Microarray analysis of single isolated cells of the female gametophyte reveals potential regulators of female germline development in *Arabidopsis thaliana*

DISSERTATION

SUBMITTED TO THE FACULTY OF NATURAL SCIENCES,
BIOLOGY AND PRECLINICAL MEDICINE,
UNIVERSITY OF REGENSBURG
FOR THE TITLE OF
DOCTOR OF NATURAL SCIENCES
(DR. RER. NAT)

by Lucija Šoljić

from Dubrovnik, Croatia

in April 2012

Das Promotionsgesuch wurde eingereicht am: 17.04.2012

Die Arbeit wurde angeleitet von: Dr. S. Sprunck

Prüfungsausschuss:

Prof. Dr. R. Wirth (Vorsitzender)

Prof. Dr. T. Dresselhaus

Prof. Dr. S. Schneuwly

Prof. Dr. F. Sprenger

TABLE OF CONTENTS

3 RESULTS

1 INTRODUCTION

1.1 The life cycle of higher plants

The plant life cycle alternates between two generations, a diploid sporophyte and a haploid gametophyte. Both male and female gametophytes develop from specialized cells within the sporophytic tissue of flowers by a series of meiotic and mitotic cell divisions, producing male and female gametes along with a number of accessory cells comprising the male gametophyte, called the pollen grain and the female gametophyte, called the embryo sac. A transition into the sporophytic generation is achieved by the process of double fertilization, typical for flowering plants (angiosperms), during which two male and two female gametes fuse. This event on one side produces an embryo, which will eventually develop into a new plant organism, and on the other side an endosperm, a tissue designated to nurture and support the developing embryo. A scheme of the transition is pictured in Figure 1.

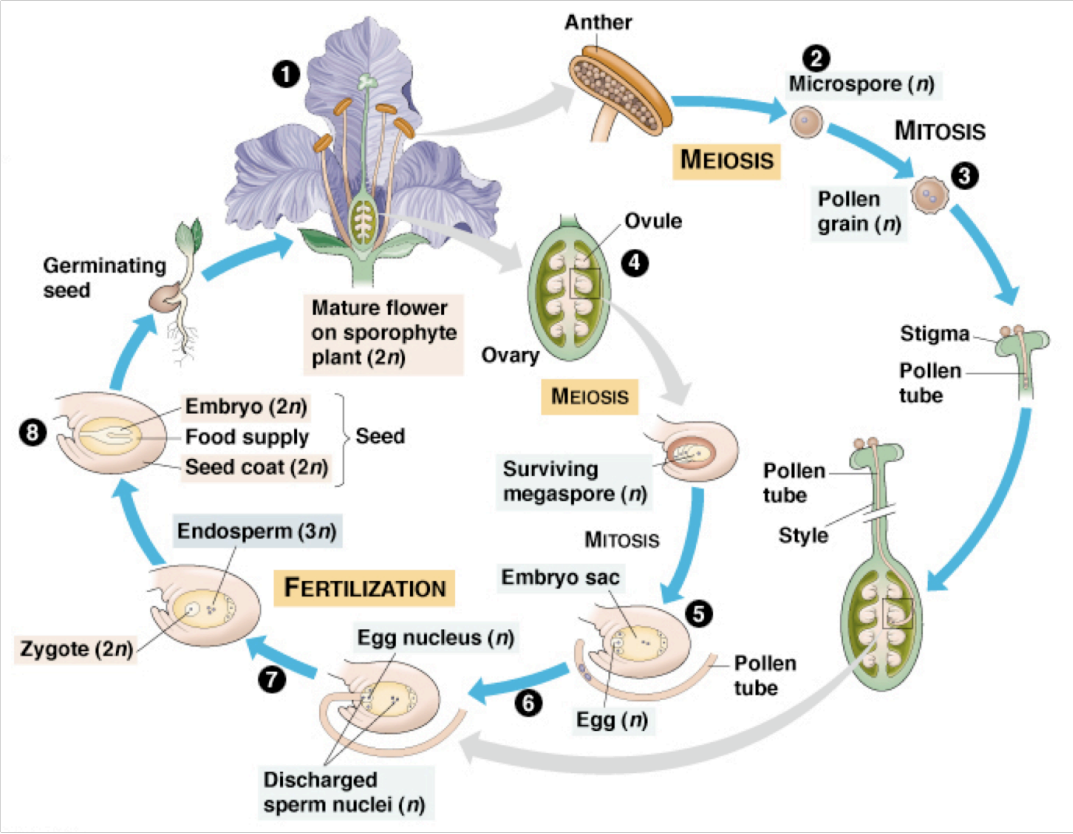


Figure 1. A scheme of the angiosperm life cycle. The picture outlines major steps in the development of both male and female gametophytes where haploid gametes are developed within the diploid sporophyte through meiosis, followed by mitosis and maturation. After sperm cell delivery to the female gametophyte by the pollen tube, they unite through the process of double fertilization, producing the diploid embryo and the

triploid endosperm, marking the step of transition back into the sporophytic stage, as the embryo develops into a new plant organism (Image from http://www.mun.ca/biology/scarr/139450_Angiospermae.jpg).

Unlike some of the earliest terrestrial plants and mosses living today, which have a dominant gametophytic generation, higher plants including angiosperms have developed a dominant sporophyte. While it reaches impressive dimensions in many species, which measure their life cycle often in hundreds of years, both male and female gametophytes have been reduced to a tissue consisting of a group of only a few cells. In angiosperms, these delicate structures are well protected inside the reproductive structures of flowers, which evolved in a multitude of extremely diverse varieties. *Arabidopsis thaliana* belongs to the *Brassicaceae* plant family and has become a major model plant, partly due to its basic flower architecture. Its flower structures are described in Figure 2.

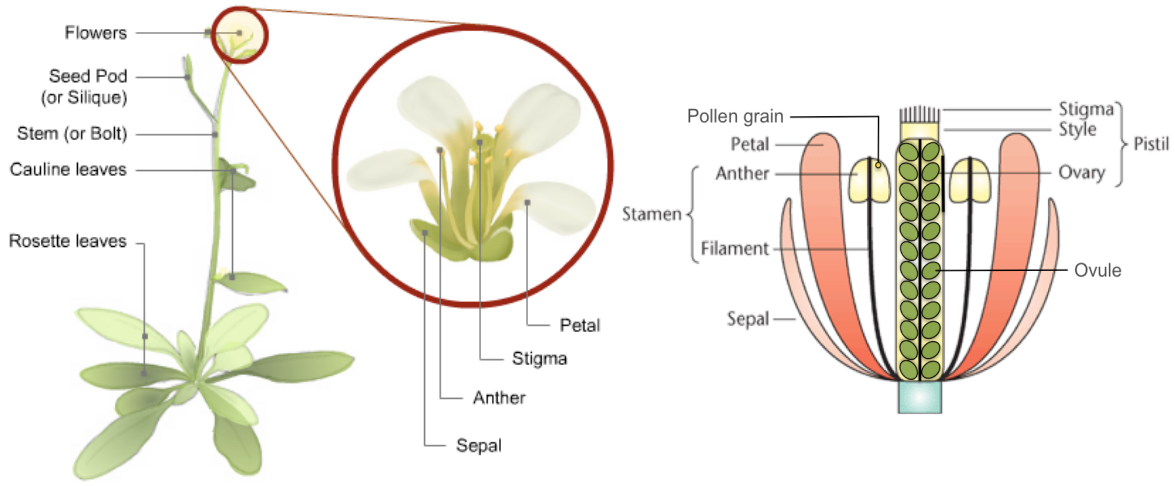


Figure 2. *Arabidopsis thaliana* basic architecture and flower anatomy. The drawing on the left side depicts a young *Arabidopsis* plant and denotes its major organs with the flower magnified in the red circle. A scheme of the flower, which holds the reproductive organs, is depicted on the right side. The male gametophyte is reduced to a single pollen grain (for illustrative purposes only one is shown within an anther in the image, but thousands develop in each anther), and the female gametophyte is reduced to an embryo sac, a group of cells embedded in the maternal sporophytic tissue of the ovule, within the pistil. Images taken and adapted from http://www.prep.biochem.vt.edu/expinfo/expinfo_anatomy.html and Ma (2009).

1.2 Female gametophyte development and cell specification

Unlike animals, which produce their gametes directly by meiosis, plants have developed a multicellular haploid structure called the gametophyte, which in addition to

male or female gametes contains several accessory cells. The female gametophyte of *Arabidopsis thaliana* develops within reproductive structures of flowers, from a specialized megasporocyte cell, also referred to as megaspore mother cell (MMC), which has differentiated from an archesporial cell within the sporophytic tissue of the ovule (Robinson-Beers *et al*., 1992; Gasser *et al.*, 1998). The exact way the transition from somatic into germline cell fate occurs remains unknown, but several genes involved in the process have been identified. One such example in *Arabidopsis* is the *SPOROCYTELESS* gene (*SPL/NOZZLE*). Its mutants are arrested after archesporial cell differentiation, when they stop further development, resulting in a complete lack of a germline on both the male and the female side (Schiefthaler *et al*., 1999; Yang *et al*., 1999). Mutations in *MAC1 (MULTIPLE ARCHESPORIAL CELL 1)* in maize and *MSP1 (MULTIPLE SPOROCYTE 1)* in rice on the other hand, produce and excessive number of archesporial cells and sporocytes, and are likely to be involved in determining germline cell fate by preventing the surrounding cells from becoming germline cells (Sheridan *et al.* 1996; Nonomura *et al*., 2003; Yang *et al*., 2010). In animals, germline fate determination is controlled by a specific PIWI-associated miRNA (piRNA) system involving a class of germline-specific ARGONAUTE (AGO) proteins (Girard *et al*., 2006; Grivna *et al*., 2006) and equivalent pathways are starting to slowly emerge in the plant field.

Once developed, a megaspore mother cell divides through meiosis, producing four megaspores (Figure 3), marking the transition to the haploid stage in the life cycle, which will be maintained up until fertilization. In the evolutionarily more advanced flowering plant species only one of the four megaspores is functional and its choice is both position- and species-dependent. In *Arabidopsis* three of the megaspores located towards the micropylar pole undergo programmed cell death leaving one surviving megaspore at the chalazal pole (Sundaresan and Alandete-Saez, 2010). This functional megaspore (FM) undergoes three nuclear mitotic divisions without cytokinesis, producing an eight-nucleate syncytium with three nuclei migrating to either side of the embryo sac in formation, the chalazal and micropylar pole, and two polar nuclei remaining in the middle, close to the large central vacuole. Events during both meiotic and mitotic divisions require precise control and are tightly regulated at designated checkpoints throughout the different stages of the cell cycle. The large syncytium ultimately cellularizes into seven cells of four different cell types (Ma and Sundaresan, 2010; Endress, 2011) as shown in Figure 3.

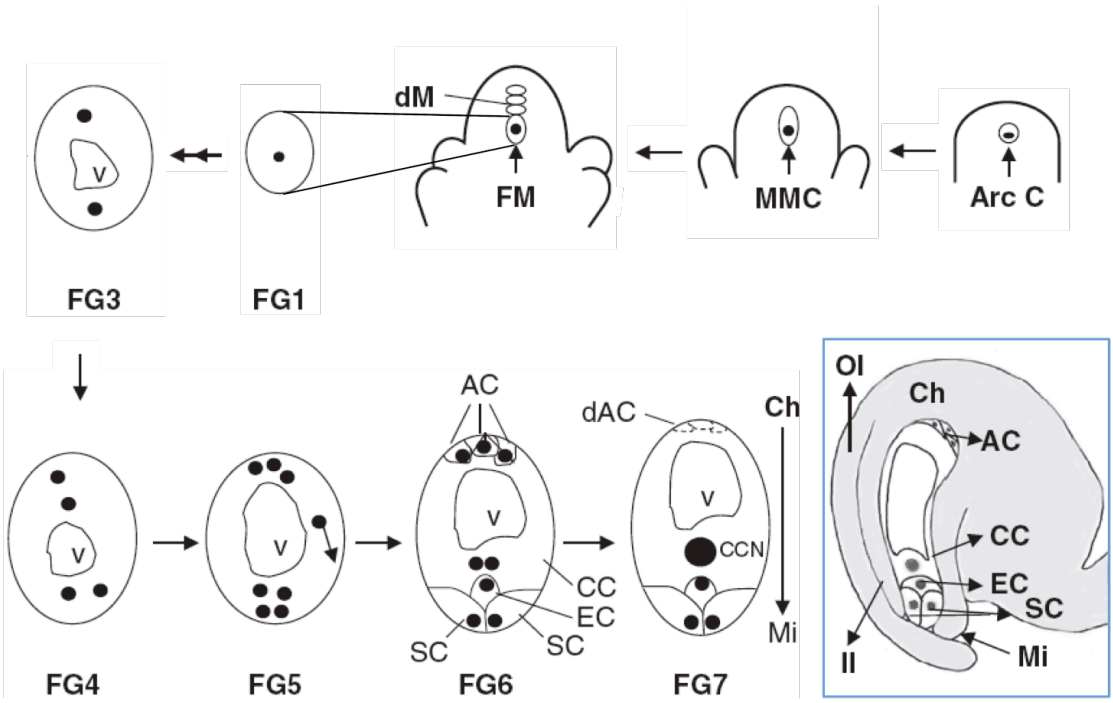


Figure 3. A schematic representation of the female gametophyte development in *Arabidopsis thaliana*. After a specialized archesporial cell (Arc C) within the ovule primordium differentiates into a megaspore mother cell (MMC), it undergoes meiosis producing four megaspores. Three of those degenerate (dM), leaving one functional megaspore (FM) to further divide mitotically (FG1 – FG7), eventually giving rise to an eight nucleate syncitium, which will cellularize to form seven cells developing into the mature embryo sac (FG7 scheme and last frame in blue). Arc C = Archesporial cell; MMC = megaspore mother cell; FM = functional megaspore; dM = degenerated megaspores; FG1 – FG7 = developmental stages of the female gametophyte; v = vacuole; SC = synergid cell; EC = egg cell; CC = central cell; AC = antipodal cell; dAC = degenerated antipodal cell; CCN = central cell nucleus; Ch = chalazal pole of the embryo sac; Mi = micropyle / micropylar pole of the embryo sac; OI = outer integument; II = inner integument; Image adapted from Ma and Sundaresan (2010).

Cellularization is a critical step in gametophytic cell fate specification and is spatially and temporally controlled. Although more is known about the process in the male gametophye, new findings concerning embryo sac cellularization continually emerge. Genes that have been implicated in this process include *GEM2*, which plays a fundamental role in coordinating karyokinesis and cytokinesis, its mutants displaying, among other, partial cellularization and occasional fusion of free nuclei in the embryo sac, as well as a variety of division defects in pollen development (Park *et al*., 2004). Mutants *gem1/mor1* and *tio*, coding for a microtubule-associated protein and an essential phragmoplast encoded protein, respectively, both have an effect on cell plate formation (Twell *et al*., 2002; Oh *et*

al., 2005). A comparable phenotype resulting in uncellularized embryo sacs with aberrant morphology, number and positioning of nuclei can be seen when *TUBG1* and two genes encoding γ-tubulin are knocked out, as well as when kinesin-like proteins *AtNACK1* and *AtNACK2* involved in the kinesin-MAPKKK pathway are affected (Tanaka *et al*., 2004; Pastuglia *et al*., 2006; Yang *et al.*, 2010).

After successful cellularization the chalazal pole of the formed embryo sac harbours three antipodal cells, whose function remains elusive. The micropylar pole holds the egg apparatus consisting of the egg cell, which represents one of the two female gametes, and two accessory cells called the synergid cells. The second female gamete, the central cell, occupies the remaining space between the two poles, and in the final stage of development its di-haploid nucleus is formed by the fusion of two nuclei, each from one pole. Athough the cellular and molecular basis of cell specification within the embryo sac is largely unknown, emerging evidence suggests that positional and lateral inhibition mechanisms are involved in determining gametophytic cell fate (Yang *et al*., 2010). Also, Pagnussat *et al*. (2009), have recently demonstrated auxin involvement in gametic cell specification, with the establishment of an asymmetric auxin gradient (Figure 4) in the course of embryo sac formation, shown through use of the synthetic *DR5:reporter* fusions. Additionally, the disruption of auxin responses by *ARF* (*Auxin Response Factor*) downregulation or biosynthesis by ectopic expression of *YUCCA1*, which is involved in an auxin biosynthesis pathway, was also found to affect gametophytic cell identities at the micropylar pole (Li *et al*., 2008; Yang *et al*., 2010). Several other mutants with disrupted cell identity, and cell-misspecification phenotypes have been identified and include *agl80, lachesis (lis), eostre, clotho (clo/gfa1), atropos (ato), diana (dia)* in *Arabidopsis* (Portereiko *et al*., 2006; Gross-Hardt *et al.,* 2007; Pagnussat *et al*., 2007; Moll *et al*., 2008; Bemer *et al*., 2008; for review see Sprunck and Gross-Hardt, 2011). In order for the female gametophyte to perform its primary functions, tight control of all necessary mechanisms must be in place throughout its development.

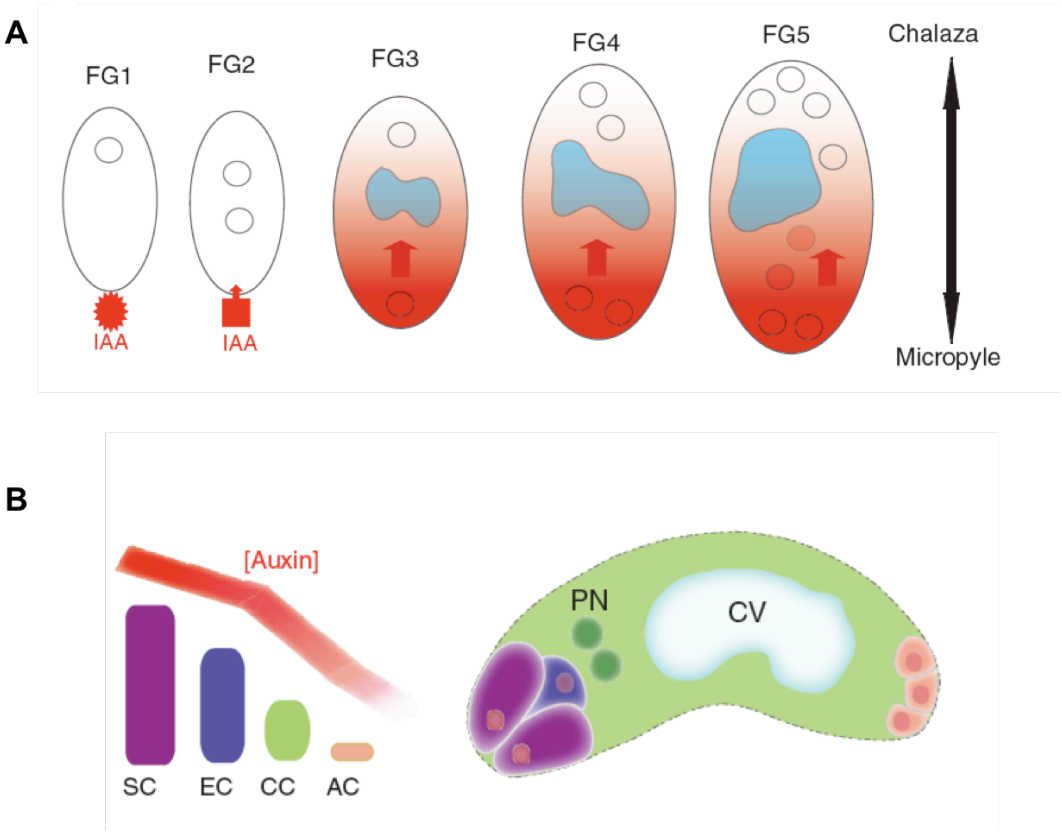


Figure 4. The role of auxin in embryo sac development and cell specification.

(**A**) A schematic depiction of the embryo sac from stages FG1 to FG5. In FG1 and FG2 the auxin source (IAA) is located outside the forming embryo sac, in the sporophytic tissue of the ovule. At stages FG3 to FG5 the micropylar end of the female gametophyte becomes the auxin source and a gradient is formed (marked in red), influencing cell specification dependent on nuclei positioning.

(**B**) A model for cell-fate specification following cellularization is shown on the left hand side, depicting the cell types that comprise the embryo sac marked in different colours in relation to the auxin gradient (thick red line), and a drawing of the female gametophyte upon cellularization at FG6, in corresponding colours on the right hand side. SC = synergid cell; EC = egg cell; CC = central cell; AC = antipodal cell; PN = polar nuclei; CV = central vacuole. (Image taken and adapted from Sundaresan and Alandete-Saez, 2010).

Therefore, in a properly developed *Arabidopsis* embryo sac, the female germ unit remaining after three antipodals in the chalazal pole of the embryo sac degenerate, completing its development into a mature female gametophyte (Figure 3), is comprised of the central cell, egg cell and two synergid cells. The latter are left to perform the tasks of pollen tube attraction and interaction with the tube in order to trigger sperm release and to successfully complete double fertilization (Yang *et al*., 2010).

1.3 Male gametophyte development

During the course of flower development the primordial cells of the microsporangia within the anthers, which are a part of the male reproductive organ - the stamen, divide and differentiate, forming several cell types (Ma and Sundaresan, 2010). The initiated gametophytic stage is confined to a few divisions, defining the male germline that produces gametes. The process can be divided into two sequential stages, microsporogenesis – during which a diploid pollen mother cell undergoes meiosis, producing four haploid microspores, and microgametogenesis, during which the microspores undergo two mitotic divisions, eventually forming three cells comprising the male gametophyte, as shown in Figure 5.

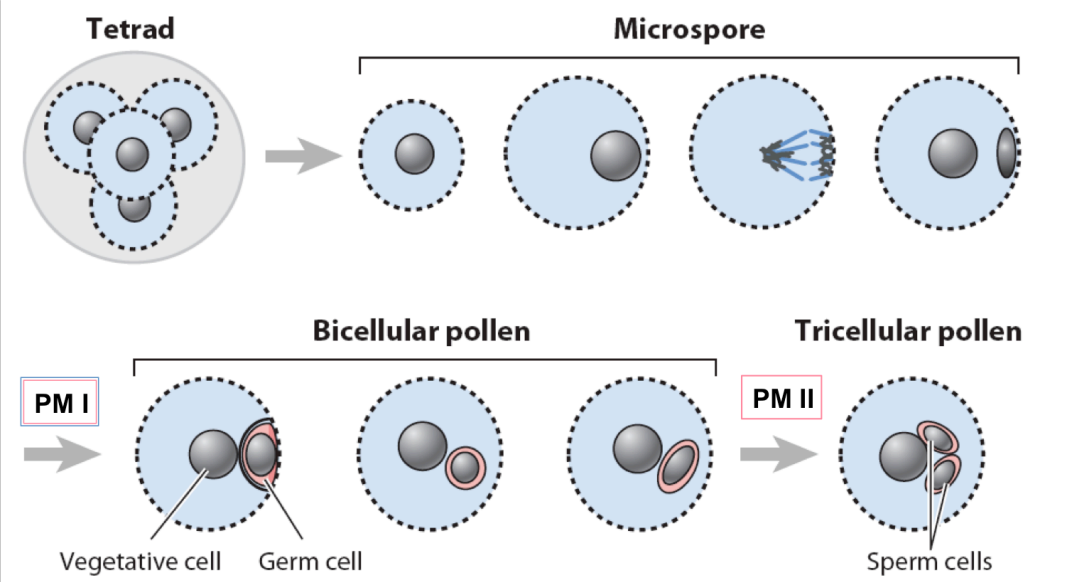


Figure 5. Male gametophyte development. A diploid pollen mother cell (PMC; not shown) undergoes meiosis to produce a tetrad of haploid microspores. Nuclear migration, asymmetric spindle formation and a highly asymmetric cell division termed pollen mitosis I (PMI) then take place, producing bicellular pollen, which harbours a small germ cell that will eventually be engulfed by the larger vegetative cell. Following elongation, the germ cell will divide again through pollen mitosis II (PMII) to produce twin sperm cells, whilst the vegetative cell differentiates, acquiring the capacity to develop a pollen tube upon pollination. (Image taken and adapted from Berger and Twell, 2011).

The first, distinctly asymmetric cell division of the microspore separates a vegetative from a generative cell, the latter then dividing again to produce the two sperm cells that will be required for double fertilization (Berger and Twell, 2011).

Even though this basic scheme of events is unchanged in most angiosperms, the time points of the second mitotic division in relation to pollination can vary, with many species producing bicellular pollen at anthesis, with the second division occurring inside the growing pollen tube, after its germination on the surface of the stigma. Some species however, including *Arabidopsis*, produce tricellular pollen, where the pollen grain already contains two sperm cells in addition to the vegetative cell at pollination. A number of biological processes play extremely important roles during male gametophyte development. Tight control of cell cycle progression is crucial to ensure twin sperm cell production and the cell cycle synchrony of male and female gametes prior to karyogamy (Friedman, 1999; Berger and Twell, 2011). Epigenetic processes as well as smallRNA pathways have also been proven to be very important, comparable to the development of animal germlines, as components in a number of important processes, and their biogenesis, regulation and metabolism have a significant impact on genome integrity and germ cell functions (Grant-Downton *et al*., 2009; reviewed by Banisch *et al*., 2012). Families of small RNAs including microRNAs, *trans*-acting siRNAs, natural antisense siRNAs and siRNAs have been shown to be involved in RNA-dependent DNA methylation (Benetti *et al.*, 2008; Daxinger *et al*., 2009; Grant-Downton, 2010; reviewed by Van Ex *et al*., 2011). Similar to animals, plants also use small RNA systems to control transposable element activity in the germline (Slotkin *et al*., 2009; Olmedo-Monfil *et al*., 2010), and in addition, recent evidence has shown that a specific regulatory module of a natural antisense gene pair, that spawns nat-siRNAs in the sperm cells, has a key role in fertilization (Dickinson and Grant-Downton, 2011).

1.4 Double fertilization

Fertilization marks the process in which female and male gametes fuse, producing a zygote that develops into an embryo and eventually a new organism, genetically different from its parents. Angiosperms have developed a unique system of double fertilization, where both male and female gametophytes produce two gametes each, which will subsequently fuse with each other in the process. On the female side two unequal gametes are produced, a haploid egg cell, which upon fertilization gives rise to a diploid zygote eventually developing into an embryo, and the diploid central cell, giving rise to the triploid

nurturing tissue called the endosperm (Berger *et al*. 2006). On the male side two haploid sperm cells are produced which, in the case of *Arabidopsis*, seem to be equally capable of fusing with either of the female gametes (Berger, 2011; Hamamura *et al*., 2011).

Even though the process was discovered at the end of the 19th century (Nawaschin, 1898; Guignard, 1899) and the first cytological observations made in 1960-ies (Jensen, 1965; Jensen, 1968), the details of the cellular and molecular mechanisms and pathways underlying it are still in the process of being resolved, largely due to the size and placement of the cells involved, making them extremely difficult to access. Unlike animals and lower plants including mosses and ferns, whose sperm cells are able to actively move towards their target, angiosperms have immobile sperm. They have therefore developed a special system, where following the transmission of a pollen grain to the surface of the stigma it is hydrated and germinates into a rapidly growing cell termed the pollen tube. The pollen tube then carries the two sperm cells through the carpel tissue to the female gametophyte, where fertilization takes place (Figure 6).

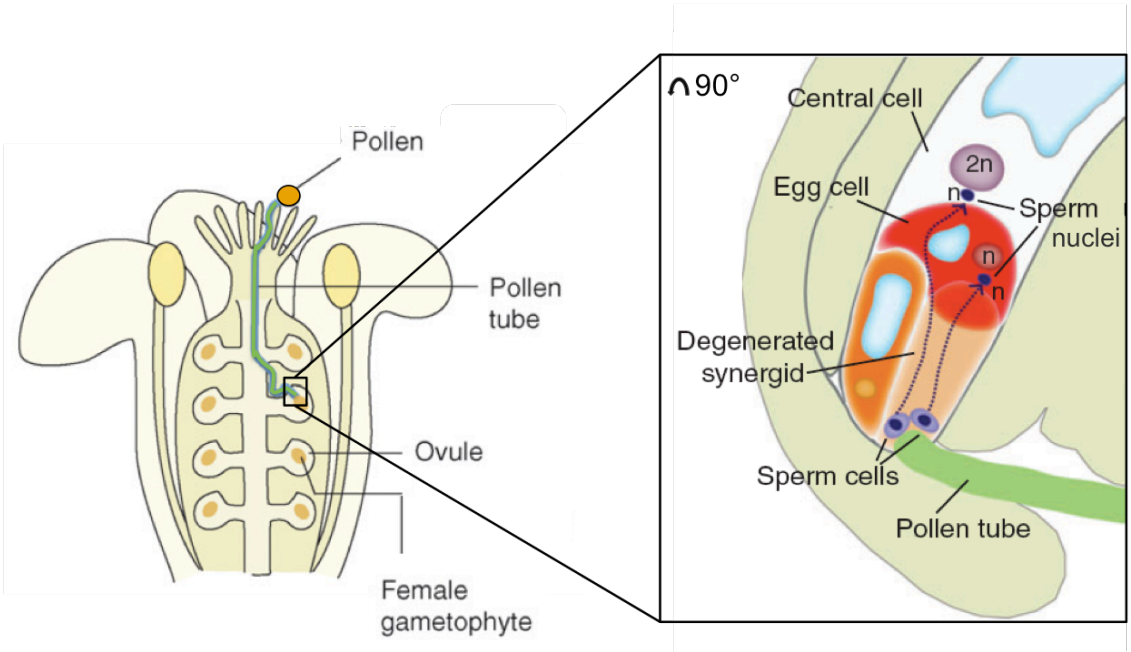


Figure 6. Sperm cell delivery and double fertilization. The image on the left depicts a scheme of a pollinated flower, where a pollen tube has just reached the ovule. The image on the right shows the magnified micropylar area of the ovule at the point of fertilization. One of the synergid cells degenerates as the pollen tube arrives and subsequently bursts, releasing the sperm cells which are then propelled towards the female gametes, where the fusion of their nuclei (karyogamy) will take place. The central cell nucleus is marked in purple, the egg cell nucleus in dark red and the sperm cell nuclei in violet. Their route to the female gamete nuclei is marked with a dotted violet line. 2n = diploid; n = haploid. (Image taken and adapted from Berger *et al*., 2008 and Sundaresan and Alandete-Saez, 2010).

In order to achieve this goal, several requirements need to be met, and a number of events have to have taken place. The pollen tube does not grow through the carpel tissue in a random, non-directional manner, it is rather accurately and precisely guided by specialized attractants. These act on both long (through sporophytic maternal tissues) and short (in the direct proximity of the micropyle and the female gametophyte, within 100 – 200 μm) distances on the way towards individual ovules and the egg apparatus. There has been substantial effort put into identifying and characterizing these attractants, and the progress made in recent years has unveiled the importance of several classes of signaling molecules involved. The two synergids have emerged quite early on as front runners in the investigation with a well defined role in attractant production in the species that have been examined, starting with *Torenia fournieri* (Higashiyama *et al*., 2001), and even when only one synergid is produced, as is the case in the *Arabidopsis* mutant *eostre* (Pagnussat *et al*., 2007), the pollen tube attraction takes place and the sperm cells are properly released (Berger *et al*., 2008). The synergids have been shown to secrete pollen tube attractants through the filiform apparatus, which is an extension of their cell wall placed directly at the micropyle, and is composed of a number of finger-like invaginations, increasing its surface area. The transcription factor MYB98, expressed selectively in the synergids of *Arabidopsis thaliana*, is required for formation of the filiform apparatus and is also involved in micropylar pollen tube guidance (Berger *et al*., 2008). Recent years have shown that small polymorphic proteins have taken on important roles in pollen tube attraction. A mycropylar pollen tube attractant secreted by the female gametophyte was the *ZmEA1* (*Zea mays EGG APPARATUS 1*), with impaired pollen tube guidance in maize upon knock-down (Marton *et al*., 2005; Marton and Dresselhaus, 2010). Other genes expressed in maize include *ZmES1-4*, a group specifically expressed in the mature embryo sac, but instantly down-regulated upon fertilization, with members shown to induce pollen tube burst *in vitro* within less than one minute of exposure (Amien *et al*., 2011). A group of cysteine-rich proteins (CRPs) play important roles in cell-cell communication and cell signaling during fertilization. Despite sharing their fundamental characteristics, they are extremely divergent, which is a factor contributing to species-specific signaling roles. A novel class of defensin-like proteins (DEFLs), a subgroup within CRPs, has been first identified in *Torenia fournieri*. They are synthesized specifically and secreted to the filiform aparatus by the synergid cells, and they reportedly act as pollen tube attractants, able to act at

nanomolar concentrations *in vitro* in a species-specific manner, and have been termed LUREs, with two members characterized in *T. fournieri* so far (Okuda *et al*., 2009). The role of the female gametes in pollen tube attraction seems to be minimal, with the egg cell implicated in the generation of the attractant in maize, and mutants in the central cell expressed *CCG* (*CENTRAL CELL GUIDANCE*) gene in *Arabidopsis* shown to be defective in pollen tube attraction, probably not coding for a direct attractant, but rather encoding a potential transcriptional regulator (Chen *et al*., 2007). Also, the gene might be acting upstream and could produce a signal either further processed by the synergid or cue the synergid to produce an attractant (Berger *et al*., 2008; Dresselhaus and Marton 2009).

Upon reaching the embryo sac, the growing pollen tube bursts, discharging its contents through interaction with the synergid cells and releasing the two sperm cells, with one of the synergid cells degenerating at the same time. Cessation of pollen tube growth and sperm cell release also seems to be controlled by the female gametophyte and require the receptor-like kinase FERONIA (Escobar-Restrepo *et al*., 2007), which accumulates on the plasma membrane in the filiform appartus of synergid cells. Mutants in the gene result in pollen tubes continuing to grow and coil around in the micropylar area after reaching the FG, failing to burst and release their content (Dresselhaus, 2006; Berger *et al*., 2008), a phenotype shared with other mutants, including *sirene* (coding for the same RLK), *scylla* (Rotman *et al*., 2008) and *lorelei* (encoding a small plant-specific CRP) (Capron *et al*., 2008). Paralogue proteins of *FERONIA* have been identified in pollen tubes as *ANXUR1* and *ANXUR2*, and their mutants also exhibit a similar phenotype, indicating they could participate in the same pathway and highlighting the fact that male and female gametes probably exchange signals regulating pollen tube arrest and sperm cell release (Miyazaki *et al*., 2009; Berger, 2011). The triggering of pollen tube discharge also requires *ACA9* (Schiott *et al*., 2004), a calcium pump on the plasma membrane of the pollen tube, suggesting a contribution of calcium signaling to this process and mutant *aca9* pollen tubes are able to reach the micropyle and embryo sac, cease growth, but fail to burst and release their contents (Berger *et al*., 2008; Dresselhaus and Marton 2009).

Data collected so far about gametophyte development, cell specification and fertilization, largely through the use of standard molecular and cytological methods and the combination of *in vivo* imaging and genetic studies, contributed to the identification and characterization of some of the genes controlling these processes. Technological advances, particularly in the field of confocal laser scanning microscopy (CLSM) and through the

application of a wide variety of fluorescent proteins available for labeling, enable researchers to visually document these events (Berger *et al*., 2008; Hamamura *et al.*, 2011).

The topics mentioned thus far, having a major impact on embryo sac development and its function, including cell cycle, silencing, auxin dynamics and signaling, will be examined in the female gametophyte through a microarray study performed in this work. Particular focus is placed on gene regulation via RNA-binding proteins (RBPs), with additional studies performed on the Pumilio family of RBPs.

As mentioned, Argonaute proteins play a major role in gene regulation by interacting with small non-coding RNAs. In *Arabidopsis* the ARGONAUTE 9 (AGO9) protein, considered to be a PIWI ortholog, was recently implicated in the control of female gamete formation by restricting the specification of gametophyte precursors. Mutations in the gene lead to the differentiation of multiple gametic cells that were able to initiate gametogenesis. Even though only a correctly positioned MMC forms an embryo sac, ectopic MMCs are able to differentiate without undergoing meiosis into a diploid megaspore that arrests at one-nucleate stage (Olmedo-Monfil *et al*., 2010). An independent smallRNA pathway, involving *AGO5* and acting in the somatic nucellar cells surrounding the megaspores, promotes the initiation of megagametogenesis in the functional megaspore (Tucker *et al*., 2012). Another example recently found in maize shows the *ARGONAUTE104* (*AGO104*) gene, likely a homolog of the *Arabidopsis AGO9*, expressed in the somatic tissue surrounding the megaspore mother cell. It seems to influence MMC fate through an AGO104-dependent mobile signal by either promoting meiosis or repressing somatic cell fate in the MMC (Singh *et al*., 2011). In rice, a germline-specific member of the *AGO* family, *MEL1 (MEIOSIS ARRESED AT LEPTOTENE 1)*, has been found to be specifically expressed in archesporial cells, disappearing at the onset of meiosis (Nonomura *et al*., 2007), most likely regulating cell division of pre-meiotic germline cells and progression of meiosis, possibly suppressing somatic gene expression during germline development (Yang *et al.*, 2010).

1.5 The role of RNA-binding proteins in embryo sac development and function

1.5.1 An overview of RNA-binding protein biology

RNA-binding proteins (RBPs) have been identified as important regulators of RNA, forming **r**ibo**n**ucleo**p**rotein (RNP) complexes through interaction with both single- and double-stranded RNA (ssRNA and dsRNA) molecules, with capacity to regulate every aspect of their biogenesis and function, and the ability to recognize both primary RNA sequences and three-dimensional structures. Their involvement has been documented in all eukaryotic organisms, extending across the entire spectrum of metabolic and developmental process in animals and plants, as diverse as neural development, sex determination, embryo patterning, cell polarity determination, stem cell maintenance, stress response and flowering, to name a few. These are all achieved by association with RNA molecules during their entire lifespan, through involvement in all steps of their processing and metabolism, including splicing, 5' capping, polyadenylation, mRNA export and localization within the cell, translation, turnover, inhibition, stability and decay (Keene, 2007; Lunde *et al*., 2007; Glisovic *et al*., 2008; Bailey-Serres *et al*., 2009; Lorković, 2009).

Although a small number of RNA-binding proteins have been functionally characterized so far, thousands have been identified across kingdoms and more than 200 have been predicted in *Arabidopsis* and rice genomes, most of which are plant-specific, indicating that they might carry out a variety of plant-specific functions (Lunde *et al*., 2007; Bailey-Serres *et al*., 2009; Lorković, 2009; Ambrosone *et al*., 2012). They offer an extremely elegant and versatile manner of post-transcriptional gene regulation, adding to the genome plasticity and enabling the plant to engage in a swift response to changing environmental conditions.

1.5.2 Structural implications in RNA-binding proteins' performance

Despite the fact that RNA-binding proteins are a large and extremely diverse group, they share common features, most importantly the presence of one or more RNA-binding domains (RBDs) through which, as their name implies, they all interact with RNA molecules in order to fulfill a variety of functions (Lorković, 2009). Even though one might expect their functional diversity to be mirrored in their architecture, most RNA-binding

proteins are built from a few modules, and the RNA-binding domains serving as their building blocks are represented in numerous structural arrangements, accommodating a multitude of different substrates and extending the functional repertoire of these proteins (Lunde *et al*., 2007). Their modularity offers multiple benefits, reflected in their ability to associate with numerous RNA molecules, which are in many cases functionally related. RBPs also have the capability of recruiting a number of different co-factors, expanding their repertoire of target transcripts in a variety of cellular contexts and increasing their potential influence to a vast array of different processes. In addition to that, RBPs may be subject to different post-translational modifications and often localize to specific parts of the cell, all of which has an influence on the downstream functional activity of these complexes (Spassov and Jurecic, 2003; Keene, 2007; Glisovic *et al*., 2008).

Representative examples of some of the most common RNA-binding protein families are shown in Figure 7, illustrating the potential functional implications gained through modular assembly of multiple domains, the most common and widespread ones being the RRM (RNA recognition motif), and the KH (K-homology) domain, in addition to the PUF domain of Pumilio proteins, which are highly represented in the *Arabidopsis* genome and will be focused on further in the text.

RRM (RNA recognition motif) is by far the most frequently found and best characterized of the RNA-binding modules, containing two highly conserved RNP motifs (RNP1 and RNP2) which mediate binding in most cases through three conserved amino-acid residues (Arg or Lys, and two aromatic residues). Individual RRMs can bind RNA with very specific affinity, but multiple domains are often required to define further specificity, because the number of nucleotides recognized by an individual RRM is generally too small to define a binding sequence, usually being four to eight nucleotides in length. More than 10 000 RRMs have been identified so far in organisms ranging from bacteria to humans, carrying out functions in the vast majority of post-transcriptional gene-expression processes.

KH (K-homology) domain, able to bind both ssRNA and dsRNA, is ubiquitous in eukaryotes, eubacteria and archaea. It contains a functionally important signature sequence near the centre of the domain and unlike the RRM, its binding platform is free of aromatic amino acids, with recognition being achieved through hydrogen bonding, electrostatic interactions and shape complementarity instead. Proteins containing the KH domain have

been implicated in transcription, mRNA stability, translational silencing and mRNA localization.

These domains are often combined with each other and other protein domains, enabling proper functionality of the RNP complexes in various processes, including protein-protein interaction, protein targeting etc. (Lunde *et al*., 2007; Lorković, 2009; Ambrosone *et al*., 2012).

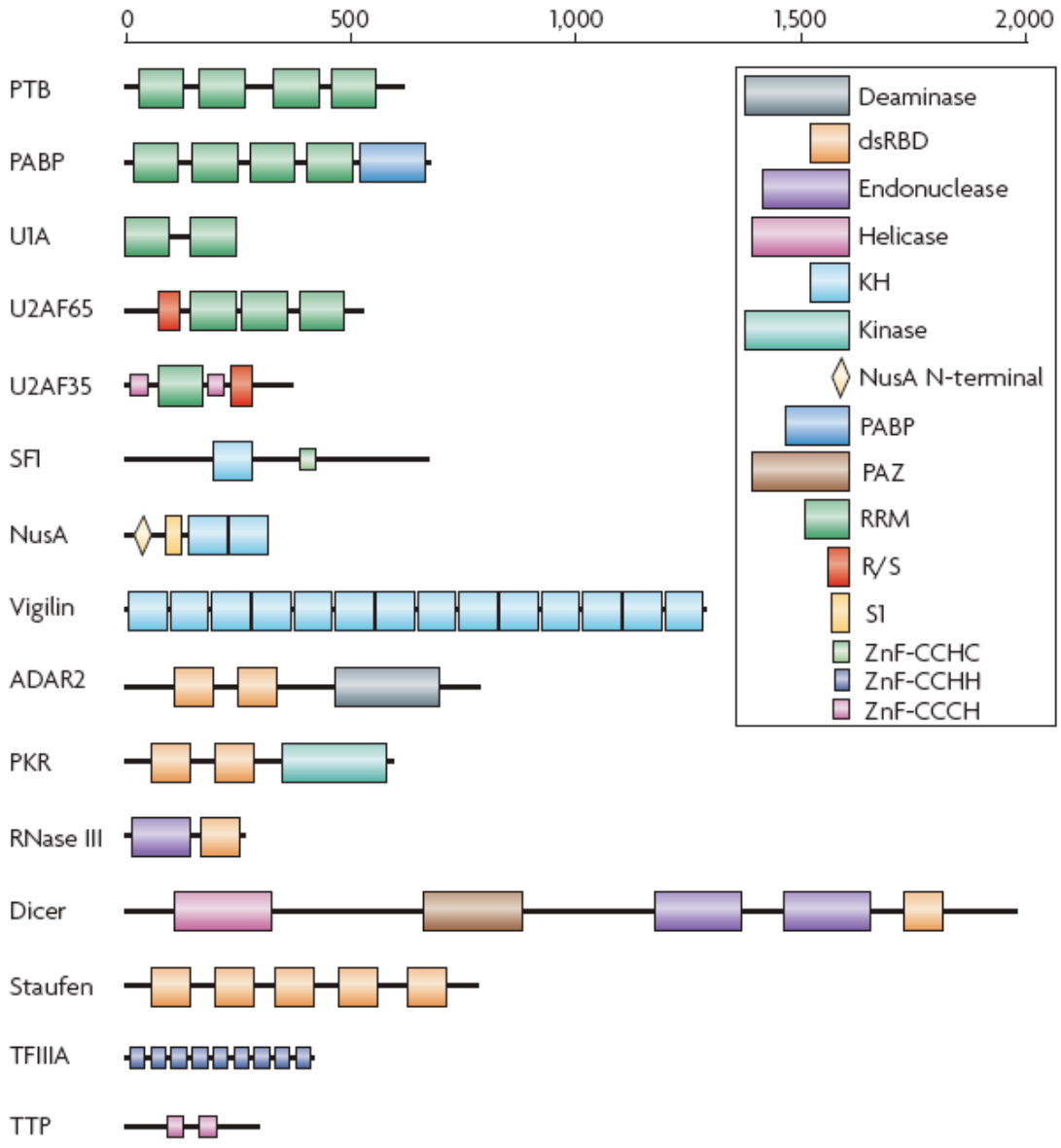


Figure 7. Modular structure of RNA binding proteins. Examples of representative members of RNA-binding protein families are depicted on the left, with the common domains, serving as their building blocks, shown in the box on the right. Image from Lunde *et al*. (2007)

In addition to the direct interaction of RNA-binding domains with RNA, interdomain arrangement is another factor playing a role as an important element in substrate recognition. A significant determinant for the affinity and specificity with which RBPs bind particular RNAs, resides in the amino acid sequence connecting their domains and the length and rigidity of linkers can have a dramatic influence on the RNA binding affinity. The length of these linker sequences is extremely variable, offering a range of features including precise domain positioning, conformational flexibility and potential for binding multiple RNA molecules, expanding the RNA-recognition interfaces of the protein (Lunde *et al*., 2007).

1.5.3 RNA-binding proteins – focus on particular functions

Although RNA-binding proteins play important roles in a wide spectrum of developmental processes, as already mentioned, this paragraph will focus on a few examples of their involvement in certain aspects of germline development, fertilization, sex determination and embryo development, particularly in the animal field, which often serves as a guideline for plant research. Many RNA-binding domains exhibit a highly similar architecture in different species, with sequences important in their binding and performance particularly highly conserved. This is especially true for proteins involved in essential cellular functions, shared across kingdoms of life, where certain parallels may serve as guidelines in finding the corresponding proteins for the analogue function in plants.

A great number of RBPs have essential functions during late germline and early embryo development because post-transcriptional control of maternal mRNAs is the dominant mode of temporal/spatial regulation of gene expression during this period. Because the genome becomes transcriptionally silent as chromosomes become condensed during the meiotic division, post-transcriptional control of pre-existing mRNAs, mainly through localization and translational regulation, is the predominant mechanism regulating protein expression in late gametogenesis, at fertilization and through early embryogenesis (Lee and Schedl, 2006).

Germline sex determination in *Drosophila* is, among other factors, heavily influenced by the *Sxl* (*Sex lethal*) gene, encoding an RNA-binding protein able to autonomously initiate female germline development (Hashiyama *et al*., 2011; Salz, 2011). A number of RNA-binding proteins have also been implicated in *Drosophila* embryo

patterning, including Pumilio, which was shown to interact with a spectrum of mRNA molecules involved in the process, including *bicoid*, *hunchback*, *oskar*, *nanos* and *caudal*, influencing their gradient along the anterior-posterior axis, as the major determinant defining general body architecture (Gamberi *et al*., 2002; Keene, 2007; Gupta *et al*., 2009). RNA-binding proteins characterized by the PUF domain (Pumilio/FBF), namely FBF-1/-2 and PUF5 - 8, are also shown be involved in both male and female gametogenesis, germline sex determination and stem cell maintenance in *C. elegans*, along with a number of other RBPs, particularly through a carefully orchestrated interaction with GLS1 and the GLD3 RNA-binding protein (Lee and Schedl, 2006; Lublin and Evans, 2007; Rybarska *et al*., 2009; Merritt and Seydoux, 2010). The expression of both human and mouse Pumilio proteins is widespread and overlapping between the two members they possess, observed in the germline, where human PUM2 was shown to interact with a germline-specific protein DAZ (Deleted in azoospermia) and with GEMIN3 in several spermatogenic stages, in fetal and adult hematopoetic cells, as well as neural stem cells, further supporting the notion of their primary function in stem cell maintenance (Moore *et al*., 2003; Spassov and Jurecic, 2003; Francischini and Quaggio, 2009; Ginter-Matuszewska *et al*., 2011). Contrary to Pumilio proteins in animals, little is known about plant Pumilios and given that they are far more numerous in plants they are likely to be involved in a number of various biological processes.

1.5.4 Pumilio family of RNA-binding proteins

1.5.4.1 General features of Pumilio proteins

In the last decade or two, members of the Pumilio family of RNA-binding proteins emerged as important translational regulators during embryo development, cell fate specification and differentiation. They are present in all eukaryotic phyla, from yeast to mammals and plants, and characterized by the highly conserved PUF domain (Pumilio/FBF), named after the founding members, *Drosophila* Pumilio and *C. elegans* FBF (fem-3 binding factor), positioned at the C-terminal and composed of several tandem repeats, usually eight (Spassov and Jurecic, 2003). These imperfect repeats comprising the domain consist of 36 amino acids and conserved N- and C-terminal flanking regions, which resemble half-repeats and are often termed repeat 1' and repeat 8'. The most conserved

amino acid residues are usually placed in the middle of each repeat, at positions 12, 13 and 16 and interact with the RNA molecules, with each repeat recognizing and binding to a specific RNA base, the common core recognition triplet being UGU, along with additional bases downstream involved in the interaction, as the Pumilio binding site is composed of eight nucleotides. The specific sequences recognized by these proteins are known as nanos response elements (NREs) and are typically situated in the 3' UTR of the target transcripts. All of the so far known mRNAs regulated by Pumilio contain a $UGURN_{1\text{-}3}AU(A/U)$ recognition sequence (where R = purine, N = any base), and an alignment of some of the examples from several species is depicted in Figure 8 (Spassov and Jurecic, 2003; Lunde *et al*. 2007; Francischini and Quaggio, 2009).

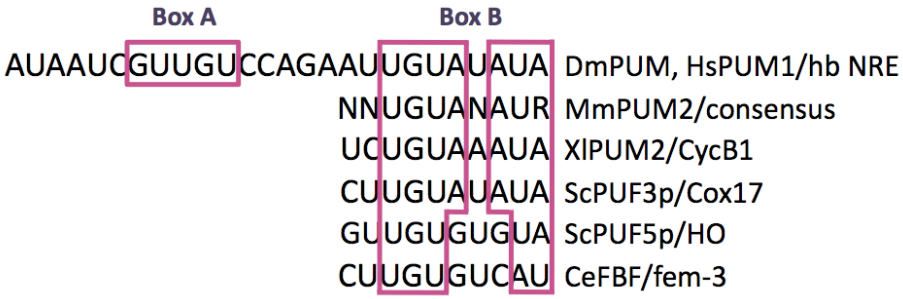


Figure 8. The conserved binding sequence of Pumilio proteins. The recognition sequences in the target transcripts are framed in pink and the Pumilio proteins from different species / their targets are listed on the right. The *Drosophila* binding sequence was thought to be bipartite and composed of Box A and Box B, however, all other members bind to only one box, reminiscent of Box B. Image adapted from Spassov and Jurecic (2003).

Very different numbers of Pumilio family members have been found in several of the model organisms examined, displaying different levels of variability in sequence and functionality within a species. To illustrate, *Drosophila melanogaster* only has a single member, all vertebrates examined so far have two members each, six members are present in yeast, *C. elegans* counts eleven, rice has 19 and *Arabidopsis* 26, which is the highest number recorded in any single species so far.

Since Pumilio proteins have been highly conserved through evolution, members from different organisms recognize RNA with same modularity. For example, human PUM1 and mouse Pum2 bind very similar, if not identical sequences, and can both bind the NRE of *Drosophila hunchback* mRNA with even higher affinity than the native *Drosophila*

PUM. The six *Arabidopsis* PUMs belonging to Group 1, are also able to specifically recognize and bind the NRE sequence of *Drosophila hunchback* mRNA, but point mutations in the NRE or amino-acid substitutions in the binding sites can drastically impair the binding (Spassov and Jurecic, 2003; Tam *et al*., 2010).

1.5.4.2 Plant Pumilio - focus on Arabidopsis

Although Pumilio proteins have been studied to some extent in animals during the last couple of decades, not much is known about plant Pumilio and the research efforts in this area are currently increasing using the model plant *Arabidopsis thaliana*. The 26 members of the Pumilio family identified in *Arabidopsis* have been categorized into four groups (Group I – IV) and their phylogenetic relationships can be seen in Figure 9. Since plants seem to have a significantly higher number of Pumilio family members in comparison to animals, it is possible that many of them possess specialized plant-specific functions and might thus display variations in their binding motifs, perhaps specialized for particular classes of RNAs. The variability in sequence and architecture observed so far in a number of members would certainly point in that direction. Both *Arabidopsis* and rice Pumilio proteins exhibit unconventional amino acid triplets in the positions important for RNA binding, demonstrating complete conservation in some of their members and extensive variability in others. The amino acid residues in question, at positions 12, 13 and 16 of each repeat, resemble human PUM1 and the *Drosophila* Pumilio in the AtPUM proteins belonging to Group I (AtPUM1 – AtPUM6). Members of Group II (AtPUM7 – AtPUM12) on the other hand, are more similar to the yeast Puf4 and Puf5. The remaining AtPUMs don't show similarity to other known Pumilio proteins and, unlike the members of the first two groups, posess less than eight repeats (Francischini and Quaggio, 2009; Tam *et al*., 2010). The basic architecture of *Arabidopsis* Pumilio proteins is schematically depicted in Figure 10.

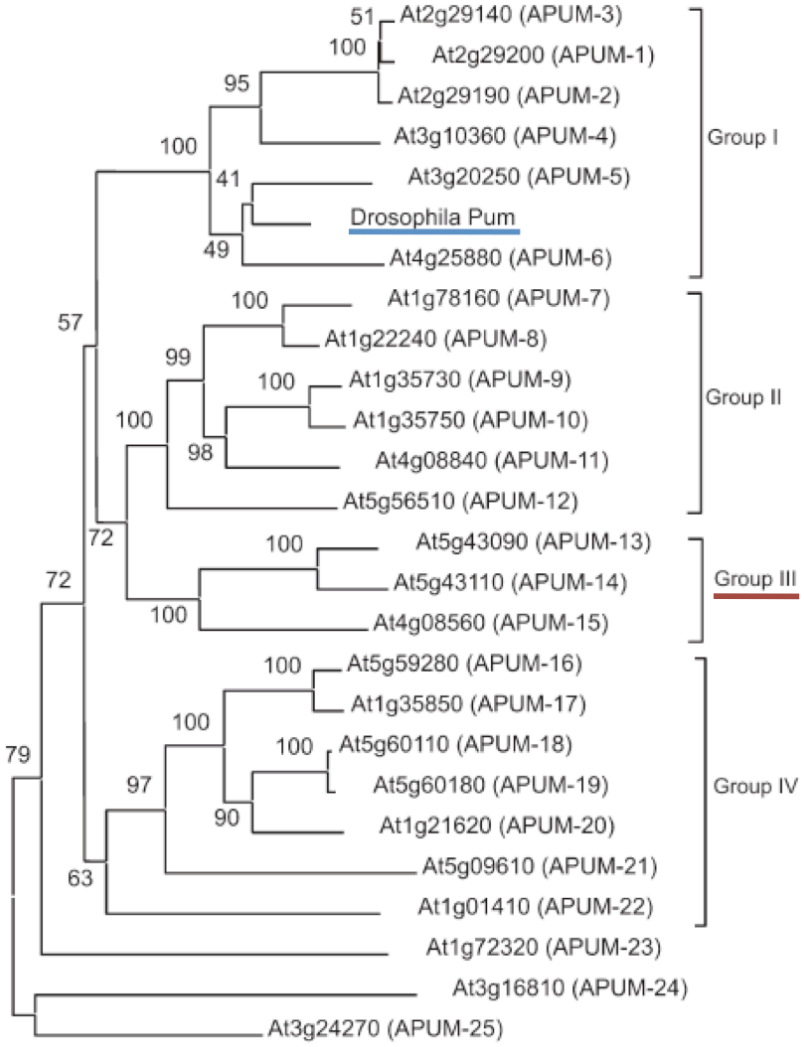


Figure 9. Phylogenetic relationships of *Arabidopsis* Pumilio proteins. The dendrogram includes 25 members of the Pumilio family of RNA-binding proteins found in *Arabidopsi*s. The proteins are divided into four groups, according to their degree of similarity, along with three proteins as outliers. Another member has been added to Group IV since (AtPUM26 / At5g64490), bearing most similarity to AtPUM22. Group I proteins share most similarity to the *Drosophila* Pumilio, which is included in the dendrogram and underlined in blue. This work will later focus more on the three Pumilio proteins comprising Group III, which is underlined in red. Image taken from Francischini and Quaggio (2009).

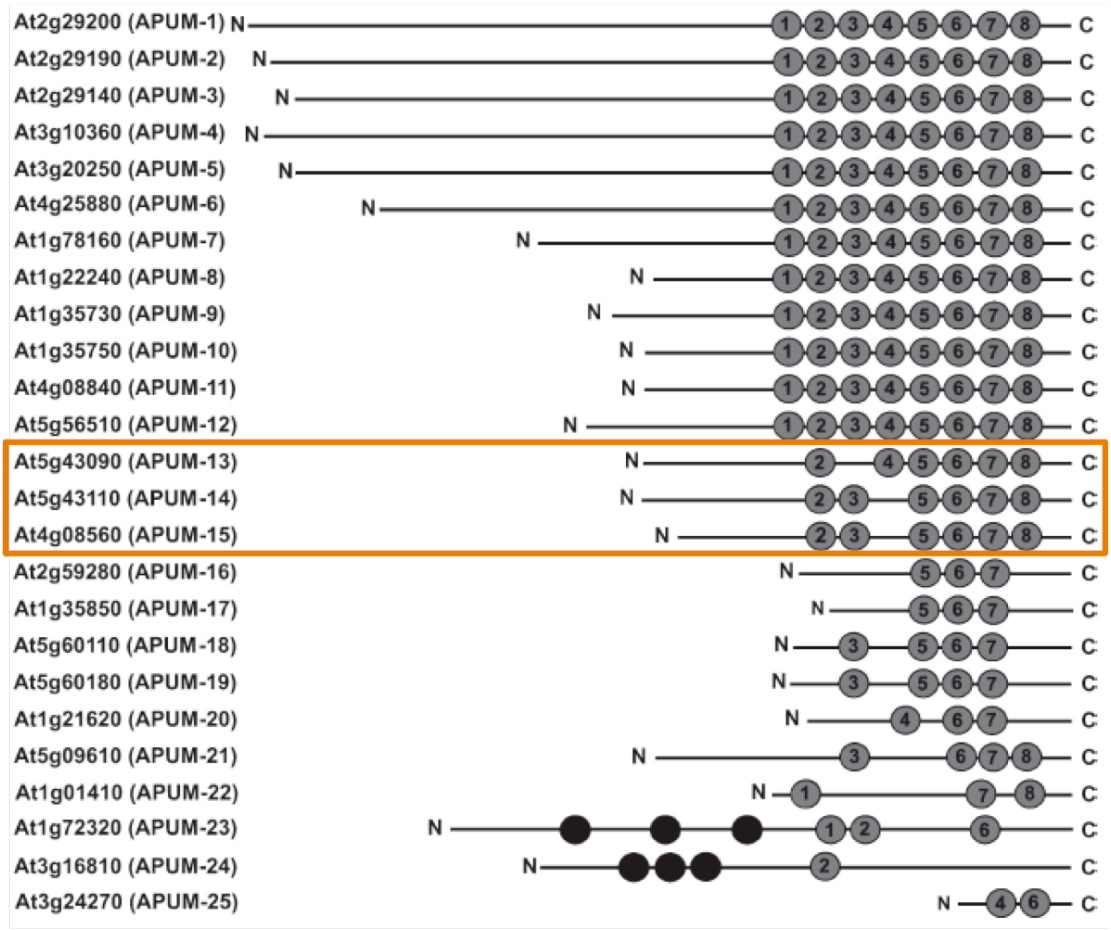


Figure 10. A scheme of basic general architecture of Pumilio proteins in Arabidopsis. The proteins are characterized by up to eight conserved repeats at the protein's C-terminus, represented by numbered grey circles. The black circles represent repeats that fall outside the C-terminal region. The three proteins framed in orange represent the Group III Pumilios, identified to be of special interest and focused on further in the scope of this work. Image from Francischini and Quaggio (2009).

A limited number of Pumilio target transcripts have been identified in the animal field and close to none in plants. The first mRNAs that are the potentially regulated by AtPUM proteins were recently identified and they code for proteins implicated in stem cell maintenance, self-renewal and cellular organization of the shoot apical meristem. The four transcripts that tested positive for interaction with Group I Pumilio proteins so far encode CLAVATA-1 (CLV-1), ZWILLE/PINHEAD (ZLL), WUSCHEL (WUS) and FASCIATA-2 (FAS-2), but the consensus binding sequence for Pumilio suggests there may be a great number of other transcripts that are also potential targets for regulation by these proteins. Two new NRE Box B-like recognition sequences were identified in the same study and were named AtPUM binding elements (APBE), confirming suspicions of novel binding

motifs in plants. A screen of all Arabidopsis sequences annotated in TAIR (The Arabidopsis Information Resource) database revealed that 43 % of all 3' UTRs have at least one binding consensus for AtPUM proteins, so more candidate transcripts are expected to be identified and characterized in the near future. Given that this information is based on research performed on *Arabidopsis* Group I Pumilio, other RNA binding sequences are likely to arise as members of other groups are examined, which might further expand the number of potential transcripts regulated by the Pumilio proteins. However, it is not expected that all of them will indeed prove to be true targets for regulation, since other factors can influence the interaction, such as the sequences flanking the consensus binding site, which have proven to play a significant role in the binding affinity (Francischini and Quaggio, 2009).

1.6 Aims of this work

Because the female gametophyte of *Arabidopsis thaliana* is deeply embedded in the maternal tissue of the ovule it has been difficult to gain insight into the molecular underpinnings of its development and function, especially in the context of individual cells that comprise it. One of the aims of this study has been to obtain single, isolated cells of the *Arabidopsis* female gametophyte by adapting a previously established method of single cell isolation from the female gametophyte of maize (Kranz *et al*., 1993), and use them for large-scale gene expression profiling through a microarray study. The focus was placed on a number of topics of particular interest, including cell cycle, gene and hormone regulation, signaling, as well as identification of genes specifically expressed in individual cell types examined, that would provide candidates for future studies into their particular function and establishment of cell identity. Moreover, at least one class of RNA-binding proteins being specifically up-regulated in the embryo sac cells shall be studied in more detail. A subclade of the Pumilio family of RNA-binding proteins was selected for futher investigation because this class of proteins is known to be involved in germline development in the animal field (Gamberi *et al*., 2002; Moore *et al*. 2003; Lublin and Evans, 2007; Ariz *et al*., 2009; Kalchhauser *et al*., 2011). The gene expression pattern, sub-cellular localization and function of these proteins shall be examined also through the use of knock-out mutants. Three family members comprising Group III Pumilios were selected for these studies.

2 MATERIALS AND METHODS

2.1. Enzymes, chemicals, consumable materials

Molecular grade chemicals, enzymes and other consumables were purchased from companies as stated in the text and prepared, where necessary, according to manufacturers' instructions, unless otherwise mentioned.

2.2 Primer sequences

All primer sequences were synthesized by the biopolymer factory "biomers.net" (http://www.biomers.net/de.html), prepared in TE buffer (10 mM Tris-HCl, pH=8.0; 0.1 mM EDTA) to a 100 µM stock solution and diluted to 10 µM in deionized H_2O before use. The stocks and working solutions were stored at -20 °C. Oligonucleotide sequences used in this work are listed in the Appendix, unless otherwise stated in the text.

2.3 Plant material and growth conditions

All plants used throughout this work were *Arabidopsis thaliana* ecotype Col-0, unless otherwise stated and their seeds were sown in a mix of soil (65 %), sand (25 %) and expansion clay (10 %), followed by stratification at 4 °C for 2-3 days. After germination, plants used for single cell isolation were first grown for four weeks at short-day conditions (9 h light, 8,500 lx, 22 °C) and then transferred to long-day conditions (16 h light, 8,500 lx, 22°C) after 4 weeks. T-DNA insertion lines mentioned in the work (SALK, SAIL and GABI Kat lines) and plants used for genotyping, promoter and phenotype analyses were grown at long-day conditions after germination.

2.4 Preparation and transformation of competent cells

2.4.1 Chemically competent *Escherichia coli*

The preparation of competent *E. coli* bacterial cells, strain DH5α (Woodcock *et al*., 1989) was performed as previously described (Inoue *et al*. 1990) using a Sorvall Evolution centrifuge (Du Pont Instruments) with the SLA-1500 rotor.

2.4.2 Chemically competent *Agrobacterium tumefaciens*

For preparation of competent *A. tumefaciens* bacterial cells, single colonies of strains *GV3101* (*pMP90RK*) and *GV3101* (*pMP90, pSOUP*) were picked and grown over-night in 5 ml of liquid YEP medium (bacto peptone 10 g/l, yeast extract 10 g/l, NaCl 5 g/l in ddH_2O) at 28 °C, with shaking at 200 rpm. The over-night culture was transferred into 200 ml YEP medium. After reaching an OD_{600}= 0.6-0.7 the cells were centrifuged at 5300 rpm, 4 °C for 20 minutes in a Sorvall Evolution centrifuge (SLA-1500 rotor). The pellet was resuspended in 200 ml of pre-cooled TE buffer (10 mM Tris-HCl, pH=7.5, 1 mM EDTA) and centrifuged again at the same conditions. The pellet was then resuspended in 2 ml 20 mM $CaCl_2$. 100 µl aliquots were frozen in liquid N_2 and stored at -80 °C until further use.

2.4.3 Transformation of chemically competent cells by "heat shock"

Competent bacterial cells were thawed on ice. In the case of *E.coli* 100 – 200 ng plasmid DNA was added to cell suspensions and incubated on ice for 5 – 30 minutes. Cells were then transferred to a 42 °C water bath for 45 seconds and subsequently supplemented with 250 – 500 µl LB medium, incubated at 37 °C for 1 h with shaking (250 rpm), plated on LB plates supplemented with appropriate antibiotics and grown over-night. For *A. tumefaciens* ≥ 1µg plasmid DNA was added to cell suspensions and incubated on ice for 5 minutes. Cells were then kept in liquid N_2 for 5 minutes and in a 37 °C water bath for 5 minutes. They were then supplemented with 500 ml YEP medium, incubated at 28 °C for 4 h with shaking (200 rpm), plated on YEP plates (YEP medium containing 15 g/l agar) supplemented with appropriate antibiotics (concentration listed in the table below) and grown at 28 °C for two days.

Antibiotic	**Final concentration in medium (µg/ml)**	
	E. coli	***A. tumefaciens***
Kanamycin	50	50
Gentamycin	/	40
Ampicilin	100	/
Spectinomycin	50	100
Rifampicin	/	20
Tetracyclin	/	30

2.5 Isolation of plasmid DNA and concentration measurement

For plasmid DNA isolation from *E.coli*, a single colony was picked and cultured over-night in 3 ml of LB medium supplemented with the appropriate antibiotic. Cells were harvested by two centrifugation steps (30' at 13,000 rpm) and plasmid DNA was isolated using the High-Speed Plasmid Mini Kit (Avegene) following manufacturers' instructions. Concentration of DNA used in all applications within the scope this work was measured using the NanoDrop™ spectrophotometer (Thermo Scientific).

2.6 Isolation of single cells of the female gametophyte

Single cells of the female gametophyte were isolated from *Arabidopsis thaliana* (ecotype Columbia-O) in the following manner. Pistils were collected 2 days after emasculating flowers at flower stage 10-11 (Smyth et al., 1990), put on glass slides with a centrally placed well inside a 50 µl droplet of 730 mOsmol mannitol and cut open lengthwise through the septum using a disposable hypodermic needle (Ø 0.4 x 20 mm, Braun). Ovules were carefully pushed out and subsequently incubated for 1 h in a humid box after adding another 50 µl of the mannitol solution and 50 µl of filter-sterilized enzyme mix containing 0.75 % Pectinase (Serva, Heidelberg, FRG), 0.25 % Pectolyase Y23 (Seishin, Tokyo, Japan), 0.5 % Hemicellulase (Sigma), 0.5 % Cellulase "Onozuka" RS (Yakult Honsha, Tokyo, Japan) with the pH value adjusted to 5.0, prepared in mannitol to a final osmolarity of 730 mOsmol. Following incubation in the enzyme solution, ovules were observed using an inverted microscope (Eclipse TE2000-S, Nikon). Numerous cells, including cells comprising the embryo sac, are released from the ovule tissue, very often still attached to the micropyle. Female gametophyte cells are then carefully detached from the ovule but usually still stay attached together and often also to other cells from the surrounding tissue (Figure 11, B – C) and therefore need to be carefully separated. This is done manually using a micro-glass needle (pulled by hand from Bo-glassrods, Ø 2 mm, Hilgenberg). Initially, homozygous marker-lines for the egg cell (*Ec1p:TaMAB2::GFP*, provided by Stefanie Sprunck), and for the synergid (*MYB98p:GFP* (Kasahara *et al*., 2005), obtained from the lab of Gary Drews) were used as an aid in identifying individual cells. This helped in determining specific features that distinguish female gametophyte cells from each other as well as from other cells found in solution following the enzyme treatment

(Figure 11 A), so that wild-type plants could also be used for isolation later. After cell separation the particular features of each cell type are more clearly visible (Figure 11, B – D). The average cell sizes are illustrated in Figure 11. The central cell is the biggest female gametophyte cell and can be identified by the presence of a large vacuole as well as numerous starch granules and chloroplasts. The synergid cells, visibly smaller than the central cell, have a uniform appearance without particularly pronounced features, a homogenously thick cytoplasm that appears somewhat darker and slightly silvery when microscopically observed, occasionally with a visible nucleus. The egg cell is similar in size to synergid cells, often slightly smaller and appears more vacuolated, usually with a visible nucleus. Its vacuole is in most cases clearly distinguishable on one side of the cell and takes up approximately half of the cell volume with most of the cytoplasm occupying the other half as well as surrounding the vacuole in a thin layer (Figure 11 B). Occasionally, the vacuole looks divided by many cytoplasmic strands, creating the appearance of numerous smaller vacuoles (Figure 11 D). Cells can be already visually assessed to determine their quality. It is often observed that synergid cells fuse together or with other cell types, becoming unusually big with more than one nuclei visible, or are permanently stretched and don't have a plump, round shape exhibited by healthy cells, which also show visible and dynamic cytoplasmic streaming. These cells are not collected nor used for further work (Figure 11 E).

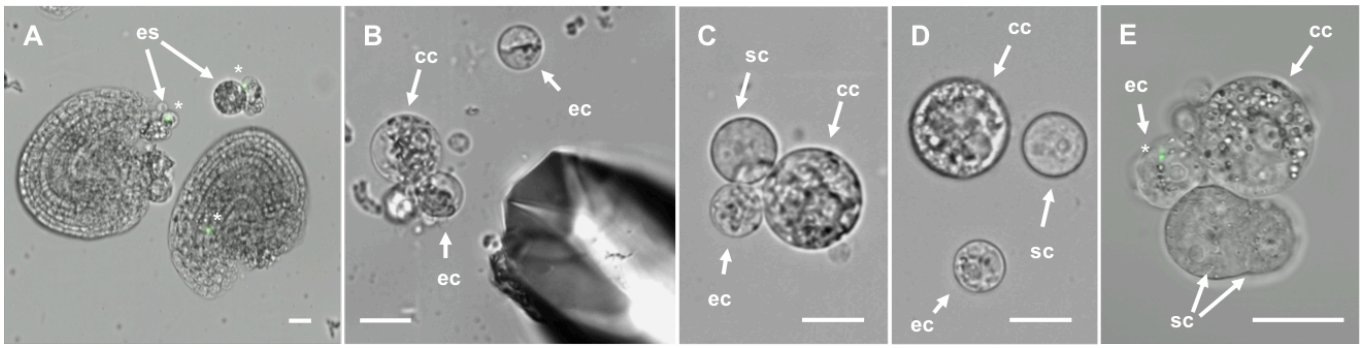


Figure 11. Illustration of the single cell isolation procedure. **(A)** After the manual detachment of ovules from the carpel tissue of Arabidopsis pistils followed by enzyme treatment, cells of the embryo sac are released in solution or still attached to the micropyle (arrows). (**B**) The tip of the needle used to separate cells is shown in the lower right corner illustrating the size ratios and the difficulty of detaching the cells from one another with such a tool without damaging them. (**C**) A wild-type embryo sac is shown before cell separation and (**D**) a single egg cell, central cell and synergid cell after separation of the same, with distinguishable morphological differences. (**E**) Cells that have rough or irregular edges or start fusing together (i.e. egg cell and two synergid cells shown) are not collected for further work. Images shown in (A) and (E) depict ovules

and cells from the *Ec1p:TaMAB2-GFP* marker-line showing a fluorescent signal in the egg cell (marked with *). Images shown in (B) – (D) depict isolated embryo sac cells from wild-type *Arabidopsis* plants. Embryo sac (es), egg cell (ec), central cell (cc), synergid cell (sc). Scale bars are 20 µm.

The next step of the procedure makes use of custom-made microscope slides, made by attaching a 24 x 40 mm cover slide to the bottom of a metal slide over a central perforation (Ø 20 mm) with melted wax. The well, made in this way, is filled with thick paraffin oil (Sigma). A fine glass capillary pulled from Disposable MicroPipettes (Hirschmann® ringcaps®) using the Flaming/Brown Micropipette Puller model P-97 (Sutter Instrument Co.) is attached to a manually operated pump (CellTram air, Eppendorf) and used to place drops (approx. volume 1-1.5 µl) of sterile 730 mOsmol mannitol solution into the paraffin oil. In the last step cells are transferred two times with the glass micro-capillary into 730 mOsmol mannitol drops made in paraffin oil for washing. Cells are collected using the glass micro-capillary together with a minimal amount of mannitol solution, placed in a 0.5 ml Eppendorf tube, immediately frozen in liquid nitrogen and stored at -80 °C until further use.

2.7 mRNA isolation and cDNA synthesis from tissues

For expression analyses of individual genes a small amount of selected vegetative or generative tissue (approx. 3 – 5 mg) was collected in 0.5 ml tubes (Eppendorf) and immediately frozen in liquid nitrogen. The tissue was ground to fine powder while still frozen and mRNA was isolated with the Dynabeads mRNA DIRECT™ Micro Kit (Invitrogen) following manufacturers' instructions. Some batches of mRNA were, after isolation and before the cDNA synthesis step, additionally treated with DNase I (MBI Fermentas) following manufacturers' instructions (cDNA synthesized from these batches was used as template in cases where the use of intron-spanning primers wasn't possible). The quality of the cDNA was checked by PCR using *Actin3* primers (see primer sequences listed in the Appendix) and subsequently used for standard PCR analysis (see 2.8) with the appropriate primers. The same kit was used for mRNA isolation from batches of single cells as described previously (Sprunck *et al.*, 2005) due to the small amount of starting material. First-strand cDNA synthesis was carried out using Oligo$(dT)_{18}$ primers (MBI

Fermentas) and RevertAid™ M–MuLV Reverse Transcriptase (MBI Fermentas), or Oligo(dT)$_{23}$ primers (SIGMA) and SuperScriptIII® Reverse Transcriptase (Invitrogen) following the manufacturers' protocols with the addition of 1 µl RiboLock™ Ribonuclease Inhibitor (MBI Fermentas).

2.8 Standard PCR analysis

Reactions were performed in a 50 µl volume containing 1x PCR buffer, 1.5 mM $MgCl_2$, 200 nM of each forward and reverse primers, 200 µM dNTP mix, 1 U *Taq* DNA Polymerase, ~ 100 – 300 ng genomic or 10 – 50 ng plasmid DNA as template. Initial denaturation step was performed at 95 °C for 2 min, followed by 28 – 38 cycles of denaturation at 95 °C for 30 s, annealing at appropriate temperatures for 30 s, extension at 72 °C for 1 min/kb. A final extension step was performed for 5 min at 72 °C to ensure complete polymerization of all products. The PCR reaction was then maintained at 4 °C until further use.

2.9 Single cell RT-PCR and Southern blot analysis

For each of the three embryo sac cell types (egg cells, central cells, synergid cells) fifteen cells were each pooled and used for mRNA isolation, as described above. The obtained mRNA was used for reverse transcription with RevertAid™ H Minus M-MuLV Reverse Transcriptase (MBI Fermentas). All reactions yielded around 20 µl cDNA of which 2 µl were used as a template for standard PCR (see 2.8) using HotStar *Taq* DNA polymerase (Qiagen). After testing the quality of the cDNA with primers for *AtCB5* (see primer sequences listed in the Appendix), encoding a constitutively expressed cytochrome b5 isoform, reactions were performed using primers for cell-specific genes (primer sequences are listed the Appendix). These were (i) *Ec1-2a*, a member of the egg cell-selective *EC1* gene family (Sprunck *et al*., unpublished), (ii) the central cell-selective *DD65* (Steffen et al., 2007) and (iii) the synergid cell-selective *DD31*gene (Steffen et al., 2007). The expression of these genes was tested with cDNA of all three embryo sac cell types as template. To increase sensitivity of detection, gels were blotted onto Hybond N+ nylon

membranes (Amersham) after agarose gel-electrophoresis and hybridized with Digoxigenin (DIG) labelled probes of *Ec1-2a, DD65* and *DD31*. Probes were generated by standard PCR (see 2.8) from genomic DNA, precipitated (0.1 Vol. 3M NaOAc, pH=5.2 and 3 Vol. 96 % EtOH) and subsequently labelled using the random priming DIG DNA Labeling Kit (Roche), following manufacturer's instructions. Hybridization with 5-25 ng probe/ml hybridization solution was carried out at 42°C using DIG Easy Hyb solution (Roche) and chemiluminescent detection was performed with CSPD® (Roche) according to the manufacturers´ guidelines. The washing buffer from the CSPD® protocol was slightly modified and was prepared with 0.1 M maleic acid and 0,3% Tween only, omitting 0.15 M NaCl. Hybridization signals were detected by exposing Amersham Hyperfilm™ MP X-ray film to membranes wrapped in plastic film for 3 min at 37°C.

2.10 Microarray hybridization

GeneChip® Arabidopsis ATH1 Genome Arrays (Affymetrix) were used for gene expression analysis of single cells of the *Arabidopsis* female gametophyte. Three biological replicates were performed for each cell type. mRNA was isolated as described (see 2.7) from 50 single cells for each replicate, with the exception of one egg cell replicate for which 30 cells were used. The mRNA was subsequently amplified using TargetAmp™ 2-Round Aminoallyl-aRNA Amplification Kit (EPICENTRE Biotechnologies) following manufacturers' instructions, stopping the procedure after Round-Two, 2nd strand cDNA Synthesis (step F.6. in the mentioned instruction manual). Cleanup of double-stranded cDNA, last round of *in vitro* transcription and labelling was performed according to standard Affymetrix GeneChip® (Affymetrix) protocols by KFB (Kompetenzzentrum für Fluoreszente Bioanalytik, Regensburg, Germany). All subsequent steps of fragmentation of cRNA (copyRNA), hybridization, washing, staining and scanning were also conducted by KFB.

2.11 Microarray data analysis

Absent and Present calls were generated using Affymetrix *GeneChip Operating Software (GCOS) 1.4.* All subsequent analyses were performed using dChip software 2008

(http://www.hsph.harvard.edu/~cli/complab/dchip/). All GeneChip® experiments were performed using three biological replicates. The data obtained from the female gametophyte cell samples was analyzed together with previously published datasets of sperm cells, pollen and seedling (Borges *et al*., 2008). For normalization, all three samples of female gametophyte cell experiments were considered a single tissue type and one array (C2) was scaled to the median median intensity (85) of all samples used in this study. The remaining arrays were then normalized to this array (baseline) applying the Invariant Set Normalization Method (Li and Wong, 2001). Sperm cell, pollen and seedling arrays were normalized in the same way to the target intensity of 85, but treating each sample type as a single tissue type. Normalized CEL (array data file format) intensities of the 18 arrays were then used to obtain model-based gene expression, based on a PM-only model (Li and Wong, 2001). Potential array outliers detected in replicates of one tissue/cell type were not called array outliers given the assumption that these genes might be expressed in a tissue-specific manner and are therefore not true array outliers. Differential expression was detected with a lower confidence bound of the fold change between experiment and baseline of 1.2. False discovery rate (FDR) was under 10 % in all cases. Genes called Present in at least two of three replicates were taken into consideration as expressed above the detection limits for further analysis and will be referred to as "called Present" in further text, unless otherwise specified. Venn diagrams of overlaps of genes called Present in individual replicates of each sample were created using online tools (http://bioinfogp.cnb.csic.es/tools/venny/index.html and http://www.venndiagram.tk). Principal Component Analysis (PCA) was computed using Partek Genomics Suite 6.4 (Partek). Comparative analysis of egg cell, central cell and synergid cell data was performed together with datasets of sperm cells, pollen and seedling (Borges *et al*., 2008). A four-way Venn diagram was generated online (http://bioinfogp.cnb.csic.es/tools/venny/index.html) (Oliveros, 2007) and manually adjusted with surface areas and overlaps approximately corresponding to the number of genes called Present in each of the samples. In order to narrow down the number of genes selectively expressed in single-cell types Excel (Microsoft) was used to manage the data using the Present and Absent calls of all arrays. The list of probe sets obtained in that manner (egg cell – 216 genes, central cell – 559 genes, synergid cell – 109 genes) was then uploaded on Genevestigator software (Zimmermann *et al*., 2004), the Digital Northern Tool, comparing expression using only high quality arrays of the AtGenExpress Database (Schmid *et al*., 2005) from experiments of developmental stages of *Arabidopsis thaliana.* Samples

containing flower tissue were omitted. Genes called Present with a p-value ≤ 0.05 in at least two of three replicates were excluded. Hierarchical clustering was performed using dChip software. The cell cycle correlation coefficient overview was made using GraphPad Prism 5 (GraphPad Software, Inc) by Jörg Becker. The bars are based on pair-wise comparisons of the sample triplicates. Receptor-like kinases (RLKs) were categorized into subfamilies according to Shiu and Bleecker (2001) and cysteine-rich proteins (CRPs) into homology subgroups according to Silverstein *et al*. (2007). Auxin-related genes (list obtained from Stefanie Sprunck) were managed using Excel, scaling colour according to the intensity of expression values.

2.12 Microarray data validation

2.12.1 Data validation by PCR analysis

After mRNA isolation from 10-20 single cells, as described above (see 2.7), cDNA was synthesized using SuperScript® III Reverse Transcriptase (Invitrogen) following manufacturer's instructions, tested for quality with *AtCB5* primers (see Appendix), and subsequently used as template for standard PCR (see 2.8) using gene-specific primers (listed in the Appendix), running 38 cycles.

2.12.2 Data validation by *promoter::GFP* fusion analysis

In order to verify the microarray expression data, promoter-reporter constructs were generated for selected candidate genes. Promoter regions were defined using TAIR-Sequence Viewer (http://www.arabidopsis.org/servlets/sv) and amplified by PCR (the primers used are listed in the Appendix) using Phusion™ High-Fidelity DNA Polymerase (Finnzymes). PCR fragments of corresponding size were cut out from agarose gels after electrophoresis and purified using QIAquick Gel Extraction Kit (Qiagen). Fragments were cloned into the pENTR/D-TOPO (Invitrogen) vector according to manufacturer´s instructions creating entry clones which were subsequently used for a recombination reaction with LR Clonase™ II enzyme mix (Invitrogen), transferring each of the promoter-fragments into the binary Gateway® destination vector *pGW:NLS:3xEGFP::nost-pGII, pGII*, a modified version of

pGII-NLS:3xEGFP::nost-pGII (Takada and Jürgens, 2007), provided by Mily Ron. Binary vectors generated in this way were delivered into *Agrobacterium tumefaciens*, strain *GV3101* (*pMP90, pSOUP)*, cultured in YEP liquid medium (bactopeptone 10 g/l, yeast extract 10 g/l, NaCl 5 g/l in ddH_2O) and used to transform *Arabidopsis thaliana* via the floral-dip method (Clough and Bent, 1998). After germination BASTA® resistant seedlings were selected by spraying with 200 mg/l BASTA® (Bayer Crop Science) supplemented with 0.1 % Tween every two days (altogether three times).

2.13 Microscopic analysis of ovules

Pistils were collected two days after emasculation and dissected using a SMZ645 binocular microscope (Nikon) and fine hypodermic needles (see 2.6) on a glass slide to remove carpels and stigma, leaving ovules preferentially still attached to the septum. For standard microscopic analyses and GFP expression, pistils were prepared in 50mM sodium-phosphate buffer. For propidium-iodide staining of ovules dissection was performed in 50 µl of freshly prepared propidium-iodide (Sigma) (0.1 mg/ml in ddH_2O). Slides were then incubated for 10-30 min in a humid box to prevent preparations from drying out, and examined with Confocal Laser Scanning Microscopy (CLSM), using Zeiss Axiovert 200 M microscope equipped with a confocal laser scanning unit LSM 510 META. The 488 nm Argon laser was used with band pass filter BP505-530 for detection of GFP fluorescent probes. Propidium-iodide stained ovules were excited with the 543 nm Helium-Neon laser using a long pass filter LP560. Pictures were made with an AxioCam HRc camera (Zeiss) and the Zeiss LSM 510 META software. Stacks of pictures and their further processing were performed with the Zeiss LSM Image Browser Version 3.5.0.359. Additional single cell images (Figure 11, A – D) were obtained using the Eclipse TE2000-S inverted microscope (Nikon) and Axiovert imaging software (Zeiss).

2.14 Expression pattern and silencing of members of the Pumilio family of RNA-binding proteins

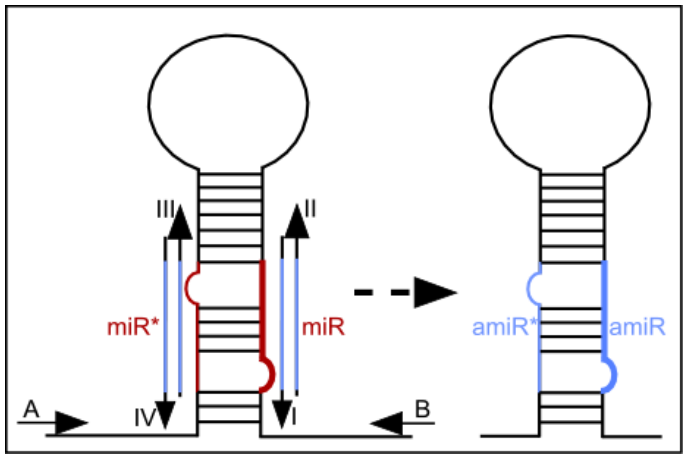
2.14.1 *Promoter::GFP* fusion analysis – generation of constructs and microscopic analysis

For promoter analysis of *AtPUM13* (*At5g43090*), *AtPUM14* (*At5g43110*) and *AtPUM15* (*At4g08560*), promoter regions of *AtPUM13* and *AtPUM15* were defined using TAIR-Sequence Viewer (http://www.arabidopsis.org/servlets/sv) and amplified from genomic DNA by PCR using Phusion™ High-Fidelity DNA Polymerase (Finnzymes). Transgenic *AtPUM14p:NLS(3x)GFP* lines were provided by Stefanie Sprunck. The open reading frame of *AtPUM14* was defined in the same manner and amplified using the peqGOLD 'Pure' *Pfu*-DNA-Polymerase (Peqlab). PCR fragments of corresponding size were purified from agarose gels after electrophoresis and purified using QIAquick Gel Extraction Kit (Qiagen). Fragments were cloned into the pENTR/D-TOPO (Invitrogen) vector according to manufacturer´s instructions creating entry clones which were subsequently used for a recombination reaction with LR Clonase™ II enzyme mix (Invitrogen). Each of the promoter-fragments was transferred into the binary Gateway® destination vector *pGII-GW:NLS:3xeEFP::nost* (Provided by Mily Ron) and the *AtPUM14* genomic sequence in the *pB-PUF14-Gate-GFP* Gateway® destination vector (provided by Mihaela Marton), behind its native promoter. Binary vectors generated in this way were delivered into *Agrobacterium tumefaciens* strain *GV3101* (*pMP90, pSOUP*) (for *pGII-GW:NLS:3xeEFP::nost*) and *GV3101* (*pMP90RK*) (for *pB-PUF14-Gate-GFP*). Strains were cultured in YEP liquid medium and used to transform *Arabidopsis thaliana* and microscopic analysis performed as described above (2.12.2 and 2.13).

2.14.2 Silencing Pumilio Group III family members via amiRNA

For the generation of constructs containing artificial miRNA sequences against *At5g43090* and *At5g43110*, the Web application for the automated design of artificial microRNAs 'WMD3-Web MicroRNA Designer' was utilized to generate appropriate sequences with the use of the 'Designer tool' http://wmd3.weigelworld.org/cgi-bin/webapp.cgi?page=Designer;project=stdwmd). After the sequences have been selected,

corresponding primers (listed in the Appendix) were designed using the 'Oligo tool' (http://wmd3.weigelworld.org/cgi-bin/webapp.cgi?page=Oligo;project=stdwmd). The 'Oligo tool' uses the sequences defined by the 'Designer tool' as input to design primers containing the appropriate sequence, making it compatible with the vector template from which they will be amplified. The procedure was then continued following the cloning protocol available online (for download - http://wmd3.weigelworld.org/cgi-bin/webapp.cgi?page=Downloads). The amiRNA-containing precursor sequences were amplified from the pRS300 (Schwab *et al*., 2006) template plasmid by overlapping PCR. A first round of PCR amplifies fragments (a), (b) and (c), listed in the table in Figure 12, which were then run on a 2 % agarose gel, excised and purified using QIAquick Gel Extraction Kit (Qiagen). The purified products were then fused in a second PCR reaction (d). All PCR reactions were performed with the high-fidelity peqGOLD 'Pure' *Pfu*-DNA-Polymerase (peqlab).



	Forward oligo	Reverse oligo	Template
(a)	A	IV	pRS300
(b)	III	II	pRS300
(c)	I	B	pRS300
(d)	A	B	(a)+(b)+(c)

Figure 12. A scheme of the amiRNA precursor cloning. Oligonucleotide primers I to IV were used to replace miRNA and miRNA* regions (red) with artificial sequences (blue). Primers A and B anneal to the template plasmid sequence. The table illustrates the primer combinations for each reaction, (a), (b) and (c) in the first round of PCR, which are then combined, serving as a template in (d). Illustration adapted from http://wmd3.weigelworld.org/downloads/Cloning_of_artificial_microRNAs.pdf

After running the PCR product (d) on a 1% agarose gel, bands were cut out and purified with QIAquick Gel Extraction Kit (Qiagen). An additional PCR step was introduced, amplifying only the fragment containing the hairpin from the template (d) produced in the last PCR reaction in order to maintain compatibility with the Gateway® cloning system (primers are listed in the Appendix). The additional fragments were also purified after electrophoresis as described above (see 2.9). Fragments were then cloned into the pENTR/D-

TOPO (Invitrogen) vector (as described in 2.14.1) and fragments transferred into the binary Gateway® destination vector *pB-FM1-Gate-GFP* (provided by Manfred Gahrtz). *Agrobacterium tumefaciens* transformation, plant infection and selection of transgenic seedlings were performed as described (see 2.12.2).

2.15 Genomic DNA isolation from plant material

A medium-sized young leaf was collected and quick-frozen in liquid N_2 in a 1.5 ml Eppendorf tube. Plant tissue was ground into a fine powder while still frozen, and 400 µl of Extraction buffer (200 mM Tris/HCl pH=7.5 – 8, 250 mM NaCl, 25 mM EDTA, 0,5 % SDS) was added to the samples, which were then vortexed for ~ 5 sec. Following centrifugation (13,000 rpm, for 1 min at room temperature) 300 µl of supernatant was transferred to a new Eppendorf tube and mixed with 300 µl of ice-cold isopropanol by vortexing for 2 min. Samples were then centrifuged (13,000 rpm for 5 min at room tempertature) and the supernatant carefully removed. Pellets were left to dry briefly and then dissolved in 100 µl TE buffer (pH=8) and stored at 4 °C. Subsequent PCR reactions were performed using 1 µl DNA as template.

2.16 T-DNA insertion lines – genotyping and crossing

Arabidopsis thaliana T-DNA insertion lines (Alonso *et al*., 2003) were selected online (http://signal.salk.edu/cgi-bin/tdnaexpress) and ordered from the Nottingham Arabidopsis Stock Center (NASC). Appropriate primers for genotyping T-DNA lines were designed using the T-DNA Primer Design Tool provided online (http://signal.salk.edu/tdnaprimers.2.html. After performing PCR reactions, genotypes were determined as described in Figure 13. Homozygous lines were then crossed by cross-pollination of pistils 2-3 days after emasculation of stage 10 – 11 flowers (Smyth *et al*., 1990).

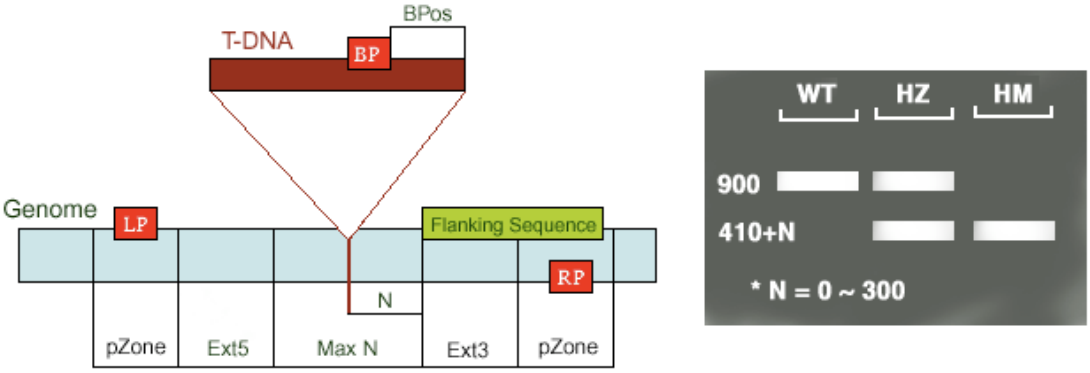


Figure 13. Genotyping T-DNA insertion lines. Three different primers used for genotyping are marked with LP (left genomic primer), BP (T-DNA insertion border primer) and RP (right genomic primer), as illustrated. Wild-type plants amplify the fragment LP – RP, heterozygous plants amplify fragment LP – BP or BP – RP in addition to the wild-type fragment and plants homozygous for the insertion amplify fragments LP – BP and/or BP – RP only. Corresponding DNA bands after gel-electrophoresis of the PCR products are shown on the right. Scheme taken from http://signal.salk.edu/tdnaprimers.2.html. N = difference of the actual insertion site and the flanking sequence position, usually 0 - 300 bases; MaxN = maximum difference of the actual insertion site and the sequence, default 300 bps; pZone = Regions used to pick up primers, default 100 bps; Ext5, Ext3 = regions between the MaxN to pZone, reserved for not picking up primers; BPos = the distance from BP to the insertion site.

3 RESULTS

3.1 Microarray analysis of single cells of the female gametophyte

3.1.1 Isolation of single female gametophyte cells and verification of cell identity

Techniques for single cell isolation from the female gametophyte (FG) of cereals such as wheat and maize have been established for some time (Kranz *et al*., 1993; Kumlehn *et al*., 1998). Up until recently (Ikeda *et al*., 2011) it was not possible to successfully establish the method for *Arabidopsis thaliana* (Figure 14), due to size restrictions, as single cells of the *Arabidopsis* female gametophyte only measure between 15 and 30 μm, and are thus significantly smaller than the FG cells in cereals, measuring between 70 and 200 μm in maize. Another challenge of the procedure, apart from manually handling sensitive plant tissue of that size, has proven to be recognizing individual cell types once in solution. To address that problem, homozygous, transgenic *Arabidopsis* marker-lines were used, labelling individual cells of the FG (the lab-generated *Ec1p:TaMAB2-GFP* line (provided by Stefanie Sprunck, expressing a wheat gene *TaMAB2* (EU360467) under the egg cell-specific *Ec1-1* promoter (Ingouff *et al*., 2009), and the *MYB98-GFP* line (Kasahara *et al*., 2004), selective for the synergid), as shown in Figure 14. This labelling has served as an aid in recognizing and distinguishing the particular features of each cell, as described in 2.6.

To ensure sample purity, verification of the cell identity was done by standard PCR performed on cDNA obtained from 15 pooled single cells of each type, using primers for cell-selective genes *Ec1-2a* (Sprunck *et al*., unpublished) for the egg cell, as well as *DD31* and *DD65* (Steffen *et al*., 2007) for the synergid and central cells, respectively. A DIG Southern-blot analysis was performed after gel-electrophoresis, with signals observed only in the corresponding cell types, confirming sample purity (Figure 15).

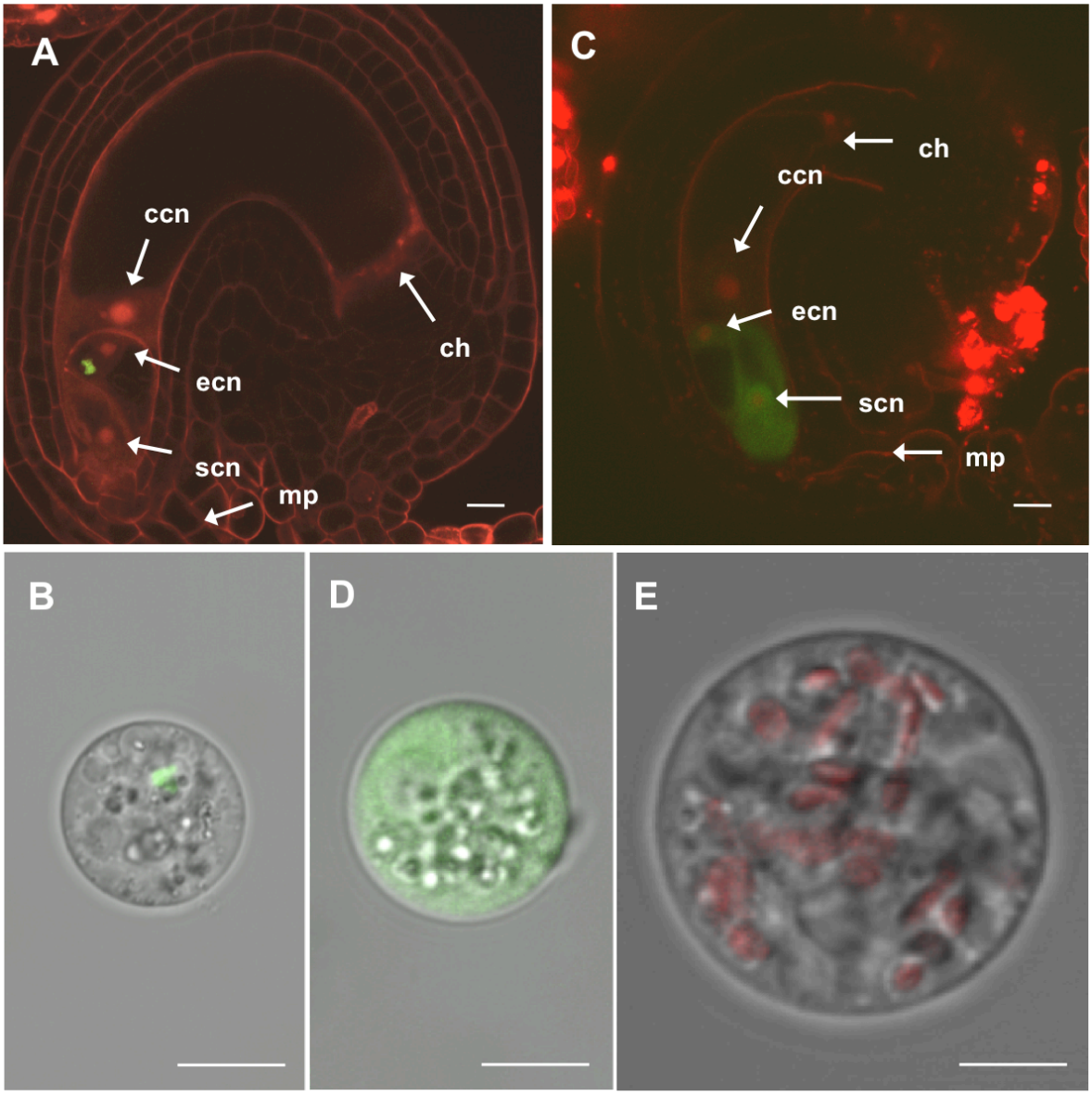


Figure 14. Single cell isolation from the female gametophyte of *Arabidopsis* facilitated by various marker-lines. (**A)** Ovule from the *Ec1p:TaMAB2-GFP* marker-line with the egg cell-specific promoter *Ec1-1*, showing expression of the wheat *TaMAB2* gene fused to *GFP*. (**B**) A single egg cell isolated from the corresponding marker-line. (**C**) The *MYB98p-GFP* marker line with the synergid-selective promoter *MYB98* fused to *GFP*. (**D**) A single synergid cell isolated from the corresponding marker-line. (**E**) A single wild-type central cell. Chloroplast auto-fluorescence is visible in red. Ovules in (A) and (C) were stained with propidium-iodide. Egg cell nucleus (**ecn**), central cell nucleus (**ccn**) and synergid nucleus (**scn**), micropyle (**mp**), chalazal region (**ch**). Size bars are 10 µm.

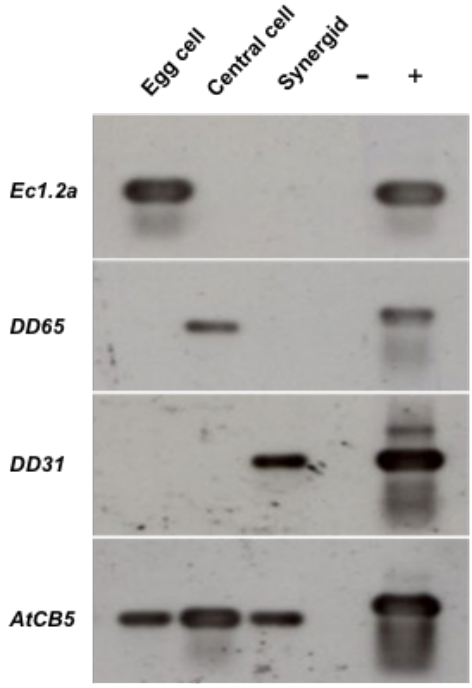


Figure 15. Southern blot analysis of cell type-selective RT-PCR. Quality and specificity of isolated cells and respective cDNAs was verified by PCR analysis using cDNA from 15-cell batches of each cell type, as indicated, using primers for cell-selective genes *Ec1-2a*, *DD65*, and *DD31*. Gels were blotted after electrophoresis and hybridized with respective radiolabeled probes. Columns represent PCR reactions with cDNA of each cell type and rows show individual primer-pairs used for the PCR reaction. *AtCB5* gene encoding the constitutively expressed *cytochrome b5* was used as an internal standard.

3.1.2 Collection and quality/quantity control of starting material

A sufficient amount of the required single-cell material needed to be collected for each of the cell types subsequently used for microarray hybridization. Because the amount of mRNA extracted from a manageable number of isolated cells is significantly below the detection limit and would be several orders of magnitude away from sufficient, a double step of RNA amplification was necessary and performed as described for three biological replicates of each cell type used. For array hybridization, 50 single cells were collected for each of the biological replicates, with the exception of one egg cell sample, where 30 cells were used. The amounts of cRNA obtained, and some standard quality control measurements in the procedure for each of the samples are shown in Table 1.

Table 1. Amount of cRNA hybridized to the Affymetrix ATH1 GeneChip® (values obtained from KFB). The quantity and quality of cRNA are displayed for the three biological replicates performed for each sample. The A_{260}/A_{280} ratios for all samples comply to that of pure RNA, which is ~ 2.0.

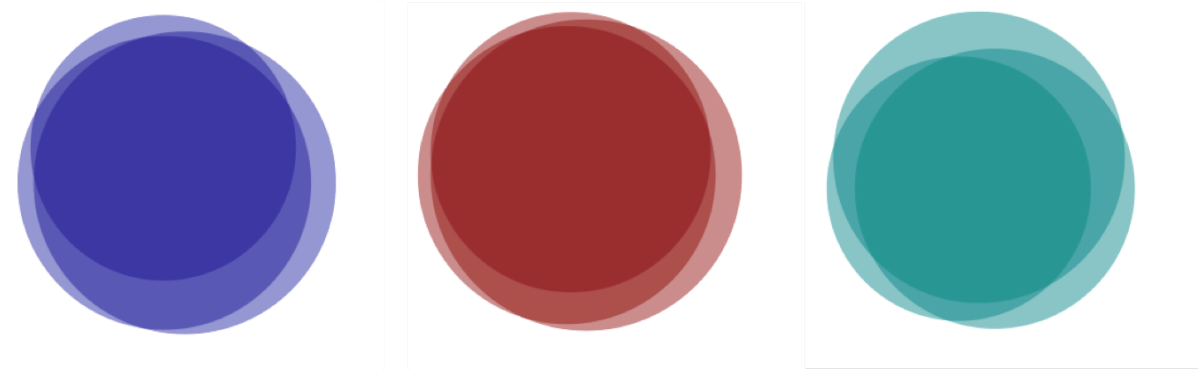
Sample	Name	Ratio (A_{260}/A_{280})	Amount (µg)
Egg cell	E1	2.06	3.33
Egg cell	E2	2.06	5.40
Egg cell	E3	2.14	15.72
Central cell	C1	2.11	6.48
Central cell	C2	2.10	14.56
Central cell	C3	2.10	20.94
Synergid	S1	2.05	8.87
Synergid	S2	2.12	7.46
Synergid	S3	2.08	6.02

3.1.3 Preliminary data analyses

3.1.3.1 Reproducibility of results and selection of an algorithm for generating Present calls

Following microarray hybridization, preliminary data analysis showed roughly half of the 22,392 genes represented on the chip being expressed in each cell type (13,572 in the egg cell, 13,690 in the central cell and 11,281 in the synergid cell). To examine the reproducibility of the expression profile among different replicates within samples, Venn diagrams were generated (Figure 16 A) showing largely overlapping surface areas representing genes called Present in all three biological replicates of each sample, indicating a high degree of reproducibility among samples. The corresponding values between individual replicate overlaps are shown in tables below each Venn diagram (Figure 16 B). A default Affymetrix algorithm MAS5 was used in generating the Present calls, as described in more detail in 2.11, and genes called Present in at least two of three replicates were taken into consideration for further analyses, being referred to as Present from now on unless otherwise specified.

A



B

E1	11365
E2	13907
E3	14766
E1/E2	10460
E1/E3	10418
E2/E3	13049
E1/E2/E3	10065

C1	12446
C2	14080
C3	15377
C1/C2	11804
C1/C3	11983
C2/C3	13441
C1/C2/C3	11641

S1	12712
S2	10446
S3	11728
S1/S2	9475
S1/S3	10283
S2/S3	9180
S1/S2/S3	8728

Figure 16. Overlap of genes called Present in the three biological replicates of single-cell samples of the FG. (**A**) Venn diagrams depict three biological replicates of egg cell (dark blue), central cell (red) and synergid cell (turquoise). (**B**) A table showing the number of genes called Present in each of the three biological replicates of a sample, as well as the number of genes shared by two, and all three replicates can be found below each diagram.

The quantity of starting material and the need to introduce amplification steps raise a concern about the 3'/5' ratios of the cRNA hybridized, as small input amounts of cRNA, combined with linear amplification steps tend to skew the signals towards the 3' end of the amplified mRNA. This can lead to a loss of 5' signal and specific probes are placed at the 5' and 3' ends as well as the middle region of internal control genes, serving as indicators of RNA integrity. As a guideline, the ratio of signal intensity should not exceed ~ 3 and was found to be within recommended values. Since the Affymetrix system is not designed for samples obtained by amplification of initial material, the MAS5 algorithm, generating Present calls isn't optimized for such cases. Another recent study (Wuest *et al.*, 2010)

illustrates an example of this problem, and the authors of the work have devised a new algorithm to address the issue. The same algorithm, called PANP (Present-Absent calls with Negative Probe sets) was also applied to the data generated in this study (performed by Jörg Becker), in addition to MAS5 to test if the same would apply. The comparison was then illustrated by Venn diagrams generated and compared between both algorithms (Figure 17). The comparisons show a higher number of Present calls generated with MAS5 than with PANP and don't display a significant change in reproducibility when overlaps of biological replicates of each sample are compared (not shown). Therefore, all data further analyzed in this study were generated following the application of the Affymetrix MAS5 algorithm, and PANP was not applied.

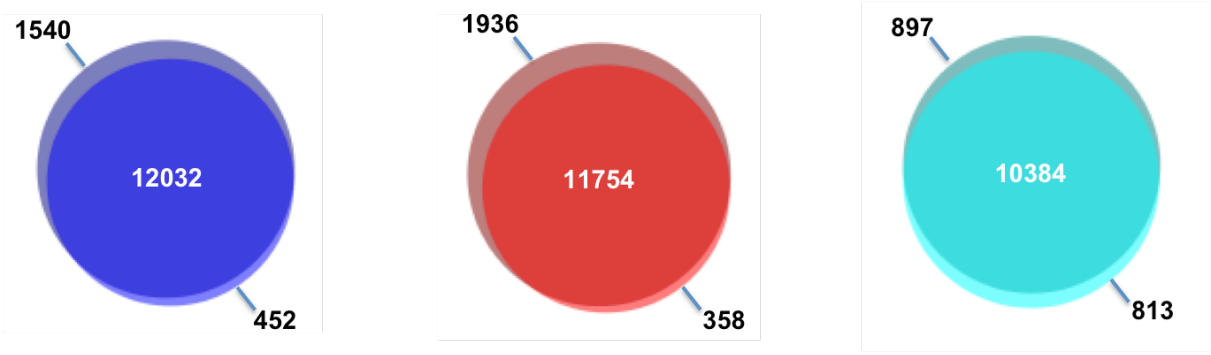


Figure 17. Comparison of Present calls generated by MAS5 and PANP algorithms. Darker shade circles represent Present calls generated using MAS5, the brighter shade overlaid circles represent the same generated using the PANP algorithm. Present calls are grouped as Present in two out of the three biological replicates in both cases. Numbers in the center of the circles represent the genes called Present using both algorithms, numbers in the upper left corner are the additional genes called Present using MAS5 and numbers in the lower right corner represent additional Present calls generated with the PANP algorithm. Dark blue diagram represents the egg cell sample, red represents the central cell, turquise represents the synergid, as before.

The data generated in this study was also compared to previously published FG genes listed in Table 2. Present calls generated by both algorithms were individually compared to a number of genes known to be expressed in the FG. Their reported expression profiles matched the MAS5 Present calls more closely than PANP Present calls (comparison not shown), further suggesting that the MAS5 algorithm outperforms PANP in this study, producing Present calls which are in partial or complete agreement with

previously published data (algorithms applied to the dataset by Jörg Becker with the help of Samuel Wuest).

Table 2. Genes previously published as selectively expressed in the FG.

The first five columns list AGI (Arabidopsis Genome Initiative) identification numbers along with the gene names and cell types where the genes are predominantly expressed (methods used and references are indicated). The last six columns list expression values for the single cells of the FG as obtained by this microarray study, along with the corresponding present (P) or absent (A) calls. EC = egg cell; CC = central cell; SC = synergid, ES = embryo sac; Pr::GFP = promoter-GFP fusion; Pr::GUS = promoter-GUS fusion; qRTPCR = quantitative RT PCR; ISH = *in situ* hybridization; ET = enhancer-trap.

AGI	Gene	Expression as described in mature Embryo sac (ES)	Method	Reference	Central cell (C)	C call	Egg cell (E)	E call	Synergid (S)	S call
At3g24220	9-Cis-Epoxycarotenoid Dioxygenase (NCED6)	SC, EC, (CC)	Pr::GFP, Pr::GUS	Lefebvre *et al*., 2006	271	P	81	A	244	P
At1g65360	AGL23	ES	Pr::GUS	Colombo *et al*., 2008	12	A	19	A	19	A
At1g01530	AGL28	ES	qRTPCR	Wang *et al*., 2010	704	P	14	A	214	P
At5g48670	AGL80	CC	Pr::GFP	Portereiko *et al.,* 2006	1453	P	51	A	444	P
At1g60280	ANAC23	SC, (EC)	Pr::GFP	Wang *et al*., 2010	9	A	182	P	17	A
At4g34940	ARO1	EC	Pr::GFP/GUS	Gebert *et al*., 2008	14	A	57	P	35	P
At5g06160	Ato	ES	Pr::GUS	Moll *et al*., 2008	128	P	173	P	107	P
At1g75250	ATRDL6	EC, CC, SC	Pr::GFP	Wang *et al*., 2010	18	A	55	A	137	P
At5g59440	AtTMPK kinase	ES	Pr::GUS	Ronceret *et al*., 2008 (b)	247	P	210	P	158	P
At3g06400	CHR11	ES	ISH	Huanca-Mamani *et al*., 2005	1177	P	1134	P	1372	P
At2g47430	CKI1	ES	Pr::GUS, ISH	Hejatko *et al*., 2003	393	P	29	A	159	P
At1g06220	Clotho/GFA1	ES	Pr::GUS	Moll *et al*., 2008	1300	P	1694	P	1269	P
At1g80370	Cyclin A2;4, Endocycle Regulation	ES	Pr::GUS	Johnston *et al*., 2007	28	A	270	P	62	A

At1g52970	DD11	ES	Pr::GFP	Steffen *et al*., 2007	561	P	310	P	13374	P
At2g21655	DD12	ES	Pr::GFP	Steffen *et al*., 2007	447	P	244	A	11994	P
At5g34885	DD17, DUF gene	ES	Pr::GFP	Steffen *et al*., 2007/Jones-Rhoades *et al*., 2007	502	P	418	P	9817	P
At1g45190	DD18	ES	Pr::GFP	Steffen *et al*., 2007	1107	P	560	P	13709	P
At2g06090	DD19	SC, CC	Pr::GFP, RT-PCR	Steffen *et al*., 2007 / Jones-Rhoades *et al*., 2007	6783	P	36	P	3727	P
At5g43510	DD2	SC	Pr::GFP	Steffen *et al*., 2007	1985	P	810	P	7438	P
At5g38330	DD22	CC	Pr::GFP, RT-PCR	Steffen *et al*., 2007/Jones-Rhoades *et al*., 2007	8547	P	44	A	4782	P
At3g04540	DD25	CC	Pr::GFP, RT-PCR	Steffen *et al*., 2007/Jones-Rhoades *et al*., 2007	8392	P	67	A	3626	P
At3g05460	DD27	SC, CC	Pr::GFP	Steffen *et al*., 2007	6530	P	98	A	5318	P
At3g46840	DD28	SC, CC	Pr::GFP	Steffen *et al*., 2007	4237	P	1301	P	9573	P
At3g56610	DD3	ES	Pr::GFP	Steffen *et al*., 2007	762	P	391	A	10984	P
At1g47470	DD31	SC, EC?	Pr::GFP	Steffen *et al*., 2007	570	P	208	P	11414	P
At3g17080	DD32; SI-protein	ES	Pr::GFP, RT-PCR, ISH	Steffen *et al*., 2007/Jones-Rhoades *et al*., 2007/Wuest *et al.*, 2010	50	A	31	A	4601	P
At2g20070	DD33	ES	Pr::GFP; Pr::GUS	Steffen *et al*., 2007/Yu *et al*., 2007	811	P	34	A	1621	P
At4g07515	DD34	ES	Pr::GFP	Steffen *et al*., 2007	1718	P	660	A	10168	P
At5g12380	DD35	SC	Pr::GFP	Steffen *et al*., 2007	93	P	73	A	8038	P
At3g24510	DD36	CC	Pr::GFP	Steffen *et al*., 2007	13896	P	323	P	8671	P
At4g20050	DD39	SC, EC	Pr::GFP	Steffen *et al*., 2007	3932	P	157	P	10185	P
At5g42955	DD4	ES	Pr::GFP	Steffen *et al*., 2007	777	P	296	P	7516	P

At1g73010	DD40	NE	Pr::GFP	Steffen *et al*., 2007	1680	P	41	A	644	P
At2g02515	DD41	CC	Pr::GFP	Steffen *et al*., 2007	5622	P	43	A	3024	P
At2g20660	DD42	SC, EC, CC	Pr::GFP, RT-PCR	Steffen *et al*., 2007/Jones-Rhoades *et al*., 2007	228	P	48	A	5322	P
At2g21740	DD45	EC	Pr::GFP, RT-PCR	Steffen *et al*., 2007/Jones-Rhoades *et al*., 2007	23	A	9542	P	2486	P
At1g22015	DD46	SC, CC	Pr::GFP	Steffen *et al*., 2007	1675	P	86	P	6324	P
At4g18770	DD53, MYB98	SC, (EC), (CC)	Pr::GFP, Pr::GUS	Steffen *et al*., 2007/Kasahara *et al*., 2005	92	A	70	A	4510	P
At3g28740	DD54	NE	Pr::GFP	Steffen *et al*., 2007	330	P	12	A	14	A
At4g30590	DD56; Plastocyanin	ES	Pr::GFP, ISH, Pr::GUS	Steffen *et al*., 2007/Johnston *et al*., 2007/ Yu *et al*., 2005	3532	P	1469	P	6442	P
At3g10890	DD65	CC	Pr::GFP	Steffen *et al*., 2007	9029	P	51	P	4377	P
At1g60985	DD66	CC	Pr::GFP	Steffen *et al*., 2007	7156	P	90	A	3101	P
At5g11940	DD67	ES	Pr::GFP	Steffen *et al*., 2007	160	P	192	A	9833	P
At2g20595	DD7; GA1-expressed protein	CC	Pr::GFP; RT-PCR, ISH	Steffen *et al*., 2007/Jones-Rhoades *et al*., 2007/Wuest *et al*., 2010	9984	P	37	A	5745	P
At5g12060	DD73	SC, CC	Pr::GFP	Steffen *et al*., 2007	2439	P	24	A	1896	P
At5g52975	DD8	ES	Pr::GFP; RT-PCR, ISH	Steffen *et al*., 2007/Jones-Rhoades *et al*., 2007/Wuest *et al*., 2010	413	P	124	P	9911	P
At1g26795	DD9; Self-Incompatibility Protein-Related	CC; ES	Pr::GFP; Pr::GUS	Steffen *et al*., 2007/Yu *et al*., 2005	5973	P	110	A	4447	P
At5g04560	Demeter	SC(early), EC? CC	Pr::GUS Pr::GFP	Choi *et al*., 2002	48	P	21	A	43	A
At1g02580	E(Z) Homologue, MEDEA	EC?, CC	Pr::GUS, ISH	Vielle-Calzada *et al*., 1999/Luo *et al*., 1999	1409	P	36	A	474	P

At2g33710	ERF/AP2 transcription factor family protein	EC, SC	Pr::GFP	Wang *et al*., 2010	112	P	744	P	117	P
At1g57760	Expressed protein	ES	RT-PCR	Jones-Rhoades *et al*., 2007	66	P	115	P	127	P
At3g51550	Feronia	SC	Pr::GFP, ISH, Pr::GUS	Escobar-Restrepo *et al*., 2007	192	P	760	P	1135	P
At3g20740	Fertilization Independent Endosperm	CC	Pr::GUS	Luo *et al*., 1999	600	P	626	P	779	P
At2g35670	Fertilization Independent Seed 2 (FIS2)	CC	Pr::GUS	Luo *et al*., 1999	1867	P	14	A	1129	P
At4g01970	Galactosyltransferase	ES	ET	Johnston *et al*., 2007	85	P	1664	P	215	P
At5g45910	GDSL-motif lipase/hydrolase-like protein	ES	RT-PCR	Jones-Rhoades *et al*., 2007	2560	P	1773	P	18793	P
At4g25530	Homeodomain Protein, FWA	CC	Pr::GFP	Kinoshita *et al*., 2004	3136	P	10	A	702	P
At5g54070	HSFA9	ES	qRT-PCR	Wang *et al*., 2007	98	P	68	P	24	A
At3g57840	hypothetical protein	ES	RT-PCR	Jones-Rhoades *et al*., 2007	86	P	63	A	5559	P
At2g41500	Lachesis	(SC), EC, CC	Pr::GUS	Gross-Hardt *et al*., 2007	968	P	1198	P	694	P
At1g21970	LEC1	ES	qRT-PCR	Wang *et al*., 2010	64	A	850	P	419	P
At3g57650	LPAT2	ES	Pr::GUS	Kim *et al*., 2005	4336	P	1836	P	1850	P
At4g02060	MADS Box Protein, PROLIFERA	ES	ET, ISH	Springer *et al*., 2000	1075	P	122	P	346	P
At5g27610	MYB family transcription factor	ES	qRT-PCR	Wang *et al*., 2010	250	P	674	P	407	P
At5g58850	MYB119	ES	qRT-PCR, Pr::GFP, Pr::GUS	Wang *et al*., 2010/Wuest *et al*., 2010	246	P	437	P	89	A
At5g11050	Myb64	ES	qRT-PCR	Wang *et al*., 2010	1534	P	2539	P	822	P
At1g56650	MYB75	ES	qRT-PCR	Wang *et al*., 2010	41	A	20	A	45	A
At1g66390	MYB90	ES	RT-PCR	Jones-Rhoades *et al*., 2007	33	A	29	A	45	A
At5g62470	MYB96 transcription factor-like protein	SC, EC	Pr::GUS	Wuest *et al*., 2010	10	A	22	A	23	A
At5g13690	N-acetyl-glucosaminidase (NAGLU)	ES	qRT-PCR	Wang *et al*., 2010	603	P	253	P	270	P
At5g61890	N.A. (AP2-EREBP Gene Family)	ES	qRT-PCR	Wang *et al*., 2010	52	A	85	A	32	A

At4g38070	N.A. (bHLH Gene Family)	ES	ISH, Pr::GUS	Ronceret *et al*., 2008 (a)	46	P	20	A	16	A
At5g40260	Nodulin MtN3 Protein	ES	ISH, Pr::GUS	Johnston *et al*., 2007/ Yu *et al*., 2005	3515	P	2474	P	3905	P
At5g48650	NTF2	ES	Pr::GUS	Wuest *et al*., 2010	157	P	1992	P	551	P
At2g37560	Origin recognition complex subunit 2 (ORC2)	SC, CC	ISH	Collinge et al., 2004.	517	P	267	P	762	P
At3g19350	PABP	EC? CC	Pr::GFP	Tiwari *et al*., 2008	25	P	16	A	21	A
At4g34110	Polyadenylate-Binding Protein 2 (PAB2)	ES	Pr::GUS	Palanivelu *et al*., 2000	5254	P	3765	P	5176	P
At1g71770	Polyadenylate-Binding Protein 5 (PAB5)	ES	Pr::GUS	Belostotsky and Meagher, 1996	2919	P	2537	P	2190	P
At1g78940	Protein Kinase, Cell Cycle Progression	ES	ISH	Johnston *et al*., 2007	139	P	77	P	99	P
At5g56510	Pumilio12	ES	Pr::GUS	Wuest *et al*., 2010	678	P	1611	P	1715	P
At1g28220	Purine Permease 3 (PUP3)	SC	Pr::GUS	Johnston *et al*., 2007	118	P	77	P	516	P
At4g17505	Putative protein	ES	RT-PCR	Jones-Rhoades *et al*., 2007	2520	P	26	A	731	P
At2g33130	RALFL18	ES	Pr::GUS	Wuest *et al*., 2010	5625	P	2298	P	4081	P
At5g60270	Receptor Kinase	EC	ISH	Johnston *et al*., 2007	109	P	58	P	35	A
At3g12280	Retinoblastoma	ES	ISH	Ebel *et al*., 2004	461	P	349	P	219	P
At3g61160	Shaggy-Like Kinase β (ASKβ)	EC	ISH	Wellmer *et al*., 2004	39	A	340	P	882	P
At2g47990	SWA1 (Slow walker 1)	ES	Pr::GUS ISH	Shi *et al*., 2005	1522	P	873	P	907	P
At5g50915	TCP Transcription Factor	EC, CC	ISH	Johnston *et al*., 2007	20	A	77	P	15	A
At5g16850	Telomerase	ES	ISH	Wuest *et al*., 2010	93	P	181	P	99	P
At3g61740	Trithorax Like Protein (ATX3)	ES	ISH	Johnston *et al*., 2007	81	P	71	P	83	P
At5g59340	WOX2	EC, CC	ISH	Haecker *et al*., 2004	54	A	33	P	37	A
At5g45980	WOX8	EC, CC	ISH, Pr::GFP	Haecker *et al*., 2003/Wang *et al*., 2010	98	P	407	P	410	P
At5g01860	zinc finger protein	ES	Pr::GFP	Wang *et al*., 2010	455	P	23	A	226	P

3.1.3.2 Grouping of samples and their expression profile overlaps

In order to create a context for differential expression of genes in the FG, particularly in reference to the male gametophyte/male gametes and a vegetative tissue, the FG samples were compared to those of sperm cell, pollen and seedling from a previous study (Borges *et al*., 2008). As described in more detail in 2.11, all six samples were normalized and modelled together, thus allowing the comparison. Array outliers detected in replicates of one tissue/cell type were not considered true array outliers, but treated as differentially/specifically expressed. In order to visualize global expression profiles of the samples, a principal component analysis (PCA) of all six samples was performed (Figure 18), followed by a four-way Venn diagram comparing the FG samples and sperm (Figure 19). PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible (the values are given in brackets, shown in Figure 18) and each succeeding component accounts for as much of the remaining variability as possible. The depiction of the samples in the plot as points separated along axes shows the degree of similarity of their global expression profiles, which is higher the closer the points appear to one another (Figure 18). The samples of the male and the female gametophyte clearly separate along the first principal component on either side of the plot, the FG cells grouping together on the right hand side. Furthermore, an additional separation can be observed between seedling and the female and male gametophytes along the second principal component as well as less pronounced differences among single-cell samples and between pollen and sperm cells themselves (Figure 18). A separation of the single FG cells along the 2nd principle component places the egg cell as the closest to the seedling, which is not surprising, considering it is it's direct descendant, followed by the central cell, having its genetic contribution limited to the endosperm, and ultimately the synergid, which doesn't genetically contribute to the next generation at all.

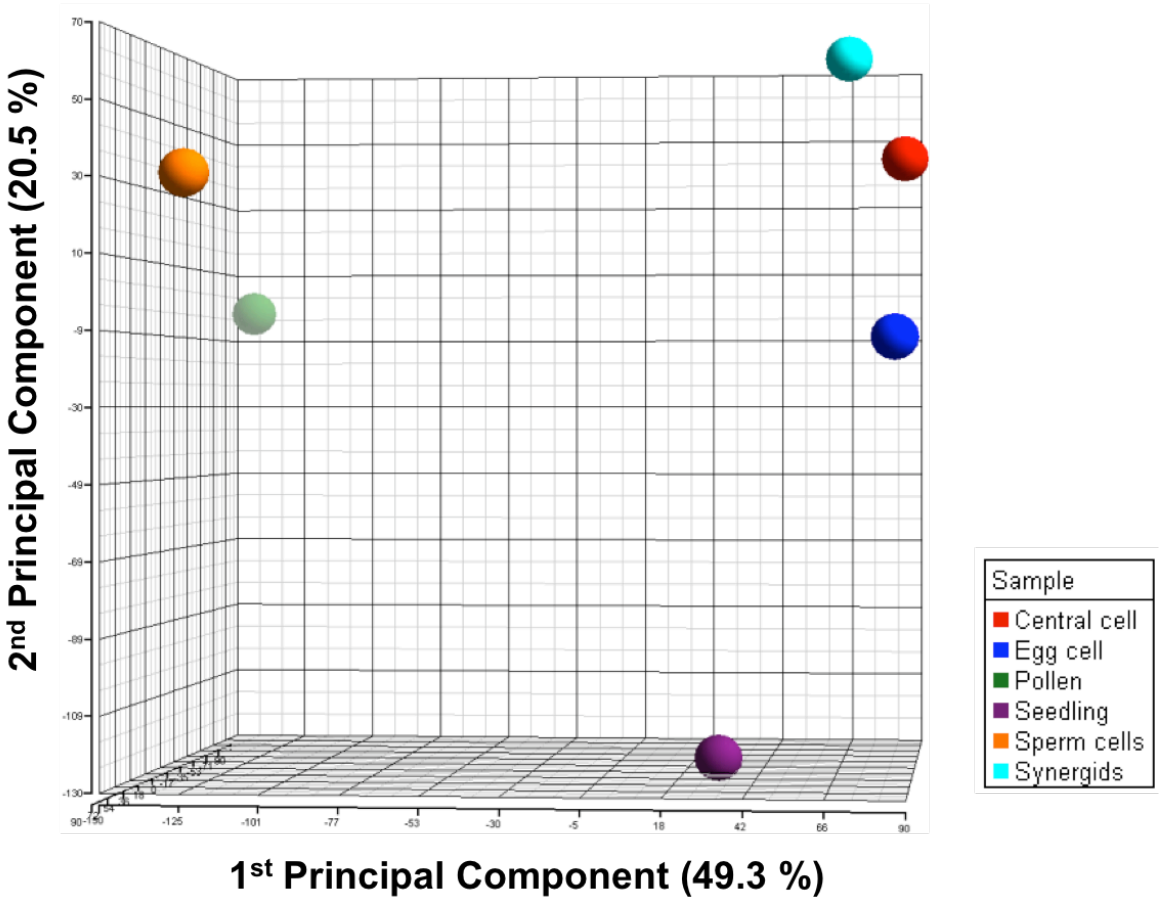


Figure 18. Principal Component Analysis (PCA) of FG cells together with pollen, sperm cells and seedling. A comparison is shown of the transcriptomes of egg cell, central cell and synergid with previously published transcriptomes of pollen, sperm cells and seedling. There is a visible separation of the female and male gametophytes along the 1st Principal Component (accounting for 49.3 % of the variability) as well as the sporophytic seedling sample separation from both of them along the 2nd Principal Component (accounting for 20.5 % of the variability).

A 4-way Venn diagram was generated using data obtained from the FG samples and previously published sperm cell samples (Borges *et al*., 2008), with surface areas and their overlaps approximately corresponding to the number of genes called Present in each cell type (Figure 19). Over 90 % of the genes expressed in each cell type represented is also present in at least one of the other cell types and all the gametes express close to 10 % genes selectively (taking only these four cell types into consideration). Only the synergid cell has a significantly lower number of selectively expressed genes (2 %) which might not

be surprising, considering it is the only of the four cell types not surviving beyond the point of fertilization.

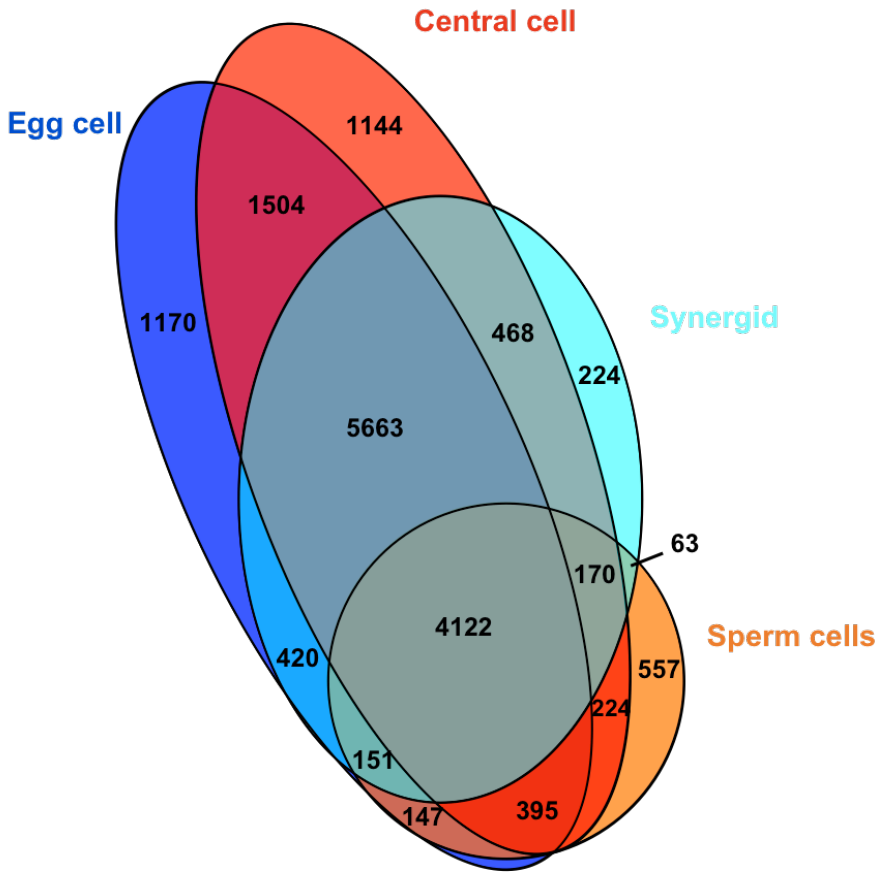


Figure 19. A 4-way Venn diagram of the number of genes called Present in single cells of FG and sperm. The surface areas of the elliptic shapes making the diagram shows the number of genes called Present in the egg cell (13572), central cell (13690), synergid cell (11281) and sperm cells (5829) and their approximate overlap.

3.1.4 Compared with the egg cell and synergid cells, central cell and sperm cells show a significant overlap in core cell cycle genes

Correct progression through the cell cycle during cellular division is critical for the formation of functional female and male gametophytes. Moreover, during fertilization the two female and two male gametes fuse, bringing their cellular components together. The gametes must therefore synchronize their cell cycle states in order to harmoniously initiate the developmental programs of the embryo and the endosperm (Berger *et al.*, 2008). Several mutants impaired in cell cycle affect both female and male gametophytes in

Arabidopsis (Springer *et al*., 2000; Pischke *et al*., 2002; Ebel *et al*., 2004; Iwakawa *et al*., 2006; Johnston *et al*., 2008; Brownfield *et al*., 2009a, b; Johnston *et al*., 2010). Therefore, analysis of genes involved in the process was included in this study in order to gain more insight into their expression profiles and cell cycle dynamics in the FG. In total, 61 genes were identified belonging to seven selected families of cell cycle regulators determining the core cell cycle (Vandepoele *et al*., 2002). A histogram was made taking these genes into consideration and comparing the correlation coefficient of the core cell cycle with the overall correlation of all genes on the array. This comparison shows a significantly higher correlation between the mentioned group of genes in sperm cells and the central cell, as well as in the egg cell and synergid samples, in comparison to the overall correlation of all genes present on the ATH1 GeneChip® between the same cell types (Figure 20) (histogram drawn by Jörg Becker).

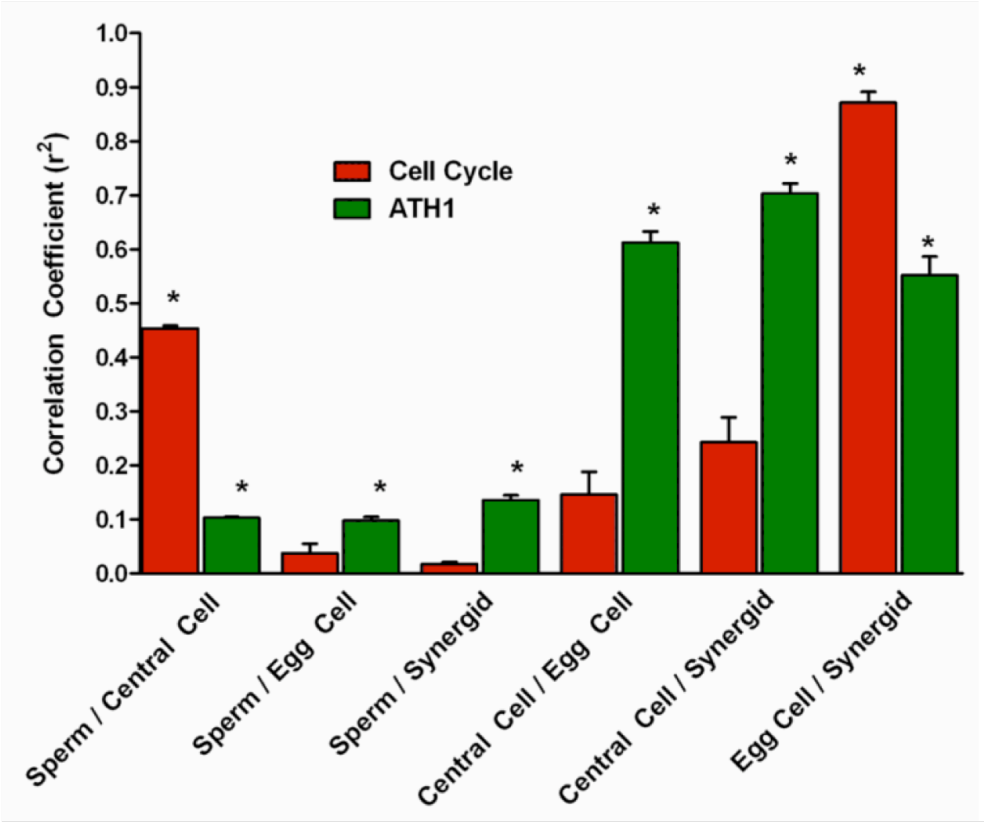


Figure 20. Correlation coefficient of core cell cycle genes in FG cells and sperm cells. The correlation coefficient of core cell cycle genes compared to all genes on the ATH1 GeneChip® in the FG cells and sperm cell samples shows a significant correlation in cell cycle between sperm and central cell as well as egg cell and synergid cell samples. The two female gametes show little overlap of core cell cycle genes compared to the overall gene overlap, indicating that they are in different stages of the cell cycle. The correlations with a significant p-value are marked with ' * '. (Histogram drawn by Jörg Becker using GraphPad Prism 5).

A series of PCRs using single cell cDNA was performed in order to further verify the array data, as well as try to determine cell cycle states of individual cells. The central cell and sperm cells, which exhibit a higher correlation in their cell cycle state, express a number of *cyclin* genes known to mark the G2/M transition, such as *CYCA1;1, CYCA2;1, CYCB1;1* and *CYCB1;3* (Menges *et al*., 2005) whereas the egg cell seems to differ in its cell cycle from them and might still be in G1 or G1/S transition (Menges *et al*., 2005) expressing genes that have their peak of expression in those stages, such as *CYCD3;3* and *CYCD4;1* (Figure 21). RT-PCR studies confirm data obtained for the FG cells through microarray analysis, thus serving as an independent confirmation of the array reliability. *CYCB2;2* represents an exception as it wasn't detected by the array in the egg cell but was detectable by RT-PCR. Another difference concerns the male side, where *CYCD3;3* was also not detected in sperm cells on the array, but produced a band in RT-PCR studies.

In addition to this, hierarchical clustering of core cell cycle genes (the same group shown in Figure 20) was performed using dChip software analysis (Liu *et al*., 2003) (Figure 22). The colour scale in the lower left corner of the figure shows expression level intensities as described. The central cell and sperm cell samples group together also in the heat map, as well as egg cell and synergid samples, pointing to more similarity between their cell cycle states (dendrogram on top of heat map).

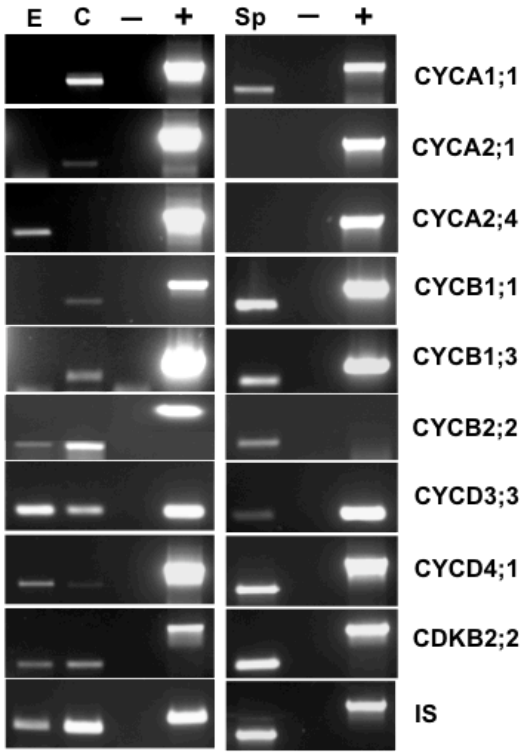


Figure 21. RT-PCR analyses of selected cell cycle genes in male and female gametes. The RT-PCR results largely overlap with the microarray findings and the cell cycle correlation between central cell and sperm cells is also evident, as most of the same genes are showing expression in those cells and only about half of them show expression also in the egg cell. The results also provide a further hint towards the cell cycle states of individual cells might be in, as most of the *cyclin* genes tested exhibit a peak of expression during G2/M transition, and are found predominantly in the central cell and sperm. The *AtCB5* gene was used as an internal standard (IS) for the egg cell (E) and central cell (C) and *Mgh3* for sperm cells (Sp).

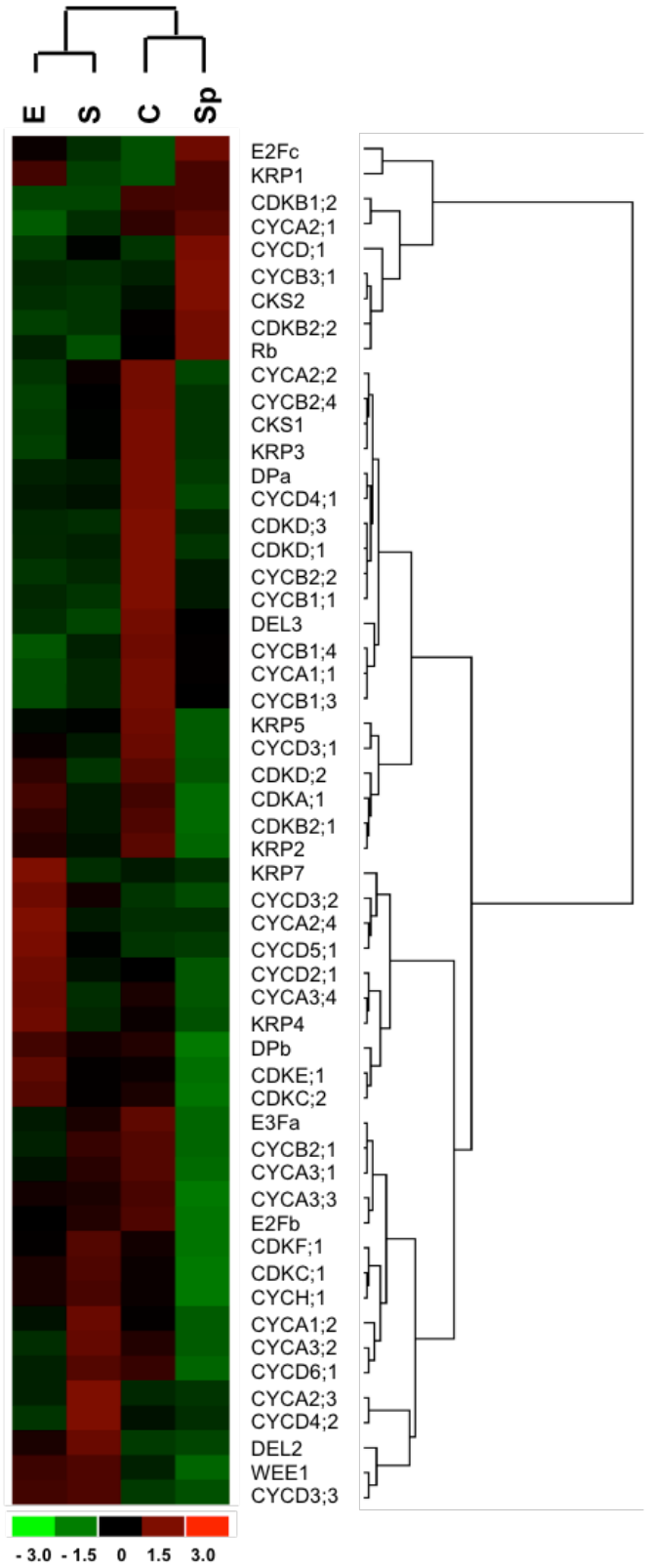


Figure 22. Hierarchical clustering of core cell cycle genes in the FG cells and sperm cells. The heat map shows the expression of 55 core cell cycle genes and the way they group within samples. A higher correlation between central cell (C) and sperm cells (Sp) is visible on top of the figure, shown by the dendogram, with the egg cell (E) and synergid cell (S) samples also grouping together. Red colour represents expression level

above mean expression of a gene across all samples, black colour represents mean expression and green represents expression level lower than the mean.

3.1.5 Small RNA pathways in the female gametophyte and sperm cells

Gene silencing has in the past decade emerged as one of the most important gene regulation mechanisms and has maintained its status as one of the most popular research topics in recent years. In the plant field however, not much was known about this process in the reproductive tissue of the plant, and until recently it was not known how dynamic, or present at all smallRNA pathways are in the FG (Martienssen, 2010). The main players in these processes are a family of proteins called Argonautes. Therefore, the expression of the ten members identified in the genome of *Arabidopsis thaliana* was examined during the microarray study (Table 3). Most members were found to be expressed in single cells of the FG, however, a few of the members had a more striking expression pattern. Three members in particular, *AGO5*, *AGO8* and *AGO9*, were strongly and almost exclusively expressed in the male and female gametophytes and *AGO8* seemed to be expressed exclusively in the egg cell. A weaker signal was also detected in the synergid cells, but the Present call couldn't be confirmed in single cells of the FG through RT-PCR experiments (Figure 23).

Table 3. Expression of *Arabidopsis Argonaute* gene family members in the FG cells according to the Affymetrix microarray data.

The expression values for all 10 *AGO* members in the FG and sperm cells are listed below. The corresponding Present (P) and Absent calls (A) can be seen in a column to the right of every sample.

		Central		Egg		Synergid		Sperm	
AGI	Gene	cell (C)	C call	cell (E)	E call	cell (S)	S call	cells (Sp)	Sp call
At1g48410	AGO1	1751	P	4317	P	1884	P	12	A
At1g31280	AGO2	315	P	938	P	434	P	21	A
At1g31290	AGO3	57	P	23	A	103	A	15	A
At2g27040	AGO4	185	P	659	P	500	P	25	A
At2g27880	**AGO5**	2217	P	4418	P	1775	P	383	P
At2g32940	AGO6	126	P	367	P	289	P	48	P
At1g69440	AGO7	35	A	65	P	96	P	21	A
At5g21030	**AGO8**	17	A	1027	P	191	P	23	A
At5g21150	**AGO9**	1745	P	5901	P	2545	P	927	P
At5g43810	AGO10	19	A	57	P	40	A	23	A

As a part of the microarray validation experiment, expression of individual members of the *Argonaute* family was investigated through RT-PCR performed on cDNA derived from single, isolated cells of the FG and sperm cells (Figure 23). The PCR results mostly agree with the microarray of the female gametes, a few genes however show minor discrepancies. These are *AGO2*, *AGO3*, *AGO6* (their expression in the central cell could not be confirmed by RT-PCR), *AGO7* (not detected in the central cell by the microarray, but showing RT-PCR signal) and *AGO8* (the expression detected in synergid cells by the array couldn't be reproduced by RT-PCR).

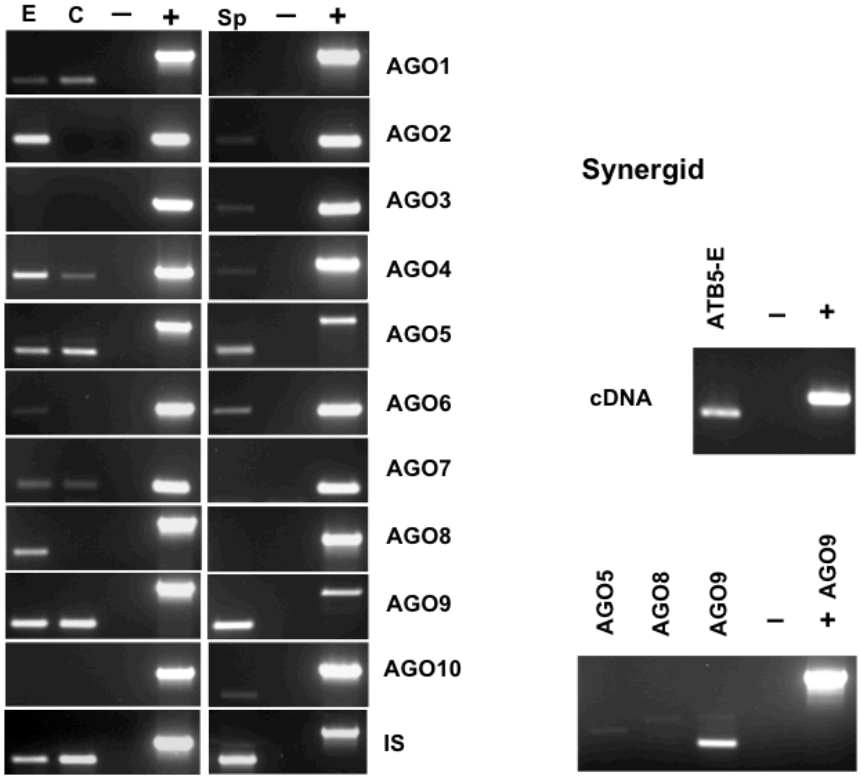


Figure 23. RT-PCR analysis of *Argonaute* genes in the gametes of *Arabidopsis*. A series of PCR reactions was performed on cDNA of all three gametes (rows on the left), examining the pattern of expression of *Argonaute* family members in the cell types indicated. An additional set of PCRs was performed on synergid cDNA for three of the most interesting members of the family (right), showing a test of cDNA quality using primers for *AtCB5* above the expression of the selected *AGO* genes in synergid cells. Of the three genes tested in synergid cells, the signal could only be observed for *AGO9*, confirming the *AGO8* expression to be restricted to the egg cell. Egg cell (E); central cell (C); sperm cells (Sp); positive genomic DNA control (+); negative control without RT reaction (-).

Expression of all *AGOs* is also depicted in a heat map, in which their relative expression levels were plotted together with 43 other genes involved in the small noncoding RNA pathways (Figure 24).

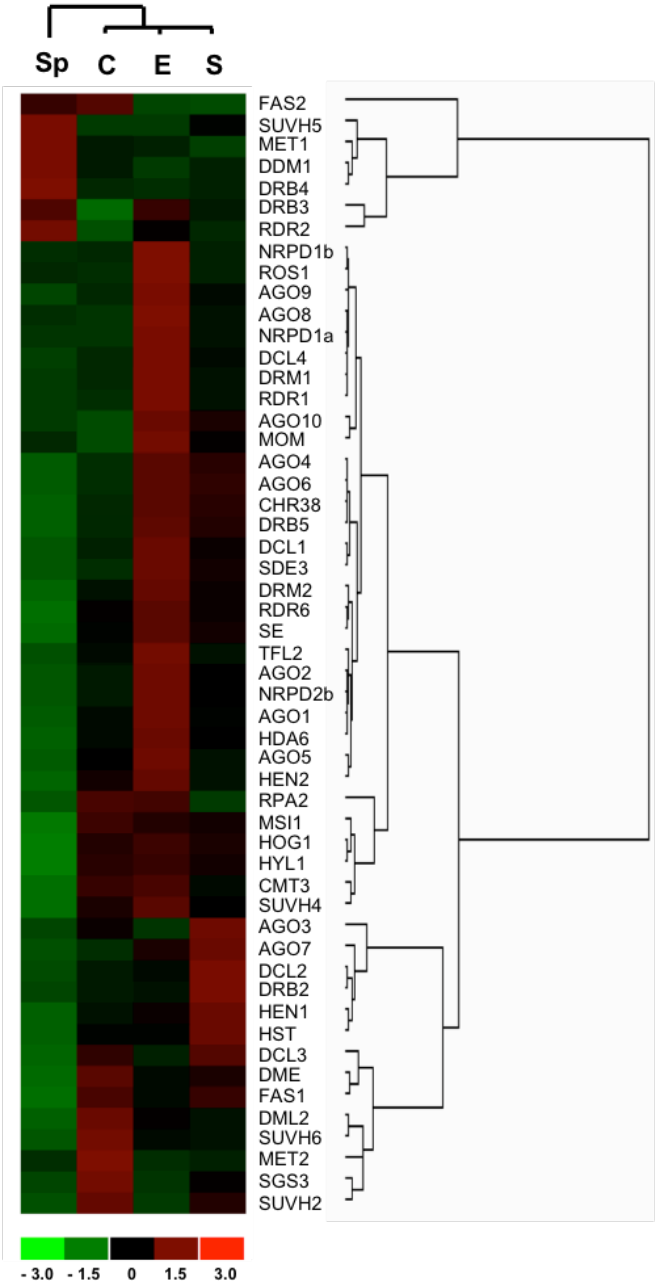


Figure 24. Hierarchical clustering of genes involved in smallRNA pathways, expressed in the FG cells and sperm cells. The samples are grouped together above the heat map according to their similarity in expression of genes involved in the pathway. FG cells group together showing more similarity within the FG than to sperm cells. The map also illustrates a dynamic activity of the silencing machinery in the individual cell types and different members more strongly expressed in particular cells, implying delicate pathway variations among them, with the egg cell displaying a prominent expression for a large number of members

(dark red). Red colour represents expression level above mean expression of a gene across all samples, black colour represents mean expression and green colour represents expression level lower than the mean.

The hierarchical clustering of the genes involved in smallRNA pathways show the expression 53 different members of the machinery. The dendrogram above the heat map groups the cells of the female gametophyte together, separated from the sperm cells, suggesting more similarity in the pathway dynamics within the FG. The red areas represent genes expressed above the mean level for all samples and suggest that smallRNA pathways play a significant role in gene regulation in the egg cell, since many of the genes examined appear to have an overall stronger expression there.

3.1.6 Auxin-related genes

Since auxin seems to play a major role in cell specification during FG development (Pagnussat *et al*., 2009), the expression of several families of auxin–related genes have been examined in the FG and sperm cells. This was done in an attempt to better understand auxin dynamics in the FG, especially in relation to its synthesis, transport and degradation. The genes examined include families of auxin influx and efflux carriers, such as the well-characterised *PIN* family, the *ARF* family of transcription factors, genes involved in *IAA* biosynthesis, as well as the *SAUR* (small auxin-up RNA) family of auxin-induced genes, the *GH3* gene family encoding a class of auxin-induced conjugating enzymes and genes related to TIR1, an F-box protein acting as an auxin receptor (Gälweiler *et al*., 1998; Dharmasiri *et al*., 2005; Guilfoyle and Hagen, 2007; Paponov *et al*., 2008;).

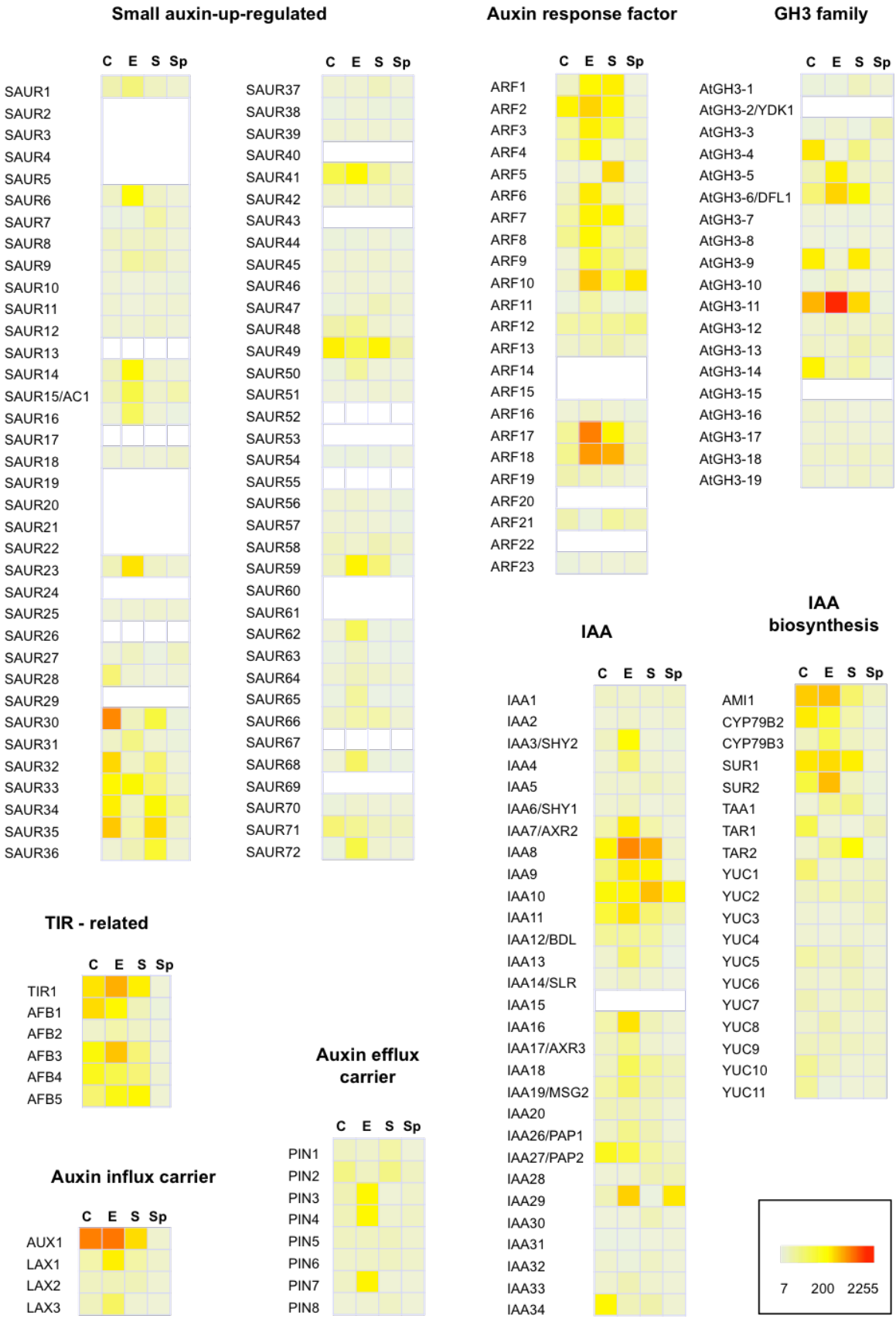

Small auxin-up-regulated
C E S Sp
SAUR1
SAUR2
SAUR3
SAUR4
SAUR5
SAUR6
SAUR7
SAUR8
SAUR9
SAUR10
SAUR11
SAUR12
SAUR13
SAUR14
SAUR15/AC1
SAUR16
SAUR17
SAUR18
SAUR19
SAUR20
SAUR21
SAUR22
SAUR23
SAUR24
SAUR25
SAUR26
SAUR27
SAUR28
SAUR29
SAUR30
SAUR31
SAUR32
SAUR33
SAUR34
SAUR35
SAUR36
C E S Sp
SAUR37
SAUR38
SAUR39
SAUR40
SAUR41
SAUR42
SAUR43
SAUR44
SAUR45
SAUR46
SAUR47
SAUR48
SAUR49
SAUR50
SAUR51
SAUR52
SAUR53
SAUR54
SAUR55
SAUR56
SAUR57
SAUR58
SAUR59
SAUR60
SAUR61
SAUR62
SAUR63
SAUR64
SAUR65
SAUR66
SAUR67
SAUR68
SAUR69
SAUR70
SAUR71
SAUR72
Auxin response factor
C E S Sp
ARF1
ARF2
ARF3
ARF4
ARF5
ARF6
ARF7
ARF8
ARF9
ARF10
ARF11
ARF12
ARF13
ARF14
ARF15
ARF16
ARF17
ARF18
ARF19
ARF20
ARF21
ARF22
ARF23
GH3 family
C E S Sp
AtGH3-1
AtGH3-2/YDK1
AtGH3-3
AtGH3-4
AtGH3-5
AtGH3-6/DFL1
AtGH3-7
AtGH3-8
AtGH3-9
AtGH3-10
AtGH3-11
AtGH3-12
AtGH3-13
AtGH3-14
AtGH3-15
AtGH3-16
AtGH3-17
AtGH3-18
AtGH3-19
IAA
C E S Sp
IAA1
IAA2
IAA3/SHY2
IAA4
IAA5
IAA6/SHY1
IAA7/AXR2
IAA8
IAA9
IAA10
IAA11
IAA12/BDL
IAA13
IAA14/SLR
IAA15
IAA16
IAA17/AXR3
IAA18
IAA19/MSG2
IAA20
IAA26/PAP1
IAA27/PAP2
IAA28
IAA29
IAA30
IAA31
IAA32
IAA33
IAA34
IAA biosynthesis
C E S Sp
AMI1
CYP79B2
CYP79B3
SUR1
SUR2
TAA1
TAR1
TAR2
YUC1
YUC2
YUC3
YUC4
YUC5
YUC6
YUC7
YUC8
YUC9
YUC10
YUC11
TIR - related
C E S Sp
TIR1
AFB1
AFB2
AFB3
AFB4
AFB5
Auxin efflux carrier
C E S Sp
PIN1
PIN2
PIN3
PIN4
PIN5
PIN6
PIN7
PIN8
Auxin influx carrier
C E S Sp
AUX1
LAX1
LAX2
LAX3
7 200 2255

Figure 25. Families of auxin-related genes and their expression in the FG and sperm cells. Several auxin-related families are expressed in the mature FG with some of the gene members having relatively high expression values. The overall most pronounced expression is in the egg cell. The blank or white lines within the colour grids represent genes within those families that are not represented on the ATH1 GeneChip®. The colour scale representing signal intensities is shown in the lower right corner.

Examining the expression of members of the above-mentioned families within the FG showed a weak or moderate expression of auxin-related genes globally, however certain members, namely *SAUR30, ARF17, ARF18, AtGH3-11, IAA8* and *AUX1*, display a significant increase of expression intensity. Also, looking at specific cell types, the egg cell seems to stand out among the other cell types, showing a pronounced dynamic, with the strongest overall expression of all the genes examined. The expression of individual genes from these families across the examined cell types is depicted in Figure 25, showing the relative intensity of expression in a colour scale.

3.1.7 Signaling

3.1.7.1 Receptor-like kinases (RLKs)

Signal perception through cell-surface proteins is a common feature among living organisms. One of the largest families in the *Arabidopsis* genome are a group of transmembrane kinases collectively named receptor-like kinases (RLKs). These proteins are reportedly involved in a diverse range of processes, including self-incompatibility, disease resistance, regulation of development, and hormone perception (Shiu and Bleecker, 2001b). More than 600 RLKs are present in the *Arabidopisis* genome and 573 of those are found on the ATH1 GeneChip® used in this study. These were identified and sorted into subfamilies according to Shiu and Bleecker (2001) and subfamilies were grouped together in individual images within the figure according to the number of genes in each family in order to provide a clearer visual overview (Figure 26). The FG expresses a high number of RLKs genes overall, and looking at individual cell types the egg cell and central cell are leading the way in these numbers compared with accessory cells (synergid cells) and sperm cells, especially in the DUF26, L-Lectin and S-Domain (SD) subfamilies (Figure 26).

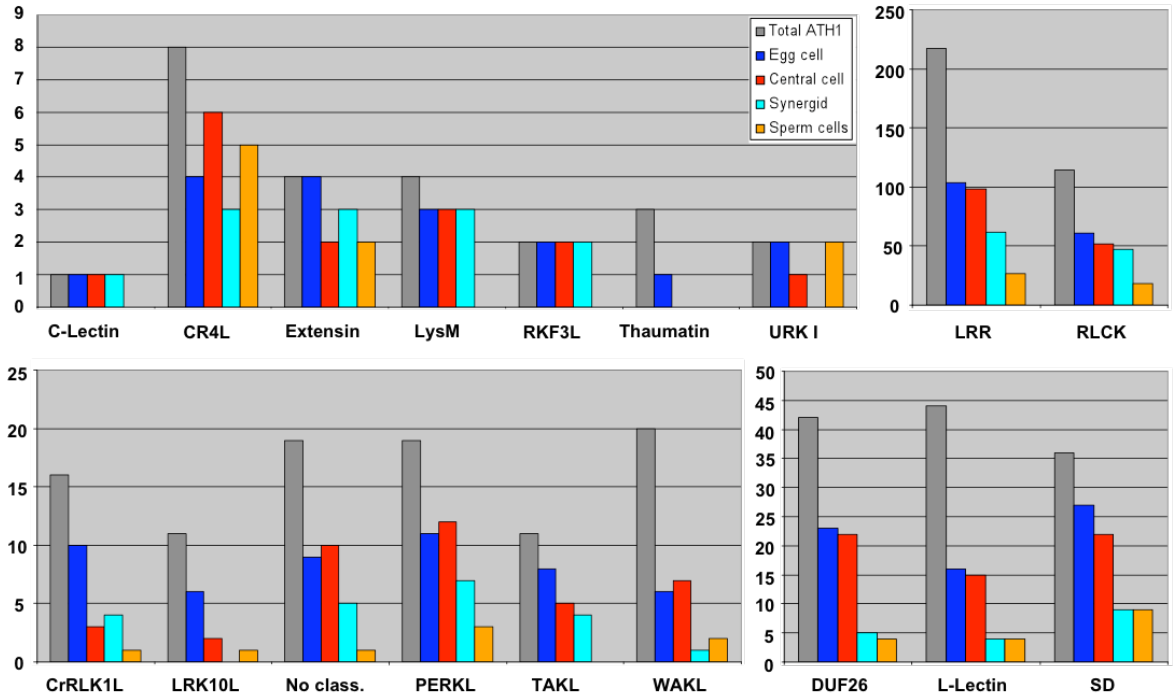


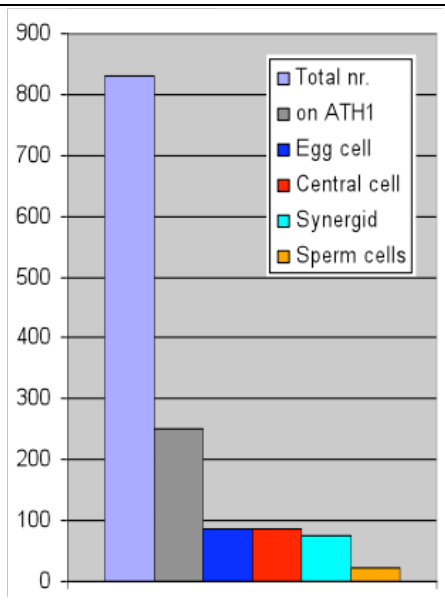
Figure 26. Expression of receptor-like kinase genes represented on the ATH1 GeneChip® in FG cells and sperm cells. Genes expressed on the chip were categorized into subfamilies according to Shiu and Bleecker (2001). Subfamilies are grouped together in individual images within the figure in relation to the number of genes they consist of.

A few members were also found to be exclusively expressed in one cell type and no other tissue in the plant (according to Genevestigator), these are three genes in the central cell, namely *At5g01560* (member of the L-Lectin subfamily), *At1g61390* (member of the S-Domain-1 subfamily) and *At5g41680* (member of the LRR III subfamily), one gene in the egg cell, namely *At3g24400* (a member of the PERKL subfamily) and one gene in the synergid cell, namely *At1g29730* (a member of the LRR VIII-2 subgroup).

3.1.7.2 Cysteine-rich peptides (CRPs)

Another group of signalling molecules are secreted small proteins and peptides, which are also important regulators of growth, development and reproduction. Because of their small size and extreme sequence divergence they have often been overlooked during genome annotations, remaining largely unidentified until recent years. Within this large, polymorphic and highly divergent group of proteins, cysteine-rich peptides (CRPs) are

particularly well represented among plants, however for the mentioned reason, they have been largely under-represented on the ATH1 GeneChip®, with only 252 of 825 members on the chip (Figure 27). The number and arrangement of cysteine residues in the primary sequence distinguishes each CRP class from the other. The members present on the chip were categorized in Homology Subgroups according to Silverstein *et al*. (2007).



Homology Subgroup	Nr. of Genes on ATH1 Gene Chip• (%)	Total number of genes in A.th.
Defensin/DEFL	35 (10,8)	323
RALF	18 (46,2)	39
Thionin	9 (12,7)	71
GASA/GAST/ Snakin	15 (100)	15
Root cap/LEA	6 (100)	6
Stig1	3 (50)	6
Pollen Ole e I	26 (74,3)	35
LTP/2S Albumin/ECA1	104 (37,7)	276
(MEG)/Ae1	5 (29,4)	17
Hevein	14 (87,5)	16
Kazal type inhibitor	2 (100)	2
Kunitz type inhibitor	6 (85,7)	7
Proteinase inhibitor II	1 (100)	1
Antimicrobial peptide MBP-1	1 (100)	1
Other	6 (60)	10

Figure 27. Cysteine-rich peptides and their representation on the *Arabidopsis* ATH1 Gene Chip•. The histogram on the left depicts a total of 825 CRPs identified so far in the *Arabidopsis* genome, and the number represented on the array as well as in each cell type indicated. Table on the right shows the representation of each Homology Subgroup (table made by Svenja Rademacher).

Among all the *CRP* genes found in the FG, there were 43 genes belonging to 27 *CRP* groups expressed only in one cell type of the FG (Table 4), with three of the genes expressed exclusively in these cells and in no other tissue or cell type (according to Genevestigator). The genes belong to six of the mentioned subgroups, namely DEFL,

RALF, Thionin, GASA/GAST/Snakin, Pollen Ole e 1 and LTP/2S Albumin/ECA1, which is the most abundantly represented with over half of the 43 genes listed belonging to it.

Table 4. A list of *CRP* genes selectively expressed in only one cell type within the FG.

Genes putatively expressed exclusively in a single cell type of the FG, and not in any other tissue or cell type (according to Genevestigator) are marked in **bold***. Colours are assigned to each cell type. Egg cell = blue, central cell = red, synergid cell = turquoise.

Homology Subgroup	Group ID	AGI	Homology Subgroup	Group ID	AGI
DEFL	CRP0000	At2g02120	LTP/2S Albumin/ECA1	CRP3860	At5g59320
DEFL	CRP0300	**At5g54220***	LTP/2S Albumin/ECA1	CRP4210	At5g55450
DEFL	CRP0700	At2g43520	LTP/2S Albumin/ECA1	CRP4240	At4g33550
DEFL	CRP0700	At2g43530	LTP/2S Albumin/ECA1	CRP4410	At3g07450
DEFL	CRP0700	At2g43535	LTP/2S Albumin/ECA1	CRP4540	At2g27130
RALF	CRP1820	At2g19040	LTP/2S Albumin/ECA1	CRP4660	At1g55260
RALF	CRP1855	At1g61566	LTP/2S Albumin/ECA1	CRP4670	At1g70250
Thionin	CRP2200	At1g66100	LTP/2S Albumin/ECA1	CRP4680	At5g13900
Thionin	CRP2200	At1g72260	LTP/2S Albumin/ECA1	CRP4690	At1g05450
GASA/GAST/Snakin	CRP2700	At1g22690	LTP/2S Albumin/ECA1	CRP4690	At3g53980
GASA/GAST/Snakin	CRP2700	At2g30810	LTP/2S Albumin/ECA1	CRP4750	At2g37870
GASA/GAST/Snakin	CRP2700	At3g02885	LTP/2S Albumin/ECA1	CRP4820	At1g62500
GASA/GAST/Snakin	CRP2700	At5g14920	LTP/2S Albumin/ECA1	CRP4820	At3g22120
Pollen Ole e I	CRP3300	At4g38770	LTP/2S Albumin/ECA1	CRP4820	At4g12490
Pollen Ole e I	CRP3350	At5g13140	LTP/2S Albumin/ECA1	CRP4820	At4g22470
Pollen Ole e I	CRP3350	At5g41050	LTP/2S Albumin/ECA1	CRP4900	At1g66850
Pollen Ole e I	CRP3390	At5g47635	LTP/2S Albumin/ECA1	CRP4900	At5g38160
Pollen Ole e I	CRP3390	**At2g41390***	LTP/2S Albumin/ECA1	CRP4900	**At5g38195***
Pollen Ole e I	CRP3495	At3g16660	LTP/2S Albumin/ECA1	CRP4920	At1g48750
Pollen Ole e I	CRP3500	At1g54970	LTP/2S Albumin/ECA1	CRP5600	At3g47540
LTP/2S Albumin/ECA1	CRP3860	At3g08770	LTP/2S Albumin/ECA1	CRP5610	At2g43620
LTP/2S Albumin/ECA1	CRP3860	At3g51600			

3.1.8 Microarray data validation through *promoter::NLS-3xGFP* fusions

Three constructs of selected promoters fused to *NLS-3xGFP* were generated as another confirmation to the microarray data obtained. The plants were analyzed whilst

flowering and GFP signals were observed in the cells of the female gametophyte (Figure 28).

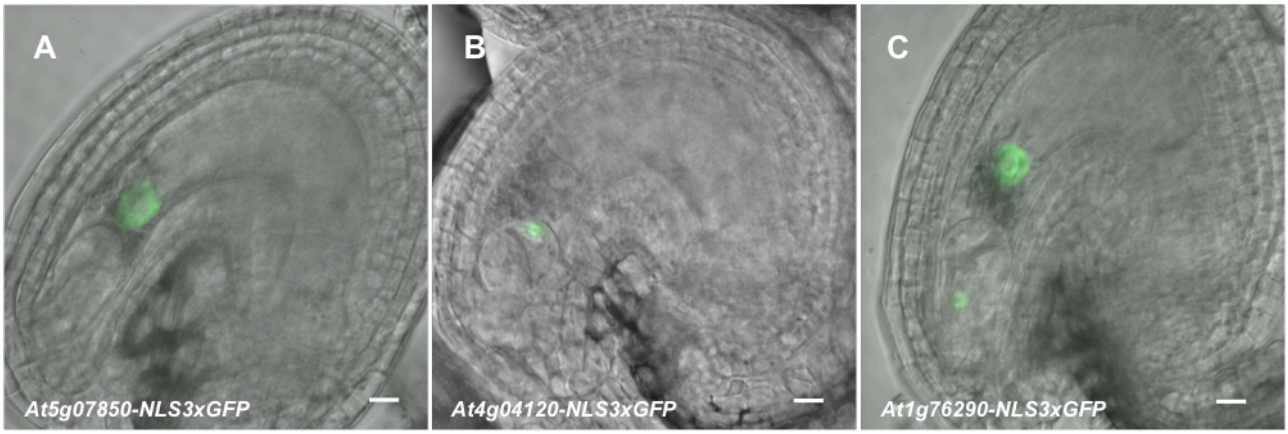


Figure 28. Promoter activity of selected candidate genes analyzed with promoter-reporter fusions. (A) Promoter of *At5g07850,* (B) promoter of *At4g04120* and (C) promoter of *At1g76290* were fused to NLS-3xGFP, showing GFP expression in nuclei of individual cells of the female gametophyte. The signals are in accordance to the data obtained in this study (details provided in the text), with the exception of the *At1g76290* showing an additional signal in the synergid cell, which wasn't detected in the array study.

The promoter fusions of the following genes: *At5g07850,* coding for an acyl-transferase family protein, expressed in the central cell (Figure 28 A), *At4g04120,* previously annotated as a putative reverse transcriptase, but recently reannotated as a transposable element gene, was found in the egg cell (Figure 28 B) and *At1g76290*, coding for an AMP-dependent synthetase and ligase family protein, found to be expressed in the central cell on the array, but promoter fusion also showed a signal in the synergid cells (Figure 28 C).

These results, taken together with the series of PCRs performed on cDNA obtained from several individual batches of single cells isolated over a longer period of time (Figures 21 and 23) and comparison with previously published information (Table 2), largely agree with the microarray data analyzed in this study and serve as a reliable, independent verification of the data.

3.2 Expression of Group III *Pumilios* in the female gametophyte

3.2.1 Identification and expression of selected *Pumilio* genes

A gene belonging to the Pumilio family of RNA-binding proteins, *AtPUM14* (*At5g43110*), has been identified through the course of a microarray study of the FG as expressed weakly, but selectively in the egg cell. The fact that members of this family are known to be involved in germline development and sex determination in animals made *AtPUM14* a particularly interesting candidate gene for research into a possible role in gametophyte development. Among 26 members found in *Arabidopsis*, *AtPUM14* was shown to separate in a group together with *AtPUM13* (*At5g43090*) and *AtPUM15* (*At4g08560*), which were therefore also included in further analyses.

Since *AtPUM15* wasn't detected as expressed in the tissues examined during the microarray analysis (Table 5) and because *AtPUM13* is not represented on the chip, an expression analysis of the three genes was performed in selected tissues by PCR, resulting in a confirmation of a weak expression in reproductive tissues (Figure 29).

Table 5. Expression values of *AtPUM14* (*At5g43110*) and *AtPUM15* (*At4g08560*) in cell types and tissues examined in the course of the microarray analysis.

AtPUM14 (*At5g43110*) was detected as expressed selectively in the egg cell, which made it an interesting candidate for further investigation. *AtPUM15* (*At4g08560*) wasn't detected as present in any of the samples in the microarray analysis. AGI = Arabidopsis Genome Initiative identifier; C = central cell; E = egg cell; S = synergid cell; Sp = sperm cells; P = pollen; Se = seedling.

AGI	C	C call	E	E call	S	S call	Sp	Sp call	P	P call	Se	Se call
At5g43110	61	A	111	P	26	A	21	A	13	A	17	A
At4g08560	22	A	21	A	29	A	33	P	12	A	9	A

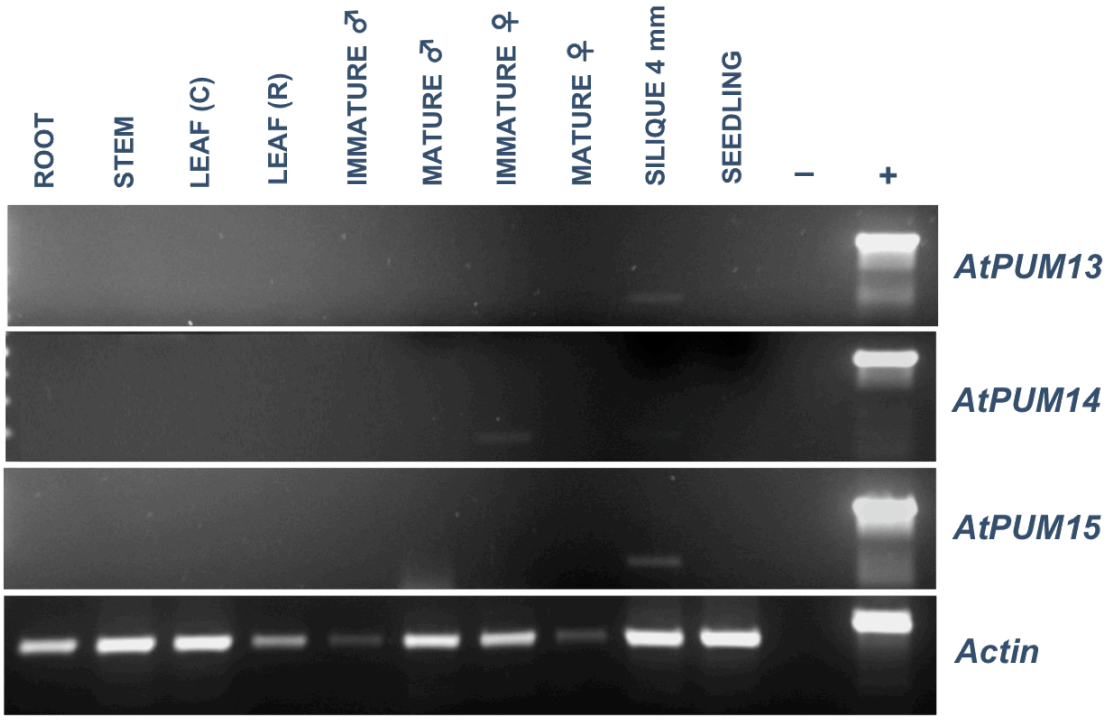


Figure 29. Expression analysis of *AtPUM13, AtPUM14* and *AtPUM15* genes in selected tissues. The PCRs performed on cDNA of selected tissues shows that *AtPUM14* expression is confined to immature ovules, and that *AtPUM13* and *AtPUM15* both seem to be expressed in young siliques, with *AtpUM15*, showing a slightly more prominent band, highlighting its potential role in the tissue after fertilization. The bottom image depicts the PCR products obtained with primers for *actin* and serves as a quality control of cDNA used in the experiment. – = negative control (water); + = positive control (gDNA).

3.2.2 *Promoter::GFP* fusions elucidate the expression patterns of candidate *AtPUM* genes during female gametophyte development

Contrary to expectation, when the *AtPUM14p:NLS(3x)GFP* plants were analyzed (14 independent lines), the expression in the egg cell was barely detectable. However, a strong GFP signal was observed in the antipodal cells (Figure 30), which was not detected before because antipodal cells degenerate in mature embryo sacs and had not been included in the microarray analysis. The earliest visible GFP signal could be observed upon cellularization of the antipodal cells within the developing embryo sac and persisted to the point of antipodal degeneration as the FG came to full maturity, in preparation for fertilization.

Promoters of *AtPUM13* and *AtPUM15* were cloned in the same manner with eight and five independent lines analyzed per construct, respectively. While GFP signals could not be observed in the embryo sac for *AtPUM13*, nor in the developing embryo, the first detectable GFP signal for *AtPUM15* was found in the mature egg cell. Comparable to *AtPUM14*, no signal could be seen prior to cellularization of the female gametophyte. The signal persisted into the early stages of embryo development (Figure 31), but was no longer detectable beyond the 4-celled stage of the developing embryo.

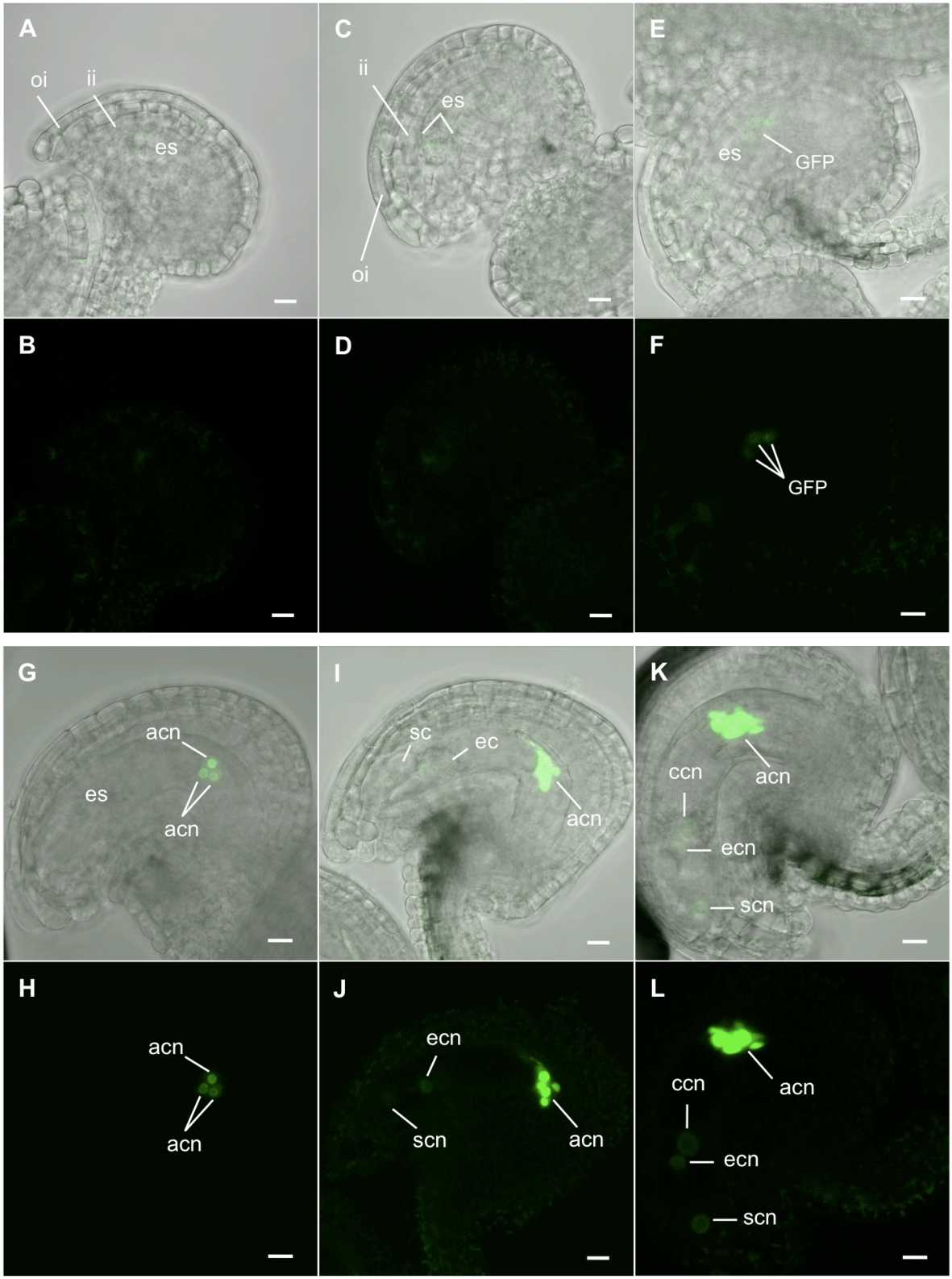


Figure 30. *AtPUM14* promoter activity. Ovules of plants transgenic for *AtPUM14p:NLS(3x)GFP* were examined at different stages of development to establish where and when the promoter was active within the developing embryo sac. (A – D) Younger stages of FG development didn't show a clear GFP signal apart from a greenish glow at the future micropylar end of the forming embryo sac. (E – F) The first distinguishable activity was detected upon cellularization, when the signal became clearly restricted to antipodal nuclei where

it remained strong until their degeneration in the final stages of embryo sac maturation. (G – H) The typical expression pattern was displayed by all examined independent plant lines, with one line displaying a very weak signal in other cells of the embryo sac in addition to the antipodals (I – L). There was no new signal observed in the male gametophyte or after fertilization. The dark images with visible GFP signal correspond to bright-field images above them. oi = outer integument; ii = inner integument; es = embryo sac; acn = antipodal cell nucleus; ec(n) = egg cell (nucleus); sc(n) = synergid cell (nucleus), ccn = central cell nucleus. Size bars are 10 μm.

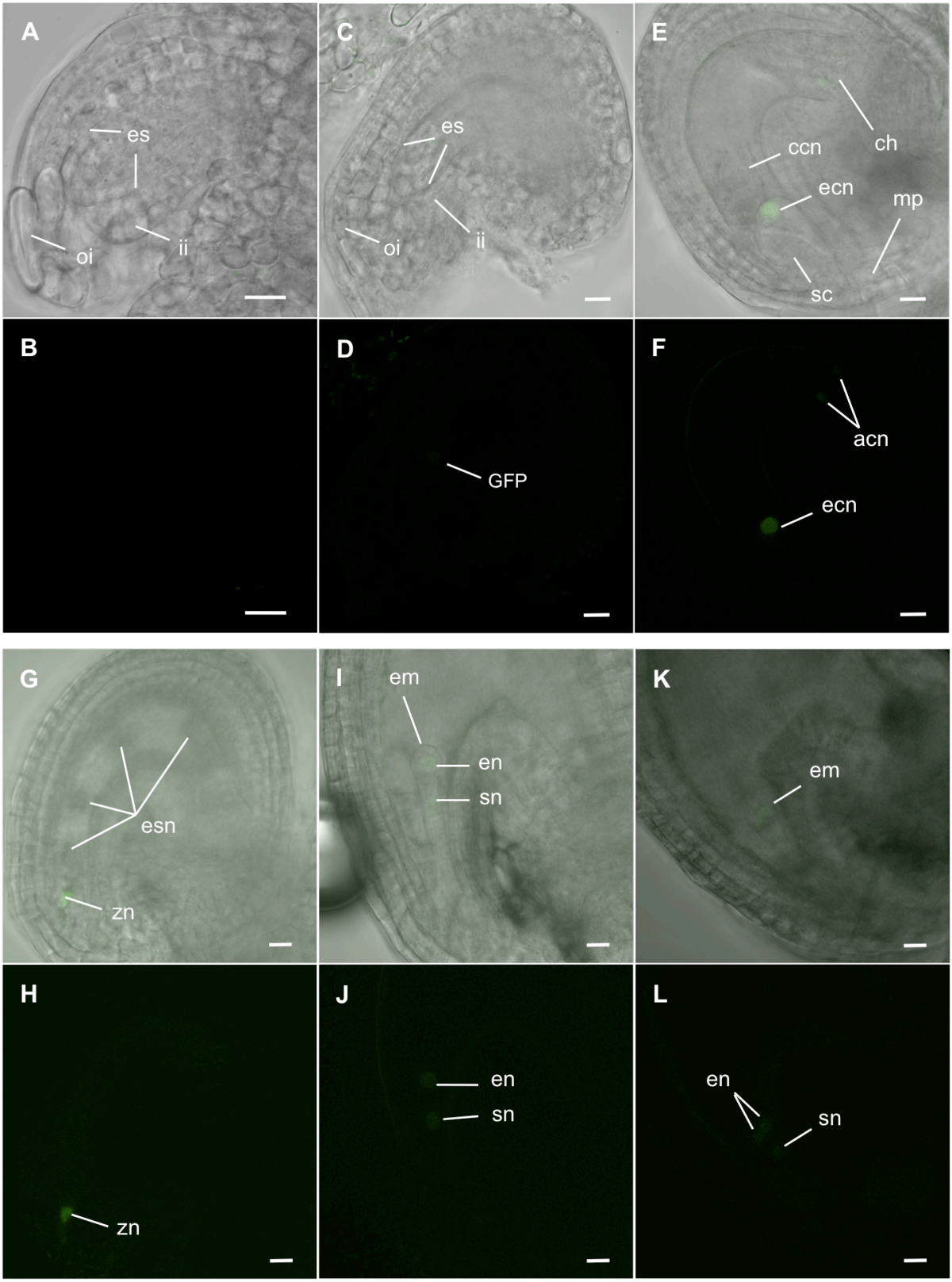


Figure 31. Promoter activity of *AtPUM15*. Ovules of plants transformed with a construct where the promoter region of *AtPUM15* was fused to *NLS-3xGFP* were examined in different stages of development to establish its spatial and temporal activity. (A – C) Younger developmental stages examined didn't display a distinguishable fluorescence. (D – F) The first GFP signals were detected in the egg cell nucleus at the point

of cellularization, with a very weak signal detectable in the antipodal cells of only one of the 5 independent lines. (I – L) After fertilization the GFP signal could be detected in the zygote nucleus, as well as the young embryo, notably the 2-celled and 4-celled stage. No signal could be observed in the male gametophyte. The dark images with visible GFP signal correspond to brightfield images above them. oi = outer integument; ii = inner integument; es = embryo sac; ecn = egg cell (nucleus); ccn = central cell nucleus; sc = synergid cell; acn = antipodal cell nucleus, ch = chalazal pole; mp = micropylar pole; esn = endosperm nuclei; zn = zygote nucleus; em = embryo; en = embryo nucleus; sn = suspensor cell nucleus. Size bars are 10 μm.

3.2.3 AtPUM14 protein localizes in the cytoplasm of antipodal cells

For the purpose of protein localization within cells, the sequence encoding the AtPUM14 protein, fused to GFP, was placed under the control of its native promoter, in order to try to further elucidate its role. The signal, once examined in mature plants, could be observed in antipodal cells, as expected, with a speckled appearance in the cytoplasm (Figure 32), implying localization in vesicles, while dark areas, marking the placement of the antipodal nuclei, indicate its absence there.

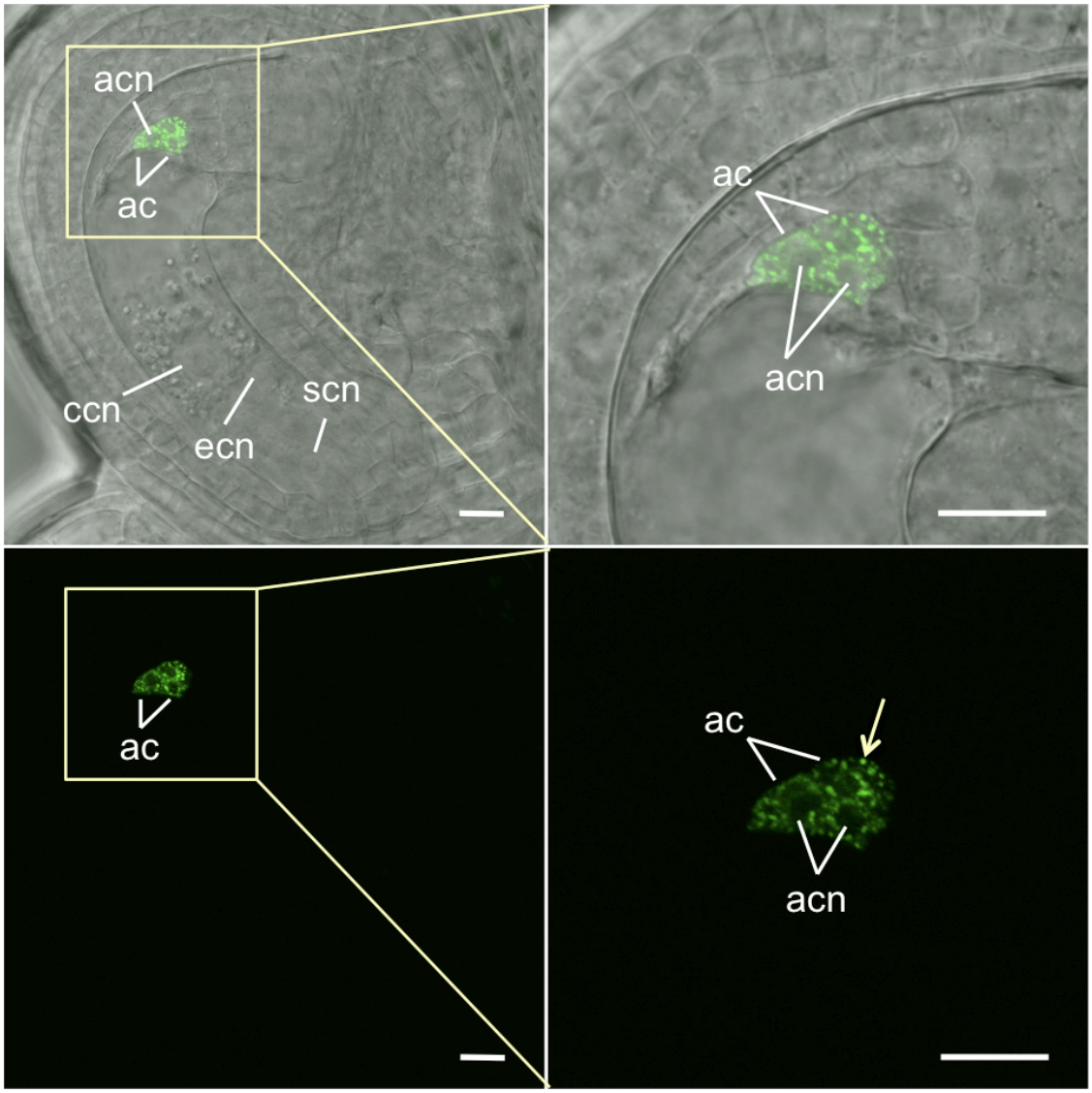


Figure 32. Imaging of *pAtPUM14::AtPUM14:GFP* for the purpose of AtPUM14 protein localization in antipodal cells. The ovule depicted on the left is from a plant transformed with the *AtPUM14* gene sequence fused to *GFP* under the control of its native promoter. All plant lines obtained (9 independent lines) showed uniform expression with punctate GFP signal (arrow) confined to the antipodal cells and seemingly absent from the nucleus. The images on the right are a magnification of the framed area in the images on the left. The upper images are a merge of the lower fluorescent images with their corresponding brightfield images. ac(n) = antipodal cell (nucleus); ecn = egg cell nucleus; ccn = central cell nucleus; scn = synergid cell nucleus. Size bars are 10 μm.

3.2.4 Functional studies on Group III *Pumilios*

3.2.4.1 Genotyping and selection of homozygous lines

Available insertion lines were ordered in order to examine potential phenotypes upon genetic knock-down (Figure 33). No adequate insertion line for *AtPUM13* was available so only those for *AtPUM14* and *AtPUM15* were included in the analysis (Figure 33).

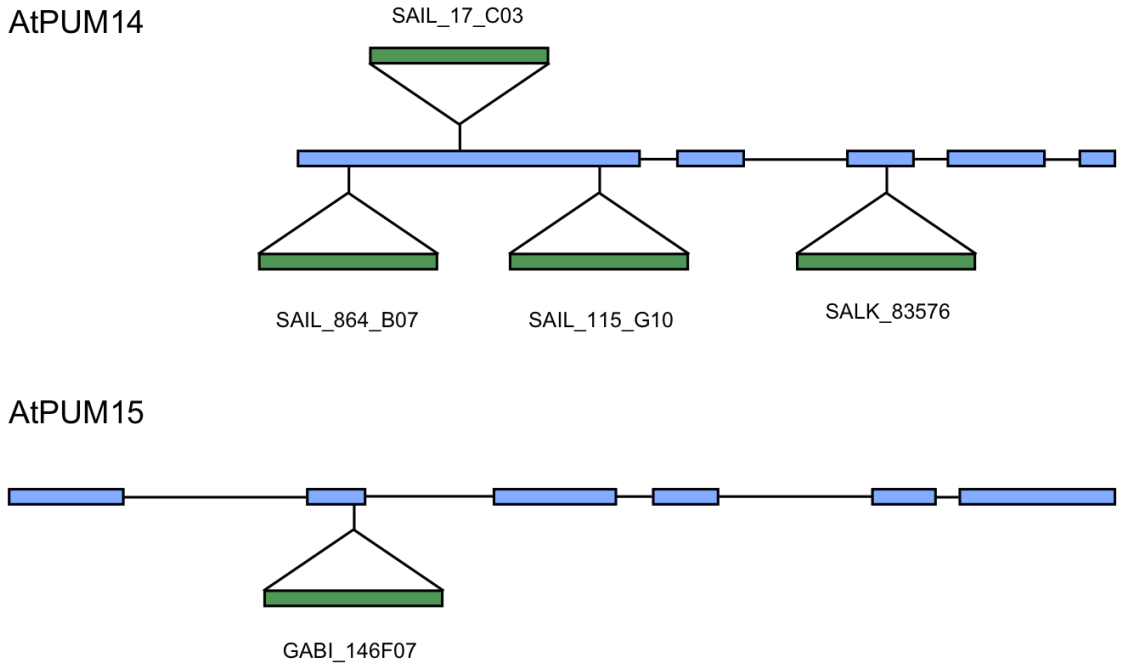


Figure 33. Available insertion lines for *AtPUM14* and *AtPUM15*. The positions of respective T-DNA insertions within genes are indicated. Four different lines were ordered for *AtPUM14*, three of which are in the first exon and one in the third. One line was ordered for *AtPUM15* with a T-DNA located in the second exon. The scheme of the two genes shows their exons marked in blue and introns between them marked as black lines. Insertion lines are marked in green, with their approximate positioning within genes indicated, and their names, as listed in the online databases, are written above or below the insertion.

The plants obtained were genotyped to determine individuals homozygous for a particular insertion. For *AtPUM14* T-DNA insertions two homozygous plants were found for SAIL_17_C03 and SAIL_115_G10, a homozygous SALK_083576 could not be found and the genotyping of SAIL_864_B07 was inconclusive. One line was found to be homozygous for the AtPUM15 GABI_146F07 insertion. The homozygous plants were then microscopically analyzed, but no obvious phenotype could be identified in the immature or

mature embryo sac nor in pollen. Siliques displayed a seed number comparable to that of wild-type plants.

3.2.4.2 Crossing homozygous insertion lines

After their identification, a line homozygous for *AtPUM14* (SAIL_17_C03) and a line homozygous for *AtPUM15* (GABI_146F07) insertion were cross-pollinated in order to obtain double-homozygous progeny that might potentially display a phenotype. All subsequent analyses were performed using these lines. A phenotype was not expected in the first generation progeny (F1) of *pum14/+* (SAIL_17_C03); *pum15/+* (GABI_146F07), as all the plants are double heterozygous. However, the developing siliques, representing the onset of the new generation, were examined and revealed a reduced seed set by approx. 50 % (Figure 34).

Crosing of *pum14/+* and *pum15/+*				
F1 double heterozgotes				
Silique	Developed	Undeveloped	Dev (%)	Undev (%)
1	20	23		
2	22	21		
3	20	25		
4	20	26		
5	30	23		
6	24	24		
7	27	21		
8	23	24		
9	29	22		
10	32	28	51%	49%
11	24	31		
12	17	35		
13	28	25		
14	19	38		
15	25	24	43%	57%
16	22	25		
17	24	25		
18	27	31		
19	23	26		
20	27	24	48%	52%
TOTAL (%)			47%	53%

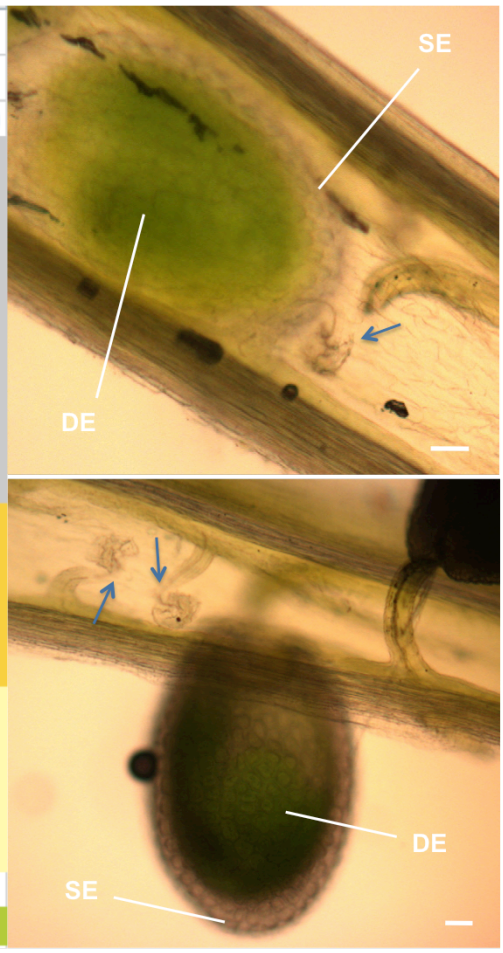


Figure 34. Seed development in the F1 generation of *pum14/+*; *pum15/+* insertion lines. The siliques produced by the progeny of the double heterozygotes had a reduced seed set. Altogether 20 siliques from three plants were examined and the number of developed/undeveloped seeds in each silique is given. Each plant is labeled in the table with a different colour and the percentage of developed/undeveloped seeds for all siliques of that plant is given (total percentages for all plants combined is marked in green). The two images on the right side of the figure show developing embryos that can be seen next to ovules arrested in development, marked with blue arrows. SE = seed; DE = developing embryo; Size bars are 10 μm.

The second generation progeny (F2) obtained from the developed seeds of *pum14/+; pum15/+* was first genotyped in order to select individual plants for further analysis and to identify a double homozygous mutant. Altogether 33 plants were examined and their genotypes are listed in two tables of Figure 35.

Plant	pum14/-	pum14/+	pum15/-	pum15/+
1				
2				
3				
4	o	o		
5				
6				
7				
8			o	o
9				
10				
11				
12				
13				
14				
15				
16				
17				
18			o	o
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32	o	o		
33			o	o

Figure 35. The genotypes of the second generation (F2) progeny of homozygous *AtPUM14* and *AtPUM15* T-DNA insertion line crossings.
In the table on the left the cells in columns marked in dark red represent **homozygosus** insertions in *AtPUM14* (***pum14/-***) or in *AtPUM15* (***pum15/-***), and the cells in orange are **heterozygous** for insertions in both genes (***pum14/+***) and (***pum15/+***). The **homo-** or **heterozygosity** of two plants of the *AtPUM14* insertion (4 and 32) and three plants of the *AtPUM15* insertion (8, 18 and 33) couldn't be determined with certainty, but all of them tested positive for the corresponding insertion. They are marked in pale yellow cells with "o". The "**Actual**" genotypes, as listed in the table below, adhere to the "**Expected**" ratios fairly well and the column titled "**Potential**" accounts for possible additions to a particular genotype if unresolved genotypes (pale yellow cells marked "o") are taken into consideration. **Homo-** and **heterozygosity** is color-coded as before, with the **out-crossed wild type** alleles marked in **pale blue**.

Genotype	Expected	Actual	Potential
pum14/- pum15/-	2	1	(+2)
pum14/- pum15/+	4	4	(+1)
pum14/+ pum15/-	4	2	(+2)
pum14/+ pum15/+	8	7	(+1)
pum14/- +/+	2	3	(+1)
pum14/+ +/+	4	7	(+2)
+/+ pum15/-	2	0	(+1)
+/+ pum15/+	4	3	
+/+ +/+	2	1	

The plant homozygous for both insertions (Plant 1, listed in the left table of Figure 35) was microscopically examined and showed neither a visible phenotype throughout embryo sac development nor after fertilization (not all stages are shown here). The stages depicted in Figure 36 display ovules directly before full maturity, fertilization and shortly after fertilization. Propidium-iodide was used to allow for a better overview of cell boundaries and their nuclei.

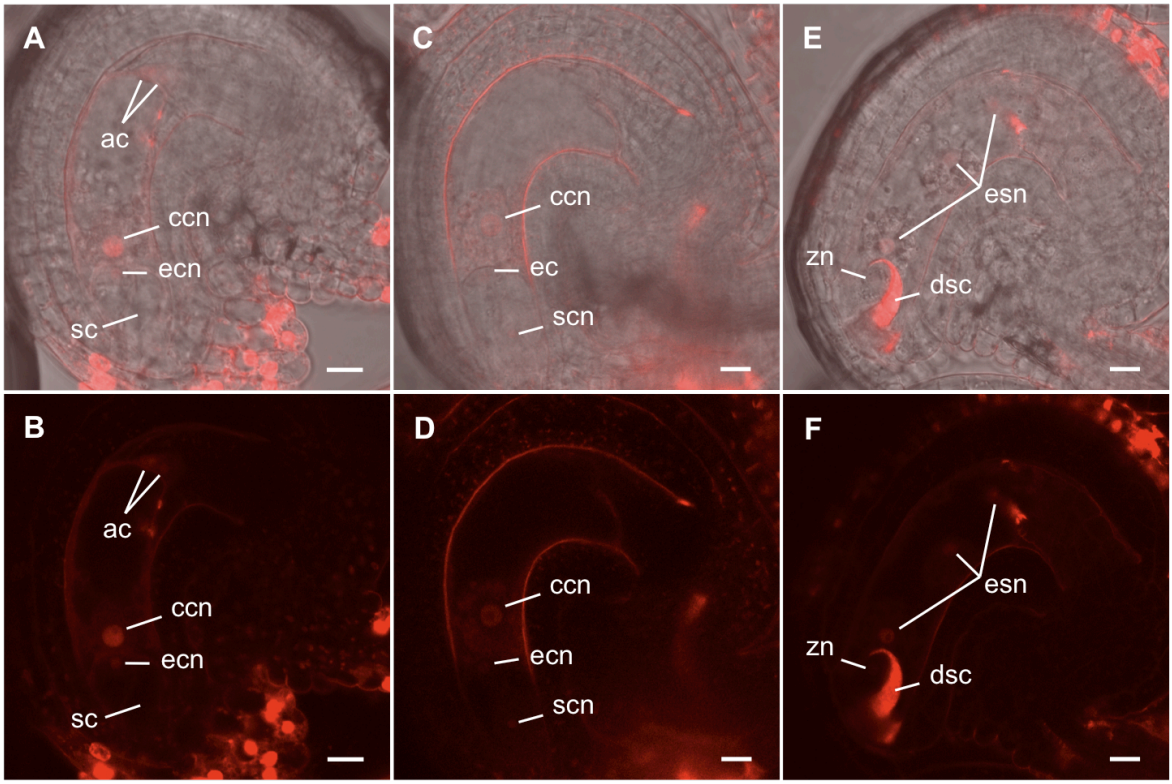


Figure 36. Propidium-iodide stained ovules of *pum14/- pum15/-* double knock-outs. The second-generation (F2) progeny displayed a normal development both before and after fertilization. (A – B) An immature embryo sac with the antipodal cells still visible at the chalazal pole of the embryo sac. (C – D) A mature embryo sac where the antipodals have already degenerated. (E – F) A recently fertilized ovule, with developing endosperm and a remainder of a degenerated synergid. All other stages examined also appeared to correspond to normal development (not shown). Cell boundaries and nuclei stained with propidium iodide. ac(n) = antipodal cell (nucleus); ccn = central cell nucleus; ec(n) = egg cell (nucleus); sc(n) = synergid cell (nucleus); zn = zygote nucleus; dsc = degenerated synergid cell; esn = endosperm nucleus. Size bars are 10 μm.

3.2.5 Silencing *AtPUM13* and *AtPUM14* using artificial microRNA (amiRNA) sequences

As an alternative approach for the knock-down strategy of Group III *AtPUM* genes, in addition to T-DNA insertion lines, two amiRNA constructs have been cloned to target *AtPUM13* and *AtPUM14* simultaneously during FG development. Two different amiRNA sequences were generated, named ami1 and ami4, each targeting a different part of the genes, as shown in Figure 37. They were cloned using the Gateway® system as described in 2.13.2 in more detail, under the control of the *FM1* promoter, which is specifically

expressed in the functional megaspore and in the developing female gametophyte (Huanca-Mamani *et al*., 2005). This produced the *pB-FM1-ami1-GFP* and *pB-FM1-ami4-GFP* constructs, which were then transformed into both wild-type as well as *pum15/-* plants, in an attempt to knock down all three genes of the Group III *Pumilios*.

Figure 37. The position of amiRNA target sequences within *AtPUM13* and *AtPUM14* genes. The figure depicts the positions of two different sequences within the two genes, with the ami1 sequence (framed in orange) targeting a region in the 3' end of both genes, and the *ami4* sequence (framed in dark blue) targeting a central region of both genes. The sequences are visible above the schematic representation of the genes, their position marked directly below the scheme in bright red. In addition to wild-type plants, the constructs *pB-FM1-ami1-GFP* and *pB-FM1-ami4-GFP*, carrying the ami sequences, were also transformed into pum15/-(GABI_146F07) insertion line mutant plants in an attempt to knock-down all three genes simultaneously.

Preliminary analyses of the progeny of ami1 and ami4 plants showed a possible arrest in embryo sac development at FG1 stage. The plants display a seed set reduced up to

50 % in all of the four independent lines produced (ami1 and ami4 in both the wt and pum15/- mutant backgrounds) as shown through the analysis performed by Maria Englhart. These results need to be examined further and in more depth in order to obtain a more conclusive result about the function of Group III *Pumilios* during FG development.

4 DISCUSSION

4.1 Female gametophyte microarray analysis in *Arabidopsis* through manual single cell isolation – an approach superior to previous methods

In this work, the technique to isolate individual cells of the female gametophyte of *Arabidopsis thaliana* was successfully established, providing the means to acquire more precise information about their transcriptional state. Similar techniques have been previously used to isolate these cells in other plants, such as maize, wheat and rice (Kranz *et al*., 1993; Kumlehn *et al*., 1998; Han *et al*., 1998), but because of inaccessibility, mostly due to size restriction, the isolation of the cells in *Arabidopsis* wasn't considered to be a possibility until recently (Ikeda *et al*., 2011). In the past years numerous strategies have been employed to investigate the genes involved in determining identity, function and fate of individual cells comprising the *Arabidopsis thaliana* female gametophyte, namely egg cell, central cell and synergid cells. Previous approaches included different molecular techniques and large-scale mutant screens (Pagnussat *et al*., 2005) as well as differential expression analyses through genetic subtraction and comparative profiling using mutants lacking embryo sacs (Yu *et al*., 2005; Jones-Rhoades *et al*., 2007; Steffen *et al*. 2007; Johnston *et al*., 2007). Another possibility to access individual cells in different tissues arose with the development of the laser-capture technique (Emmert-Buck *et al*., 1996), however due to positioning of cells within the tissue of the female gametophyte it is virtually impossible to capture entire cells and prevent including parts of others, thus possibly excluding some transcripts that might be specifically targeted to a particular part of the cell or "contaminating" samples with transcripts from other cell types.

Once cDNA was generated from manually isolated embryo sac cells and hybridized to the ATH1 GeneChip® the reproducibility of the results was examined since it is one of the most important prerequisites for the analysis of data obtained in a microarray experiment. For that purpose the experiment is performed in three biological replicates for each sample. To display their agreement visually, a Venn diagram (Figure 16) was generated for each sample, with individual circles comprising each diagram representing one of the biological replicates. In a recent study done by Wuest *et al.* (2010) a similar experiment was performed on samples obtained through the application of the laser-capture technique. Even though the number of cells used in that study was significantly higher (300 - 800 per sample compared to 30 – 50 cells used for each replicate in this study), the

resulting data has proven to be highly variable (Figure 38) compared with data generated in this study (Figure 16) probably due to the restriction of the technique itself.

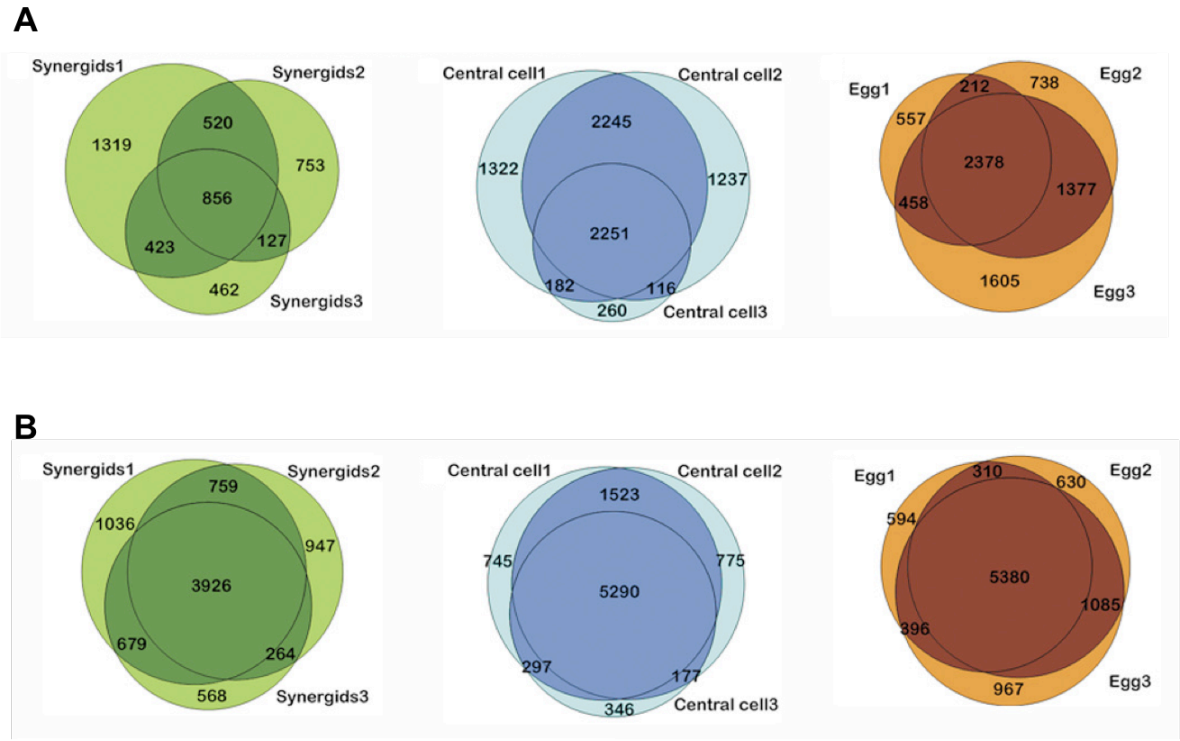


Figure 38. Venn diagrams from Wuest *et al.* (2010) showing overlaps of sample replicates obtained using two different algorithms. (**A**) The Venn diagrams were generated with the MAS5 algorithm, showing relatively low surface-area overlaps of Present calls of biological replicates for each of the samples (synergid, central cell, egg cell). (**B**) After developing a novel algorithm named PANP (Present-Absent calls with Negative Probe sets), the group managed to improve the overlapping surface area of biological replicates of individual samples. However, they obtained a significantly lower number of overall Present calls in comparison to this study, which found 13572 Present in the egg cell, 13690 in the central cell and 11281 in the synergid cell.

It is virtually impossible to precisely collect entire cells of the female gametophyte by LCM because of their irregular three-dimensional positioning within the embryo sac, thus inevitably contaminating samples with parts of other cells, and obtaining incomplete cells, therefore losing the potentially rare transcripts that might have a polar distribution within those cell types and/or may be transcribed in low amounts. Another issue with working with such minute quantities of starting material is the need to introduce amplification steps in order to obtain quantities sufficient for hybridization. The problem that often arises during amplification is a potential bias in 3'/5' signal intensity, towards the

3' end and a loss of signal at the 5' end. This ratio acts as an indication towards the integrity of the input RNA, as well as the efficiency of first-strand cDNA synthesis and amplification by *in-vitro* transcription (Dumur *et al*., 2004; Heber and Sick, 2006). The probes on the Affymetrix ATH1 GeneChip® are designed with a 3' end bias, but also contain probes in the middle of genes as well as the 5' end, the default algorithm however, is not designed to generate Present calls from amplified material, so special caution must be used when dealing with such samples (Casson *et al*., 2005). An illustration of this problem is provided in the afore mentioned study by Wuest *et al.* (2010*)*, in which they devised a novel algorithm, termed PANP (Present-Absent calls with Negative Probe sets), for the generation of Present calls, that significantly improved the overlaps between biological replicates of their samples, which were, before its application hardly reliable (Figure 38). Both algorithms were also applied in this study, but MAS5 already displayed higher overlaps than shown in the Wuest *et al.* (2010) study after the application of the PANP algorithm, and applying PANP here did not show a very significant change when comparing overlaps of individual replicates, nor comparing data with those previously published (Figure 16, 17; Table 2). Also, the number of Present calls that Wuest *et al.* (2010) report remained significantly lower than the number detected in this study, and the group concluded that the FG and sperm are comparable in the number of expressed genes, noting it is a conservative estimate and that the real number of expressed genes might be under-estimated. The microarray study performed in the scope of this work found the FG expressing ~ 90 % of genes found to be expressed in sperm cells (Borges *et al*., 2008) with more than 50 % of these genes found in each individual cell type of the FG. It is quite likely that this difference between the two studies is due to technical limitations, i.e. sample-collection by the laser-capture method, and statistical constraints applied for that reason. Transcription profiling of the central cell, obtained by laser-capture, has been performed recently by high-throughput RNA sequencing (RNA-Seq) (Schmid *et al*., 2012). This novel gene expression platform exhibits superior sensitivity to microarray and offers a variety of new possibilities such as the transcriptional profiling of organisms lacking sequence information, or the identification of novel loci, alternative splicing events, and sequence variation. The group assessed their data in correlation to the previous study (Wuest *et al*., 2010) comparing expression values and number of transcripts detected in both studies. Through this new methodology they managed to increase the number of detected transcripts in comparison in the previous study from 7,633 identified through

microarray to 17,419 through RNA-Seq. This number encompasses 13,849 sequences annotated on the ATH1 GeneChip® and is therefore comparable to the number of transcripts detected in the central cell in the present study (13,690). This confirms the notion that valuable sample purity possible through isolation of single live cells, eliminating cross-contamination from other cell types, contributes significantly to the high resolution of detected transcripts.

4.2 Correlation studies indicate that the cell cycle state is similar between central cell and sperm cells as well as between egg cell and synergid cells

Regulation of the cell cycle progression is extremely important during gametophyte development, ensuring tight coordination throughout the series of meiotic and mitotic events taking place in its formation. A number of mutants, both in the male as well as the female gametohphyte, are known (Ebel *et al*., 2004; Park *et al*., 2004; Guitton *et al*., 2005; Iwakawa *et al*., 2006; Liu *et al*., 2008; Ravi *et al*., 2008; Brownfield *et al*., 2009a, b; Li *et al*., 2009; Gusti *et al*., 2009) leading to errors in the number of cells comprising the mature gametophyte, or misspecification of cell identity.

The precise cell cycle state of the female gametes has been a matter of debate for some time and experiments performed in the scope of this work (Figures 20 – 22) add valuable insight to the topic. In *Arabidopsis*, the male gametes reinitiate the S phase at the end of pollen maturation and are likely to be arrested at the G2/M transition before their release into the female gametophyte (Durbarry *et al*., 2005; Berger *et al*., 2008). During fertilization the cytoplasm and nuclei of the gametes involved are brought together, and their cellular components have to synchronize their cell-cycle states in order to successfully initiate the embryo and endosperm development (Berger *et al*., 2008). Upon fertilization, the initiation of the cell cycle is nearly immediate in the endosperm, with its nuclei dividing first, whereas the first embryonic division takes place around 16 h after fertilization. The division of the zygote, but not of the endosperm, strictly requires the expression of the thymidilate kinase, expressed at the G1/S transition (Faure *et al*., 2002; Ronceret *et al*., 2008; Berger *et al*., 2008). The histogram in Figure 20 clearly shows a high correlation of core cell cycle genes of sperm cells with the central cell, but not the egg cell, supporting the notion that, at fertilization, the central cell is in the G2/M transition phase (Berger and Twell 2011), just like the sperm cells are (Berger *et al*. 2008), and egg cell might be

arrested in G1/S transition as also implied by Berger and Twell (2011), which was additionally supported by RT-PCR (Figure 21).

4.3 The egg cell leads the way in auxin dynamics and signaling in the embryo sac

The determination of egg cell fate, as well as other cells comprising the embryo sac, likely depends on location-specific cues and asymmetric distribution of morphogenetic factors. Since the various roles of auxin are so extremely diverse within the plant organism, examining its presence, distribution and role in the female gametophyte has been particularly interesting, as its presence in this tissue places it at the point both closely preceding and coinciding with the beginning of new life. Even though the overall expression of auxin-related genes in the mature embryo sac doesn't seem to be particularly strong, there is an obvious trend of a higher expression on the female side, especially in the egg cell (Figure 25). Several *IAA* members, as well as genes involved in *IAA* biosynthesis exhibit significantly strong expression in the egg cell, as well as some members in the central cell and synergid, supporting the claims of *de novo* auxin synthesis (Pagnussat *et al*., 2009). All members of the *TIR1* auxin receptor family, shown to be involved in determination of cell identity, switching from synergid to egg cell fate in quadruple mutants (Dharmasiri *et al*., 2005; Pagnussat *et al*., 2009), are found to be moderately expressed throughout the FG. The expression of these genes, including *TIR1* and *AFB1* and *AFB3-5* (Figure 25) is visibly stronger in the egg cell, affirming their roles in maintenance of cell identities assumed by the egg cell and synergid in the egg apparatus. Also auxin influx carriers, *AUX1* in particular, are found to be relatively strongly expressed with the egg cell exhibiting the strongest signal followed by the central cell and synergid cells having a somewhat weaker expression. This could indicate that these cells still accumulate auxin from the surrounding tissue in addition to *de novo* synthesis observed within the FG from FG3 to FG5 (Sundaresan and Alandete-Saez, 2010). Taken together with the very weak expression of genes from the *PIN* family, well-known auxin efflux carriers, this supports existing evidence of its accumulation in the embryo sac prior to fertilization. The overall auxin maximum in the egg cell demonstrated through the dominant expression of auxin responsive genes (in particular *SAUR30, ARF17, ARF18, GH3-11*) implies that the egg cell

is likely the main player involved in the maintenance of cell identities throughout the embryo sac, as recently suggested (Völz *et al*., 2012). Since auxin has been known to be involved in an immense variety of plant processes, it was no surprise to find its role unravelling in the female gametophyte. Although auxin dynamics in the embryo sac has been more closely examined in recent years, this work also provides valuable insight into individual cell particularities, encompassing a number of auxin-related families including almost 200 genes.

Since the discovery of the first RLK almost two decades ago, substantial efforts have been devoted to their characterization, revealing their involvement in a diverse range of processes, including self-incompatibility, disease resistance, regulation of development, and hormone perception (Shiu and Bleecker, 2001). The number of RLK transcripts is well represented among the FG cells and more then half of them are found in at least one cell type, with the egg cell and central cell dominating the overall transcript abundance (Figure 26), which is not surprising given that they are expected to be fertilized, and the complex process itself includes numerous signalling events, many of which are yet to be elucidated. Several RLKs mentioned in 3.1.7.1 and belonging to several different subfamilies, have been found in this work to be expressed in a single cell type of the female gametophyte. These genes in particular offer an intriguing source for research into their role and function in processes occurring in this specialized tissue.

Even though CRPs are poorly represented on the ATH1 GeneChip®, the members that are reveal that a relatively large percentage of certain subgroups is dedicated specifically to the FG. Although the number and diversity of characterized plant CRPs is large, it is thought that many more members of this superfamily remain to be discovered. Some CRPs have non-defense roles, and several types of these proteins are expressed specifically in flowers and seeds, where they play reproductive regulatory roles. These include cell-cell communication in pollen-pistil interactions, such as self-incompatibility, pollen germination, pollen tube growth and attraction (Silverstein *et al*., 2007; Higashiyama, 2010). For four of the six CRP subgroups represented among the 43 CRP genes selectively expressed in one of the FG cells this amounts to 22 – 27 % of the members of each subgroup, namely Thionin (22 %), GASA/GAST/Snakin (27 %), Pollen Ole e 1 (27 %) and LTP/2S Albumin/ECA1 (22 %). For the two remaining subgroups, DEFL and RALF, this number is somewhat lower and adds up to 14 and 11 %, respectively. The relatively high number of LTP/2S Albumin/ECA1 subgroup members

among the single FG cell selective genes, claiming more than half of the 43 members found, is mirrored in their general abundance in *Arabidopsis*, having 276 members identified so far. The *CRPs* identified as FG cell-selective in this study make up an extremely interesting group of candidate genes for further study of signalling dynamics in the FG, especially considering the fact that very few *CRP* genes have been functionally characterized to date. One of the more well-known examples are LUREs, which are CRPs with six cysteines, belonging to a subgroup of defensin-like proteins, expressed specifically in the synergid, and two members in particular, LURE1 and LURE2, were shown to be involved in pollen tube attraction in *Torenia fournieri* (Higashiyama, 2010). A recent study demonstrated that the spatio-temporal dynamics of these peptides could be examined in an *in vitro* pollen tube attraction assay, using direct imaging of chemically labelled LURE with the low molecular mass Alexa Fluor 488 fluorescent dye (Goto *et al*., 2011), which is an exciting example serving as a model for future plant studies.

4.4 The role of RNA-binding proteins and smallRNAs in female germline development in *Arabidopsis*

Small, non-coding RNA molecules have been known for more than two decades. They have the capability to suppress gene expression through various mechanisms once they are incorporated in the multi-protein RNA-Induced Silencing Complex (RISC) and interact with their target mRNA (van den Berg *et al*., 2008). Although the mechanisms used to control gene expression by RISC are quite diverse, two central themes are common to all. At its core, a RISC contains a member of the Argonaute (AGO) protein family that binds to the small regulatory RNA, which then functions as a guide that leads RISC to its target. Argonaute proteins are ubiquitous in plants and animals, common in many fungi and protists, and also present in some archaea (Pratt and MacRae, 2009). The hierarchical clustering heat map (Figure 24), implies that silencing plays a very important role in the FG and smallRNA pathways are likely to be among the most prominent mechanisms of gene regulation in the embryo sac, with the egg cell displaying a particular dynamic through up-regulation of more than half of the genes included in the analysis. The microarray analysis revealed that the egg cell is transcriptionally very active, as are the other FG cells. It is likely that numerous transcripts are deposited in the egg cell prior to fertilization and

potentially subject to translational regulation through interaction with RNA-binding proteins and the silencing machinery. An RT-PCR analysis has been performed in this work, examining the expression pattern of all *Arabidopsis AGOs* in isolated cells of the FG and sperm cells, finding a number of members showing preferential expression between those cell types. Several members were identified to be of special interest given their expression pattern being restricted to reproductive tissues. *AGO5, AGO8* and *AGO9* are perhaps paving the way in *Arabidopsis* for germline-specific gene regulation pathways known in animals, where a group of Argonaute proteins, called PIWI proteins, have been found in different species (Aravin *et al*., 2006; Girard *et al*., 2006; Grivna *et al*., 2006; Lau *et el.*, 2006; Watanabe *et al*., 2006) and the long-awaited analogues regulating germline development in plants have slowly started emerging in recent years (Nonomura *et al*., 2007; Olmedo-Monfil *et al*., 2010; Singh *et al*., 2011; Tucker *et al*., 2012).

The *AtPUM14* gene, a member of Group III Pumilio RNA-binding proteins, also implicated in germline development, spermatogenesis-oogenesis switch and embryo patterning in animals (Lublin and Evans, 2007; Ariz *et al*., 2009; Ginter-Matuszewska *et al*., 2009; Francischini and Quaggio, 2009; Merritt and Seydoux, 2010; Ginter-Matuszewska *et al*., 2011; Kalchhauser *et al*., 2011) was identified as selectively expressed in the egg cell and was examined along with two other members comprising the Group III Pumilios. Further studies revealed a much stronger expression level of the gene in antipodal cells, whose exact role in the FG is still unclear (Friedman *et al*., 2008). Protein localization studies done in this work revealed the protein distributed relatively uniformly in the cytoplasm of antipodal cells in a punctate pattern, seemingly absent from their nuclei. A study done by Tam *et al*. (2010) showed expression of *AtPUM14-GFP* in fava bean under the control of the cauliflower mosaic virus (CaMV) 35S promoter, with the protein also confined to cytoplasmic particles of epidermal cells, however unlike in its native cell type in *Arabidopsis*, quite a strong GFP signal could also be observed in the nuclei of the examined cells. This could be due to the absence of the appropriate nuclear export machinery in the fava bean epidermal cells, where the protein was expressed. It is likely that nuclear export is a common feature of most AtPUM proteins, given that 24 out of 26 members (all except AtPUM1 and AtPUM4) possess a predicted leucine-rich nuclear export signal (NES) located in the variable amino terminal of the protein (Tam *et al*., 2010), which would explain the localization pattern of the protein as shown in this study, i.e. its absence in the nuclei of *Arabidopsis* antipodal cells. The observed cytoplasmic particles

might be P-bodies (processing bodies), which are dynamic centres of mRNA sorting, storage and degradation, or the equivalent of P-granules (P-body like, germline-specific mRNP granules found in *C. elegans*), where a number of RBPs along with certain mRNAs localize, at least transiently. In *C. elegans,* the interacting partners of Pumilio proteins in germline sex determination, *gls-1* and *gld-3*, localize to P-granules, and PUF-8 is also found in P-granules of mitotic germ cells (Lee and Schedl, 2006; Noble *et al*., 2008; Ariz *et al*., 2009; Bailey-Serres *et al*., 2009; Rybarska *et al*., 2009).

Weak expression of all three genes belonging to the Group III *Pumilios* in the FG was confirmed in this study by RT-PCR either prior to, or soon after fertilization, supporting the notion that they are indeed FG specific and might be involved either in germline development or early embryo development. It is possible that Pumilio proteins are able to influence cell identity and function by sequestering one or more transcripts, impeding their further processing, metabolism or potential onward function by repressing translation or triggering mRNA decay. Further studies are necessary to address this issue, primarily identifying the transcripts that Group III Pumilio proteins actually interact with, and elucidating their modes of action. Unfortunately promoter studies didn't produce additional insight in the expression pattern of *AtPUM13* within the embryo sac in any of the plant lines obtained for the promoter fusions. It is possible that the promoter has a very narrow window of expression or the signal might just be too weak to be visually detected. *Promoter::AtPUM15* fusion however, produced a distinct signal in the egg cell nucleus despite the fact it wasn't detected on the chip during the microarray experiment, perhaps also due to signal intensity being below the detection limit on the array. Since the signal persisted through early stages of embryo development there is a possibility that *AtPUM15* is involved in early embryo patterning. In *Drosophila melanogaster*, Pumilio regulates several mRNAs that are involved in embryo polarity (Keene, 2007), is involved in translational repression of *hunchback* mRNA and also interacts with *bicoid* mRNA, (Gamberi *et al*., 2002; Gupta *et al*., 2009) which encode major proteins responsible for anterior-posterior body patterning of the embryo in *Drosophila*. Finding potential AtPUM15 interaction partners in the young embryo stages and functionally characterizing them would help elucidate the situation, determining if the pathways responsible for embryo patterning in plants are comparable to those in some animals or if the mechanisms are perhaps entirely different. The expression pattern of other Pumilio proteins in the individual cells of the female gametophyte remains to be confirmed.

Knock-down of Group III Pumilio was attempted through T-DNA insertion lines for *AtPUM14* and *AtPUM15* and amiRNA for all three members. The absence of an obvious phenotype for homozygous insertion line crossings might be due to the fact they are not expressed in the same cell type or perhaps there is functional redundancy with another member of the family. The phenotype found at FG1, when ~ 50 % of the embryo sacs arrest in development, might not be resulting from the knock-down of the target genes, given they seem not to be expressed at the developmental time-point the phenotype was detected or it might be caused by other factors, such as the unlikely but possible interaction of the amiRNA sequences with an off-target mRNA that might be involved in embryo sac development in that stage. No such target was identified however during the amiRNA sequence design, despite the software feature to find potential off targets in the genome. It is also possible that the expression of Group III Pumilios in the FG prior to cellularization is below the detection limit and even low transcript levels are able to produce the knock-down effect. The high number of Pumilio proteins found in *Arabidopsis*, counting 26 members, implies a potentially high versatility in their functions, with individual members perhaps specialized in different tissues, or a high redundancy of a number of them. Several members of other Pumilio Groups were also found to be weakly expressed in the FG through microarray (*AtPUM2* - *AtPUM8*, *AtPUM12* and *AtPUM23*-*AtPUM24*). Among them *AtPUM12*, belonging to Group II Pumilios, could be a particularly interesting candidate for future examination in the FG. Its expression appears to be quite high in all three FG cells examined, egg cell and synergid cell in particular and it is the only member of Group II detected in the FG.

4.5 Summary and outlook

This work has used single-cell isolation and transcriptome profiling to further our knowledge about the cells comprising the female gametophyte of *Arabidopsis thaliana* which have thus far been highly inaccessible. The study produced valuable candidate genes for future work in a variety of exciting areas of research. Some of these are highlighted throughout the text, found to be expressed exclusively in one embryo sac cell type and are of particular interest as they are likely to be involved in differentiation and determination of cell identity or in specific signalling pathways playing a major role in gamete interaction

during double fertilization. The functional characterization of these genes should be performed, along with the identification of their potential interaction partners through yeast two hybrid screens, co-immunoprecipitation or pulldown assays. The results obtained through the microarray study have been confirmed through several independent methods and offer insight into a number of important processes such as the cell cycle, gene regulation, signalling and hormone dynamics, to name a few. They also provide a database that offers insight into a vast number of different topics through a number of gene lists (some not analyzed in the scope of this thesis) that remain a valuable source of information for years to come.

Pumilio proteins, a family of RNA-binding proteins, were proven to be important factors involved in germline development and sex determination in animals. During the course of this work, beginning with the microarray study, several members of the family were shown to be selectively expressed in the female gametophyte, hinting that pathways comparable to those in the animal field might indeed be present in plants. Future work should be able to address this question through a more in-depth study of these genes, including all remaining family members, their expression profiles, and through identification and characterization of their interaction partners. In addition to Group III *Pumilios*, the most promising new candidate for these studies at this point is *AtPUM12,* as it is strongly expressed in the embryo sac as the only member of Group II *Pumilios*. Since the family in *Arabidopsis* numbers 26 genes, multiple knock-outs would hardly be feasible. Depending on their specific expression profiles and potential functional redundancy it could be possible to silence or knock-down certain members or even entire Groups, which might then result in a phenotype.

Bibliography

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science, 301*, 653-657.

Ambrosone, A., Costa, A., Leone, A., & Grillo, S. (2012). Beyond transcription: RNA-binding proteins as emerging regulators of plant response to environmental constraints. *Plant Sci, 182*, 12-18.

Amien, S., Kliwer, I., Marton, M. L., Debener, T., Geiger, D., Becker, D., et al. (2010). Defensin-like ZmES4 mediates pollen tube burst in maize via opening of the potassium channel KZM1. *PLoS Biol, 8*, e1000388.

Aravin, A. et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature, 442*, 203–207.

Ariz, M., Mainpal, R., & Subramaniam, K. (2009). C. elegans RNA-binding proteins PUF-8 and MEX-3 function redundantly to promote germline stem cell mitosis. *Dev Biol, 326*, 295-304.

Bailey-Serres, J., Sorenson, R., & Juntawong, P. (2009). Getting the message across: cytoplasmic ribonucleoprotein complexes. *Trends Plant Sci, 14*, 443-453.

Banisch, T. U., Goudarzi, M., & Raz, E. (2012). Small RNAs in germ cell development. *Curr Top Dev Biol, 99*, 79-113.

Belostotsky, D. A., & Meagher, R. B. (1996). A pollen-, ovule-, and early embryo-specific poly(A) binding protein from Arabidopsis complements essential functions in yeast. *Plant Cell, 8*, 1261-1275.

Bemer, M., Wolters-Arts, M., Grossniklaus, U., & Angenent, G. C. (2008). The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. *Plant Cell, 20*, 2088-2101.

Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., et al. (2008). A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat Struct Mol Biol, 15*, 998.

Benkova, E., Michniewicz, M., Sauer, M., Teichmann, T., Seifertova, D., Jurgens, G., et al. (2003). Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell, 115*, 591-602.

Berger, F. et al. (2006). Endosperm: an integrator of seed growth and development. *Curr Opin Plant Biol,* 9, 664–670

Berger, F. (2011). Imaging fertilization in flowering plants, not so abominable after all. *J Exp Bot, 62*, 1651-1658.

Berger, F., Hamamura, Y., Ingouff, M., & Higashiyama, T. (2008). Double fertilization - caught in the act. *Trends Plant Sci, 13*, 437-443.

Berger, F., & Twell, D. (2011). Germline specification and function in plants. *Annu Rev Plant Biol, 62*, 461-484.

Bies-Etheve, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., et al. (2008). Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in Arabidopsis thaliana. *Plant Mol Biol, 67*, 107-124.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijo, J. A., et al. (2008). Comparative transcriptomics of Arabidopsis sperm cells. *Plant Physiol, 148*, 1168-1181.

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 Genet, 5*, e1000430.

Brownfield, L., Hafidh, S., Durbarry, A., Khatab, H., Sidorova, A., Doerner, P., et al. (2009). Arabidopsis DUO POLLEN3 is a key regulator of male germline development and embryogenesis. *Plant Cell, 21*, 1940-1956.

Capron, A., Gourgues, M., Neiva, L. S., Faure, J. E., Berger, F., Pagnussat, G., et al. (2008). Maternal control of male-gamete delivery in Arabidopsis involves a putative GPI-anchored protein encoded by the LORELEI gene. *Plant Cell, 20*, 3038-3049.

Casson, S., Spencer, M., Walker, K., & Lindsey, K. (2005). Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis. *Plant J, 42*, 111-123.

Chen, Y. H., Li, H. J., Shi, D. Q., Yuan, L., Liu, J., Sreenivasan, R., et al. (2007). The central cell plays a critical role in pollen tube guidance in Arabidopsis. *Plant Cell, 19*, 3563-3577.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., et al. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell, 110*, 33-42.

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J, 16*, 735-743.

Collinge, M. A., Spillane, C., Kohler, C., Gheyselinck, J., & Grossniklaus, U. (2004). Genetic interaction of an origin recognition complex subunit and the Polycomb group gene MEDEA during seed development. *Plant Cell, 16*, 1035-1046.

Colombo, M., Masiero, S., Vanzulli, S., Lardelli, P., Kater, M. M., & Colombo, L. (2008). AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. *Plant J, 54*, 1037-1048.

Cordts, S., Bantin, J., Wittich, P. E., Kranz, E., Lorz, H., & Dresselhaus, T. (2001). ZmES genes encode peptides with structural homology to defensins and are specifically expressed in the female gametophyte of maize. *Plant J, 25*, 103-114.

Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A. J., et al. (2009). A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J, 28*, 48-57.

De Smet, I., Voss, U., Jurgens, G., & Beeckman, T. (2009). Receptor-like kinases shape the plant. *Nat Cell Biol, 11*, 1166-1173.

Dharmasiri, N., Dharmasiri, S., Weijers, D., Lechner, E., Yamada, M., Hobbie, L., et al. (2005). Plant development is regulated by a family of auxin receptor F box proteins. *Dev Cell, 9*, 109-119.

Dickinson, H., & Grant-Downton, R. (2011). Small RNA pathways and their function in the male gametophye. In V.A. Erdmann and J. Barciszewski (eds.), *Non Coding RNAs in Plants*, RNA Technologies (pp 175-191). Berlin Heidelberg, CT: Springer Verlag

Dresselhaus, T. (2006). Cell-cell communication during double fertilization. *Curr Opin Plant Biol, 9*, 41-47.

Dresselhaus, T., & Marton, M. L. (2009). Micropylar pollen tube guidance and burst: adapted from defense mechanisms? *Curr Opin Plant Biol, 12*, 773-780.

Dumur, C. I., Nasim, S., Best, A. M., Archer, K. J., Ladd, A. C., Mas, V. R., et al. (2004). Evaluation of quality-control criteria for microarray gene expression analysis. *Clin Chem, 50*, 1994-2002.

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.

Ebel, C., Mariconti, L., & Gruissem, W. (2004). Plant retinoblastoma homologues control nuclear proliferation in the female gametophyte. *Nature, 429*, 776-780.

Edwards, T. A., Pyle, S. E., Wharton, R. P., & Aggarwal, A. K. (2001). Structure of Pumilio reveals similarity between RNA and peptide binding motifs. *Cell, 105*, 281-289.

Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., et al. (1996). Laser capture microdissection. *Science, 274*, 998-1001.

Endress, P. K. (2011). Angiosperm ovules: diversity, development, evolution. *Ann Bot, 107*, 1465-1489.

Escobar-Restrepo, J. M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, J., Yang, W. C., et al. (2007). The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science, 317*, 656-660.

Francischini, C. W., & Quaggio, R. B. (2009). Molecular characterization of Arabidopsis thaliana PUF proteins--binding specificity and target candidates. *FEBS J, 276*, 5456-5470.

Friedman, W. E. (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.

Friedman, W. E., Madrid, E. N. & Williams, J. H. (2008). Origin of the fittest and survival of the fittest: relating female gametophyte development to endosperm genetics. *Int J Plant Sci,* 169, 79-92.

Friml, J., Vieten, A., Sauer, M., Weijers, D., Schwarz, H., Hamann, T., et al. (2003). Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis. *Nature, 426*, 147-153.

Galweiler, L., Guan, C., Muller, A., Wisman, E., Mendgen, K., Yephremov, A., et al. (1998). Regulation of polar auxin transport by AtPIN1 in Arabidopsis vascular tissue. *Science, 282*, 2226-2230.

Gamberi, C., Peterson, D. S., He, L., & Gottlieb, E. (2002). An anterior function for the Drosophila posterior determinant Pumilio. *Development, 129*, 2699-2710.

Gasser, C. S., Broadhvest, J., & Hauser, B. A. (1998). Genetic Analysis of Ovule Development. *Annu Rev Plant Physiol Plant Mol Biol, 49*, 1-24.

Gebert, M., Dresselhaus, T., & Sprunck, S. (2008). F-actin organization and pollen tube tip growth in Arabidopsis are dependent on the gametophyte-specific Armadillo repeat protein ARO1. *Plant Cell, 20*, 2798-2814.

Gerber, C. A., Relich, A., & Driscoll, D. M. (2004). Isolation of an mRNA-binding protein involved in C-to-U editing. *Methods Mol Biol, 265*, 239-249.

Ginter-Matuszewska, B., Kusz, K., Spik, A., Grzeszkowiak, D., Rembiszewska, A., Kupryjanczyk, J., et al. (2011). NANOS1 and PUMILIO2 bind microRNA biogenesis factor GEMIN3, within chromatoid body in human germ cells. *Histochem Cell Biol, 136*, 279-287.

Ginter-Matuszewska, B., Spik, A., Rembiszewska, A., Koyias, C., Kupryjanczyk, J., & Jaruzelska, J. (2009). The SNARE-associated component SNAPIN binds PUMILIO2 and NANOS1 proteins in human male germ cells. *Mol Hum Reprod, 15*, 173-179.

Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature, 442,* 199-202

Glisovic, T., Bachorik, J. L., Yong, J., & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett, 582*, 1977-1986.

Goto, H., Okuda, S., Mizukami, A., Mori, H., Sasaki, N., Kurihara, D., et al. (2011). Chemical visualization of an attractant peptide, LURE. *Plant Cell Physiol, 52*, 49-58.

Grant-Downton, R., & Dickinson, H. (2007). Germlines: Argonautes go full cycle. *Curr Biol, 17*, R919-921.

Grant-Downton, R., Hafidh, S., Twell, D., & Dickinson, H. G. (2009). Small RNA pathways are present and functional in the angiosperm male gametophyte. *Mol Plant, 2*, 500-512.

Grant-Downton, R. T. (2010). Through a generation darkly: small RNAs in the gametophyte. *Biochem Soc Trans, 38*, 617-621.

Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev, 20*, 1709–1714.

Gross-Hardt, R., Kagi, C., Baumann, N., Moore, J. M., Baskar, R., Gagliano, W. B., et al. (2007). LACHESIS restricts gametic cell fate in the female gametophyte of Arabidopsis. *PLoS Biol, 5*, e47.

Guilfoyle, T. J., & Hagen, G. (2007). Auxin response factors. *Curr Opin Plant Biol, 10*, 453-460.

Guitton, A. E., & Berger, F. (2005). Loss of function of MULTICOPY SUPPRESSOR OF IRA 1 produces nonviable parthenogenetic embryos in Arabidopsis. *Curr Biol, 15*, 750-754.

Guignard, M.L. (1899). Sur les anthe´rozoı¨ des et la double copulation sexuelle chez les ve´ge´taux angiospermes. *Rev Ge´n de Botanique, 11*, 129–135

Gupta, G. D., Swetha, M. G., Kumari, S., Lakshminarayan, R., Dey, G., & Mayor, S. (2009). Analysis of endocytic pathways in Drosophila cells reveals a conserved role for GBF1 in internalization via GEECs. *PLoS One, 4*, e6768.

Gusti, A., Baumberger, N., Nowack, M., Pusch, S., Eisler, H., Potuschak, T., 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.

Haecker, A., Gross-Hardt, R., Geiges, B., Sarkar, A., Breuninger, H., Herrmann, M., et al. (2004). Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in Arabidopsis thaliana. *Development, 131*, 657-668.

Haffani, Y. Z., Silva, N. F., & Goring, D. R. (2004). Receptor kinase signalling in plants. *Can J Bot 82*: 1-15.

Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., et al. (2011). Live-cell imaging reveals the dynamics of two sperm cells during double fertilization in Arabidopsis thaliana. *Curr Biol, 21*, 497-502.

Hashiyama, K., Hayashi, Y., & Kobayashi, S. (2011). Drosophila Sex lethal gene initiates female development in germline progenitors. *Science, 333*, 885-888.

Havecker, E. R. (2011). Detection of small RNAs and microRNAs using deep sequencing technology. *Methods Mol Biol, 732*, 55-68.

He, X. J., Hsu, Y. F., Zhu, S., Wierzbicki, A. T., Pontes, O., Pikaard, C. S., et al. (2009). An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell, 137*, 498-508.

Heber, S., & Sick, B. (2006). Quality assessment of Affymetrix GeneChip data. *OMICS, 10*, 358-368.

Hejatko, J., Pernisova, M., Eneva, T., Palme, K., & Brzobohaty, B. (2003). The putative sensor histidine kinase CKI1 is involved in female gametophyte development in Arabidopsis. *Mol Genet Genomics, 269*, 443-453.

Higashiyama, T. (2010). Peptide signaling in pollen-pistil interactions. *Plant Cell Physiol, 51*, 177-189.

Higashiyama, T., Yabe, S., Sasaki, N., Nishimura, Y., Miyagishima, S., Kuroiwa, H., et al. (2001). Pollen tube attraction by the synergid cell. *Science, 293*, 1480-1483.

Hock, J., & Meister, G. (2008). The Argonaute protein family. *Genome Biol, 9*, 210.

Huanca-Mamani, W., Garcia-Aguilar, M., Leon-Martinez, G., Grossniklaus, U., & Vielle-Calzada, J. P. (2005). CHR11, a chromatin-remodeling factor essential for nuclear proliferation during female gametogenesis in Arabidopsis thaliana. *Proc Natl Acad Sci U S A, 102*, 17231-17236.

Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol, 9*, 22-32.

Ikeda, Y., Kinoshita, Y., Susaki, D., Iwano, M., Takayama, S., Higashiyama, T., et al. (2011). HMG domain containing SSRP1 is required for DNA demethylation and genomic imprinting in Arabidopsis. *Dev Cell, 21*, 589-596.

Ingouff, M., Sakata, T., Li, J., Sprunck, S., Dresselhaus, T., & Berger, F. (2009). The two male gametes share equal ability to fertilize the egg cell in Arabidopsis thaliana. *Curr Biol, 19*, R19-20.

Inoue, H., Nojima, H., & Okayama, H. (1990). High efficiency transformation of Escherichia coli with plasmids. *Gene, 96*, 23-28.

Iwakawa, H., Shinmyo, A., & Sekine, M. (2006). Arabidopsis CDKA;1, a cdc2 homologue, controls proliferation of generative cells in male gametogenesis. *Plant J, 45*, 819-831.

Jain, M., Kaur, N., Tyagi, A. K., & Khurana, J. P. (2006). The auxin-responsive GH3 gene family in rice (Oryza sativa). *Funct Integr Genomics, 6*, 36-46.

Jensen, W.A. (1965). The ultrastructure and composition of the egg and central cell of cotton. *Am J Bot, 52*, 781–797

Jensen, W.A. (1968). Cotton embryogenesis: the sperm. *Protoplasma, 65*, 277–286

Johnston, A. J., Kirioukhova, O., Barrell, P. J., Rutten, T., Moore, J. M., Baskar, R., et al. (2010). Dosage-sensitive function of retinoblastoma related and convergent epigenetic control are required during the Arabidopsis life cycle. *PLoS Genet, 6*, e1000988.

Johnston, A. J., Matveeva, E., Kirioukhova, O., Grossniklaus, U., & Gruissem, W. (2008). A dynamic reciprocal RBR-PRC2 regulatory circuit controls Arabidopsis gametophyte development. *Curr Biol, 18*, 1680-1686.

Johnston, A. J., Meier, P., Gheyselinck, J., Wuest, S. E., Federer, M., Schlagenhauf, E., et al. (2007). Genetic subtraction profiling identifies genes essential for Arabidopsis reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. *Genome Biol, 8*, R204.

Jones-Rhoades, M. W., Borevitz, J. O., & Preuss, D. (2007). Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. *PLoS Genet, 3*, 1848-1861.

Kalchhauser, I., Farley, B. M., Pauli, S., Ryder, S. P., & Ciosk, R. (2011). FBF represses the Cip/Kip cell-cycle inhibitor CKI-2 to promote self-renewal of germline stem cells in C. elegans. *EMBO J, 30*, 3823-3829.

Kasahara, R. D., Portereiko, M. F., Sandaklie-Nikolova, L., Rabiger, D. S., & Drews, G. N. (2005). MYB98 is required for pollen tube guidance and synergid cell differentiation in Arabidopsis. *Plant Cell, 17*, 2981-2992.

Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., & Jin, H. (2007). A novel class of bacteria-induced small RNAs in Arabidopsis. *Genes Dev, 21*, 3123-3134.

Keene, J. D. (2007). RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet, 8*, 533-543.

Kim, H. U., Li, Y., & Huang, A. H. (2005). Ubiquitous and endoplasmic reticulum-located lysophosphatidyl acyltransferase, LPAT2, is essential for female but not male gametophyte development in Arabidopsis. *Plant Cell, 17*, 1073-1089.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., et al. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science, 303*, 521-523.

Kranz, E., & Lorz, H. (1993). In Vitro Fertilization with Isolated, Single Gametes Results in Zygotic Embryogenesis and Fertile Maize Plants. *Plant Cell, 5*, 739-746.

Kumlehn, J., Lorz, H., & Kranz, E. (1998). Differentiation of isolated wheat zygotes into embryos and normal plants. *Planta 205*, 327-333.

Lau, N. C. et al. (2006). Characterization of the piRNA complex from rat testes. *Science, 313,* 363–367

Lease, K. A., & Walker, J. C. (2006). The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics. *Plant Physiol, 142*, 831-838.

Lee, M. H., & Schedl, T. (2006). RNA-binding proteins. *WormBook*, 1-13.

Lefebvre, V., North, H., Frey, A., Sotta, B., Seo, M., Okamoto, M., et al. (2006). Functional analysis of Arabidopsis NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. *Plant J, 45*, 309-319.

Li, C., & Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A, 98*, 31-36.

Li, L. C., Qin, G. J., Tsuge, T., Hou, X. H., Ding, M. Y., Aoyama, T., et al. (2008). SPOROCYTELESS modulates YUCCA expression to regulate the development of lateral organs in Arabidopsis. *New Phytol, 179*, 751-764.

Li, N., Yuan, L., Liu, N., Shi, D., Li, X., Tang, Z., et al. (2009). SLOW WALKER2, a NOC1/MAK21 homologue, is essential for coordinated cell cycle progression during female gametophyte development in Arabidopsis. *Plant Physiol, 151*, 1486-1497.

Liu, G., Loraine, A. E., Shigeta, R., Cline, M., Cheng, J., Valmeekam, V., et al. (2003). NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Res, 31*, 82-86.

Liu, J., & Qu, L. J. (2008). Meiotic and mitotic cell cycle mutants involved in gametophyte development in Arabidopsis. *Mol Plant, 1*, 564-574.

Lorkovic, Z. J. (2009). Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci, 14*, 229-236.

Lublin, A. L., & Evans, T. C. (2007). The RNA-binding proteins PUF-5, PUF-6, and PUF-7 reveal multiple systems for maternal mRNA regulation during C. elegans oogenesis. *Dev Biol, 303*, 635-649.

Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol, 8*, 479-490.

Luo, M., Bilodeau, P., Koltunow, A., Dennis, E. S., Peacock, W. J., & Chaudhury, A. M. (1999). Genes controlling fertilization-independent seed development in Arabidopsis thaliana. *Proc Natl Acad Sci U S A, 96*, 296-301.

Ma, H. (2009). Regulatory Genes in Plant Development: MADS. eLS

Ma, H., & Sundaresan, V. (2010). Development of flowering plant gametophytes. *Curr Top Dev Biol, 91*, 379-412.

Martienssen, R. (2010). Molecular biology. Small RNA makes its move. *Science, 328*, 834-835.

Marton, M. L., Cordts, S., Broadhvest, J., & Dresselhaus, T. (2005). Micropylar pollen tube guidance by egg apparatus 1 of maize. *Science, 307*, 573-576.

Marton, M. L., & Dresselhaus, T. (2010). Female gametophyte-controlled pollen tube guidance. *Biochem Soc Trans, 38*, 627-630.

Menges, M., de Jager, S. M., Gruissem, W., & Murray, J. A. (2005). Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly

specific profiles of expression and provides a coherent model for plant cell cycle control. *Plant J, 41*, 546-566.

Merritt, C., & Seydoux, G. (2010). The Puf RNA-binding proteins FBF-1 and FBF-2 inhibit the expression of synaptonemal complex proteins in germline stem cells. *Development, 137*, 1787-1798.

Miyazaki, S., Murata, T., Sakurai-Ozato, N., Kubo, M., Demura, T., Fukuda, H., et al. (2009). ANXUR1 and 2, sister genes to FERONIA/SIRENE, are male factors for coordinated fertilization. *Curr Biol, 19*, 1327-1331.

Mockaitis, K., & Estelle, M. (2008). Auxin receptors and plant development: a new signaling paradigm. *Annu Rev Cell Dev Biol, 24*, 55-80.

Moll, C., von Lyncker, L., Zimmermann, S., Kagi, C., Baumann, N., Twell, D., et al. (2008). CLO/GFA1 and ATO are novel regulators of gametic cell fate in plants. *Plant J, 56*, 913-921.

Moore, F. L., Jaruzelska, J., Fox, M. S., Urano, J., Firpo, M. T., Turek, P. J., et al. (2003). Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. *Proc Natl Acad Sci U S A, 100*, 538-543.

Nawaschin, S.G. (1898) Resultate einer revision der befruchtungsvorgaenge bei Lilium martagon und Fritillaria tenella. *Bul Acad Imp Sci, St Petersburg 9*, 377–382

Noble, S. L., Allen, B. L., Goh, L. K., Nordick, K., & Evans, T. C. (2008). Maternal mRNAs are regulated by diverse P body-related mRNP granules during early Caenorhabditis elegans development. *J Cell Biol, 182*, 559-572.

Nonomura, K., Miyoshi, K., Eiguchi, M., Suzuki, T., Miyao, A., Hirochika, H., et al. (2003). The MSP1 gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. *Plant Cell, 15*, 1728-1739.

Nonomura, K., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H., et al. (2007). A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *Plant Cell, 19*, 2583-2594.

Oh, S. A., Johnson, A., Smertenko, A., Rahman, D., Park, S. K., Hussey, P. J., et al. (2005). A divergent cellular role for the FUSED kinase family in the plant-specific cytokinetic phragmoplast. *Curr Biol, 15*, 2107-2111.

Okuda, S., Tsutsui, H., Shiina, K., Sprunck, S., Takeuchi, H., Yui, R., et al. (2009). Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature, 458*, 357-361.

Oliveros, J. C. (2007). VENNY. An interactive tool for comparing lists with Venn diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Olmedo-Monfil, V., Duran-Figueroa, N., Arteaga-Vazquez, M., Demesa-Arevalo, E., Autran, D., Grimanelli, D., et al. (2010). Control of female gamete formation by a small RNA pathway in Arabidopsis. *Nature, 464*, 628-632.

Pagnussat, G. C., Alandete-Saez, M., Bowman, J. L., & Sundaresan, V. (2009). Auxin-dependent patterning and gamete specification in the Arabidopsis female gametophyte. *Science, 324*, 1684-1689.

Pagnussat, G. C., Yu, H. J., Ngo, Q. A., Rajani, S., Mayalagu, S., Johnson, C. S., et al. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development, 132*, 603-614.

Pagnussat, G. C., Yu, H. J., & Sundaresan, V. (2007). Cell-fate switch of synergid to egg cell in Arabidopsis eostre mutant embryo sacs arises from misexpression of the BEL1-like homeodomain gene BLH1. *Plant Cell, 19*, 3578-3592.

Palanivelu, R., Belostotsky, D. A., & Meagher, R. B. (2000). Conserved expression of Arabidopsis thaliana poly (A) binding protein 2 (PAB2) in distinct vegetative and reproductive tissues. *Plant J, 22*, 199-210.

Paponov, I. A., Paponov, M., Teale, W., Menges, M., Chakrabortee, S., Murray, J. A., et al. (2008). Comprehensive transcriptome analysis of auxin responses in Arabidopsis. *Mol Plant, 1*, 321-337.

Park, S. K., Rahman, D., Oh, S. A., & Twell, D. (2004). gemini pollen 2, a male and female gametophytic cytokinesis defective mutation. *Sex Plant Reprod, 17*, 63-70.

Park, S. O., Hwang, S., & Hauser, B. A. (2004). The phenotype of Arabidopsis ovule mutants mimics the morphology of primitive seed plants. *Proc Biol Sci, 271*, 311-316.

Pastuglia, M., Azimzadeh, J., Goussot, M., Camilleri, C., Belcram, K., Evrard, J. L., et al. (2006). Gamma-tubulin is essential for microtubule organization and development in Arabidopsis. *Plant Cell, 18*, 1412-1425.

Pischke, M. S., Jones, L. G., Otsuga, D., Fernandez, D. E., Drews, G. N., & Sussman, M. R. (2002). An Arabidopsis histidine kinase is essential for megagametogenesis. *Proc Natl Acad Sci U S A, 99*, 15800-15805.

Portereiko, M. F., Lloyd, A., Steffen, J. G., Punwani, J. A., Otsuga, D., & Drews, G. N. (2006). AGL80 is required for central cell and endosperm development in Arabidopsis. *Plant Cell, 18*, 1862-1872.

Pratt, A. J., & MacRae, I. J. (2009). The RNA-induced silencing complex: a versatile gene-silencing machine. *J Biol Chem, 284*, 17897-17901.

Ravi, M., Marimuthu, M. P., & Siddiqi, I. (2008). Gamete formation without meiosis in Arabidopsis. *Nature, 451*, 1121-1124.

Reinhardt, D. (2003). Vascular patterning: more than just auxin? *Curr Biol, 13*, R485-487.

Robinson-Beers, K., Pruitt, R. E., & Gasser, C. S. (1992). Ovule Development in Wild-Type Arabidopsis and Two Female-Sterile Mutants. *Plant Cell, 4*, 1237-1249.

Ronceret, A., Gadea-Vacas, J., Guilleminot, J., & Devic, M. (2008). The alpha-N-acetyl-glucosaminidase gene is transcriptionally activated in male and female gametes prior to fertilization and is essential for seed development in Arabidopsis. *J Exp Bot, 59*, 3649-3659.

Ronceret, A., Gadea-Vacas, J., Guilleminot, J., Lincker, F., Delorme, V., Lahmy, S., et al. (2008). The first zygotic division in Arabidopsis requires de novo transcription of thymidylate kinase. *Plant J, 53*, 776-789.

Rotman, N., Gourgues, M., Guitton, A. E., Faure, J. E., & Berger, F. (2008). A dialogue between the SIRENE pathway in synergids and the fertilization independent seed pathway in the central cell controls male gamete release during double fertilization in Arabidopsis. *Mol Plant, 1*, 659-666.

Rotman, N., Rozier, F., Boavida, L., Dumas, C., Berger, F., & Faure, J. E. (2003). Female control of male gamete delivery during fertilization in Arabidopsis thaliana. *Curr Biol, 13*, 432-436.

Rybarska, A., Harterink, M., Jedamzik, B., Kupinski, A. P., Schmid, M., & Eckmann, C. R. (2009). GLS-1, a novel P granule component, modulates a network of conserved RNA regulators to influence germ cell fate decisions. *PLoS Genet, 5*, e1000494.

Salz, H. K. (2011). Sex determination in insects: a binary decision based on alternative splicing. *Curr Opin Genet Dev, 21*, 395-400.

Schiefthaler, 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 organ development in Arabidopsis thaliana. *Proc Natl Acad Sci U S A, 96*, 11664-11669.

Schiott, M., Romanowsky, S. M., Baekgaard, L., Jakobsen, M. K., Palmgren, M. G., & Harper, J. F. (2004). A plant plasma membrane Ca2+ pump is required for normal pollen tube growth and fertilization. *Proc Natl Acad Sci U S A, 101*, 9502-9507.

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., et al. (2005). A gene expression map of Arabidopsis thaliana development. *Nat Genet, 37*, 501-506.

Sheridan, W. F., Avalkina, N. A., Shamrov, II, Batygina, T. B., & Golubovskaya, I. N. (1996). The mac1 gene: controlling the commitment to the meiotic pathway in maize. *Genetics, 142*, 1009-1020.

Shi, D. Q., Liu, J., Xiang, Y. H., Ye, D., Sundaresan, V., & Yang, W. C. (2005). SLOW WALKER1, essential for gametogenesis in Arabidopsis, encodes a WD40 protein involved in 18S ribosomal RNA biogenesis. *Plant Cell, 17*, 2340-2354.

Shi, D. Q., & Yang, W. C. (2011). Ovule development in Arabidopsis: progress and challenge. *Curr Opin Plant Biol, 14*, 74-80.

Shiu, S. H., & Bleecker, A. B. (2001a). Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci STKE, 2001*, re22.

Shiu, S. H., & Bleecker, A. B. (2001b). Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci U S A, 98*, 10763-10768.

Silverstein, K. A., Moskal, W. A., Jr., Wu, H. C., Underwood, B. A., Graham, M. A., Town, C. D., et al. (2007). Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J, 51*, 262-280.

Singh, M., Goel, S., Meeley, R. B., Dantec, C., Parrinello, H., Michaud, C., et al. (2011). Production of viable gametes without meiosis in maize deficient for an ARGONAUTE protein. *Plant Cell, 23*, 443-458.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J. D., Feijo, J. A., et al. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell, 136*, 461-472.

Smyth, D. R., Bowman, J. L., & Meyerowitz, E. M. (1990). Early flower development in Arabidopsis. *Plant Cell, 2*, 755-767.

Spassov, D. S., & Jurecic, R. (2003). The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? *IUBMB Life, 55*, 359-366.

Springer, P. S., Holding, D. R., Groover, A., Yordan, C., & Martienssen, R. A. (2000). The essential Mcm7 protein PROLIFERA is localized to the nucleus of dividing cells during the G(1) phase and is required maternally for early Arabidopsis development. *Development, 127*, 1815-1822.

Sprunck, S., Baumann, U., Edwards, K., Langridge, P., & Dresselhaus, T. (2005). The transcript composition of egg cells changes significantly following fertilization in wheat (Triticum aestivum L.). *Plant J, 41*, 660-672.

Sprunck, S., & Gross-Hardt, R. (2011). Nuclear behavior, cell polarity, and cell specification in the female gametophyte. *Sex Plant Reprod, 24*(2), 123-136.

Srilunchang, K. O., Krohn, N. G., & Dresselhaus, T. (2010). DiSUMO-like DSUL is required for nuclei positioning, cell specification and viability during female gametophyte maturation in maize. *Development, 137*, 333-345.

Stals, H., & Inze, D. (2001). When plant cells decide to divide. *Trends Plant Sci, 6*, 359-364.

Steffen, J. G., Kang, I. H., Macfarlane, J., & Drews, G. N. (2007). Identification of genes expressed in the Arabidopsis female gametophyte. *Plant J, 51*, 281-292.

Sundaresan, V., & Alandete-Saez, M. (2010). Pattern formation in miniature: the female gametophyte of flowering plants. *Development, 137*, 179-189.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., & Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant Cell Physiol, 49*, 493-500.

Tam, P. P., Barrette-Ng, I. H., Simon, D. M., Tam, M. W., Ang, A. L., & Muench, D. G. (2010). The Puf family of RNA-binding proteins in plants: phylogeny, structural modeling, activity and subcellular localization. *BMC Plant Biol, 10*, 44.

Tanaka, H., Ishikawa, M., Kitamura, S., Takahashi, Y., Soyano, T., Machida, C., et al. (2004). The AtNACK1/HINKEL and STUD/TETRASPORE/AtNACK2 genes, which encode functionally redundant kinesins, are essential for cytokinesis in Arabidopsis. *Genes Cells, 9*, 1199-1211.

Tiwari, S., Schulz, R., Ikeda, Y., Dytham, L., Bravo, J., Mathers, L., et al. (2008). MATERNALLY EXPRESSED PAB C-TERMINAL, a novel imprinted gene in Arabidopsis, encodes the conserved C-terminal domain of polyadenylate binding proteins. *Plant Cell, 20*, 2387-2398.

Tucker, M. R., Okada, T., Hu, Y., Scholefield, A., Taylor, J. M., & Koltunow, A. M. (2012). Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in Arabidopsis. *Development, 139*, 1399-1404.

Twell, D., Park, S. K., Hawkins, T. J., Schubert, D., Schmidt, R., Smertenko, A., et al. (2002). MOR1/GEM1 has an essential role in the plant-specific cytokinetic phragmoplast. *Nat Cell Biol, 4*, 711-714.

Ulmasov, T., Murfett, J., Hagen, G., & Guilfoyle, T. J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *Plant Cell, 9*, 1963-1971.

van den Berg, A., Mols, J., & Han, J. (2008). RISC-target interaction: cleavage and translational suppression. *Biochim Biophys Acta, 1779*, 668-677.

Van Ex, F., Jacob, Y., & Martienssen, R. A. (2011). Multiple roles for small RNAs during plant reproduction. *Curr Opin Plant Biol, 14*, 588-593.

Vandepoele, K., Raes, J., De Veylder, L., Rouze, P., Rombauts, S., & Inze, D. (2002). Genome-wide analysis of core cell cycle genes in Arabidopsis. *Plant Cell, 14*, 903-916.

Vaucheret, H. (2008). Plant ARGONAUTES. *Trends Plant Sci, 13*, 350-358.

Vielle-Calzada, J. P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M. A., & Grossniklaus, U. (1999). Maintenance of genomic imprinting at the Arabidopsis medea locus requires zygotic DDM1 activity. *Genes Dev, 13*, 2971-2982.

Volz, R., von Lyncker, L., Baumann, N., Dresselhaus, T., Sprunck, S., & Gross-Hardt, R. (2012). LACHESIS-dependent egg-cell signaling regulates the development of female gametophytic cells. *Development, 139*, 498-502.

Watanabe, T. et al. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev, 20*, 1732–1743

Wang, D., Zhang, C., Hearn, D. J., Kang, I. H., Punwani, J. A., Skaggs, M. I., et al. (2010). Identification of transcription-factor genes expressed in the Arabidopsis female gametophyte. *BMC Plant Biol, 10*, 110.

Wang, X., Zamore, P. D., & Hall, T. M. (2001). Crystal structure of a Pumilio homology domain. *Mol Cell, 7*, 855-865.

Wang, Y., Opperman, L., Wickens, M., & Hall, T. M. (2009). Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proc Natl Acad Sci U S A, 106*, 20186-20191.

Wellmer, F., Riechmann, J. L., Alves-Ferreira, M., & Meyerowitz, E. M. (2004). Genome-wide analysis of spatial gene expression in Arabidopsis flowers. *Plant Cell, 16*, 1314-1326.

Woodcock, D. M., Crowther, P. J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., et al. (1989). Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res, 17*, 3469-3478.

Wuest, S. E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., et al. (2010). Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr Biol, 20*, 506-512.

Yadegari, R., & Drews, G. N. (2004). Female gametophyte development. *Plant Cell, 16 Suppl*, S133-141.

Yang, H., Kaur, N., Kiriakopolos, S., & McCormick, S. (2006). EST generation and analyses towards identifying female gametophyte-specific genes in Zea mays L. *Planta, 224*, 1004-1014.

Yang, W. C., Shi, D. Q., & Chen, Y. H. (2010). Female gametophyte development in flowering plants. *Annu Rev Plant Biol, 61*, 89-108.

Yang, W. C., 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 Dev, 13*, 2108-2117.

Yu, H. J., Hogan, P., & Sundaresan, V. (2005). Analysis of the female gametophyte transcriptome of Arabidopsis by comparative expression profiling. *Plant Physiol, 139*, 1853-1869.

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

Appendix

Primer lists

RT-PCR analysis

AGI	Gene	Forward primer (5'-3')	Reverse primer (5'-3')
At1g44110	*CYCA1;1*	TTGGTGTGGCTTGTATGATG	CTCTTGTGGGATTACGGATG
At5g25380	*CYCA2;1*	ATCGATCCAACGATGCGCGG	GGCGCGAATGAACCGCCTGA
At1g80370	*CYCA2;4*	CAGCCTTGGATGAGAAGAAA	CGAAGCAATGAGCATACAA
At4g37490	*CYCB1;1*	CAGTACCAAAGCCGAAGAAA	TGCACATCAATCAACCACTC
At3g11520	*CYCB1;3*	TGCTTGCTGCATCAGCGGTCT	CGTAGCTTGCTCTCCCCCGC
At4g35620	*CYCB2;2*	CGTTACTCGCCGCCACTGCC	AGAGGGAAAATGTTAGCAGCGAGC
At3g50070	*CYCD3;3*	CCCGCTAGTCCGATCGGTGTGT	GCCTCATCTGCTGCTCTTGCACT
At5g65420	*CYCD4;1*	CCCCTTCCTCCACGAGGGCA	TGCCACAGCAGCAGCAACTTCA
At1g20930	*CDKB2;2*	TGATGGACCGGAAGACAATG	TCCAGGCCAAACTTCTTCGT
At1g48410	*AGO1*	GGCACTTCTCGACCTGCTCAT	AGCTGGGAGTGGCCTCACTG
At1g31280	*AGO2*	ACTTTAAGCAGCCGCGGGGA	CTGCGTTATTGGGGCGTTCCGA
At1g31290	*AGO3*	GGCAATGTGCCTTCAGGTACGGT	CGCTCCACGCGACTGCTTAGAATT
At2g27040	*AGO4*	AAGCACCAGTGCCATTTCTG	AACCCCAACAGGCAAACAAG
At2g27880	*AGO5*	CCCTGAGCAACACGGGAATC	TAGTAGCGGGCACGGAATGC
At2g32940	*AGO6*	GCAGCAGCTCAAGTTGCGCAAT	AGCTAAACGCAAAAATCGCTGCCG
At1g69440	*AGO7*	CACGAGCAGGCCAACGCATT	AGCCTTCCTCTGTACGCAGCAAG
At5g21030	*AGO8*	GCGGTTGTGAGCTCCAGAGAG	GGGCTTTCGGTTTGGAAGAAC
At5g21150	*AGO9*	TCGAGGCCCTGATAATGTTCC	CTGTGCAGCTGCCAAATGAG
At5g43810	*AGO10*	ACAGAAGCGTCACCACTCG	CGTGCTCGAAATGCTGCAAG
At5g53560	*AtCB5*	AGGCGATGAAGTCTTGTTGTCC	CCTTTGGCTTCTTCTAGTCTTTCT
At1g19890	*MGH3*	ACTAGACGACCGTACCGTGGT	CGATTCTATCTCACCCATCAA

PCR / Southern hybridization

AGI	Gene	Forward primer (5'-3')	Reverse primer (5'-3')
At2g21740	*Ec1-2a*	AACTCTCCCGGAGACGGAAG	AAGAACGTTGGTTTCATCAGAG
At1g47470	*DD31*	CAGGAGGTGCCAAAAGATGT	AAAAGAGACCGGAGCACCAAT
At3g10890	*DD65*	AAAACCCTTACGGCTTGGATGAAT	TTGCGTTGAACACTGCGTCTCT
At5g53560	*AtCB5*	AGGCGATGAAGTCTTGTTGTCC	CCTTTGGCTTCTTCTAGTCTTTCT

Promoter-fusion analyses

AGI	Gene	Forward primer (5'-3')	Reverse primer (5'-3')
At5g07850	*HXXXD-type acyl-transferase*	CACCCACTTTAAAATCGAATTTGC	ATAAAGAACAAGAGTGTCTCTACT
At1g76290	*AMP-dependent synthetase/ligase*	CACCAGCTTCTCTCGTTAAAGCTC	TTCTTTTCTATTGTTATTCTCTAT
At4g04120	*transposable element gene*	CACCACGGTGCTTTTGTATCTAAG	AGCTATATCGTATAACTTGGTTAG
At5g43090	*AtPUM13*	CACCAGACAACTTTTGGTAATCGG	GGTTAAAGAAAGAAGAAGAAGCTGA
At5g43110	*AtPUM14*	CACCTTATACTCCTAATTTTACTCCGAGCTAAACTC	CATGGTTTTGAAAATCAAGGAGGAGC
At4g08560	*AtPUM15*	CACCAATCTTATGCTTACTTCTATT	TTTCATTTTCTTTTTGGTAGATGTT

Protein localization

AGI	Gene	Forward primer (5'-3')	Reverse primer (5'-3')
At5g43110	*AtPUM14*	CACCATGGACAAGAATTTTCGTGTCAACACTAAC	GATATTGAGTTTCTCCAGAACTTTGTTCCC

amiRNA cloning

AGI	Target Genes	Primer	Sequence (5'-3')
		amiA + CACC	CACCCTGCAAGGCGATTAAGTTGGGTAAC
		amiB	GCGGATAACAATTTCACACAGGAAACAG
		I miR-s_1	gaTTTACGCTCTATATGATGCCTtctctcttttgtattcc
		II miR-a_1	gaAGGCATCATATAGAGCGTAAAtcaaagagaatcaatga
At5g43090	*AtPUM13*	III miR*s_1	gaAGACATCATATAGTGCGTAATtcacaggtcgtgatatg
At5g43110	*AtPUM14*	IV miR*a_1	gaATTACGCACTATATGATGTCTtctacatatatattcct
		I miR-s_4	gaTTGACAATTGAAGTGCGACCGtctctcttttgtattcc
		II miR-a_4	gaCGGTCGCACTTCAATTGTCAAtcaaagagaatcaatga
		III miR*s_4	gaCGATCGCACTTCATTTGTCATtcacaggtcgtgatatg
		IV miR*a_4	gaATGACAAATGAAGTGCGATCGtctacatatatattcct
	Gateway® compatibility	HAU62	CACCAAACACACGCTCGGACGCATATTAC
		HAU63	CATGGCGATGCCTTAAATAAAGATAAACC

Acknowledgements

I would like to thank Prof. Thomas Dresselhaus for the opportunity to work in his research group at the Department of Cell Biology and Plant Biochemistry and especially for his constructive comments during the correction of my Thesis.

I would also like to thank my supervisor, Dr. Stefanie Sprunck, for the correction of my thesis. I have learned a lot during our time together which has helped me grow as a person, for which I will always be very grateful.

I kindly acknowledge Dr. Stefanie Sprunck, Dr. Manfred Gahrtz, Dr. Mihaela Marton (University of Regensburg) and Dr. Mily Ron (UC Davis, USA) for providing vectors used in this work. Thank you Dr. Marina Gebert for teaching me how to isolate single FG cells. I am grateful to all group leaders in our Department who helped me with their scientific advice and encouraged me throughout my PhD, as well as to Jörg Becker for being available for my inquiries, for our discussions and his support during microarray analysis. Also to Dunja Leljak-Levanić, for her scientific and moral support throughout my thesis. It means a lot to have you on my side.

Thank you to the staff at the Department, Veronika Mrosek for all the administration, Gunther Peissig for endless *Arabidopsis* plants and of course for the 'kuchen', Monika Kammerer, Annemarie Taffner, Angelika Rechemacher, Ingrid Fuchs and to everyone who I haven't mentioned by name, members past and present.

To all my colleagues who have made the hard times easier and all the fun we had in and outside the lab. Marina Gebert, Birgit Bellmann, Jeab (Kanok-orn Srilunchang), Nadia Krohn, Irina Kliwer, Svenja Rademacher, Andreas Lausser, Philipp Alter. To everyone in our Department outside our group and to the students – especially my dance partner Klemens Michel and my Roman friend Daniele Hassler. To all the guest students and researchers visiting our lab. Especially to my Croatian girl Martina Juranić and my Portuguese girl Ana Marta Pereira. To my other favourite Portuguese person, Filipe Borges, thank you for your support, our endless chats and above all for the music.

To all of my friends in Croatia and elsewhere, to all the people who played a role in my life one way or another. You helped me become who I am today and will always be a part of me... (some of) you know who you are ;)

I would like to dedicate this work to my friend and mentor Srećko Jelenić.

I will always remember you.

To Daniel Neal. Words are insufficient and I will not waste them....

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe des Literaturzitats gekennzeichnet.

Lucija Šoljić
Regensburg, den 17 April 2012