; wt: $d = 15.87 \pm 0.66$ μm; $n = 30$) than wild type pollen. Moreover, *atbzip34* pollen contained characteristic cytoplasmic inclusions evoking lipid or oil bodies (Fig. 3) that were examined in more detail by electron microscopy.

Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) were employed to observe cell wall patterning, membrane structure and ultrastructure of developing *atbzip34* pollen. The most obvious differences from wild type pollen observed by SEM were irregular pollen shape and abnormal exine patterning (Fig. 4). Aberrant exine patterning appeared as regions of collapsed baculae and tecta together with areas with extra material deposited onto them. This phenotype was observed in all *atbzip34* pollen grains. Some pollen grains were found still attached together (Fig. 4E). There were no significant differences in the frequency of exine patterning defects in wild type pollen and pollen from heterozygous *atbzip34* plants (data not shown), consistent with the sporophytic control of exine patterning defects.

More thorough ultrastructural analysis was performed by TEM. Because of presumed sporophytic nature of cell wall patterning defects, the ultrastructure of both tapetum and spores was examined at several developmental stages (tetrads, uninucleate microspores, bicellular pollen; Supplemental Fig. 1). When observing tapetum development, apart from the general ultrastructure of tapetal cells (Ariizumi et al. 2004; Vizcay-Barrena and Wilson 2006; Yang et al. 2007), special attention was paid to the number

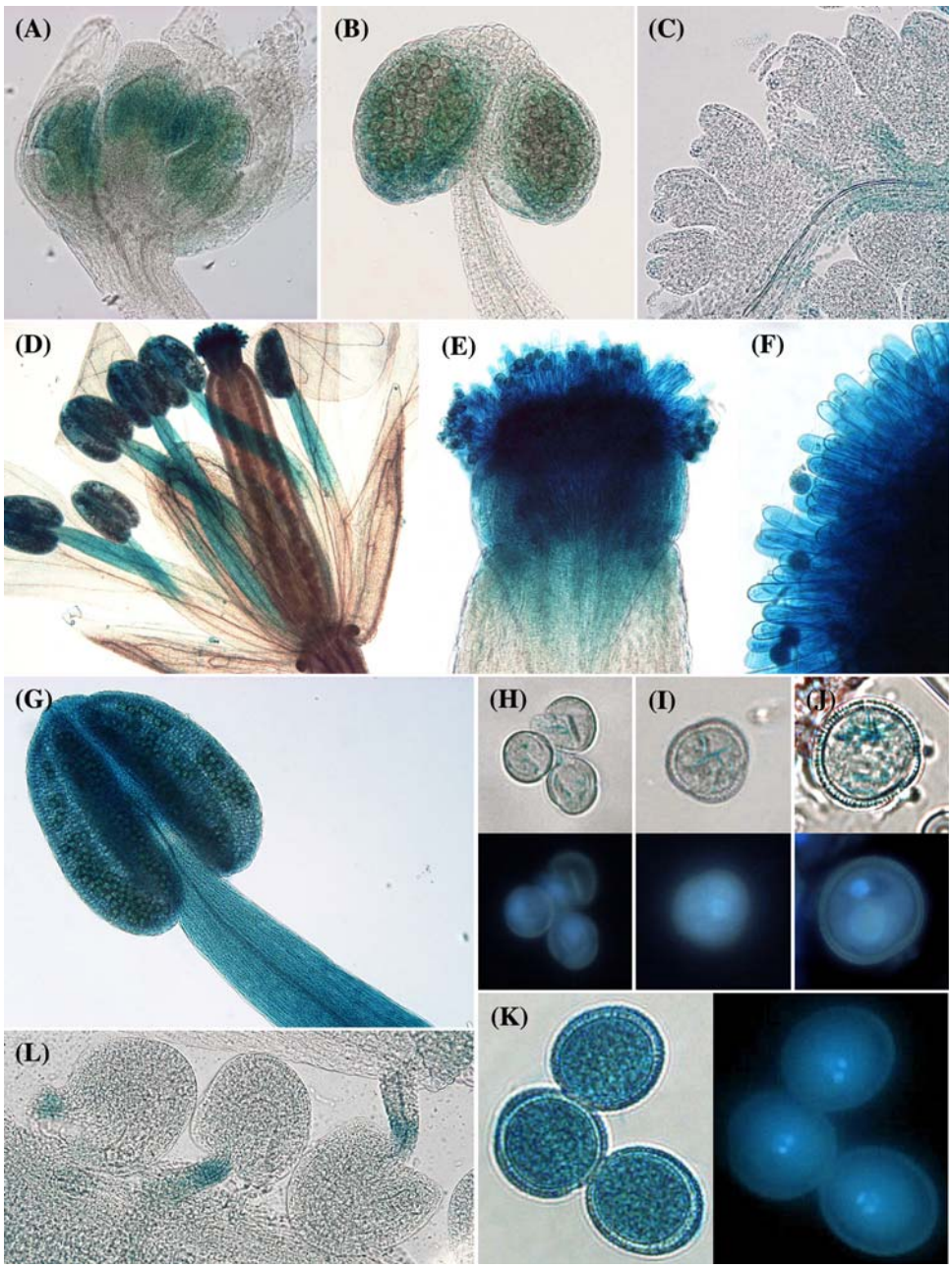


Fig. 2 Activity of the *AtbZIP34* promoter. Bright field microscopy of flower bud (stage 8, **A**) with detailed view of anther (**B**) and ovary (**C**). Later developmental stages are represented by an open flower (stage 15, **D**) with details of anther (**G**), pistil (**E**), papillar cells with attached pollen (**F**) and three ovules (**L**). Bright field and epifluorescent micrographs of several stages of pollen development including uninucleate microspores (**H**), early bicellular (**I**), late bicellular (**J**) and mature pollen (**K**)

and organization of secretory vesicles, vacuolization, plastid development (number and size of plastoglobules, lipid bodies, elaioplasts) and cell wall degeneration. In developing spores, cell wall structure and subcellular organization including endomembrane system and lipid bodies was analysed as phenotypic defects in these structures were suggested by bright field observations.

TEM observations confirmed differences in pollen wall structure between wt and *atbzip34* pollen (Fig. 5). Mature a*tbzip34* pollen possessed a characteristic wrinkled intine (Fig. 5D, F), which in wt is smooth and closely connected to the inner side of nexine (Fig. 5C, E). Mutant pollen also showed sparse and deformed baculae and tecta (Fig. 5D, F) that correspond with regions of unusual exine patterning observed by SEM. There were no apparent differences in cell wall structure of microspores in tetrads; the first differences were found in bicellular stage. Mutant pollen grains had wrinkled intine (and malformed exine characteristic of mature pollen grains) and vacuoles were increased in number and size. Unlike the exine-patterning defect, the unusual intine shape was observed also in approximately one half of pollen grains from *atbzip34* heterozygous plants. Moreover, there were differences in endomembrane systems together with the appearance of

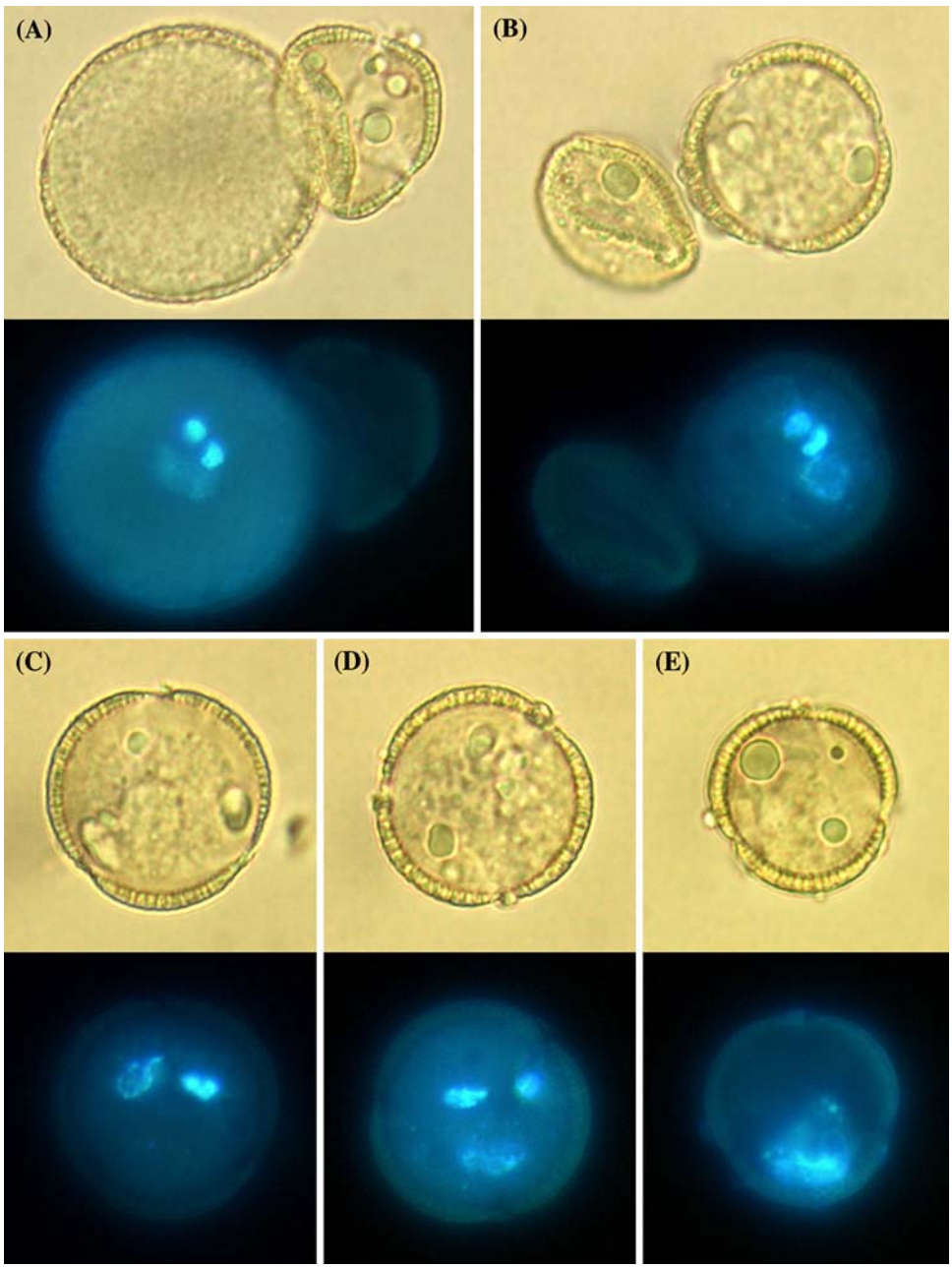


Fig. 3 Phenotypic defects in *atbzip34* pollen. Bright field and fluorescence images after DAPI-staining are shown, (**A**) wild type and *atbzip34* collapsed pollen, (**B**–**E**) *atbzip34* pollen

dense round inclusions (Fig. 5A, B). Generally, endoplasmic reticulum (ER) was underdeveloped in *atbzip34* pollen. Electron dense round inclusions were present both in wild type and *atbzip34* pollen. Their structure and co-localisation with rough ER correspond to lipid bodies that are formed during pollen maturation (Murphy 2001; Van Aelst et al. 1993). In wt pollen, lipid bodies were enclosed by one to mostly several layers of rough endoplasmic reticulum (Fig. 5E, G). However, more than one ER layer surrounding lipid bodies was rare in *atbzip34* pollen and often no surrounding ER was present (Fig. 5D, H). Lipid bodies were also more numerous in *atbzip34* pollen grains and localized in clusters in a cortical regions of the vegetative cell cytoplasm.

Tapetum development seemed less affected by *atbzip34* mutation. Tapetal cells of wild type and mutant were similar throughout development (Supplementary Fig. 1). The only apparent difference was the organization of round electron-dense inclusions at microspore stage. These structures were more numerous and clustered into larger groups (Supplementary Fig. 1h) clearly distinguishable from other structures found in tapetal cells, especially plastids. In heterozygotes, the ultrastructure of tapetal cells was unchanged (data not shown).

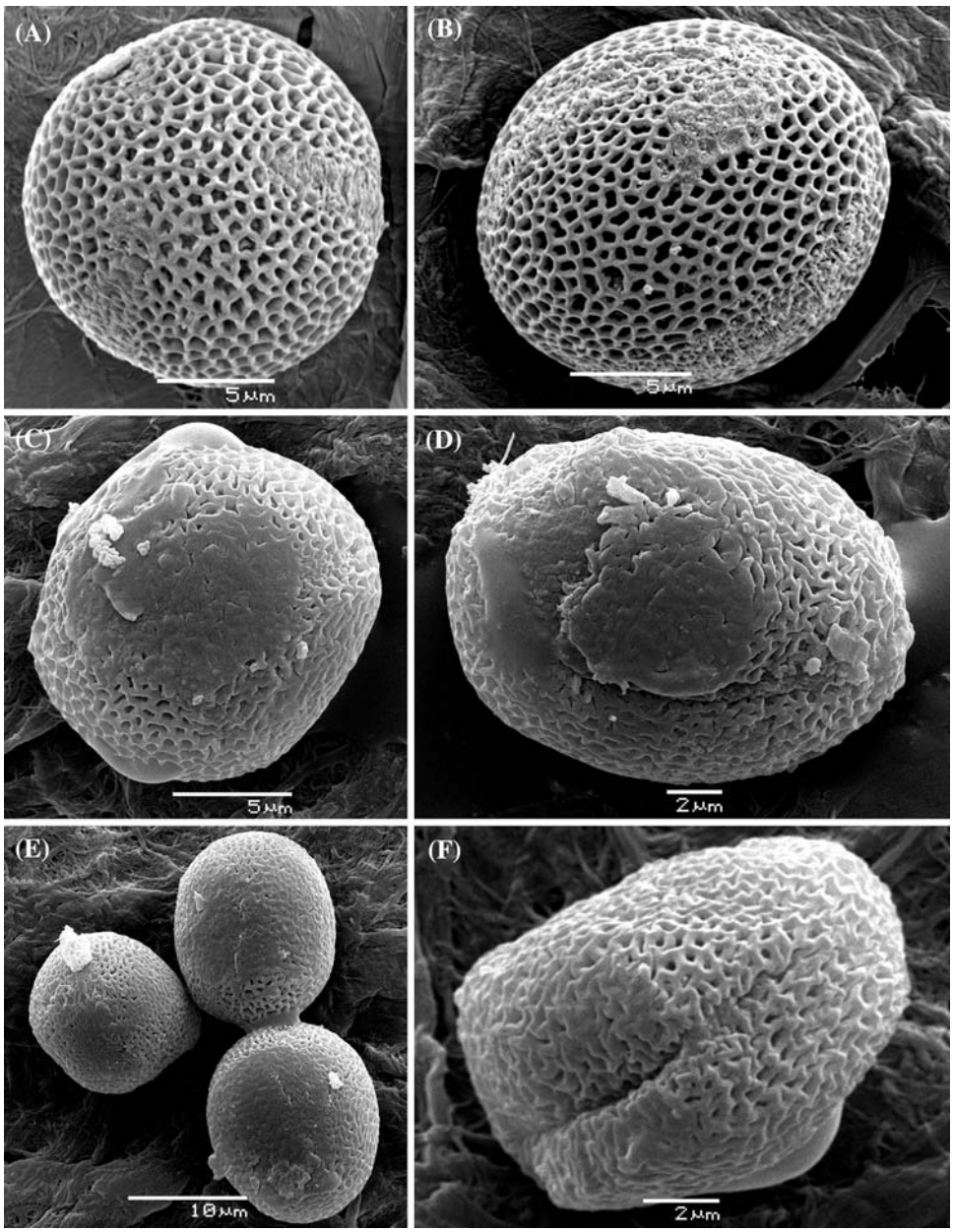


Fig. 4 Scanning electron micrographs of wild type pollen (**A**), *atbzip34* pollen complemented with At2g42380 genomic fragment (**B**) and *atbzip34* pollen grains (**C**–**F**). *atbzip34* pollen is defective in exine pattern formation (**C**–**F**) with often irregular shape (**F**). Pollen grains are frequently attached (**E**)

atbzip34 pollen shows reduced viability and progamic phase defects

Since AtbZIP34 affects early and late stages of pollen development, defects in the progamic phase were expected. First, pollen viability was calculated after FDA staining. In wild type plants over 90% of pollen was viable (91.8 ± 2.3%; $n = 412$). In *atbzip34* pollen population, this percentage was lower, 72.2 ± 4.3%; $n = 386$). To examine progamic phase defects we monitored pollen tube growth in vitro (Fig. 6A, B) and in vivo (Fig. 6C, D). Significant differences between *wt* and mutant pollen were observed in both assays. The in vitro germination rate of mutant *atbzip34* pollen was reduced by 85% compared to that of wild type pollen ($n = 300$). Moreover, mutant pollen tube growth rate was slower than that of wild type and after 10 h, mutant pollen tubes were ~53% shorter than wild type tubes ($n = 100$). In vivo pollen tubes growth tests confirmed slower growth rate of *atbzip34* mutant pollen tubes to the embryo sac when compared to wild type (Fig. 6C, D). However, resulting differences in length were less dramatic than observed in vitro. After 7 h postpollination, the longest *atbzip34* pollen tubes only reached the ninth ovule from the base of the ovary ($l = 1{,}438 \pm 53$ µm; $n = 5$ pistils), whereas wild type pollen tubes had reached the third ovule from the base ($l = 1{,}818 \pm 65$ µm; $n = 5$ pistils).

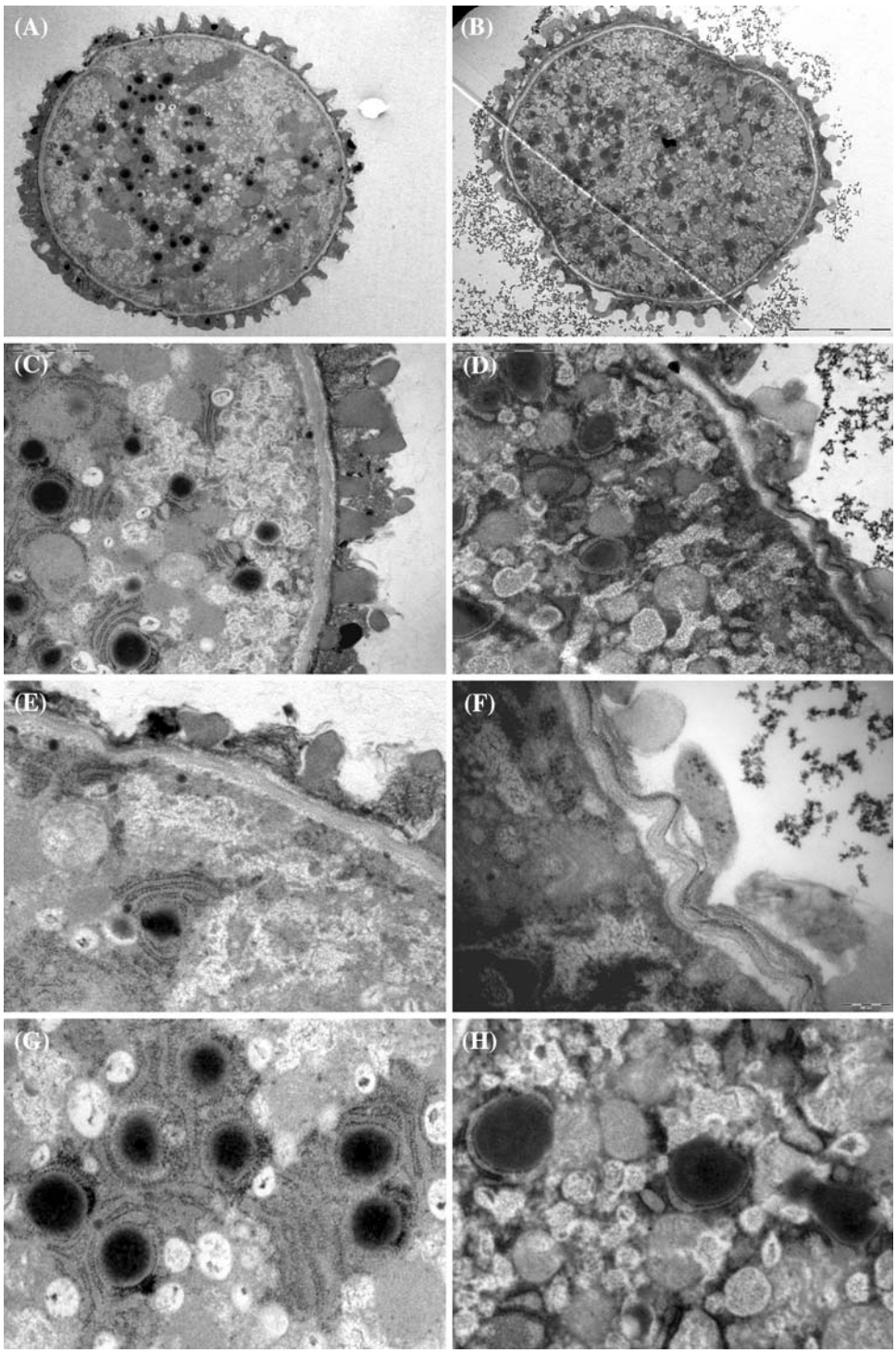


Fig. 5 Transmission electron micrographs of mature wild type (**A**, **C**, **E**, **G**) and *atbzip34* (**B**, **D**, **F**, **H**) pollen grains. *atbzip34* pollen has an irregular, wrinkled intine and exine with misplaced tecta and baculi (**D**, **F**). Mutant pollen has less developed endomembrane system and higher number of clustered lipid bodies that are surrounded by one or very rarely more layers of ER (**D**, **H**). In wild type, these lipid bodies are enclosed by several layers of ER (**E**, **G**)

atbzip34 shows gametophytic transmission defects

The transmission efficiency of mutant alleles and segregation ratio was examined. Heterozygous *atbzip34* plants were allowed to self-fertilize and seeds were sown onto kanamycin-containing plates. A non-Mendelian segregation ratio 1.87:1 (R:S) was observed among self progeny ($n = 448$) indicating reduced gametophytic transmission. Analysis of progeny from reciprocal crosses demonstrated that both gametophytes were affected. Through the male, *atbzip34* was transmitted with moderately reduced efficiency resulting in a distorted segregation ratio of 0.66:1 ($n = 186$). Through the female, the transmission of *atbzip34* was reduced further to 0.55:1 ($n = 219$). Thus

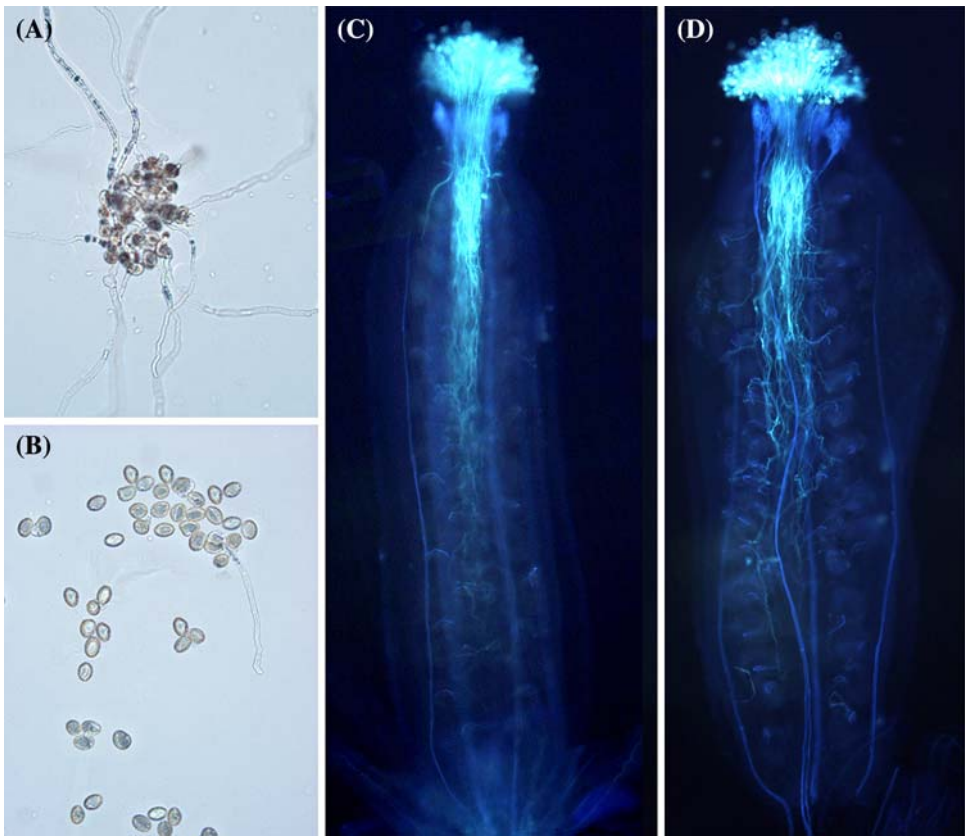


Fig. 6 Pollen tube growth tests. Wild type (**A**) and *atbzip34* (**B**) pollen tubes grown in vitro for 10 h. *atbzip34* pollen tubes were indistinguishable from wild-type pollen tubes, but there was markedly impaired germination. Wild type (**C**) and *atbzip34* (**D**) pollen tubes grown in wild type pistils. Tubes were observed 7 h after pollination

gametophytic transmission of *atbzip34* is reduced by 34% through the male and 45% through the female compared with the wild type *AtbZIP34* allele.

Complementation analysis was performed in which homozygous *atbzip34* plants were transformed with a vector containing a 3,232 bp *AtbZIP34* genomic fragment (pKGW:*AtbZIP34*). Pollen from 12 independent transformed lines was analysed by bright field and fluorescence microscopy after DAPI staining. Ten out of twelve pKGW:AtbZIP34 lines showed a reduced frequency of aberrant pollen. The percentage of normal pollen in *atbzip34* plants complemented with pKGW:*AtbZIP34* ranged between 95 and 99%, with only 1–5% of pollen exhibiting phenotypic defects characteristic of *atbzip34* pollen (Fig. 4B).

AtbZIP34 directly or indirectly affects the expression of genes involved in metabolic pathways

Some characteristics of *atbzip34* pollen analysed suggested impairment of certain metabolic pathways such as lipid metabolism and cellular transport during pollen maturation. To test this hypothesis, Affymetrix Arabidopsis ATH1 Genome Arrays were used to explore gene expression in *atbzip34* pollen in comparison with wt. Microarrays were hybridized with cRNA probes made from total RNA isolated from mature pollen of wild type and homozygous mutant *atbzip34* plants.

In *atbzip34* pollen, 813 genes were downregulated at least two-fold (Supplementary Table 3). Accordingly, another 1,379 genes were at least two-fold upregulated (Supplementary Table 4). Although nearly 70% more genes were upregulated in *atbzip34* pollen, the downregulated subset represented more distinguishable group for several reasons. First, downregulated transcripts were more highly expressed with average expression signal 863 compared to 225 in the upregulated set. Second, 760 of all downregulated genes (over 93%) had a relative expression signal over the threshold value of 100 compared to 937 genes (only 68%) in the upregulated subset. Third, the level of fold change was higher in the downregulated group reaching an average ratio of 2.71. On the contrary, the average upregulation was 2.42-fold. Finally, functional categorization of both subsets revealed specific composition of the group of downregulated genes (Fig. 7). Functional classes were defined as published previously (Honys and Twell 2004). Most significant changes were observed in these functional categories: protein synthesis (8.49% in downregulated subset to 2.94% in total pollen transcriptome; 2.89-fold change), transport (9.59 to 5.94; 1.61-fold), metabolism (19.43 to 13.05; 1.49-fold) and protein fate (5.04 to 3.92; 1.29-fold). However, the distribution of

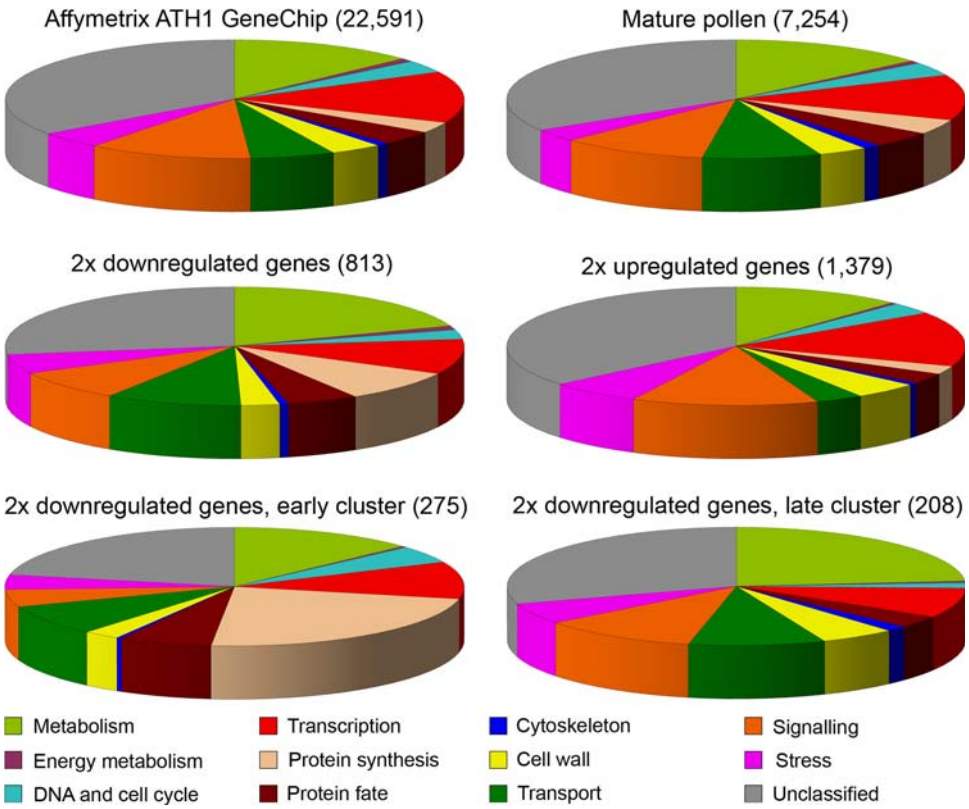


Fig. 7 Proportional representation of expressed mRNA among gene function categories. Data is presented for up and down regulated genes in *atbzip34* pollen in comparison with wt

genes among functional categories in the upregulated subset was very similar to that of the complete mature pollen transcriptome. The only exceptions were stress-related genes (6.53 to 4.59%) and those involved in cell wall metabolism (4.57 to 3.63%).

The weak activity of the *AtbZIP34* promoter was first detected in microspores and gradually increased until pollen maturity (Fig. 2H–K), so late pollen genes were more likely to be affected in its absence. This assumption was confirmed by cluster analysis of transcripts that were both two-fold down regulated, and upregulated according to their developmental expression profiles (Honys and Twell 2004). This led to the identification of three distinct groups within each geneset. These comprised genes with early, constitutive and late expression patterns. Among downregulated genes, 208 showed a late expression profile (25.6%), whereas within the upregulated set it was only 19 genes (1.4%). All affected late pollen genes are listed in Supplementary Tables 5 and 6. Gene ontology (GO) analysis of late downregulated genes revealed that several categories were affected more than others. Moreover, the GO profile of late genes is quite distinguishable from early genes (275 genes; Fig. 7E, F). All downregulated genes encoding ribosomal proteins (protein synthesis) were early. Similarly, most genes involved in cell cycle control comprised the early cluster. On the contrary, the most overrepresented categories in the late cluster were metabolism (23.6%), signalling (11.1%), transport (9.6%) and cell wall (5.3%).

The set of *AtbZIP34*-downstream genes shared several characteristic features. First, it was enriched with membrane-associated proteins as 49 out of 100 most highly downregulated genes in *atbzip34* pollen fell into this category. A fraction of these genes encoded various transporters including the ATP-binding cassette (ABC) transporter, AtABCB9 (At4g18050, 14.7X downregulated), lipid transfer proteins (At4g08670, 6.6X; At1g18280, 4.3X), mitochondrial import inner membrane translocase (At3g46560, 5.5X), lysine and histidine specific transporter (At1g67640, 5X), potassium transporter family protein (At4g19960, 4.57X), sugar transporter family protein (At4g16480, 4X), sucrose transporter (At1g71880, 3.8X), porin (At5g15090, 3.95X), cation/hydrogen exchanger (At3g17630, 3.7X), acyl carrier protein (At3g05020, 3.7X). These proteins were involved in transport of ions and various metabolites. The importance of membrane-associated transporters for male gametophyte development was already demonstrated (Bock et al. 2006; Sze et al. 2004). Moreover, there were two lipid transfer proteins and ABC transporter AtABCB9 is also likely involved in lipid transport (Martinoia et al. 2002; Verrier et al. 2008) and all

three genes were amongst those most downregulated in *atbzip34* pollen.

The complete Sec61 translocon complex was downregulated in *atbzip34* pollen. Translocons are sites of cotranslational protein translocation through ER membrane to its luminal compartment. They consist of core heterotrimeric Sec61 complexes (Sec61$\alpha\beta\gamma$) and associated proteins forming a cylindrical channel aligning with ribosomal large subunit during translocation (Beckmann et al. 1997). Selectivity of translocon function is facilitated by gating protein, luminal HSP70 chaperone BIP1, that seals its luminal side (Alder et al. 2005). In *Arabidopsis* each Sec61 subunit is encoded by three genes, whereas BIP1 by a single gene. All ten genes are expressed in male gametophyte and all but one are downregulated in *atbzip34* pollen (Fig. 8). For all subunits, the most abundant genes showed the highest level of downregulation.

Another set of proteins overrepresented among *atbzip34* pollen-downregulated genes included those involved in several steps of lipid catabolism: aspartate aminotransferase (At2g30970, 5.09X), family II extracellular lipase (At5g42170, 4.77X), malate dehydrogenase (At3g15020, 4.17X) (Kindl 1993; Pracharoenwattana et al. 2007; Teller et al. 1990; Zhou et al. 1995). All these genes were abundantly expressed in wild type pollen and significantly downregulated in *atbzip34* pollen.

We investigated potential metabolic pathways that may be controlled by AtbZIP34 factor using the MapMan visualization tool (http://gabi.rzpd.de/projects/MapMan//; Thimm et al. 2004). Most down- or up-regulated genes were scattered amongst various pathways. However, several metabolic pathways contained overrepresented down- or up-regulated genes with absolute value of ln(wt/*atbzip34*) over 1. These were transporters (Supplementary Fig. 2a) and genes involved in stress responses and development (Supplementary Fig. 2c). However, most genes with altered expression were associated with transport and cell wall-related pathways (Supplementary Fig. 2b). Although most visualized pathways contained both down- and up-regulated genes, genes involved in transport and metabolism of cell wall precursors were predominantly downregulated in *atbzip34* pollen.

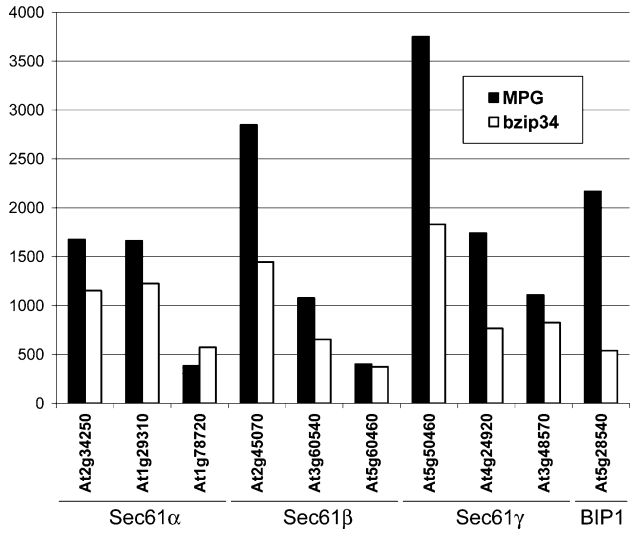


Fig. 8 Relative expression levels of nine genes encoding all Sec61 translocon subunits and gating protein AtBIP1 in wild type and *atbzip34* pollen

To verify microarray data visualised by Map Man tool, we selected several genes for quantitative RT-PCR tests. We selected the whole cluster formed by four genes from the metabolic pathway leading to the cell wall precursors, which were predominantly downregulated in *atbzip34* pollen (Supplementary Fig. 2b). This cluster was selected because it comprised two highly downregulated genes and two genes showing little or no change in expression. The selected genes encoded orthologues of UDP-glucose epimerases that are linked to the interconversion of sugar nucleotides UDP-glucose and UDP-galactose via UDP-4-hexo ketose intermediate. According to microarray data, two genes (At1g12780 and At1g63180) were downregulated and the remaining two (At4g23920 and At1g64440) showed no change (Supplementary Table 7). The expression of all four genes was verified by quantitative RT-PCR and related to the expression of the KAPP control gene (At5g19280). Expression profiles of genes putatively downregulated in bZIP microarray experiments were verified by qRT-PCR (Supplementary Table 7). Two genes, At1g12780 and At1g63180, downregulated on the microarrays were shown to be considerably downregulated also by RT-PCR. Expression of the third gene, At1g64440 was unchanged. The only exception was the At4g23920 gene that was expected to be unchanged as suggested by transcriptomic results. By quantitative RT-PCR, it was shown to be 4.4-fold downregulated in *atbzip34* pollen. This can be explained by the low expression signal (mean 135 in mature pollen, compared to 1,586 for At1g12780) and its low reliability (detection call P in only four out of nine mature pollen datasets). This low expression signal was confirmed by quantitative RT-PCR.

In addition to the above mentioned protein classes, we looked for TFs with changes in expression pattern in *atbzip34* mutant pollen. Putative TFs were defined according to MapMan. In general, TFs followed the above described scheme. Most genes were upregulated in *atbzip34* pollen (22 genes with ln(*atbzip34*/MPG) >1, Supplementary Fig. 3). However, both the relative expression signal and the ln-change were low, and only five upregulated genes had reliable detection calls in both biological replicates. On the contrary, twelve genes downregulated in *atbzip34* pollen with ln(MPG/*atbzip34*) >1 were identified. Again,

downregulated genes were more abundant; the mean relative expression signal of 22 upregulated TF genes was 166, whereas that of 12 downregulated ones was over 530.

Although downregulated TFs belonged among several gene families, there was no apparent cluster since in each family no more than one member was affected (Table 1). These families were as follows; AP2, ARR, bZIP, several zinc finger families (C2H2, C2C2-CO-like, C2C2-GATA), GeBP, MYB, NAC and general or unspecified factors. The expression of individual TF genes showed significant variation. MYB97 was the most abundant with a mean signal over 1,500. All the others had signals under 1,000 and only three reached values over 500. Moreover, MYB97 was the only downregulated TF with a late expression profile. According to microarray data, At4g26930 encoding MYB97, is also male gametophyte-specific, thus MYB97 may be controlled by AtbZIP34. The distribution of upregulated TFs was wider; they belonged to families AP2, APRR, bZIP, several zinc finger families (C2H2, C2C2-CO-like, C2C2-Dof, C3H), GeBP, MADS-box, MYB, NAC, WRKY and several general or unspecified factors. Most of these do not show reliable expression throughout male gametophyte development. Moreover, none of the reliably expressed genes are male gametophyte-specific nor have late expression profiles. The data presented seem to confirm previously published results (Jakoby et al. 2002) indicating that bZIP family TFs do not act as master regulators of TFs networks, but mainly act to regulate the expression of metabolic and structural genes.

Discussion

We have functionally characterized the role of TF AtbZIP34 in male gametophyte development. Characteristic phenotypic and genetic transmission defects provide several lines of evidence that support sporophytic and gametophytic roles for AtbZIP34 in male gametophyte development and function.

The T-DNA insertion line, SALK_18864 harbours an insertion in exon four of *AtbZIP34* gene, but *atbzip34* pollen express 3′ truncated transcripts (Fig. 1). Therefore, the corresponding truncated protein may be expressed in *atbzip34* pollen. The missing 3′-region encodes the bZIP dimerization domain and the truncated polypeptide would lack the dimerization potential of wild type protein. This T-DNA insertion was sufficient to cause transmission and phenotypic defects. However, the non-lethal nature of the mutation in *AtbZIP34* may also stem from redundancy among bZIP TFs co-expressed in pollen.

bZIP TFs possess a basic DNA binding domain adjacent to a leucine zipper region and act as homo- or heterodimers. In general, bZIP proteins do not heterodimerize promiscuously; specific interactions are preferred (Newman and Keating 2003). In *Arabidopsis*, GBF1-3 belonging to the G-group can form both homo- and heterodimers (Schindler et al. 1992), other G-group bZIP TFs AtbZIP16 and AtbZIP68, could form homodimers and heterodimers with other G-group members (Shen et al. 2008). S-group bZIP TFs can heterodimerize with C-group bZIPs (Ehlert et al. 2006; Weltmeier et al. 2009). AtbZIP43 (a member of S group bZIP TFs) can form heterodimers with members of E group (Shen et al. 2007). The basic region is relatively similar between members of groups E and I. E-group bZIPs (bZIP34 and bZIP61) and I-group bZIP51 were already shown to heterodimerize (Shen et al. 2007). Moreover, bZIP34 and bZIP61 could not form homodimers because they have a proline residue in the third heptad of their basic region distorting its α-helix structure (Shen et al. 2007). Comparative analyses of dimerization domains suggested that the most likely interactors of subfamily E AtbZIP proteins belong to subfamily I (Shen et al. 2007).

atbzip34 pollen showed characteristic phenotypic defects affecting cell wall as well as pollen ultrastructural organization (Figs. 3, 4, 5). There were misshaped and misplaced nuclei, inclusions in the cytoplasm and, most significantly, defects in cell wall patterning and endomembrane systems. Severe pollen surface defects were observed with scanning electron microscopy. These data were confirmed using transmission electron microscopy demonstrating that exine patterning is affected in *atbzip34* mutant pollen (Fig. 5F). A number of mutants have been characterized in *Arabidopsis* that show defects in exine structure and sporopollenin deposition that often lead to pollen abortion and male sterility (*dex1* (Paxson-Sowders et al. 2001), *ms2* (Aarts et al. 1997), *nef1* (Ariizumi et al. 2004), *tde1* (Ariizumi et al. 2008), *rpg1* (Guan et al. 2008)). Some of these pollen wall mutants affect callose accumulation (*calS5;* Dong et al. 2005; Nishikawa et al. 2005), or wax biosynthesis (*flp1/cer3–7;* Ariizumi et al. 2003). Mutant *atbzip34* pollen is distinguished by the characteristic wrinkled nexine and rare and deformed baculae and tecta, but does not lead to high levels of pollen abortion or male sterility. Interestingly, our transcriptomic analyses revealed that *RPG1* (ruptured pollen grain 1; At5g40620; Guan et al. 2008) is approximately four-fold downregulated in *atbzip34* pollen. Considering the expression profiles of both genes, RPG1 could represent a direct target of AtbZIP34 as the *RPG1* gene contains three copies of the core ACGT motif recognized by bZIP TFs within 1 kb of 5′ flanking sequence.

Exine pattern malformations were not the only phenotypic defects observed in *atbzip34 pollen*. Characteristic inclusions observed in the cytoplasm of *atbzip34* pollen (Fig. 3) suggested disturbance of metabolic pathways,

Table 1 Transcription factor genes with altered expression in *atbzip34* pollen

AGI	Family	Annotation	MS		BC		TC		MP1		MP2	
			Sig	DC	Sig	DC	Sig	DC	Sig	DC	Sig	DC
At5g15150	Homeobox	HAT7/HB-3	6	AA	6	AA	10	AA	4	AA	14	AAA
At3g04410	NAC	Hypothetical protein	135	PA	103	PA	157	PA	196	AA	309	AAA
At1g13370	Histone	Histone H3, putative	145	PA	125	AA	88	AA	99	AA	78	AAA
At1g52890	NAC	No apical meristem (NAM) family protein	198	AA	203	AA	125	AA	95	AA	123	AAA
At4g00940	C2C2-Dof	Dof-type zinc finger domain-containing protein	134	AA	117	AA	190	AA	276	AA	218	PAA
At3g61150	Homeobox	Homeobox-leucine zipper family protein HD-GL2-1	111	AA	100	PA	133	AA	154	AA	153	AAA
At2g41380	Unspecified	Embryo-abundant protein-related	24	AA	23	AA	58	AA	50	AA	131	AAA
At1g73870	C2C2-CO-like	Zinc finger family protein	28	AA	23	AA	36	AA	52	AA	112	PAA
At4g22950	MADS-box	MADS-box protein (AGL19)	66	PP	49	PA	64	PA	68	PA	88	AAA
At1g30650	WRKY	WRKY transcription factor	69	AA	54	AA	111	AA	102	AA	108	AAA
At1g07840	Unspecified	Leucine zipper factor-related	716	PP	607	PP	255	AA	264	AA	241	AAA
At5g06500	MADS-box	MADS-box family protein	260	AA	226	PA	121	PA	121	AA	180	PAA
At5g15310	MYB	MYB family transcription factor	65	AA	54	AA	77	AA	112	AA	131	PAA
At5g60890	MYB	MYB34	105	AA	83	AA	166	AA	209	AA	170	AAA
At5g06839	bZIP	bZIP family transcription factor	222	PA	183	PA	270	PA	355	PA	381	PAA
At1g27730	C2H2	ZAT10	129	AA	173	PP	190	AA	208	PA	244	PAA
At4g18020	APRR/GARP	APRR2/TOC2	253	AA	234	AA	276	AA	362	AA	401	AAA
At5g09240	General	Transcriptional coactivator p15 (PC4) family protein	490	PP	435	PP	383	PA	478	AA	532	AAA
At5g58890	MADS-box	MADS-box family protein	137	PP	128	PP	213	PP	232	PA	128	PAA
At3g55980	C3H	Zinc finger family protein	51	AA	72	PA	147	PP	86	AA	139	PAA
At3g17730	NAC	No apical meristem fam. protein	54	AA	51	AA	72	AA	107	AA	76	AAA
At2g25650	GeBP	DNA-binding storekeeper protein-related	15	AA	19	AA	22	AA	27	AA	59	AAA
At4g26930	MYB	MYB97	108	AA	122	PA	556	PP	1,466	PP	1,640	PPP
At1g61730	GeBP	DNA-binding storekeeper protein-related	1,924	PP	1,586	PP	330	AA	344	PA	243	PPA
At5g17580	Unspecified	Phototropic-responsive NPH3 family protein	1,303	PP	1,113	PP	170	AA	221	AA	230	AAA
At1g05290	C2C2-CO-like	Hypothetical protein	1,181	PP	1,543	PP	529	PP	204	PA	210	PPA
At1g26610	C2H2	Zinc finger family protein	1,771	PP	2,609	PP	1,737	PP	583	PP	624	PPP
At2g40670	ARR	ARR16	1,024	PP	1,605	PP	275	PP	265	PA	216	PPP
At1g48630	Unspecified	Guanine nucleotide-binding family protein	4,408	PP	3,951	PP	674	PP	822	PP	719	PPP
At5g25830	C2C2-GATA	Zinc finger family protein	895	PP	1,631	PP	1,353	PP	257	PA	375	PPA
At4g04830	General	Methionine sulfoxide reductase domain-containing protein	35	AA	41	AA	51	AA	67	AA	220	PPP
At5g05410	AP2	DREB2A	191	PP	338	PP	252	PP	240	PP	1,092	PPP
At5g44080	bZIP	bZIP family protein	445	PP	603	PP	545	PP	147	AA	337	PPA
At3g49530	NAC	No apical meristem fam. protein	228	PP	372	PP	250	AA	169	AA	847	PPP

Table 1 continued

AGI	Family	Annotation	MP3		MP4		MPG	bZIP		Fold change	
			Sig	DC	Sig	DC		Sig	DC	Down	Up
At5g15150	Homeobox	HAT7/HB-3	6	AA	42	AA	8	47	AA	0.17	6.03
At3g04410	NAC	Hypothetical protein	111	AA	188	AA	165	652	AA	0.25	3.95
At1g13370	Histone	Histone H3, putative	100	AA	129	AA	92	329	AA	0.28	3.57
At1g52890	NAC	No apical meristem (NAM) family protein	98	AA	1,279	PP	105	347	PP	0.3	3.29
At4g00940	C2C2-Dof	Dof-type zinc finger domain-containing protein	207	AA	241	AA	222	716	AA	0.31	3.23
At3g61150	Homeobox	Homeobox-leucine zipper family protein HD-GL2-1	170	AA	194	AA	159	510	AA	0.31	3.21
At2g41380	Unspecified	Embryo-abundant protein-related	138	AA	556	AA	106	340	AA	0.31	3.2
At1g73870	C2C2-CO-like	Zinc finger family protein	55	PA	208	PP	73	228	AA	0.32	3.12
At4g22950	MADS-box	MADS-box protein (AGL19)	98	PA	79	PA	88	274	AA	0.32	3.1
At1g30650	WRKY	WRKY transcription factor	166	AA	232	AA	126	388	PP	0.32	3.09
At1g07840	Unspecified	Leucine zipper factor-related	210	AA	362	AA	238	736	AA	0.32	3.09
At5g06500	MADS-box	MADS-box family protein	241	PA	149	AA	150	458	PA	0.33	3.05
At5g15310	MYB	MYB family transcription factor	149	AA	92	AA	130	398	AA	0.33	3.05
At5g60890	MYB	MYB34	172	AA	217	AA	184	551	AA	0.33	3
At5g06839	bZIP	bZIP family transcription factor	362	PP	419	PA	366	1,086	AA	0.34	2.97
At1g27730	C2H2	ZAT10	176	AA	2,331	PP	209	607	PP	0.34	2.9
At4g18020	APRR/GARP	APRR2/TOC2	354	AA	672	AA	372	1,079	AA	0.35	2.9
At5g09240	General	Transcriptional coactivator p15 (PC4) family protein	414	AA	922	PA	474	1,360	AA	0.35	2.87
At5g58890	MADS-box	MADS-box family protein	110	PP	120	AA	119	340	PP	0.35	2.85
At3g55980	C3H	Zinc finger family protein	119	AA	1,351	PP	114	320	PP	0.36	2.8
At3g17730	NAC	No apical meristem fam. protein	182	AA	145	AA	109	306	AA	0.36	2.79
At2g25650	GeBP	DNA-binding storekeeper protein-related	27	AA	71	AA	38	105	AA	0.36	2.79
At4g26930	MYB	MYB97	1,864	PP	886	PP	1,656	608	PP	2.72	0.37
At1g61730	GeBP	DNA-binding storekeeper protein-related	210	AA	356	PP	314	115	AA	2.73	0.37
At5g17580	Unspecified	Phototropic-responsive NPH3 family protein	222	AA	140	AA	224	80	AA	2.81	0.36
At1g05290	C2C2-CO-like	Hypothetical protein	70	PA	163	PP	192	65	AA	2.97	0.34
At1g26610	C2H2	Zinc finger family protein	314	PP	314	PP	404	129	PP	3.12	0.32
At2g40670	ARR	ARR16	177	AA	156	PP	183	58	AA	3.18	0.31
At1g48630	Unspecified	Guanine nucleotide-binding family protein	471	PP	974	PP	839	262	PA	3.2	0.31
At5g25830	C2C2-GATA	Zinc finger family protein	87	PA	274	PP	302	94	PA	3.21	0.31
At4g04830	General	Methionine sulfoxide reductase domain-containing protein	398	PP	280	PP	299	93	PA	3.23	0.31
At5g05410	AP2	DREB2A	1,243	PP	624	PP	986	273	PP	3.62	0.28
At5g44080	bZIP	bZIP family protein	304	AA	208	AA	283	77	AA	3.66	0.27
At3g49530	NAC	No apical meristem fam. protein	550	PP	637	PP	678	134	PA	5.06	0.2

The first three columns show AGI number, gene family and protein annotation. Following columns show mean expression value (Sig) and Detection call in all replicates (DC) of pollen developmental stages (MS, microspores; BC, bicellular pollen; TC, tricellular pollen), wild type mature pollen (MP1 (Honys and Twell 2004); MP2 (Zimmerman et al. 2005); MP3 (Pina et al. 2005); MP4 (this study); MPG (mean)) and *atbzip34* pollen (bZIP). Fold change columns show gene down- and up-regulation in *atbzip34* pollen

possibly related to cellular transport and/or lipid metabolism. This assumption was supported by transmission electron microscopy showing differences in structure of endomembrane systems and lipid bodies (Fig. 5). Lipid bodies were frequently found enclosed by an extensive network of ER especially at later developmental stages (see Murphy 2001 and references therein). The encirclement of pollen cytosolic lipid bodies by ER is proposed to prevent coalescence of lipid bodies (Piffanelli et al. 1998). This ER network persists throughout pollen release and has been proposed to facilitate the direct mobilization of the lipid-body TAGs required to support the rapid pollen tube growth (Murphy 2001; Piffanelli et al. 1998). In *atbzip34* pollen, lipid bodies were more numerous, localized in clusters close to cell surface and, most interestingly surrounded by a less dense ER network. In this respect the cytoplasm of *atbzip34* mature pollen also resembles that of immature pollen (Van Aelst et al. 1993; Yamamoto et al. 2003) that could indicate retarded pollen maturation. The observed higher frequency of unicellular and bicellular pollen as well as the smaller size of tricellular pollen supports this assumption.

It has been repeatedly demonstrated that pollen exine defects are generally of sporophytic origin, mainly caused by impaired tapetal cells development and/or function (Aarts et al. 1997; Ariizumi et al. 2003, 2004, 2008; Dong et al. 2005; Guan et al. 2008; Nishikawa et al. 2005; Paxson-Sowders et al. 2001). On the contrary, general pollen metabolism is likely to be under gametophytic control. To address these issues, developing male gametophyte and tapetum of both homozygous and heterozygous plants were observed for phenotypic defects. Light and electron microscopy observations (Figs. 3, 4, 5) confirmed the coordinated sporophytic and gametophytic modes of action already suggested by *AtbZIP34* expression pattern (Fig. 2). Exine patterning defects were shown to be under sporophytic control as indicated by the presence of defective exine in all pollen grains from *atbzip34* (-/-) plants and normal pollen shape from *atbzip34* heterozygous and wt plants. On the contrary, other metabolic defects as well as wrinkled intine were observed in around one half of pollen isolated from heterozygous plants and thus appear to be under gametophytic control. A similar mode of action can be expected for recently published AtbZIP60 that was also expressed in tapetum and male gametophyte besides number of other tissues. This membrane-bound TF was demonstrated to be involved in general ER stress response and its possible role in ER stress response function in normal development of secretory cells was suggested (Iwata et al. 2008).

To independently address the hypothesis of impaired cellular transport and metabolism, microarray analysis was performed to identify *AtbZIP34*-downstream genes. The reliability of microarray data was verified by quantitative RT-PCR of four orthologues of UDP-glucose epimerases. Treated microarray data were further analysed using Map Man to visualize metabolic pathways possibly affected by *atbzip34* mutation (Supplementary Fig. 2). Two diagrams showed relevant results, grouping transporters and genes involved in cell wall and lipid metabolism. There was apparent correlation between observed *atbzip34* pollen phenotype and microarray data supported by MapMan (Supplementary Fig. 2). First, there were no marked differential expression of other TFs (Fig. 7, Supplementary Fig. 3). This seems to confirm previously published findings of bZIP proteins mainly acting as "effector" TFs, controlling the expression of structural or metabolic genes, rather than other TFs. The only strong exception was the late pollen-expressed MYB97 gene. Although MYB97 is putatively pollen-specific, it is not a close orthologue of the MYB factor DUO1, that has an essential role in male germline development (Brownfield et al. 2009), moreover, its downregulation in *atbzip34* mutants does not support an important role in germ cell development.

On the contrary, six genes encoding proteins involved in lipid metabolism and/or transport were identified among the most highly downregulated genes. The most affected was the ABC transporter AtABCB9 that was downregulated 14.67-fold. AtABCB9 protein (synonyms AtMDR9 and AtPGP9) is a member of multidrug resistance subfamily of full ABC transporters (Sanchez-Fernandez et al. 2001a, b) and is also likely to be involved in lipid transport (Martinoia et al. 2002; Verrier et al. 2008). AtABCB9 expression was abundant and specific to the male gametophyte according to microarray data. Two seed storage/lipid transfer proteins downregulated in *atbzip34* pollen 6.6- and 4.4-fold are involved in lipid transfer from liposomes or microsomes to mitochondria and play a major role in membrane biogenesis by conveying phospholipids from their site of biogenesis to target membranes (Ohlrogge et al. 1991; Wirtz 1991). Aspartate aminotransferase (5.1-fold downregulated) is mainly involved in energy metabolism, in aspartate catabolic processes. However, besides its predominant role it was reported to be active also in fatty acid uptake in mitochondria in animal cells in a similar manner as its closely related plasma membrane fatty acid binding protein (FABPpm; Zhou et al. 1995). Moreover, aspartate aminotransferase was shown to interact with another downregulated enzyme, malate dehydrogenase (4.2-fold downregulated) in the inner mitochondrial membranes in various animal tissues (Teller et al. 1990). Apart from this, one study showed that malate dehydrogenase might be involved in plant peroxisomal fatty acid degradation (Kindl 1993) that in *Arabidopsis* seeds causes slow triacylglycerol mobilization and impaired growth (Pracharoenwattana et al. 2007). The last

affected enzyme involved in lipid catabolism was extracellular lipase that catalyses hydrolysis of triacylglycerols (Svendsen 2000).

Lipid storage, transport and catabolism were not the only affected metabolic pathways. The function of bZIP TFs RSG, RF2a and VSF-1 in vascular tissue development has been demonstrated in different species (Fukazawa et al. 2000; Ringli and Keller 1998; Yin et al. 1997). All these factors belong to group I bZIP proteins that heterodimerize with group E bZIP TFs (Shen et al. 2007). These facts together with the expression pattern of *AtbZIP34* (Fig. 2) provide indirect support for its involvement in regulation of transport tissue development and/or functions alongside its activity in gametophyte development.

Microarray data analyses further revealed that all subunits of Sec61 translocon were downregulated in *atbzip34* pollen (Fig. 8). Sec61 is an ER membrane protein translocator consisting of three subunits Sec61α, Sec61β and Sec61γ (Beckmann et al. 1997). Each subunit is encoded by three genes and all nine genes gave reliable signals in the male gametophyte. All genes but one (the least abundant) encoding all three Sec61 subunits were downregulated in *atbzip34* pollen. Moreover, gating protein AtBIP1 associated with the luminal side of Sec61 complex was downregulated to a greater extent that any Sec61 subunit. The orchestrated downregulation of almost all Sec61 subunits suggests common regulatory mechanism of Sec61 translocon synthesis in the male gametophyte. Many ribosomal proteins and proteins involved in protein posttranslational modifications were also downregulated in *atbzip34* pollen (Supplementary Table 3). Thus, the rate of protein synthesis together with protein translocation to ER and subsequent processing may be affected in *atbzip34* pollen, although its impact is not critical. Recently, a novel adenine nucleotide transporter (ER-ANT1, At5g17400) was identified in *Arabidopsis* ER (Leroch et al. 2008), but this gene is not significantly downregulated in *atbzip34* pollen. Among other phenotype defects, delayed flower bud development was observed in *er-ant1* knock-out lines. Moreover, several genes downregulated in *atbzip34* pollen were also downregulated in *er-ant1* plants (AtBIP1 (At5g28540), Sec61γ (At5g50460)). ER-ANT1 is involved in ATP/ADP antiport on ER membranes thus maintaining ATP concentration in the luminal space. Although unlikely because of coordinated downregulation of Sec61 subunits, the possibility of impaired protein translocation in *atbzip34* resulting from shortage of ATP cannot be completely ruled out.

Conclusions

Collectively our results indicate that AtbZIP34 has multiple roles in the development of gametophytic and sporophytic reproductive tissues. AtbZIP34 is the first bZIP-family TF with a demonstrated role in male gametophyte development. Characteristic phenotype and genetic transmission defects demonstrate a requirement for AtbZIP34 for correct formation of pollen cell walls. Although sporophytic control of exine patterning has been repeatedly shown in a number of mutants, analyses of *atbzip34* revealed sporophytic and gametophytic roles for AtbZIP34 in exine and intine formation. Wrinkled intine in $\sim$50% pollen of *+/atbzip34* plants, the presence of large inclusions in *atbzip34* pollen and the altered structure of ER in contact with lipid bodies indicate that AtbZIP34 is involved in gametophytic control of lipid metabolism, cellular transport and/or intine synthesis. This hypothesis is further supported by the downregulation of distinct subsets of genes. Moreover, altered cellular transport from the tapetum could also explain the defects observed in exine synthesis and cell wall patterning. The investigation of putative downstream genes including the regulator MYB97 will help to reveal new features in the cellular networks that control pollen wall development in relation to cellular transport.

Acknowledgments Authors thank Dr Milada Čiamporová (Institute of Botany, Slovak Academy of Sciences) and Dr Aleš Soukup (Department of Plant Physiology, Charles University in Prague) for their expertise with evaluation of transmission electron micrographs. Authors gratefully acknowledge financial support from Grant Agency of the Czech Republic (GACR grants 522/06/0896 and 522/09/0858) and Ministry of Education of the Czech Republic (MSMT grant LC06004). DT acknowledges support from the Biotechnology and Biological Sciences Research Council (BBSRC) and the University of Leicester.

References

Aarts MG, Hodge R, Kalantidis K, Florack D, Wilson ZA, Mulligan BJ, Stiekema WJ, Scott R, Pereira AP (1997) The *Arabidopsis* MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. Plant J 12:615–623. doi:10.1046/j.1365-313X.1997.00615.x

Alder NN, Shen Y, Brodsky JL, Hendershot LM, Johnson AE (2005) The molecular mechanisms underlying BiP-mediated gating of the Sec61 translocon of the endoplasmic reticulum. J Cell Biol 168:389–399. doi:10.1083/jcb.200409174

Allen RS, Li J, Stahle MI, Dubroue A, Gubler F, Millar AA (2007) Genetic analysis reveals functional redundancy and the major target genes of the *Arabidopsis* miR159 family. Proc Natl Acad Sci USA 104:16371–16376. doi:10.1073/pnas.0707653104

Ariizumi T, Hatakeyama K, Hinata K, Sato S, Kato T, Tabata S, Toriyama K (2003) A novel male-sterile mutant of *Arabidopsis thaliana*, faceless pollen-1, produces pollen with a smooth surface and an acetolysis-sensitive exine. Plant Mol Biol 53:107–116. doi:10.1023/B:PLAN.0000009269.97773.70

Ariizumi T, Hatakeyama K, Hinata K, Inatsugi R, Nishida I, Sato S, Kato T, Tabata S, 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*. Plant J 39:170–181. doi:10.1111/j.1365-313X.2004.02118.x

Ariizumi T, Kawanabe T, Hatakeyama K, Sato S, Kato T, Tabata S, Toriyama K (2008) Ultrastructural characterization of exine development of the transient defective exine 1 mutant suggests the existence of a factor involved in constructing reticulate exine architecture from sporopollenin aggregates. Plant Cell Physiol 49:58–67. doi:10.1093/pcp/pcm167

Beckmann R, Bubeck D, Grassucci R, Penczek P, Verschoor A, Blobel G, Frank J (1997) Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. Science 278:2123–2126. doi:10.1126/science.278.5346.2123

Boavida LC, McCormick S (2007) Temperature as a determinant factor for increased and reproducible in vitro pollen germination in *Arabidopsis thaliana*. Plant J 52:570–582. doi:10.1111/j.1365-313X.2007.03248.x

Bock KW, Honys D, Ward JM, Padmanaban S, Nawrocki EP, Hirschi KD, Twell D, Sze H (2006) Integrating membrane transport with male gametophyte development and function through transcriptomics. Plant Physiol 140:1151–1168. doi:10.1104/pp.105.074708

Borg M, Brownfield L, Twell D (2009) Male gametophyte development: a molecular perspective. J Exp Bot. doi:10.1093/jxb/ern355

Brownfield L, Hafidh S, Borg M, Sidorova A, Mori T, Twell D (2009) A plant germ cell-specific integrator of cell cycle progression and sperm specification. PLoS Genet 5 doi:10.1371/journal.pgen.1000430

Chen YN, Slabaugh E, Brandizzi F (2008) Membrane-tethered transcription factors in *Arabidopsis thaliana*: novel regulators in stress response and development. Curr Opin Plant Biol 11:695–701. doi:10.1016/j.pbi.2008.10.005

Choi H, Hong J, Ha J, Kang J, Kim SY (2000) ABFs, a family of ABA-responsive element binding factors. J Biol Chem 275:1723–1730. doi:10.1074/jbc.275.3.1723

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16:735–743. doi:10.1046/j.1365-313x.1998.00343.x

Cluis CP, Mouchel CF, Hardtke CS (2004) The *Arabidopsis* transcription factor HY5 integrates light and hormone signaling pathways. Plant J 38:332–347. doi:10.1111/j.1365-313X.2004.02052.x

Correa LG, Riano-Pachon DM, Schrago CG, dos Santos RV, Mueller-Roeber B, Vincentz M (2008) The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. PLoS ONE 3:e2944. doi:10.1371/journal.pone.0002944

Darlington GJ, Wang N, Hanson RW (1995) C/EBP alpha: a critical regulator of genes governing integrative metabolic processes. Curr Opin Genet Dev 5:565–570. doi:10.1016/0959-437X(95)80024-7

Darlington GJ, Ross SE, MacDougald OA (1998) The role of C/EBP genes in adipocyte differentiation. J Biol Chem 273:30057–30060. doi:10.1074/jbc.273.46.30057

Deppmann CD, Alvania RS, Taparowsky EJ (2006) Cross-species annotation of basic leucine zipper factor interactions: insight into the evolution of closed interaction networks. Mol Biol Evol 23:1480–1492. doi:10.1093/molbev/msl022

Dong X, Hong Z, Sivaramakrishnan M, Mahfouz M, Verma DP (2005) Callose synthase (CalS5) is required for exine formation during microgametogenesis and for pollen viability in *Arabidopsis*. Plant J 42:315–328. doi:10.1111/j.1365-313X.2005.02379.x

Dupl'áková N, Reňák D, Hovanec P, Honysová B, Twell D, Honys D (2007) *Arabidopsis* gene family profiler (aGFP)—user-oriented transcriptomic database with easy-to-use graphic interface. BMC Plant Biol 7:39. doi:10.1186/1471-2229-7-39

Durbarry A, Vizir I, Twell D (2005) Male germ line development in *Arabidopsis*. duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. Plant Physiol 137:297–307. doi:10.1104/pp.104.053165

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

Eferl R, Sibilia M, Hilberg F, Fuchsbichler A, Kufferath I, Guertl B, Zenz R, Wagner EF, Zatloukal K (1999) Functions of c-Jun in liver and heart development. J Cell Biol 145:1049–1061. doi:10.1083/jcb.145.5.1049

Ehlert A, Weltmeier F, Wang X, Mayer CS, Smeekens S, Vicente-Carbajosa J, Droge-Laser W (2006) Two-hybrid protein–protein interaction analysis in *Arabidopsis* protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. Plant J 46:890–900. doi:10.1111/j.1365-313X.2006.02731.x

Finkelstein RR, Lynch TJ (2000) The *Arabidopsis* abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. Plant Cell 12:599–609

Fukazawa J, Sakai T, Ishida S, Yamaguchi I, Kamiya Y, Takahashi Y (2000) Repression of shoot growth, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. Plant Cell 12:901–915

Guan YF, Huang XY, Zhu J, Gao JF, Zhang HX, Yang ZN (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 Physiol 147:852–863. doi:10.1104/pp.108.118026

Honys D, Twell D (2003) Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiol 132:640–652. doi:10.1104/pp.103.020925

Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. Genome Biol 5:R85. doi:10.1186/gb-2004-5-11-r85

Honys D, Reňák D, Twell D (2006) Male gametophyte development and function. In: Teixeira da Silva J (ed) Floriculture, ornamental and plant biotechnology: advances and topical issues, 1st edn. Global Science Books, London, pp 76–87

Ito T, Nagata N, Yoshiba Y, Ohme-Takagi M, Ma H, Shinozaki K (2007) *Arabidopsis* MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. Plant Cell 19:3549–3562. doi:10.1105/tpc.107.054536

Iwata Y, Koizumi N (2005) An *Arabidopsis* transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. Proc Natl Acad Sci USA 102:5280–5285. doi:10.1073/pnas.0408941102

Iwata Y, Fedoroff NV, Koizumi N (2008) *Arabidopsis* bZIP60 is a proteolysis-activated transcription factor involved in the endoplasmic reticulum stress response. Plant Cell 20(11):3107–3121

Jakoby M, Weisshaar B, Droge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F (2002) bZIP transcription factors in *Arabidopsis*. Trends Plant Sci 7:106–111. doi:10.1016/S1360-1385(01)02223-3

Karimi M, De Meyer B, Hilson P (2005) Modular cloning in plant cells. Trends Plant Sci 10:103–105

Kindl H (1993) Fatty acid degradation in plant peroxisomes: function and biosynthesis of the enzymes involved. Biochimie 75:225–230. doi:10.1016/0300-9084(93)90080-C

Leroch M, Neuhaus HE, Kirchberger S, Zimmermann S, Melzer M, Gerhold J, Tjaden J (2008) Identification of a novel adenine nucleotide transporter in the endoplasmic reticulum of *Arabidopsis*. Plant Cell 20:438–451. doi:10.1105/tpc.107.057554

Li C, Wong WH (2001a) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci USA 98:31–36. doi:10.1073/pnas.011404098

Li C, Wong WH (2001b) Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol 2:R32

Liu JX, Srivastava R, Che P, Howell SH (2007a) An endoplasmic reticulum stress response in *Arabidopsis* is mediated by proteolytic processing and nuclear relocation of a membrane-associated transcription factor, bZIP28. Plant Cell 19:4111–4119. doi:10.1105/tpc.106.050021

Liu JX, Srivastava R, Che P, Howell SH (2007b) Salt stress responses in *Arabidopsis* utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. Plant J 51:897–909. doi:10.1111/j.1365-313X.2007.03195.x

Lu G, Gao C, Zheng X, Han B (2008) Identification of OsbZIP72 as a positive regulator of ABA response and drought tolerance in rice. Epub ahead of print, Planta

Martinoia E, Klein M, Geisler M, Bovet L, Forestier C, Kolukisaoglu U, Muller-Rober B, Schulz B (2002) Multifunctionality of plant ABC transporters—more than just detoxifiers. Planta 214:345–355. doi:10.1007/s004250100661

McCormick S (2004) Control of male gametophyte development. Plant Cell 16(Suppl):S142–S153. doi:10.1105/tpc.016659

Millar AA, Gubler F (2005) The *Arabidopsis* GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. Plant Cell 17:705–721. doi:10.1105/tpc.104.027920

Murphy DJ (2001) The biogenesis and functions of lipid bodies in animals, plants and microorganisms. Prog Lipid Res 40:325–438. doi:10.1016/S0163-7827(01)00013-3

Newman JR, Keating AE (2003) Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science 300:2097–2101. doi:10.1126/science.1084648

Nijhawan A, Jain M, Tyagi AK, Khurana JP (2008) Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. Plant Physiol 146:333–350. doi:10.1104/pp.107.112821

Nishikawa S, Zinkl GM, Swanson RJ, Maruyama D, Preuss D (2005) Callose (beta-1, 3 glucan) is essential for *Arabidopsis* pollen wall patterning, but not tube growth. BMC Plant Biol 5:22. doi:10.1186/1471-2229-5-22

Ohlrogge JB, Browse J, Somerville CR (1991) The genetics of plant lipids. Biochim Biophys Acta 1082:1–26

Park SK, Howden R, Twell D (1998) The *Arabidopsis thaliana* gametophytic mutation gemini pollen1 disrupts microspore polarity, division asymmetry and pollen cell fate. Development 125:3789–3799

Paxson-Sowders DM, Dodrill CH, Owen HA, Makaroff CA (2001) DEX1, a novel plant protein, is required for exine pattern formation during pollen development in *Arabidopsis*. Plant Physiol 127:1739–1749. doi:10.1104/pp.010517

Piffanelli P, Ross JHE, Murphy DJ (1998) Biogenesis and function of the lipidic structures of pollen grains. Sex Plant Reprod 11:65–80. doi:10.1007/s004970050122

Pina C, Pinto F, Feijo JA, Becker JD (2005) Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol 138:744–756. doi:10.1104/pp.104.057935

Pracharoenwattana I, Cornah JE, Smith SM (2007) *Arabidopsis* peroxisomal malate dehydrogenase functions in beta-oxidation but not in the glyoxylate cycle. Plant J 50:381–390. doi:10.1111/j.1365-313X.2007.03055.x

Ringli C, Keller B (1998) Specific interaction of the tomato bZIP transcription factor VSF-1 with a non-palindromic DNA sequence that controls vascular gene expression. Plant Mol Biol 37:977–988. doi:10.1023/A:1006030007333

Rotman N, Durbarry A, Wardle A, Yang WC, Chaboud A, Faure JE, Berger F, Twell D (2005) A novel class of MYB factors controls sperm-cell formation in plants. Curr Biol 15:244–248. doi:10.1016/j.cub.2005.01.013

Sanchez-Fernandez R, Davies TG, Coleman JO, Rea PA (2001a) The *Arabidopsis thaliana* ABC protein superfamily, a complete inventory. J Biol Chem 276:30231–30244. doi:10.1074/jbc.M103104200

Sanchez-Fernandez R, Rea PA, Davies TG, Coleman JO (2001b) Do plants have more genes than humans? Yes, when it comes to ABC proteins. Trends Plant Sci 6:347–348. doi:10.1016/S1360-1385(01)02038-6

Sanyal S, Sandstrom DJ, Hoeffer CA, Ramaswami M (2002) AP-1 functions upstream of CREB to control synaptic plasticity in Drosophila. Nature 416:870–874. doi:10.1038/416870a

Schindler U, Menkens AE, Beckmann H, Ecker JR, Cashmore AR (1992) Heterodimerization between light-regulated and ubiquitously expressed *Arabidopsis* GBF bZIP proteins. EMBO J 11:1261–1273

Seo PJ, Kim SG, Park CM (2008) Membrane-bound transcription factors in plants. Trends Plant Sci 13(10):550–556

Shen H, Cao K, Wang X (2007) A conserved proline residue in the leucine zipper region of AtbZIP34 and AtbZIP61 in *Arabidopsis thaliana* interferes with the formation of homodimer. Biochem Biophys Res Commun 362:425–430. doi:10.1016/j.bbrc.2007.08.026

Shen H, Cao K, Wang X (2008) AtbZIP16 and AtbZIP68, two new members of GBFs, can interact with other G group bZIPs in *Arabidopsis thaliana*. BMB Rep 41:132–138

Smyth DR, Bowman JL, Meyerowitz EM (1990) Early flower development in *Arabidopsis*. Plant Cell 2:755–767

Svendsen A (2000) Lipase protein engineering. Biochim Biophys Acta 1543:223–238

Sze H, Padmanaban S, Cellier F, Honys D, Cheng NH, Bock KW, Conejero G, Li X, Twell D, Ward JM, Hirschi KD (2004) Expression patterns of a novel AtCHX gene family highlight potential roles in osmotic adjustment and K+ homeostasis in pollen development. Plant Physiol 136:2532–2547. doi:10.1104/pp.104.046003

Tajima H, Iwata Y, Iwano M, Takayama S, Koizumi N (2008) Identification of an *Arabidopsis* transmembrane bZIP transcription factor involved in the endoplasmic reticulum stress response. Biochem Biophys Res Commun 374:242–247. doi:10.1016/j.bbrc.2008.07.021

Takeda T, Toda T, Kominami K, Kohnosu A, Yanagida M, Jones N (1995) Schizosaccharomyces pombe atf1+ encodes a transcription factor required for sexual development and entry into stationary phase. EMBO J 14:6193–6208

Takeda T, Amano K, Ohto MA, Nakamura K, Sato S, Kato T, Tabata S, Ueguchi C (2006) RNA interference of the *Arabidopsis* putative transcription factor TCP16 gene results in abortion of early pollen development. Plant Mol Biol 61:165–177. doi:10.1007/s11103-006-6265-9

Tateda C, Ozaki R, Onodera Y, Takahashi Y, Yamaguchi K, Berberich T, Koizumi N, Kusano T (2008) NtbZIP60, an endoplasmic reticulum-localized transcription factor, plays a role in the defense response against bacterial pathogens in *Nicotiana tabacum*. J Plant Res 121(6):603–611

Teller JK, Fahien LA, Valdivia E (1990) Interactions among mitochondrial aspartate aminotransferase, malate dehydrogenase, and the inner mitochondrial membrane from heart, hepatoma, and liver. J Biol Chem 265:19486–19494

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37:914–939. doi:10.1111/j.1365-313X.2004.02016.x

Twell D, Oh S-A, Honys D (2006) Pollen development, a genetic and transcriptomic view. In: Malho R (ed) The pollen tube, vol 3. Springer–Verlag, Berlin, Heidelberg, pp 15–45

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K (2000) *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci USA 97:11632–11637. doi:10.1073/pnas.190309197

Van Aelst AC, Pierson ES, Van Went JL, Cresti M (1993) Ultrastructural changes of *Arabidopsis thaliana* pollen during final maturation and rehydration. Zygote 1:173–179. doi:10.1017/S096719940000143X

Verelst W, Saedler H, Munster T (2007a) MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific *Arabidopsis* promoters. Plant Physiol 143:447–460. doi:10.1104/pp.106.089805

Verelst W, Twell D, de Folter S, Immink R, Saedler H, Munster T (2007b) MADS-complexes regulate transcriptome dynamics during pollen maturation. Genome Biol 8:R249. doi:10.1186/gb-2007-8-11-r249

Verrier PJ, Bird D, Burla B, Dassa E, Forestier C, Geisler M, Klein M, Kolukisaoglu U, Lee Y, Martinoia E, Murphy A, Rea PA, Samuels L, Schulz B, Spalding EJ, Yazaki K, Theodoulou FL (2008) Plant ABC proteins—a unified nomenclature and updated inventory. Trends Plant Sci 13:151–159. doi:10.1016/j.tplants.2008.02.001

Vizcay-Barrena G, Wilson ZA (2006) Altered tapetal PCD and pollen wall development in the *Arabidopsis* ms1 mutant. J Exp Bot 57:2709–2717. doi:10.1093/jxb/erl032

Wang ZQ, Ovitt C, Grigoriadis AE, Mohle-Steinlein U, Ruther U, Wagner EF (1992) Bone and haematopoietic defects in mice lacking c-fos. Nature 360:741–745. doi:10.1038/360741a0

Watanabe Y, Yamamoto M (1996) Schizosaccharomyces pombe pcr1+ encodes a CREB/ATF protein involved in regulation of gene expression for sexual development. Mol Cell Biol 16:704–711

Weigel D, Glazebrook J (2002) *Arabidopsis*. A laboratory handbook. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Weltmeier F, Rahmani F, Ehlert A, Dietrich K, Schutze K, Wang X, Chaban C, Hanson J, Teige M, Harter K, Vicente-Carbajosa J, Smeekens S, Droge-Laser W (2009) Expression patterns within the *Arabidopsis* C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. Plant Mol Biol 69:107–119. doi:10.1007/s11103-008-9410-9

Wirtz KW (1991) Phospholipid transfer proteins. Annu Rev Biochem 60:73–99. doi:10.1146/annurev.bi.60.070191.000445

Xiang Y, Tang N, Du H, Ye H, Xiong L (2008) Characterization of OsbZIP23 as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. Plant Physiol 148:1938–1952. doi:10.1104/pp.108.128199

Yamaguchi S, Mitsui S, Yan L, Yagita K, Miyake S, Okamura H (2005) Role of DBP in the circadian oscillatory mechanism. Mol Cell Biol 20:4773–4781. doi:10.1128/MCB.20.13.4773-4781.2000

Yamamoto Y, Nishimura M, Hara-Nishimura I, Noguchi T (2003) Behavior of vacuoles during microspore and pollen development in *Arabidopsis thaliana*. Plant Cell Physiol 44:1192–1201. doi:10.1093/pcp/pcg147

Yang C, Vizcay-Barrena G, Conner K, Wilson ZA (2007) MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. Plant Cell 19:3530–3548. doi:10.1105/tpc.107.054981

Yin Y, Zhu Q, Dai S, Lamb C, Beachy RNP (1997) RF2a, a bZIP transcriptional activator of the phloem-specific rice tungro bacilliform virus promoter, functions in vascular development. EMBO J 16:5247–5259. doi:10.1093/emboj/16.17.5247

Zhang ZB, Zhu J, Gao JF, Wang C, Li H, Zhang HQ, Zhang S, Wang DM, Wang QX, Huang H, Xia HJ, Yang ZN (2007) Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. Plant J 52:528–538. doi:10.1111/j.1365-313X.2007.03254.x

Zhou SL, Stump D, Kiang CL, Isola LM, Berk PD (1995) Mitochondrial aspartate aminotransferase expressed on the surface of 3T3–L1 adipocytes mediates saturable fatty acid uptake. Proc Soc Exp Biol Med 208:263–270

Zhu J, Chen H, Li H, Gao JF, Jiang H, Wang C, Guan YF, Yang ZN (2008) Defective in Tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. Plant J 55:266–277. doi:10.1111/j.1365-313X.2008.03500.x

Zimmermann P, Hennig L, Gruissem W (2005) Gene-expression analysis and network discovery using Genevestigator. Trends Plant Sci 10:407–409. doi:10.1016/j.tplants.2005.07.003

8.5. C2-domain protein DEPOLL is required for intine development in Arabidopsis pollen

Dupľáková Nikoleta, Novotný Marián, Fíla Jan, Levkaničová Zuzana, Honys David

Planta 2010 (in preparation)

The article describes the functional characterization of the DEPOLL gene (At1g70790). This gene was originally annotated as transcription factor with C2H2-type zinc finger domain. However, *in-silico* analyses made this presumption unlikely. *depoll* mutation resulted in increased number of dead pollen and defects in the intine ultrastructure and organisation. Similarly, transcripotomic experiment led to the identification of several gene families affected in mature pollen by *depoll* mutation. Therefore DEPOLL is considered to function in proper intine and plasma membrane organisation.

My contribution: As a first author I was responsive for most of the work in this paper. I am being involved in the manuscript preparation.

Supporting online material on CD

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Nikoleta Dupľáková, Marian Novotný, Jan Fíla, Zuzana Levkaničová, David Honys

N. Dupľáková, D. Honys, Jan Fíla, Zuzana Levkaničová
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

N. Dupľáková, D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

M. Novotný
Department of Cell Biology,
Faculty of Science, Charles University in Prague,
Viničná 7, 128 44, Praha 2, Czech Republic

Author for correspondence:
D. Honys
e-mail: honys@ueb.cas.cz
tel. number: +420225106450
fax number: +420225106456

Abstract

Developmental processes responsible for the proper male gametophyte formation are one of the most critical steps in plant reproduction. Here we present the functional analysis of the At1g70790 gene (*DEPOLL*). This gene was originally annotated as transcription factor with C2H2-type zinc finger domain. However, *in-silico* analyses made this presumption unlikely. We confirmed only the C2 domain in N-terminal region. The mature pollen of *depoll* plants exhibits the plentiful phenotypic defects including increased number of dead pollen grains. The electron microscopy revealed further defects in the intine ultrastructure and organisation. In *depoll* pollen, intine was greatly reduced and often nearly missing, passing to granular structure leaving only aperture regions unaffected. Accordingly, the *depoll* mutant showed no changes in pollen germination. The microarray experiment led to the identification of several gene families affected in mature pollen by *depoll* mutation including genes related to membrane transport on the base their ability to facilitate *in vitro* transfer of phospholipids between membranes (lipid transfer proteins) and cell wall metabolism (invertase/pectin methylesterase inhibitor family proteins, polygalacturonase family proteins, xyloglucan endotransglucosylase/hydrolases family proteins). We consider DEPOLL function in proper intine and/or plasma membrane structure.

Key words: *Arabidopsis thaliana*, cell wall, intine, electron microscopy, transcriptomics, pollen development.

Abbrevations

MGU, male germ unit; PKC protein kinase C; PMII, pollen mitosis II; qRT-PCR, quantitative real time PCR; SEM, scanning electron microscopy; TEM, transmission electron microscopy; TF, transcription factor, nsLTP, non-specific lipid transfer protein; PG, polygalacturonase, PME, pectin methylesterases, XTH, xyloglucan endotransglucosylase/hydrolases.

Introduction

Life cycle of flowering plants consists of dominant vegetative diploid phase and markedly reduced but vital haploid phase, the male or female gametophyte with specialized function in plant fertility and crop production. Male gametophyte development can be divided into two consequent phases, early and late. The early phase lies between the end meiosis and the first haploid mitosis that represents also the beginning of the late phase finished at pollen maturation. Both developmental phases are characterised by specific sets of expressed genes. The percentage of specific genes increases from unicellular microspores to mature pollen (Honys and Twell 2004).

Broad spectrum of gametophytic mutants affecting pollen development in many aspects had been described and characterised (for review see (Honys et al. 2006)). Most mutations were related to the first and second haploid mitoses (ie. *duo1*, *gem1*), to the male germ unit (MGU) structure (ie. *mud*, *gum*) and very often to pollen cell wall structure (ie. *atbzip34* (Gibalova et al. 2009).

Unique pollen cell wall composition and structure reflect its specialized functions, namely high resistance towards environmental conditions and cell-cell recognition reaction on the stigma. Pollen grain wall is composed of two main layers, the pecto-cellulosic intine and exine formed by sporopollenin. Sporopollenin is secreted from metabolically active tapetum into anther loculi. It has significant influence on pollen wall patterning and represent one of the most known resistant biopolymers. This mixed polymers of fatty acid and phenylpropanoid derivatives is responsible for the elaborately sculpted, complex structures and many properties of the pollen cell wall. It serves as protective barier for the survival of the pollen grain free from the sporophyte prior to fertilization (Bedinger 1992), it has critical function in pollen dispersal and pollen–stigma interactions (Piffanelli et al. 1998; Zinkl et al. 1999).

Following chemical fixation with glutaraldehyde and post-fixation with osmium tetroxide, exine represents an electron-dense structure in contrast to electron-translucent intine (Van Aelst and Van Went 1992). In general, sporophytic exine consists of two main layers: inner and featureless nexine and the outer sexine with the fine elaborate structure of multiple baculae and tecta. On the contrary, intine is relatively simple layer adjacent to the plasma membrane. It is composed of two sub-layers: granular exintine with pectin and protein inclusions facing the exterior and microfibrillar cellulosic endintine facing the interior. The intine is of the gametophytic origin and is secreted by the microspore during the ring-vacuolated microspore stage (Owen and Makaroff 1995).

The importance of pollen wall is reflected in number of characterised pollen cell-wall mutants with defects that can be so tremendous that they lead to pollen sterility or even pollen development aborition at various developmental stages. Many male-sterile or non-sterile gametophytic or sporophytic mutants were revealed in previous years. Some mutations were related to the exine patterning (Suzuki et al. 2008; Guan et al. 2008; Ariizumi et al. 2008; Ariizumi et al. 2003; Ariizumi et al. 2004a). Among sporophytic mutations that impair microspore development and show pollen wall-patterning defects, belong *Arabidopsis thaliana* mutants *ms9 and ms12* showing primexine defects during pollen development

(Taylor et al. 1998) and *dex1* mutant with deviations in the microspore membrane and primexine formation (Paxson-Sowders et al. 2001). DEX1 is a membrane associated protein containing several potential calcium-binding domains (Paxson-Sowders et al. 2001). Another gene important for early stage of microgametogenesis is *MALE STERILITY 2* gene (*MS2*). It is expressed in tapetum layer at stage when the microspore are released from tetrads or shortly after that and affect the exine layer deposition in microspore development. The exine in *ms2* pollen is absent and thus the pollen grain are very poorly constructed. The function of this gene is proposed in fatty acid processing for exine formation (Aarts et al. 1997). NEF1 is *non exine formation 1* mutants with completly absent pollen at anthesis in *Arabidopsis thaliana*. It belongs to male-sterile mutant with pollen degeneration after microspore release (Ariizumi et al. 2004b). Another male-sterile mutant in *Arabidopsis* is *faceless pollen1* which shows conditional sterility. Mutant pollen grains are with smooth surface and reticulate pattern that is not prominent. This phenotype is elicited by tryphine filling in the exine cavities and covering the pollen surface. The *FLP1* gene is probably involved in synthesis of components of tryphine, sporopollenin of exine and the wax of siliques and stem (Ariizumi et al. 2003).

On the contrary, only limited number of gametophytic mutations affecting intine development have been identified. One of them is a mutation in *BcMF2* gene encoding putative polygalacturonase in *Brassica campestris.* This mutant shows defects in the shape and position of apertures caused by overdevelopment of the exintine that subsequently occupies most of the intine (Huang et al. 2009). The last one mutant, that shows incomplete intine in microspore stage is *ms8* mutant. It is proposed, in this case intine formation sporofytically controlled in stage microspore development.

T-DNA insertion mutagenesis represents the efficient tool for studying the male gametophyte development and function. Our previous global analyses of mutations affecting male gametophyte development (unpublished results) based on pollen transcriptomic data (Honys and Twell 2003, 2004) led to the identification of several gametophytic genes. One of these genes, At1g70790, was shown to cause defects in pollen wall development. At1g70790 was annotated as a C2 domain-containing protein. C2 domain is a widespread intracellular protein domain originally described in Ca^{2+}-dependent isoforms of protein kinase C (PKC) (Corbolán-García and Gómez-Fernández 2010). Members of the protein kinase C family are central proteins in downstream signal transduction pathway and, as such, are involved in the regulation of a multitude of cellular processes (Krauss 2003) such as in permeability, contraction, migration, hypertrophy, proliferation, apoptosis, secretion and etc. (Dempsey et al. 2000). C2 domain was primarily identified as a Ca^{2+}-binding motif with phospholipid binding to many C2 domains being regulated by Ca^{2+} (Nalefski and Falke 1996). Subsequently, C2 domain was shown to evolutionarily diverge into Ca^{2+} -dependent and Ca^{2+} -independent forms interacting with multiple targets. C2 domain displayed the remarkable property of binding a variety of different ligands and substrates, including Ca^{2+}, phospholipids, inositol polyphosphates and intracellular proteins (Nalefski and Falke 1996). Generally it is supposed that most C2 domains are putative Ca^{2+}-binding domains that represent a family of versatile protein modules with diverse functions (Rizo and Sudhof 1998). Single or multiple copies of C2 domains have been identified in a growing number of signalling proteins that interact with cellular membranes and mediate a broad array of critical intracellular processes, including membrane trafficking (Perin et al. 1990; Shirataki et al. 1993) and signal transduction - the generation of lipid second messengers (PLC (Rhee et al. 1989)), activation of GTPases (Ras-GAP (Trahey et al. 1988)), the control of protein phosphorylation (PKC (Coussens et al. 1986)), ubiquitin ligation (Nalefski and Falke 1996; Rizo and Sudhof 1998).

Only few C2 domain-containing proteins have been identified and/or even functionally characterized in plants. One of them, SRC2 protein in *Capsicum annum,* has implicated role

in resistance against host and non-host pathogens and abiotic stresses and plays a critical role in SRC2 localization on plasma membrane (Kim et al. 2008). In *Oryza sativa,* two small C2-domain proteins of 156 and 159 amino acids were identified. One of them, OsERG1 is believed to be involved in plant defense signaling systems. It binds to phospholipid vesicles in a Ca^{2+}-dependent manner and is translocated to the plasma membrane of plant cells after ungal elicitor treatment (Kim et al. 2003). The last example of single C2-domain protein with relation to stress, Vr-PLC3, was identified in *Vigna radiata*. Its C2 domain was also proven to be essential for the Vr-PLC3 targeting to the plasma membrane in *Arabidopsis* protoplasts.

Here we present the cytological and functional data for the *depoll* mutant supported by the transcriptomic experiment. Collectively, our results suggest a putative role of DEPOLL protein in intine development although its previously annotated function as a transcription factor is very unlikely.

Materials and Methods

Plant material and growth conditions

Arabidopsis thaliana T-DNA line SAIL_1168_C11 (insertion in At1g70790 gene; Basta resitance) was obtained from NASC (The Nottingham Arabidopsis Stock Centre). Columbia-0 plants were used as a wild type. Plants were grown in pots containing mixture of vermiculite, soil and sand under controlled conditions(growth room, 21°C, illumination of 150 μmol m^{-2} s^{-1}, 16-h photoperiod, 50% relative humidity)**.** Pollen for microarray experiments was harvested from two independently grown populations according to (Honys and Twell 2003). The purity of isolated pollen was determined by light microscopy and 4',6-diamino-phenylindole (DAPI) -staining according to (Park et al. 1998).

Genomic DNA was isolated by a CTAB DNA extraction method modified from (Weigel and Glazebrook 2002) in which initial grinding of leaf tissue frozen in liquid nitrogen was carried out ground for 12 s with glass beads in a dental amalgam mixer ESME Cap mix (3M, Maplewood, MN).

Genotyping of T-DNA insertion lines

Appropriate gene-specific (1168_C11_F, 1168_C11_R_new) and insert-specific primer LB3_25 were designed using Primer3 software (http://www-genome.wi.mit.edu/ cgi-bin/primer/primer3_www.cgi) and are listed in Supplementary Table 1. The PCR program was as follows: 40 cycles of 2 min at 94°C, 30 sec at 94°C, 30 sec at 55°C, 90 sec at 72°C, 3 min at 72°C.

RNA extraction, probe preparation and DNA chip hybridization

Mature pollen grains from *Arabidopsis thaliana* Col-0 (WT) and *depoll* plants were isolated from flowers (developmental stage 14) by 0.3 M mannitol as described previously (Honys and Twell 2003). Total RNA was extracted from 100 mg of isolated pollen using the RNeasy Plant Mini Kit according to the manufacturer's instructions (Qiagen, Valencia, CA). For microarray experiments, RNA integrity was checked using an Agilent 2100 Bioanalyser (Agilent Technologies, Boblingen, Germany) at NASC. Biotinylated target RNA was prepared from 20 ìg of total RNA as described in the Affymetrix GeneChip Technical Analysis Manual (Affymetrix, Santa Clara, CA). Preparation of cRNA probes and hybridization to ATH1 Genome Arrays and scanning were carried out as decribed (Honys and Twell 2003). Publicly available transcriptomic datasets were downloaded from arabidopsisGFP database (http://aGFP.ueb.cas.cz; (Dupľáková et al. 2007)). All transcriptomics datasets were normalized using freely available dChip 1.3 software (http://www.dchip.org). The reliability and reproducibility of analyses was ensured by the use

of duplicates or triplicates in each experiment, the normalization of all arrays to the median probe intensity level and the use of normalized CEL intensities of all arrays for the calculation of model-based gene-expression values based on the Perfect Match-only model (Li and Hung Wong 2001; Li and Wong 2001). For increasing realibility of data interpretation we processed all seven available mature pollen transcriptomic datasets of wild-type as a reference labeled MP1 (Ler, 2 repeats; (Honys and Twell 2004)), MP2 (Col-0, 2 repeats; (Pina et al. 2005)), MP3 (Col-0, 3 repeats; (Zimmermann et al. 2005)), MP4 (Col-0, 3 repeats; (Borges et al. 2008; Qin et al. 2009)), MP5 (Col-0, 4 repeats, (Qin et al. 2009)), MP6 (Col-0, 2 repeats, (Wang et al. 2008)) and MP7 (2 repeats, wild type in this study). For each gene, the two most deviant expression values were eliminated and mean from remaining five was calculated (MP).

Quantitative real-time PCR

Quantitative real-time PCR (qRT-PCR) was carried out on a LightCycler 480 using LightCycler FastStart DNA Master^{PLUS} SYBR Green I and LightCycler 480 Multiwell Plate 96 (Roche Applied Science, Manheim, Germany) according manufacturer´s instructions. Reaction was performed in 96-well plastic plate (Roche, Manheim, Germany in total reaction volume of 10 µl with 500 nM of each HPLC-purified primers (Supplementary Table 1) and 2.5 ul of 50x diluted first strand cDNA. qRT-PCR data were collected on the light cycler with cycling conditions: 10 min of initial activation at 95˚C, then 50 cycles of 10 s at 95˚C, 10 s at 60˚C, and 10 s at 72˚C. PCR efficiencies were estimated from calibration curves generated from serial dilution of cDNAs. Real time PCR expression measurements were normalized with the expression of reference transcript (RNA-binding protein 45; At1g11650). The calibrator normalized relative ratio of the relative amount of the target and reference gene was calculated as follows: E_R^{CpR} / E_T^{CpT} (E_T, E_R: efficiency for target or reference gene qRT PCR assay; C_pT, C_pR: a crossing point for target or reference genes).

Complementation analysis

A 3,193 bp genomic fragment including the complete *DEPOLL* gene, 1,363 bp upstream region and 426 bp trailer was PCR-amplified using C2H2_70790 attB1 and C2H2_70790 attB2 primers (Supplementary Table 1) and recombined into the pDONR201 vector (Invitrogen, Carlsbad, CA). This entry clone was further recombined into pKGW GATEWAY–compatible destination vector (VIB, Ghent, Belgium, (Karimi et al. 2005)). Constructs were verified by restriction analysis and sequenced. Homozygous *depoll* plants were transformed using the floral dip (Clough and Bent 1998) and *Agrobacterium tumefaciens* strain GV3101. Transformants were selected on ½ MS medium-300mL (0.66 g Murashige and Skoog basal medium, 3g sucrose, 30 mg Myo-inositol, 150mg MES-2-(N-morpholino)ethanesulfonic acid, 0.8% agar, pH 5.7 with KOH) containing 50 µg/ml kanamycin. Transformants were verified by PCR with 1168_C11 F/1168_C11_R_new primers (Supplementary Table 1). Phenotypic complementation was examined by bright field and fluorescence microscopy after DAPI staining as described (Park et al. 1998).

Light and fluorescent microscopy

For pollen phenotype observations, flowers from *depoll* or wild type plants were collected to GUS buffer (0.1 M phosphate buffer, pH 7.0; 10 mM EDTA, pH8.0; 0.1% triton X-100) supplemented with 100 ng/ml DAPI as described (Park et al. 1998). After 30-min incubation at room temperature, samples were analyzed by bright field and fluorescence microscopy with Nikon TE2000-E microscope (objective Nikon CFI Plan Fluor ELWD 10x/0.60, eyepiece Nikon CFI 10x/22, intermediate magnification 1x-1.5x).

For mature pollen scoring, three fully open flowers (stage 14 (Smyth et al. 1990)) were cut with scissors from *depoll/depoll* or wild type plants and pollen was transfered to the agarose pad by tapping. The bright field microscopy was performed on the Nikon TE2000-E microscope as specified above.

Pollen germination in vitro

Pollen was collected from just-open flowers of wild type and *depoll* plants. Pollen grains were germinated on a germination medium on microscope slide according to Boavida and McCormick (2007) with several modifications. Pollen germination medium (final volume 25 ml) was always prepared fresh from 0.5 M stock solutions of the main components (5 mM KCl, 0.01% H3BO3, 5 mM CaCl2, 1 mM MgSO4) using autoclaved water. Sucrose (10%) was added and dissolved and pH was then adjusted to 7.5 with NaOH. About 1.5% of low-melting agarose (Amresco, Solon, Ohio) was added and briefly heated in a microwave oven, just long enough for the agarose to melt. Glass slide was then filled with 500 ml melted germination media. Pollen from individual flowers was spread on the surface of germination pad by inverting the flower with the help of tweezers and gently bringing it onto agarose surface after its solidification. The whole flower could be used as a ‘‘brush’’ to spread pollen uniformly on the surface of the germination medium. Glass slides were immediately placed inside a moisture incubation chamber to avoid media dehydration and incubated for 10 h in the dark at 24_C. Samples were examined by bright field and fluorescence microscopy with an Olympus DP50-CU microscope.

Scanning electron microscopy

For scanning electron microscopy, freshly harvested pollen from Col-0 or *depoll* plants was fixed in a 2.5% (w/v) glutaraldehyde in 0.1M PBS for 24 h at room temperature, washed with 4% glucose in 0.1M PBS for 15 min., dehydrated through an inscreasing ethanol series (30%-100% ethanol), and, via ethanol:acetone, to acetone. Pollen samples for scanning electron microscopy were then critical point dried in CO_2, mounted on a stub, sputter coated with gold, and observed and photographed with a JEOL 6300 scanning microscope.

Transmnission electron microscopy

Freshly harvested pollen from Col-0 or *depoll* plants from flower developmental stage 14 (Smyth et al. 1990) was was fixed in a 2.5% (w/v) glutaraldehyde in a 0.1 M cacodylate buffer (pH7.2) for 24 h at room temperature, washed with 4% glucose in 0.1M PBS (NaH_2PO_4 x H_2O, pH=7.0) for 15 min., post-fixed in 2% (w/v) osmium tetroxide in 0.1M PBS buffer, washed with 4% glucose in 0.1M PBS for 15 min, dehydrated through an ascending ethanol series (30%-100% ethanol), and, via ethanol:acetone, to acetone. Samples were embedded in Poly/Bed® 812/Araldite 502 resins. Thin sections (70 nm) were cut on a Reichert-Jung Ultracut E ultra-microtome and stained using uranyl acetate and lead citrate. Sections were analyzed and photographed using the JEM-1011 electron microscopes with Megaview III camera and analySIS 3.2 software (Soft Imaging System®).

Results

DEPOLL is unlikely to encode a transcription factor

Reverse-genetic screen for transcription factors affecting pollen development led to the identification of the At1g70790 gene. The appropriate line SAIL_1168_C11 carried the T-DNA insertion in its first intron (Fig. 1A). One of the most prominent phenotypic defects in

this line was the higher proportion of aborted pollen. That is why the protein encoded by At1g70790 was named DEPOLL (DEath POLLen grain).

The putative DEPOLL protein is 185-amino-acids long and its molecular weight is 21 kDa. The molecular, biological function and cellular localization of the protein was completely unknow. The *DEPOLL* gene was annotated as a transcription factor (TF) with C2H2-type zinc finger domain according to AtTFDB (Arabidopsis thaliana transcription factor database) on AGRIS server (Arabidopsis Gene Regulatory Information Server; http://arabidopsis.med.ohio-state.edu/AtTFDB/AtTFBrowseResults.jsp?fam=C2H2) (Palaniswamy et al. 2006; Davuluri et al. 2003). However, PROSITE (http://www.expasy.ch/prosite/) (Gasteiger et al. 2003) and I-TASSER (http://zhang.bioinformatics.ku.edu/I-TASSER/) (Zhang 2008) computational analysis of DEPOLL protein domains revealed only C2 domain with conserved calcium-membrane binding motif in the N-terminal region (amino acids 1-to-88, Fig. 1B, C). Neither C2H2-type zinc finger domain nor other protein domain with nucleic acid binding properties was identified.

Protein multiple sequence alignment of 61 C2H2 zinc finger binding domain proteins and related sequences defined in AGRIS C2H2 TF family by ClustalW2 at EMBL-EBI with default analysis parameters is presented as a Supplementary Fig. 1. Protein domains were verified and confirmed by PROSITE (http://www.expasy.ch/prosite/) (Gasteiger et al. 2003) (Supplementary Table 2). DEPOLL was clustered into distinct subcluster together with proteins At4g05330, At4g21160 and At3g07940 (red box, Supplementary Fig. 1). It is remarkable that all these genes have the strongest expression signal in late stages of pollen development (aGFP database, http://agfp.ueb.cas.cz) (Dupľáková et al. 2007). Moreover, the alignment also revealed the Arf-GAP (the ADP-ribosylation factor GTPase-activating proteins) domain in all proteins except DEPOLL. The Arf-GAP domain contains the characteristic zinc finger motif (Cys-x2-Cys-x(16,17)-x2-Cys) similar to the C4 type GATA zinc fingers. So far we found no evidence for the zinc finger domain(s) of any type in DEPOLL protein.The only positively identified domain remained C2. Therefore the putative function of DEPOLL as transcription factor is very unlikely.

DEPOLL has a specific expression profile in late stages of pollen development

According to transcriptomic data, At1g70790 expression profile was predicted as late pollen-specific. This assumption was confirmed by quantitative real-time RT-PCR using RNA isolated from four stages of male gametophyte development (unicellular microspores, bicellular pollen, immature tricellular pollen and mature pollen), three sporophytic tissues (root, leaf, stem) and whole inflorescences. qRT-PCR analysis revealed strong and cumulative *DEPOLL* expression in tricellular and mature pollen. On the contrary, mean *DEPOLL* expression in sporophytic tissues was calculated to be ~300-fold lower than in late stages of pollen development and ~120-fold lower than in microspores. Simultaneously, it was confirmed that T-DNA insertion in At1g70790 gene caused significant, ~15-fold decrease of *depoll* transcript abundance in comparison with wild type pollen. These data confirmed *DEPOLL* as a late pollen-specific gene.

DEPOLL microscopic and ultrastructural characterization

The original population of the SAIL_1168_C11 line was heterozygous. The identification of numerous the *depoll/depoll* plants within the tested F1 progeny confirmed that the presence of T-DNA insertion in *DEPOLL* gene was not lethal. However, pollen

produced by both homozygous and heterozygous plants showed characteristic phenotypic defects.

First, we defined and categorised individual phenotypic defects and counted their presence in *Arabidopsis thaliana* Col-0 wild-type mature pollen as a reference (Tab. 1). In the wild type population, the total proportion of pollen deviant from standard phenotype is about 29.36% (100 flowers analysed). In total, 10,000 pollen grains were counted. In *depoll/depoll* and *depoll/+* plants, the percentage of mutant pollen phenotypes was significantly higher. In 100 flowers on heterozygous plants, 10,000 pollen grains were observed. Of them, 41.53% showed phenotypic defects. The most prominent deviations were the presence of aborted pollen (11.83±14.01%) and the PMII defect (presence of bicellular pollen, 11.38±4.46%). In 100 *depoll* homozygous flowers (10,000 observed pollen grains), the above phenotype classes (aborted pollen, 14.06±14.57% and the PMII defect, 18±6.32%) were supplemented with two classes of pollen with misarranged MGU – misshaped (19.7±7.71%) and linear (16.9±5.73%) MGU. In total, the proportion of aberrant pollen was more than twice higher 77.42%. More thorough analysis of the distribution of pollen deviant phenotypes frequency in each category is given in Supplementary table 3. There, percentage of plants carrying the respective phenotype defect in the particular occurrency range is shown. All defects observed by bright and epifluorescence microscopy are depicted on Fig. 2.

In F1 and F2 generations, the characterised phenotype defects co-segregated with the T-DNA insertion. Moreover, in the SAIL_1168_C11 line, the T-DNA insertion in the At1g70790 gene is in the *quartet1* (*qrt1*) background that enables more thorough analysis of the distribution of phenotype defects. The genetic and phenotypic data data as well as the quartet analysis demonstrated the gametophytic mode of action of the *depoll* mutation.

Subsequently, scanning electron microscopy (SEM) and transmission electron microscopy (TEM) were employed to reveal possible phenotype defects of mature *depoll* pollen at the ultrastructural level. *depoll* pollen exine surface structure visualised by SEM did not show significant deviation from the wild type (Fig. 3). The only difference was the presence of depressions in a cell wall (Fig. 3B, D, E, F) or even collapsed pollen grains (Fig. 3C). The representation of aborted pollen was similar to that observed by bright field microscopy – 15%. The absence of visible exine ultrastructure defects corresponded with the above stated gametophytic nature of *depoll* mutation.

More interesting phenotype defects were observed by TEM. The ultrastructure of wild type and *depoll* mature pollen is shown in the figure 4. General ultrastructure of *depoll* pollen was similar to that of wild type. The representation and appearance of most organelles and other structures (mitochondria, plastids, dictyosomes, spherosomes and rough endoplasmic reticulum) was normal. However, there were significant changes observed in the intine ultrastructure and organisation. However, there were significant changes observed in the intine ultrastructure and organisation. In *depoll* pollen, the intine was greatly reduced and often nearly missing (Fig. 3D, F), passing to granular structure (Fig. 3H). In pollen grains with reduced or missing intine, plasma membrane was detached from cell wall forming so called “ghost space” (Fig. 3D, F, H). Interestingly, there was no plasma membrane detachment observed in aperture regions. This can provide an explaination why *depoll* pollen showed no germination defects (see below).

depoll plants produce less pollen

Further we investigated how the frequent pollen abortion influenced the overall pollen production in *depoll* plants. The pollen production was examined in the same number of individual anthers (45) by the calculation of pollen grains on agarose cushions later used for pollen in vivo germination assays (Boavida and McCormick 2007). Collectively, homozygous

depoll plants produced 2,767±84.8 pollen grains per 45 anthers, what stand 61,5 pollen grains per anther in comparison to 6,221±235,2 pollen grains per 45 anthers in wild type plants what stand 138,2 pollen grains per anthers. It represented over 55% reduction of pollen production in *depoll* plants.

On the base lower number of pollen grains in *depoll* plants it was suspected that plants will be carry shorter siliques than wild type ones. To verify this observation, the silique lengths were measured on wild type and *depoll* homozygous plants. In each assay, 50 siliques were measured on 50 individual plants. The mean lenght of siliques on control Col-0 plants was 13.28 ±2.81 mm that falls into the published standard range of 8-18 mm (Kubát et al. 2002). Self-pollinated of heterozygous plants were produced siliques with the mean length of 10.4±1.64 mm. We believe that such production of significantly truncated siliques could be caused by reduced pollen production in *depoll* plants.

In pollen germination assays, 5023 wild type and 2,767 *depoll* pollen grains were examined. In wild type, the average germination percentage was 83.1±20.3%. The presence of *depoll* mutant allele did not resulted in the reduction of germination efficiency as the average value was 88.4±11.4%.

depoll shows transmission distortion

The seeds produced by self-fertilised heterozygous *depoll* plants were sown on soil and the segregation ratio was calculated by genotyping. A non-Mendelian segregation (Tab. 2) suggested transmission distortion in gametophytic transmission efficiency. In analysed population (385 plants) was increased the presence of *depoll* mutant alele. This was confirmed by reciprocal crosses (Tab. 4) showing that transmission through the male was increased by 39% (128 plants) and whereas through female only by 21.6% (153 plants).

The functional complementation analysis was performed to test whether the At1g70790 genomic fragment expression is able to restore the wild type phenotype of *depoll* pollen. Homozygous *depoll* plants were transformed with a construct pKGW:gDEPOLL containing At1g70790 genomic squence (3,193 bp) cloned under its own promoter (1363 bp upstream region) and containing also 426-bp trailer. Five resistant plants from four independent transformed lines (B2/line 2, B4/line 3, G2/line4, G3/line 4, H1/line 5) with confirmed insertion were selected for pollen phenotype analyses by bright field and fluorescence microscopy. All five analysed lineds showed significant reduction of aberrant pollen frequency in all observed phenotype classes (Fig. 5). Wild type plants contained 14.4% aberrant pollen whereas in *depoll* plants this proportion was as high as 58.2%. The frequency of aberrant pollen in pKGW:gDEPOLL-complemented *depoll* lines ranged between 17.2 (line B4) to 24.8% (H1). Thus the mean complementation efficiency was 84% (76.3-93.6%).

DEPOLL knock-down triggers the activation of genes associated with cell wall metabolism

Previous sequence analyses disconfirmed DEPOLL as a transcription factor and subsequent pollen phenotype observations suggested its role in intine formation. To test this hypothesis, Affymetrix Arabidopsis ATH1 Genome Arrays were employed to compare *depoll* mature pollen gene expression with that of wild type pollen. Micorarrays were hybridised with cRNA probes synthesised from total RNA isolated from mature pollen isolated from wild type and homozygous *depoll* plants. For both samples, two biological replicates were made.

After data normalisation, fold changes in gene expression between wild type and *depoll* pollen were calculated. At1g70790 gene was on the top of the downregulated genes in *depoll* pollen confirming the effective gene knock-down by T-DNA insertion.

In *depoll* pollen, 75 genes were downregulated at least two-fold (Supplementary Table 4). On the other side, another 265 genes were at least two-fold upregulated (Supplementary Table 5). Moreover in both up- and down regulated gene subset over 2/3 of all genes had relative expression signal over the treshold value of 100 (72% genes in downregulated subset and over 68% genes in upregulated subset). These groups of down- or up- regulated genes represented only small fraction of *Arabidopsis* genome (0.37% genes in downregulated subset and 1.31% genes in upregulated subset). Accordingly, the number of upregulated genes was ~3.5-times higher than that of downregulated ones.Subsequently, we carried out the functional characterization of both gene subsets with relative expression signal over the treshold value of 100 (Supplementary Tables 6 and 7, Fig. 6). Functional classes were defined as published previously (Honys and Twell 2004). The most significant changes in the downregulated subset were observed in these functional categories: transport (24.07% in downregulated subset to 6.21% in total pollen transcriptome; 3.88-fold change) and energy (1.85 to 0.99; 1.87-fold). However, the structure of functional categories and fold changes in the upregulated subset in *depoll* pollen were more distinctive than in the downregulated subset. The most significant changes in the upregulated subset were observed in these functional cateories: genes associated with cell wall metabolism and function (9.89 to 3.81; 2.6-fold), energy (2.2 to 0.99; 2.22-fold), stress-related genes (8.79 to 4.66; 1.89-fold) and transport (8.24 to 6.21; 1.33-fold).

As *DEPOLL* belongs among late pollen genes, we speculated that late genes would be more affected by its knock-down. Therefore we functionally analysed genes with the expression signal in late stages at least 2- or 4-fold higher than in the early ones (Supplementary Tables 8 and 9). The sub-selection of late pollen genes highlighted trends suggested when all down- or upregulated genes were taken into account. The functional characterization of downregulated late genes revealed these categories: transport (25 to 6.21; 4.03-fold), energy (3.13 to 0.99; 3.16-fold), cell wall (6.25 to 3.81; 1.64-fold). The same analysis of upregulated late genes highlighted: stress (10 to 4.66; 2.15-fold), metabolism (25 to 13.59; 1.84-fold), transport (7.5 to 6.21; 1.21-fold) and protein fate (5 to 4.12; 1.21-fold change) (Fig. 6). Apparently, transport and cell wall categories were relevant in both subsets.

The assumption that late pollen genes would belong amongst the most affected by *depoll* knock-down was confirmed. Moreover, the gene ontology (GO) annotation of both down- and upregulated subsets using WEGO web tool (http://wego.genomics.org.cn/cgi-bin/wego/index.pl, (Ye et al. 2006)) (Fig. 7) provided another indirect evidence for the DEPOLL function. It highlighted DEPOLL as a putative inhibition factor in signal transduction pathway(s) or as a transporter. This categorisation was in accordance with the annotation of top genes specifically up/down-regulated in *depoll* pollen - lipid transfer proteins, invertase/pectin methylesterase inhibitor family proteins and polygalacturonase and xyloglucan endotransglucosylase/hydrolases (XTHs). Potential metabolic pathways thay may be controlled by DEPOLL protein were investigated also using the MapMan visualization tool (http://mapman.gabipd.org/web/guest (Thimm et al. 2004)). The MapMan outcome was consistent with previous computational and GO analyses; the most abundantly overexpressed up-regulated genes were scaterred in transport pathways concretely in lipid transfer proteins and cell wall metabolism (Supplementary Fig. 2 and 3).

Eleven members of the lipid transfer protein (LTP) gene family were upregulated in *depoll* pollen; several of them even belonged amongst the top upregulated genes (At2g38530, 10.68X upregulated; At3g08770, 8.40X; At4g22490, 8.28X; At5g59320, 6.67X; At5g59310, 6.18X; At5g38160, 4.72X; At5g38170, 4.69X; At5g38180, 4.51X; At3g22600, 2.84X; At2g38540, 2.44X; At4g22470, 2.36X). It represented over 22% of 49 known *Arabidopsis* nsLTP genes (Boutrot et al. 2008; Arondel et al. 2000). Most of above mentioned LTPs in *depoll* mutant belonged among type I and II nsLTPs that apparently lack the specificity for the

various phospholipids (At2g38530, At2g38540, At3g08770, At5g38160, At5g38170, At5g38180, At5g59310, At5g59320) (Boutrot et al. 2008). The others, namely At3g22600, At4g22470, At4g22490, were refused as non-specific LTPs (Boutrot et al. 2008). .Non-specific lipid transfer proteins (nsLTPs) are abundant proteins in higher plants where they often constitute up to 4% of the total soluble proteins (Kader 1996). Initially, based on their ability to facilitate *in vitro* transfer of phospholipids between membranes and to bind acyl chains, they are considered to be involved in membrane biogenesis, turnover and intracellular lipid trafficking (Molina and Garcia-Olmedo 1993; Kader 1996; Wirtz 1991; Kader et al. 1984; Kader 1975; Kader 1997).

The second distinct upregulated gene subset in *depoll* pollen comprised enzymes involved in cell wall carbohydrate metabolism, namely invertase/pectin methylesterase inhibitor proteins (At4g15750 (13.38X upregulated), At1g56100 (12.21X), At1g09370 (3.83X), At5g62350 (2.43X). Invertases are glycoproteins that cleave α1-β2-glycosidic bond in sucrose with glucose and fructose as final products. So they play a crucial role in carboydrate metabolism and sugar signaling (Godt and Roitsch 1997; Smeekens 2000; Rausch and Greiner 2004; Rolland et al. 2002; Sturm 1999). Invertase inhibitor proteins (InvI) are ubiquitously present in plants (Rausch and Greiner 2004) and probably control their activity. Pectin methylesterases (PME) belong among carbohydrate-modifying enzymes catalysing pectin deesterification (Micheli 2001). Pectin methylesterase inhibition proteins (PMEIs) represent a base building block in PME post-translational regulation (Jiang et al. 2002; Lionetti et al. 2007; An et al. 2008; Balestrieri et al. 1990). However, the precise biological function of PMEIs is almost unknown. It is highly probable that their function is the suppression of the PME-catalyzed processes.

Not only inhibitors but also carbohydrate-modyfying enzymes themselves were regulated in *depoll* pollen. Upregulated At2g41850 (9.66X) and downregulated At4g35670 (2.92X) genes belonged to the polygalacturonase (pectinase) protein family (PGs) catalyzing the hydrolysis of 1,4-alpha-D-galacturonan backbone of pectic polysaccharides and belonging among the pectin modification and disassembly enzymes. PGs are likely to be involved in modifications of the structural properties of wall-bound pectins during plant growth and developmental processes. Three other *depoll*-upregulated genes (At4g25810 (2.93X), At5g57560 (4.63X) and At4g30280 (2.69X)) were members of the xyloglucan endotransglucosylase/hydrolase (XTHs) gene family. Hemicelluloses, together with cellulose and pectins, form the main building units of the polysaccharide network in cell walls. They represent the network of xyloglucans, xylans, mannans, and arabinogalactans. As xyloglucan modifying enzymes, XTHs cut the xyloglucan backbone and ligate the fragments either to another xyloglucan (endotransglucosylases, XETs) or to water (endohydrolases, XEHs) (Nishitani and Tominaga 1992; *Rose* 2003; Farkas et al. 1992).

Taken together, *depoll* knock-down in pollen caused the upregulation of significant subset of male gametophytic genes. Most of the genes were suggested to be involved in the cell wall metabolism and communication with plasma membrane. Number of other genes were downregulated but their number and abundance was limited. Moreover, among them there were no prominent functional clusters.

Discussion

At1g70790 gene (*DEPOLL*) was described and functionally characterised within the scope of the reverse-genetic screen for transcription factors regulating the male gametophyte development. According to transcriptomic data, *DEPOLL* is a late pollen-specific gene. Its late expression profile was confirmed by quantitative RT-PCR analysis. Before identification of its downstream genes, we aimed to analyze DEPOLL domain structure and to confirm its

position amongst C2H2 zinc finger transcription factors as suggested by AtTFDB. Surprisingly, the presence of neither C2H2 zinc finger domain nor any other DNA-binding was not confirmed. On the contrary, C2 domain was identified in the DEPOLL N-terminal region (amino acids 1 to 88). C2 domain was originally described as a Ca^{2+} dependent phospholipids-binding domain identified in protein kinase C (Nalefski and Falke 1996). Subsequent finding that the C2 domain is able to bind also number of other substrates as well as the discovery of Ca^{2+} independent forms of C2 domain indicated high degree of functional variability among C2 proteins (Nalefski and Falke 1996). This variability was later confirmed by the role of C2 proteins in many signaling transduction pathways and membrane trafficking processes [5]. Currently, over a hundred of C2 domain-containing proteins were described in animals and yeast. In plants, only a few C2 proteins are known.

However, although unlikely, the putative existence of still unknown DNA-binding domain cannot be completely ruled out because of the similarity of glycine and proline distribution between DEPOLL C-terminal region and the AlkB DNA-repair enzyme from Escherichia coli that catalyzes the oxidative demethylation of 1-methyladenine and 3-methylcytosine (Trewick et al. 2002; Aas et al. 2003; Falnes et al. 2002). Currently, the possible DEPOLL DNA-binding properties are being investigated.

DEPOLL functional characterisation was performed on Syngenta T-DNA insertion line SAIL_1168_C11. The presence of the appropriate T-DNA insertion in this line was confirmed as well as the co-segregation of this T-DNA insertion with observed phenotypic defects. *depoll* pollen showed wide range of phenotypic defects (Tab. 1, Fig. 2, 3, 4) with generally high occurrence. In wild type, the frequency of aberrant pollen ranged between 14-29% and this value increased to 24-42% (*depoll*/+ plants) and even to 58-77% in homozygous *depoll* plants. The observed frequency of mutant phenotypes in pollen produced by homozygous and heterozygous plants as well as transmission data highlighted the gametophytic nature of *depoll* mutation. The most distinct phenotypic deviation was high number of dead pollen that was also reflected in the mutation name. The high penetrance of mutant phenotype was in accordance with the limited pollen production (55% reduction) in *depoll* plants, although *depoll* pollen germination efficiency was not affected.

Although frequent, observed phenotypic deviations of *depoll* pollen were not massive. This could be caused by the co-expression of orthologous, functionally redundant genes. The most likely candidates were two genes, At1g70800 and At1g70810, that share a high degree of homology with the At1g70790 gene are located in its immediate proximity forming a short array. All three genes showed similar late gametophytic expression profile. Moreover, althoung lacking the C2 domain, both At1g70800 and At1g70810 contained amino acid sequence homologus to DEPOLL C-terminal region (amino acids 98-165).

Protein multiple sequence alignment clustered DEPOLL in a distinct subcluster with Arf-GAP domain-containing proteins with DEPOLL being the only protein lacking this domain. Of co-clustered proteins, only ZAC protein (At4g21160) from *Arabidopsis thaliana* was functionally characterized. ZAC protein was localised mainly in association with plasma membrane and Golgi apparatus. The ZAC C2 domain was responsible for non-specific binding to negatively charged phospholipids, with some variance in Ca^{2+} and salt-concentration-dependent manner (Jensen et al. 2000). Moreover, most amino acids associated with ZAC C2 domain binding to phospholipids are strongly conserved in DEPOLL (Supplementary Fig. 4). Accordingly, TEM analysis revealed specific ultrastructural defect in *depoll* pollen – intine reduction and plasma membrane detachment from pollen cell wall. The only regions where plasma membrane remained attached to pollen wall were apertures and this phenomenon could explain normal pollen germination efficiency. Intine is synthesised by the gametophyte (developing microspore) after microspore release. This fact fits with above demonstrated gametophytic mode of action of DEPOLL protein.

Subsequent transcriptomic experiment identified genes with altered expression profile in *depoll* pollen. Two sets of differentially expressed genes were found – smaller group of downregulated genes and larger group comprising upregulated genes. The upregulated cluster comprised more consistent geneset encoding proteins associated with plasma membrane and cell wall metabolism and modifications, namely lipid transfer proteins, invertase/pectin methylesterase inhibitors, polygalacturonases and xyloglucan endotransglucosylases /hydrolases. As DEPOLL is unlikely to act as a transcription factor, upregulation of a specific gene subset can result from the partial complementation of cellular processes affected by *depoll* mutation. Such complementation cannot be complete because a major fraction of *depoll* pollen still showed apparent phenotype defects and reduced fertility.

The role of lipid transfer proteins in sexual plant reproduction has been repeatedly demonstrated. nsLTP were detected in the tapetum (Koltunow et al. 1990; Foster et al. 1992; Lauga et al. 2000; Ariizumi et al. 2002), stigma (Thoma et al. 1994), pistil transmitting tract (Chae et al. 2009), microspores (Foster et al. 1992), mature pollen (Thoma et al. 1994; Chae et al. 2009) and pollen tubes (Chae et al. 2009). They were shown to be involved in pollen tube adhesion on an artificial stylar matrix (Park et al. 2000; Park and Lord 2003) or in the pistil transmitting tract (Chae et al. 2009) and in pollen tube tip growth (Chae et al. 2009) as a signal transducer (Kim et al. 2006; Chae et al. 2007). Although these findings together with our data suggested the importance of lipid transfer proteins for flawless pollen development, their physiological function still remains unclear. It has been suggested that nsLTPs were involved in the deposition or transport of lipophilic material (ie. sporopollenin) in the developing pollen wall (Foster et al. 1992). Extracellular LTPs were further suggested to be the structural component of cell walls (Renan et al. 2003).

The knowledge of cell wall-bound invertase inhibitors (InvI-CW) biological function *in vivo*, like that of LTPs, is still rather limited. Interestingly, northern blot analysis one of the best characterized plant invertase inhibitors, NtCIF (cell wall inhibitor of b-fructosidase or cell wall invertase inhibitor from tobacco suspension-cultured cells) revealed its particularly strong expression in ovaries, stamens and petals (Greiner et al. 1998). NtCIF protein accumulated only during later stages of flower development and thus was co-expressed with its putative target Inv-CW (Rausch and Greiner 2004). This co-expression of InvI-CW and Inv-CW in later stages of flower development was confirmed also in tomato (Greiner et al. 2000).

Pectin methylesterase inhibition proteins represent a base building block in PME post-translational regulation (Jiang et al. 2002; Lionetti et al. 2007; An et al. 2008; Balestrieri et al. 1990) as PME gene expression is strongly regulated in a tissue-specific manner (Micheli et al. 1998; Li et al. 2002). In *Arabidopsis* pollen, two PMEI types (AtPMEI1 and 2) were discovered. Their functional analysis revealed that both PMEIs were able to inhibit PME activity in flowers and siliques (Wolf et al. 2003). Again, PMEI biological function is almost unknown. However, their function in processes opposite to those realised by PMEs (ie. termination of PME-catalyzed reactions) is highly probable. Both invertase-and pectin methylesterase inhibitors are also secreted into the cell wall and thus are presumably extracellular proteins (Hothorn et al. 2004).

Polygalacturonases represent other important protein family associated with the pollen cell wall. Their activity was long known to increase substantially during fruit ripening (Gross and Wallner 1979; *Rose* 2003) so it was deduced that PGs mediated the pectin depolymerization and constitute the enzymatic basis of softening. PGs are expressed in a wide range of plant tissues including very abundant late expression in pollen (Hadfield and Bennett 1998; *Rose* 2003; Hruba et al. 2005). Significant fraction of PGs were also shown to be pollen specific (Honys and Twell 2004). The likely role of polygalacturonases in pollen development and function involves the degradation the cell walls of the stylar cells allowing pollen tube

penetration and/or provide wall precursors for tube growth. In addition, their putative function in oligosaccharide release during pollen/pistil interactions was speculated (Brown and Crouch 1990).

The last family overrepresented among *depoll*-upregulated genes comprised xyloglucan endotransglucosylase/hydrolases that catalyse cell wall loosening reactions resulting in cell wall extension (Nishitani and Tominaga 1992; Kaku et al. 2002). Moreover, XTHs with XET activity are considered to function in cell wall reinforcing (stiffening). Further XTH roles may include cell wall restructuring during cell growth and fruit ripening, and integration of newly synthesized xyloglucans into the existing cell wall (Rose et al. 2002; Hayashi et al. 1994). This can be the function of TCH4 protein (At5g57560) (Xu et al. 1995), one of proteins significantly unpregulated in *depoll* pollen. TCH4, a member of XET-related (XTR) gene family has been shown to be strongly expressed in response to environmental stimuli (ie. touch, darkness) and the growth-stimulating hormones, auxin and brassinosteroids (Xu et al. 1996; Braam and Davis 1990).

Both polygalacturonases and pectin methylesterase inhibitor proteins are linked to the pectin metabolism within the cell wall. Both protein families contribute to the cell wall loosening and relaxation, PMEIs by the inhibition of PME expression whereas by PGs direct pectin hydrolysis. As pectins are integral components of intine, distortion of their metabolism results in intine formation defects. Although *DEPOLL* expression starts already in microspores (so DEPOLL can be involved in intine synthesis), it peaks in mature pollen. Therefore its main role is likely to be accomplished during pollen tube growth, possibly in relation to the synthesis of pectin pollen tube wall as suggested also by progamic phase abnormalities of *depoll* pollen.

In *depoll* population was increased the presence of *depoll* mutant allele. In agreement with this results we suggested that not presence of DEPOLL protein lead to higher expression of many genes with relevance to cell wall metabolism. Their coordinated activities are distinguishable in elevated mutant pollen tube fitness.

Taken together, DEPOLL (At1g70790) represents the first C2 domain protein with demonstrated function in male gametophyte development. DEPOLL activity is essential for the correct intine formation and its knock-down 1) results in wide range of pollen phenotypic defects including abortion and increased pollen tube growth and fertility and 2) is accompanied by abundant expression of number of lipid transfer proteins and proteins involved in pollen cell wall metabolism. DEPOLL is suggested to be extracellular membrane-associated or cell wall protein involved in several metabolic pathways either as structural but more likely as a regulatory protein active during pollen dvelopment and subsequent progamic phase.

Acknowledgments

Authors gratefully acknowledge the financial support from the **Czech Science Foundation** (grant 522/09/0858) and **Ministry of Education, Youth and Sports of the Czech Republic** (grants LC06004 and OC10054).

References

Aarts MG, Hodge R, Kalantidis K, Florack D, Wilson ZA, Mulligan BJ, Stiekema WJ, Scott R, Pereira A (1997) The arabidopsis male sterility 2 protein shares similarity with reductases in elongation/condensation complexes. Plant J 12 (3):615-623

Aas PA, Otterlei M, Falnes PO, Vagbo CB, Skorpen F, Akbari M, Sundheim O, Bjoras M, Slupphaug G, Seeberg E, Krokan HE (2003) Human and bacterial oxidative

demethylases repair alkylation damage in both rna and DNA. Nature 421 (6925):859-863. doi:10.1038/nature01363 nature01363 [pii]

An SH, Sohn KH, Choi HW, Hwang IS, Lee SC, Hwang BK (2008) Pepper pectin methylesterase inhibitor protein capmei1 is required for antifungal activity, basal disease resistance and abiotic stress tolerance. Planta 228 (1):61-78. doi:10.1007/s00425-008-0719-z

Ariizumi T, Amagai M, Shibata D, Hatakeyama K, Watanabe M, Toriyama K (2002) Comparative study of promoter activity of three anther-specific genes encoding lipid transfer protein, xyloglucan endotransglucosylase/hydrolase and polygalacturonase in transgenic arabidopsis thaliana. Plant cell reports 21 (1):90-96

Ariizumi T, Hatakeyama K, Hinata K, Inatsugi R, Nishida I, Sato S, Kato T, Tabata S, Toriyama K (2004a) 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. Plant J 39 (2):170-181. doi:10.1111/j.1365-313X.2004.02118.x TPJ2118 [pii]

Ariizumi T, Hatakeyama K, Hinata K, Inatsugi R, Nishida I, Sato S, Kato T, Tabata S, Toriyama K (2004b) 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

Ariizumi T, Hatakeyama K, Hinata K, Sato S, Kato T, Tabata S, Toriyama K (2003) A novel male-sterile mutant of arabidopsis thaliana, faceless pollen-1, produces pollen with a smooth surface and an acetolysis-sensitive exine. Plant molecular biology 53 (1-2):107-116. doi:10.1023/B:PLAN.0000009269.97773.70 5149836 [pii]

Ariizumi T, Kawanabe T, Hatakeyama K, Sato S, Kato T, Tabata S, Toriyama K (2008) Ultrastructural characterization of exine development of the transient defective exine 1 mutant suggests the existence of a factor involved in constructing reticulate exine architecture from sporopollenin aggregates. Plant & cell physiology 49 (1):58-67. doi:pcm167 [pii] 10.1093/pcp/pcm167

Arondel VV, Vergnolle C, Cantrel C, Kader J (2000) Lipid transfer proteins are encoded by a small multigene family in arabidopsis thaliana. Plant Sci 157 (1):1-12. doi:S0168945200002326 [pii]

Balestrieri C, Castaldo D, Giovane A, Quagliuolo L, Servillo L (1990) A glycoprotein inhibitor of pectin methylesterase in kiwi fruit (actinidia chinensis). Eur J Biochem 193 (1):183-187

Bedinger P (1992) The remarkable biology of pollen. The Plant cell 4 (8):879-887. doi:10.1105/tpc.4.8.879 4/8/879 [pii]

Boavida LC, McCormick S (2007) Temperature as a determinant factor for increased and reproducible in vitro pollen germination in arabidopsis thaliana. Plant J 52 (3):570-582

Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijo JA, Becker JD (2008) Comparative transcriptomics of arabidopsis sperm cells. Plant physiology 148 (2):1168-1181. doi:pp.108.125229 [pii] 10.1104/pp.108.125229

Boutrot F, Chantret N, Gautier MF (2008) Genome-wide analysis of the rice and arabidopsis non-specific lipid transfer protein (nsltp) gene families and identification of wheat nsltp genes by est data mining. BMC genomics 9:86. doi:1471-2164-9-86 [pii] 10.1186/1471-2164-9-86

Braam J, Davis RW (1990) Rain-, wind-, and touch-induced expression of calmodulin and calmodulin-related genes in arabidopsis. Cell 60 (3):357-364. doi:0092-8674(90)90587-5 [pii]

Brown SM, Crouch ML (1990) Characterization of a gene family abundantly expressed in oenothera organensis pollen that shows sequence similarity to polygalacturonase. The Plant cell 2 (3):263-274. doi:10.1105/tpc.2.3.263 2/3/263 [pii]

Clough SJ, Bent AF (1998) Floral dip: A simplified method for agrobacterium-mediated transformation of arabidopsis thaliana. Plant J 16 (6):735-743

Corbolán-García S, Gómez-Fernández JC (2010) The c2 domains of classical and novel pkcs as versatile decoders of membrane signals. BioFactors 36 (1):1-7

Coussens L, Parker PJ, Rhee L, Yang-Feng TL, Chen E, Waterfield MD, Francke U, Ullrich A (1986) Multiple, distinct forms of bovine and human protein kinase c suggest diversity in cellular signaling pathways. Science (New York, NY 233 (4766):859-866

Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003) Agris: Arabidopsis gene regulatory information server, an information resource of arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics 4 (1):25

Dempsey EC, Newton AC, Mochly-Rosen D, Fields AP, Reyland ME, Insel PA, Messing RO (2000) Protein kinase c isozymes and the regulation of diverse cell responses. Am J Physiol Lung Cell Mol Physiol 279 (3):L429-438

Dupľáková N, Reňák D, Hovanec P, Honysová B, Twell D, Honys D (2007) Arabidopsis gene family profiler (agfp)--user-oriented transcriptomic database with easy-to-use graphic interface. BMC plant biology 7:39. doi:1471-2229-7-39 [pii] 10.1186/1471-2229-7-39

Falnes PO, Johansen RF, Seeberg E (2002) Alkb-mediated oxidative demethylation reverses DNA damage in escherichia coli. Nature 419 (6903):178-182. doi:10.1038/nature01048 nature01048 [pii]

Farkas V, Sulova Z, Stratilova E, Hanna R, Maclachlan G (1992) Cleavage of xyloglucan by nasturtium seed xyloglucanase and transglycosylation to xyloglucan subunit oligosaccharides. Arch Biochem Biophys 298 (2):365-370

Foster GD, Robinson SW, Blundell RP, Roberts MR, Hodge R (1992) A brassica napus mrna encoding a protein homologous to phospholipid transfer proteins, is expressed specifically in the tapetum and developing microspores. Plant Sci 84:184-192

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) Expasy: The proteomics server for in-depth protein knowledge and analysis. Nucleic acids research 31 (13):3784-3788

Gibalova A, Renak D, Matczuk K, Dupl'akova N, Chab D, Twell D, Honys D (2009) Atbzip34 is required for arabidopsis pollen wall patterning and the control of several metabolic pathways in developing pollen. Plant molecular biology 70 (5):581-601. doi:10.1007/s11103-009-9493-y

Godt DE, Roitsch T (1997) Regulation and tissue-specific distribution of mrnas for three extracellular invertase isoenzymes of tomato suggests an important function in establishing and maintaining sink metabolism. Plant physiology 115 (1):273-282. doi:115/1/273 [pii]

Greiner S, Köster U, Lauer K, Rosenkranz H, Vogel R, Rausch T (2000) Plant invertase inhibitors: Expression in cell culture and during plant development. Aust J Plant Physiol 27:807-814

Greiner S, Krausgrill S, Rausch T (1998) Cloning of a tobacco apoplasmic invertase inhibitor. Proof of function of the recombinant protein and expression analysis during plant development. Plant physiology 116 (2):733-742

Gross KC, Wallner SJ (1979) Degradation of cell wall polysaccharides during tomato fruit ripening. Plant physiology 63 (1):117-120

Guan YF, Huang XY, Zhu J, Gao JF, Zhang HX, Yang ZN (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 (2):852-863. doi:pp.108.118026 [pii] 10.1104/pp.108.118026

Hadfield KA, Bennett AB (1998) Polygalacturonases: Many genes in search of a function. Plant physiology 117 (2):337-343

Hayashi T, Takeda T, Ogawa K, Mitsuishi Y (1994) Effects of the degree of polymerization on the binding of xyloglucans to cellulose. Plant & cell physiology 35 (6):893-899

Honys D, Reňák D, Twell D (2006) Male gametophyte development and function. In: Teixeira da Silva JA (ed) Floriculture, ornamental and plant biotechnology, vol I. Global Science Books, London

Honys D, Twell D (2003) Comparative analysis of the arabidopsis pollen transcriptome. Plant physiology 132 (2):640-652. doi:10.1104/pp.103.020925 132/2/640 [pii]

Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in arabidopsis. Genome biology 5 (11):R85

Hothorn M, Wolf S, Aloy P, Greiner S, Scheffzek K (2004) Structural insights into the target specificity of plant invertase and pectin methylesterase inhibitory proteins. The Plant cell 16 (12):3437-3447. doi:tpc.104.025684 [pii] 10.1105/tpc.104.025684

Hruba P, Honys D, Twell D, Capkova V, Tupy J (2005) Expression of beta-galactosidase and beta-xylosidase genes during microspore and pollen development. Planta 220 (6):931-940. doi:10.1007/s00425-004-1409-0

Huang L, Cao J, Zhang A, Ye Y, Zhang Y, Liu T (2009) The polygalacturonase gene bcmf2 from brassica campestris is associated with intine development. Journal of experimental botany 60 (1):301-313. doi:ern295 [pii] 10.1093/jxb/ern295

Chae K, Kieslich CA, Morikis D, Kim SC, Lord EM (2009) A gain-of-function mutation of arabidopsis lipid transfer protein 5 disturbs pollen tube tip growth and fertilization. The Plant cell 21 (12):3902-3914. doi:tpc.109.070854 [pii] 10.1105/tpc.109.070854

Chae K, Zhang K, Zhang L, Morikis D, Kim ST, Mollet JC, de la Rosa N, Tan K, Lord EM (2007) Two sca (stigma/style cysteine-rich adhesin) isoforms show structural differences that correlate with their levels of in vitro pollen tube adhesion activity. The Journal of biological chemistry 282 (46):33845-33858. doi:M703997200 [pii] 10.1074/jbc.M703997200

Jensen RB, Lykke-Andersen K, Frandsen GI, Nielsen HB, Haseloff J, Jespersen HM, Mundy J, Skriver K (2000) Promiscuous and specific phospholipid binding by domains in zac, a membrane-associated arabidopsis protein with an arf gap zinc finger and a c2 domain. Plant molecular biology 44 (6):799-814

Jiang CM, Li CP, Chang JC, Chang HM (2002) Characterization of pectinesterase inhibitor in jelly fig (ficus awkeotsang makino) achenes. J Agric Food Chem 50 (17):4890-4894. doi:jf011568+ [pii]

Kader JC (1975) Proteins and the intracellular exchange of lipids. I. Stimulation of phospholipid exchange between mitochondria and microsomal fractions by proteins isolated from potato tuber. Biochimica et biophysica acta 380 (1):31-44. doi:0005-2760(75)90042-9 [pii]

Kader JC (1996) Lipid-transfer proteins in plants. Annu Rev Plant Physiol Plant Mol Biol 47:627-654. doi:10.1146/annurev.arplant.47.1.627

Kader JC (1997) Lipid-transfer proteins: A puzzling family of plant proteins. Trends in plant science 2:66-70

Kader JC, Julienne M, Vergnolle C (1984) Purification and characterization of a spinach-leaf protein capable of transferring phospholipids from liposomes to mitochondria or chloroplasts. Eur J Biochem 139 (2):411-416

Kaku T, Tabuchi A, Wakabayashi K, Kamisaka S, Hoson T (2002) Action of xyloglucan hydrolase within the native cell wall architecture and its effect on cell wall extensibility in azuki bean epicotyls. Plant & cell physiology 43 (1):21-26

Karimi M, De Meyer B, Hilson P (2005) Modular cloning in plant cells. Trends in plant science 10 (3):103-105. doi:S1360-1385(05)00024-5 [pii] 10.1016/j.tplants.2005.01.008

Kim CY, Koo YD, Jin JB, Moon BC, Kang CH, Kim ST, Park BO, Lee SY, Kim ML, Hwang I, Kang KY, Bahk JD, Cho MJ (2003) Rice c2-domain proteins are induced and translocated to the plasma membrane in response to a fungal elicitor. Biochemistry 42 (40):11625-11633. doi:10.1021/bi034576n

Kim ST, Zhang K, Dong J, Lord EM (2006) Exogenous free ubiquitin enhances lily pollen tube adhesion to an in vitro stylar matrix and may facilitate endocytosis of sca. Plant physiology 142 (4):1397-1411. doi:pp.106.086801 [pii] 10.1104/pp.106.086801

Kim YC, Kim SY, Choi D, Ryu CM, Park JM (2008) Molecular characterization of a pepper c2 domain-containing src2 protein implicated in resistance against host and non-host pathogens and abiotic stresses. Planta 227 (5):1169-1179. doi:10.1007/s00425-007-0680-2

Koltunow AM, Truettner J, Cox KH, Wallroth M, Goldberg RB (1990) Different temporal and spatial gene expression patterns occur during anther development. The Plant cell 2 (12):1201-1224. doi:10.1105/tpc.2.12.1201 2/12/1201 [pii]

Krauss G (2003) Biochemistry of signal transduction and regulation. WILEY-VCH Verlag

Kubát K, Hrouda L, Chrtek Jj, Kaplan Z, Kirschner J, Štěpánek J (2002) Klíč ke květeně české republiky. Academia, Praha

Lauga B, Charbonnel-Campaa L, Combes D (2000) Characterization of mzm3-3, a zea mays tapetum-specific transcript. Plant Sci 157 (1):65-75

Li C, Hung Wong W (2001) Model-based analysis of oligonucleotide arrays: Model validation, design issues and standard error application. Genome biology 2 (8):RESEARCH0032

Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. Proceedings of the National Academy of Sciences of the United States of America 98 (1):31-36

Li YQ, Mareck A, Faleri C, Moscatelli A, Liu Q, Cresti M (2002) Detection and localization of pectin methylesterase isoforms in pollen tubes of nicotiana tabacum l. Planta 214 (5):734-740. doi:10.1007/s004250100664

Lionetti V, Raiola A, Camardella L, Giovane A, Obel N, Pauly M, Favaron F, Cervone F, Bellincampi D (2007) Overexpression of pectin methylesterase inhibitors in arabidopsis restricts fungal infection by botrytis cinerea. Plant physiology 143 (4):1871-1880. doi:pp.106.090803 [pii] 10.1104/pp.106.090803

Micheli F (2001) Pectin methylesterases: Cell wall enzymes with important roles in plant physiology. Trends in plant science 6 (9):414-419. doi:S1360-1385(01)02045-3 [pii]

Micheli F, Holliger C, Goldberg R, Richard L (1998) Characterization of the pectin methylesterase-like gene atpme3: A new member of a gene family comprising at least 12 genes in arabidopsis thaliana. Gene 220 (1-2):13-20

Molina A, Garcia-Olmedo F (1993) Developmental and pathogen-induced expression of three barley genes encoding lipid transfer proteins. Plant J 4 (6):983-991

Nalefski EA, Falke JJ (1996) The c2 domain calcium-binding motif: Structural and functional diversity. Protein Sci 5 (12):2375-2390. doi:10.1002/pro.5560051201

Nishitani K, Tominaga R (1992) Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyzes transfer of a segment of xyloglucan molecule to

another xyloglucan molecule. The Journal of biological chemistry 267 (29):21058-21064

Owen HA, Makaroff CA (1995) Ultrastructure of microsporogenesis and microgametogenesis in *arabidopsis thaliana* (l.) heynh. Ecotype wassilewskija (brassicaceae). Protoplasma 185:7-21

Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri RV, Grotewold E (2006) Agris and atregnet. A platform to link cis-regulatory elements and transcription factors into regulatory networks. Plant physiology 140 (3):818-829

Park SK, Howden R, Twell D (1998) The arabidopsis thaliana gametophytic mutation gemini pollen1 disrupts microspore polarity, division asymmetry and pollen cell fate. Development (Cambridge, England) 125 (19):3789-3799

Park SY, Jauh GY, Mollet JC, Eckard KJ, Nothnagel EA, Walling LL, Lord EM (2000) A lipid transfer-like protein is necessary for lily pollen tube adhesion to an in vitro stylar matrix. The Plant cell 12 (1):151-164

Park SY, Lord EM (2003) Expression studies of sca in lily and confirmation of its role in pollen tube adhesion. Plant molecular biology 51 (2):183-189

Paxson-Sowders DM, Dodrill CH, Owen HA, Makaroff CA (2001) Dex1, a novel plant protein, is required for exine pattern formation during pollen development in arabidopsis. Plant physiology 127 (4):1739-1749

Perin MS, Fried VA, Mignery GA, Jahn R, Sudhof TC (1990) Phospholipid binding by a synaptic vesicle protein homologous to the regulatory region of protein kinase c. Nature 345 (6272):260-263. doi:10.1038/345260a0

Piffanelli P, Ross JHE, Murphy DJ (1998) Biogenesis and function of the lipidic structures of pollen grains. Sexual plant reproduction 11:65-80

Pina C, Pinto F, Feijo JA, Becker JD (2005) Gene family analysis of the arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant physiology 138 (2):744-756. doi:pp.104.057935 [pii] 10.1104/pp.104.057935

Qin Y, Leydon AR, Manziello A, Pandey R, Mount D, Denic S, Vasic B, Johnson MA, Palanivelu R (2009) Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. PLoS genetics 5 (8):e1000621. doi:10.1371/journal.pgen.1000621

Rausch T, Greiner S (2004) Plant protein inhibitors of invertases. Biochimica et biophysica acta 1696 (2):253-261. doi:10.1016/j.bbapap.2003.09.017 S1570963903003352 [pii]

Renan M, Francois J, Marion D, Axelos MAV, Douliez JP (2003) Study of the interaction between end-capped telechelic polymers and the wheat lipid transfer protein ltp1, in solution and at the air/water interface. Colloids Surf B.32:213-221

Rhee SG, Suh PG, Ryu SH, Lee SY (1989) Studies of inositol phospholipid-specific phospholipase c. Science (New York, NY 244 (4904):546-550

Rizo J, Sudhof TC (1998) C2-domains, structure and function of a universal ca2+-binding domain. The Journal of biological chemistry 273 (26):15879-15882

Rolland F, Moore B, Sheen J (2002) Sugar sensing and signaling in plants. The Plant cell 14 Suppl:S185-205

Rose JK, Braam J, Fry SC, Nishitani K (2002) The xth family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature. Plant & cell physiology 43 (12):1421-1435

Rose JKC (2003) The plant cell wall. Blackwell Publishing

Shirataki H, Kaibuchi K, Sakoda T, Kishida S, Yamaguchi T, Wada K, Miyazaki M, Takai Y (1993) Rabphilin-3a, a putative target protein for smg p25a/rab3a p25 small gtp-

binding protein related to synaptotagmin. Molecular and cellular biology 13 (4):2061-2068
Smeekens S (2000) Sugar-induced signal transduction in plants. Annu Rev Plant Physiol Plant Mol Biol 51:49-81. doi:10.1146/annurev.arplant.51.1.49
Smyth DR, Bowman JL, Meyerowitz EM (1990) Early flower development in arabidopsis. The Plant cell 2 (8):755-767. doi:10.1105/tpc.2.8.755 2/8/755 [pii]
Sturm A (1999) Invertases. Primary structures, functions, and roles in plant development and sucrose partitioning. Plant physiology 121 (1):1-8
Suzuki T, Masaoka K, Nishi M, Nakamura K, Ishiguro S (2008) Identification of kaonashi mutants showing abnormal pollen exine structure in arabidopsis thaliana. Plant & cell physiology 49 (10):1465-1477. doi:pcn131 [pii] 10.1093/pcp/pcn131
Taylor PE, Glover JA, Lavithis M, Craig S, Singh MB, Knox RB, Dennis ES, Chaudhury AM (1998) Genetic control of male fertility in arabidopsis thaliana: Structural analyses of postmeiotic developmental mutants. Planta 205 (4):492-505
Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) Mapman: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37 (6):914-939. doi:2016 [pii]
Thoma S, Hecht U, Kippers A, Botella J, De Vries S, Somerville C (1994) Tissue-specific expression of a gene encoding a cell wall-localized lipid transfer protein from arabidopsis. Plant physiology 105 (1):35-45. doi:105/1/35 [pii]
Trahey M, Wong G, Halenbeck R, Rubinfeld B, Martin GA, Ladner M, Long CM, Crosier WJ, Watt K, Koths K, et al. (1988) Molecular cloning of two types of gap complementary DNA from human placenta. Science (New York, NY 242 (4886):1697-1700
Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B (2002) Oxidative demethylation by escherichia coli alkb directly reverts DNA base damage. Nature 419 (6903):174-178. doi:10.1038/nature00908 nature00908 [pii]
Van Aelst AC, Van Went JL (1992) Ultrastructural immuno-localization of pectins and glycoproteins in *arabidopsis thaliana* pollen grains. Protoplasma 168:14-19
Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in arabidopsis. Plant physiology 148 (3):1201-1211. doi:pp.108.126375 [pii] 10.1104/pp.108.126375
Weigel D, Glazebrook J (2002) Arabidopsis: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
Wirtz KW (1991) Phospholipid transfer proteins. Annu Rev Biochem 60:73-99. doi:10.1146/annurev.bi.60.070191.000445
Wolf S, Grsic-Rausch S, Rausch T, Greiner S (2003) Identification of pollen-expressed pectin methylesterase inhibitors in arabidopsis. FEBS Lett 555 (3):551-555. doi:S0014579303013449 [pii]
Xu W, Campbell P, Vargheese AK, Braam J (1996) The arabidopsis xet-related gene family: Environmental and hormonal regulation of expression. Plant J 9 (6):879-889
Xu W, Purugganan MM, Polisensky DH, Antosiewicz DM, Fry SC, Braam J (1995) Arabidopsis tch4, regulated by hormones and the environment, encodes a xyloglucan endotransglycosylase. The Plant cell 7 (10):1555-1567. doi:10.1105/tpc.7.10.1555 7/10/1555 [pii]
Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L (2006) Wego: A web tool for plotting go annotations. Nucleic acids research 34 (Web Server issue):W293-297. doi:34/suppl_2/W293 [pii] 10.1093/nar/gkl031

Zhang Y (2008) I-tasser server for protein 3d structure prediction. BMC Bioinformatics 9:40. doi:1471-2105-9-40 [pii] 10.1186/1471-2105-9-40

Zimmermann P, Hennig L, Gruissem W (2005) Gene-expression analysis and network discovery using genevestigator. Trends in plant science 10 (9):407-409. doi:S1360-1385(05)00170-6 [pii] 10.1016/j.tplants.2005.07.003

Zinkl GM, Zwiebel BI, Grier DG, Preuss D (1999) Pollen-stigma adhesion in arabidopsis: A species-specific interaction mediated by lipophilic molecules in the pollen exine. Development (Cambridge, England) 126 (23):5431-5440

Tables

Tab. 1 The level of phenotypical defects in *depoll* mutant, *Arabidopsis thaliana* Col-0 as a reference sample and heterozygous *depoll*/+ plants (%) with calculated standard deviation (MGU male germ unit)

Tab. 2 Self progeny segregation of heterozygous *depoll*/+ plants

Tab. 3 Male gametophyte-specific effect of *depoll* mutation

Figure Legends

Fig. 1 (A) At1g70790 gene model including T-DNA insertion site (*triangle*) and positions of introns-*white boxes*, exon-*black boxes*, untranslated regions-*light grey boxes*, LB and RB-*left* and *right borders* of T-DNA **(B)** Fasta format of At1g70790 protein sequence **(C)** Graphical view of At1g70790 C2 protein domains in 1 – 88 aa region by Prosite

Fig.2 Phenotypic defects in *depoll* pollen by bright field and epifluorescence microscopy. **(A, D, G, J)** *depoll* mutant with 2, 3 or all collapsed pollen grains in tetrads, **(B)** fluorescence images after DAPI-staining of wild type, (**C, E-F, H-I, K-L**) *depoll* pollen with various type of defects

Fig. 3 Scanning electron micrographs of wild type pollen (**A**) and *depoll* pollen grains (**B-F**). *depoll* pollen has depressions in a cell wall (**B, D, E, F**) or even collapsed pollen grains (**C**)

Fig.4 Transmission electron micrographs of mature wild type (**A, C, E, G**) and *depoll* (**B, D, F, H**) pollen grains. *depoll* pollen grains are totally collapsed (**B**). Mutant pollen has greatly reduced and often nearly missing intine (**D, F**), sometimes passing to granular structure (**H**). In *depoll* pollen is plasma membrane detach from cell wall forming so called "ghost space" (**D, F, H**)

Fig. 5 Maximal pollen defects (%) in aech complement lines (B2, B4, G2, G3, H1) in compare to *dpl* mutant and *Arabidopsis thaliana* Col-0.

Fig. 6 Proportional representation of expressed mRNA among gene function categories. Data is presented for up and down regulated genes in *depoll* pollen in comparison with wt

Fig. 7 Gene ontology (GO) annotation of both down- and upregulated subsets using WEGO web tool (http://wego.genomics.org.cn/cgi-bin/wego/index.pl)

Figure 1

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

(A)

(B)

MEDKPLGILRVHVKRGINLAIRDATTSDPYVVITLANQKLKTRVINNNCNPVWNEQL
TLSIKDVNDPIRLTVFDKDRFSGDDKMGDAEIDFRPFLEAHQMELDFQKLPNGCAIKR
IRPGRTNCLAEESSITWSNGKIMQEMILRLKNVECGEVELMLEWTDGPGCKGLGREG
SKKTPWMPTKRLD

(C)

Figure 2

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

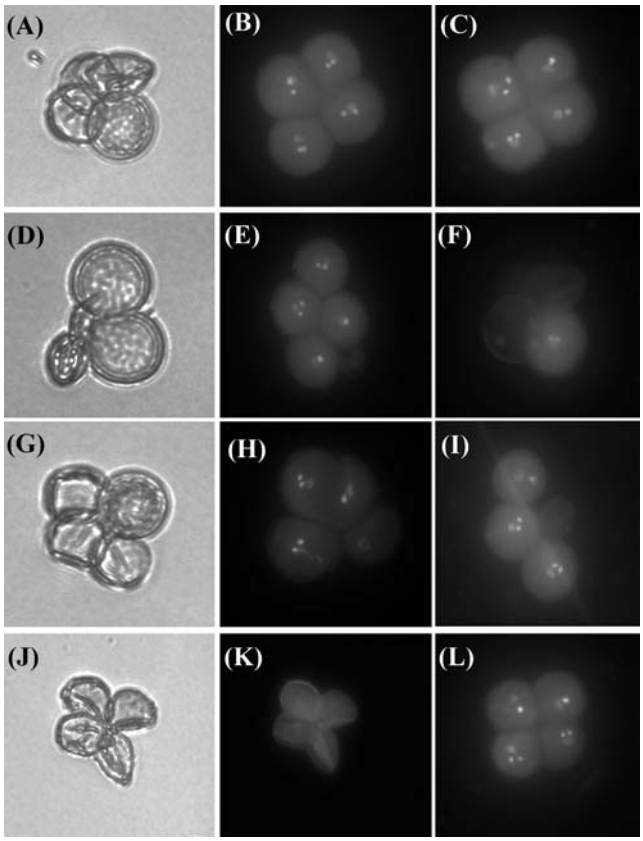

(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
(K)
(L)

Figure 3

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

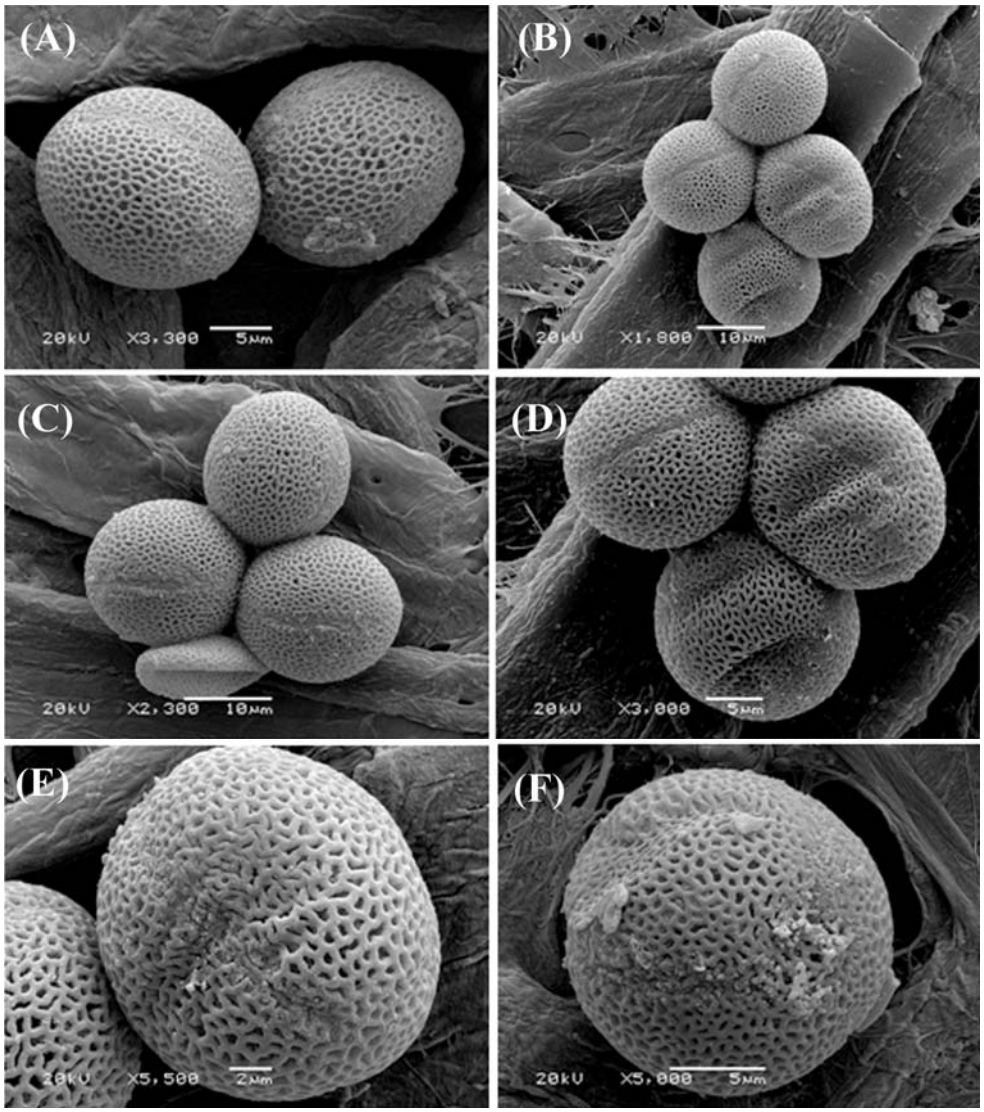

(A)
20kV X3,300 5μm
(B)
20kV X1,800 10μm
(C)
20kV X2,300 10μm
(D)
20kV X3,000 5μm
(E)
20kV X5,500 2μm
(F)
20kV X5,000 5μm

Figure 4

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

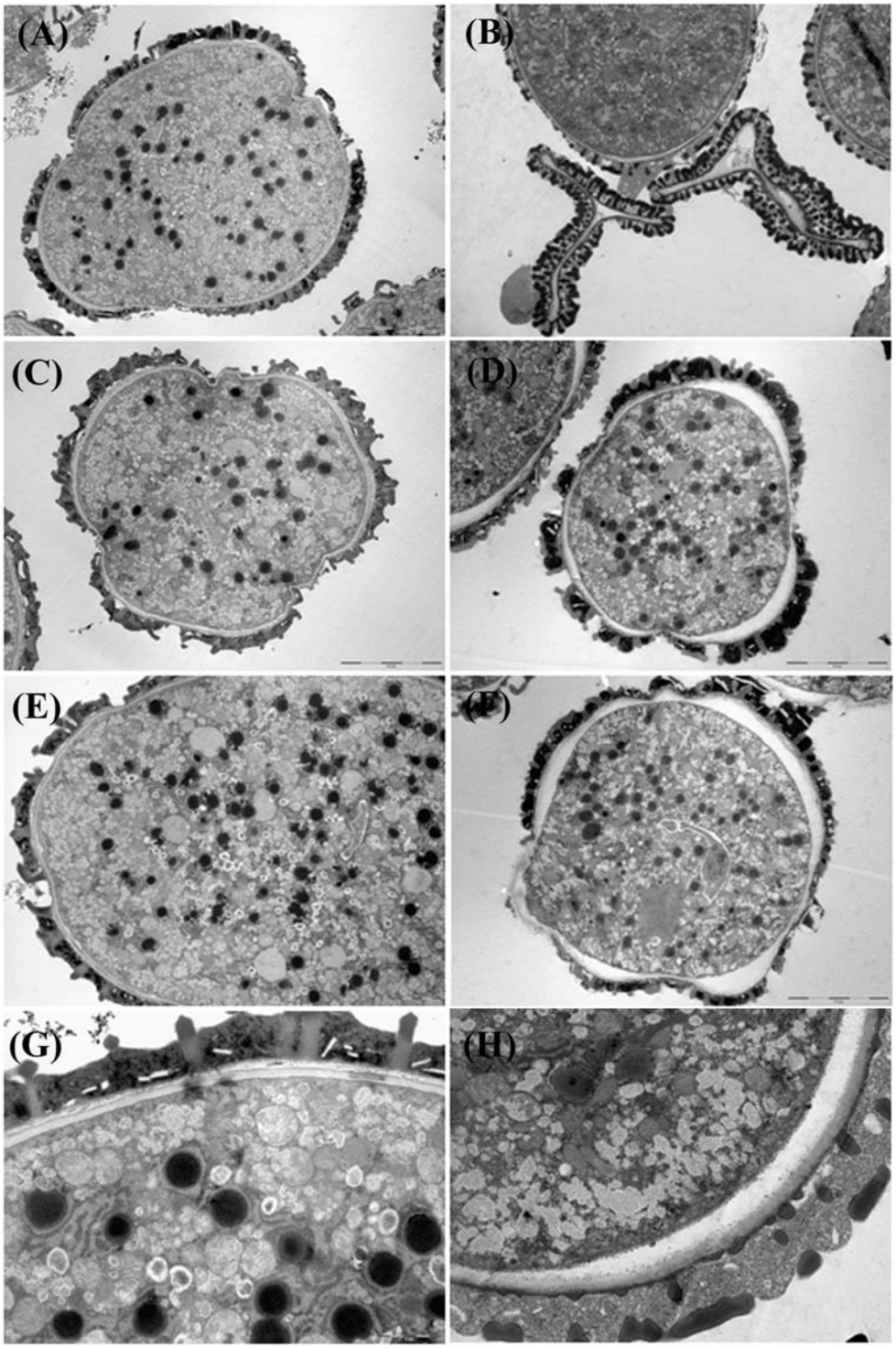

(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)

Figure 5

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

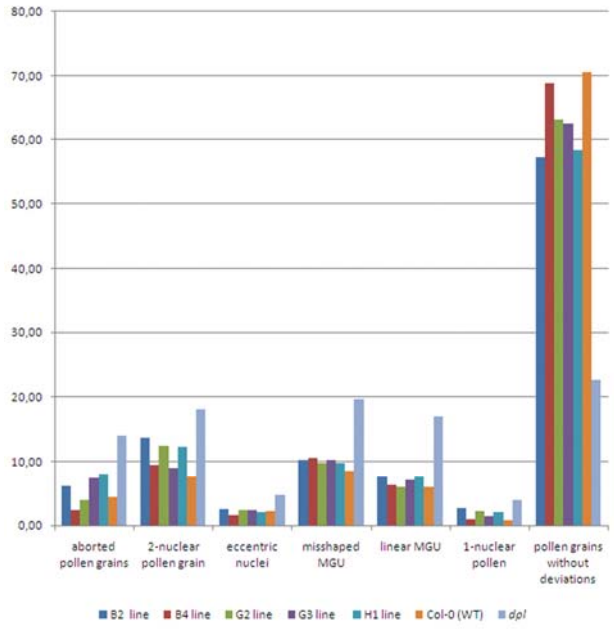

80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
aborted pollen grains
2-nuclear pollen grain
eccentric nuclei
misshaped MGU
linear MGU
1-nuclear pollen
pollen grains without deviations
B2 line
B4 line
G2 line
G3 line
H1 line
Col-0 (WT)
dpl

Figure 6

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

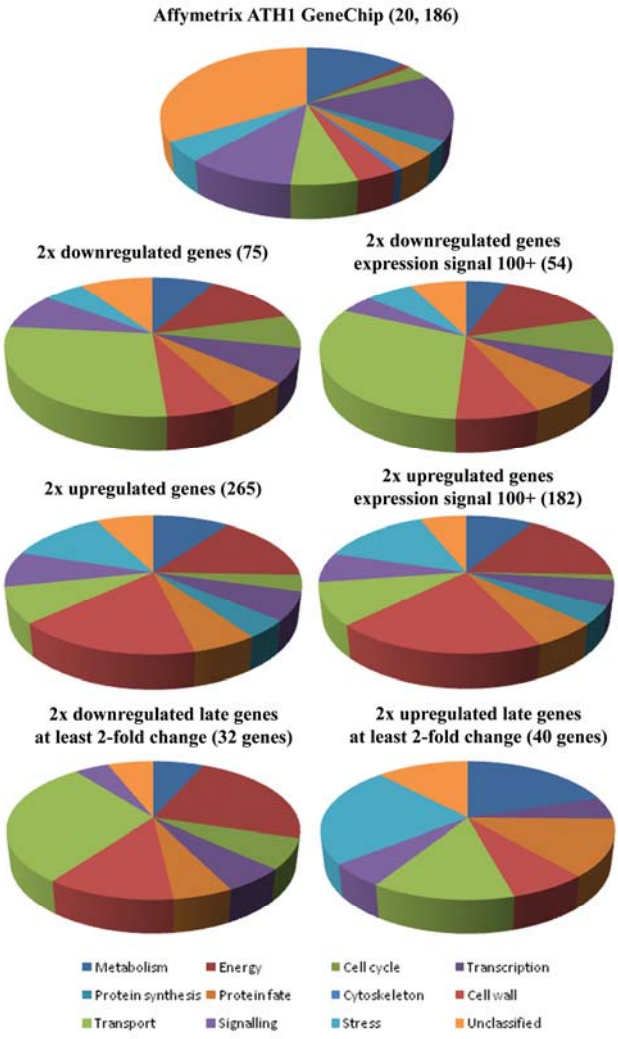

Affymetrix ATH1 GeneChip (20, 186)
2x downregulated genes (75)
2x downregulated genes expression signal 100+ (54)
2x upregulated genes (265)
2x upregulated genes expression signal 100+ (182)
2x downregulated late genes at least 2-fold change (32 genes)
2x upregulated late genes at least 2-fold change (40 genes)
Metabolism
Energy
Cell cycle
Transcription
Protein synthesis
Protein fate
Cytoskeleton
Cell wall
Transport
Signalling
Stress
Unclassified

Figure 7

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

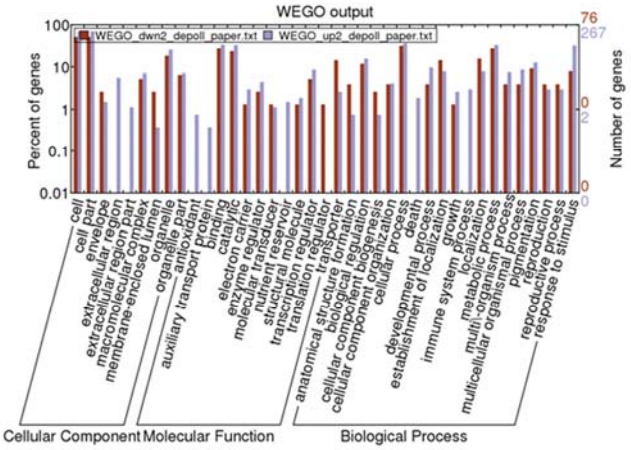

WEGO output
WEGO_dwn2_depoll_paper.txt
WEGO_up2_depoll_paper.txt
Percent of genes
Number of genes
100
10
1
0.1
0.01
76
267
0
2
0
0
cell
cell part
envelope
extracellular region
extracellular region part
macromolecular complex
membrane enclosed lumen
organelle
organelle part
antioxidant
auxiliary transport protein
binding
catalytic
electron carrier
enzyme regulator
molecular transducer
nutrient reservoir
structural molecule
transcription regulator
translation regulator
transporter
anatomical structure formation
biological regulation
cellular component biogenesis
cellular component organization
cellular process
developmental process
establishment of localization
death
immune system process
growth
localization
metabolic process
multi-organism process
multicellular organismal process
pigmentation
reproduction
reproductive process
response to stimulus
Cellular Component
Molecular Function
Biological Process

Table 1

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

Phenotypical defects	*depoll*	Col-0 (WT)	*depoll*/+
aborted pollen grains	14.06±14.57	4.44±2.26	11.83±14.01
2-nuclear pollen grain	18±6.32	7.56±2.67	11.38±4.46
eccentric nuclei	4.76±5.21	2.15±2.37	3.09±2.87
misshaped MGU	19.7±7.71	8.56±3.65	7.91±3.29
linear MGU	16.9±5.73	5.92±3.78	4.15±2.42
1-nuclear pollen	4±3.97	0.73±1.41	3.17±2.60
pollen grains without deviations	22.58	70.64	58.47

Table 2

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

Parental cross	Number of WT (+/+) progeny	Number of *depoll*/+ progeny	Number of *depoll*/*depoll* progeny	Total
depoll/+ × *depoll*/+	73	197	115	385

Table 3

C2-domain protein DEPOLL is required for intine development in *Arabidopsis* pollen

Journal name: Planta

Author names: Nikoleta Dupľáková

Author for correspondence: D. Honys, e-mail: honys@ueb.cas.cz

D. Honys
Laboratory of Pollen Biology,
Institute of Experimental Botany AS CR, v.v.i.,
Rozvojová 263, 165 00 Praha 6, Czech Republic

D. Honys
Department of Plant Experimental Biology,
Faculty of Science, Charles University in Prague,
Viničná 5, 128 44, Praha 2, Czech Republic

Parental cross	Number of WT (+/+) progeny	Number of *depoll*/+ progeny	Total
+/+ female × *depoll*/+ male	39	89	128
depoll/+ female × +/+ male	60	93	153

9. References

Alonso R, Onate-Sanchez L, Weltmeier F, Ehlert A, Diaz I, Dietrich K, Vicente-Carbajosa J, Droge-Laser W (2009) A pivotal role of the basic leucine zipper transcription factor bzip53 in the regulation of arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. The Plant cell 21 (6):1747-1761. doi:tpc.108.062968 [pii] 10.1105/tpc.108.062968

Alves-Ferreira M, Wellmer F, Banhara A, Kumar V, Riechmann JL, Meyerowitz EM (2007) Global expression profiling applied to the analysis of arabidopsis stamen development. Plant physiology 145 (3):747-762. doi:pp.107.104422 [pii] 10.1104/pp.107.104422

Aoki K, Ogata Y, Shibata D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. Plant & cell physiology 48 (3):381-390. doi:pcm013 [pii] 10.1093/pcp/pcm013

Ariizumi T, Hatakeyama K, Hinata K, Sato S, Kato T, Tabata S, Toriyama K (2005) The hkm gene, which is identical to ms1 gene of arabidopsis thaliana, is essential for primexine formation and exine pattern formation. Sex Plant Reprod 18:1-7

Arondel V, Kader JC (1990) Lipid transfer in plants. Experientia 46 (6):579-585

Assuncao AG, Herrero E, Lin YF, Huettel B, Talukdar S, Smaczniak C, Immink RG, van Eldik M, Fiers M, Schat H, Aarts MG (2010) Arabidopsis thaliana transcription factors bzip19 and bzip23 regulate the adaptation to zinc deficiency. Proceedings of the National Academy of Sciences of the United States of America 107 (22):10296-10301. doi:1004788107 [pii] 10.1073/pnas.1004788107

Aya K, Ueguchi-Tanaka M, Kondo M, Hamada K, Yano K, Nishimura M, Matsuoka M (2009) Gibberellin modulates anther development in rice via the transcriptional regulation of gamyb. The Plant cell 21 (5):1453-1472. doi:tpc.108.062935 [pii] 10.1105/tpc.108.062935

Baginsky S, Hennig L, Zimmermann P, Gruissem W (2010) Gene expression analysis, proteomics, and network discovery. Plant physiology 152 (2):402-410. doi:pp.109.150433 [pii] 10.1104/pp.109.150433

Ball CA, Awad IA, Demeter J, Gollub J, Hebert JM, Hernandez-Boussard T, Jin H, Matese JC, Nitzberg M, Wymore F, Zachariah ZK, Brown PO, Sherlock G (2005) The stanford microarray database accommodates additional microarray platforms and data formats. Nucleic acids research 33 (Database issue):D580-582

Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R (2005) Ncbi geo: Mining millions of expression profiles--database and tools. Nucleic acids research 33 Database Issue:D562-566

Baud S, Wuilleme S, To A, Rochat C, Lepiniec L (2009) Role of wrinkled1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in arabidopsis. Plant J 60 (6):933-947. doi:TPJ4011 [pii] 10.1111/j.1365-313X.2009.04011.x

Becker JD, Boavida LC, Carneiro J, Haury M, Feijo JA (2003) Transcriptional profiling of arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant physiology 133 (2):713-725. doi:10.1104/pp.103.028241 pp.103.028241 [pii]

Bedinger P (1992) The remarkable biology of pollen. The Plant cell 4 (8):879-887. doi:10.1105/tpc.4.8.879 4/8/879 [pii]

Borg M, Brownfield L, Twell D (2009) Male gametophyte development: A molecular perspective. Journal of experimental botany 60 (5):1465-1478. doi:ern355 [pii] 10.1093/jxb/ern355
Borg M, Twell D (2010) Life after meiosis: Patterning the angiosperm male gametophyte. Biochem Soc Trans 38 (2):577-582. doi:BST0380577 [pii] 10.1042/BST0380577
Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijo JA, Becker JD (2008) Comparative transcriptomics of arabidopsis sperm cells. Plant physiology 148 (2):1168-1181. doi:pp.108.125229 [pii] 10.1104/pp.108.125229
Bouche N, Bouchez D (2001) Arabidopsis gene knockout: Phenotypes wanted. Current opinion in plant biology 4 (2):111-117. doi:S1369-5266(00)00145-X [pii]
Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Gorlach J (2001) Growth stage-based phenotypic analysis of arabidopsis: A model for high throughput functional genomics in plants. The Plant cell 13 (7):1499-1510
Brady SM, Provart NJ (2009) Web-queryable large-scale data sets for hypothesis generation in plant biology. The Plant cell 21 (4):1034-1051. doi:tpc.109.066050 [pii] 10.1105/tpc.109.066050
Brazma A (2009) Minimum information about a microarray experiment (miame)--successes, failures, challenges. ScientificWorldJournal 9:420-423. doi:10.1100/tsw.2009.57
Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M (2001) Minimum information about a microarray experiment (miame)-toward standards for microarray data. Nat Genet 29 (4):365-371. doi:10.1038/ng1201-365 ng1201-365 [pii]
Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca-Serra P, Sansone SA (2003) Arrayexpress--a public repository for microarray gene expression data at the ebi. Nucleic acids research 31 (1):68-71
Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (mpss) on microbead arrays. Nat Biotechnol 18 (6):630-634
Brownfield L, Hafidh S, Borg M, Sidorova A, Mori T, Twell D (2009a) A plant germline-specific integrator of sperm specification and cell cycle progression. PLoS genetics 5 (3):e1000430. doi:10.1371/journal.pgen.1000430
Brownfield L, Hafidh S, Durbarry A, Khatab H, Sidorova A, Doerner P, Twell D (2009b) Arabidopsis duo pollen3 is a key regulator of male germline development and embryogenesis. The Plant cell 21 (7):1940-1956. doi:tpc.109.066373 [pii] 10.1105/tpc.109.066373
Brownfield L, Twell D (2009) A dynamic duo of regulatory proteins coordinates gamete specification and germ cell mitosis in the angiosperm male germline. Plant Signal Behav 4 (12). doi:9950 [pii]
Campbell P, Braam J (1999) Xyloglucan endotransglycosylases: Diversity of genes, enzymes and potential wall-modifying functions. Trends in plant science 4 (9):361-366. doi:S1360-1385(99)01468-5 [pii]
Correa LG, Riano-Pachon DM, Schrago CG, dos Santos RV, Mueller-Roeber B, Vincentz M (2008) The role of bzip transcription factors in green plant evolution: Adaptive

features emerging from four founder genes. PLoS ONE 3 (8):e2944. doi:10.1371/journal.pone.0002944
Coussens L, Parker PJ, Rhee L, Yang-Feng TL, Chen E, Waterfield MD, Francke U, Ullrich A (1986) Multiple, distinct forms of bovine and human protein kinase c suggest diversity in cellular signaling pathways. Science (New York, NY 233 (4766):859-866
Craigon DJ, James N, Okyere J, Higgins J, Jotham J, May S (2004) Nascarrays: A repository for microarray data generated by nasc's transcriptomics service. Nucleic acids research 32 (Database issue):D575-577
Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK (2004) Real-time rt-pcr profiling of over 1400 arabidopsis transcription factors: Unprecedented sensitivity reveals novel root- and shoot-specific genes. Plant J 38 (2):366-379. doi:10.1111/j.1365-313X.2004.02051.x TPJ2051 [pii]
Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003a) Agris: Arabidopsis gene regulatory information server, an information resource of arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics 4 (1):25
Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003b) Agris: Arabidopsis gene regulatory information server, an information resource of arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics 4:25. doi:10.1186/1471-2105-4-25
Dickinson HG, Grant-Downton R (2009) Bridging the generation gap: Flowering plant gametophytes and animal germlines reveal unexpected similarities. Biol Rev Camb Philos Soc 84 (4):589-615. doi:BRV88 [pii] 10.1111/j.1469-185X.2009.00088.x
Dietz KJ, Vogel MO, Viehhauser A (2010) Ap2/erebp transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. Protoplasma. doi:10.1007/s00709-010-0142-8
Donson J, Fang Y, Espiritu-Santo G, Xing W, Salazar A, Miyamoto S, Armendarez V, Volkmuth W (2002) Comprehensive gene expression analysis by transcript profiling. Plant molecular biology 48 (1-2):75-97
Dupľáková N, Reňák D, Hovanec P, Honysová B, Twell D, Honys D (2007) Arabidopsis gene family profiler (agfp)--user-oriented transcriptomic database with easy-to-use graphic interface. BMC plant biology 7:39. doi:1471-2229-7-39 [pii] 10.1186/1471-2229-7-39
Durbarry A, Vizir I, Twell D (2005) Male germ line development in arabidopsis. Duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. Plant physiology 137 (1):297-307. doi:pp.104.053165 [pii] 10.1104/pp.104.053165
Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: Ncbi gene expression and hybridization array data repository. Nucleic acids research 30 (1):207-210
Englbrecht CC, Schoof H, Bohm S (2004) Conservation, diversification and expansion of c2h2 zinc finger proteins in the arabidopsis thaliana genome. BMC genomics 5 (1):39
Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D (1991) Light-directed, spatially addressable parallel chemical synthesis. Science (New York, NY 251 (4995):767-773
Francis KE, Lam SY, Copenhaver GP (2006) Separation of arabidopsis pollen tetrads is regulated by quartet1, a pectin methylesterase gene. Plant physiology 142 (3):1004-1013. doi:pp.106.085274 [pii] 10.1104/pp.106.085274

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) Expasy: The proteomics server for in-depth protein knowledge and analysis. Nucleic acids research 31 (13):3784-3788

Geitmann A (2010) How to shape a cylinder: Pollen tube as a model system for the generation of complex cellular geometry. Sex Plant Reprod 23 (1):63-71

Gibalova A, Renak D, Matczuk K, Dupl'akova N, Chab D, Twell D, Honys D (2009) Atbzip34 is required for arabidopsis pollen wall patterning and the control of several metabolic pathways in developing pollen. Plant molecular biology 70 (5):581-601. doi:10.1007/s11103-009-9493-y

Gil-Humanes J, Piston F, Martin A, Barro F (2009) Comparative genomic analysis and expression of the apetala2-like genes from barley, wheat, and barley-wheat amphiploids. BMC plant biology 9:66. doi:1471-2229-9-66 [pii] 10.1186/1471-2229-9-66

Ginsberg AM, King BO, Roeder RG (1984) Xenopus 5s gene transcription factor, tfiiia: Characterization of a cdna clone and measurement of rna levels throughout development. Cell 39 (3 Pt 2):479-489. doi:0092-8674(84)90455-0 [pii]

Gollub J, Ball CA, Binkley G, Demeter J, Finkelstein DB, Hebert JM, Hernandez-Boussard T, Jin H, Kaloper M, Matese JC, Schroeder M, Brown PO, Botstein D, Sherlock G (2003) The stanford microarray database: Data access and quality assessment tools. Nucleic acids research 31 (1):94-96

Gollub J, Ball CA, Sherlock G (2006) The stanford microarray database: A user's guide. Methods in molecular biology (Clifton, NJ 338:191-208

Gong W, Shen YP, Ma LG, Pan Y, Du YL, Wang DH, Yang JY, Hu LD, Liu XF, Dong CX, Ma L, Chen YH, Yang XY, Gao Y, Zhu D, Tan X, Mu JY, Zhang DB, Liu YL, Dinesh-Kumar SP, Li Y, Wang XP, Gu HY, Qu LJ, Bai SN, Lu YT, Li JY, Zhao JD, Zuo J, Huang H, Deng XW, Zhu YX (2004) Genome-wide orfeome cloning and analysis of arabidopsis transcription factor genes. Plant physiology 135 (2):773-782. doi:10.1104/pp.104.042176 135/2/773 [pii]

Grant-Downton R, Hafidh S, Twell D, Dickinson HG (2009) Small rna pathways are present and functional in the angiosperm male gametophyte. Mol Plant 2 (3):500-512. doi:ssp003 [pii] 10.1093/mp/ssp003

Guo A, He K, Liu D, Bai S, Gu X, Wei L, Luo J (2005) Datf: A database of arabidopsis transcription factors. Bioinformatics 21 (10):2568-2569. doi:bti334 [pii] 10.1093/bioinformatics/bti334

Hadfield KA, Bennett AB (1998) Polygalacturonases: Many genes in search of a function. Plant physiology 117 (2):337-343

Hess J, Laumen H, Wirth T (1998) Application of differential cdna screening techniques to the identification of unique gene expression in tumours and lymphocytes. Curr Opin Immunol 10 (2):125-130. doi:S0952-7915(98)80239-2 [pii]

Honys D, Dupľáková N, Reňák D (2008) Online tools for presentation and analysis of plant microarray data. In: Moretti MK, Rizzo LJ (eds) Oligonucleotide array sequence analysis. Nova Science Publishers, pp 265-295

Honys D, Oh SA, Renak D, Donders M, Solcova B, Johnson JA, Boudova R, Twell D (2006a) Identification of microspore-active promoters that allow targeted manipulation of gene expression at early stages of microgametogenesis in arabidopsis. BMC plant biology 6:31. doi:1471-2229-6-31 [pii] 10.1186/1471-2229-6-31

Honys D, Reňák D, Twell D (2006b) Male gametophyte development and function. In: Teixeira da Silva JA (ed) Floriculture, ornamental and plant biotechnology, vol I. Global Science Books, London,
Honys D, Twell D (2003) Comparative analysis of the arabidopsis pollen transcriptome. Plant physiology 132 (2):640-652. doi:10.1104/pp.103.020925 132/2/640 [pii]
Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in arabidopsis. Genome biology 5 (11):R85. doi:gb-2004-5-11-r85 [pii] 10.1186/gb-2004-5-11-r85
Hossain MA, Cho JI, Han M, Ahn CH, Jeon JS, An G, Park PB (2010) The abre-binding bzip transcription factor osabf2 is a positive regulator of abiotic stress and aba signaling in rice. Journal of plant physiology. doi:S0176-1617(10)00234-8 [pii] 10.1016/j.jplph.2010.05.008
Hothorn M, Wolf S, Aloy P, Greiner S, Scheffzek K (2004) Structural insights into the target specificity of plant invertase and pectin methylesterase inhibitory proteins. The Plant cell 16 (12):3437-3447. doi:tpc.104.025684 [pii] 10.1105/tpc.104.025684
Howden R, Park SK, Moore JM, Orme J, Grossniklaus U, Twell D (1998) Selection of t-DNA-tagged male and female gametophytic mutants by segregation distortion in arabidopsis. Genetics 149 (2):621-631
Charon C, Moreno AB, Bardou F, Crespi M (2010) Non-protein-coding rnas and their interacting rna-binding proteins in the plant cell nucleus. Mol Plant. doi:ssq037 [pii] 10.1093/mp/ssq037
Chen YC, McCormick S (1996) Sidecar pollen, an arabidopsis thaliana male gametophytic mutant with aberrant cell divisions during pollen development. Development (Cambridge, England) 122 (10):3243-3253
Chen Z, Tan JL, Ingouff M, Sundaresan V, Berger F (2008) Chromatin assembly factor 1 regulates the cell cycle but not cell fate during male gametogenesis in arabidopsis thaliana. Development (Cambridge, England) 135 (1):65-73. doi:dev.010108 [pii] 10.1242/dev.010108
Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, Konagaya A, Shinozaki K (2005) Rartf: Database and tools for complete sets of arabidopsis transcription factors. DNA Res 12 (4):247-256. doi:12/4/247 [pii] 10.1093/dnares/dsi011
Ito T, Nagata N, Yoshiba Y, Ohme-Takagi M, Ma H, Shinozaki K (2007) Arabidopsis male sterility1 encodes a phd-type transcription factor and regulates pollen and tapetum development. The Plant cell 19 (11):3549-3562. doi:tpc.107.054536 [pii] 10.1105/tpc.107.054536
Ito T, Shinozaki K (2002) The male sterility1 gene of arabidopsis, encoding a nuclear protein with a phd-finger motif, is expressed in tapetal cells and is required for pollen maturation. Plant & cell physiology 43 (11):1285-1292
Iven T, Strathmann A, Bottner S, Zwafink T, Heinekamp T, Guivarc'h A, Roitsch T, Droge-Laser W (2010) Homo- and heterodimers of tobacco bzip proteins counteract as positive or negative regulators of transcription during pollen development. Plant J. doi:TPJ4230 [pii] 10.1111/j.1365-313X.2010.04230.x
Iwakawa H, Shinmyo A, Sekine M (2006) Arabidopsis cdka;1, a cdc2 homologue, controls proliferation of generative cells in male gametogenesis. The Plant Journal 45:819-831
Iwata Y, Fedoroff NV, Koizumi N (2008) Arabidopsis bzip60 is a proteolysis-activated transcription factor involved in the endoplasmic reticulum stress response. The Plant cell 20 (11):3107-3121. doi:tpc.108.061002 [pii] 10.1105/tpc.108.061002

Jen CH, Manfield IW, Michalopoulos I, Pinney JW, Willats WG, Gilmartin PM, Westhead DR (2006) The arabidopsis co-expression tool (act): A www-based tool and database for microarray-based gene expression analysis. Plant J 46 (2):336-348. doi:TPJ2681 [pii] 10.1111/j.1365-313X.2006.02681.x

Jiang Y, Deyholos MK (2006) Comprehensive transcriptional profiling of nacl-stressed arabidopsis roots reveals novel classes of responsive genes. BMC plant biology 6:25. doi:1471-2229-6-25 [pii] 10.1186/1471-2229-6-25

Jiang Y, Deyholos MK (2009) Functional characterization of arabidopsis nacl-inducible wrky25 and wrky33 transcription factors in abiotic stresses. Plant molecular biology 69 (1-2):91-105. doi:10.1007/s11103-008-9408-3

Jiang Y, Yang B, Deyholos MK (2009) Functional characterization of the arabidopsis bhlh92 transcription factor in abiotic stress. Mol Genet Genomics 282 (5):503-516. doi:10.1007/s00438-009-0481-3

Johnson MA, von Besser K, Zhou Q, Smith E, Aux G, Patton D, Levin JZ, Preuss D (2004) Arabidopsis hapless mutations define essential gametophytic functions. Genetics 168 (2):971-982. doi:168/2/971 [pii] 10.1534/genetics.104.029447

Kaneko M, Inukai Y, Ueguchi-Tanaka M, Itoh H, Izawa T, Kobayashi Y, Hattori T, Miyao A, Hirochika H, Ashikari M, Matsuoka M (2004) Loss-of-function mutations of the rice gamyb gene impair alpha-amylase expression in aleurone and flower development. The Plant cell 16 (1):33-44. doi:10.1105/tpc.017327 tpc.017327 [pii]

Kang IH, Steffen JG, Portereiko MF, Lloyd A, Drews GN (2008) The agl62 mads domain protein regulates cellularization during endosperm development in arabidopsis. The Plant cell 20 (3):635-647. doi:tpc.107.055137 [pii] 10.1105/tpc.107.055137

Kang SG, Price J, Lin PC, Hong JC, Jang JC (2010) The arabidopsis bzip1 transcription factor is involved in sugar signaling, protein networking, and DNA binding. Mol Plant 3 (2):361-373. doi:ssp115 [pii] 10.1093/mp/ssp115

Kapoor S, Kobayashi A, Takatsuji H (2002) Silencing of the tapetum-specific zinc finger gene taz1 causes premature degeneration of tapetum and pollen abortion in petunia. The Plant cell 14 (10):2353-2367

Kapoor S, Takatsuji H (2006) Silencing of an anther-specific zinc-finger gene, mez1, causes aberrant meiosis and pollen abortion in petunia. Plant molecular biology 61 (3):415-430

Kapushesky M, Kemmeren P, Culhane AC, Durinck S, Ihmels J, Korner C, Kull M, Torrente A, Sarkans U, Vilo J, Brazma A (2004) Expression profiler: Next generation--an online platform for analysis of microarray data. Nucleic acids research 32 (Web Server issue):W465-470. doi:10.1093/nar/gkh470 32/suppl_2/W465 [pii]

Kato N, He H, Steger AP (2010) A systems model of vesicle trafficking in arabidopsis pollen tubes. Plant physiology 152 (2):590-601. doi:pp.109.148700 [pii] 10.1104/pp.109.148700

Kim HJ, Oh SA, Brownfield L, Hong SH, Ryu H, Hwang I, Twell D, Nam HG (2008) Control of plant germline proliferation by scf(fbl17) degradation of cell cycle inhibitors. Nature 455 (7216):1134-1137. doi:nature07289 [pii] 10.1038/nature07289

Kim OK, Jung JH, Park CM (2010) An arabidopsis f-box protein regulates tapetum degeneration and pollen maturation during anther development. Planta 232 (2):353-366. doi:10.1007/s00425-010-1178-x

Kobayashi A, Sakamoto A, Kubo K, Rybka Z, Kanno Y, Takatsuji H (1998) Seven zinc-finger transcription factors are expressed sequentially during the development of anthers in petunia. Plant J 13 (4):571-576

Krysan PJ, Young JC, Jester PJ, Monson S, Copenhaver G, Preuss D, Sussman MR (2002) Characterization of t-DNA insertion sites in arabidopsis thaliana and the implications for saturation mutagenesis. Omics 6 (2):163-174. doi:10.1089/153623102760092760
Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in arabidopsis. The Plant cell 11 (12):2283-2290
Lee S, Jeong HJ, Kim SA, Lee J, Guerinot ML, An G (2010) Oszip5 is a plasma membrane zinc transporter in rice. Plant molecular biology 73 (4-5):507-517. doi:10.1007/s11103-010-9637-0
Leubner-Metzger G, Petruzzelli L, Waldvogel R, Vogeli-Lange R, Meins F, Jr. (1998) Ethylene-responsive element binding protein (erebp) expression and the transcriptional regulation of class i beta-1,3-glucanase during tobacco seed germination. Plant molecular biology 38 (5):785-795
Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. Nat Genet 21 (1 Suppl):20-24
Liu JX, Howell SH (2010) Bzip28 and nf-y transcription factors are activated by er stress and assemble into a transcriptional complex to regulate stress response genes in arabidopsis. The Plant cell 22 (3):782-796. doi:tpc.109.072173 [pii] 10.1105/tpc.109.072173
Liu Z, Bao W, Liang W, Yin J, Zhang D (2010) Identification of gamyb-4 and analysis of the regulatory role of gamyb in rice anther development. J Integr Plant Biol 52 (7):670-678. doi:JIPB959 [pii] 10.1111/j.1744-7909.2010.00959.x
Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14 (13):1675-1680
Mandaokar A, Browse J (2009) Myb108 acts together with myb24 to regulate jasmonate-mediated stamen maturation in arabidopsis. Plant physiology 149 (2):851-862. doi:pp.108.132597 [pii] 10.1104/pp.108.132597
Mandaokar A, Kumar VD, Amway M, Browse J (2003) Microarray and differential display identify genes involved in jasmonate-dependent anther development. Plant molecular biology 52 (4):775-786
Mandaokar A, Thines B, Shin B, Lange BM, Choi G, Koo YJ, Yoo YJ, Choi YD, Choi G, Browse J (2006) Transcriptional regulators of stamen development in arabidopsis identified by transcriptional profiling. Plant J 46 (6):984-1008
Manfield IW, Jen CH, Pinney JW, Michalopoulos I, Bradford JR, Gilmartin PM, Westhead DR (2006) Arabidopsis co-expression tool (act): Web server tools for microarray-based gene expression analysis. Nucleic acids research 34 (Web Server issue):W504-509. doi:34/suppl_2/W504 [pii] 10.1093/nar/gkl204
McCormick S (1993) Male gametophyte development. The Plant cell 5 (10):1265-1275. doi:10.1105/tpc.5.10.1265 5/10/1265 [pii]
McCormick S (2004) Control of male gametophyte development. The Plant cell 16 Suppl:S142-153. doi:10.1105/tpc.016659 tpc.016659 [pii]
McGall GH, Barone D, Diggelmann M, Fodor SPA, Gentalen E, Ngo N (1997) The efficiency of light-directed synthesis of DNA arrays on glass substrates. *JAmChemSoc* 119:5081-5090
Meyers BC, Tej SS, Vu TH, Haudenschild CD, Agrawal V, Edberg SB, Ghazal H, Decola S (2004a) The use of mpss for whole-genome transcriptional analysis in arabidopsis. Genome research 14 (8):1641-1653. doi:10.1101/gr.2275604 14/8/1641 [pii]

Meyers BC, Vu TH, Tej SS, Ghazal H, Matvienko M, Agrawal V, Ning J, Haudenschild CD (2004b) Analysis of the transcriptional complexity of arabidopsis thaliana by massively parallel signature sequencing. Nat Biotechnol 22 (8):1006-1011. doi:10.1038/nbt992 nbt992 [pii]

Mitsuda N, Hiratsu K, Todaka D, Nakashima K, Yamaguchi-Shinozaki K, Ohme-Takagi M (2006) Efficient production of male and female sterile plants by expression of a chimeric repressor in arabidopsis and rice. Plant Biotechnol J 4 (3):325-332. doi:PBI184 [pii] 10.1111/j.1467-7652.2006.00184.x

Mohnen D (2008) Pectin structure and biosynthesis. Current opinion in plant biology 11 (3):266-277. doi:S1369-5266(08)00063-0 [pii] 10.1016/j.pbi.2008.03.006

Money T, Reader S, Qu LJ, Dunford RP, Moore G (1996) Aflp-based mrna fingerprinting. Nucleic acids research 24 (13):2616-2617

Morcillo F, Gallard A, Pillot M, Jouannic S, Aberlenc-Bertossi F, Collin M, Verdeil JL, Tregear JW (2007) Egap2-1, an aintegumenta-like (ail) gene expressed in meristematic and proliferating tissues of embryos in oil palm. Planta 226 (6):1353-1362. doi:10.1007/s00425-007-0574-3

Nalefski EA, Falke JJ (1996) The c2 domain calcium-binding motif: Structural and functional diversity. Protein Sci 5 (12):2375-2390. doi:10.1002/pro.5560051201

O'Malley RC, Ecker JR (2010) Linking genotype to phenotype using the arabidopsis unimutant collection. Plant J 61 (6):928-940. doi:TPJ4119 [pii] 10.1111/j.1365-313X.2010.04119.x

Okamuro JK, Caster B, Villarroel R, Van Montagu M, Jofuku KD (1997) The ap2 domain of apetala2 defines a large new family of DNA binding proteins in arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 94 (13):7076-7081

Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri RV, Grotewold E (2006) Agris and atregnet. A platform to link cis-regulatory elements and transcription factors into regulatory networks. Plant physiology 140 (3):818-829

Parinov S, Sundaresan V (2000) Functional genomics in arabidopsis: Large-scale insertional mutagenesis complements the genome sequencing project. Curr Opin Biotechnol 11 (2):157-161. doi:S0958-1669(00)00075-6 [pii]

Park SK, Howden R, Twell D (1998) The arabidopsis thaliana gametophytic mutation gemini pollen1 disrupts microspore polarity, division asymmetry and pollen cell fate. Development (Cambridge, England) 125 (19):3789-3799

Park SK, Rahman D, Oh SA, Twell D (2004) Gemini pollen 2, a male and female gametophytic cytokinesis defective mutation. Sexual plant reproduction 17 (2):63-70. doi:10.1007/s00497-004-0216-x

Parkinson H, Kapushesky M, Kolesnikov N, Rustici G, Shojatalab M, Abeygunawardena N, Berube H, Dylag M, Emam I, Farne A, Holloway E, Lukk M, Malone J, Mani R, Pilicheva E, Rayner TF, Rezwan F, Sharma A, Williams E, Bradley XZ, Adamusiak T, Brandizi M, Burdett T, Coulson R, Krestyaninova M, Kurnosov P, Maguire E, Neogi SG, Rocca-Serra P, Sansone SA, Sklyar N, Zhao M, Sarkans U, Brazma A (2009) Arrayexpress update--from an archive of functional genomics experiments to the atlas of gene expression. Nucleic acids research 37 (Database issue):D868-872. doi:gkn889 [pii] 10.1093/nar/gkn889

Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara GG, Holloway E, Kapushesky M, Lilja P, Mukherjee G, Oezcimen A, Rayner T, Rocca-Serra P, Sharma A, Sansone S, Brazma A (2005) Arrayexpress--a public

repository for microarray gene expression data at the ebi. Nucleic acids research 33 Database Issue:D553-555
Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proceedings of the National Academy of Sciences of the United States of America 91 (11):5022-5026
Perin MS, Fried VA, Mignery GA, Jahn R, Sudhof TC (1990) Phospholipid binding by a synaptic vesicle protein homologous to the regulatory region of protein kinase c. Nature 345 (6272):260-263. doi:10.1038/345260a0
Piffanelli P, Ross JHE, Murphy DJ (1998) Biogenesis and function of the lipidic structures of pollen grains. Sexual plant reproduction 11:65-80
Pina C, Pinto F, Feijo JA, Becker JD (2005) Gene family analysis of the arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant physiology 138 (2):744-756. doi:pp.104.057935 [pii] 10.1104/pp.104.057935
Preston J, Wheeler J, Heazlewood J, Li SF, Parish RW (2004) Atmyb32 is required for normal pollen development in arabidopsis thaliana. Plant J 40 (6):979-995. doi:TPJ2280 [pii] 10.1111/j.1365-313X.2004.02280.x
Qin Y, Leydon AR, Manziello A, Pandey R, Mount D, Denic S, Vasic B, Johnson MA, Palanivelu R (2009) Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. PLoS genetics 5 (8):e1000621. doi:10.1371/journal.pgen.1000621
Qu LJ, Zhu YX (2006) Transcription factor families in arabidopsis: Major progress and outstanding issues for future research. Current opinion in plant biology 9 (5):544-549. doi:S1369-5266(06)00111-7 [pii] 10.1016/j.pbi.2006.07.005
Rao KP, Richa T, Kumar K, Raghuram B, Sinha AK (2010) In silico analysis reveals 75 members of mitogen-activated protein kinase kinase kinase gene family in rice. DNA Res 17 (3):139-153. doi:dsq011 [pii] 10.1093/dnares/dsq011
Rensink WA, Buell CR (2005) Microarray expression profiling resources for plant genomics. Trends in plant science 10 (12):603-609. doi:S1360-1385(05)00252-9 [pii] 10.1016/j.tplants.2005.10.003
Rhee SG, Suh PG, Ryu SH, Lee SY (1989) Studies of inositol phospholipid-specific phospholipase c. Science (New York, NY 244 (4904):546-550
Rhee SY, Somerville CR (1998) Tetrad pollen formation in quartet mutants of arabidopsis thaliana is associated with persistence of pectic polysaccharides of the pollen mother cell wall. Plant J 15 (1):79-88
Riechmann JL (2002) Transcriptional regulation: A genomic overview. In: Somerville CR, Meyerowitz EM (eds) The arabidopsis book. American Society of Plant Biologists, Rockville, MD. doi:doi/10.1199/tab.0085
Riechmann JL (2006) Transcription factors of *arabidopsis* and rice: A genomic perspectives. In: Grasser KD (ed) Regulation of transcription in plants, annual plant reviews, vol 29. Blackwell Publishing, pp 28-43
Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. Science (New York, NY 290 (5499):2105-2110. doi:9054 [pii]
Riechmann JL, Meyerowitz EM (1998) The ap2/erebp family of plant transcription factors. Biol Chem 379 (6):633-646

Riechmann JL, Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. Current opinion in plant biology 3 (5):423-434. doi:S1369-5266(00)00107-2 [pii]
Rizo J, Sudhof TC (1998) C2-domains, structure and function of a universal ca2+-binding domain. The Journal of biological chemistry 273 (26):15879-15882
Rong J, Song Z, de Jong TJ, Zhang X, Sun S, Xu X, Xia H, Liu B, Lu BR (2010) Modelling pollen-mediated gene flow in rice: Risk assessment and management of transgene escape. Plant Biotechnol J 8 (4):452-464. doi:PBI488 [pii] 10.1111/j.1467-7652.2009.00488.x
Rose JKC (2003) The plant cell wall. Blackwell Publishing
Rotman N, Durbarry A, Wardle A, Yang WC, Chaboud A, Faure JE, Berger F, Twell D (2005) A novel class of myb factors controls sperm-cell formation in plants. Curr Biol 15 (3):244-248. doi:S0960982205000394 [pii] 10.1016/j.cub.2005.01.013
Russell SD, Gou X, Wei X, Yuan T (2010) Male gamete biology in flowering plants. Biochem Soc Trans 38 (2):598-603. doi:BST0380598 [pii] 10.1042/BST0380598
Saez-Vasquez J, Gadal O (2010) Genome organization and function: A view from yeast and arabidopsis. Mol Plant. doi:ssq034 [pii] 10.1093/mp/ssq034
Sandhu AP, Randhawa GS, Dhugga KS (2009) Plant cell wall matrix polysaccharide biosynthesis. Mol Plant 2 (5):840-850. doi:ssp056 [pii] 10.1093/mp/ssp056
Sarkans U, Parkinson H, Lara GG, Oezcimen A, Sharma A, Abeygunawardena N, Contrino S, Holloway E, Rocca-Serra P, Mukherjee G, Shojatalab M, Kapushesky M, Sansone SA, Farne A, Rayner T, Brazma A (2005) The arrayexpress gene expression database: A software engineering and implementation perspective. Bioinformatics 21 (8):1495-1501. doi:bti157 [pii] 10.1093/bioinformatics/bti157
Scott RJ, Spielman M, Dickinson HG (2004) Stamen structure and function. The Plant cell 16 Suppl:S46-60. doi:10.1105/tpc.017012 tpc.017012 [pii]
Shirataki H, Kaibuchi K, Sakoda T, Kishida S, Yamaguchi T, Wada K, Miyazaki M, Takai Y (1993) Rabphilin-3a, a putative target protein for smg p25a/rab3a p25 small gtp-binding protein related to synaptotagmin. Molecular and cellular biology 13 (4):2061-2068
Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (New York, NY 270 (5235):467-470
Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of arabidopsis thaliana development. Nat Genet 37 (5):501-506. doi:ng1543 [pii] 10.1038/ng1543
Singh MB, Bhalla PL (2007) Control of male germ-cell development in flowering plants. Bioessays 29 (11):1124-1132. doi:10.1002/bies.20660
Soares MB (1997) Identification and cloning of differentially expressed genes. Curr Opin Biotechnol 8 (5):542-546. doi:S0958-1669(97)80026-2 [pii]
Sorensen AM, Krober S, Unte US, Huijser P, Dekker K, Saedler H (2003) The arabidopsis aborted microspores (ams) gene encodes a myc class transcription factor. Plant J 33 (2):413-423. doi:1644 [pii]
Steiner-Lange S, Unte US, Eckstein L, Yang C, Wilson ZA, Schmelzer E, Dekker K, Saedler H (2003) Disruption of arabidopsis thaliana myb26 results in male sterility due to non-dehiscent anthers. Plant J 34 (4):519-528. doi:1745 [pii]
Stewman SF, Jones-Rhoades M, Bhimalapuram P, Tchernookov M, Preuss D, Dinner AR (2010) Mechanistic insights from a quantitative analysis of pollen tube guidance. BMC plant biology 10:32. doi:1471-2229-10-32 [pii] 10.1186/1471-2229-10-32

Stolc V, Samanta MP, Tongprasit W, Sethi H, Liang S, Nelson DC, Hegeman A, Nelson C, Rancour D, Bednarek S, Ulrich EL, Zhao Q, Wrobel RL, Newman CS, Fox BG, Phillips GN, Jr., Markley JL, Sussman MR (2005) Identification of transcribed sequences in arabidopsis thaliana by using high-resolution genome tiling arrays. Proceedings of the National Academy of Sciences of the United States of America 102 (12):4453-4458. doi:0408203102 [pii] 10.1073/pnas.0408203102

Suzuki G (2009) Recent progress in plant reproduction research: The story of the male gametophyte through to successful fertilization. Plant & cell physiology 50 (11):1857-1864. doi:pcp142 [pii] 10.1093/pcp/pcp142

Takatsuji H (1998) Zinc-finger transcription factors in plants. Cell Mol Life Sci 54 (6):582-596

Takatsuji H (1999) Zinc-finger proteins: The classical zinc finger emerges in contemporary plant science. Plant molecular biology 39 (6):1073-1078

Takeda T, Amano K, Ohto MA, Nakamura K, Sato S, Kato T, Tabata S, Ueguchi C (2006) Rna interference of the arabidopsis putative transcription factor tcp16 gene results in abortion of early pollen development. Plant molecular biology 61 (1-2):165-177. doi:10.1007/s11103-006-6265-9

Thoma S, Kaneko Y, Somerville C (1993) A non-specific lipid transfer protein from arabidopsis is a cell wall protein. Plant J 3 (3):427-436

Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The botany array resource: E-northerns, expression angling, and promoter analyses. Plant J 43 (1):153-163. doi:TPJ2437 [pii] 10.1111/j.1365-313X.2005.02437.x

Trahey M, Wong G, Halenbeck R, Rubinfeld B, Martin GA, Ladner M, Long CM, Crosier WJ, Watt K, Koths K, et al. (1988) Molecular cloning of two types of gap complementary DNA from human placenta. Science (New York, NY 242 (4886):1697-1700

Tsuji H, Aya K, Ueguchi-Tanaka M, Shimada Y, Nakazono M, Watanabe R, Nishizawa NK, Gomi K, Shimada A, Kitano H, Ashikari M, Matsuoka M (2006) Gamyb controls different sets of genes and is differentially regulated by microrna in aleurone cells and anthers. Plant J 47 (3):427-444. doi:TPJ2795 [pii] 10.1111/j.1365-313X.2006.02795.x

Twell D (2002) Pollen developmental biology. In: O'Neil SD, Robert JA (eds) Plant reproduction. Annual plant reviews vol 6. Sheffield Academic Press, Sheffield, pp 86-153

Twell D, Oh SA, Honys D (2006) Pollen development, a genetic and transcriptomic view. In: Malhó R (ed) The pollen tube. Plant Cell Monograph 3 edn. Springer-Verlag, Heidelberg,

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science (New York, NY 270 (5235):484-487

Vizcay-Barrena G, Wilson ZA (2006) Altered tapetal pcd and pollen wall development in the arabidopsis ms1 mutant. Journal of experimental botany 57 (11):2709-2717. doi:57/11/2709 [pii] 10.1093/jxb/erl032

Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in arabidopsis. Plant physiology 148 (3):1201-1211. doi:pp.108.126375 [pii] 10.1104/pp.108.126375

Willats WG, McCartney L, Mackie W, Knox JP (2001) Pectin: Cell biology and prospects for functional analysis. Plant molecular biology 47 (1-2):9-27

Wilson ZA, Marroll SM, Dawson J, Swarup R, Tighe PJ (2001) The arabidopsis male sterility (ms1) gene is a transcriptional regulator of male gametogenesis, with homology to the phd-finger family of transcription factors. Plant J 28:27-39

Wilson ZA, Yang C (2004) Plant gametogenesis: Conservation and contrasts in development. Reproduction 128 (5):483-492. doi:128/5/483 [pii] 10.1530/rep.1.00306

Xiong Y, Liu T, Tian C, Sun S, Li J, Chen M (2005) Transcription factors in rice: A genome-wide comparative analysis between monocots and eudicots. Plant molecular biology 59 (1):191-203. doi:10.1007/s11103-005-6503-6

Yang C, Vizcay-Barrena G, Conner K, Wilson ZA (2007) Male sterility1 is required for tapetal development and pollen wall biosynthesis. The Plant cell 19 (11):3530-3548. doi:tpc.107.054981 [pii] 10.1105/tpc.107.054981

Yang X, Makaroff CA, Ma H (2003) The arabidopsis male meiocyte death1 gene encodes a phd-finger protein that is required for male meiosis. The Plant cell 15 (6):1281-1295

Zhang H, Liang W, Yang X, Luo X, Jiang N, Ma H, Zhang D (2010) Carbon starved anther encodes a myb domain protein that regulates sugar partitioning required for rice pollen development. The Plant cell 22 (3):672-689. doi:tpc.109.073668 [pii] 10.1105/tpc.109.073668

Zhang Y (2008) I-tasser server for protein 3d structure prediction. BMC Bioinformatics 9:40. doi:1471-2105-9-40 [pii] 10.1186/1471-2105-9-40

Zhang ZB, Zhu J, Gao JF, Wang C, Li H, Zhang HQ, Zhang S, Wang DM, Wang QX, Huang H, Xia HJ, Yang ZN (2007) Transcription factor atmyb103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in arabidopsis. Plant J 52 (3):528-538. doi:TPJ3254 [pii] 10.1111/j.1365-313X.2007.03254.x

Zimmermann P, Hennig L, Gruissem W (2005) Gene-expression analysis and network discovery using genevestigator. Trends in plant science 10 (9):407-409. doi:S1360-1385(05)00170-6 [pii] 10.1016/j.tplants.2005.07.003

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) Genevestigator. Arabidopsis microarray database and analysis toolbox. Plant physiology 136 (1):2621-2632

Zinkl GM, Zwiebel BI, Grier DG, Preuss D (1999) Pollen-stigma adhesion in arabidopsis: A species-specific interaction mediated by lipophilic molecules in the pollen exine. Development (Cambridge, England) 126 (23):5431-5440