

# A decade of pollen transcriptomics

Nicholas Rutley · David Twell

Received: 9 January 2015 / Accepted: 24 February 2015 / Published online: 12 March 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

## Abstract

### **Key message** Overview of pollen transcriptome studies.

Pollen development is driven by gene expression, and knowledge of the molecular events underlying this process has undergone a quantum leap in the last decade through studies of the transcriptome. Here, we outline historical evidence for male haploid gene expression and review the wealth of pollen transcriptome data now available. Knowledge of the transcriptional capacity of pollen has progressed from genetic studies to the direct analysis of RNA and from gene-by-gene studies to analyses on a genomic scale. Microarray and/or RNA-seq data can now be accessed for all phases and cell types of developing pollen encompassing 10 different angiosperms. These growing resources have accelerated research and will undoubtedly inspire new directions and the application of system-based research into the mechanisms that govern the development, function and evolution of angiosperm pollen.

**Keywords** Pollen · Male gametophyte · Transcriptome · Gene expression · Development

## Introduction

The haploid male gametophyte (or pollen) of flowering plants is a well-understood and intriguing cell system in which to study gene expression and its regulation, as its development involves single-cell ontogeny and the cooperation of two-cell lineages to enable double fertilisation (Twell 2011). Pollen develops from haploid microspores that are produced following meiosis within the anthers (Fig. 1). The four haploid microspores are initially associated in a tetrad, but typically separate and undergo vacuolation and expansion, with the microspore nucleus migrating towards the cell wall. In this polarised arrangement, pollen mitosis I (PMI) results in a bicellular pollen (BCP) grain composed of a generative cell (representing the male germline) enclosed within the vegetative cell cytoplasm. The vegetative cell exits the cell cycle, but the generative cell elongates and divides at pollen mitosis II (PMII), giving rise to a pair of sperm cells. Upon pollination of a receptive stigma, pollen grains hydrate and germinate to produce a pollen tube. The pollen tube grows through the pistil by tip extension, guided by sporophytic and female gametophyte-derived signals, to deliver the sperm cells into the ovule where double fertilisation takes place.

In this review, we provide a brief survey of historical advances in understanding of gene expression in the male gametophyte and review the scale and diversity of the transcriptome data that have accumulated in the past decade. Some allied topics of importance including meiosis, pollen proteome studies, and the emerging role of small RNAs are not covered in detail, and the reader is referred to other recent and topical reviews (Le Trionnaire and Twell 2010; Twell 2011; Baroux et al. 2011; Feng et al. 2013; Dukowic-Schulze and Chen 2014; Fu and Yang 2014; Kawashima and Berger 2014).

---

Communicated by Silvia Coimbra and Lucia Colombo.

---

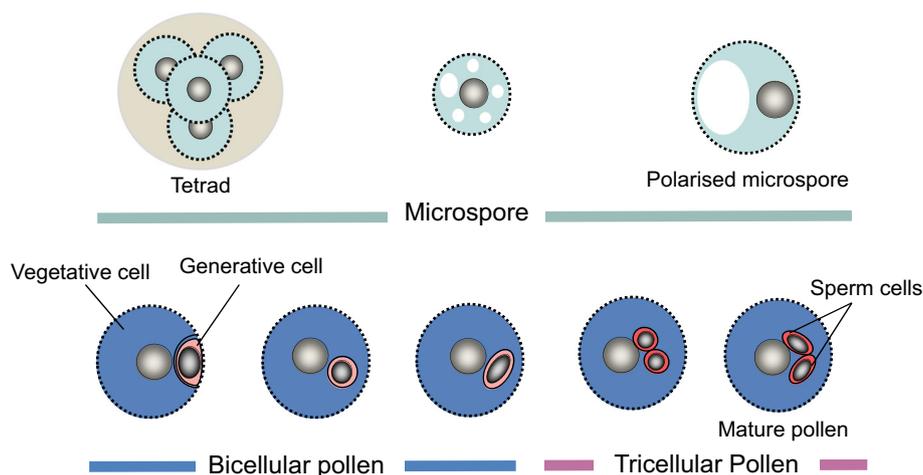
A contribution to the special issue 'From Gametes to Seeds'.

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00497-015-0261-7) contains supplementary material, which is available to authorized users.

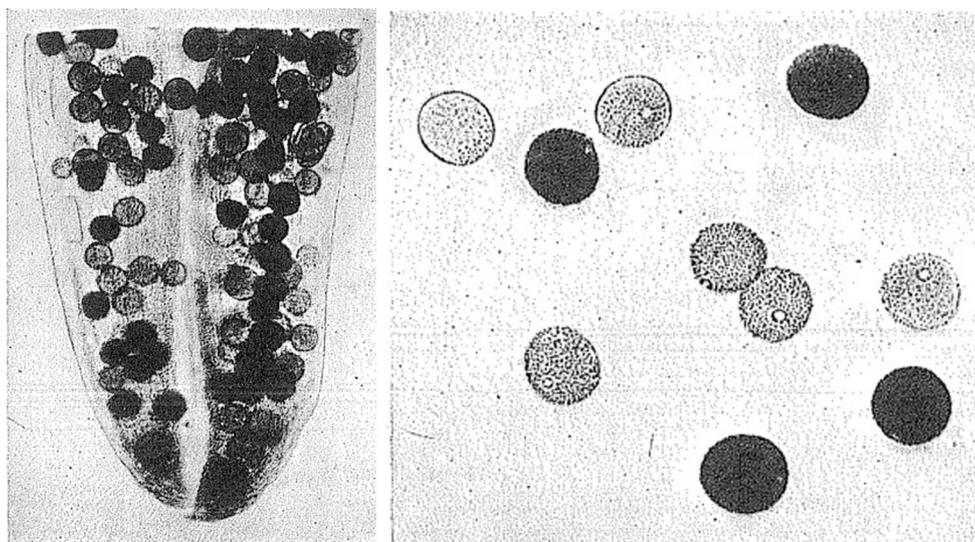
---

N. Rutley · D. Twell (✉)  
Department of Biology, University of Leicester,  
Leicester LE1 7RH, UK  
e-mail: twe@le.ac.uk



**Fig. 1** Distinct cytological stages that accompany male gametophyte development in *Arabidopsis*. Diploid pollen mother cells undergo meiotic division to produce a tetrad of haploid microspores. Microspores released from the tetrad undergo a highly asymmetric cell division (pollen mitosis I) to produce a bicellular pollen grain

with a small generative (germline) cell within the cytoplasm of the larger vegetative cell. The generative cell undergoes a further mitotic division (pollen mitosis II) during pollen maturation to produce a pair of sperm cells



**Fig. 2** Images from an original plate by Parnell (1921) showing segregation of F1 pollen into *starchy* (dark) and *glutinous* (light) types. Anther (left) and free pollen (right). Parnell describes the result as follows: ‘With a view to distinguishing the two genetic types, the

pollen was treated with iodine. The result was most satisfactory—two distinct types became evident, one giving the *dark blue* reaction of ordinary starch and the other the reddish reaction of amyloextrine’. With kind permission from Springer Science + Business Media

### Pollen gene expression in the pre-genomic era

The first evidence for haploid gene expression in pollen can be traced back to the first half of the twentieth century. In 1921, Parnell observed that half of the pollen from rice plants, which were heterozygous for the recessive *glutinous* (or *waxy*) endosperm phenotype, stained reddish, rather than dark blue, with iodine (Fig. 2). The phenotypic segregation of the *Waxy* locus, which is responsible for

amylose synthesis and encodes granule bound starch synthase I (Hirano and Sano 1991), thus provided genetic evidence for gene activity in pollen. Similar observations were made for *waxy* maize pollen and endosperm by Brink and MacGillivray (1924), who observed reduced pollen transmission of *waxy* alleles and hypothesised that this may be caused by reduced pollen tube growth, ‘by the action of certain factors active in the tube nucleus’. Male-biased and distorted Mendelian segregation ratios have been detected

repeatedly and include examples where mutant alleles affect pollen development, germination, or pollen tube growth (reviewed in Ottaviano and Mulcahy 1989; Twell 1994).

More recently, the directed isolation of developmental mutants and their associated genes has provided insight into the spatiotemporal expression of important genes in the male gametophyte. An innovation in the search for such genes was the use of the fluorescent DNA stain 4',6-diamidino-2-phenylindole (DAPI) to screen for abnormal pollen in mutagenised populations of *Arabidopsis* by fluorescence microscopy. This approach was used to identify the LOB domain family protein SIDECAR POLLEN (SCP), required for correct timing and organisation of PMI (Chen and McCormick 1996; Oh et al. 2010); the MAP215-family GEMINI POLLEN1 (GEM1), required for microspore polarity and asymmetric division at PMI (Park et al. 1998; Twell et al. 2002), as well as the TWO-IN-ONE (TIO) fused kinase required for pollen cytokinesis (Oh et al. 2005, 2010) and the germline-specific transcription factor DUO POLLEN1 (DUO1) (Durberry et al. 2005; Rotman et al. 2005). A second strategy made use of T-DNA or transposon-derived resistance markers in segregation distortion screens for deviations in the ratio of resistant to sensitive progeny from 3:1 to  $\leq 1:1$ . They delivered developmental mutants, such as *limpet pollen* in which the germ cells remain attached to the pollen wall (Howden et al. 1998), and progamic phase mutants affecting pollen tube growth and/or guidance (Procissi et al. 2001; Lalanne et al. 2004; Johnson et al. 2004). *Arabidopsis* mutants affecting all stages of pollen development, from the development of the microspore to the pollen tube as well as male–female gamete interactions, have now been isolated (reviewed by Twell 2011; Mori et al. 2014).

While segregating pollen mutants provided compelling evidence for developmentally regulated haploid gene expression, the first steps towards describing the RNA landscape of pollen began with the characterisation of different classes of RNA. Studies in *Tradescantia*, *Lilium*, and tobacco showed the rate of accumulation of rRNA and tRNA to be dynamic and variable between species (see Mascarenhas 1990). In *Tradescantia* and *Lilium*, rRNA and tRNA levels peaked around PMI, while in tobacco, levels of rRNA, tRNA, and mRNA increased progressively after PMI (Schrauwen et al. 1990). The detection of large amounts of mRNA in mature pollen further suggested that transcripts are stored for use during pollen germination and tube growth.

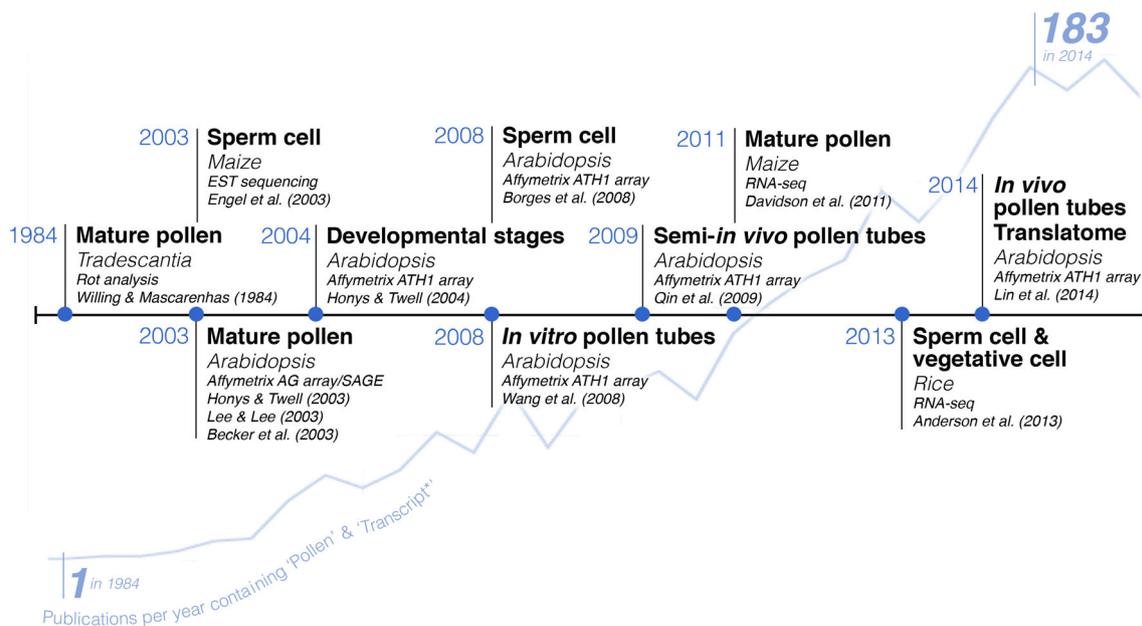
Initial studies to explore the composition of mRNA populations involved *in vitro* translation of mRNA from developing pollen of tobacco and *Lilium* (Schrauwen et al. 1990). For both species, the resulting 2D protein profiles revealed the presence of different transcripts before and

after PMI, with the greatest number of new transcripts appearing in mature pollen. These and similar findings in maize added weight to the evidence for post-meiotic transcription, based on the accumulation profiles of cloned pollen-expressed and pollen-specific transcripts, such as actin and the familiar tomato late pollen gene *LAT52*, respectively (Stinson et al. 1987; Twell et al. 1989). Two major groups of genes were recognised: ‘early’ genes first expressed in the microspore, which decreased in abundance before pollen maturation, and ‘late’ genes expressed after PMI, which accumulated until maturity, although more complex patterns were apparent (reviewed by Mascarenhas 1990, 1992; Twell 1994, 2002). A further level of cell-specific control of transcription was recognised when the tomato *LAT52* promoter was linked to a nuclear-targeted GUS fusion protein gene, which revealed that the *LAT52* promoter was active in the vegetative nucleus, but not in the generative cell (Twell 1992).

The first predictions of pollen transcriptome size were based on poly(A)RNA–cDNA association kinetic ( $R_{ot}$ ) analysis, from which maize and *Tradescantia* pollen were estimated to express around 20,000 different transcripts, representing around 30 % fewer than in shoots (Willing and Mascarenhas 1984; Willing et al. 1988). It is only in the past decade, however, that genomic platforms have been used to validate these estimates and to explore the full repertoire of genes expressed throughout the different phases of pollen development. Some of the important advances are charted in Fig. 3, together with the remarkable increase in the annual rate of publications reporting on pollen and transcription.

### Pollen gene expression in the post-genomic era

By the turn of this century, gene-by-gene studies had identified around 150 different pollen-expressed genes associated with a range of functions, including cell wall metabolism, cytoskeleton, transcription, and signalling (see Twell 2002). The plants investigated were numerous (>28 species) and diverse contrasting with the studies of the subsequent decade, which initially were focussed on the model *Arabidopsis thaliana*. The obvious limitations of gene-by-gene studies were overcome with advances in transcriptome profiling, which included serial analysis of gene expression (SAGE) and microarray technology, allowing users to simultaneously monitor the expression of thousands of genes. For example, the first-generation Affymetrix *Arabidopsis* AG array or 8-K GeneChip harboured probe sets for around 8100 genes, approximately one-third of the *Arabidopsis* gene models known from EST collections and cDNA libraries (Richmond and Somerville 2000).



**Fig. 3** A timeline charting some of the important advances in male gametophyte transcriptome studies. The developmental stage, cell type, technique, species, and associated publications are indicated. The remarkable increase in the rate of published work associated with

the terms 'pollen' and 'transcript\*' is illustrated by the *line graph* indicating the numbers of publications per year (Web of Science™ version 5.16)

The transcriptome of mature pollen was first investigated in three studies, which provided a coherent overview of pollen gene expression, two utilising 8-K microarrays (Honys and Twell 2003; Becker et al. 2003) and the third SAGE technology (Lee and Lee 2003). The microarray studies led to the identification of 992 (Honys and Twell 2003) and 1587 (Becker et al. 2003) genes expressed in mature pollen estimating the total number of pollen-expressed genes in *Arabidopsis* between 3500 and 5500. The pollen transcriptome showed a unique profile and was enriched for genes involved in signalling, cell wall metabolism, and the cytoskeleton, but under-represented for energy metabolism, transport, transcription, and translation. The presence of transcripts for a large number of cell wall-associated proteins (e.g. glycoside hydrolases, polygalacturonases, and cellulases) and cytoskeletal components (e.g. actin and profilin), highlighted a significant role for storage and translation of mRNAs after pollination, where pollen germination and tube growth require significant modifications to the cell wall and dynamic regulation of the actin cytoskeleton (Honys and Twell 2003). The enrichment of signalling components in the pollen transcriptome was also in accordance with the important role of pollen–pistil interactions and female gametophytic signals, which support and guide the pollen tube to the ovule (reviewed by Palanivelu and Tsukamoto 2012).

With the release of the *Arabidopsis* genome, the second-generation ATH1 Genome Array (Affymetrix) was

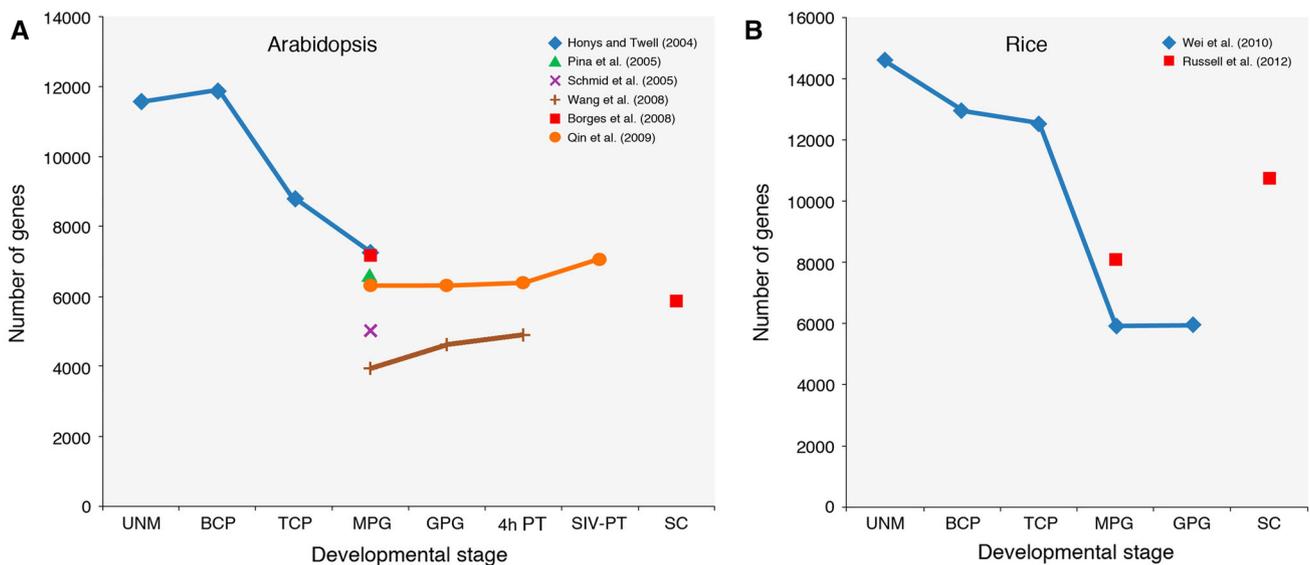
developed, harbouring probe sets for around 24,000 genes (Redman et al. 2004). However, around 3000 genes had no representation on the ATH1 platform, including important regulators, such as the male germline-specific transcription factor DUO1 (Rotman et al. 2005; Brownfield et al. 2009a, b). The ATH1 platform represented a major advance and a substantial body of data now exists for sporophyte-free *Arabidopsis* pollen, with >130 raw data files available in Gene Expression Omnibus and ArrayExpress (Tables S1 and S2). These data have not only provided the essential raw materials required to decipher the contribution of individual genes and gene families to male gametophyte development, but also guided the design of reverse genetics approaches and the analysis of gametophytic pollen mutants (Table S1). For example, pollen developmental transcriptome data were initially exploited for the analysis of cation/H<sup>+</sup> exchanger proteins (Sze et al. 2004) and more widely for all putative transporters (Bock et al. 2006), enabling functionally redundant proteins, such as CHX21 and CHX23, required for pollen tube guidance to the ovules to be identified (Lu et al. 2011).

Given that near-comprehensive transcriptome data sets have been generated in *Arabidopsis* for different phases of pollen development, including the germline and progamic phase (Table 1), we review these to highlight the major advances, followed by a discussion of the impact of comparative pollen transcriptome studies (see previous reviews by Twell et al. 2006; Becker and Feijo 2007; Schmidt et al. 2012).

**Table 1** A summary of male gametophyte stage-specific transcriptome studies for different plant species

	UNM	BCP	TCP	MPG	PT	GC	SC	VC
<i>A. thaliana</i>	X	X	X	X	X		X	
<i>O. sativa</i>	X	X	X	X	X		X	X
<i>N. tabacum</i>	X	X	X <sup>1</sup>	X	X		X <sup>EST</sup>	
<i>Z. mays</i>				X			X <sup>EST</sup>	
<i>L. longiflorum</i>				X	X	X <sup>EST</sup>		
<i>G. max</i>				X				
<i>V. vinifera</i>				X				
<i>F. vesca</i>				X				
<i>S. lycopersicum</i>				X				
<i>P. zeylanica</i>							X <sup>EST</sup>	

Developmental stages: UNM, microspore; BCP, bicellular pollen; TCP, tricellular pollen; MPG, mature pollen; PT, pollen tube; GC, generative cell; SC, sperm cell; VC, vegetative cell. X, signifies at least one microarray or RNA-seq data set. X<sup>EST</sup>, EST data is published for *N. tabacum* (Xin et al. 2011) and *Z. mays* (Engel et al. 2003) sperm cells and *L. longiflorum* generative cells (Okada et al. 2006). X<sup>1</sup>, generative cell division occurs during pollen tube growth in *N. tabacum*. See Table S1 and Table S2 for additional references



**Fig. 4** Plots of transcriptome sizes at different stages of male gametophyte development based on studies using Arabidopsis ATH1 (a) or rice 57 K (b) microarrays. UNM, uninucleate microspores;

BCP, bicellular pollen; TCP, tricellular pollen; MPG, mature pollen; GPG, germinating pollen; 4-h PT, 4-h in vitro-grown pollen tubes, SIV-PT, semi-in vivo pollen tubes. SC, purified sperm cells

### The Arabidopsis pollen transcriptome

Several groups have provided independent estimates of the number of genes expressed in mature pollen, which range from 3954 to 7235 genes, with an average of 6044 genes (Hony and Twell 2004; Pina et al. 2005; Schmid et al. 2005; Borges et al. 2008; Wang et al. 2008; Qin et al. 2009—see Fig. 4a). The use of different Arabidopsis accessions, pollen states (desiccated or hydrated), sample collection methods, and detection call algorithms may have contributed to the large variation in these estimates. Similarly, estimates of the percentage of genes that are

pollen-specific also vary, from 4 to 11 %, depending on the normalisation algorithms used and the number and diversity of sporophytic data sets used for comparison (Boavida et al. 2005; Twell et al. 2006). There are however two major findings common to all studies: first, the unique composition of the pollen transcriptome; second, its striking reduction in complexity compared with sporophytic tissues and purified sporophytic cell types, such as root hair cells (11,696 genes; Becker et al. 2014) and stomatal guard cells (13,222 genes; Bates et al. 2012).

RNA-seq is a powerful alternative approach to study the pollen transcriptome, with higher sensitivity and a

broader dynamic range (Zhao et al. 2014), which overcomes the limitations of promiscuous probe sets and gene models not represented on the ATH1 microarray. In the first RNA-seq analysis of Arabidopsis pollen, Loraine et al. (2013) reported at least 4172 protein-coding genes to be expressed. Although the number of genes detected was lower than estimates based on the majority of ATH1 studies, a highly conservative threshold of five reads per million (RPM) was applied, and reads mapping to multiple locations in the genome assembly were discarded. Reprocessed data now available in the online version of Loraine et al. (2013) provide normalised data as reads per kilobase per million (RPKM). In pollen, 6722 genes were expressed at 1 RPKM or above, including 6473 protein-coding genes, which is greater than the average number of pollen-expressed genes (6044) detected using ATH1 arrays. Illustrating the detection sensitivity of this threshold, we found that transcripts for two well-known male germline-specific genes, *DUO1* and *GCS1/HAP2*, were absent from seedlings but present in pollen at 2.4 and 3.0 RPKM, respectively. Nevertheless, the overlap between pollen ATH1 and RNA-seq data was almost complete, with <1 % of genes reliably detected on the ATH1 array not being found in RNA-seq data. Interestingly, 11 % of the genes detected by RNA-seq had no corresponding probe sets on the ATH1 array, revealing a previously unknown group of pollen-expressed genes. Pollen-expressed genes with additional exons and genes with previously unannotated 5' and 3' untranslated regions were also identified, although fewer than 20 genes annotated as alternatively spliced in TAIR10 were differentially spliced between pollen and seedlings. The RNA-seq data set of Loraine et al. (2013) thus provides a useful resource for the evaluation of annotation accuracy, alternative splicing, and a point of reference for future RNA-seq-based studies of Arabidopsis pollen gene expression.

### The Arabidopsis male germline transcriptome

The male germline makes a unique contribution to the pollen transcriptome (Engel et al. 2003). This was first revealed in Arabidopsis by profiling RNA from fluorescence-activated cell sorting (FACS)-purified sperm cells on the ATH1 platform (Borges et al. 2008). The sperm cell transcriptome of 5829 genes was smaller than that of mature pollen (7177 genes) and showed considerable divergence, with 65.4 % of genes in common (Fig. 4a). Compared with seedlings and pollen, around 2400 transcripts were enriched in sperm, with 642 (11 %) detected only in sperm. Wider comparisons with sporophytic data sets refined a set of 74 sperm-preferential transcripts likely to be sperm cell-specific. While the major GO categories

enriched in pollen were membrane transport, signalling, and vesicle trafficking, among sperm cell-preferential transcripts, those showing the most enrichment were DNA repair, ubiquitin proteasome system, and cell cycle, consistent with the propagation and maintenance of germline integrity (Pina et al. 2005; Borges et al. 2008). Compelling evidence for the role of the ubiquitin proteasome system in male germline cell cycle progression includes the role of the SCF<sup>FBL17</sup> E3 ubiquitin ligase complex (Kim et al. 2008; Gusti et al. 2009) and a pair of Ub-specific proteases (UBP3/UBP4) (Doelling et al. 2007). Small non-coding RNA pathways and DNA methylation pathways were also upregulated in sperm compared with the vegetative cell (Borges et al. 2008). For example, the DNA methyltransferase (MET1) is enriched in sperm, consistent with the active role of MET1 in the maintenance and epigenetic inheritance of CG-context methylation (Saze et al. 2003; Saze 2008; Calarco et al. 2012).

The sperm cell transcriptome has been used to inform the selection of genes of interest for further study. For example, it was used effectively to help distinguish spurious from genuine target genes, following ectopic expression of the male germline-specific MYB transcription factor, *DUO1*, in seedlings (Borg et al. 2011). Application of a sperm cell filter reduced the number of potential *DUO1*-activated target genes (DATs), from 124 to 63 genes, including those shown to play a role in gamete fusion (*GCS1/HAP2*; Mori et al. 2006; von Besser et al. 2006), sperm–egg adhesion (*GEX2*; Mori et al. 2014), and more recently, germ cell division and sperm cell specification (*DAZ1* and *DAZ2*; Borg et al. 2014).

### The Arabidopsis developmental transcriptome

Developmental resolution to the transcriptome was enabled by the use of density gradient centrifugation to separate four stages from microspore to mature pollen (Honys and Twell 2004): uninucleate microspore (UNM), bicellular pollen (BCP), tricellular pollen (TCP), and mature pollen grain (MPG). In total, 13,977 genes showed male gametophyte expression, with transcriptome size decreasing progressively from 11,565 genes in UNM to 7235 genes in MPG (Fig. 4a). In contrast, the percentage of pollen-specific genes increased from 6.9 % at UNM to 8.6 % at MPG, reflecting the differentiation and functional specialisation of mature pollen. Pairwise comparisons showed expression profiles to be well correlated for UNM and BCP ( $R = 0.96$ ) and for TCP and MPG ( $R = 0.86$ ); however, profiles were less similar for BCP and TCP ( $R = 0.54$ ) stages.

This study was the first to document dynamic changes in gene expression during male gametophyte development

and to quantify the numbers of genes contributing to early and late expression programs on a genomic scale (reviewed in Twell et al. 2006). This study was also the first to perform hierarchical cluster analysis, allowing co-regulated genes to be identified. Different co-expressed clusters of genes were found in UNM-BCP compared with TCP-MPG stages, revealing a phase shift in gene expression between BCP and TCP stages. Core cell cycle genes and transcription factors were enriched in UNM-BCP, while genes for signalling and cell wall metabolism were overrepresented in TCP-MPG, consistent with the early proliferative and late differentiation phases of pollen development. Moreover, the transcriptome of cell suspension cultures was more similar to UNM ( $R = 0.47$ ) than to MPG ( $R = 0.13$ ), reflecting the undifferentiated and proliferative state of microspores. Here, it is interesting to note the role of auxin in maintaining dedifferentiation in cell cultures with the peak levels of bioactive auxin, which occurs in microspores (Cecchetti et al. 2008).

The reported profiles of 607 transcription factors (TFs) expressed during pollen development (Honys and Twell 2004) have been exploited in reverse genetic approaches to study regulatory networks (Verelst et al. 2007a; Gibalova et al. 2009; Leydon et al. 2013, Liang et al. 2013; Xia et al. 2014), as well as in the broad phenotypic screening of T-DNA insertions in TF genes (Renak et al. 2012). For example, while the MADS family are collectively under-represented in the pollen transcriptome (Honys and Twell 2004), the non-classical lineages, including the AtMIKC\* genes, are overrepresented (Pina et al. 2005). In the first application of these data, Verelst et al. (2007a) explored the function of five MIKC\* MADS-box proteins (AGL66, AGL104, AGL67, AGL65, and AGL94), which are co-expressed in late stages of pollen development. MIKC\* heterodimer pairs were shown to bind MEF2-type CArG-box motifs, which were found to be highly enriched in the promoters of genes selectively expressed in TCP/MPG during pollen maturation (Verelst et al. 2007a). Moreover, *agl66/104* double mutants showed functional redundancy with severe in vitro germination defects and reduced male transmission. In a further pioneering study, Verelst et al. (2007b) profiled the pollen of double and triple mutants, providing a unique insight into the complexity of the MIKC\* TF network that directs a cellular differentiation network during pollen maturation (Table S1). This was further elaborated upon to include phenotypic analysis and profiling of pollen from quadruple mutants, highlighting the large-scale changes in transcripts associated with the absence of MIKC\* and supporting the model of Verelst et al. (2007b), whereby the MIKC\* network represses genes for early development and activates pollen maturation genes (Adamczyk and Fernandez 2009).

## The *Arabidopsis* progamic phase pollen transcriptome

In the first study to show transcriptome changes during pollen germination and tube growth of *Arabidopsis*, Wang et al. (2008) used ATH1 arrays to profile mature (dry) pollen (MP), germinating pollen (GP) cultured in vitro for 45 min, and pollen tubes (PT) after 4 h of in vitro culture. The size of the transcriptome increased from 3945 genes in MP to 4892 genes in PT (Fig. 4a). The percentage of stage-specific genes also increased progressively, from 4.1 % at MP to 14.9 % in PT, with more genes being up- and down-regulated from GP to PT than from MP to GP. Some biological processes enriched in germinating pollen were also overrepresented in pollen tubes, including cell wall metabolism, signalling, and cellular transport. Processes showing the greatest change during pollen germination were stress response and transcription, while in pollen tubes, metabolism and signalling components showed the greatest change. These observations reflect physiological differences, with germinating pollen activated from a quiescent state and pollen tubes actively growing by tip extension.

Communication between pollen and pistil and the stylar environment has important roles in pollen tube growth in vivo (see Taylor and Hepler 1997; Palanivelu and Tsukamoto 2012). For example, it takes around 4 h for the pollen tube to reach the ovule in *Arabidopsis* (a distance >400  $\mu\text{m}$ ), but at 4 h, in vitro-grown pollen tubes are only about 150  $\mu\text{m}$  (Wang et al. 2008). In addition, the growth of pollen tubes in vitro is non-directional and targeting of ovule explants is low, unlike ‘conditioned’ pollen tubes grown through the pistil (Palanivelu and Preuss 2006).

Three studies have examined the transcriptional landscape of *Arabidopsis* pollen tubes, factoring in pollen tube crosstalk with the pistil and pollen tube guidance (Qin et al. 2009; Chen et al. 2014; Lin et al. 2014).

To examine the role of ‘pistil conditioning’, Qin et al. (2009) compared the transcriptome of semi-in vivo pollen tubes (SIV-PT) grown through stigma–style explants with those of mature pollen and in vitro-grown pollen tubes (PT) at 0.5 h and 4 h. Although similar numbers of genes were expressed in mature pollen and in vitro pollen tubes, SIV-PT expressed a greater number (7044 genes; Fig. 4a). SIV-PT also had the largest number of pollen stage-specific genes (1254), suggesting that growth through the pistil elicits a significant change in the pollen tube transcriptome. Further, a set of 383 genes, uniquely expressed in SIV-PT, was enriched for genes involved in signalling (e.g. transmembrane receptors and protein kinases), defence response (e.g. TIR-NBS-LRR receptors), and cell extension (transporters and antiporters). Reverse genetic analysis of 33 genes induced during pollen tube growth identified five required for optimal tube growth in vitro, while mutations

in two further genes showed pollen tube navigation defects *in vivo*. One of the *in vivo* effect genes, At3g18000 (*XIPOTL*), was specifically up-regulated in SIV-PT and is required for phosphatidylcholine synthesis (Cruz-Ramirez et al. 2004), implicating lipid signalling and/or plasma membrane composition in pollen tube growth through the pistil.

Chen et al. (2014) extended the semi-*in vivo* approach to factor in the role of diffusible pollen tube attractants and monitored the transcriptome of semi-*in vivo*-guided pollen tubes (SIV-PG) by positioning ovules below pollen tubes emerging from pistil explants. Using a 12 × 135 K Arabidopsis gene chip (Roche-NimbleGen), 719 genes were identified to be specifically expressed in SIV-PG. Gene families, such as defensin-like (DEFL), leucine-rich repeat receptor like kinases (LRR-RLKs), and TIR-NBS-LRR receptors, were specifically enriched in SIV-PG, but surprisingly there was little overlap with candidate pollen tube guidance genes from the SIV-PT system of Qin et al. (2009). Phenotypic screening of 18 confirmed T-DNA insertion lines for pollen tube guidance defects was unsuccessful, highlighting potential redundancy among the often large, gene families involved. However, in a dominant negative approach involving the expression of a ‘kinase-deleted’ variant, phenotypes consistent with pollen tube guidance defects were observed for one LRR-RLK gene (Chen et al. 2014).

Lin et al. (2014) used a novel application of the ‘TRAP’ method (translating ribosomes affinity purification) to the male gametophyte, allowing *in vivo* gene expression data to be gathered for the progamic phase. This approach involves immunopurification (IP) of ribosome-associated mRNA to identify transcripts undergoing active translation—the ‘translatome’ (Zanetti et al. 2005). Transgenic plants in which epitope-tagged ribosomal protein L18 was expressed from the pollen-specific LAT52 promoter were used for mRNA–ribosome complex isolation from un-pollinated buds (IP-bud), open flowers (IP-*in vivo*), as well as 4.5 h *in vitro*-grown pollen tubes (IP-*in vitro*). The LAT52 promoter is active in the late microspores and after PMI in the vegetative cell of Arabidopsis and is expected to capture the vegetative cell translatome (Eady et al. 1994; Grant-Downton et al. 2013). In IP-bud samples, 8140 genes had corresponding transcripts undergoing translation with a similar number in IP-*in vivo* samples, but significantly fewer (5188 genes) than in the IP-*in vitro* pollen tubes. In the IP-*in vivo* data set, 519 genes were specifically enriched, implicating these genes in the growth and guidance of the pollen tube in the gynoecium and in response to fertilisation. This set of genes was enriched for molecular functions associated with heme, which has been implicated in sperm cell discharge through pollen tube rupture (Lin et al. 2014). Whereas an overlap of 13 % (67/519) was

observed between IP-*in vivo*-enriched and pollination-induced mRNAs (Boavida et al. 2011), there was only 4 % (17/383) overlap between semi-*in vivo* and IP-*in vivo*, pollination-induced mRNAs, indicating a significant distinction between enriched transcripts from semi-*in vivo* and true *in vivo* pollen tubes. Reverse genetic analysis identified mutations in three IP-*in vivo*-enriched genes, *iv2* (methyl esterase 8, MES8), *iv4* (glutathione s-transferase, GSTU26), and *iv6* (xyloglucan endotransglucosylase/hydrolase 19), which showed reduced male transmission. These mutants showed defective micropylar guidance, pollen tube burst, and ovules receiving multiple pollen tubes (polytubey).

Collectively, the transcriptome analyses outlined above establish some unifying features of Arabidopsis pollen development. First, mature pollen grains possess the most unique transcriptome with the least transcript diversity when compared with the sporophyte. Second, the transition from proliferating microspores to differentiated pollen is associated with large-scale repression and *de novo* transcription of an increased number of pollen-specific genes. Third, sperm cells possess a similarly reduced, but diverse and unique transcriptome. Fourth, the germination and directional growth of the pollen tubes involves stored transcripts and *de novo* transcriptional responses associated with signalling and communication with the gynoecium. The distinctive transcriptomes associated with the different phases of pollen development and the germline are contrasted with those of the sporophyte in the principal component analysis shown in Fig. 5.

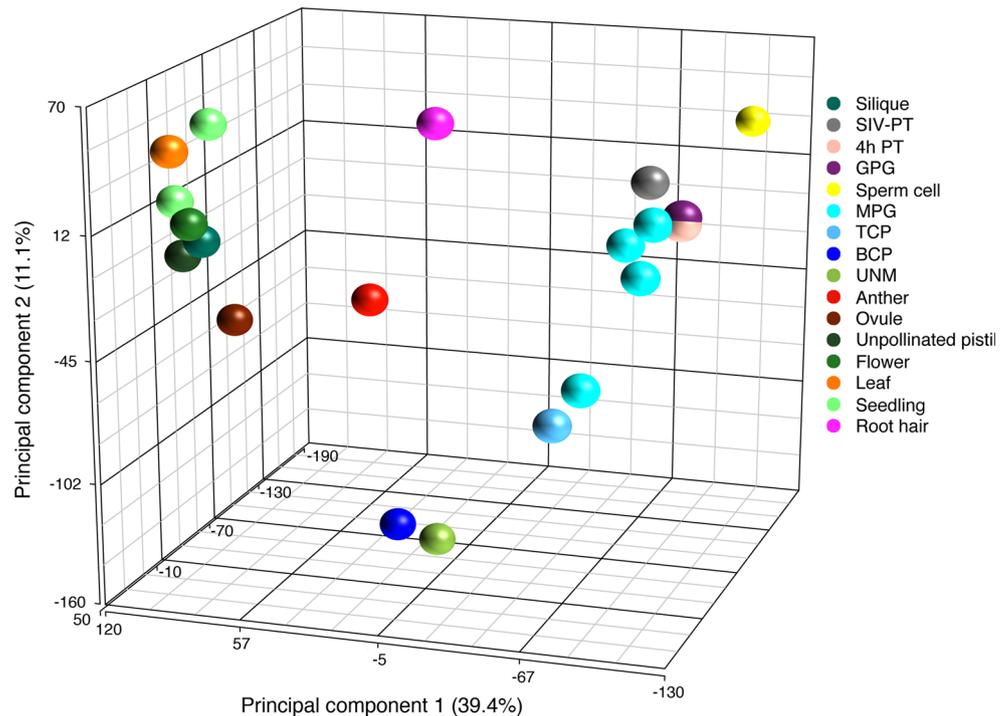
### Comparative pollen transcriptomics

The analysis of pollen transcriptome profiles for diverse angiosperms enables the conservation of the molecular mechanisms underlying the different phases of pollen development to be explored. Pollen microarray and/or RNA-seq data are available for at least 10 different species including Arabidopsis. Lists of the current data sets are compiled in Tables S1 and S2. Transcriptome profiling at multiple points in development has only been reported for Arabidopsis, cultivated rice, and tobacco (see Table 1), which differ in specific features, such as the timing of germ cell division and length of the progamic phase.

#### Rice

The transcriptome of developing rice pollen has been investigated in two independent studies, the first using laser microdissection of tapetum, microspores, and pollen, and the 44-K rice oligo microarray platform (Suwabe et al. 2008; Hobo et al. 2008). K-means cluster analysis of the

**Fig. 5** Principal component analysis (PCA) of ATH1 microarray data for male gametophyte and selected sporophytic samples. The data sets included are from the following publications: Honys and Twell (2004), Borges et al. (2008), Qin et al. (2009), Phan et al. (2011), Becker et al. (2014). PCA was generated using RMA-normalised data in Partek® Genomics Suite® (Partek Inc., Missouri)



data revealed eight gene clusters with synchronous gene expression in tapetum and microspores involving a total of 10,810 genes. Within these clusters, there was enrichment for transcription, secondary metabolism (e.g. chalcone synthase, chalcone reductase, and phytoene synthase), and lipid synthesis and metabolism processes associated with construction of the pollen wall. These observations counter the view of the tapetum simply as a nurse cell layer, recognising shared pathways and the common developmental origins of tapetal cells and microspores (Scott et al. 2004; Hobo et al. 2008).

A subsequent study using the 57 K Affymetrix Rice Genome Array (Wei et al. 2010) showed a progressive decrease in the number of expressed genes, from 14,590 genes in uninucleate microspores to 5939 genes in mature pollen and a similar number (5945 genes) in germinated pollen (Fig. 4b). Similar to Arabidopsis, the rice pollen transcriptome was highly reduced compared to the sporophyte, where 17,383 genes were detected in roots and 17,242 in leaves. The overall patterns of genes expressed mirrored observations from Arabidopsis, with highly correlated early (UNM and BCP;  $r = 0.82$ )- and late (TCP and MPG;  $r = 0.76$ )-stage profiles and a distinct phase shift between. The authors described a ‘U-type’ change for pollen-preferential or stage-specific transcripts in rice and Arabidopsis, with the least number of genes preferentially expressed at bicellular stage. This trend reflects the bicellular stage as transitional between early proliferative and late differentiation phases. Compared with Arabidopsis,

rice showed a steeper decline in the number of genes expressed between TCP and MPG stages, suggesting differences in the onset of large-scale transcriptional repression and/or in the rates of transcript turnover (Fig. 4).

Despite overall similarities, significant differences were noted in the distribution of GO categories for genes expressed at different stages between rice and Arabidopsis. For example, rice expressed more stage-enriched transcripts associated with defence and stress response in mature pollen, while rice pollen was more enriched for signalling at bicellular stage. Overall comparisons revealed that 62.4 % (1195 genes) of stage-enriched genes from rice had homologues in the Arabidopsis genome; however, only 56.6 % (677 genes) of these were expressed, with a very small proportion, 3.2 % (22 genes), being preferentially expressed in Arabidopsis pollen. Moreover, the analysis of regulatory proteins identified many TF classes, which were shared between rice and Arabidopsis, but the expression of a set of unique TFs in rice pollen indicates that regulatory networks will show significant differences (Wei et al. 2010).

Direct evidence for the conservation of an important TF network between rice and Arabidopsis includes a study of the MIKC\* MADS-box network, which is required for pollen maturation (Liu et al. 2013). While Arabidopsis encodes five pollen-expressed MIKC\* genes, the rice genome has three genes divided between S- (OsMADS62, OsMADS63) and P- (OsMADS68) clades, which show conserved expression in pollen. Similar to observations in

Arabidopsis, mutants affecting rice P- or S-clade MIKC\* genes displayed defects in pollen germination in vitro, but also showed reduced pollen viability and abnormal starch accumulation (Liu et al. 2013). Moreover, the transcriptome data of Wei et al. (2010) were used to show that the promoters of genes expressed during late pollen development were enriched for MIKC\* binding sites. This work indicates that the function of the MIKC\* regulatory network in pollen development has been conserved, since the divergence of eudicots and monocots, some 150 million years ago (Liu et al. 2013).

#### Tobacco

Transcriptome analysis of the male gametophyte from tobacco using the Agilent 44-K tobacco microarray, originally consisting of three stages (mature pollen, 4, and 24 h in vitro pollen tubes; Hafidh et al. 2012a, b), was recently extended to include earlier stages of development (uninucleate microspores, early bicellular pollen, and late bicellular pollen; Bokvaj et al. 2015). Similar to the observations in Arabidopsis and rice (Fig. 4), transcriptome complexity was reduced during pollen maturation in tobacco, although the onset of the decline was later, with an unusually low diversity of transcripts (16,700) detected in microspores relative to early and late bicellular pollen (22,872 and 19,750, respectively). Although some features, such as the abundance of translational components at early stages and the overrepresentation of transport and cytoskeleton at late stages of pollen development (Bokvaj et al. 2015), reflected patterns observed in Arabidopsis and rice, the temporal shift in maximum transcriptome complexity from microspore to the bicellular stage may reflect different demands on the rate of development, and it will be interesting to examine whether this is a general feature of species that shed bicellular pollen. Division of the germ cell in tobacco occurs 8–12 h after pollen germination, such that 4-h pollen tubes harbour M-phase generative cells, while 24-h pollen tubes contain sperm. During tobacco pollen tube growth, a moderate increase in the number of detectable transcripts was observed, from 13,966 in mature pollen to 14,420 in 24-h pollen tubes, with 3597 transcripts in common between these three stages. Transcripts of 699 genes (4.8 %) accumulated significantly (>fivefold) compared to mature pollen and the 4-h pollen tube, while 320 genes (2.2 %) accumulated de novo after 4 h, highlighting the extended transcriptional capacity of pollen tubes cultivated in vitro (Hafidh et al. 2012a, b). The delay in germ-cell-cycle progression, until 10–12 h of in vitro pollen tube growth, was linked with the presence of transcripts for cell cycle inhibitors, such as RBR1 and DEL3. Transcripts for the G2/M-phase activator CYCB1;1 peaked during the transition to PMII, together

with the expression of the tobacco orthologue of *AtDUO1* (*NtDUO1*; B25, Kyo et al. 2003), which is required for germline CYCB1;1 accumulation in Arabidopsis (Brownfield et al. 2009a). Although the expression of cell cycle repressors and activators requires validation in tobacco germ cells, these data support the conservation of DUO1-mediated mitotic progression and differentiation between Arabidopsis and tobacco.

#### Other species

Other species for which mature pollen transcriptome data have been published include maize (Ma et al. 2008; Davidson et al. 2011; Chettoor et al. 2014), soybean (Haerizadeh et al. 2009), grapevine (Fasoli et al. 2012), potato (Sanetomo and Hosaka 2013), woodland strawberry (Hollender et al. 2014), and most recently lily (Lang et al. 2015).

The maize pollen transcriptome has been analysed in three independent studies focussed on reproductive development. In the microarray study of Ma et al. (2008), 10,545 different transcripts were detected in maize pollen, which was extended in the two RNA-seq experiments to 13,418 genes (Davidson et al. 2011) and 14,591 genes (Chettoor et al. 2014). The maize pollen transcriptome was typically reduced in complexity, expressing, for example, only 57 % of the mean number of genes expressed in 12 other sporophytic tissues (Davidson et al. 2011). Moreover, pollen was also among tissues with the most extreme range of transcript levels, showing the highest maximum expression levels (Davidson et al. 2011). The classification of array data (Ma et al. 2008) into high (237 transcripts) and medium (5547) abundance classes was also strikingly congruent with the original estimates (240 and 6000, respectively) for maize pollen based on  $R_{0t}$  curves (Willing et al. 1988). However, the low copy number, slow to hybridise transcript class, appears to be substantially overestimated in  $R_{0t}$  hybridisations (Ma et al. 2008). Comparative analysis by Ma et al. (2008) indicated significant conservation in gene expression programs between Arabidopsis and maize pollen. Of 4407 homologues of Arabidopsis pollen-expressed genes represented on the maize array, 3444 (78 %) were expressed in maize pollen and highly expressed genes in each species showed substantial overlap.

Since maize is an allotetraploid consisting of two subgenomes, the relative contributions of each subgenome to the transcriptome can be determined to explore potential bias between the gametophyte and sporophyte. Previous studies revealed that subgenome 2 is characterised by reduced expression and a higher rate of gene loss, relative to subgenome 1 (Schnable et al. 2011). However, in their study, Chettoor et al. (2014) discovered, that the pollen

transcriptome was overrepresented for subgenome 2, compared to the other tissues assessed. The basis for this difference was linked to the retention of a greater proportion of duplicate gene pairs derived from both subgenomes for pollen, with both members of duplicate pairs also more likely to be pollen-enriched. This suggests that selective pressure could be acting on pollen to maintain functional copies of both homeologues following tetraploidisation, due to increased sensitivity of pollen to gene dosage effects (Chettoor et al. 2014).

The soybean (*Glycine max*) pollen transcriptome is the first described for a legume and for a bicellular pollen species (Haerizadeh et al. 2009). Soybean pollen expressed 10,299 transcripts representing only 37 % of those detected in sporophytic tissues, while 7.9 % were pollen-specific. The most abundant transcripts were enriched for cell wall-modifying enzymes, signalling genes, and transporters, typical for *Arabidopsis* pollen. There was divergence, however, in the number and diversity of regulatory proteins, and small RNA pathway proteins were underrepresented. Heat-shock proteins and heat-shock TFs were up-regulated in mature soybean pollen, in contrast to *Arabidopsis*, in which they were up-regulated specifically in germinating pollen and proposed to act as molecular chaperones to accommodate the intense physiological activities associated with pollen germination and tube growth (Wang et al. 2008).

Pollen microarray data for grapevine (*Vitis vinifera*) were generated as part of a gene expression atlas for this species (Fasoli et al. 2012). Similar to other studies, mature pollen showed a distinctive transcriptome including enrichment for cell wall-modifying enzymes. Although reduced in complexity compared to some sporophytic samples, in contrast to previous studies, the pollen transcriptome was not among the least complex, with berries, tendrils, and stems showing lower complexity, perhaps reflecting the specialised activities of some tissues and the diversity of tissue samples analysed.

The pollen transcriptome of woodland strawberry (*Fragaria vesca*) was generated by as part of a study to develop *F. vesca* into a model plant for the Rosaceae (Shulaev et al. 2011). The transcriptomes of mature pollen and microspores isolated by laser capture microdissection were determined by RNA-seq (Hollender et al. 2014). Similar to studies in *Arabidopsis* and rice, *F. vesca* mature pollen and microspores showed distinctive transcriptomes, with mature pollen showing the least complexity (11,540 genes) compared with microspores (33,109 genes) or all floral tissues combined (34,527 genes).

Lily (*Lilium longiflorum*) provides an established model for pollen germination and tube growth, with significant advantages for physiological and biochemical analyses; however, only limited molecular data are available. The

recent analysis by Lang et al. (2015), involving the RNA-seq of a normalised cDNA library (pooled from dry, hydrated, germinating pollen and pollen tubes), provides the first comprehensive overview of the lily pollen transcriptome. Assembled transcripts revealed conserved features when visualised with MAPMAN software tools and compared with RNAseq data from *Arabidopsis* pollen (Loraine et al. 2013). This study provides useful data sets and tools to search for lily sequences of interest enabling comparative studies.

### Comparative male germline data

Exploring gene expression in the male germline has benefitted from high-throughput screens using EST libraries. This approach overturned previous notions of transcriptionally inert male gametes and delivered valuable large-scale data for generative cells of lily (Okada et al. 2006), and for sperm cells of maize (Engel et al. 2003), *Plumbago* (Gou et al. 2009), and tobacco (Xin et al. 2011). The comparative analysis of maize sperm EST data identified the conserved gamete-expressed genes, *GEX1* and *GEX2* in *Arabidopsis* (Engel et al. 2005), which have important roles and use as research tools in gamete biology (Brownfield et al. 2009a, b; Alandete-Saez et al. 2011; Mori et al. 2014).

The in-depth profiling of purified sperm cells from a crop species was first achieved for rice using the 57-K microarray (Russell et al. 2012). Strikingly, rice sperm cells expressed 10,732 genes, twice as many as sperm cells from *Arabidopsis*. Rice sperm cells were also found to express more genes than mature pollen (8101 genes), in contrast to *Arabidopsis* sperm, which express fewer than in mature pollen (Fig. 4). Similar to *Arabidopsis*, rice sperm cells were enriched for genes associated with ubiquitin pathways, DNA modification and repair, and chromatin remodelling. The enrichment of cell cycle genes observed in *Arabidopsis* sperm (Borges et al. 2008), however, was not reported in rice, consistent with the arrest of Poaceae sperm in G1 (Friedman 1999). The expression of more than 70 sperm-enriched TFs was reported in rice, and homologues of several genes in the *Arabidopsis* DUO1 regulatory network were also selectively expressed, including OsDUO1 (although not present on the array), supporting the conservation of a key germline regulatory network (Russell et al. 2012).

A simple cell isolation procedure involving the fractionation of sperm cells from burst pollen grains has allowed the sperm cell transcriptome to be compared with that of sperm cell-depleted vegetative cells (Anderson et al. 2013). The number of genes reliably detected was 16,985 in sperm and 18,611 in vegetative cells, which reflects the reduced complexity of sperm compared with mature pollen of

*Arabidopsis* (Borges et al. 2008). This is in contrast to the findings in rice using the 57-K microarray, where more genes were detected in sperm than in mature pollen (Russell et al. 2012). The top 50 transcripts in sperm showed a broader range of expression with higher peak levels than those of the vegetative cell, indicating that the reduced size and transcriptional output of sperm cells do not limit the capacity of sperm to express individual genes (i.e. ATPase) at very high levels. Furthermore, the overlap between the vegetative and sperm cells was found to be much higher (by around 25 %) than between either of these and the egg cell, reflecting the common lineage of the sperm cells and vegetative cell.

In both studies of the rice sperm cell transcriptome, diversity in the complement of histones was noted, particularly type H2B and H3, with enrichment for at least one member of each of the five major histone types (Russell et al. 2012). All three homologues of HTR10 (MGH3), the *Arabidopsis* male germline-specific histone H3.3, are expressed in rice sperm cells, together with histone-modifying proteins. Histone modification genes were differentially expressed between the vegetative cell and sperm cells (Anderson et al. 2013). For the JUMONJI (JMJ) histone demethylase family, which uses histone H3 as a substrate and is involved in transcriptional repression, vegetative cells lacked most of the H3K36-modifying JMJ proteins expressed in sperm cells, with sperm cells expressing the H3K27 demethylase ZOS1-20 to higher levels than in the vegetative cell. In common with *Arabidopsis*, rice sperm cells express MET1, while transcript levels are very low in vegetative cells, consistent with the maintenance of CG methylation in the male germline and its absence in the vegetative cell (Calarco et al. 2012). Equally, DRM2 is expressed in the vegetative cell implying that RNA-directed CHH methylation is conserved. However, unlike *Arabidopsis*, components of the RNA-directed DNA methylation (RdDM) machinery are expressed in rice sperm, including DCL3, DRM2, and RDR2, although the latter is only reported by Anderson et al. (2013). Interestingly, a similar number of sequence reads was found to map to transposon-associated repeats in vegetative cells and in the sperm cells, and previous reports have described a high proportion (8 % of transcripts) of retroposon sequences in maize sperm (Turcich and Mascarenhas 1994; Engel et al. 2003). These findings indicate that differences may exist between rice and *Arabidopsis* in the mode of epigenetic regulation of siRNA-mediated gene silencing in the male germline.

### Integration and analysis of pollen transcriptome data

With the wealth of transcriptome data now available for the male gametophyte, meta-analyses can be used to identify

genes with common expression and activities between multiple data sets. The data can be used to construct *in silico*-derived co-expression networks and are based on the assumption that genes with similar expression patterns are likely to interact with each other. In a pioneering study in rice, Aya et al. (2011) adopted a co-expression analysis to identify ‘meiosis’ and ‘pollen wall synthesis’ sub-networks. The study analysed 176 microarray data sets, including LM-microarray data separating tapetum and male gametophyte cells, to identify a robust pollen wall synthesis co-expression network, enriched for sporopollenin-associated genes together with genes not previously implicated in exine formation. This study demonstrates the value of co-expression network analysis and how the separate transcriptomes of pollen and tapetum cells can increase the precision and resolving power of network construction.

Integrative approaches utilising data within and between species can also be used to explore common cellular processes in different cell types. Two independent studies have searched for a common set of genes associated with polar cell growth. As part of a study of the tobacco pollen tube transcriptome, Hafidh et al. (2012a) incorporated cross-species comparisons, including transcriptome and proteome data to identify *Arabidopsis* homologues of genes co-expressed in tobacco pollen tubes and roots (3264 genes), of which 78 genes overlapped with genes known to possess root hair-specific promoter motifs (Won et al. 2009). This subset, extended to a candidate list of 104 genes, showed enrichment for vesicle transport, signal transduction, translation, and cytoskeleton. Evidence for the function of three candidate ‘root hair–pollen tube’ genes in the progamic phase was shown, by transfection of antisense RNA, resulting in reduced *in vitro* pollen tube growth. In a different strategy, Becker et al. (2014) generated purified root hair and pollen transcriptomes in *Arabidopsis*, incorporating a distinction between apical (polar) and diffuse (non-polar) cell growth to identify a molecular signature for polar cell growth. This identified 4989 genes co-expressed in root hairs and pollen, with 277 showing enrichment compared to diffuse growth cell types and 105, which were exclusive to root hairs and pollen. An interesting finding based on promoter analysis of a subset of 49 co-expressed apical cell growth-specific genes was the identification of a sequence motif enriched in hypoxia-induced genes, supporting the hypothesis that tip growing cells experience anaerobic conditions (Becker et al. 2014). Despite the different strategies employed in the two studies, there was notable overlap, with 48 out of the 104 candidate polar cell expansion genes of Hafidh et al. (2012a) detected in *Arabidopsis* root hairs and pollen, but only three of these genes were found to be enriched compared with cell types showing non-polar growth (Becker et al. 2014).

Computational approaches involving meta-analyses have also been used to construct and explore a regulatory network of TF activity during pollen development (Wang et al. 2011). Network component analysis (NCA) was applied to a total of 23 data sets encompassing developmental and progamic phases to construct a regulatory network involving 19 TFs, 101 target genes, and 319 regulatory interactions. Although NCA can infer hidden TF activities by taking advantages of the prior knowledge of network structure, most of the regulatory information and regulatory pairs retrieved from co-expression analysis remain hypothetical. Nevertheless, NCA did include the pollen-specific factor WRKY34 (Honys et al. 2006), which is negatively regulated by MIKC\* transcription factors and involved in cold sensitivity and pollen maturation (Verelst et al. 2007b; Zou et al. 2010; Guan et al. 2014). Currently, none of the network predictions including TF interactions and target genes are reflected by experimental data, and further effort is required to gather empirical knowledge of system components and to validate and improve network models.

Most recently, pollen transcriptome data have been used in two phylostratigraphic studies, to address a novel hypothesis in plants concerning the origin of ‘new’ or ‘orphan’ genes. In animals, new genes, which provide material for evolutionary innovation, show testes-biased expression and are thus regarded as the birthplace for new genes in an “out of testes” hypothesis (Kaessmann 2010). Wu et al. (2014) propose a corresponding “out of pollen” hypothesis and Cui et al. (2015) a “young genes out of the male” hypothesis for the origin of new genes in flowering plants. In the studies of Wu et al. (2014) in *Arabidopsis* and Cui et al. (2015), which included *Arabidopsis* and rice, the phylogenetically youngest transcriptome was found in pollen, based on the analysis of transcriptome age index (TAI; Domazet-Lošo and Tautz 2010). Moreover, the sperm transcriptomes of *Arabidopsis* and rice were shown to be similarly young (Cui et al. 2015). Both studies highlight the enrichment of TEs in the new genes and the importance of epigenetics and chromatin states in the vegetative cell and/or the germ cells in shaping transcriptome age.

### Concluding remarks and outlook

Transcriptome studies of the male gametophyte have not only increased knowledge and understanding, but also improved the efficacy of experimental strategies: first, in describing transcript profiles (throughout development and in specific cell types); second, by informing experimental design (such as by gene selection for reverse genetics), through query-based and co-expression analysis; and third,

by providing the raw materials to build gene networks and a methodology to understand how these are affected in pollen mutants. All three approaches have been exploited (to a greater or lesser extent) by the *Arabidopsis* community, and progress encompassing all three approaches in other model and non-model species is expected as more data are accumulate, together with improved methods for exploring gene function, such as genome editing (Belhaj et al. 2013).

Microarrays have been the dominant profiling technology, providing a uniform platform from which to explore the expression of individual genes and gene families, to build co-expression networks and to aid the design of reverse genetics experiments. For multiple reasons, including reduced costs of replication, higher sensitivity, and information content, RNA-seq is set to become the dominant profiling technology in future studies (Liu et al. 2014). Since RNA-seq requires little a priori knowledge of the genome, this enables transcriptome-based studies of non-model species, such as *Capsella grandiflora*, in which pollen-specific genes were identified to show stronger purifying selection and higher rates of positive selection than sporophytic genes (Arunkumar et al. 2013). RNA-seq has also opened the way for the small RNA landscape of pollen to be explored (Grant-Downton et al. 2009; Slotkin et al. 2009; Zhang et al. 2009; Borges et al. 2011; Li et al. 2013). By integration of mRNA transcriptome data and other genome-wide information, such as bisulphite sequencing, our understanding of the epigenomic landscape of the gametophyte and male germline has also advanced (Calarco et al. 2012). Pollen transcriptomics can also be applied more widely to develop tools important to the biotechnology sector (Oo et al. 2014) and to address agronomically important questions such as tolerance to environmental stresses linked to global climate change. For example, pollen heat stress can cause sterility and associated crop losses, and efforts are underway to explore the molecular networks associated with thermotolerance in a broader system-level analysis (Bokszczanin et al. 2013; Giorno et al. 2013; Loraine et al. 2015).

**Author contribution statement** NR and DT wrote the manuscript. Both authors read and approved the manuscript.

**Acknowledgments** We thank the Biotechnology and Biological Science Research Council (BBSRC) for grant support and for the award of a studentship to NR. We also thank Matt Blades from the Bioinformatics and Biostatistics Analysis Support Hub (BBASH) for help with PCA, Ann Loraine for helpful discussions on *Arabidopsis* and tomato pollen RNA-seq data, and Jerome Twell for help in drafting Figs. 1 and 3.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Adamczyk BJ, Fernandez DE (2009) MIKC\* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiol* 149:1713–1723
- Alandete-Saez M, Ron M, Leiboff S, McCormick S (2011) *Arabidopsis thaliana* GEX1 has dual functions in gametophyte development and early embryogenesis. *Plant J* 68:620–632
- Anderson SN, Johnson CS, Jones DS, Conrad LJ, Gou X, Russell SD, Sundaresan V (2013) Transcriptomes of isolated *Oryza sativa* gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. *Plant J* 76:729–741
- Arunkumar R, Josephs EB, Williamson RJ, Wright SI (2013) Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. *Mol Biol Evol* 30:2475–2486
- Aya K et al (2011) Comprehensive network analysis of anther-expressed genes in rice by the combination of 33 laser microdissection and 143 spatiotemporal microarrays. *PLoS ONE* 6:e26162
- Baroux C, Raissig MT, Grossniklaus U (2011) Epigenetic regulation and reprogramming during gamete formation in plants. *Curr Opin Genet Dev* 21:124–133
- Bates GW et al (2012) A comparative study of the *Arabidopsis thaliana* guard-cell transcriptome and its modulation by sucrose. *PLoS ONE* 7:e49641
- Becker JD, Feijo JA (2007) How many genes are needed to make a pollen tube? Lessons from transcriptomics. *Ann Bot* 100:1117–1123
- 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
- Becker JD, Takeda S, Borges F, Dolan L, Feijo JA (2014) Transcriptional profiling of Arabidopsis root hairs and pollen defines an apical cell growth signature. *BMC Plant Biol* 14:197
- Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* 9:39
- Boavida LC, Becker JD, Feijo JA (2005) The making of gametes in higher plants. *Int J Dev Biol* 49:595–614
- Boavida LC, Borges F, Becker JD, Feijo JA (2011) Whole genome analysis of gene expression reveals coordinated activation of signaling and metabolic pathways during pollen-pistil interactions in Arabidopsis. *Plant Physiol* 155:2066–2080
- 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
- Bokszczanin KL, Solanaceae Pollen Thermotolerance Initial Training Network (SPOT-ITN) Consortium, Frangkostefanakis S (2013) Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Front Plant Sci* 4:315
- Bokvaj P, Hafidh S, Honys D (2015) Transcriptome profiling of male gametophyte development *Nicotiana tabacum*. *Genom Data* 3:106–111
- Borg M, Brownfield L, Khatab H, Sidorova A, Lingaya M, Twell D (2011) The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. *Plant Cell* 23:534–549
- Borg M et al (2014) An EAR-dependent regulatory module promotes male germ cell division and sperm fertility in Arabidopsis. *Plant Cell* 26:2098–2113
- 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
- Borges F, Pereira PA, Slotkin RK, Martienssen RA, Becker JD (2011) MicroRNA activity in the Arabidopsis male germline. *J Exp Bot* 62:1611–1620
- Brink RA, MacGillivray JH (1924) Segregation for the waxy character in maize pollen and differential development of the male gametophyte. *Am J Bot* 11:465–469
- 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 Genet* 5:e1000430
- Brownfield L, Hafidh S, Durbarray A, Khatab H, Sidorova A, Doerner P, Twell D (2009b) Arabidopsis DUO POLLEN3 is a key regulator of male germline development and embryogenesis. *Plant Cell* 21:1940–1956
- Calarco JP et al (2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* 151:194–205
- 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 YCS, McCormick S (1996) Sidecar pollen, an *Arabidopsis thaliana* male gametophytic mutant with aberrant cell divisions during pollen development. *Development* 122:3243–3253
- Chen Y, Zou M, Cao Y (2014) Transcriptome analysis of the Arabidopsis semi-in vivo pollen tube guidance system uncovers a distinct gene expression profile. *J Plant Biol* 57:93–105
- Chettoor AM et al (2014) Discovery of novel transcripts and gametophytic functions via RNA-seq analysis of maize gametophytic transcriptomes. *Genome Biol* 15:414
- Costa M, Nobre MS, Becker JD, Masiero S, Amorim MI, Pereira LG, Coimbra S (2013) Expression-based and co-localization detection of arabinogalactan protein 6 and arabinogalactan protein 11 interactors in Arabidopsis pollen and pollen tubes. *BMC Plant Biol* 13:7
- Cruz-Ramirez A et al (2004) The *xip1l* mutant of Arabidopsis reveals a critical role for phospholipid metabolism in root system development and epidermal cell integrity. *Plant Cell* 16:2020–2034
- Cui X, Lv Y, Chen M, Nikoloski Z, Twell D, Zhang D (2015) Young genes out of the male: an insight from evolutionary age analysis of the pollen transcriptome. *Mol Plant*. doi:10.1016/j.molp.2014.12.008
- Davidson RM et al (2011) Utility of RNA sequencing for analysis of maize reproductive transcriptomes. *Plant Genome* 4:191–203
- Doelling JH et al (2007) The ubiquitin-specific protease subfamily UBP3/UBP4 is essential for pollen development and transmission in Arabidopsis. *Plant Physiol* 145:801–813
- Domazet-Loso T, Tautz D (2010) A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 468:815–818
- Dukowicz-Schulze S, Chen C (2014) The meiotic transcriptome architecture of plants. *Front Plant Sci* 5:220
- Durbarray 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
- Eady C, Lindsey K, Twell D (1994) Differential activation and conserved vegetative cell-specific activity of a late pollen promoter in species with bicellular and tricellular pollen. *Plant J* 5:543–550

- Engel ML, Chaboud A, Dumas C, McCormick S (2003) Sperm cells of *Zea mays* have a complex complement of mRNAs. *Plant J* 34:697–707
- Engel ML, Davis RH, McCormick S (2005) Green sperm. Identification of male gamete promoters in Arabidopsis. *Plant Physiol* 138:2124–2133
- Fasoli M et al (2012) The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *Plant Cell* 24:3489–3505
- Feng X, Zilberman D, Dickinson H (2013) A conversation across generations: soma-germ cell crosstalk in plants. *Dev Cell* 24:215–225
- Friedman WE (1999) Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. *Development* 126:1065–1075
- Fu Z, Yang P (2014) Proteomics advances in the understanding of pollen-pistil interactions. *Proteomes* 2:468–484
- 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 Mol Biol* 70:581–601
- Giorno F, Wolters-Arts M, Mariani C, Rieu I (2013) Ensuring reproduction at high temperatures: the heat stress response during anther and pollen development. *Plants* 2:489–506
- Gou X, Yuan T, Wei X, Russell SD (2009) Gene expression in the dimorphic sperm cells of *Plumbago zeylanica*: transcript profiling, diversity, and relationship to cell type. *Plant J* 60:33–47
- Grant-Downton R, Hafidh S, Twell D, Dickinson HG (2009) Small RNA pathways are present and functional in the angiosperm male gametophyte. *Mol Plant* 2:500–512
- Grant-Downton R, Kourmpetli S, Hafidh S, Khatib H, Le Trionnaire G, Dickinson H, Twell D (2013) Artificial microRNAs reveal cell-specific differences in small RNA activity in pollen. *Curr Biol* 23:R599–R601
- Guan YF, Meng XZ, Khanna R, LaMontagne E, Liu YD, Zhang SQ (2014) Phosphorylation of a WRKY transcription factor by MAPKs is required for pollen development and function in Arabidopsis. *PLoS Genet* 10:e1004384
- Gusti A et al (2009) The *Arabidopsis thaliana* F-box protein FBL17 is essential for progression through the second mitosis during pollen development. *PLoS ONE* 4:e4780
- Haerizadeh F, Wong CE, Bhalla PL, Gresshoff PM, Singh MB (2009) Genomic expression profiling of mature soybean (*Glycine max*) pollen. *BMC Plant Biol* 9:25
- Hafidh S, Breznenova K, Ruzicka P, Fecikova J, Capkova V, Honys D (2012a) Comprehensive analysis of tobacco pollen transcriptome unveils common pathways in polar cell expansion and underlying heterochronic shift during spermatogenesis. *BMC Plant Biol* 12:24
- Hafidh S, Breznenova K, Honys D (2012b) De novo post-pollen mitosis II tobacco pollen tube transcriptome. *Plant Signal Behav* 7:918–921
- Hirano H, Sano Y (1991) Molecular characterization of the *waxy* locus of rice (*Oryza sativa*). *Plant Cell Physiol* 32:989–997
- Hirano K et al (2008) Comprehensive transcriptome analysis of phytohormone biosynthesis and signaling genes in microspore/pollen and tapetum of rice. *Plant Cell Physiol* 49:1429–1450
- Hobo T et al (2008) Various spatiotemporal expression profiles of anther-expressed genes in rice. *Plant Cell Physiol* 49:1417–1428
- Hollender CA et al (2014) Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks. *Plant Physiol* 165:1062–1075
- 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
- Honys D et al (2006) Identification of microspore-active promoters that allow targeted manipulation of gene expression at early stages of microgametogenesis in Arabidopsis. *BMC Plant Biol* 6:31
- 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* 149:621–631
- Johnson MA et al (2004) Arabidopsis *hapless* mutations define essential gametophytic functions. *Genetics* 168:971–982
- Kaessmann H (2010) Origins, evolution, and phenotypic impact of new genes. *Genome Res* 20:1313–1326
- Kawashima T, Berger F (2014) Epigenetic reprogramming in plant sexual reproduction. *Nat Rev Genet* 15:613–624
- Kim HJ et al (2008) Control of plant germline proliferation by SCF(FBL17) degradation of cell cycle inhibitors. *Nature* 455:1134–1137
- Kyo M, Hattori S, Yamaji N, Pechan P, Fukui H (2003) Cloning and characterization of cDNAs associated with the embryogenic dedifferentiation of tobacco immature pollen grains. *Plant Sci* 164:1057–1066
- Lalanne E et al (2004) *SETH1* and *SETH2*, two components of the glycosylphosphatidylinositol anchor biosynthetic pathway, are required for pollen germination and tube growth in Arabidopsis. *Plant Cell* 16:229–240
- Lang V, Usadel B, Obermeyer G (2015) De novo sequencing and analysis of the lily pollen transcriptome: an open access data source for an orphan plant species. *Plant Mol Biol* 87:69–80
- Le Trionnaire G, Twell D (2010) Small RNAs in angiosperm gametophytes: from epigenetics to gamete development. *Genes Dev* 24:1081–1085
- Lee JY, Lee DH (2003) Use of serial analysis of gene expression technology to reveal changes in gene expression in Arabidopsis pollen undergoing cold stress. *Plant Physiol* 132:517–529
- Leydon AR et al (2013) Three MYB transcription factors control pollen tube differentiation required for sperm release. *Curr Biol* 23:1209–1214
- Li XM, Sang YL, Zhao XY, Zhang XS (2013) High-throughput sequencing of small RNAs from pollen and silk and characterization of miRNAs as candidate factors involved in pollen-silk interactions in maize. *PLoS ONE* 8:e72852
- Liang Y et al (2013) MYB97, MYB101 and MYB120 function as male factors that control pollen tube-synergid interaction in *Arabidopsis thaliana* fertilization. *PLoS Genet* 9:e1003933
- Lin S-Y, Chen P-W, Chuang M-H, Juntawong P, Bailey-Serres J, Jauh G-Y (2014) Profiling of translatoemes of in vivo-grown pollen tubes reveals genes with roles in micropylar guidance during pollination in *Arabidopsis*. *Plant Cell* 26:602–618
- Liu Y et al (2013) Functional conservation of MIKC\*-Type MADS box genes in Arabidopsis and rice pollen maturation. *Plant Cell* 25:1288–1303
- Liu Y, Zhou J, White KP (2014) RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* 30:301–304
- Lorraine AE, McCormick S, Estrada A, Patel K, Qin P (2013) RNA-seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. *Plant Physiol* 162:1092–1109
- Lorraine AE, Blakley IC, Jagadeesan S, Harper J, Miller G, Firon N (2015) Analysis and visualization of RNA-seq expression data using RStudio, bioconductor and integrated genome browser, vol 1284. *Plant functional genomics: methods and protocols*. Springer, New York. doi:10.1007/978-1-4939-2444-8\_24
- Lu Y, Chanroj S, Zulkifli L, Johnson MA, Uozumi N, Cheung A, Sze H (2011) Pollen tubes lacking a pair of K<sup>+</sup> transporters fail to target ovules in Arabidopsis. *Plant Cell* 23:81–93
- Ma J, Skibbe DS, Fernandes J, Walbot V (2008) Male reproductive development: gene expression profiling of maize anther and pollen ontogeny. *Genome Biol* 9:R181

- Mascarenhas JP (1990) Gene activity during pollen development. *Ann Rev Plant Physiol Plant Mol Biol* 41:317–338
- Mascarenhas JP (1992) Pollen gene expression: molecular evidence. *Int Rev Cytol* 140:3–18
- Miyazaki S, Murata T, Sakurai-Ozato N, Kubo M, Demura T, Fukuda H, Hasebe M (2009) ANXUR1 and 2, sister genes to FERONIA/SIRENE, are male factors for coordinated fertilization. *Curr Biol* 19:1327–1331
- Mori T, Kuroiwa H, Higashiyama T, Kuroiwa T (2006) Generative cell specific 1 is essential for angiosperm fertilization. *Nat Cell Biol* 8:64–71
- Mori T, Igawa T, Tamiya G, S-y Miyagishima, Berger F (2014) Gamete attachment requires GEX2 for successful fertilization in *Arabidopsis*. *Curr Biol* 24:170–175
- Oh SA, Johnson A, Smertenko A, Rahman D, Park SK, Hussey PJ, Twell D (2005) A divergent cellular role for the FUSED kinase family in the plant-specific cytokinetic phragmoplast. *Curr Biol* 15:2107–2111
- Oh SA, Park KS, Twell D, Park SK (2010) The *SIDECAR POLLEN* gene encodes a microspore-specific LOB/AS2 domain protein required for the correct timing and orientation of asymmetric cell division. *Plant J* 64:839–850
- Ohr H, Bui AQ, Le BH, Fischer RL, Choi Y (2007) Identification of putative *Arabidopsis* DEMETER target genes by GeneChip analysis. *Biochem Biophys Res Commun* 364:856–860
- Okada T, Bhalla PL, Singh MB (2006) Expressed sequence tag analysis of *Lilium longiflorum* generative cells. *Plant Cell Physiol* 47:698–705
- Okada T, Singh MB, Bhalla PL (2007) Transcriptome profiling of *Lilium longiflorum* generative cells by cDNA microarray. *Plant Cell Rep* 26:1045–1052
- Oo MM et al (2014) Evaluation of rice promoters conferring pollen-specific expression in a heterologous system, *Arabidopsis*. *Plant Reprod* 27:47–58
- Ottaviano E, Mulcahy DL (1989) Genetics of angiosperm pollen. *Adv Genet* 26:1–64
- Palanivelu R, Preuss D (2006) Distinct short-range ovule signals attract or repel *Arabidopsis thaliana* pollen tubes in vitro. *BMC Plant Biol* 6:7
- Palanivelu R, Tsukamoto T (2012) Pathfinding in angiosperm reproduction: pollen tube guidance by pistils ensures successful double fertilization. *Wiley Interdiscip Rev Dev Biol* 1:96–113
- Park SK, Howden R, Twell D (1998) The *Arabidopsis thaliana* gametophytic mutation *gemin1 pollen1* disrupts microspore polarity, division asymmetry and pollen cell fate. *Development* 125:3789–3799
- Parnell FR (1921) Note on the detection of segregation by examination of the pollen of rice. *J Genet* 11:209–212
- Peng H et al (2012) MicroRNA profiles and their control of male gametophyte development in rice. *Plant Mol Biol* 80:85–102
- Phan HA, Iacuone S, Li SF, Parish RW (2011) The MYB80 transcription factor is required for pollen development and the regulation of tapetal programmed cell death in *Arabidopsis thaliana*. *Plant Cell* 23:2209–2224
- 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
- Procissi A, de Laissardiere S, Ferault M, Vezon D, Pelletier G, Bonhomme S (2001) Five gametophytic mutations affecting pollen development and pollen tube growth in *Arabidopsis thaliana*. *Genetics* 158:1773–1783
- Qin Y et al (2009) Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genet* 5:e1000621
- Redman JC, Haas BJ, Tanimoto G, Town CD (2004) Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *Plant J* 38:545–561
- Renak D, Dupl'akova N, Honys D (2012) Wide-scale screening of T-DNA lines for transcription factor genes affecting male gametophyte development in *Arabidopsis*. *Sex Plant Reprod* 25:39–60
- Richmond T, Somerville S (2000) Chasing the dream: plant EST microarrays. *Curr Opin Plant Biol* 3:108–116
- Rotman N et al (2005) A novel class of MYB factors controls sperm-cell formation in plants. *Curr Biol* 15:244–248
- Russell SD et al (2012) Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. *New Phytol* 195:560–573
- Sanetomo R, Hosaka K (2013) Pollen transcriptome analysis of *Solanum tuberosum* (2n=4x=48), *S. demissum* (2n=6x=72), and their reciprocal F1 hybrids. *Plant Cell Rep* 32:623–636
- Saze H (2008) Epigenetic memory transmission through mitosis and meiosis in plants. *Semin Cell Dev Biol* 19:527–536
- Saze H, Scheid OM, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet* 34:65–69
- Schmid M et al (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37:501–506
- Schmidt A, Schmid MW, Grossniklaus U (2012) Analysis of plant germline development by high-throughput RNA profiling: technical advances and new insights. *Plant J* 70:18–29
- Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci USA* 108:4069–4074
- Schrauven JAM, Degroot PFM, Vanherpen MMA, Vanderlee T, Reynen WH, Weterings KAP, Wullems GJ (1990) Stage-related expression of messenger-RNAs during pollen development in lily and tobacco. *Planta* 182:298–304
- Scott RJ, Spielman M, Dickinson HG (2004) Stamen structure and function. *Plant Cell* 16:S46–S60
- Shen Y et al (2011) Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing. *Genome Res* 21:1478–1486
- Shulaev V et al (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet* 43:109–116
- Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136:461–472
- Stinson JR, Eisenberg AJ, Willing RP, Pe ME, Hanson DD, Mascarenhas JP (1987) Genes expressed in the male gametophyte of flowering plants and their isolation. *Plant Physiol* 83:442–447
- Suwabe K et al (2008) Separated transcriptomes of male gametophyte and tapetum in rice: validity of a laser microdissection (LM) microarray. *Plant Cell Physiol* 49:1407–1416
- Sze H et al (2004) Expression patterns of a novel AtCHX gene family highlight potential roles in osmotic adjustment and K<sup>+</sup> homeostasis in pollen development. *Plant Physiol* 136:2532–2547
- Taylor LP, Hepler PK (1997) Pollen germination and tube growth. *Ann Rev Plant Physiol Plant Mol Biol* 48:461–491
- Turcich MP, Mascarenhas JP (1994) PREM-1, a putative maize retroelement has LTR (long terminal repeat) sequences that are preferentially transcribed in pollen. *Sex Plant Reprod* 7:2–11
- Twell D (1992) Use of a nuclear-targeted  $\beta$ -glucuronidase fusion protein to demonstrate vegetative cell-specific gene expression in developing pollen. *Plant J* 2:887–892
- Twell D (1994) The diversity and regulation of gene expression in the pathway of male gametophyte development. *Molecular and*

- cellular aspects of plant reproduction, vol 55. Cambridge University Press, Cambridge
- Twell D (2002) Pollen developmental biology. In: O'Neill SD, Roberts JA (eds) Plant reproduction, vol 6. Blackwell Publishing, Sheffield, pp 86–153
- Twell D (2011) Male gametogenesis and germline specification in flowering plants. *Sex Plant Reprod* 24:149–160
- Twell D, Wing R, Yamaguchi J, McCormick S (1989) Isolation and expression of an anther-specific gene from tomato. *Mol Gen Genet* 217:240–245
- Twell D, Park SK, Hawkins TJ, Schubert D, Schmidt R, Smertenko A, Hussey PJ (2002) MOR1/GEM1 plays an essential role in the plant-specific cytokinetic phragmoplast. *Nat Cell Biol* 4:711–714
- Twell D, Oh S-A, Honys D (2006) Pollen development, a genetic and transcriptomic view. In: Malhó R (ed) The pollen tube, vol 3. Plant cell monographs. Springer, Berlin, pp 15–45. doi:10.1007/7089\_042
- Verelst W, Saedler H, Muenster T (2007a) MIKC\* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters. *Plant Physiol* 143:447–460
- Verelst W, Twell D, de Folter S, Immink R, Saedler H, Muenster T (2007b) MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biol* 8:R249
- von Besser K, Frank AC, Johnson MA, Preuss D (2006) Arabidopsis *HAP2 (GCSI)* is a sperm-specific gene required for pollen tube guidance and fertilization. *Development* 133:4761–4769
- 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
- Wang J, Qiu X, Li Y, Deng Y, Shi T (2011) A transcriptional dynamic network during *Arabidopsis thaliana* pollen development. *BMC Syst Biol* 5:S8
- Wei LQ, Xu WY, Deng ZY, Su Z, Xue Y, Wang T (2010) Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa*. *BMC Plant Biol* 11:338
- Willing RP, Mascarenhas JP (1984) Analysis of the complexity and diversity of mRNAs from pollen and shoots of *Tradescantia*. *Plant Physiol* 75:865–868
- Willing RP, Bashe D, Mascarenhas JP (1988) An analysis of the quantity and diversity of messenger RNAs from pollen and shoots of *Zea mays*. *Theor Appl Genet* 75:751–753
- Won S-K, Lee Y-J, Lee H-Y, Heo Y-K, Cho M, Cho H-T (2009) Cis-element- and transcriptome-based screening of root hair-specific genes and their functional characterization in Arabidopsis. *Plant Physiol* 150:1459–1473
- Wu D-D, Wang X, Li Y, Zeng L, Irwin DM, Zhang Y-P (2014) “Out of pollen” hypothesis for origin of new genes in flowering plants: study from *Arabidopsis thaliana*. *Genome Biol Evol* 6:2822–2829
- Xia C et al (2014) The ARID-HMG DNA-binding protein AtHMGB15 is required for pollen tube growth in *Arabidopsis thaliana*. *Plant J* 79:741–756
- Xin H-P, Peng X-B, Ning J, Yan T-T, Ma L-G, Sun M-X (2011) Expressed sequence-tag analysis of tobacco sperm cells reveals a unique transcriptional profile and selective persistence of paternal transcripts after fertilization. *Sex Plant Reprod* 24:37–46
- Xu XH, Chen H, Sang YL, Wang F, Ma JP, Gao X-Q, Zhang XS (2012) Identification of genes specifically or preferentially expressed in maize silk reveals similarity and diversity in transcript abundance of different dry stigmas. *BMC Plant Biol* 13:294
- Xu XH, Wang F, Chen H, Sun W, Zhang XS (2013) Transcript profile analyses of maize silks reveal effective activation of genes involved in microtubule-based movement, ubiquitin-dependent protein degradation, and transport in the pollination process. *PLoS ONE* 8:e53545
- Zanetti ME, Chang IF, Gong FC, Galbraith DW, Bailey-Serres J (2005) Immunopurification of polyribosomal complexes of Arabidopsis for global analysis of gene expression. *Plant Physiol* 138:624–635
- Zhang L et al (2009) A genome-wide characterization of microRNA genes in maize. *PLoS Genet* 5:e1000716
- Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE* 9:e78644
- Zou C, Jiang W, Yu D (2010) Male gametophyte-specific WRKY34 transcription factor mediates cold sensitivity of mature pollen in *Arabidopsis*. *J Exp Bot* 61:3901–3914