Proteome and Transcriptional Analysis of *Arabidopsis thaliana* Sperm Cells

Ajaraporn Sriboonlert

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2008

Abstract

Angiosperm sexual reproduction is unique and remarkable. Unlike for other living organisms, fertilisation involves the fusion of two sperm cells to two cell types of female gametophyte, the egg and central cell. This fertilization process is called double fertilisation. Fusion of the egg and the sperm yields a zygote whereas the interaction of the central cell and the sperm gives the nutritive endosperm. This fertilization process has been studied extensively for many years. However, despite these studies, relatively little is known at the molecular level about either of the plant gametes. This is primarily due to the difficulties in plant gamete isolation. Both plant gametes reside within enclosed tissues, pollen grains and embryo sacs. In this study, sperm isolation techniques were successfully developed for *Brassica oleracea* and *Arabidopsis thaliana* which ultimately utilised fluorescence-activated cell sorting (FACS) to obtain pure cell samples. Proteomic studies utilising two-dimensional protein electrophoresis and mass spectrometry (MS) were carried out on both semi-purified and FACS-purified sperm cells. In parallel with the proteomic studies a bioinformatics approach was taken which used sperm transcriptome data of maize, *Plumbago zeylanica*, rice and tobacco to identify homologues in Arabidopsis. Transcriptional analyses, RT-PCR and GFP translational fusion experiments were used to investigate these Arabidopsis sperm cell-expressed gene candidates. As a result, two sperm cell-expressed genes (At1g10090 and At5g39650) were identified and these are being analysed to confirm their functions in the reproductive process. Moreover, the sperm cell-expressed gene candidates derived from the bioinformatics were also screened for roles in plant reproduction by a reverse genetics approach (Arabidopsis T-DNA insertion mutagenesis plant screening) and eight genes were identified. In addition, for the first time, sperm cell size dimorphism was identified for Arabidopsis in this study utilising a GFP-labelled sperm line and confocal microscopy. Overall the techniques for sperm cell-expressed gene candidate selection were proven to be effective and will certainly facilitate further sperm cell-specific gene identification studies. Further the Arabidopsis sperm purification technique successfully developed in this project will surely be useful for any plant sperm cell studies in the future.

Acknowledgements

I would like to thank my supervisor Dr. James Doughty for his support throughout my Ph.D. Without his advice and encouragement this thesis would not have been possible. My thanks also go to Dr. Alex Jeffries for his help throughout the bioinformatics part of this project, Dr. Ursula Gerike for her assistance on the 2DE and MS, Dr. Ursula Potter for her assistance on the TEM, Dr. Timothy Karr for the useful information on the sperm proteomics, Dr. Andrew Herman (University of Bristol), his expertise in FACS operation was tremendously helpful for the sperm purification technique development in this study. Further thanks to Dr. Scott D Russell (University of Oklahoma) and Dr. Rod Scott for providing the plant materials.

I would also like to thank all of my friends and colleagues for their advices academically and personally. Most importantly, I would like to thank my parents, without their support none of this could have been done.

Finally, I would like to thank the Royal Thai Government for funding.

Contents

Table of Figures

List of Tables

Chapter 1:

Introduction

The sperm cell is a crucial element in sexual reproduction in both animals and plants. The haploid sperm cell carries genetic material from the male to combine with that of the female, generating a progeny cell called the zygote. Studies on sexual reproduction in both plants and animals in the past decades have revealed some similarities in this process between the two kingdoms (Marton and Dresselhaus, 2008). However it is noteworthy that in higher plants two fusion events occur in sexual reproduction, a process termed double fertilisation. In plants the male reproductive unit, the pollen grain, carries two sperm cells to fertilise two female reproductive cells in the embryo sac, the egg cell (female gamete) and the central cell. Gamete fusion between one sperm cell and egg cell generates the zygote (embryo) where the fusion between the other sperm cell and central cell generates endosperm which provides nutrients for the developing embryo (for more detail see section 1.2.2). Sperm cell studies in animals have progressed dramatically as the sperm are generally easily harvested or relatively accessible in testis thus facilitating isolation and purification for molecular analyses. Indeed a recent proteomic study on the purified sperm cells of the fruit fly, *Drosophila melanogaster*, was successfully performed utilising new whole-sperm mass spectrometry (WSMS) technique and revealed 381 proteins presented in the sperm cell (Dorus et al., 2006; Karr, 2007). In contrast to animals systems, plant sperm cells are more difficult to study and thus relatively little is known about them, particularly at the molecular level. So far only a few sperm cell-specific genes have been identified in plants and only one gene called *HAP2* (Mori et al., 2006; von Besser et al., 2006) identified to be involved in gamete fusion (detail in section 1.2.2). As every pair of sperm cells is contained within a strong pollen grain, sperm cell isolation and purification is problematic. Importantly, to fully understand sperm cell biology and hence the plant reproductive process, pure sperm cells are required. Therefore, this project aimed to develop

protocols to isolate and purify sperm cells from the model species Arabidopsis to provide material for sperm cell proteome and molecular studies. Moreover, bioinformatics and transcriptional studies were also applied in parallel to identify sperm cell-specific genes. In this chapter, an overview of plant and animal sperm structures and fertilisation processes will be reviewed and compared. Further, plant sexual reproductive processes in both primitive and modern flowering plants, including double fertilisation will also be reviewed. In addition, the angiosperm reproduction process, from gametogenesis to gamete fusion, will be reviewed to provide a solid contextual background for the research carried out in this project.

1.1 Animal sperm biology

Both animal and plant sperm cells have identical functions in reproduction, that of delivering a haploid complement of genetic material to the female gamete. Although animal and flowering plant sperm cells are different in many respects they also share some similarities (Marton and Dresselhaus, 2008) (see reviews later in this section) which could inform studies such as those covered in this project. Therefore, current knowledge relating to animal sperm cells will be reviewed in this chapter. Moreover, some techniques utilised in animal sperm cell studies will also be mentioned as they could be adapted for use in plant models.

1.1.1 Animal sperm cells: History, morphology, and functions

Sperm cells were first discovered by Leeuwenhoek – the Dutch microscopist and Hartsoeker in 1678 (Gilbert, 2000; Hill, 1985). Leeuwenhoek believed that sperm were seeds which contained a preformed embryo and the female provided the soil for them to grow. However, no one knew the importance of sperm in reproduction until 1824; Prevost and Dumas proposed that sperm were the active agent for fertilization. They also noted that sperm existed in all sexually mature males and were absent in immature and aged individuals (Gilbert, 2000).

As mentioned earlier sperm cells have a crucial function in fertilisation acting as a vehicle to deliver genetic material from the male to the female to produce the zygote. Therefore, animal sperm cells consist of a haploid nucleus, and an enzyme compartment that enables the nucleus to enter the egg (Ainsworth, 2005). The head of the sperm is comprised of tightly packed DNA and little cytoplasm (Ainsworth, 2005). During spermatogenesis in animals, a basic nuclear protein called protamine is responsible for nuclear DNA condensation by replacing histones in the chromosome (Balhorn, 1982; Rooney et al., 2000; Wouters-Tyrou et al., 1998). Behind the head is the midpiece,

containing more than 50 power units called mitochondria that drive the lashing motion of the attached tail (Ainsworth, 2005). The tail movement is controlled by using lamellipodial protrusion, a process analogous to the amoeboid motility of other eukaryotic cells, or in most species by flagella (Bottino et al., 2002; Gilbert, 2000). Sperm cell morphology is diverse with huge ranges in size and shape – for example, the fruit fly *Drosophila bifurca* has giant sperm cells, the termite *Mastotermes darwiniensis* has multi-flagellate sperm similar to that found in some plant species such as the whisk fern, *Psilotum nudum* (Renzaglia et al., 2001) and the nematode *C. elegans* possesses aflagellate sperm which are similar in terms of structure to those of higher plant species where sperm cells have no flagella. These differences in sperm cell morphology among animal species probably occurred due to sperm competition forces in each species (Morrow, 2004). Therefore, in monandrous species (a species that has one male partner at a time) where sperm competition is not required, immotile sperm cells are produced to conserve energy for other purposes of reproduction as flagella production is considered costly (Morrow, 2004). However, in contrast with most of the animal species, the sperm cells of flowering plants are immotile and contain no flagella thus a special gamete delivery mechanism is required for successful reproduction (see detail in section 1.2.2).

As sperm were known to have a function in fertilisation it was long believed that these specialised cells were isomorphic in animal species. However, in some animals, sperm cells were discovered to be polymorphic. In plants, sperm cell polymorphism is relatively common and refers to the differences between the two sperm cells in each pollen grain. As fertilisation in angiosperms occurs twice (double fertilisation) with two different fusion partners, plant sperm dimorphism was believed to be related to the fate of their fusion partner i.e. central cell generating endosperm, and the egg cell generating a zygote (see detail in section 1.2.2). Sperm polymorphism in animals has been proposed to be involved in sperm cell competition between individuals (Bellis et al., 1990). An interesting discovery on sperm cell competition in animals was reported in rats (Bellis et al., 1990) where sperm cells of a different functional class to that of fertilisation are produced to facilitate competition with sperm from other males (Bellis et al., 1990; Kura and Nakashima, 2000). This class of sperm cell is termed "soldier sperm" or "kamikaze sperm" (Bellis et al., 1990; Kura and Nakashima, 2000). Such a system is present in species where females interact with many males resulting in ferocious sperm competition (Kura and Nakashima, 2000). The soldier sperm are hypothesised to block or kill the sperm from other males (Bellis et al., 1990; Kura and Nakashima, 2000). However it has been argued by some that organisms cannot afford to produce such sperm cells as sperm production is costly (Harcourt, 1991), nonetheless there is a suggestion that both morphs of sperm cells are capable of fertilising eggs

under certain circumstances (Kura and Nakashima, 2000). Polymorphic sperm are also present in lepidopterans, where two types of sperm are produced, one called apyrene – anucleate sperm reported to delay female remating behaviour, and eupyrene – nucleate sperm that fertilise the egg (Snook, 1997). Thus, in summary, the function of this 'helper' sperm seems to be varied among species. Selection pressure is considered to be the main reason for sperm polymorphism either associated with sperm competition (Snook, 1997).

1.1.2 Gamete fusion (syngamy) in animals

An interaction between an ovum (egg cell) and a sperm in animals generally begins with the chemoattraction of sperm cells by soluble molecules secreted by the egg cell (Wassarman et al., 2004). In plants, chemoattraction occurs during the final stages of pollen tube growth (see section 1.2.2). The second step involves the penetration of the egg cell by the sperm. In animals, the acrosomal region of the sperm cell provides the enzymes to help sperm get through the egg envelope, an extracellular coat of the egg cell. Sperm cells bind to the vitelline envelope (zona pellucida – ZP) of the egg which contains glycoproteins ZP1, ZP2, and ZP3 that interact with each other using non-covalent bonds (Wassarman et al., 2004). Species-specific binding of sperm to eggs is mediated by oligosaccharides linked to ZP3. Glycoproteins have also been identified on the egg cell surface of Fucus and were demonstrated to be crucial for gamete fusion (Callow, 1985; Catt et al., 1983) (see section 1.2.1 for detail). Further, a Concanavalin A binding site was also identified on the egg cell surface of the higher plant *Nicotiana tabaccum*. Similar to the case in animals, this lectin binding site demonstrated polar distribution which is proposed to determine the site of gamete interaction (Ke-Feng et al., 2005a; Ke-Feng et al., 2005b) (see detail in section 1.2.2). The last step of egg-sperm interaction is the fusion of sperm and egg plasma membranes (Gilbert, 2000) (Figure 1.1a).These five steps of gamete interactions are illustrated in figure 1.1b.

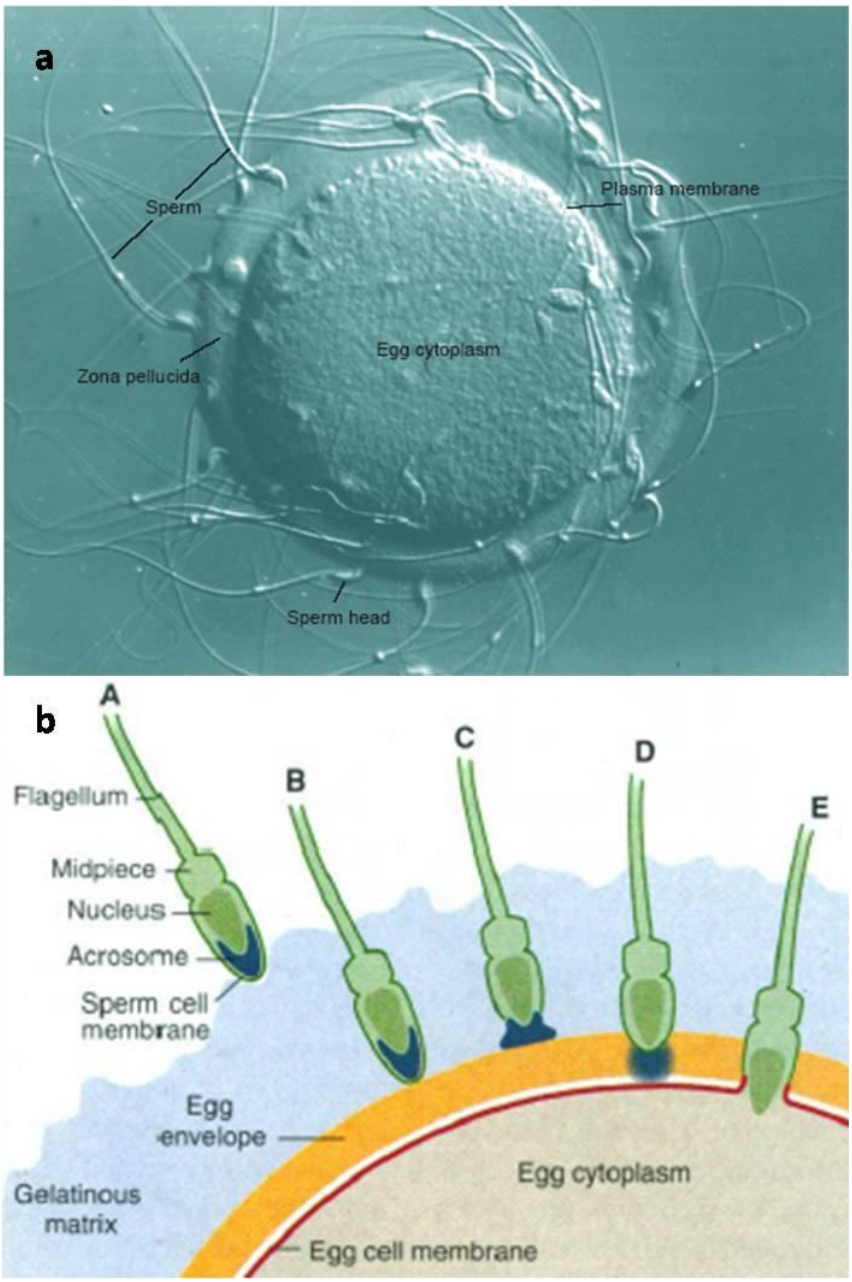


Figure 1.1. Gamete fusion in mammals.

a) Light photomicrograph of mouse sperm bound to the ZP of an unfertilized mouse egg *in vitro* taken from Wassarman et al., 2001.

b) Five crucial steps of animal fertilisation starting from A) sperm cell is attracted to the egg by chemical cues, B) the sperm binds to the egg envelope, C) the acrosomal reaction occurs releasing the contents of acrosomal vesicles, D) the sperm reaches the egg cytoplasm, E) the membranes and nucleus of the two cells fuse and activate zygote development. Taken from Vacquier, 1998.

Binding of sperm to the egg plasma membrane is thought to be mediated by a member of the ADAM family of transmembrane proteins on mammal sperm and their receptors, oocyte integrins (Kaji and Kudo, 2004; McLeskey et al., 1998; Wassarman et al., 2001). Moreover, CD9, a member of the tetraspan superfamily of integral plasmamembrane proteins and GPI-anchored proteins on the oocyte have been found to be candidate molecules involved in sperm-egg fusion (Wassarman et al., 2001). CD9 homozygous-null female mice, but not males, exhibited severely reduced fertility (Miyado et al., 2000). *TRP-3* (transient receptor potential), a member of the cation channel superfamily of proteins, has recently been identified in the sperm cells of *C. elegans* and to be involved in the egg-sperm interaction. TRP-3 is initially localised in intracellular vesicles, and then translocates to the plasmamembrane during sperm activation (spermiogenesis), a final stage of spermatogenesis where spermatids differentiate into motile spermatozoa (Xu and Sternberg, 2003). Mutations in *TRP-3* lead to sterility in both hermaphrodites and males due to a defect in their sperm. This TRPC family of channels has been proposed to function in mediating calcium influx during sperm-egg plasma membrane interactions (Xu and Sternberg, 2003). The proteins mentioned above represent a few examples of important egg-sperm fusion molecules in animals, many more have been identified but cannot be covered in detail here - the reader is referred to the following review articles for more details (Evans and Kopf, 1998; McLeskey et al., 1998; Myles and Primakoff, 1997; Rubinstein et al., 2006; Wassarman, 1999; Wassarman et al., 2001; Wassarman et al., 2005). Unfortunately, gamete fusion in plants is still relatively poorly understood. Only one Arabidopsis gene, *HAP2*, which encodes a transmembrane protein, has been identified to be involved in the egg-sperm fusion (Mori et al., 2006; von Besser et al., 2006), and the mechanism by which HAP2 operates in this event is still unclear.

In vivo eggs cells normally encounter multiple sperm simultaneously and the entrance of multiple sperm into an egg (polyspermy) leads to disastrous consequences in most organisms. Therefore the prevention of polyspermy is crucial. There are two blocks to polyspermy in animal systems which act at different rates. The fast block involves electrical alteration of the egg cell membrane by sodium ions and occurs immediately after sperm entry. The slow block is mediated by calcium ions and causes the egg extracellular matrix to be hardened, creating the fertilization envelope (Gilbert, 2000; Wong and Wessel, 2006). Although the potential for polyspermy in angiosperms is much reduced two sperms are still released from the pollen tube into the vicinity of the egg cell. Interestingly, *in vitro* fertilisation assays in maize have revealed that cell wall formation is initiated by Ca^{2+} influx 30s after egg-sperm fusion suggesting that a similar block to polyspermy may operate in plants (Kranz and Scholten, 2008; Kranz et al., 1995).

1.1.3 Animal sperm cell studies and applications

Animal sperm cells have been studied intensively in various species for many years. Numerous transciptome studies have been performed (Wu et al., 2004) and various proteins in both sperm and egg cells involved in gamete interaction have been identified, as mentioned previously (section 1.1.2). Recently developed proteomic techniques are becoming more widely used in such studies as these provide accurate information on the protein content of the cell unlike transriptomics where untranslated products of the genome are also identified. Sperm cells have been suggested to be an ideal cell type for proteomic studies as they contain only one set of chromosomes (Ainsworth, 2005; Karr, 2007). Moreover, large numbers of animal sperm cells can be easily obtained for such studies (Dorus et al., 2006). On the other hand, study of plant sperm is more problematic as these cells reside within a robust pollen grain, each of which contains only two sperm cells. Proteomic studies of isolated sperm cells free from seminal fluid have been reported in human (Johnston et al., 2005) and fruit fly, *Drosophila melanogaster* (Dorus et al., 2006). The *D. melanogaster* study was a breakthrough in sperm proteomics as it utilised whole sperm mass spectrometry (WSMS) of sperm cells (Dorus et al., 2006). This study not only identified all 16 sperm proteins previously reported to affect sperm morphology and function but also identified 60 fold more sperm proteins that has been previously characterised (Dorus et al., 2006). Importantly the WSMS technique used by Dorus and colleagues could potentially be adapted for the analysis of plant sperm cells - plant sperm protein yields tend to be low due to complicated process of sperm isolation and purification and thus WHMS, which does not require 2DE, could minimize protein loss experienced during conventional 2DE/MS.

Interestingly, animal sperm cells have been utilised for introducing foreign DNA into egg cells (Birnstiel and Busslinger, 1989; Lavitrano et al., 1989; Pittoggi et al., 2006). This technique involved introducing a DNA fragment (>5kb) into live sperm cells by simple co-incubation (Lavitrano et al., 1989). The technique was demonstrated in many animal species, e.g. mouse and *Xenopus laevis* (Lavitrano et al., 1989; Pittoggi et al., 2006). However, whether this sperm-mediated gene transfer technique will be possible in plant systems is still an unknown. In Arabidopsis both sperm and egg cells are difficult to isolate and manipulate, moreover, Agrobacterium-mediated gene transfer is efficient and well established in this species. Nonetheless, this technique could possibly be used for gene transfer in crops species such as rice and maize as they contained relatively large gametes that are easily manipulated and *in vitro* fertilisation has also been proven to be successfully performed in maize (Kranz and Lorz, 1993). Further, a

plant regeneration system has been successfully developed for *in vitro* fertilised maize embryos which utilises a feeder system (Kranz and Kumlehn, 1999; Kranz and Lorz, 1993; Kranz et al., 1995).

1.2 Plant sperm biology

As mentioned earlier in section 1.1, animal sperm cell studies could potentially facilitate flowering plant sperm cell studies as some similarities were well documented between these two kingdoms. Similarly, by understanding sperm cell structure and the fertilisation process of non-seed plants could also be useful for flowering plant sperm cell research not only that the non seed plant and angiosperm are more closely related but the sperm cell of non-seed plants are also more accessible. Therefore, the knowledge of sexual reproduction process either from animal or lower plants could be tremendously helpful for understanding flowering plant fertilisation process and could facilitate sperm cell studies. Thus, the structure and reproduction process of non-seed plants and gymnosperms will be described and compared to that of the animal and angiosperm systems in order to provide an insight in reproduction process across kingdoms.

1.2.1 Primitive plants

Sperm cells in primitive non-seed (algae, mosses, ferns) and primitive seed plants (gymnosperms; *Ginkgo*, cycads) are more similar to animal sperm cells than those of angiosperms in terms of their structure. The sperm cells of non-seed plants contain flagella required for movement. Fern sperm cells are coiled with an elongated nucleus extending along much of the cell, whereas ginkgo and cycad sperm cells are spherical with a flagellated band at the anterior end (Southworth and Cresti, 1997). The nuclear structure of fern sperm cells is condensed and like animal sperm, some ferns i.e. *Marsilea*, *Pteridium*, and *Marchantia* have highly basic protamine replacing histones for chromosome packing (Southworth and Cresti, 1997). Interestingly, fern spermatogenesis also involves a microtubule organising centre (MTOC) that is biochemically similar to animal MTOCs but is unique among plant cells as the MTOC is absent in flowering plants (Banks, 1999; Brown and Lemmon, 2007; Murata et al., 2007). Moreover, in primitive plants water is required for flagellated sperm to reach their target, the egg cell. Similar to the situation in animals, the sperm cells of ferns were found to be attracted to the egg cell by chemoattractants. In *Pteridium aquilinum* and *Ceratopteris*, malic and other organic acids act as the sperm chemoattractants (Banks, 1999). In gymnosperms, sperm cells are typically non-motile and located in the

pollen grains. However, in some gymnosperms, such as cycads and ginkgo, sperm cells are motile even though they are still sited in the pollen. The sperm cells of these plants are released from the pollen tubes and freely swim to the egg in the fluid of the egg-containing archegonia suggesting that chemoattraction is likely to play a significant role in fertilisation (Pettitt, 1977).

Although now not considered as plants the algae have been extensively studied at the reproductive level. In particular sperm-egg recognition in the brown marine algae (fucoids) received much attention in the 1980's and revealed the importance of lectins and lectin receptors in this process (Callow, 1985; Catt et al., 1983). The binding of two lectins, concanavalin A and RCA120 (*Ricinus communis* agglutinin), and FBP (fucose-binding protein from *Lotus tetragonobolus*) to gametes of *Fucus serratus* was studied. Concanavalin A and RCA120 were found to bind strongly to egg surfaces, as detected using fluorescent labels, but not to sperm and their binding to eggs inhibited fertilization. On the other hand, fucose-binding protein bound only weakly to eggs but strongly to sperm, again causing inhibition of fertilisation (Callow, 1985; Catt et al., 1983). These inhibition experiments suggest that specific sugar residues may be involved in egg-sperm recognition (Callow, 1985; Catt et al., 1983). However, to date there is still no concrete evidence to support a specific role for lectins and lectin receptors in egg-sperm recognition.

The polyspermy block has also been investigated in algae and plants. Such a block was found in the brown alga Fucus and to operate by a similar mechanism to that of animals. A fast block to polyspermy is electrical in nature and involves changes in the membrane potential of the egg cell (Brawley, 1991). An intermediated block may also operate involving enzymatic destruction of sperm receptors in the egg plasma membrane. Finally, a slow block operates via cell wall formation (Brawley, 1991; Santelices, 2002) which is similar to the proposed polyspermy block in angiosperms (see section 1.2.2). In the fern, *Marsilea vestita***,** after spermatozoid penetration, a new extracellular layer appears above the surface of the egg, beginning in the region of sperm penetration and spreading across the top of the egg. This layer may be important in preventing other spermatozoids from fusing with the egg (Myles, 1978) and again is similar to the rapid formation of a cell wall in maize eggs following fertilisation (see section 1.2.2).

Thus non-seed plant sperm cells and gamete interactions demonstrate many similarities to the animals. However in the evolutionary jump to gymnosperms, sperm cells become resident in the pollen grains although flagella often still remain. This special characteristic of sperm cells residing in the pollen grain clearly has an advantage in fertilisation. The most advanced sperm-pollen system is possessed by angiosperms has proven to be highly evolutionarily successful and along with other reproductive traits

has facilitated rapid adaptive radiation of this plant group. It is highly likely that some of the sperm proteins present in primitive plants will still be conserved in angiosperms along with mechanisms involved in gamete interaction. By understanding these processes in primitive plant groups their study in angiosperms studies are likely to be facilitated.

1.2.2 Angiosperms

Aflagellated sperm cells of angiosperms have evolved to reside in the pollen grains. The male gametophyte (pollen), which provides nutritional support for the male gametes, protects the sperm cells from the external environment and also acts as the vehicle to deliver the sperm to the female gamete to facilitate successful fertilization (Russell, 2003; Weterings and Russell, 2004b). Mature pollen grains are composed of three cells, a vegetative cell that encloses the two sperm cells (Weterings and Russell, 2004b). Pollen and sperm cells are generated by the processes called microsporogenesis and microgametogenesis (Figure 1.2). Microsporogenesis begins with the pollen mother cell undergoing meiosis I and II to give rise to the tetrad. Following meiosis II microgametogenesis starts with each uninucleate cell in the tetrad separating from the tetrad and undergoing asymmetric mitosis I producing the immature pollen grain which contains unequal daughter cells – the pollen vegetative cell and generative cell. These two daughter cells are completely different in their structure and developmental fates. The vegetative cell is large compared to the generative cell, and accumulates an abundance of stored metabolites required for pollen tube growth. The generative cell is enclosed in the vegetative cell and by comparison contains relatively few organelles and stored metabolites (Park et al., 2000). The generative cell in the immature pollen grain then undergoes mitosis to give rise to the two sperm. The timing of this last step varies among species. For example in the Brassicaceae (including Arabidopsis) and *Zea may*s (maize), mitosis II occurs before pollen germination resulting in tricellular mature pollen, whereas in plants with bicellular pollen mitosis II occurs during pollen tube growth through the stylar tissue, for instance in *Lilium longiforum* (lily) and *Nicotiana tabaccum* (tobacco) (Russell, 1991; Xu et al., 2002). Many genes have been discovered to be involved in the microsporogenesis process by mutational analyses. These male gametophyte mutants are summarised in table 1.1.

As female gametes inevitably interact with the male gametes, gaining a detailed understanding of these cells, particularly the proteins present on their surfaces will be of crucial importance for research into sperm biology. Moreover, some of the genes involved in female gametogenesis could possibly also be involved in male

gametogenesis (Mathilde et al., 2003). However, so far, no egg cell proteins have been identified to be involved directly in gamete fusion. Nonetheless, female gametogenesis and some female gametophytic mutants will be briefly mentioned in this chapter to provide a general knowledge of the sperm fusion partners, the egg and central cell. The female gametophyte typically comprises seven cells of four different cell types, three antipodal cells, one central cell with two nuclei, two synergids and one egg cell (Yadegari and Drews, 2004). As for the production of the male gametes female gametophyte development occurs in two major phases, megasporogenesis and megagametogenesis (Christensen et al., 1998). Megasporogenesis initiates with a diploid megaspore mother cell. This megaspore, in the most common type of plant female gametophyte development (monosporic), undergoes meiosis and gives rise to four haploid megaspores (Yadegari and Drews, 2004) - three megaspores then undergo cell death. Megagametogenesis proceeds with this remaining haploid megaspore undergoing three rounds of mitosis without cytokinesis. Two nuclei from each pole migrate to the centre of the cell and fuse together forming the central cell. The three nuclei at the micropylar end form the egg cell and synergid cells. The other three nuclei at the chalazal axis become the antipodal cells which undergo programmed cell death before fertilisation in Arabidopsis (Marton and Dresselhaus, 2008; Murgia et al., 1993). During cellularisation, cell walls develop around these nuclei creating a seven-celled female gametophyte (Christensen et al., 1998; Yadegari and Drews, 2004) (Figure 1.3). Several female gametophyte mutants have been isolated from many plant species including Arabidopsis. Some important examples of female gametic mutants are described in table 1.1.

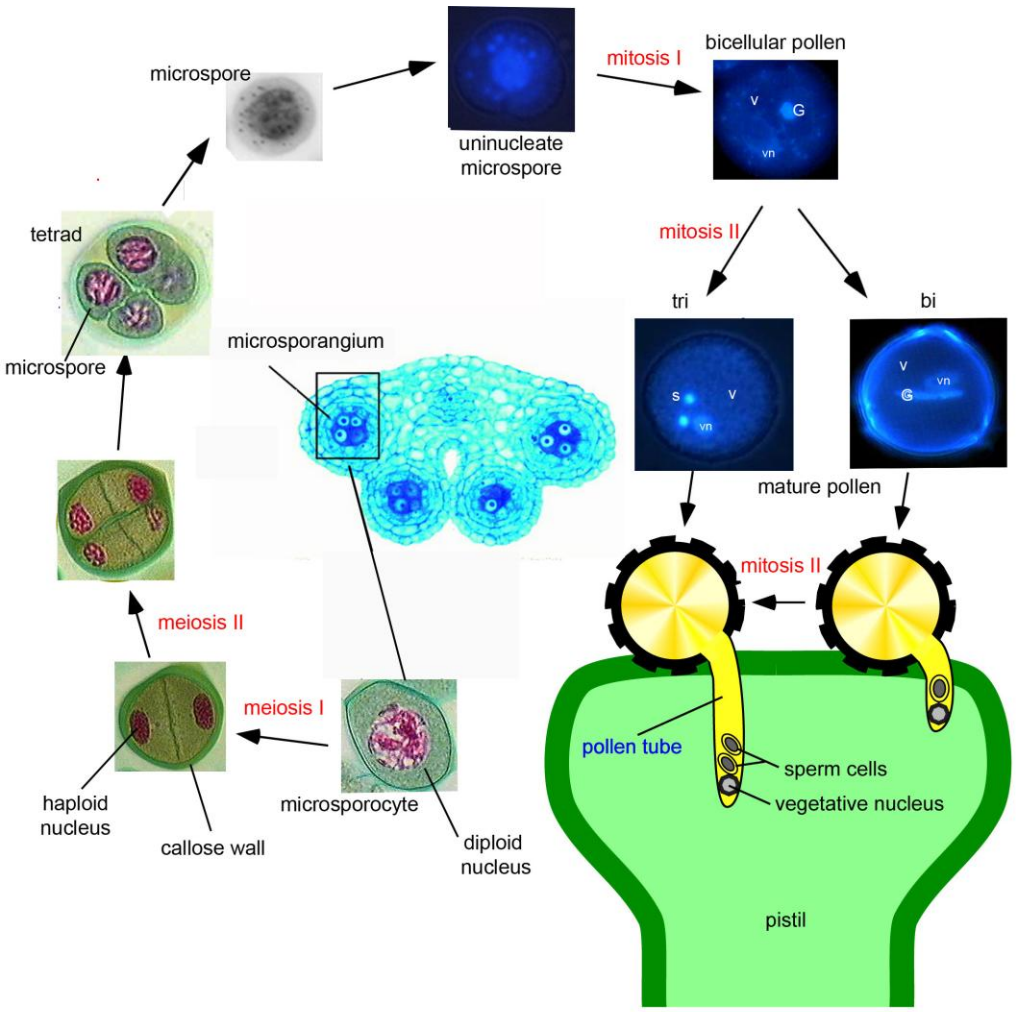


Figure 1.2. Male gametophyte development (Honys and Twell, 2004a).

Microsporogenesis starts with the microspore mother cell undergoing meiosis I and meiosis II to generate the tetrad. This tetrad separates releasing four haploid microspores; these uninucleate microspores undergo an asymmetric mitosis I to give rise to the small generative cell residing in the large vegetative cell (bicellular microspore). A second division of the generative cell, mitosis II, may then follow resulting in two sperm cells in mature pollen (tricellular pollen). However, in species with bicellular pollen this generative cell division (mitosis II) occurs after the pollen grain germinates and initiates growth down female stylar tissue (V: vegetative cell, vn: vegetative nucleus, G: generative cell, S: sperm cell, tri: tricellular pollen species, bi: bicellular pollen species) (Figure adapted from Honys and Twell, 2004a).

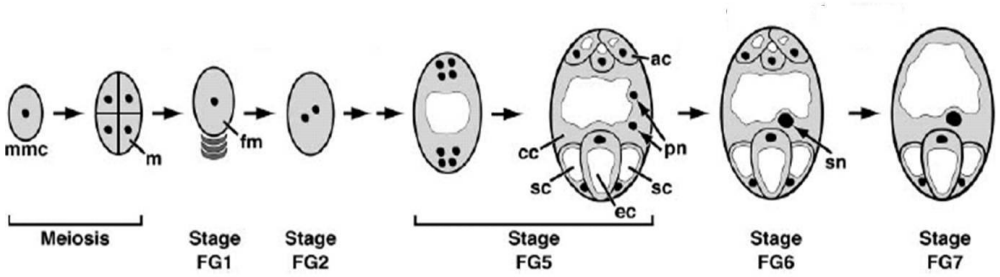


Figure 1.3. Female gametophyte development of the most common 'monosporic' pattern.

Megasporogenesis initiates when a diploid megaspore mother cell undergoes meiosis and gives rise to four haploid megaspores. Three megaspores then undergo cell death therefore only one haploid megaspore remains. Megagametogenesis occurs when this remaining haploid megaspore undergoes mitosis without cytokinesis. Two nuclei from each pole migrate to the center of the cell and later fuse together forming the central cell. The three nuclei at the micropylar end become the egg cell and synergid cells. The other three nuclei at the chalazal axis turn into antipodal cells (mmc: megaspore mother cell, m: megaspore, fm: functional megaspore, cc: central cell, sc: synergid cell, ec: egg cell, pn: polar nuclei, ac: antipodal cell, sn: secondary nucleus figure reproduced from Yadegari and Drews, 2004.

1.2.2.1 Sperm cell structure

Seed-plant sperm cells are small, frequently less than 5 µm in diameter, and plasma membrane bound without cell walls (Engel et al., 2003; Russell, 1996). These specialised plant cells are unique not only in that they are lacking cell walls but also because they contain a distinct type of microtubule bundles, small cytoplasmic volume, and condensed chromatin (Southworth and Cresti, 1997). Unlike non-seed plants and some gymnosperms, angiosperm sperm cells lack flagella, therefore are non-motile and require the pollen tube to deliver them to the female gametophyte (Russell, 1996). The sperm cells are typically connected to one another and physically associated with the vegetative nucleus as one male germ unit (Xu et al., 2002). This male germ unit clearly increases the effectiveness of gamete delivery and ensures nearly simultaneous transmission of the two sperm cells into the receptive embryo sac. This association is present in most species of flowering plants studied to date or is formed soon after pollen germination (Russell, 1991). The two sperm cells in several species have been found to be physically different. In these dimorphic sperm species, one sperm cell called "Svn" is physically associated with the vegetative nucleus, whereas the other sperm cell named "Sua" is unassociated with the vegetative nucleus, but usually linked to the Svn. (Tian et al., 2001). Dimorphic sperm have been reported in many plant species, for example in *Nicotiana tabaccum* Sua sperm was measured to be 30% bigger than Svn sperm. In *Plumbago zeylanica*, the most completely described dimorphic species, the Sua receives most (and frequently all) of the plastids, whereas the Svn contains few (typically no) plastids and five times the number of mitochondria (Tian et al., 2001). Moreover, different transcript levels of ubiquitin between the two sperm cells was also detected in *Plumbago zeylanica* reflecting metabolism occurrence in these cells (Singh et al., 2002). This polyubiquitin gene was found to be preferentially expressed in the Svn (Singh et al., 2002). Sperm dimorphism has been shown to link to preferential fusion between sperm-egg and sperm-central cell during double fertilization. In *P. zeylanica* the smaller sperm cell with plastids (Sua) prefers to fuse with the egg cell, whereas the bigger sperm containing more mitochondria (Svn) fuses with the central cell. This pattern occurs in over 95% of the fertilisation events observed in *P. zeylanica* (Russell, 2003). Considering the significant differences in surface area and volume of sperm egg and central cells, there may be biophysical reasons for preferential fusion of the sperm cells (Tian et al., 2001). In *Zea mays*, the two sperm cells are morphologically identical but the sperm cell that tends to fertilize the egg cell contains B chromosomes, the supernumerary chromosomes that are not homologous to any of the normal set of chromosomes (A chromosomes) (Engel et al., 2003; Faure et al., 2003; Shi et al., 1996). Even though sperm dimorphism has been detected in some species, the gene expression differences in the two sperm cells are still broadly unknown (except

recently in *P. zeylanica*) due to the inaccessibility and low transcriptional activity of the sperm cells (Xu et al., 2002; Zhang et al., 1998). In an Arabidopsis relative, *Brassica oleracea*, sperm size dimorphism has also been reported (McConchie et al., 1987b). The Svn was reported to have 18.4% larger in volume and 41% larger in surface area than Sua(McConchie et al., 1987b). However, whether preferential fusion occurs in this species is still unknown. As *B. oleracea* has been demonstrated to exhibit sperm cell dimorphism, it raises the interesting question whether its relative, *A. thaliana* may also have dimorphic sperm - sperm dimorphism in Arabidopsis has not been reported to date and forms part of the investigation in this thesis (Chapter3).

1.2.2.2 Pollen tube guidance

As mentioned above angiosperm sperm cells are delivered to female gametophyte via the pollen tube (Figure 1.4). Complex processes of male gamete delivery via the pollen tube to the female gamete are well established. Many steps are involved in this process starting from pollen-stigma interaction to the final step when the pollen tube enters the micropyle of the ovule. Various molecules on both the male and female sides are involved in this complex process. One exceedingly interesting protein, HAP2, identified as sperm cell-specific was discovered to function not only in gamete fusion but also to be involved in pollen tube guidance (Mori et al., 2006; von Besser et al., 2006). This dual functionality could be a way for the plant to prevent pollen grains carrying defect sperm cells from fertilising the female gamete. Therefore, other sperm cell specific proteins yet to be identified could possibly have a similar dual function to HAP2. By understanding the process of gamete delivery, it could shed light on the connection between these two functions of HAP2 and facilitate identification of similar sperm cell-specific proteins potentially involved in gamete fusion.

The early processes of pollen germination and tube growth will not be discussed in detail here as these events mainly involve the physical interaction between sporophytic tissues of the female and the pollen, see reviews (Higashiyama et al., 2003; Kim et al., 2004; Lord and Russell, 2002; Russell, 1996; Sanchez et al., 2004). Briefly, pollen grains arrive at the stigma surface and become hydrated. Then the germinating tubes penetrate the stigmatic tissue and grow through the style towards the ovary. In Arabidopsis, pollen tubes emerge from the septum and then navigate up the funiculus and into the micropyle of the ovule (Palanivelu et al., 2003; Swanson et al., 2004). The plant ovule encloses seven cells within, three antipodal cells at the chalazal pole, the central cell in the middle, and the egg cell at the micropylar pole between the two synergid cells (Barrell and Grossniklaus, 2005). This final stage of pollen tube growth is

directed by an as yet uncharacterised chemoactractant released from the ovule / female gametophyte, a situation similar in concept to the animal systems mentioned earlier. This signal required for attracting the pollen tube to the micropyle (Swanson et al., 2004) was demonstrated by Higashiyama and colleagues (2001) to be released from the synergid cells. Laser ablation of synergids elegantly demonstrated that they are the source of the final attraction signal (Higashiyama et al., 2001; Swanson et al., 2004). Moreover, previous morphological studies of synergid cells demonstrated that they have properties of secretory cells and could therefore be a source of guidance molecules (Shimizu, 2000).

After the pollen tube enters the micropyle of the female gametophyte and arrives at one of synergid cells, the tip of pollen tube is arrested and ruptures releasing the sperm cells, the synergid cell also ruptures. The timing of synergid degeneration has received some attention and in some species occurs before pollen tube arrival. In *Torenia fournieri*, synergid degeneration may be triggered by the force of pollen tube discharge (Higashiyama et al., 2000). However, in the Arabidopsis *feronia* mutant, pollen tubes fail to rupture and release the sperm cells despite normally timed degeneration of the synergid cells suggesting that, in this species at least, the pollen tube does not trigger synergid degeneration (Huck et al., 2003). Interestingly this mutant did not only fail to discharge sperm cells but also failed to prevent other pollen tubes entering the ovule indicating that this mutant lacks communication signals between the male and female (Huck et al., 2003). The proposal of pollen tube-synergid signalling was confirmed when FERONIA was identified as a receptor kinase expressed specifically in synergids (Berger, 2008; Escobar-Restrepo et al., 2007). It is envisaged that FERONIA perceives a signal from the pollen tube which then triggers a signalling pathway leading to release of the sperm. After pollen tube bursting at the synergid cell, the mechanism by which the two sperm cells reach their targets is still undefined. The force of tube disruption was believed to be sufficient to send both cells to their partners (Berger, 2008). However, actin coronas have been detected in the receptive synergid extending to the both the egg and central cell. These are proposed to be involved in actomyosin-mediated sperm transport suggesting an active migratory mechanism of the two sperm cells (Marton and Dresselhaus, 2008; Ye et al., 2002). However whether there are any actin interacting molecules expressed specifically on the sperm cell surface has yet to be determined.

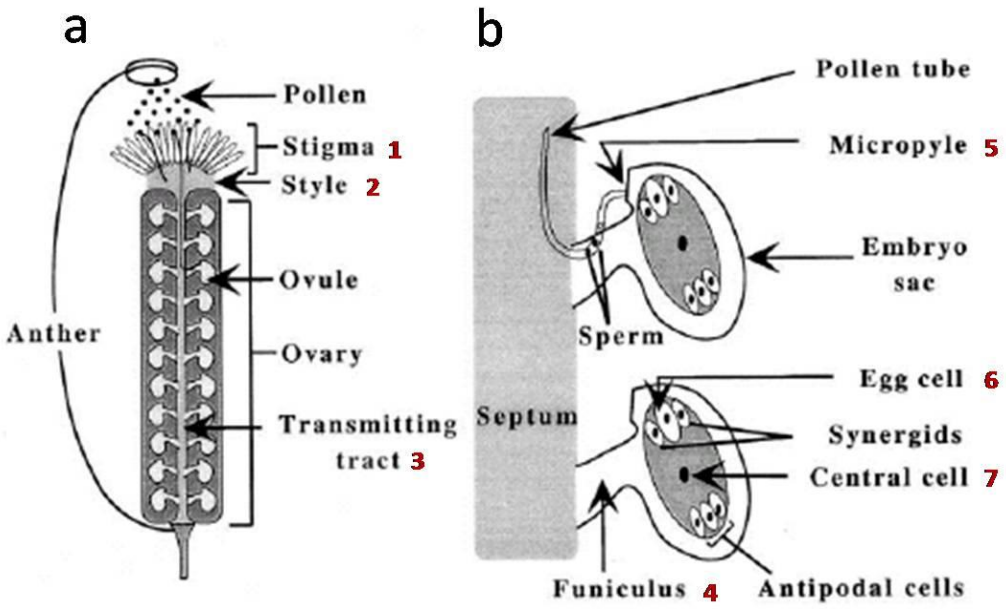


Figure 1.4. Pollen tube guidance in a model plant, Arabidopsis.

Pollen grains released from the anther arrive at the stigma surface in a dehydrated state and become hydrated via papilla cells (1). Then the germinating tubes penetrate the stigma and grow through the style (2). After leaving the style, pollen tubes of many species navigate across the surfaces of the pistil cells until they reach the ovule. In Arabidopsis, pollen tubes travel down the transmitting tract (3) and emerge from transmitting tract near the unfertilized ovules and continue to grow on the funiculus surface (4). The pollen tube then enters the micropyle (5) of the female gametophyte and arrives at one of the synergid cells. The tip of pollen tube is arrested and ruptures, the synergid cell also ruptures and sperm cells are then able to fuse with the egg (6) and central cell (7). Figure taken from Palanivelu et al., 2003.

1.2.2.3 Egg-sperm fusion

After the two sperm cells have been delivered into the embryo sac, double fertilization occurs. This double fertilization phenomenon is the main feature of sexual reproduction in angiosperms and occurs only in this group of plants. It involves the interaction of the male gametes (sperm cells) with the female gamete, the egg, and the central cell (Russell, 2003; Weterings and Russell, 2004a). One sperm fuses with the egg cell to form the diploid embryo (zygote), whereas the other sperm combines with the central cell to form nutritive endosperm (Russell, 2003). A mutation in the A-type cyclin-dependent kinase (CDKA1) gene of Arabidopsis leads to pollen that is unable to perform mitosis II and therefore only one sperm-like cell is produced in place of normal two sperm cells. Interestingly this sperm-like cell can still fuse to the egg cell and generate an embryo, however, the polar nuclei are left unfertilised and thus endosperm fails to develop resulting in seed abortion (Iwakawa et al., 2006). One important question relating to double fertilization is whether the sperm cells fuse randomly to their fusion partners or whether there is preferential fusion. The *cdka1* mutation demonstrated that the sperm cell preferentially fuses to the egg cell rather than the central cell. This apparent preference is possibly due to the position of the egg cell being in close proximity to the micropylar opening and thus would be the first fusion partner that the sperm cell would come into contact with (Iwakawa et al., 2006; Marton and Dresselhaus, 2008). In plants with dimorphic sperm, the experimental data indicates that the sperm cell with plastids (Sua) prefers to fuse with the egg cell, whereas the other sperm (Svn) fuses with the central cell as described earlier (Russell, 2003). The explanation for this preferential fusion could either be that the sperm cell fate was determined during cell cycle or as simple as Sua being the first sperm cell to arrive in the ovule as was suggested in *cdka1* mutant. A similar situation may well occur in plants with monomorphic sperm, Faure suggests that fusion could be random and depend largely on the positioning of sperm cells in the pollen tube on discharge into embryo sac (Faure et al., 2003). Nonetheless, the other mutation affecting pollen mitosis II (PMII) similar to *cdka1*called *msiI* demonstrated non-preferential fusion of sperm cells (Chen et al., 2008). The defects of *msiI* pollen was proven to be a result from chromatin assembly factor1 (CAF) loss of function. As this mutation exhibited no preferential fusion, the sperm cell fate was concluded not to be determined during cell cycle (Chen et al., 2008). However, interestingly, the PMII in this *msiI* mutant is not completely disrupted but delayed as demonstrated in some of the mutant pollen tubes, the single sperm cells in the germinating pollen tube can be divided into two sperm cells which are able to normally fuse to egg and central cells (Chen et al., 2008). Therefore, it is likely that the single sperm cells produced in this mutant are incompletely developed and their fates are still undetermined unlike in *cdka1* mutant.

Sperm-egg interaction in animals is characterised by rapid increases in intracellular Ca^{2+} associated with egg activation. Remarkably Ca^{2+} ion fluxes have also been described on sperm-egg interaction *in vitro* for maize (Digonnet, 1997; Franklin-Tong, 2002; Kranz and Scholten, 2008). This Ca^{2+} event during gamete fusion suggests possible fundamental similarities between plant and animal sperm-egg fusion thus knowledge of the more highly characterised mechanism in animals could facilitate studies on syngamy in plants. In plants, very few sperm-borne proteins have been identified that potentially have a role in egg-sperm fusion. AtGEX2 was the first Arabidopsis sperm cell-expressed protein reported and being a transmembrane protein located in the sperm cell plasmamembrane the authors speculated that it might be involved in fertilisation (Engel et al., 2005). However, as mentioned previously to date only one sperm cell-specific protein GCS1 (HAP2) (Mori et al., 2006; von Besser et al., 2006) has been confirmed to be crucial for fertilisation. Interestingly, this protein also appears to have a crucial role in pollen tube guidance - a novel role for sperm cells (von Besser et al., 2006) (see detail in section 1.2.3). Many questions concerning sperm cell-specific proteins and fertilisation events are still waiting to be answered; for example whether the cytoplasmic content of sperm cells are transmitted to the next generation following sperm-egg fusion (Russell, 1983), whether the dimorphic sperms are common in angiosperms, whether preferential fusion really exists in plants and whether the preferential fusion is determined by the molecules on the sperm cell surface, and whether the two sperm cells carrying the same cargo for each of their fusion partners.

In vitro sperm egg fusion experiments have been successfully carried out in several plant species, for example rice and maize (Khalequzzaman and Haq, 2005; Kranz and Lorz, 1993). An artificial *in vitro* fertilization technique used widely is electrofusion, however in order to mimic the natural process of fertilisation Ca^{2+} has been used to facilitate the process (Digonnet, 1997; Faure et al., 1994; Khalequzzaman and Haq, 2005). Further research utilising *in vitro* fusion assays will undoubtedly advance understanding of egg-sperm interactions, for instance morphological changes occurring during fertilisation as well as signalling cascades could be monitored as real-time events. Importantly Con A binding sites have been found on the surface of egg and central cells in *Nicotiana tabaccum* which could be the site of sperm recognition (Ke-Feng et al., 2005b). Similar lectin binding sites have also found in the animals and lower plants as described earlier. However, despite these tantalising similarities between different biological systems very little is known about the sperm-egg recognition in plants. Therefore there is clearly an urgent need for research in this area utilising a genetically tractable plants system.

Blocks to polyspermy in angiosperms occur at several stages of gamete delivery presumably to ensure that multiple sperm cells do not fuse with the egg. The first block operates at the final stage of pollen tube growth - a diffusible signal from synergid cell of the female gametophyte guides the pollen tube to enter the micropyle (see section 1.2.2.2) upon which the synergid degenerates, the attractant signal is lost and thus additional pollen tubes are effectively prevented from seeking the egg (Huck et al., 2003). However loss of attractants is not the only factor responsible for the prevention of polyspermy but 'repulsion' signals also play a role (Scott, 2007; Shimizu and Okada, 2000). These repulsion signals were hypothesised to be released either from competing pollen tubes or from female gamete to prevent polyspermy as demonstrated in *magatama (maa)* mutants where the pollen tubes lost their way just before entering the micropyle and the mutant female gametophytes often attracted more than one tubes (Scott, 2007; Shimizu and Okada, 2000). Similar to animal systems, a final block to polyspermy occurs in plants after gamete fusion involving changes to the fertilised egg cell membrane - fertilised egg cells of maize were demonstrated to undergo rapid cell wall formation preventing further sperm fusion events polyspermy (Kranz and Scholten, 2008; Kranz et al., 1995). This block could possibly occur particularly to prevent the second sperm from the same pollen tube fusing with the fertilised egg cell.

1.2.3 Plant sperm cells studies and applications

Plant sperm cell studies have been carried out in many species i.e. *Zea mays* (Dupuis et al., 1987; Engel et al., 2003), *Plumbago zeylanica* (Russell, 1984; Russell, 1986), *Nicotiana tabacum* (Yu et al., 1992), *Lilium longiflorum* (Okada et al., 2006; Okada et al., 2007; Xu et al., 1999b), *Brassica oleracea* (Matthys-Rochon et al., 1988; Matthys-Rochon et al., 1987; McConchie et al., 1987b), and *Arabidopsis thaliana* (Engel et al., 2005; Mori et al., 2006; von Besser et al., 2006). Construction of cDNA libraries has been performed from semi-purified sperm cells of lily (Xu et al., 1999b; Okada et al., 2006), tobacco (Xu et al., 2002), and *P. zeylanica* (Gou et al., 2005). Unfortunately these libraries represented not only sperm cell transcripts but also the transcripts derived from contaminating pollen vegetative cells. However, a few sperm cell-specific genes i.e. *LGC1* (Xu et al., 1999b) and *gcH2A* and *gcH3* (Xu et al., 1999a) of lily, have been characterised from these studies. Fluorescence-activated cell sorting was utilised in maize to obtain pure sperm cells free from pollen cytoplasmic (vegetative cell) contamination (Engel et al., 2003). EST data were generated from these Fluorescence-activated cell sorting (FACS) purified maize sperm cells (Engel et al., 2003). These available EST data were utilised for a cross-species *in silico* comparative study in this project which will be covered in detail in chapter 5. Two sperm cell-expressed

homologues in Arabidopsis, *GEX1* and *GEX2*, were identified subsequently to this study (Engel et al., 2005). Sperm cell-specific gene *GCS1* (*HAP2*) (Mori et al., 2006; von Besser et al., 2006) was discovered using a differential display approach in lily and found to have a crucial function in reproduction. Pollen tubes carrying defective sperm cells in *+/gsc1* Arabidopsis mutants entered the ovule but the sperm cells failed to fuse with the female gametes (Mori et al., 2006). GCS1 function was also identified to be involved in pollen tube funicular/micropyle guidance as twofold fewer *hap2* pollen tubes fail enter the micropyle and burst within synergids (von Besser et al., 2006). Various T-DNA insertion mutants affecting male and female gametes at various developmental stages and different stages of the reproductive process have been identified (Table 1.1). As mentioned earlier, to date only one gene has been proven to be involved in the gamete fusion event. The vast majority of the proteins identified are involved in gametogenesis possibly because of **i**) a very limited nember of proteins are required for gamete fusion, **ii**) mutations in genes involved in gamete fusion affect both gametes thus plant lines for this particular gene knockout are unable to be recovered. In many cases mutated genes impacting on the female gamete also affect the pollen tube, however most of this effect is probably explained by a disruption of signals released from the female gamete to attract pollen tubes.

Plant sperm cell studies clearly provide a better understanding of complex plant reproductive processes. The *hap2* mutant mentioned above revealed exciting novel dual roles of a sperm cell-specific protein in gamete fusion and prevention of the defective gamete reaching the target (thus avoiding a wasted pollination and permitting pollen carrying a normal gamete to effect fertilisation). Nonetheless whether this is a common feature of proteins crucial for gamete fusion or not, other proteins involved in gamete fusion are needed to be identified. These sperm cell studies also offer valuable tools for other aspects of reproductive biology. In 1996, Southworth and Kwiatkowski generated monoclonal antibodies, JIM8 and JIM13 from *Brassica oleracea* and lily that bound to the sperm cell surface (Southworth and Kwiatkowski, 1996). Moreover, in 1997, Xu and Tsao isolated proteins from maize sperm cells that interacted with peroxidase-conjugated Con A (Xu and Tsao, 1997). These tools can be utilised as germ cell-specific markers that could facilitate other reproductive studies (Wang et al., 2006). Moreover, plant gamete isolation techniques have been greatly developed for both the male and female over the past decade. The main purpose for these developments was to extend our knowledge of plant reproduction i.e. double fertilisation. *In vitro* fertilisation techniques have also been improved to facilitate gamete interaction studies. One potentially important application of the *in vitro* fertilisation technique is for zygote microinjection. Microinjection has been used for horizontal gene transferred in various plant species. This technique has been successfully performed with protoplasts, calli

and young embryos. The basic principle of the technique is that the recipient cells need to have ability to divide and differentiate. Thus, the zygote is the ideal target cell for this manipulation (Wang et al., 2006). Another possibility of gene transfer technique that can possibly be adapted to be used in some plant species that are difficult to be transformed is through sperm cells. As described earlier (section 1.1.3) that animal sperm cells can be used as a vector for gene transfer, this technique together with *in vitro* fertilisation could possibly be utilised for plant transformation. In addition, analysis of DNA content and structure of sperm cells was utilised in evolutionary studies in non-seed plant due to the cell unique characters i.e. the small size lacking of cell wall, low cytoplasmic content , and its haploid property (Renzaglia et al., 1995). As demonstrated in this section, plant sperm research clearly not only provides a basic knowledge of sperm cell and reproduction biology but can also be applied for use in other molecular technologies.

Table 1.1. Mutation affecting plant reproduction.

Mutant	Type of mutation	Affected stage of gametophyte development	References
Male			
sidecar pollen (Scp)	pollen development	prior to the first asymmetric postmeiotic mitosis	Chen and Mccormick, 1996
gametophytic male sterile-1 (gaMS-1)	pollen development	second microspore mitosis	Sari-Gorla et al., 1996
gametophytic male sterile-2 (gaMS-2)	pollen development	second microspore mitosis	Sari-Gorla et al., 1996
limpet pollen (lip)	pollen development	generative cell inward migration after pollen mitosis	Howden et al., 1998
germini pollen 1 (gem1)	pollen development	asymmetric division and cytokinesis at pollen mitosis I	Park et al., 1998
male-sterile mutants (ms7-ms13)	pollen development	postmeiotic defect of microspore development	Taylor et al., 1998
T-DNA transmission defect (Ttd38)	pollen development	pollen meiosis I	Procissi et al., 2001
Ttd40	pollen development	pollen meiosis I	Procissi et al., 2001
gift-wrapped pollen	pollen development	early events in the pollination pathway (hydration and germination)	Johnson and Mccormick, 2001
raring-to-go (rtg)	pollen development	early events in the pollination pathway (hydration and germination)	Johnson and Mccormick, 2001
mor1/gem1	pollen development	cytokinesis	Twell et al., 2002
halfman (ham)	pollen development	mid-bicellular pollen stage after detachment of the generative cell from the pollen wall	Oh et al., 2003
duo pollen (duo)	pollen development	pollen mitosis II at G2-M transition	Durbarry et al., 2005
abnormal gametophytes (agm)	pollen development	pollen mitosis II	Sorensen et al., 2004
A-type cyclin-dependent kinase (cdka1)	pollen development	pollen mitosis II	Iwakawa et al., 2006
Atmlh1	pollen development	mitotic reconbination	Dion et al., 2007
no pollen germination1 (npg1)	pollen germination	initiation of pollen germination	Golovkin and Reddy, 2003
sec8	pollen germination	initiation of pollen germination	Cole et al., 2005
npg1	pollen germination	no germination	Golovkin and Reddy, 2003
apy1-1/apy2-1	pollen germination	no germination	Steinebrunner et al., 2003
mad4	pollen germination or tube growth	slow pollen tube growth	Grini et al., 1999
seth1, seth2	pollen germination or tube growth	reduced germination rate and short pollen tube	Lalanne et al., 2004
hapless (hap1, 18, 22)	pollen tube growth	pollen tubes fail to leave the septum	Johnson et al., 2004
hapless (hap2, 4, 24, 27)	pollen tube growth	pollen tubes growth is chaotic in the ovary	Johnson et al., 2004
hapless (hap3-9, 13, 15, 17, 20, 21, 28, 32)	pollen tube growth	pollen tubes fail to exit style	Johnson et al., 2004
Ttd26/ kinky pollen1 (kip1)	pollen tube growth	abnormal pollen tube	Procissi et al., 2001, 2003
Ttd34/ kinky pollen2 (kip2)	pollen tube growth	abnormal pollen tube (short/twisted)	Procissi et al., 2001, 2003

Ttd42/ kinky pollen3 (kip3)	pollen tube growth	abnormal pollen tube (branched at tube tip)	Procissi et al., 2001, 2003
vanguard1 (vgd1)	pollen tube growth	pollen tube growth down the style and transmitting tract tissues	Jiang et al., 2005
attaf6	pollen tube growth	slow pollen tube growth	Lago et al., 2005
Female			
short integument1 (sin1)	female gametophyte development	abnormal ovule integument development and aberrant differentiation of the megagametophyte	Lang et al., 1994; Ray et al., 1996
prolifera (prl)	female gametophyte development	embryo sac arrested at various stages	Springer et al., 1995
andarta (ada)	female gametophyte development	ovules arrested at viable megaspore stage	Howden et al. 1998
tiatrya (tya)	female gametophyte development	ovules arrested at one-nucleate megagametophyte	Howden et al. 1998
Aintegumenta (ant)	female gametophyte development	ovules arrested at megasporogenesis (single megaspore)	Klucher et al., 1996
arabinogalactan proteins18 (AGP18)	female gametophyte development	functional megaspore fails to enlarge and mitotically divide	Acosta-Garcia and Vielle-Calzada, 2004
EDA1-9	female gametophyte development	arrested at two or four-nuclear stages	Pagnussat et al., 2005
EDA10-15	female gametophyte development	abnormal nuclear numbers and positions	Pagnussat et al., 2005
EDA16-23	female gametophyte development	arrested at varying stages of embryo sac development	Pagnussat et al., 2005
EDA24-41	female gametophyte development	unfused polar nuclei	Pagnussat et al., 2005
gfa4, 5	female gametophyte development	one-nucleate female gametophyte	Feldmann et al., 1997; Christensen et al., 1998
gfa2, 3, 7	female gametophyte development	polar nuclei failed to fuse	Feldmann et al., 1997; Christensen et al., 1998
female gametophyte2, 3 (fem2, fam3)	female gametophyte development	one-nucleate female gametophyte	Christensen et al., 1998
female gametophyte1 (fem1)	female gametophyte development	female gametophyte absent	Christensen et al., 1998
female gametophyte4 (fem4)	female gametophyte development	irregular female gametophyte	Christensen et al., 1998
megatama (maa)	female gametophyte development/pollen tube attraction	polar nuclei failed to fuse/pollen failed to enter micropyle, more than one pollen tubes traveled up to an ovule (loss of polyspermy block)	Shimizu and Okada, 2000
feronia (fer)	female gametophyte development/pollen tube attraction	synergid-pollen tube signaling/pollen tube discharge	Norbert et al., 2002; Huck et al., 2003
sirene (srn)	female gametophyte development/pollen tube attraction	synergid-pollen tube signaling/pollen tube discharge	Rotman et al., 2003; 2008
Both male and female			
syn1	male and female meiosis	defects in meiotic chromosome cohesion and condensation and in maintaining cohesion at the centromeres during late stages of meiosis I	Peirson et al., 1997; Cai et al., 2003

asy1	male and female meiosis	fail to undergo extensive synapsis during early prophase I	Caryl et al., 2000
Arabidopsis homologue pairing 2 (ahp2)	male and female meiosis	defective in bivalent forming and in chromosome segregation during meiosis	Schommer et al., 2003
mei1	both gametes development	defective in DNA damage repair during meiosis	He and Mascarenhas, 1998; Mathilda et al., 2003
succinate dehydrogenase (sdh1-1)	both gametes development	embryo sacs arrested either at the two-nucleate stage or before polar nuclei fusion/microspores fail to undergo mitosis I	Leon et al., 2007
atxrcc3	both gametes development	defective in DNA damage repair during meiosis	Bleuyard and White 2004
two-in-one (tio)	male/female gametophytes development	cytokinesis at pollen mitosis I and cellularization (cytokinesis) during female gametogenesis	Oh et al., 2005

1.3 Perspectives and Aims

Despite intensive studies on plant sperm biology, the understanding of this specialised cell including the syngamy event is still very poor, particularly in the model plant Arabidopsis. To date, only one Arabidopsis sperm cell-specific protein GCS1 (HAP2) has been identified to be involved in the gamete fusion event (Mori et al., 2006; von Besser et al., 2006). The primary aim of this thesis was to further the characterisation of sperm cells from the model plant *Arabidopsis thaliana* by utilising a range of molecular and bioinformatics techniques. Sperm cell isolation and purification techniques for *A. thaliana* were intended to be developed utilising its much larger relative, *Brassica oleracea*, as a facilitator. A further aim was to obtain morphological and quantitative data for Arabidopsis sperm cells by transmission electron microscopy (TEM) and advanced confocal microscopy techniques, as this has not been described in detail before. Moreover, as the sperm cells of *Brassica oleracea* have been reported to be dimorphic (McConchie et al., 1987b), the question of Arabidopsis sperm dimorphism would hopefully be addressed through the sperm quantitative cytological analyses in this project. In addition the development of purification protocols involving Fluorescence-activated cell sorting (FACS) was seen as a priority to obtain pure sperm fractions for proteomic and future molecular analyses. In parallel, a bioinformatic approach was taken to identify Arabidopsis sperm protein candidates by carrying out a cross species comparative study. Available plant sperm ESTs (mainly from maize, but also from Plumbago, rice and tobacco ESTs) and pollen microarray data were utilised for Arabidopsis sperm-specific protein identification *in silico*. Several candidate genes obtaining from this study were analysed further to determine their precise spatial and temporal expression profiles by RT-PCR and the generation of promoter::GFP-fusion transgenic lines. A reverse-genetics approach, utilising available T-DNA insertion lines, was also taken to identify the function of candidate genes obtained from the bioinformatics study. Despite the inherent difficulties in the study of sperm cells, a combination of these techniques will almost certainly reveal meaningful information on plant sperm cell biology and also should assist in the development of useful techniques to be used in plant sperm cell research in the future.

Chapter 2:

Materials and Methods

All reagents were supplied by Sigma-Aldrich unless otherwise stated.

2.1 Sperm cell isolation

2.1.1 Plant material

Arabidopsis thaliana Col-0 ecotype, GFP-positive sperm cell plants (a generous gift from Dr. Scott D. Russell, University of Oklahoma), and *Brassica oleracea* were grown under greenhouse conditions 16 hr day/8 hr night at 22˚C, 60% humidity. Arabidopsis seeds were stratified in 0.3% agarose at 4˚C for 2-3 days. The soil for sowing Arabidopsis seeds was treated with 0.04% Intercept 70WG (Scotts). The stratified Arabidopsis seeds were then pipetted onto the soil. Thipex-plus (Koppert Biological systems) was placed on the plant tray approximately 2-3 weeks after sowing to control thrips. The Arabidopsis pollen grains were collected using a vacuum cleaner (Numatic Charles CVC 370 vacuum cleaner) through two different size filters, 55 and 10 µm (Figure 2.1) every two days. The 55 µm filter was used as the barrier for plant debris. The 10 µm filter was used for collecting the pollen of approximately 23 µm in diameter. The trapped pollen was scraped from the filter into 1.5 ml tubes using a razor blade. The Brassica pollen grains (35 µm in diameter) were collected by agitating the flowers in a standard nylon kitchen flour sieve. Both types of pollen grains were collected into 1.5 ml microfuge tubes (Eppendorf) and stored at -80 ˚C until required.

Figure 2.1. Arabidopsis pollen collection utilising vacuum cleaner equipped with filters.

Arabidopsis pollen collection was performed using a vacuum cleaner equipped with two different sized filters, 55 and 10 µm. The 55 µm filter was used to exclude plant debris. The 10 µm was used to collect the ~23 µm *Arabidopsis* pollen.

2.1.2 *Brassica oleracea* pollen disruption and sperm cell isolation

Sperm cells from *Brassica oleracea* were isolated using various techniques to determine the best method to obtain the highest quantity and quality of isolated sperm cells. These techniques were then adapted for use in the model plant *Arabidopsis thaliana* which will be described in section 2.1.3.

B. oleracea pollen grains were collected from the flower by agitating using a standard nylon kitchen flour sieve as mentioned above. Once collected, 15 mg of pollen was transferred to a new 1.5 ml microcentrifuge tube (Eppendorf). The pollen was hydrated in 1 ml 25% sucrose for 30 min on a rotator (Stuart scientific). The pollen sample was then centrifuged at 13,000*g* for 5 min. at room temperature and the supernatant removed. The pollen was subjected to osmotic shock in 1 ml pollen germination media containing 12.5% sucrose (see appendix B for recipe) for a further 30 min. The pollen grains were then physically disrupted using different methods as follows;

2.1.2.1 Grinding method

Hydrated and osmotically shocked pollen grains were centrifuged at 13,000*g* for 5 min. and the supernatant removed. The pollen grains were then ground using a plastic grinder. The disrupted pollen grains were then transferred to a glass slide for microscopical observation (Nikon ECLIPSE 90i).

2.1.2.2 Sonication

Hydrated and osmotically shocked pollen grains were sonicated at 12 and 30 µm amplitude for 10, 15 and 30 sec. The pollen samples were transferred immediately onto ice after sonication. Samples were then pipetted onto glass slides and stained with DAPI to be observed under a microscope equipped with UV.

2.1.2.3 Physical bursting using a metal roller

Hydrated and osmotically shocked *B. oleracea* pollen grains were transferred into a small plastic bag and the bag was heat-sealed. The pollen samples were placed onto a hard smooth surface and rolled with a stainless steel roller 5 times. The sample was then transferred to a 1.5 ml microfuge tube. The sample was then transferred to a glass slide for assessment.

2.1.2.4 PARR® pressurisation/depressurisation pollen disruption method

Hydrated and osmotically shocked *B. oleracea* pollen grains were transferred into an ice-chilled PARR® cell disruption compartment (Parr Instrument Company) containing 9 ml of sperm isolation buffer (Appendix B). The sample was the pressurised under nitrogen gas. The pressure gauge was set at 2,250 psi for 20 min. The burst pollen sample was collected in 15 ml tube (Fisher scientific) as the pressure was released. The sample was agitated on a rotator (Stuart scientific) for an additional 5 min to release the sperm from the enclosing pollen wall and then centrifuged at 13,000*g* for 5 min to pellet the pollen material including sperm and the buffer volume was adjusted to 1 ml. The burst pollen sample was transferred onto a glass slide for assessment.

2.1.2.5 Enzyme treatment and PARR® pressure/depressurisation pollen disruption

Hydrated pollen were enzyme treated at the same time as osmotic shock in 10 ml of 12.5% sucrose sperm isolation media (Appendix B) containing 1% w/v cellulose and 0.5% w/v pectinase for 1 hour. The pollen samples were centrifuged at 13,000*g* for 5 min. to pellet them and the enzyme-containing media was removed. The pollen was resuspended in sperm isolation media (Appendix B) and transferred into the ice-chilled PARR® cell disruption compartment (Parr Instrument Company) and pressured under nitrogen gas. The pressure gauge was set at 1,750 psi for 15 min. The burst pollen sample was collected into a 15 ml tube (Fisher scientific) as the pressure was released. The sample was agitated on a rotator for an additional 5 min. to release the sperm from the burst pollen. The sample was then centrifuged at 13,000*g* for 5 min. to pellet the pollen-derived material and the buffer volume was adjusted to 1 ml. The burst pollen sample was transferred onto a glass slide for microscopical assessment.

2.1.2.6 En masse B. oleracea sperm isolation for FACS

B. oleracea pollen grains (150 mg) were hydrated in 20 ml of 25% sucrose for 30 minutes on a rotator (Stuart scientific). The pollen was then osmotically shocked for 30 min in sperm isolation media containing 12.5% sucrose again on a rotator. The pollen samples were then placed into an ice-chilled PARR® cell disruption compartment (Parr Instrument Company) and pressured under nitrogen gas to 2,250 psi for 20 min. The burst pollen sample was collected into a 50 ml tube (Fisher scientific) as the pressure was released. The burst pollen sample was pushed through two different mesh sizes, 20 and 10 µm nylon meshes respectively to remove pollen wall debris with a 5 ml syringe (Terumo). The sieved sperm sample was centrifuged in swing arm centrifuge

(Eppendorf) at 3,320*g* for 5 min. The sperm enriched pellet was resuspended in 12.5% sperm isolation media and centrifuged again at 1,000*g* for 7 min. to separate sperm cells from the vegetative cell cytoplasm. The sperm sample was resuspended again in 5ml sperm isolation media and stained with 400 ng/ml of Hoechst 33342. The stained sperm sample was then observed by fluorescence microscopy with UV excitation.

2.1.3 *Arabidopsis thaliana* pollen disruption and sperm isolation

Once the pollen disruption and sperm isolation techniques were successfully developed in *B. oleracea*, these techniques were adapted to be used in *Arabidopsis thaliana*. Arabidopsis pollen grains from Col-0 ecotype and GFP-positive sperm line were collected by vacuum containing two filters 55 and 10 µm as mentioned in section 2.1.1. Once collected, 10 mg of pollen from both plant lines was weighed and transferred to a new 1.5 ml microcentrifuge tube (Eppendorf). The pollen was hydrated in 1 ml of 25% sucrose for 30 min on a rotator. The pollen sample was then centrifuged at 13,000*g* for 5 min. at room temperature to remove supernatant. The pollen was osmotically shocked in 1 ml pollen germination media containing 8.5% sucrose (see appendix B for recipe) for a further 30 min. The pollen grains were then physically disrupted using different methods i.e. grinding, sonication, metal roller, and PARR® pressure/depressurisation as in Brassica or otherwise stated as follows;

2.1.3.1 Enzyme treatment and PARR® pressure/depressurisation

Hydrated pollen was enzyme treated at the same time as being osmotically shocked in 10 ml of 8.5% sucrose sperm isolation media (Appendix B) containing 1% w/v cellulose and 0.5% w/v pectinase for 30 min. The pollen samples were then prepared as for Brassica pollen disruption by PARR® pressure/depressurisation technique (see section 2.1.2.4). The pressure gauge was set at 1,750 psi for 10 min. The burst pollen sample was collected in a 15 ml tube (Fisher scientific) as the pressure was released. The sample was agitated on a rotator for an additional 5 min. to release the sperm. The sample was then centrifuged at 13,000*g* for 5 min. and the pellet resuspended in 1 ml sperm isolation media (Appendix B). The burst pollen sample was transferred onto a glass slide for assessment by microscopical observation.

2.1.3.2 En masse Arabidopsis sperm isolation for FACS

Arabidopsis thaliana pollen grains (100 mg) were hydrated in 20 ml of 25% sucrose for 30 minutes on a rotator (Stuart scientific). The pollen grains were then osmotically shocked in 8.5% sucrose sperm isolation media containing 1% w/v cellulose and 0.5% w/v pectinase for 30 min. again on a rotator. The pollen samples were then placed into an ice-chilled PARR® cell disruption compartment (Parr Instrument Company) and pressured under nitrogen gas. The pressure gauge was set at 1,750 psi for 10 min. The burst pollen sample was collected in a 50 ml tube (Fisher scientific) as the pressure was released. The burst pollen sample was pushed through two different mesh sizes, 20 and 10 µm nylon meshes respectively to remove pollen wall debris with a 5 ml syringe (Terumo). The pollen debris trapped on the filter was resuspended in 20 ml 8.5% sucrose sperm isolation media and burst again in the pressure compartment at 1,750 psi for 10 min. The sample were released and sieved through 20 and 10 µm meshes and pooled with the first sample. The sieved sperm sample was centrifuged in swing arm centrifuge (Eppendorf) at 3,320*g* for 5 min. The sperm enriched pellet was resuspended in sperm isolation media and centrifuged again at 1,000*g* for 7 min. to separate sperm cells from the vegetative cell cytoplasm. The sperm sample was resuspended again in 5ml of sperm isolation media and stained with 400 ng/ml Hoechst 33342. The stained sperm sample was then observed with fluorescence microscope under UV excitation.

2.1.3.3 Assessment of sperm isolation buffers

Different sperm isolation buffers (M1-M7 [see Appendix B for recipes]) were assessed for their effectiveness in sperm isolation and storage. GFP-labelled sperm cells were used to assess these buffers. The GFP sperm from 1 mg of pollen were isolated in each buffer utilising the metal roller technique. The sperm cells were counted at isolation and at 1, 2, 6, 12, 24, 48 hours post-isolation. The cells were stored at 4°C throughout the experiment.

2.1.4 Semi-purification and enrichment of sperm cells

2.1.4.1 Percoll discontinuous gradient

Sperm cells were isolated from burst pollen grains by using a discontinuous gradient of Percoll (Amersham Bioscience). The gradient was set by layering 2 ml of 60%, 40%, and 20% Percoll prepared in 8.5% sperm isolation media in 15 ml centrifuge tube. 2 ml of the crude sperm sample, prepared as described in section 2.1.3.1 was layered on the

top using a glass pipette. The gradient was centrifuged in a swing arm rotor (Eppendorf centrifuge 5810R) (3,320*g* for 60 min). The sperm-containing fraction was identified as a white band at the 40%-20% interface and removed gently using a glass pipette. The sperm cells were stained with DAPI and observed with a fluorescence microscope under UV excitation.

2.1.4.2 Percoll cushion

Sperm cells were isolated from burst pollen grains by using a Percoll (Amersham Bioscience) cushion. The cushion was set by layering 2 ml of 40% Percoll prepared in 8% pollen germination media with 2 ml of the isolated sperm sample using a glass pipette. The sample was centrifuged in a swing arm rotor (Eppendorf centrifuge 5810R) (3,320*g* for 60 min). The sperm-containing fraction was identified as a white band and was removed from the top portion of the 40% Percoll (Amersham Bioscience) using a glass pipette. The sperm cells were stained with DAPI and observed by fluorescence light microscopy with UV excitation.

2.1.4.3 Filtration and centrifugation

Burst pollen grains were gently pushed through two different mesh sizes, 20 and 10 µm respectively using a 5 ml syringe. Filtrated sperm cells were centrifuged at 1,000*g* at 4°C for 5 min. in a 50 ml falcon tube. The sperm pellet was then resuspended in 15 ml of sperm isolation media and centrifuged again at 500*g* at 4°C for 7 min. The enriched semi-purified sperm sample was stained with DAPI and observed by fluorescence light microscopy with UV excitation.

2.1.5 Sperm cell purification by Fluorescence-activated cell sorting (FACS)

Sperm cells obtained using the *en masse* sperm isolation techniques (section 2.1.2.6 and 2.1.3.2) were stained with Hoechst 33342. The cell concentration was identified to be approximately 1×10^6 to 2×10^6 cells/ml utilising a haemocytometer. The stained sperm sample was then processed using a FACScan™ Flow Cytometer (Becton Dickinson) with the unstained sample being used as a negative control for calibration purposes. An argon laser was tuned to 488 nm and the UV laser was set to 50 mW. Stained and unstained sperm enriched samples were distinguished using the linear mode on the scatter channel (FSC/SSC [forward scatter – measuring cell size/ side scatter –

measuring cell granularity]) and fluorescence channel (FL4). The sperm cells from the Hoechst 33342 stained sample were sorted according to fluorescence scatter (FL4) versus forward scatter (FSC) (Figure 2.2). The sorted cells were collected directly onto a glass slide for visualisation under a microscope equipped with UV. The rest of the cells were sorted into 1.7 ml MµlTI SafeSeal low-binding microcentrifuge tubes (Sorenson Bioscience) containing 500 µl sperm cell storage buffer to preserve cell osmolarity and viability from BD FACSFlow™ sheath fluid. The sorted cells were frozen in liquid nitrogen and stored at -80°C until required for further analysis.

2.1.6 Purity assessment of FACS-sorted sperm cells by reverse transcription-PCR (RT-PCR)

Arabidopsis FACS-sorted sperm cells were enriched from 1ml FACS sheath buffer mixed with sperm isolation buffer by centrifugation at 12,000*g* for 10 min. at 4°C. The cell pellet was resuspended in 100 µl lysis buffer of the Cell-to-Signal™ kit (Ambion). First strand synthesis and PCR were performed according to the manufacturer's manual for 2 step RT-PCR. Pollen specific – *TUA1* (Carpenter et al., 1992), *AT59* (Kulikauskas and McCormick, 1997), and *PTEN1*(Gupta et al., 2002) and sperm cell specific gene – *GEX2* (Engel et al., 2005) primers were used to assess the purity of the FACS-sorted cells. Both first strand synthesis and PCR steps were performed utilising gene specific primers (Appendix A) for each gene.

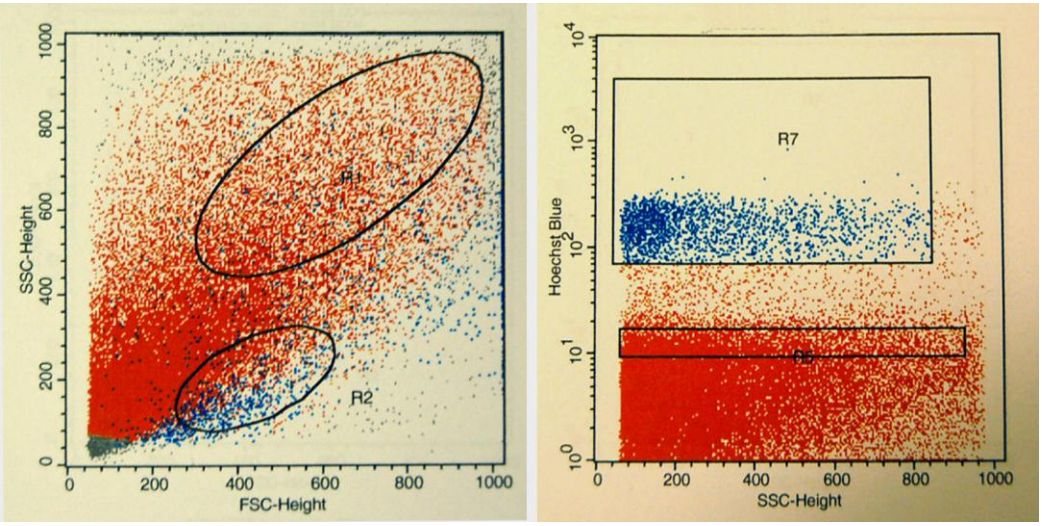


Figure 2.2. Fluorescence-activated Arabidopsis sperm cell sorting

Fluorescence-activated Arabidopsis sperm cell sorting was performed based on cell size (FSC), intra-cellular complexity (SSC), and Hoechst signal. R2 population was selected based on low granulosity (low SSC) and small particles (low FSC) together with R7 population which has Hoechst positive signals to identifythe spermcell population from the total population.

2.1.7 Transmission Electron Microscopy (TEM)

2.1.7.1 Sample preparation

Pollen grains of Arabidopsis wild-type flowers were pollinated onto male sterile A9-barnase flowers. These pollinated flowers were placed onto a 96-well plate (Eppendorf) filled with MilliQ water where parafilm (Pechiney Plastic Packaging Company) was used to cover the water reservoirs. The plate was then placed under a glass jar to be exposed to osmium tetroxide crystals (OsO_4) for 2 hours at 22°C. Samples were then washed 3 times in 0.1 M sodium cacodylate buffer (SCB) pH7.4 for 5 minutes. The final fixation step was performed by an addition of 2.5% glutaraldehyde in 0.1 M SCB pH7.4 for 16-20 hours. The samples were post-fixed in 1% osmium tetroxide in 0.1 M SCB pH7.4 for 2 hours at room temperature and washed 3 times with 0.1M SCB pH7.4 for 5 minutes. The samples were then washed again 2 times with 0.05 M maleate buffer pH6.0 for 3 minutes. This was followed by a wash with 1% uranyl acetate in maleate buffer pH 6.0 for 1 hour in the dark followed by 3 washes in 0.05 M maleate buffer pH 6.0 for 5 minutes.

The samples were dehydrated with -20°C pre-chilled graded ethanol washes of 50%, 70%, 85% and 95% for 5 minutes each wash. 100% ethanol washes were performed at room temperature for 3x10 minutes. The dehydration stage was completed with 2 washes of 100% Propylene Oxide (PO) for 5 minutes. Samples were then infiltrated by submersion in a 3:1 mixture of 100% PO : Spurr's Resin for 2x15 minutes followed by submersion in a 1:1 mixture of PO : Spurr's Resin for 2x15 minutes. The samples were then treated under vacuum for 45 minutes and left overnight in 100% Spurr's Resin. Samples were vacuum treated again for 30 minutes the next day before the samples were placed in moulds and cast in Spurr's Epoxy Resin. The moulds were then placed at 70°C for 7 hours to allow the resin to harden.

2.1.7.2 Transmission Electron Microscopy

The resin block containing the pollen grains was trimmed using a razor blade to establish the best position for pollen grain sectioning. Semi-thin sections were cut using an Ultracut-E ultramicrotome (Leica) and visualised by light microscopy to confirm sectioning of the pollen grains. Ultra-thin sections of 100nm were prepared using a diamond knife for TEM. The cut sections were placed on slot grids and viewed and photographed using a JEM1200EXII transmission electron microscope (JEOL Ltd.) equipped with digital camera at 80KV.

2.1.8 Pollen and sperm cell measurements of *B. oleracea* and *A. thaliana*

2.1.8.1 Pollen and isolated sperm cell measurement utilising DIC microscopy and UDRuler (AVPsoft)

B. oleracea and *A. thaliana* pollen were photographed utilising DIC microscopy. Cell diameter was measured using UDRuler (AVPsoft). Cell volume and surface area were calculated using ellipsoid equations as follows;

$$\text{Ellipsoid volume } (\mu m3) = 4/3\pi abc$$

($\pi = 3.1415$, a = equatorial radius along x axis, b = equatorial radius along y axis, c = polar radius along z axis).

Isolated Arabidopsis sperm cells were stained with DAPI to facilitate identification of the cells under microscope. The cells were photographed under DIC for cell size assessment. Cell diameter was manually measured using UDRuler (AVPsoft). Cell volume and surface area were calculated using following equation;

$$\text{Sphere volume } (\mu m^3) = 4/3\pi r^3$$

$$\text{Sphere surface area } (\mu m^2) = 4\pi r^2$$

($\pi = 3.1415$, r = radius).

2.1.8.2 Sperm cell measurement of GFP-positive sperm utilising confocal microscopy and 3D reconstruction by Imaris 6.1 (Bitplane)

GFP-tagged sperm cells in hydrated pollen grains were visualised and photographed utilising confocal microscopy (Nikon ECLIPSE 90i). 3D images of the sperm cell image sections were reconstructed with Imaris 6.1 imaging software (Bitplane). Volume and surface area of the selected sperm cells were automatically calculated by the software.

2.2 Proteomic studies

2.2.1 SDS-PAGE

2.2.1.1 Sample preparation

Protein samples from each step of the sperm isolation method using the PARR® pressure compartment were subjected to analysis by SDS-PAGE. Protein samples from 10 mg of whole pollen, homogenized pollen, PARR® burst pollen, sieved sperm pellet, sieved supernatant and pollen debris in 1 ml pollen germination media were centrifuged at 14,000*g* for 1 min. 900 µl of the supernatant was taken out from each tubes. The pellet was resuspended in the remaining supernatant. 100 µl of SDS sample buffer (Appendix B) was then added to each sample. The samples were heated to 100˚C for 5 min. and then centrifuged at 14,000*g* for 1 min. 15 µl of the supernatant of each sample was loaded into each well of the acrylamide gel including Precision Plus Protein prestained standards (BioRad) as a marker.

2.2.1.2 Acrylamide gel preparation

Acrylamide gels were prepared by pipetting 3.75 ml of the 12% resolving gel (see Appendix B for recipe) into the gel apparatus (BioRad). The 12% gel was overlaid with butanol and was left to polymerise for 45 minutes. 1ml of 5% stacking gel (Appendix B) was pipetted onto the top of the resolving gel. The gel comb was then put in place to create the wells. The arcrylamide gel was left for another 15 minutes to polymerise. The gel was then placed into the gel tank and TGS running buffer (Appendix B) was poured into the tank. Following loading of the protein samples and the protein standard (BioRad) (15 µl into each well) the samples were run for 50 minutes at 200 volts.

2.2.1.3 Protein gel staining with Coomassie R-250

The gel was stained with the protein dye, Coomassie R-250 (Appendix B) overnight. The stained gel was destained in destaining solution (Appendix B) until the blue protein bands were visible against a clear background.

2.2.1.4 Protein gel stain by silver stain

The protein gel was stained using a Silver Stain Kit (Bio-Rad) as described in the manufacturer's protocol.

2.2.2 Two-dimensional protein gel electrophoresis

Sperm cell-enriched and whole pollen samples obtained as described in section 2.1 were centrifuged at 14,000*g* 4°C for 5 min. The pellet was resuspended in CHAPS cell lysis buffer 100 µl (Appendix B). The resuspended cells were disrupted by freeze-thaw cycles in liquid nitrogen (3 times) followed by vortexing for 1 min.

2.2.2.1 Protein quantification

All protein samples were then quantified using PlusOne 2-D Quant kit (Amersham) as described in the handbook. The samples were then cleaned using 2-D Clean-up kit (Amersham) as described in the user manual. Samples were then quantified again using EZQ protein quantification kit (Molecular probes) as described in the manufacturer's handbook. The membrane containing the protein spots for quantification, prepared using EZQ protein quantification kit (Molecular probes), was scanned using a Fuji BAS1000 phosphoimager (Fuji) at an excitation at 473 nm 400V. The AIDA software (Imagenes) was used to analyse the protein spot intensity.

2.2.2.2 First-dimension Isoelectric Focusing (IEF)

Two-dimensional polyacrylamide gel electrophoresis was performed using a Multiphor II Electrophoresis Unit (Amersham). 8 and 83 µg of protein was used for silver and Coomassie staining respectively. The protein samples were added to 200 µl of rehydration solution (Appendix B) containing 1 µl IPG buffer (Amarsham). The Immobiline™ DryStrip strip gels (Amersham), pH 3-10 and 4-7 were incubated with the sample for 16 hours as described in the manufacturer's 2-D electrophoresis handbook. IEF was conducted at 2mA with a power setting of 5W using the following steps:

Phase 1	300V	1 min	0.001kVh
Phase 2	500V	1.5 hours	2.9kVh
Phase 3	500V	2.5 hours	7.1kVh for pH 3-10 strip, 8.1kVh for pH 4-7 strip

The strips were then stored at -80°C until required or used immediately for the second dimension gel electrophoresis.

2.2.2.3 Second-dimension SDS-PAGE

After IEF, strips were equilibrated in 10 ml SDS equilibration solution (Appendix B) containing 100 mg DTT for 15 min with gentle shaking. The strips were transferred into new 10 ml equilibration solution containing 250 mg iodoacetamide and incubated for another 15 min. The samples were separated in the second dimension on ExcelGel SDS gels (Amersham) as described in the manufacturer's 2-D electrophoresis handbook.

2.2.2.4 Visualisation of protein spots by silver staining

The gels were removed from the system and stained with PlusOneTM silver staining kit (Amersham) as described in the user manual. The stained gels were immersed in glycerol (87% w/w) and were dried on glass plates overnight.

2.2.2.5 Visualisation of protein spots by Coomassie staining

The gels were removed from the system and stained with Coomassie^{TM} brilliant blue (Amersham) as described in the manufacturer's instructions. The stained gels were immersed in glycerol (87% w/w) and were dried on glass plates overnight.

2.2.3 Protein identification by mass spectrometry

2.2.3.1 Silver stained gel

Protein spots chosen to be identified by mass spectrometry were excised from the rehydrated gel. Each protein spot was cut and removed with sharp forceps and transferred into separate eppendorf tubes. The protein spots were destained with equal amount of 30 mM potassium ferricyanide and 100 mM Sodium thiosulphate, a total of 100 µl for 20 min. The spots were washed 4 times with distilled water for 20 min each. The spots were then kept in -20°C freezer until used.

2.2.3.2 Coomassie stained gel

Protein spots chosen to be identified by mass spectrometry were excised from the rehydrated gel. Each protein spot was cut and removed with sharp forceps and

transferred into separate eppendorf tubes. The spots were then kept in -20°C freezer until used.

The spots from both gels were automatedly destained and dehydrated and digested with trypsin. The peptides were then extracted from the gel and transferred to a 96-well plate by a robotic Ettan digester. The plates were then dehydrated. The peptides were resuspended and mixed with a matrix required for running on the MALDI-TOF (Waters Micromass) and spotted onto the MALDI target plate by the robotic Ettan spotter. The samples were run on the MALDI-TOF. A peptide mass fingerprint (PMF) was then generated and compared against the Swissprot and MSDB database using the MASCOT search engine (http://www.matrixscience.com/).

2.3 Bioinformatics

5,068 Maize EST sequences (derived from maize sperm cells, Engel *et al.*, 2003) were searched against Genbank Arabidopsis protein database using BLASTX, e-value 10^3, and blosum45, a matrix which assigns a score for aligning any possible pair of residues. The ESTs that did not match proteins in the database were searched again against the Arabidopsis genome using TBLASTX with blosum45, and e-value 10^{-3}.

2.3.1 Analysis of hits to unannotated regions of the Arabidopsis genome

The sequences that matched unannotated regions of the *Arabidopsis* genome were analyzed further using the Artemis tool (Sanger) to aid the search for potential open reading frames and putative sperm specific candidates. The ESTs that could potentially be a novel Arabidopsis gene were searched for promoter, coding sequences, and open reading frames using Softberry (http://linux1.softberry.com/berry.phtml) and Genescan (http://genes.mit.edu/GENSCAN.html) sequence analysis tools.

2.3.2 Analysis of annotated Arabidopsis hits

The protein sequences that matched the annotated Arabidopsis genome were made non-redundant and searched against Arabidopsis root and leaf ESTs to cull out the sequences unlikely to be sperm specific using e-value equal to 10^{-100} as an indicator. The sequences that had an e-value less than 10^{-100} were considered to be expressed in root or leaf. The rest was checked manually using Gene Atlas (Genevestigator), the

microarray data, to search for the sequences that may be specifically expressed in the pollen.

The 1,522 *Plumbago* sperm ESTs, 61 *Oryza* sperm ESTs, and 3 *Nicotiana* sperm ESTs were searched against the Arabidopsis protein database using blastx, blosum45, with an e-value of 10^{-3}. The sequences that matched Arabidopsis proteins were analysed further using Gene Atlas (Genevestigator), the microarray data, to search for the sequences that were specifically expressed in pollen.

2.4 Transcriptional analysis of putative sperm-expressed gene candidates

2.4.1 RT-PCR analysis

2.4.1.1 RNA extraction

RNA samples were isolated from the sperm cell-enriched fraction and the whole pollen sample (see section 2.1.3 and 2.1.4.3) using the Illustra RNAspin Mini RNA isolation kit (GE healthcare) as described in the manufacturer's handbook. RNAs from both samples were then quantified using a UV spectrometer at absorption settings of 260 and 280 nm. The amount of RNA from both samples was then calculated using the following equation:

$$\text{RNA } (\mu g/\mu l) = 40 \times \text{dilution factor} \times A260 / 1000$$

2.4.1.2 First strand synthesis and PCR

First strand cDNA was synthesised from 1µg RNA using Omniscript RT kit (Qiagen) as described in the handbook. PCR was performed by mixing 1µl of each of the gene specific primers (see appendix A), 1µl 10×Taq polymerase buffer, 0.8 µl 20 mM dNTPs, 1 µl cDNA, 11.1 µl ddH2O, and 0.1 µl Taq polymerase (NEB). PCR was then performed using the following conditions in a thermocycler (Mastercycler Personal, Eppendorf); step 1 - initialisation at 95°C for 2 min; step 2 - denaturation at 95°C for 20 secs; step 3 - annealing at 55°C for 20 sec; step 4 - extension at 72°C for 1 min; step 5 - final extension at 72°C for 10 min; steps 2 to 4 were repeated 28 times. The PCR products were then run on a 1% agarose gel at 120V for 15 minutes with 100bp DNA ladder (NEB) as a marker.

2.4.2 Construction of GFP-translational fusions

2.4.2.1 PCR amplification of genes of interest

The promoter region (approximately 1000-1500bp) and part of the genes of interest were amplified using gene specific primers (Appendix A) containing restriction enzyme sites for cloning. PCR reactions were performed using KOD polymerase (Invitrogen) as described in the manufacturer's handbook. The PCR reaction was performed using the following conditions; step 1 - initialization at 95°C for 2 mins; step 2 - denaturation at 95°C for 20 secs; step 3 - annealing at 55°C for 20 secs; step 4 - extension at 72°C for 2 mins; step 5 - final extension at 72°C for 10 mins; steps 2 to 4 were repeated 35 times. The PCR products were run on a 1% agarose gel at 120V for 15 minutes with a 1Kb DNA ladder (NEB) as a marker.

2.4.2.2 pGEM-T easy cloning

PCR products were excised from the agarose gel for clean up using a QIAquick PCR Purification Kit (Qiagen). The purified PCR products were A-tailed by adding 1µl 10× Taq polymerase buffer, 1µl 2mM dATP, 0.1µl Taq polymerase and 8 µl of purified PCR product. The mixture was incubated at 70°C for 30 minutes. The A-tailed PCR products were now cloned into pGEM-T easy vector (Promega) and transformed into JM109 competent cells (Promega) as described in the manufacturer's handbook. The transformed cells were checked by colony PCR. Individual colonies on the LB agar plate were picked and transferred into 0.2ml tube containing 25µl TE buffer and 1µl protease K (Promega). The mixture was digested at 55°C for 15 min and 80°C for additional 15 min. The PCR reactions were performed by adding 1 µl of T7 and SP6 primers, 1 µl 10× Taq polymerase buffer, 0.4 µl 20mM dNTPs, 1 µl of digested cells, 6.05 µl of ddH20 and 0.05 µl of Taq polymerase (NEB). The PCR mixtures were then incubated as follows; step 1 - initialisation at 95°C for 2 mins; step 2 - denaturation at 95°C for 20 secs; step 3 - annealing at 50°C for 20 secs; step 4 - extension at 72°C for 2 mins; step 5 - final extension at 72°C for 10 mins; steps 2 to 4 were repeated 35 times. The PCR products were run on a 1% agarose gel at 120V for 15 mins with a 1Kb DNA ladder (NEB) as a marker.

2.4.2.3 pBI-GFP cloning

Positive colonies were cultured in 5 ml of liquid LB medium (recipe in Appendix B) containing 100 µg/µl ampicillin overnight at 37°C. The plasmid containing the PCR product was isolated from the cultured cells using Wizard Plus Minipreps DNA

Purification System (Promega) as described in the handbook. Once the plasmid was isolated the PCR insert was digested out using two restriction enzymes, SalI and BamHI (NEB). The digested PCR product was then run on a 1% agarose gel and purified using QIAquick PCR Purification Kit (Qiagen). pBI-GFP plasmid (Kinoshita et al., 2004) (see Appendix D for vector map) was digested with SalI and BamHI restriction enzymes (NEB) and dephosphorylated with shrimp alkaline phosphatase (Promega). The plasmid was then purified using pellet paint (Novagen) as described in the manufacturer's protocol. Both digested PCR product and pBI-GFP plasmid were run on 1% agarose gel to estimate the amount of both DNAs. The PCR product was ligated into pBI-GFP plasmid in a 3:1 ratio with T4 DNA ligase (NEB). The ligation reaction was incubated at 4°C overnight and 2µl was used in transformations with 30 µl of EC100^{TM} electrocompetent E. coli (Epicentre Biotechnologies) as described in manufacturer's protocol. The transformed cells were then plated out on LB agar plates containing 50µg/µl kanamycin. The plates were incubated at 37°C overnight. Colony PCR was then performed as described above using pBI-sense and sGFP-antisense primers (see Appendix A for sequence). The PCR reaction was performed as follows; step 1 - initialisation at 95°C for 2 mins; step 2 - denaturation at 95°C for 20 secs; step 3 - annealing at 50°C for 20 secs.; step 4 - extension at 72°C for 2 mins; step 5 - final extension at 72°C for 10 mins; steps 2 to 4 were repeated 35 times. The PCR products were run on a 1% agarose gel at 120V for 15 minutes. with 1Kb DNA ladder (NEB) as a marker.

2.4.2.4 PCR product sequence analysis by DNA sequencing

The PCR products that were cloned into pGEM-Teasy and pBI-GFP plasmids were analysed by DNA sequencing. The selected *E. coli* colonies that contained plasmid with the correct size insert were cultured in 5 ml LB liquid medium (Appendix B) overnight. The plasmids were then isolated and purified using QIAprep Spin Miniprep Kit (Qiagen). The purified plasmids were sent for sequencing (Cogenics). The DNA sequences obtained from Cogenics were compared with the cloned gene sequence (NCBI) using a DNA sequence alignment tool (http://www.ebi.ac.uk/Tools/emboss /align/index.html).

2.4.2.5 Agrobacterium tumefaciens electrocompetent cell preparation

Agrobacterium tumefaciens GV3101 strain was inoculated into 10ml 2×YT (Appendix B) containing 100 µg/µl ampicillin, 50 µg/µl kanamycin and 100 µg/µl rifampicin and

incubated on a shaker at 28°C for 2-3 days. The 1ml culture was then inoculated into 50 ml 2×YT containing antibiotics and incubated overnight at 28°C until OD_{600} > 0.5. The cells were collected by centrifugation at 8,500 rpm for 5 minutes and resuspended in 50 ml ice cold 10% glycerol twice. The pellet was then resuspended in 1 ml of ice cold 10% glycerol. 30 µl of the cells were divided into 1.5 ml tubes and stored at -80°C.

2.4.2.6 Agrobacterium transformation

Colonies containing the correct PCR products were cultured in 5 ml liquid LB medium (recipe in Appendix B) containing 50 µg/µl kanamycin overnight at 37°C and plasmids were isolated from the cultured cells using the Wizard Plus Minipreps DNA Purification System (Promega) as described in the handbook. 1 µl of purified pBI-GFP plasmid containing the expected PCR product was transformed into 30µl of *Agrobacterium tumefaciens* competent cells. The plasmid and the cells were gently mixed by flicking in a chilled 1.5 ml eppendorf tube. The mixture was then transferred into pre-chilled electroporation cuvettes (Eppendorf). The electroporation device (BioRad) was set at Ec2 for bacterial transformation. The cuvette was placed into the electroporator and the electric pulse was applied for a few seconds. 500 µl of SOC medium was then immediately added to the cells. The transformed Agrobacterium was incubated at 28°C for 3 hours on the shaker. The cells were then plated out onto 2×YT agar plates (see appendix B) containing 100 µg/µl ampicillin, 50 µg/µl kanamycin and 100 µg/µl rifampicin and incubated at 28°C for 2-3 days. Colony PCR was then performed as described in the pBI-GFP cloning section above (insert section number here). The PCR products were run on 1% agarose gel at 120V for 15 minutes with a 1Kb DNA ladder (NEB) as a marker.

2.4.2.7 Floral dipping for Agrobacterium-mediated transformation of Arabidopsis

Transformed Agrobacterium were grown to stationary phase (OD_{600} = 2.0) in 300 ml 2×YT medium containing 100 µg/µl ampicillin, 50 µg/µl kanamycin and 100 µg/µl rifampicin at 28°C for 2-3 days. The cells were collected by centrifugation at 4000*g* for 20 minutes. The cell pellet was resuspended in 500 ml floral dipping solution (Appendix B). The inflorescences of one month old Arabidopsis were immersed in the dipping solution and agitated for 10 sec. The plant was left to dry for 1 hour, covered in a plastic bag and then left in the dark overnight. The plant was dipped again after five days to increase the transformation efficiency. The plant was grown normally until seeds were mature. The dry seeds were then collected for selection.

2.4.2.8 Selection of Agrobacterium transformants

1g of seeds collected from the previous step were weighed out (~50,000 seed) into 50 ml tubes. The seeds were sterilised in 25 ml 70% EtOH for 5 minutes followed by 25 ml 50% bleach containing 0.05% Tween 20 for an additional 5 min. The seeds were then washed in sterile water six times. Seeds were then resuspended in 45 ml of 0.3% agarose and plated out onto MS agar plates containing 50µg/µl kanamycin. The plates were sealed with parafilm and stratified at 4°C for 2 days. The plates were then placed in the growth room for 4 hours and left in the dark for 2 days. The plates were then placed back in the growth room until the transformants appeared. The transformed plant were then transferred to soil and allowed to grow normally.

2.4.2.9 Microscopy of GFP transformed plants

Pollen of the transfomants were dispersed onto glass slides by dipping the opened flower onto a drop of 25% sucrose and covered with cover slip. The pollen was then observed with a fluorescence microscope (Nikon ECLIPSE 90i) under UV excitation with a GFP filter and compared to wild type Col-0 pollen. Different stages of pollen development were also studied to see GFP expression in uni cellular, bicellular and mature pollen. Flower buds of different sizes were dissected using forceps under dissecting microscope. Anthers were isolated and transferred onto a drop of mounting media (Appendix A) on a glass slide and covered with cover slip. A small amount of pressure was placed on the cover slip to release the pollen out of the anther without bursting the pollen. The pollen grains of different developmental stages were observed with a fluorescence microscope under UV excitation. Different tissues of the transformed plant i.e. sepal, petal, carpel, anther, seed, leaf, and pollen of different stages were also observed using a confocal microscope for GFP expression. Each tissue type was dissected and placed on a drop of mounting media on the glass slide and covered with cover slip. The samples were then observed and compare with wild type plants by confocal microscopy.

2.5 Analysis of T-DNA insertion plant lines

2.5.1 Silique phenotype screening

Arabidopsis thaliana T-DNA insertion lines (123 lines) were grown under growth room (Sanyo) conditions (16 hr day/8 hr night at 22˚C). Siliques of 5-6 week-old plants were

opened with forceps. The numbers of seeds in the silique were counted with the aid of a dissecting microscope.

2.5.2 Segregation distortion assays

A range of *Arabidopsis thaliana* T-DNA insertion lines - SALK, SAIL and GABI-Kat lines with insertions in genes of interest obtained from the bioinformatics studies were ordered from NASC. Approximately 25 seeds per line were transferred into 1.5 ml tubes and sterilised with 25 ml 70% EtOH for 5 minutes followed by 25 ml of 50% bleach containing 0.05% Tween 20 for an additional 5 minutes. The seeds were then washed in sterile water for six times. The seeds were pipetted out onto MS agar plates (Appendix B) containing the appropriate antibiotic (100 µg/µl kanamycin for SALK lines, 7 µg/µl BASTA for SAIL lines and 75 µg/µl sulfadiazine for GABI-Kat lines). The plates were sealed with Parafilm® M (Pechiney Plastic Packaging Company) and stratified at 4°C for 2 days. The plates were then placed in a growth room until the resistant and sensitive plants were visible. The resistant plants from each line were removed to soil and allowed to grow normally. The seeds were then collected from each individual plant. Two hundred seeds from each plant line collected from a single, probably heterozygous, plant were transferred to a 1.5 ml tube and processed as mentioned above. The number of resistant plants was scored after two weeks.

2.5.3 Genotypic analysis of Arabidopsis T-DNA insertion lines

Arabidopsis thaliana T-DNA insertion lines were grown under growth room (Sanyo) conditions (16 hr day/8 hr night at 22˚C). Genomic DNA was extracted from the leaves of 2-3 week-old plants.

2.5.3.1 DNA extraction

Young leaves (approximately 1 cm in diameter) of the plant were ground with a plastic grinder attached to a drill in 100µl DNA extraction buffer (Appendix B) in a 1.5 ml tube containing glass beads (Sigma). The ground leaf sample was then incubated at 65°C for 5 minutes, 100 µl of chloroform was then added to the sample. The sample was vortexed for 30 seconds and centrifuged at 13,000*g* for 10 minutes. The clear upper phase was then carefully pipetted into a new 1.5 ml tube. The sample was mixed with 100 µl isopropanol and incubated at room temperature for 15 min. The samples were then centrifuged at 13,000*g* for 20 minutes. The pellet was then washed with 200 µl

70% EtOH. The sample was centrifuged at 13,000*g* for 20 minutes. The pellet was left to dry at room temperature. The DNA pellet was then dissolved in 50 µl ddH2O.

2.5.3.2 Genotyping

The extracted genomic DNA from each plant was used PCR to identify the genotype of each individual plant. Two PCR reactions were performed for each plant. The first reaction was a mixture of 1 µl of genomic DNA, 1µl of left and right gene specific primers, 3 µl of 2× Reddy mix (ABgene), and 4 µl ddH2O. The second reaction was a mixture of 1 µl of genomic DNA, 1µl of T-DNA left border and right gene specific primers, 3 µl of 2× Reddy mix (ABgene) and 4 µl of ddH2O. Both PCR reactions were performed under following conditions; step 1 - initialisation at 95°C for 2 min; step 2 – denaturation at 95°C for 20 secs; step 3 - annealing at 50°C for 20 secs; step 4 - extension at 72°C for 1 min; step 5 - final extension at 72°C for 10 mins; steps 2 to 4 were repeated 35 times. The PCR products were run on a 1% agarose gel at 120V for 20 minutes with a 1Kb DNA ladder (NEB) as a marker.

2.5.4 Reciprocal crosses

Arabidopsis thaliana T-DNA insertion lines, wild-type col-0, and Arabidopsis male-sterile A9 were grown under conditions mentioned in section 2.1.1. Two set of crosses were performed in each line in order to identify problems with male and/or female gametophytes. Firstly, the wild-type pollen was pollinated onto flowers of the mutant lines. The unopened flowers of T-DNA knockout lines were emasculated and left in the growth room for 2-3 days. The emasculated flowers were then pollinated with wild-type pollen. Secondly, the mutant line pollen was pollinated onto male-sterile A9 stigmas. The plants were then transferred to the growth room and allowed to set seed. The 2-3 weeks old siliques were then opened with forceps. The seeds were counted and observed by microscopy.

2.5.5 Pollen staining and pollen germination assays

2.5.5.1 Pollen viability assessment

A Fluorochromatic reaction (FCR) assay was used to verify pollen viability in T-DNA knockout lines compared to wild-type pollen. The Arabidopsis pollen was dispersed onto a drop of 20 µl 25% sucrose solution containing 1 µl of 5 mg/ml fluorescein

diacetate (FDA) on a glass slide. The pollen sample was covered with a cover slip and incubated for 5 min. The slide was observed with a fluorescence microscope under UV excitation with an FITC filter.

2.5.5.2 DAPI staining

Arabidopsis pollen was stained with DAPI (4',6-diamidino-2-phenylindole) to assess the presence and arrangement of the sperm cells and vegetative nucleus inside the pollen grain. The Arabidopsis pollen was dispersed onto a drop of 20 µl 25% sucrose solution containing 1 µl of 5 mg/ml DAPI and 1 µl of 0.5% Triton X-100 on a glass slide. The solution was covered with a cover slip and incubated for 10 min. The slide was observed with fluorescence microscope under UV excitation with DAPI filter.

2.5.5.3 In vitro pollen germination assay

Arabidopsis pollen was carefully dispersed onto a thin layer of solidified pollen germination media (Appendix B) on a glass slide. The carpel was also placed on the media to facilitate pollen germination. The slide was placed on top of a wet paper tissue in a small plastic box containing water to maintain a humid atmosphere for pollen germination (this humidity chamber was adopted from Johnson-Brousseau, S. A. and McCormick, S. [2004]). The pollen was placed in the growth room and observed by light microscopy after 6 and 24 hours.

2.5.5.4 In vivo pollen germination assay

Arabidopsis flowers were crossed as described in the reciprocal crosses section above (section xxx). One day-old hand pollinated carpels were excised and fixed in 100 µl fixative (see appendix B) overnight. The fixed carpel was then transferred to 100 µl 8M NaOH and incubated for 2 days. The softened carpel was stained with 0.1% aniline blue in 100 mM K_3PO_4 buffer pH 11 for 1.5 hour. The stained carpel was transferred onto a glass slide and mounted with mounting solution (Appendix B). A cover slip was carefully placed on top of the sample without causing any air bubbles. The sample was observed with a fluorescence microscope under UV excitation with a DAPI filter.

2.5.5.5 Excision and examination of the female gametophyte

The unfertilised ovules of the mutant plants that demonstrated a semi-infertility phenotype were dissected from un-opened flowers with the aid of a dissecting microscope (Nikon) with forceps and a hypodermic needle. The ovules were isolated into a few drops of Hoyer's solution (Liu and Meinke, 1998) (Appendix B) on a glass slide and covered with a cover slip. The ovules were incubated on the slide for 2-3 minutes and observed under DIC microscope (Nikon ECLIPSE 90i).

2.5.6 RT-PCR analysis of T-DNA insertion lines

RNA samples were extracted from the leaves of homozygous plants from each line as well as wild-type to be used as a control. 100 mg of leaf material was collected into a 1.5ml microfuge tube and ground with a plastic grinder attached to a drill in 200µl Tri reagent. 800µl Tri regent was added to the homogenised material. The rest of the procedure was performed according to the manufacturer's protocol. First strand cDNA synthesis and PCR was performed as described in section 2.4.1 above. The transcripts of the gene of interest were verified to be present in the leaves of wt plants before PCR was performed on each knockout line.

Chapter 3:

Sperm Cell Morphological Studies, Isolation and Purification Technique Development

3.1 Introduction

Despite a long history of plant sperm cell studies, knowledge of sperm cells, particularly of the model plant *Arabidopsis thaliana*, is still in its infancy. Very few genes have been identified that are expressed specifically in this type of cell. Only one gene, *GCS1* (also known as *HAP2*) (Mori et al., 2006; von Besser et al., 2006) has been identified as being involved in egg-sperm fusion (see review in chapter 1). The main obstacle to plant sperm cell studies is that the sperm cells reside in a strong sporopollenin wall that protects the pollen, a vehicle for gamete delivery. Therefore, effective release of sperm cells from the pollen grain is the first barrier in the studies. In this project, sperm cell isolation and purification techniques were developed for the purpose of molecular analyses of these cells. Moreover, as Arabidopsis sperm cells have not been reported on thoroughly with respect to their cytology and morphology, in this project, 3D images of the sperm pair were constructed utilising confocal microscopy and advanced imaging software to gain a better understanding of their overall structure. In addition, transmission electron microscopy (TEM) was utilised to visualise sperm cell cellular components in order to investigate the possibility of dimorphism of the two sperm cells of Arabidopsis. Details and discussion of these analyses will be described in this chapter.

3.1.1 Plant sperm cell morphology

Flowering plant sperm cells are unique as they reside in another cell type, the pollen grain, and are delivered to their fusion partner by the pollen tube, a mechanism that greatly facilitates successful fusion between the sperm cells and their female partners, the egg and central cell (Boavida et al., 2005; Lord and Russell, 2002). The overall structure of plant sperm cells has already been described in chapter 1. In this chapter, sperm cell ultrastructure and dimorphism will be reviewed and discussed in more detail as these studies could provide more understanding of preferential sperm fusion in angiosperms.

3.1.1.1 Ultrastructure of plant sperm cells

In the late 1900s, plant sperm ultrastructural studies were widely performed in various species, e.g. *Brassica oleracea, Brassica campestris* (McConchie et al., 1987b), *Zea mays* (Mogensen et al., 1990), and *Plumbago zeylanica* (Russell, 1984) to reveal the structure of the two sperm cells. The two sperm cells of *P. zeylanica* differ in size and morphology (Russell, 1984). One sperm cell (Svn) is physically associated with pollen vegetative nucleus at one end and with the other sperm cell (Sua) at the other end (Russell, 1984). Svn was observed to be morphologically polarised, consisting of the main cell body at one end and a narrow cellular projection at the other end which wraps around the vegetative nucleus (Russell, 1984). The other sperm cell, Sua, is spindle shaped with one end attached to Svn and the other possessing a blunt tip (Russell, 1984). A similar dimorphic sperm cell structure was observed in *Brassica oleracea* and *Brassica campestris* (McConchie et al., 1987b). The sperm cells contained normal cellular organelles e.g. endoplasmic reticulum, Golgi complex, ribosomes, microtubules and vesicles. However, unlike *P. zeylanica*, *Brassica* sperm cells lack plastids (McConchie et al., 1987b). The bigger sperm cell (Svn) of *B. oleracea* possesses a blunt evagination which wraps around, and lies within, shallow furrows of the vegetative nucleus. Sua on the other hand does not link to the vegetative cell nucleus but is connected to Svn by a common cell junction (McConchie et al., 1987b). The junction between the two sperm cells was observed to consist of an elaborate pairing of finger like evaginations held within a common periplasm (McConchie et al., 1987b). These cytological studies have provided valuable information about sperm cell size, their organelles, the connection between the two sperm cells and the pollen vegetative nucleus and importantly these studies revealed that many plant sperm cells are dimorphic.

3.1.1.2 Plant sperm cell dimorphism

As mentioned in the previous section plant sperm cells from various species have been reported to be dimorphic. However, the characteristics of dimorphism differ between species. In *P. zeylanica* where dimorphic sperm are reported, the smaller sperm cell (Sua) contains more plastids and the larger (Svn) contains few plastids but 5 times more mitochondria (Russell, 1984). As mentioned in chapter 1, these dimorphic sperm from *Plumbago* display preferential fusion to their female partners. The plastid-rich sperm cell (Sua) fused with the egg cell and the mitochondria-rich sperm cell (Svn) fuses with the central cell (Russell, 1984; Southworth et al., 1997). Another species having dimorphic sperm is *Zea mays* (Wagner et al., 1989). In this case the sperm cells are of similar size, and the type and number of organelles contained within them are also similar. However, the difference between the sperm of this species lies in the nucleus where two types of nuclei, heterochromatic and non-heterochromatic, were identified (Wagner et al., 1989). It has been suggested that differences between the plasma membranes of the sperm cells could be the basis of preferential fusion. However freeze fracture analyses of sperm cells from various species i.e. *Plumbago zeylanica* (Southworth et al., 1997), *Zea mays* (Southworth et al., 1988), *Amaryllis* (Southworth et al., 1994), *Spinacia oleracea* (van Aelst et al., 1990), and *Lolium perenne* (Van der Maas et al., 1994) revealed no dissimilarities between the two sperm cell plasma membranes. However, interestingly, a study investigating the surface charge of sperm cells in *P. zeylanica* revealed charge differences between the two sperm cells and the authors speculated that this could be a factor for determining sperm cell fusion fate (Zhang and Russell, 1999). It was speculated that the different surface charge of each sperm in the pair (determined by electrophoretic mobility) was due to the different cellular composition of each cell (Zhang and Russell, 1999). These different negative charges were proposed to affect the sperm's ability to absorb myosin at the sperm cell surface which could consequently determine each sperm cell's fate by affecting the actomyosin interaction in the embryo sac (Huang and Russell, 1994; Zhang and Russell, 1999). In Arabidopsis, sperm cell dimorphism has never been reported. However, as mentioned above an Arabidopsis relative, *Brassica oleracea* was reported to have dimorphic sperm cells with different sizes by McChonchie et al (1987b). Thus it is possible that the sperm cells of Arabidopsis are also dimorphic and could display preferential fusion to their partners. The potential dimorphism of Arabidopsis sperm cells was investigated in this project and will be described in this chapter.

3.1.2 Sperm cell isolation

3.1.2.1 Pollen disruption

Plant sperm cell research has been carried out for decades in various plant species and one of the crucial stages for molecular studies is the release of sperm cells from the pollen grain. Sperm cells of angiosperms are embedded in the vegetative cells of the pollen and therefore the isolation of intact sperm cells can be troublesome. However, several reports have documented successful sperm cell isolation for various species utilising different approaches. In the 1970's attempts to isolate the generative cell and pollen vegetative nuclei from the pollen tubes of the flowering plant *Tradescanti* were successful (LaFountain and Mascarenhas, 1972). However, the described method, which involved osmostic shock, provided a very small number of sperm cells which were insufficient for sperm cell studies (LaFountain and Mascarenhas, 1972; Russell, 1986). Nevertheless, the osmotic shock technique was becoming a standard basic procedure to isolate sperm cells. The buffers used in these studies are a crucial factor for this procedure. The osmolarity of the solution needs to be suitable for disruption of the pollen grain and simultaneously for preserving the sperm cells. Sucrose was often used as the main component of the sperm isolation media as its concentration is easily adjusted, has neutral pH and is readily available. The percentage of sucrose used in the isolation buffer is adjusted to suit the size and osmolarity of the pollen and sperm cells in each plant species. Variations on pollen germination media were often used to isolate sperm as they were designed to maintain the health and integrity of these specialised cell types. Brewbaker-Kwack (BKS) pollen germination media containing 20% sucrose was first used to isolate the sperm cells of barley by Cass (1973). This was the first tricellular pollen species from which sperm cells were reported to be isolated by osmotic shock. Sperm cell viability was assessed by observing the intactness of the cells (Cass, 1973). Freshly isolated sperm cells were observed in pairs which assumed a spindle shape (Cass, 1973; Russell, 1991). However, after 30 minutes sperm cells were seen to separate and became spherical (Cass, 1973; Russell, 1991). Since reports that the sperm cells could be successfully isolated in BKS media, the use of this media as a sperm isolation solution became universal (as shown in Table 3.1). However, the buffer system has since been adapted and improved to enhance sperm cell longevity. MOPS and MES buffer have commonly been used to adjust pH in both plant and animal cell cultures (Dunphy and Nolan, 1979; Satoh et al., 1990; Zhang et al., 1992a). Both MOPS and MES were used and compared in the *Entomophthora egressa* protoplast culture to maintain retension and viability of the protoplasts (Dunphy and Nolan, 1979). MES was found to work the best in this study providing the best balance between buffering capacity and toxicity (Dunphy and Nolan, 1979). In 1992, Zhang and colleagues

developed new buffers from conventional BKS media to improve longevity of maize sperm cells (Zhang et al., 1992b). Different factors including 6 different types of sugar, 10 buffers, 5 pH values, and 3 membrane protective agents were adjusted to improve the cell viability and longevity (Zhang et al., 1992b). The best conditions found for maize sperm cell storage in this study were media containing 0.55 M galactose with 2 mM 2-(N-morpholino) ethanesulfonic acid and 0.1% bovine serum albumin pH 6.7. However the 0.55 M galactose could be substituted with 0.44 M sucrose without any significant changes to the data (Zhang et al., 1992b). However osmotic disruption with BKS media is still the most commonly used technique to disrupt the pollen grains for sperm cells isolation (Table 3.1). In some species e.g. *Spinacia oleracea*, *Brassica oleracea*, and *Lolium longiforum*, osmotic bursting alone was found to be insufficient to release sperm cells from the pollen grain and therefore other physical pollen disruption techniques were utilized to facilitate the rupture of the pollen (summarized in Table 3.1). Isolated sperm cells reviewed so far have been released from the tricellular pollen species in which the sperm cells are produced during microsporogenesis in the anther. Nevertheless, sperm cells have also been successfully isolated from species that have bicellular pollen in which the two sperm cells are derived from the dividing generative cell in the growing pollen tube. In 1988, Shivanna *et al*. developed a new sperm isolation technique for the bicellular pollen species *Rhododendron macgregoriae* and *Gladiolus gandavensis* (Shivanna et al., 1988) . Pollen tubes were allowed to grow through segments of style on to an agar plate and sperm cells were then released from the pollen tubes either by osmotic shock or pollen wall degrading enzymes (Shivanna et al., 1988).

3.1.2.2 Sperm cell enrichment and viability assessment

After disruption of the pollen samples are usually incubated for a further 10-20 minutes to facilitate the separation of sperm cells from one another and the vegetative cell cytoplasm. Purification of the sperm cells is then performed by various methods, as shown in table 3.1. Filtration and centrifugation techniques have frequently been used in studies. These techniques are generally sufficient for separating pollen components from the desired sperm cells and can also be easily adapted to suit the plant species under study. Centrifugation techniques have normally been carried out in conjunction with a discontinuous sucrose or Percoll gradient which enriches sperm cells in the appropriate layer that has the same density as the cells (Russell, 1991; Schwitzguebel and Siegenthaler, 1984). In 1986, the two dimorphic sperm cells of *Plumbago zeylanica* were isolated by osmotic disruption and a sperm enriched fraction has been obtained by utilising a sucrose gradient (Russell, 1986). The sperm cells were separated on a 30%

sucrose layer when centrifuged at 1300*g*. Sperm cell viability was then determined by Evans blue exclusion staining (Gaff, 1971) of the dead cells and a fluorochromatic reaction (FCR) using fluorescein diacetate (FDA) which is cleaved in intact living cells liberating fluoroscein (Heslop-Harrison et al., 1984; Rotman and Papermaster, 1966; Russell, 1986). Moreover, viability of the sperm cells can also be tested using an assay for ATP, an indicator of life, involving luciferin which yields measurable luminescence (Karl, 1980; Matthys-Rochon et al., 1987; Russell, 1991). Sperm cell yield was reported to be up to 75% with concentrations of up to 8.8×10^6 cells/ml in *P. zeylanica* (Russell, 1986). Dupuis and colleagues were able to collect 90% viable sperm with a concentration of 3×10^6 cells/ml from maize (Dupuis et al., 1987). Since this work a number of studies have demonstrated the successful isolation of reasonable quality sperm cells and in quantities sufficient for study (various isolation techniques in different plant species were summarised in table 3.1). However, the isolated sperm cells reported in these publications were only semi-purified and still contaminated with pollen vegetative cell components. Therefore the development of techniques to obtain pure sperm cells was crucial for molecular studies.

Table 3.1. Sperm cell isolation from various plant species (format used in Russell 1991 with many additional entries).

Plant species	Method	Medium	Purification method	Cell viability assessment	Sperm yield	Half life	Reference
Tri cellular pollen							
Monocotyledons							
Hordeum vulgare	OS	BKS Suc30	N/A	DIC	N/A	N/A	Cass, 1973
Lolium perenne	OS/G	RY-2 Suc20	P(0/*/30%)	FDA, DAPI, PCM	2%	O/N	van der Maas and Zaal, 1990
Lolium perenne	OS/G	BKS Suc15	P(0/*/30%)	DIC	N/A	N/A	van der Maas et al., 1994
Secale cereale	GH	BKS Suc5 Sob8	Fil 40, P(30%)	FDA, Ho	N/A	N/A	Yang and Zhou, 1989
Triticum aestivum	OS/G	BKS Suc30	P(0/*/30%)	FDA, DAPI, EtBr, PCM	N/A	N/A	Matthys-Rochon et al., 1987
Triticum aestivum	OS	BKS Suc20	Sor 20/40%	FDA, DAPI	N/A	15min	Szakacs and Barnabas, 1990
Zea mays	OS	N/A	Fil, 1.22g/m NCD	DAPI, PCM	N/A	N/A	Cass et al., 1986
Zea mays	OS/G	BKS Suc15	N/A	FDA, DAPI, EtBr, PCM	N/A	N/A	Matthys-Rochon et al., 1987
Zea mays	OS	BKS Suc15	Fil 50/20, P(15*/40%)	FDA, DAPI, TEM, SEM	3×10^6 cells/ml, 20%	20h	Dupuis et al., 1987
Zea mays	OS	BKS Suc15	N/A	FDA, DAPI, PCM	N/A	N/A	Matthys-Rochon et al., 1988
Zea mays	OS	BKS Suc15	Fil 40, P(30%)	FDA, PCM	N/A	N/A	Roeckel et al., 1988
Zea mays	OS	BKS Suc15	Fil, 1.22g/m NCD	EvB, DIC, PCM, TEM,ABB, CBB	1.5×10^6 cells/ml, 30%	3h	Cass and Fabi., 1988
Zea mays	OS/G	BKS Suc5 Sob8/K	Fil 40, P(30%)	FDA, Ho	N/A	N/A	Yang and Zhou, 1989
Zea mays	OS	BKS Suc15	Fil 50/20, P(15*/40%)	FDA, EM	N/A	N/A	Wagner et al., 1989
Zea mays	OS/pH	BKS Suc15	Fil 50/20, P(15*/40%)	FDA, DAPI	N/A	N/A	Kranz et al., 1991
Zea mays	OS	BKS Suc15	Fil 40, P(30%)	FDA, PI, FACScan, DIC	N/A	>72h	Zhang et al., 1992 a
*Zea mays**	OS	0.44M Suc/0.55M Gal (+B)	Fil 40, P(30%)	FDA	N/A	>72h	Zhang et al., 1992 b
Zea mays	OS	BKS Suc15	N/A	FDA, DAPI	N/A	N/A	Roeckel and Dumas, 1993
Zea mays	OS	BKS Suc15	Fil 40, P(30%)	FDA, DAPI	N/A	N/A	Zhang et al., 1993
Zea mays	OS	BKS Suc15	Fil 40, P(30%)	FDA, DAPI	N/A	N/A	Zhang et al., 1995
Zea mays	OS/pH	BKS Suc15	Fil 50/20, P(10%)	DIC	2×10^5 cells/30 min	N/A	Engel et al., 2003
Zea mays	OS/pH	520mM Man	P(10%), FACS	DIC, DAPI	2×10^4 cells	N/A	Faure et al., 2003
Oryza sativa	OS/GR	Mod1	Fil 25, P(10*/30/70%)	DIC, FDA	153.6 cells/434 pollen grains (17.8%)	5h (RT)	Khalequzzaman and Haq, 2005

Oryza sativa	OS	0.3M Man	Mic	DIC	N/A	N/A	Uchiumi et al., 2006; Uchiumi et al., 2007
Dicotyledons							
Ambrosia sp.	OS	BKS	N/A	FDA, EtBr, PCM	N/A	N/A	Matthys-Rochon and Dumas, 1988
Artemesia sp.	OS	BKS	N/A	FDA, EtBr, PCM	N/A	N/A	Matthys-Rochon and Dumas, 1988
Bellis sp.	OS	BKS	N/A	FDA, EtBr, PCM	N/A	N/A	Matthys-Rochon and Dumas, 1988
Beta vulgaris	OS	20+d	P(20/30*/50%)	FDA, PI, Ho, DIC	7.3×10^6 cells/ml, 30%	N/A	Nielsen and Olesen, 1988
Brassica campestris	GH	N/A	Fil, 1.07 g/ml NCD	Ho, PCM	N/A	N/A	Cass et al., 1986
Brassica campestris	GH	Man-Tris	P	Ho, PCM	N/A	N/A	Hough et al., 1986
Brassica campestris	OS, GH*	Suc12.5+a	Fil, Cent	FDA	332 cells/34%, 760 cells/86%	36h, 96h	Mo and Yang, 1991
Brassssica napus	GH	BKS Suc12.5	N/A	FDA, PCM	N/A	N/A	Matthys-Rochon and Dumas, 1988
Brassssica napus	GH	RM+b	N/A	FDA, Ho	N/A	N/A	Yang and Zhou, 1989
Brassica oleracea	OS/G	BKS Suc15	N/A	FDA, DAPI, EtBr, PCM	N/A	N/A	Matthys-Rochon et al., 1987
Gerbera jamesonii	GH	Man-Suc+c	P	FDA, SEM, TEM	N/A	N/A	Southworth and Knox, 1988
Gerbera jamesonii	GH	1M Suc+e	P	EvB, DAPI, DIC, SEM	N/A	N/A	Southworth and Knox, 1989
Plumbago zeytanica	OS	BKS Suc5-50	N/A	DIC	N/A	N/A	Russell and Cass, 1981
Plumbago zeytanica	OS	Suc 20	Suc 30%	EvB, FDA, Ho, DIC	8.8×10^6 cell/ml, 60%	8h	Russell, 1986
Plumbago zeytanica	OS/pH	MOPS/Man	Mic	FDA	100 cells/h	>2h	Zhang et al., 1998
Senecio sp.	OS	BKS	N/A	FDA, EtBr, PCM	N/A	N/A	Matthys-Rochon and Dumas, 1988
Spinacia oleracea	PS	BKS Suc25	P(*/10/30/50%)	FDA, PCM	4×10^6 cells/ml, 5-10%	18h	Theunis et al., 1988
Spinacia oleracea	PS	BKS Suc25	P(20%)	FDA, PCM, TEM	1×10^6 cells/ml, 5-10%	20h	Theunis and Went, 1988, 1990
Taraxacum sp.	OS	BKS	N/A	FDA, EtBr, PCM	N/A	N/A	Matthys-Rochon and Dumas, 1988
<u>Bicellular pollen</u>							
Monocotyledons							
Lilium longiflorum	Enz[j], GH	BKS Suc10 in vitro1	N/A	My, PCM	N/A	N/A	Southworth and Knox, 1988
Gladiolus gandavensis	OS[k], Enz[l]	Semi-vivo1	N/A	DAPI, EtBr, Ho, SEM	60%	N/A	Shivanna et al., 1988
Gladiolus gandavensis	OS, Enz	Semi-vivo3	N/A	DIC, FDA, DAPI	538 cells, 71.9%	N/A	Mo and Yang, 1992
Hemerocallis minor	OS, Enz	Semi-vivo3	N/A	DIC, FDA, DAPI	627 cells, 78.2%	24h	Mo and Yang, 1992

Hippeastrum vittatum	OS, Enz	Semi-vivo3	N/A	DIC, FDA, DAPI	566 cells, 57.2%	N/A	Mo and Yang, 1992
Iris tectorum	OS, Enz	Semi-vivo3	N/A	DIC, FDA, DAPI	485 cells, 55%	N/A	Mo and Yang, 1992
Zephyranthes candida	OS, Enz	Semi-vivo3	N/A	DIC, FDA, DAPI	812 cells, 82.9%	N/A	Mo and Yang, 1992
Dicotyledons							
Rhododendron macgregoriae	OS[k], Enz[l]	Semi-vivo1	N/A	DAPI, EtBr, Ho, SEM	N/A	N/A	Shivanna et al., 1988
Nicotiana tabacum	OS	Semi-vivo2	N/A	PCM, FDA	N/A	N/A	Tian and Russell, 1997 a, b
Nicotiana tabacum	OS	Semi-vivo2	N/A	PCM	N/A	N/A	Tian and Russell, 1998
Nicotiana tabacum	OS, Enz	In vitro2	Fil, P (15*/22%)	DIC, DAPI	N/A	N/A	Xu et al., 2002
Nicotiana tabacum	OS	Semi-vivo2	Mic	DIC	100 sp/30 min	N/A	Yang et al., 2005
Torenia fournieri.	OS, Enz	Semi-vivo[1]	N/A	FDA, DAPI	N/A	N/A	Vagi et al., 2004
Torenia fournieri.	OS	Semi-vivo[1]	Mic	FDA, DIC	100 sp/30 min	N/A	Chen et al., 2006

N/A: not mentioned in the publication; O/N: over night; OS: osmotic shock; Enz: enzyme treatment; PS: pressure; G: grinding; GH: glass homogeniser; GR: glass roller; Mic: micromanipulator; P: Percoll; Fil: filtration; Man: mannitol; Suc: sucrose; Sob: sorbitol; K: macronutrients; Mod1: 1.3 mM boric acid, 0.74 mM potassium phosphate, 3.6 mM calcium chloride, 438 mM sucrose and 7mM{3-(N-morpholino) propane sulphonic acid monosodium salt (MOPS) pH 6.0 with a osmolarity of 0.075 mos mol kg^{-3} H2O; +a: 0.1g/l KNO_3, 0.36 g/l $CaCl_2$. $2H_2O$, 0.3% PDS, 0.6% BSA, 0.3% PVP; +b: 0.3% potassium dextrand sulphate (PDS) 10 µg/ml FDA; +c: 0.5 mg/ml nicotinic acid, 0.5 mg/ml pyridoxine, 0.1 mg/ml thiamine HCl, 2 mg/ml glycine; +d: Burst medium contained 150 mg/l $CaCl_2.2H_2O$, isolation medium contained 30% sucrose, 150 mg/l $CaCl_2$.2H2O, 100 mg/l H_3BO_3; +e: Pollen hydrated in 2.1 mM $Ca(NO_3)_2$, 1.6 mM H_3BO_3, then grinded in 1M mannitol, 0.2 M sucrose, 10 mM HEPES pH 7.2, 0.3-1% BSA, 0.3% polyvinylpyrrolidone, 10 mM cysteine, 2.1 mM $Ca(NO_3)_2$; Semi vivo1: Styles implanted in BKS12with 0.6% agar in the dark until pollen tubes emerged, then incubated in BKS Suc5 or BKS Suc7.5 or in enzymes; Semi-vivo2: Pollen tubes were germinated through cut style floated in a culture medium (Tian and Russell, 1997; Yang et al., 2005) and transferred to the 9% (w/v) mannitol solution,; Semi vivo3: Semi vivo1+c; In vitro1: Pollen tubes grew in vitro for 2 hr then enzymes – 2% cellulysin, 11 U/mg pectinase were added; In vitro2: the pollen tubes were incubated in a cell-wall-degrading enzyme mixture (Xu et al., 2002) for 20 min at room temperature; BKS: Brewbaker-Kwack salt solution; RY-2: medium (van der Maas and Zaal, 1990); RM: Roberts medium (Yang and Zhou, 1989); EtBr: ethidium bromide, Ho: Hoechst; DAPI: 4',6-diamidino-2-phenylindole , PI: propidium iodide; My: mythramycin; EvB-: Evans blue; ABB: Aniline blue black; CBB: Coomassie brilliant blue; FDA: fluorescein diacetate SEM: scanning electron microscopy; TEM: transmission electron microscopy; PCM: phrase contrast microscopy; DIC: differential interference contrast microscopy.

3.1.2.3 Sperm cell purification

For the past decade of sperm cell research only semi-purified sperm cells have been obtained in the vast majority of studies and thus molecular analyses of sperm cell fractions have had to be done in comparative manner. In 1988, Geltz and Russell isolated protein from three fractions, 'cytoplasmic-paticulate', 'water soluble' and 'male germ unit (MGU)-rich', from *Plumbago zeylanica* for two-dimensional electrophoresis studies. Of 1,227 spots obtained from all three fractions, 133 most conspicuous spots were selected for comparison. Of these 133 spots, 18 spots were found to be unique to the MGU-rich fraction, 3 to the cytoplasmic-particulate fraction, and 14 to the water soluble fraction (Geltz and Russell, 1988). Later in 2002, sperm cells of tobacco were isolated from the pollen tube by osmotic shock in combination with enzymatic digestion (Xu et al., 2002). Sperm cells were initially semi-purified by filtration and Percoll discontinuous gradient centrifugation. The sperm cell fraction was then subsequently treated with RNase to eliminate possible RNA contamination from the pollen vegetative cell (Xu et al., 2002). However, some significant contaminants were present in the isolated sperm cell fraction including the vegetative cell nucleus (Xu et al., 2002). The cells were used to construct a cDNA library. An initial screen of 396 clones uncovered only 2 cDNAs representing sperm-cell-expressed transcripts, one of which was found to be involved in sperm cell differentiation (Xu et al., 2002). In 2003, fluorescence-activated cell sorting (FACS) was used with semi-purified sperm samples from *Zea mays*. FACS has long been used to purify particular cell types from mixed populations, including sperm cells (Aslam et al., 1998; Loken and Herzenberg, 1975; Mays-Hoopes et al., 1995). A human spermatogenic cell suspension obtained from testicular tissue was fractionated using two different methods, velocity sedimentation under unit gravity (VSUG) combined with discontinuous Percoll centrifugation (DPC) and FACS for comparison (Aslam et al., 1998). The FACS method gave a higher yield than VSUG and more than 99% of the sorted cells retained their viability after FACS separation (Aslam et al., 1998). However, FACS had not been used to purify plant sperm cells until 2003 when two independent research groups optimized the FACS technique to be used with a Percoll semi-purified maize sperm sample (Engel et al., 2003; Faure et al., 2003). Sperm cells were stained with the fluorescent DNA dye Hoechst 33342 and sorted using a FACS machine equipped with a double laser (argon and UV lasers). Sorted sperm cell quality was assessed by DIC microscopy (Engel et al., 2003; Faure et al., 2003) and sperm cell purity by RT-PCR for transcripts specific to vegetative cell (Engel et al., 2003). The first plant sperm cell cDNA library, free from vegetative cell contamination, was generated from this purified maize sperm cell population (Engel et al., 2003).

Research on sperm biology in plants is still in its infancy and the identification of proteins that are crucial for sperm function is the major challenge of this project. The inaccessibility of ovules and the small size of both male and female gametes, especially of Arabidopsis, have proved a challenge to biologists interested in plant reproduction (Ye et al., 2002). So far no research has been reported on how to isolate sperm cells from the pollen of *Arabidopsis thaliana*. This project aims to develop techniques to isolate and purify sperm cells and to characterize the sperm proteome of the angiosperm *A. thaliana* by using the much larger *Brassica oleracea*, an Arabidopsis relative (Ayele et al., 2005), as a facilitator. Comparative genetic analyses of *B. oleracea* and *A. thaliana* have revealed the similarity between these two species, suggesting that knowledge gained in one species can be applied to the other (Ayele et al., 2005; Babula et al., 2003). The nucleotide conservation between these two species has been reported to be 75-90% in exonic regions (Ayele et al., 2005; Quiros et al., 2001).

In this chapter, pollen disruption and sperm cell purification methods will be reported that were developed in *B. oleracea* and adapted to be used in *A. thaliana*. In addition sperm cell isolation techniques using fluorescence-activated cell sorting (FACS) were developed for both plant species with the aim of using FACS-purified Arabidopsis sperm cells for further studies. Moreover, Arabidopsis sperm cell morphology was investigated utilising transmission electron microscopy (TEM) to determine whether sperm dimorphism was evident in this species due to the fact that its relative *B. oleracea* was reported to have dimorphic sperm (McConchie et al., 1987b). Putative dimorphism of Arabidopsis sperm cells was assessed by TEM and by confocal microscopy utilising imaging software (Imaris; Bitplane) which permitted relatively accurate sperm cell volume measurements.

3.2 Results

3.2.1 Comparative analysis of pollen grains and sperm cells of *Brassica oleracea* and *Arabidopsis thaliana*

Brassica oleracea and *Arabidopsis thaliana* pollen collected from flowers (Figure 3.1a) by agitation and vacuum methods respectively were weighed and counted. *B. oleracea* and *A. thaliana* pollen were approximately 1.3×10^5 and 1.6×10^5 grains per mg respectively. Both *A. thaliana* and *B. oleracea* pollen grains were visualised using light microscopy (Figure 3.1c, d) and measured using the Universal Desktop Ruler software

(AVPsoft). The volume of *B. oleracea* and *A. thaliana* pollen grains were calculated using following ellipsoid equation;

$$\text{Ellipsoid volume } (\mu m3) = 4/3\pi abc$$

(π = 3.1415, a = equatorial radius along x axis, b = equatorial radius along y axis, c = polar radius along z axis). The average volumes of *B. oleracea* and *A. thaliana* pollen were 11,281.459 μm^3 and 3,642.583 μm^3 respectively. The larger pollen of *B. oleracea* was approximately 22% larger in diameter than *A. thaliana* pollen and over three times the volume. Viability of Arabidopsis pollen grains collected by the vacuum method was assessed by the FDA procedure (Figure 3.1b) which demonstrated that the pollen was not adversely affected by this collection technique. Pollen vegetative nuclei and sperm cells were observed in the pollen grains when visualised using fluorescence light microscopy coupled with DAPI staining. The two sperm cells from both species have condensed chromatin and stain more brightly than the pollen vegetative nucleus when stained with DAPI. The sperm were regularly observed to be close to one another and to the pollen vegetative cell nucleus (Figure 3.1e, f).

3.2.2 *Brassica oleracea* and *Arabidopsis thaliana* sperm isolation and purification

B. oleracea was used in this project to facilitate the development of protocols for pollen disruption, sperm isolation and purification for *A. thaliana*. *B. oleracea* produces much larger pollen grains in great quantity whereas Arabidopsis pollen grains are very scarce. Moreover, these two species are both members of the Brassicaceae and possess very similar flower and pollen morphologies. In addition at the molecular genetic level these species are similar, as mentioned in the introductory section of this chapter. Pollen grains of *B. oleracea* were ruptured by a range of physical disruption techniques. The percentage of ruptured pollen grains compared to the number of viable isolated sperm cells was assessed (Table 3.2). Osmotic shock coupled with pressurisation /depressurisation utilising the Parr apparatus was found to be the best way to disrupt the pollen grains. Once the best method had been obtained for *B. oleracea* sperm isolation, the techniques were applied to Arabidopsis. However, in Arabidopsis, osmotic shock coupled with enzymatic treatment before Parr disruption was found to be more efficient in bursting the pollen and releasing the sperm cells (Table 3.2). Sperm cells were then semi-purified and enriched using various methods i.e. Percoll cushion, Percoll gradient, centrifugation, and filtration.(Table 3.3) However, filtration through 20 µm nylon mesh and centrifugation at 3200*g* for 5 minutes followed by 1000*g* for 7 minutes was found to be efficient for isolating sperm cells with the least contamination.

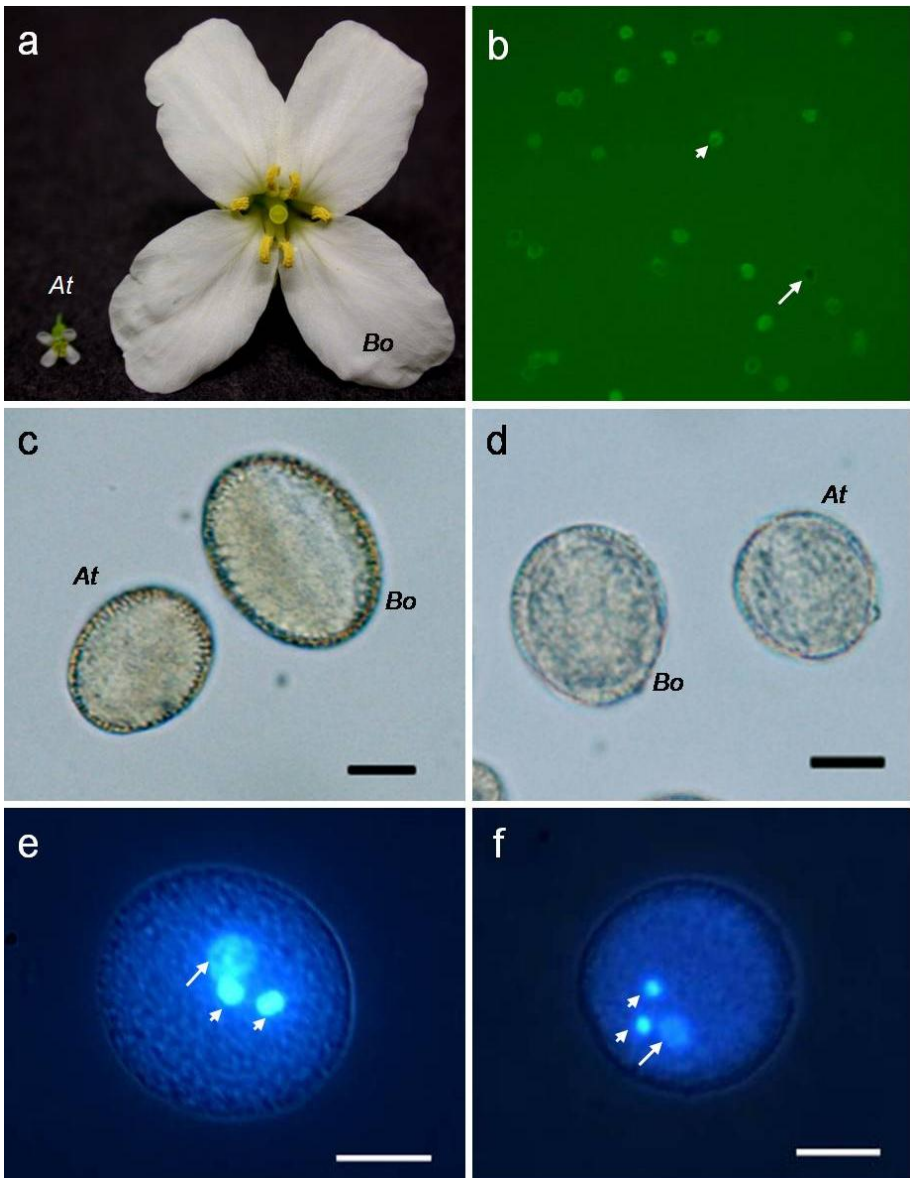


Figure 3.1. Comparative analysis of *Brassica oleracea* and *Arabidopsis thaliana* pollen.

Flowers and pollen grains collected from *B. oleracea* and *A. thaliana* were observed and compared. *A. thaliana* (*At*) flowers and the much larger *B. oleracea* (*Bo*) flowers (a). Pollen grains of *A. thaliana* collected by vacuum were stained with FDA for assessment of viability (b: arrow head - green viable pollen; arrow - nonviable pollen). Hydrated pollen grains of Brassica and Arabidopsis were compared by light microscopy (c - x axis; d - y axis). Hydrated pollen grains of *B. oleracea* (e) and *A. thaliana* (f) were stained with DAPI and visualised by fluorescence light microscopy with UV excitation. DAPI stained pollen grains demonstrated two sperm cells (e and f arrow head) and a less condensed vegetative cell nucleus (e and f arrow) (bar 10 µm).

Sperm cell viability and sample purity were assessed by differential interference contrast (DIC) microscopy and the data is presented in table 3.3. Moreover, for Arabidopsis a GFP-positive sperm line (a gift from Dr. Scott D. Russell, University of Oklahoma) was used to facilitate the sperm viability and longevity assessments. The longevity of isolated sperm cells was assessed for 7 different sperm isolation media (M1-M7, see Appendix B, for media composition). The quality of the isolated sperm cells was similar in all isolation media at 1-2 hours after isolation. However the longevity of the sperm cells was different for each of the isolation media. 6 hours following isolation the numbers of sperm cells surviving dramatically decreased in all media except M5 which contained sorbitol and MES buffer. Sucrose concentration was also found to be crucial for sperm cell longevity as shown in table 4 where the percentage of sucrose in M5 media clearly affected the long term viability of sperm. The half-life of sperm in the most effective media (M5 containing 15% sucrose) was 12 h after isolation (Table 4). To maintain sperm cell viability, the number of steps and duration of the centrifugation procedure were also reduced to minimum. The buffer chosen to be used in sperm isolation experiments was balanced at an optimum percentage of 12.5% sucrose in M5 media to allow efficient hydration of the pollen grains and subsequently 8.5% was used to apply osmotic shock to the pollen. The 8.5% sucrose M5 media was then used throughout the isolation and purification processes, including FACS, as it allowed intact sperm cells to be pelleted at relatively low centrigugal forces. The pH of this M5 sperm storage media was stabilised by MES and adjusted to 6.8 as a stable buffer environment was particularly important for the FACS sorting step.

En masse isolated and semi-purified *B. oleracea* and *A. thaliana* sperm cells utilising Parr pressurisation/depressurisation technique (Figure 3.2a, b) followed by filtration (Figure 3.2c, d) and centrifugation (Figure 3.2e, f) were counted using a haemocytometer. From 100mg and 150mg pollen grains, approximately 7.6×10^6 cells/ml and 5×10^6 cells/ml were collected from *B. oleracea* and *A. thaliana* respectively. The spherical isolated sperm cells of both *A. thaliana* and *B. oleracea* were observed using light microscopy, photographed and their diameters measured using the Universal Desktop Ruler (UDRuler, AVPsoft). The surface area and volume of *B. oleracea* and *A. thaliana* sperm cells were calculated using the following equations;

$$\text{Sphere volume } (\mu m^3) = 4/3\pi r^3$$

$$\text{Sphere surface area } (\mu m^2) = 4\pi r^2$$

(π = 3.1415, r = radius). The volumes and areas of the sperm cells are reported in table 3.5.

Table 3.2. A comparison of pollen disruption techniques for *Brassica oleracea*.

Pollen disruption techniques were developed in *B. oleracea* as the quantity of pollen that could be obtained for analysis was much larger than that for Arabidopsis. The techniques yielding the greatest number of viable sperm cells *from B. oleracea* were later assessed in Arabidopsis.

Pollen disruption techniques coupled with osmotic shock (10 mg/ml)					
Pollen disruption methods		No. of disrupted pollen grains per µl (total ~1100 grains/µl)	Pollen disruption rates (%)	No. of intact sperm cells isolated (cells/µl)	Intact sperm yield (%)
Grinding		440	40	448	51
Metal roller		748	68	1032	69
Parr pressure bomb		605	55	690	57
Enzymatic coupled with pressure bomb		693	63	568	41
Sonication					
Amplitudes (µM)	Time (sec.)				
12	10	594	54	274	23
12	15	715	65	444	31
12	30	770	70	600	39
30	10	737	67	412	28
30	15	781	71	468	30
30	30	836	76	402	24

Table 3.3. Comparison of sperm enrichment and semi-purification methods for *B. oleracea*.

Various sperm cell isolation and semi-purification techniques were developed in *B. oleracea* in order to be applied to Arabidopsis. The yield of viable sperm cells and degree of contamination after purification were the main criteria for choosing the best method for further studies in Arabidopsis.

Sperm cell enrichment and semi-purification techniques (total 350 cells/µl)			
Methods	No. of intact sperm cells (cells/µl)	Intacted sperm yield (%)	Degree of cytoplasmic contamination*
Percoll gradient	92	38	1.5
Percoll cushion	143	41	3
Centrifugation	220	63	4
Filtration	227	65	4
Filtration coupled with centrifugation	182	52	1

* Degree of cytoplasmic contamination 1-4 (based on an observation under fluorescence light microscopy), 1 represents the least cytoplasmic contamination, 4 represents the most contaminated.

Table 3.4. Assessment of sperm isolation media.

Various sperm isolation media with different sucrose concentrations were assessed. The medium that was able to sustain the cell viability for the longest time was chosen to be used in the study which was M5. Different sucrose concentrations (8.5%, 12.5%, and 15%) of the M5 medium were assessed. The highest sucrose concentration was found to be the best for maintaining viable sperm cells. The viable sperm yield was counted in 20 random fields under a 22 mm^2 cover slip and calculate using the volume of liquid present under each field (method adapted from Russell, 1986). The isolated sperm cells were stored at 4°C in between each count.

Sperm isolation by metal roller method (1 mg *Arabidopsis* sperm:GFP pollen)								
Media*	No. of viable sperm cells (cells/µl)	Sperm yield (%)	Survival rate of sperm cells (%)					
			Storage duration (hours)					
			1	2	6	12	24	48
M1	161	70	100	100	67	45	6	0
M2	158	69	100	100	65	42	4	0
M3	117	51	98	97	63	48	10	0
M4	154	67	100	99	62	40	0	0
M5								
8.5% Suc	147	64	100	98	80	30	0	0
12.5% Suc	156	68	100	100	95	60	16	0
15% Suc	156	68	100	100	95	62	19	0
M6	122	53	100	98	67	37	0	0
M7	118	51	100	100	68	39	0	0

* see media recipes in Appendix B

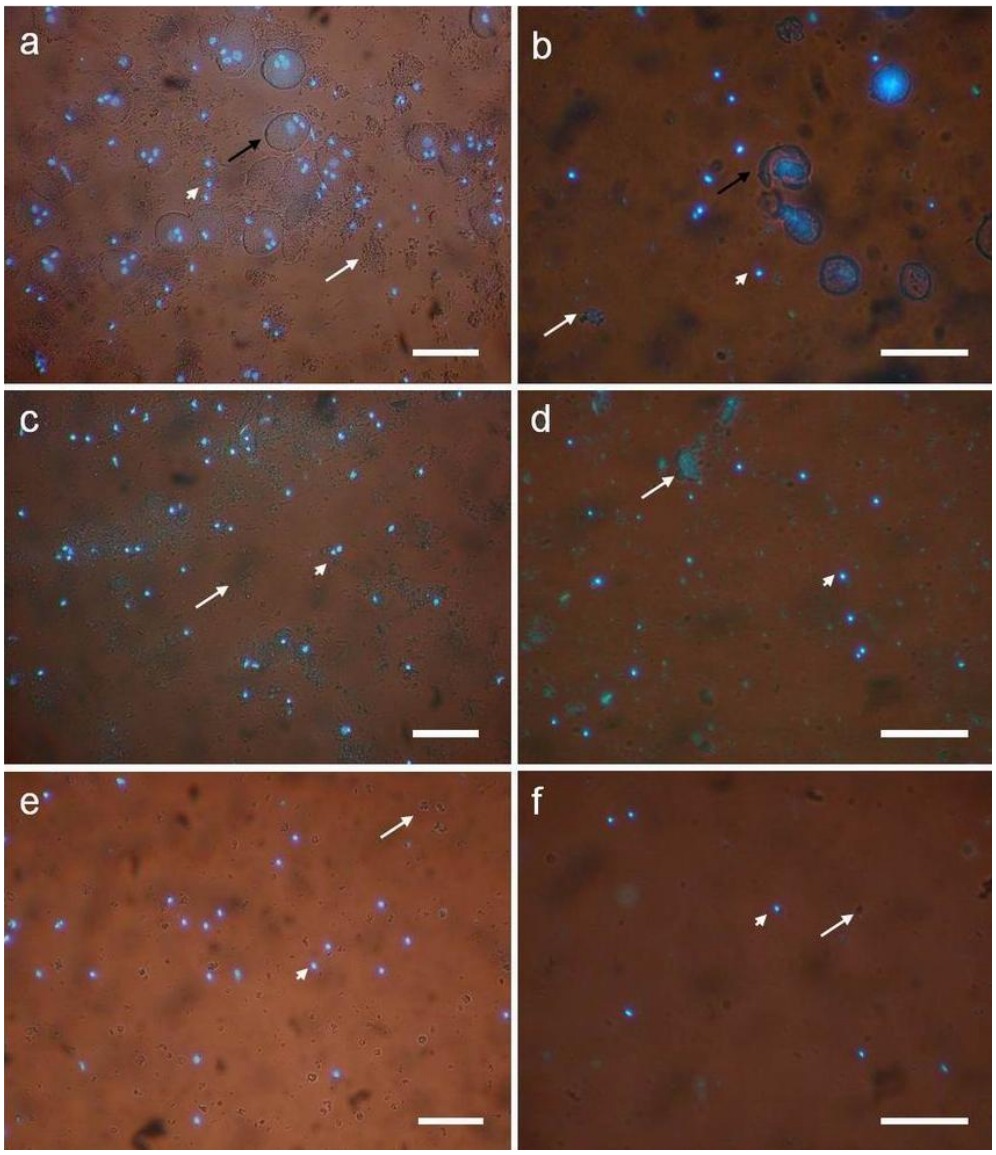
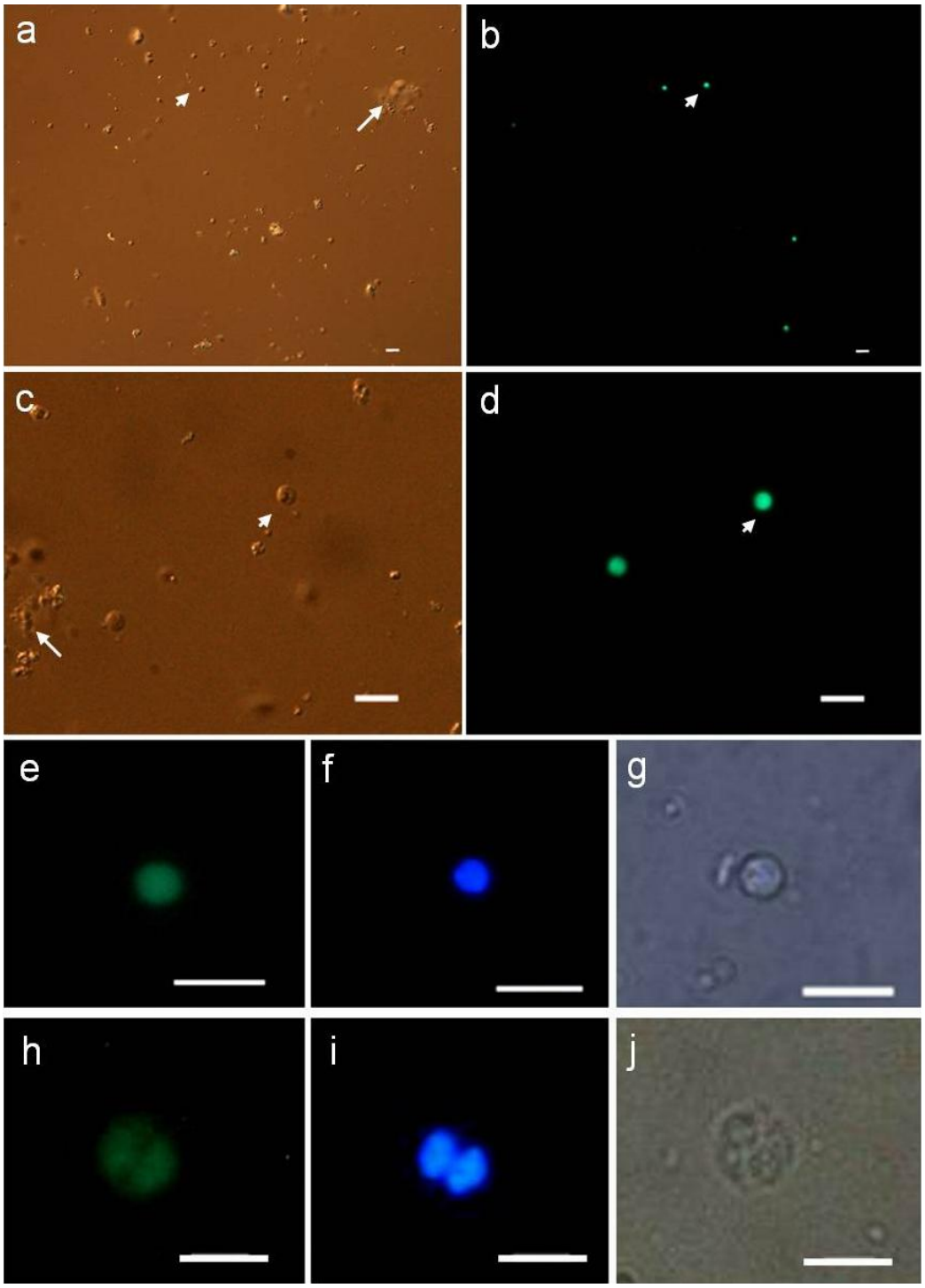


Figure 3.2. *B. oleracea* and *A. thaliana* sperm isolation, partial purification and enrichment.

En masse B. oleracea (left panel) and *A. thaliana (right panel)* isolated sperm cells using osmotic shock, enzymatic treatment and the Parr pressure compartment (a, b) were sieved sequentially through 20 and 10 µm meshes (c, d) and centrifuged at 3220*g* for 5 minutes followed by a second centrifugation step at 1000*g* for 7 minutes (e, f). The samples were stained with DAPI and visualised by fluorescence light microscopy for an assessment of each step. The pollen wall (black arrow) was excluded in the first filtration step. Pollen cytoplasmic contamination (white arrow) was reduced gradually in each step. The partially purified sperm cell sample in the final step contained intact sperm cells (arrow head) and comparatively little visible pollen cytoplasmic contamination (bar 50 µm).

Figure 3.3. Sperm cell quality and viability assessment utilising GFP-tagged sperm cells.

En masse isolation of GFP-tagged sperm cells of Arabidopsis was performed using osmotic shock, enzymatic treatment and the Parr pressure compartment and purified by filtration through 20 and 10 µm filters and centrifuged at 3220*g* for 5 minutes followed by second centrifugation step at 1000*g* for 7 minutes for further purification and enrichment. Pollen cytoplasmic contamination (arrow) and sperm cells (arrow head) could be visualised when the sample was observed under low magnification (20×) by DIC microscopy (a). The GFP signal of viable isolated sperm cells was observed under UV with an FITC filter (b). Intact spherical isolated sperm cells were visualised at higher magnification (60×) by DIC microscopy (c). The viability of this spherical sperm cell was demonstrated by the bright GFP signal (d). Single isolated sperm cell of approximately 2.4 µm across visualised at 100× (e - GFP sperm cell; f - slightly smaller sperm nucleus stained with DAPI; g - sperm cell when visualised by DIC microscopy). Arabidopsis sperm pairs are often seen immediately after isolation, the two sperm cells were physically connected (h - GFP sperm cell; I - smaller sperm nucleus stained with DAPI; j - sperm cells when visualised by DIC microscopy (bar 5 µm).


a
b
c
d
e
f
g
h
i
j

3.2.3 Purification of sperm cells by FACS

Semi-purified sperm cells obtained following the filtration and centrifugation techniques described in previous sections were stained with Hoechst 33342 dye and further purified using fluorescence-activated cell sorting (FACS) (Figure 3.4). Approximately 1.6×10^6 cells/ml (from 100 mg pollen) were obtained from Brassica and 2×10^6 cells/ml from Arabidopsis (from 150 mg pollen). FACS-sorted sperm cells were free from cytoplasmic contamination as assessed by DIC microscopy (Figure 3.4). However, a further more accurate assessment of sperm cell purity was required to confirm this result and will be described in section 3.2.4.

3.2.4 Assessment of FACS-sorted sperm cell purity

FACS-sorted sperm cells were visualised by DIC microscopy and appeared free from cytoplasmic contamination (see Figure 3.4c, d). However, reverse transcription-PCR (RT-PCR) was adopted to be utilised for a more rigorous assessment of purity. Arabidopsis pollen specific genes, *ALPHA-1 TUBULIN (TUA1)* (Carpenter et al., 1992), *pollen specific phosphatase (PTEN1)* (Gupta et al., 2002), and an Arabidopsis homolog of tomato *LAT59* (*AT59)* (Kulikauskas and McCormick, 1997) were used as a control to identify any pollen cytoplasmic contamination if present. Sperm-expressed *GEX2* was utilised as a positive control for sperm cell transcripts. Despite the fact that the reported transcription levels of all these pollen-specific genes are exceptionally high (verified by Gene Atlas) in comparison to *GEX2* only *GEX2* was able to be amplified when RT-PCR was performed. This result clearly indicated that the FACS-sorted sperm cells were free from pollen cytoplasmic contamination (Figure 3.5).

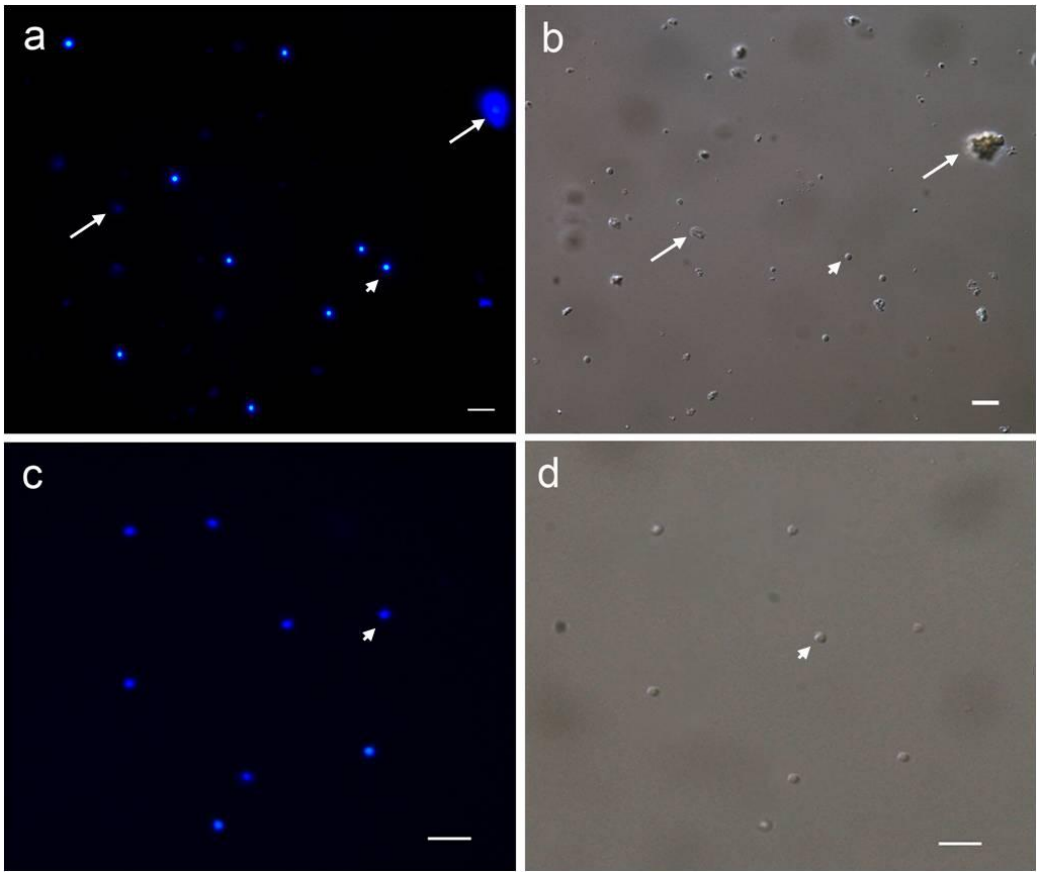


Figure 3.4. FACS-sorted sperm cells.

Arabidopsis thaliana sperm cells were released from pollen by a combination of osmotic shock, enzymatic treatment and Parr pressure compartment disruption. The sample was then sieved and centrifuged. The semi-purified sperm sample was stained with Hoechst 33342 and sorted using FACS (bar 10 µm). (a) Semi-purified Arabidopsis sperm sample stained with Hoechst 33342. Sperm cells (arrow head) are visible as bright blue fluorescent dots and less intense blue fluorescence signals represent pollen cytoplasmic contamination (arrow). (b) DIC microscopy of a semi-purified sperm sample; sperm cells are indicated with an arrow head and pollen cytoplasmic contamination indicated with arrow. (c) FACS-sorted sperm cells stained with DAPI. (d) FACS-sorted sperm cells (approximately 2.3 µm across) visualised by DIC microscopy.

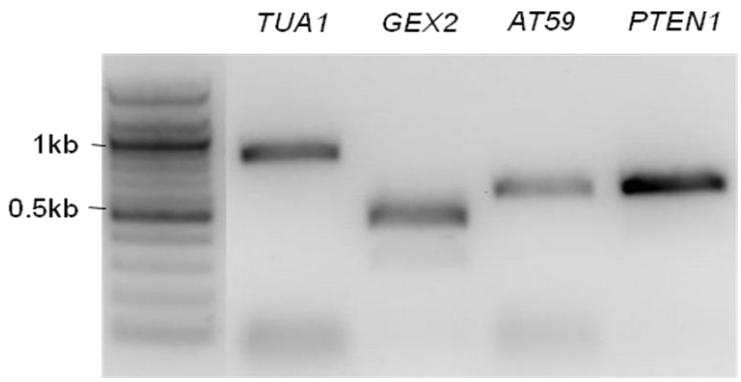


Figure 3.5. Purity assessments of FACS-sorted sperm cells by RT-PCR.

The purity of FACS-sorted Arabidopsis sperm cells was assessed by reverse transcription-PCR (RT-PCR). Primers were designed to enable genomic DNA and cDNA PCR products to be distinguished. Only one cDNA PCR product of the expected 312bp was amplified when primers for the sperm cell-specific gene *GEX2* were used (upper band: genomic DNA product of 466bp, lower band: cDNA product of 312bp). Three vegetative cell-specific genes, *TUA1*, *AT59*, and *PTEN1* were selected to assess for pollen cytoplasmic contamination. Only genomic DNA products of these genes were amplified despite the fact that they are expressed at a high level in the pollen. This data suggests that the FACS-sorted sperm cells were free from pollen cytoplasmic contamination.

3.2.5 Transmission Electron Microscopy (TEM) of *Arabidopsis thaliana* sperm cells in the pollen grain

The two sperm cells of *Arabidopsis thaliana* were visualised and examined by transmission electron microscopy (TEM) to enable a detailed analysis of their cellular components and to determine whether there is any evidence of sperm dimorphism in this species. Serial thin sections of *A. thaliana* pollen grains were collected using an ultra microtome for TEM. Serial sections derived from five grains were examined in order to compare the morphology of the two sperm cells to one another. In each sperm cell a range of organelles were identified i.e. mitochondria, the nucleus, endoplasmic reticulum (ER), Golgi complex, mitochondria and vesicles (Figure 3.6). The two sperm cells were often observed to align in the different axes (Figure 3.6a, b). Detailed assessment of the serial sections indicated that the two sperm cells contain relatively similar types and numbers of organelles and thus no obvious morphological differences were identified. Mitochondria were found to be the most abundant organelle in the Arabidopsis sperm cells and typically five or more were counted per cell. The connections between the two sperm cells and between the sperm cell and vegetative nucleus were observed to be along their plasma membranes which seemed to form a very close association (Figure 3.7). However the nature of this association formed in the male germ unit is still undefined and is probably quite weak and it was only observed only when the cells were in the pollen grain. Shortly after sperm cells are released from the grain, the two sperm cells were often disconnected. However, the precise moment of disconnection *in vivo* and *in vitro* of the male germ unit has never been successfully identified

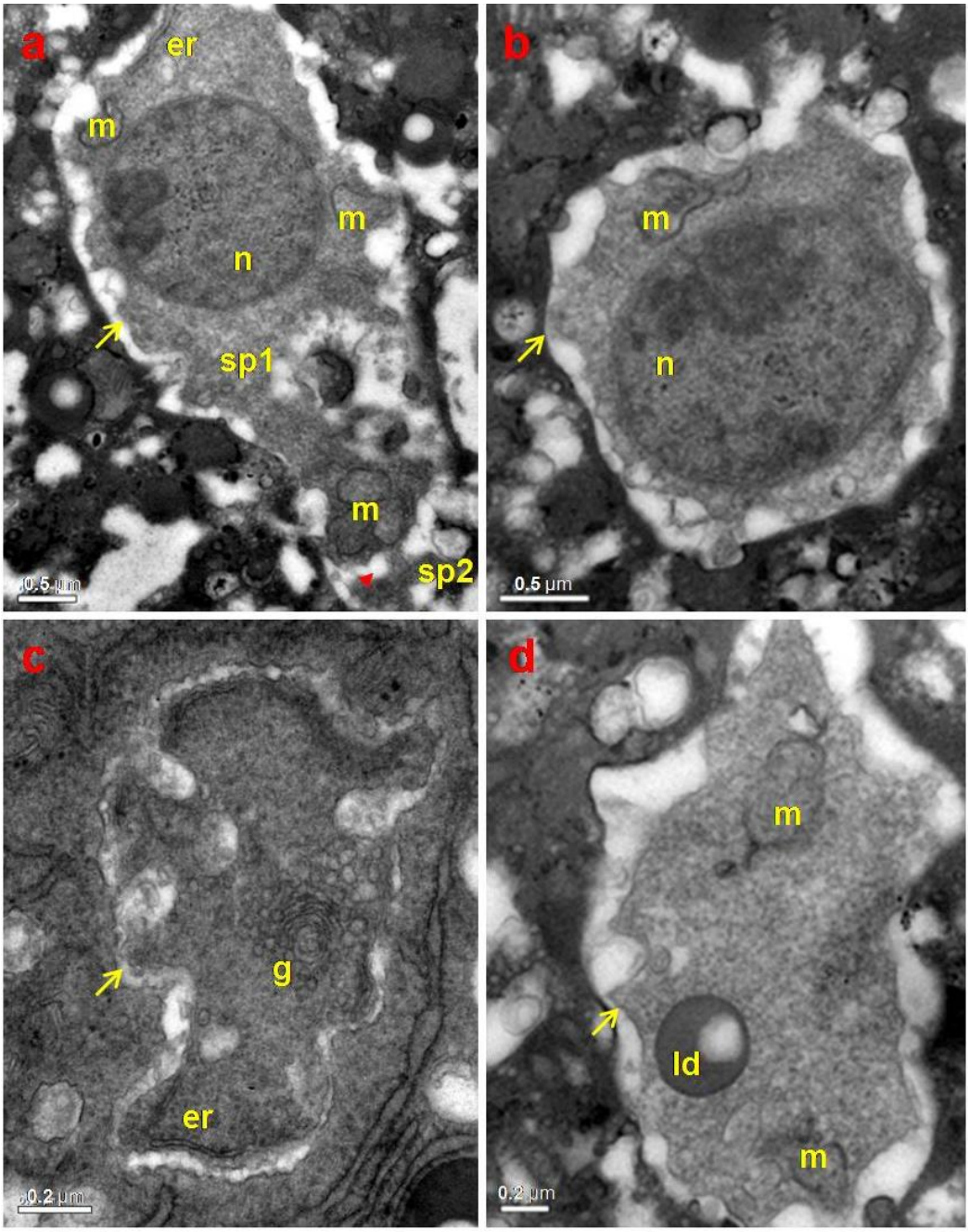
3.2.6 Sperm cell measurement utilising confocal microscopy and Imaris 6.1 software (Bitplane)

The sperm cell morphology studied utilising TEM described in the previous section revealed the organellar composition and structure of the two sperm cells when packed in the pollen grain. However, due to the limited numbers of pollen grains that can be observed utilising TEM the measurement of sperm cells could not be accurately performed. Therefore an alternative approach was taken utilising a method that allowed the visualisation of a large number of pollen grains to produce data that could be validated by statistical analysis. A GFP-tagged sperm cell line (a generous gift from Dr. Scott D. Russell, University of Oklahoma) was used in this project for the measurement of Arabidopsis sperm cells. The two GFP-positive sperm cells in the hydrated pollen grain were visualised and photographed as serial images by confocal microscopy

(Nikon ECLIPSE 90i) for 3D reconstruction (Figure 3.8). These serial photos were analysed utilising Imaris 6.1 software (Bitplane) for volume and area calculations. The two sperm cells were measured as two separated objects, however unable to determine the Sua and Svn. A total of 200 sperm cells from 100 grains were measured and the data is presented as histrogram (Figure 3.10). The volume and surface area data for the sperm cells is reported in the table below (Table 3.5).

Figure 3.6. Transmission electron micrographs of *Arabidopsis thaliana* sperm cells.

Hydrated pollen grains of *Arabidopsis thaliana* were fixed and sectioned for TEM. Serial sections of five grains were observed under TEM. The micrographs demonstrated the different axis of the position of the two sperm cells (sp) in the same grain. One sperm cell was sectioned longitudinally (a) whereas the other sectioned transversely (b). The sperm cell micrographs (a-d) illustrated a lobed plasma membrane (arrow) and various organelles i.e. nucleus (n), Golgi body (g), endoplasmic reticulum (er), lipid droplet (ld), and mitochondria (m). Mitochondria were the most abundant organelle found in the sperm cell with at least five being noted per cell (red arrow head = connection between the two sperm cells, a and b bar = 0.5µm, c and d bar = 0.2µm).


a
er
m
m
n
sp1
m
sp2
0.5 µm
b
m
n
0.5 µm
c
g
er
0.2 µm
d
m
ld
m
0.2 µm

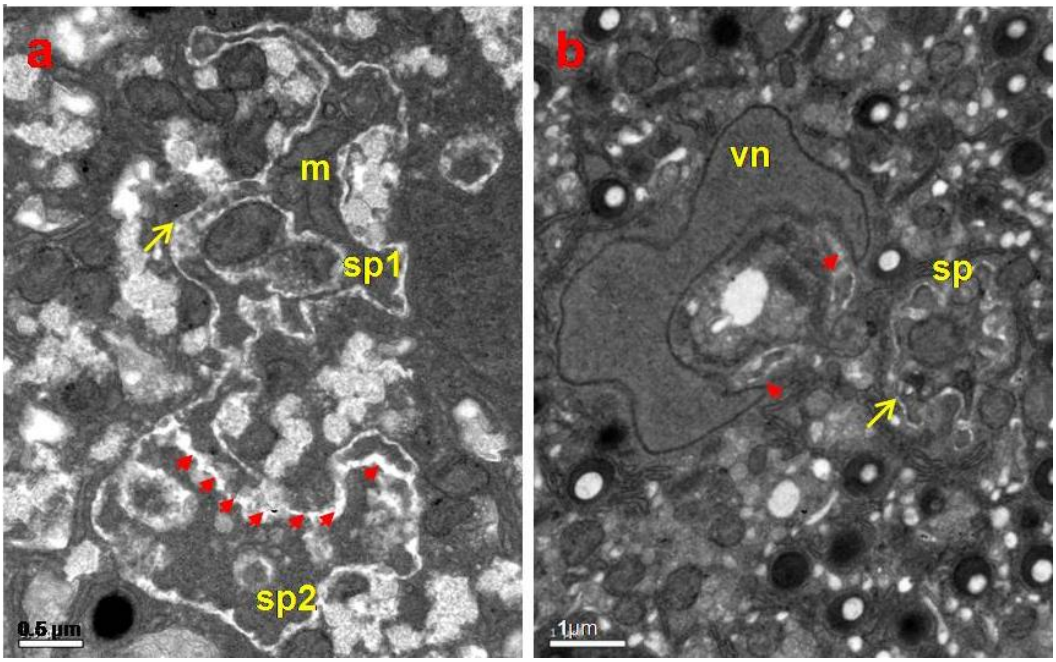


Figure 3.7. Transmission electron micrographs of *Arabidopsis thaliana* sperm cells demonstrating the connections of the male germ unit.

TEM micrographs of a hydrated pollen grain demonstrating the male germ unit (two sperm cells and vegetative cell nucleus). a) The 'connections' between the two sperm cells appeared to be made up of the plasma membranes of the two sperm cells aligning closely with one another (a, red arrows). b) The link between one sperm cell and the vegetative nucleus also appeared as a simple attachment between the sperm cell plamsa membrane and the vegetative cell nuclear membrane (b, red arrows) a bar = 0.5µm, b bar = 1µm, sp: sperm cell, vn: vegetative cell nucleus, m: mitochondria, yellow arrow: sperm plasma membrane).

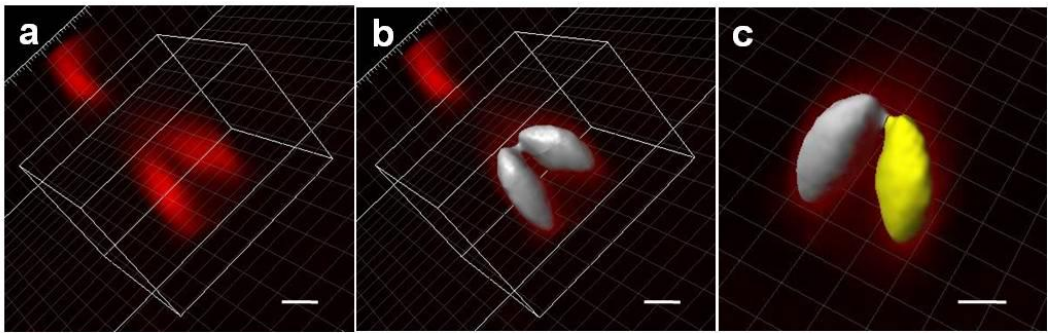


Figure 3.8. 3D reconstruction of sperm cell serial sections obtained by confocal microscopy.

Serial sections of GFP-positive sperm cells (a) were captured using a confocal microscope (Nikon ECLIPSE 90i). 3D images were then reconstructed (b) using Imaris 6.1 software (Bitplane). Cell volume and surface area of the two sperm cells of *Arabidopsis thaliana* were automatically measured and reported as separate objects (c) highlighted in two different colours by the software (scale = 2 μm).

Table 3.5. Cell volume and surface area measurement of the two sperm cells of *Arabidopsis thaliana* utilising two different techniques.

Cell volume and surface area of the sperm cells of *Arabidopsis thaliana* were measured by two different methods. The first method was performed utilising UDRuler software (AVsoft) of isolated sperm cells (metal roller method) visualised by DIC microscopy. The diameter of spherical isolated sperm cells was measured using the software and the volume and area of the sperm were calculated using standard equations. The second technique was automated utilising Imaris 6.1 software (Bitplane). This technique was performed with GFP-positive sperm cells still residing in intact pollen grains and visualised by confocal microscopy. Volume and surface area of the sperm cells were plotted into histograms to test the size distribution.

Parameter	**UDRuler**		**Imaris 6.1**
	Bo-Sp (n=100)	At-Sp (n=200)	At-Sp (n=200)
Cell surface (μm^2)			
Range	47.78−95.03	16.62−58.09	22.23−65.98
Mean	76.64 ± 9.58	35.38 ± 17.94	40.88 ± 8.28
Cell volume (μm^3)			
Range	31.06−87.12	6.37−41.63	9.03−40.95
Mean	63.47 ± 11.56	20.27 ± 16.84	19.10 ± 5.56

Bo; Brassica oleracea, At; Arabidopsis thaliana, Sp; sperm cell

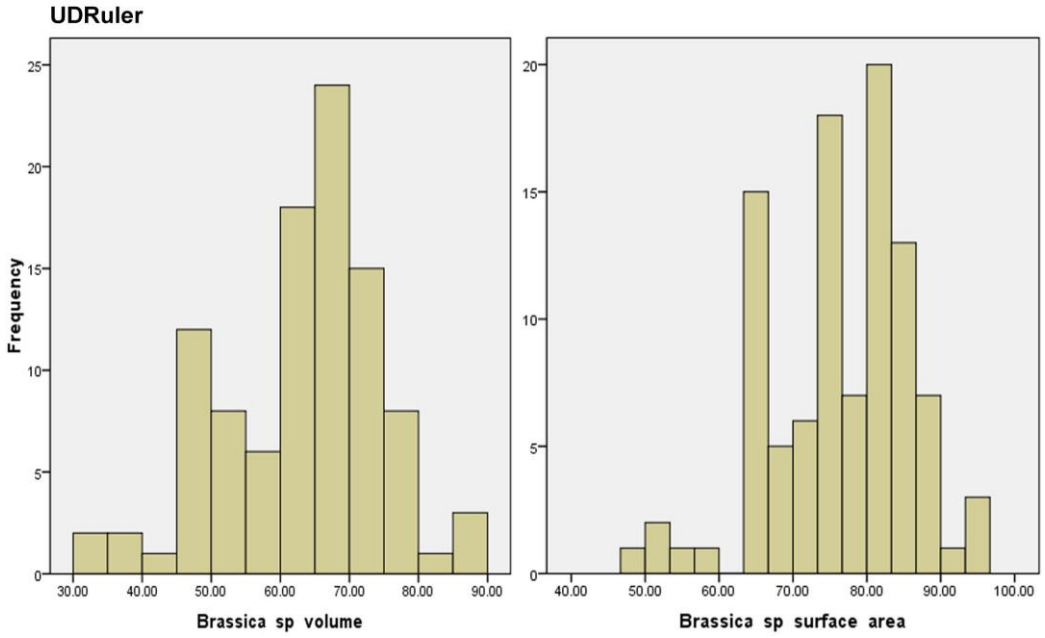


Figure 3.9. Histograms of Brassica sperm cell volume and surface area measured by UDRuler software.

The isolated Brassica sperm cell (n=100) micrographs were measure utilising UDRuler. The sperm cell volume and surface area were calculated and plotted into a histogram to determine the sperm size dimorphism. The distributions of all histograms indicates no differences between the two sperm cell population.

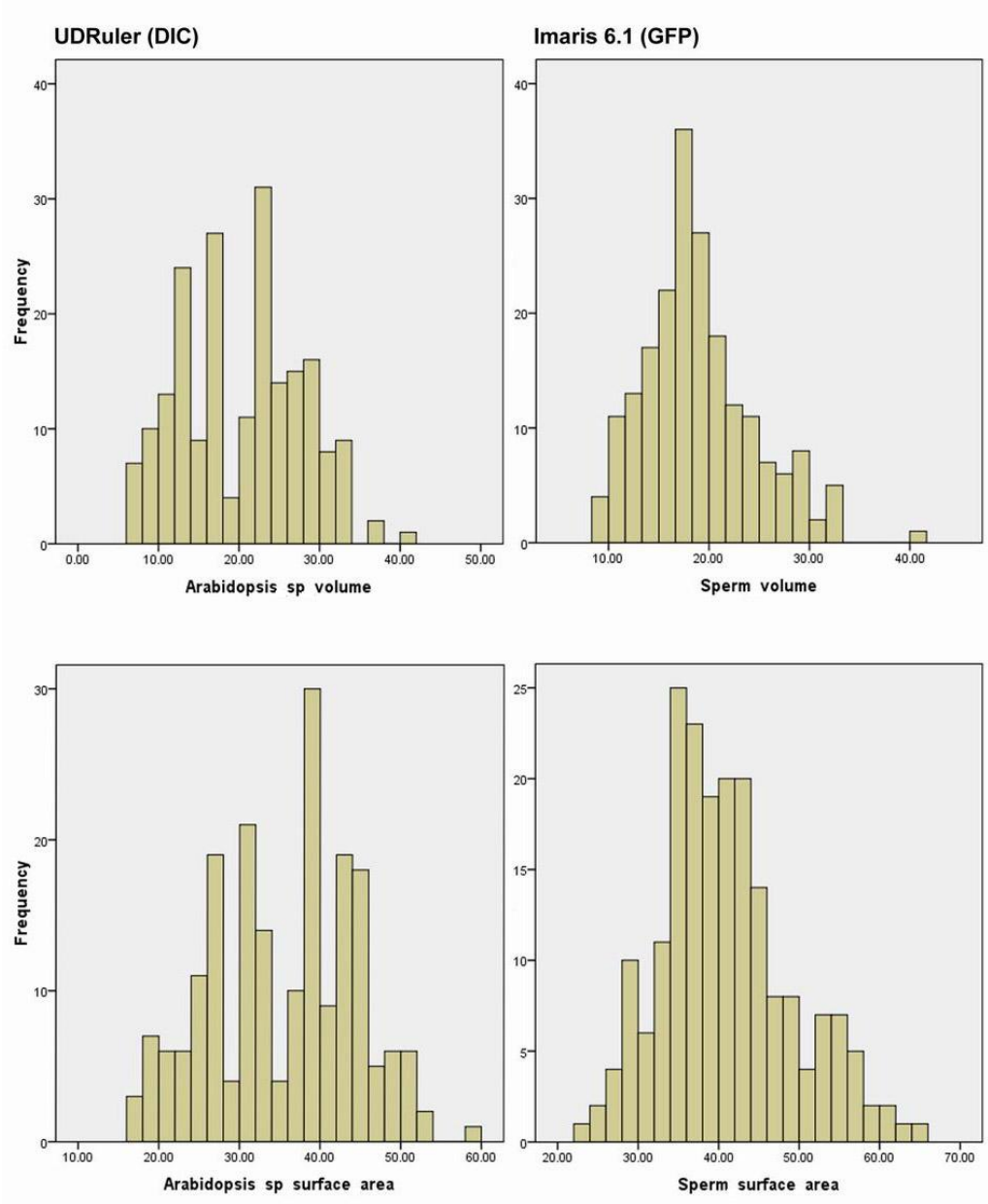


Figure 3.10. Histograms of Arabidopsis sperm cell volume and surface area measured by UDRuler and Imaris 6.1 software.

The isolated (n=200) and unisolated (n=200) GFP sperm cell micrographs were captured using DIC and confocal microscopies respectively. The sperm cell volumes and surface areas were measured utilising UDRuler (DIC) and Imaris 6.1 (GFP). The cell volume and surface area were calculated and plotted into a histogram to determine the sperm size dimorphism. The distributions of all the graphs indicate that there were no differences between the two sperm cells measured by both techniques.

3.3 Discussion

3.3.1 *B. oleracea* and *A. thaliana* pollen disruption and sperm cell isolation

3.3.1.1 Brassica oleracea

Various techniques have been used in this study to disrupt the pollen grain of *Brassica oleracea* in order to allow the release of sperm cells. Osmotic shock has been used as the main procedure, as the sperm cells have lower solute concentration than vegetative cells (Southworth and Morningstar, 1992), coupled with other techniques to efficiently rupture the pollen grain. The disruption techniques, including physical disruption (homogenization, sonication and pressurisation), and enzymatic treatment of the grain were found to have different efficiencies regarding sperm release. As is apparent in table 3.2, the best method of pollen disruption is sonication at 30 µm amplitude with the longest duration (30 seconds) being optimal. However this technique resulted in low sperm viability, assessed in this case by viewing the intactness of the sperm cells by DIC microscopy. On the other hand, the method which gave the best results for isolating and releasing viable sperm cells was rupturing the grains with a metal roller. This method not only efficiently disrupted the pollen grain but was also found to be the best way of displacing the sperm cells from the broken pollen grain. In other disruption methods sperm cells were frequently found to remain in the ruptured pollen grain. The release of such sperm cells from the pollen grain was found to be problematic even though the disrupted pollen sample was allowed to be agitated in sperm cell isolation media for a short period of time after disruption. This problem was also reported for other species (Russell, 1991). One factor reported to affect coagulation of the pollen cytoplasm and consequently sperm cell release was the high sucrose concentration of the isolation media (Russell, 1986). The sucrose concentration for *B. oleracea* sperm isolation in this study was optimised to be ideal for pollen bursting and also for preserving the viability of the isolated sperm cells. Despite the metal roller method being the best technique for sperm cell release it was inconsistent and proved difficult to apply to larger sample sizes. The best method to obtain a large number of disrupted pollen grains and a relatively good yield of intact sperm cells was the pressurization / depressurisation technique using the Parr pressure compartment. This technique not only provided a very efficient way of isolating sperm cells but could also be used with a large sample size which is ideal for sperm cells study in the future. Once the techniques were successfully developed in *B. oleracea* they were adapted to be used for Arabidopsis.

3.3.1.2 Arabidopsis thaliana

In Arabidopsis the percentage of successfully disrupted pollen was lower following processing with the Parr nitrogen pressure compartment compared to *B. oleracea*. This is probably due to the fact that the pollen of Arabidopsis is smaller and therefore less likely to be disrupted on pressure release. To overcome this problem the pollen was treated with the cell wall degrading enzymes, cellulase and pectinase, followed by osmotic shock to aid the disruption. Together these treatments were successful in improving the release of sperm from the pollen grain. Sperm cells isolated from both species were often found in pairs immediately after pollen disruption (Figure 3.3 h-j). However, the sperm cells separated from one another approximately 5-10 minutes later as described by Russell, 1991. After successfully developing this relatively good method for pollen disruption (osmotic shock coupled with enzymatic treatment and pressurisation/depressurisation), the next challenge was to find techniques to facilitate sperm cell isolation and purification. After pollen disruption the sperm containing fraction is heavily contaminated with other pollen debris. In order to obtain relatively clean sperm cell fractions for further studies various techniques have been developed to isolate the sperm from the pollen debris including filtration, fractionation on Percoll gradients and centrifugation (Russell, 1991).

3.3.2 Partial purification of sperm cell

In this study Percoll discontinuous gradient centrifugation was one of the methods used to separate the sperm cells from other pollen debris. Sperm cells were successfully trapped on the 40% layer of the gradient. Following this, separation of sperm onto a 40% Percoll cushion was performed and its effectiveness compared with separation on a Percoll gradient. Sperm cell yield was found to be similar between the two methods however the level of contamination in the sperm sample on the cushion was higher compared to the sample arrested on the Percoll gradient. In addition a filtration method was tested for the purification of sperm cells. Nylon mesh sizes of 20 and 10 µm were used that permitted sperm cells to pass through whilst capturing pollen wall material. This filtration technique was not only effective at excluding the pollen wall but also facilitated the separation of sperm cell pairs and removal of pollen cytoplasmic strands which often remained attached to the isolated sperm cells. The number of intact sperm cells isolated from this method was also higher than for the Percoll techniques. However although sperm yield was high the sample was heavily contaminated with pollen cytoplasmic material as revealed by DIC microscopy. A centrifugation technique was then used to reduce the level of contamination from small cellular debris. Although this

was relatively successful the number of sperm cells also reduced as the purification steps proceeded. The longevity of isolated sperm cells was then assessed in the 7 different sperm storage solutions (see appendix B). All buffers used in this project were sufficient to preserve sperm viability immediately after isolation. However, once the time progressed to 6 hours post isolation, sperm cell viability dramatically decreased in all buffers except M5. Moreover the percentage of sucrose contained in the buffer proved to be the main factor affecting sperm cell longevity. Despite better sperm cell preservation at higher sucrose concentrations, the sucrose concentration was adjusted to 12% for pollen hydration and then reduced to 8% for pollen bursting – dropping the concentration to 8% also facilitated enrichment of Arabidopsis sperm cells by centrifugation in this study. MES was also included in the media as a buffer since stabilisation of pH has been proven to be very important in sperm cell survival (Zhang et al., 1992a).

3.3.3 Isolated sperm cell quality assessment and sperm cell purification by FACS

The intactness of purified *B. oleracea* sperm cells was assessed by DIC microscopy and Hoechst staining which together were considered good indicators of cell viability. In Arabidopsis the assessment was carried out by both DIC microscopy, Hoechst staining and by the use of the Arabidopsis GFP-positive sperm plant line obtained from Dr. Scott D. Russell (University of Oklahoma). The GFP signal emitted by the sperm cells could be detected throughout the whole process of pollen disruption, sperm cell isolation and purification. However, the percentage of the GFP-positive sperm cells was approximately 25% lower compared to the number of cells staining with Hoechst indicating that some of the Hoechst stained cells observed were non-living sperm cells or free sperm cell nuclei. Semi-purified Brassica and Arabidopsis sperm cells were successfully purified further using FACS. Again the purity of the sample and quality of the FACS-sorted sperm were analysed using DIC microscopy. Moreover, a more accurate method for purity assessment – reverse transcriptase-PCR (RT-PCR) was used to assess the purity of FACS-sorted sperm cells. The sperm cell preferentially expressed gene, *GEX2* (Engel et al., 2005) was used along with three pollen vegetative cell specific genes, *TUA1*(Carpenter et al., 1992), *PTEN1* (Gupta et al., 2002) and *AT59* (Kulikauskas and McCormick, 1997) to assess the sorted sperm sample. Only a *GEX2* product was amplified when RNA extracted from sorted sperm sample was used as template. This RT-PCR result confirms that the FACS-sorted sperm cells were essentially free from pollen cytoplasmic contamination. These purified sperm cells were used in a proteome study in this project which will be described in chapter 4. Moreover,

these FACS-purified sperm cells will also be useful in the future for other molecular studies e.g. transcriptional profiling by microarray and cDNA library construction.

3.3.4 Morphological study of *Arabidopsis thaliana* sperm cells utilising transmission electron microscopy (TEM)

As mentioned in section 3.1.1, cytological and morphological studies of plant sperm cells have been performed in various plant species utilising EM e.g. *Plumbago zeylanica* (Southworth et al., 1997), *Zea mays* (Mogensen et al., 1990; Southworth et al., 1988), *Spinacia oleracea* (van Aelst et al., 1990), *Nicotiana tabacum* (Yu and Russell, 1994b), and *Brassica oleracea* (McConchie et al., 1987b). However sperm cell cytological studies have not been reported in Arabidopsis. In this project an ultrastructural analysis of Arabidopsis sperm was carried out by TEM. Various plant species i.e. *Plumbago zeylanica, Zea mays, Nicotiana tabacum*, and the Arabidopsis relative *Brassica oleracea*, have been reported to have dimorphic sperm cells (McConchie et al., 1987c; Russell, 1984b; Shi et al., 1996; Wagner et al., 1989; Yu and Russell, 1993). The two sperm cells in these dimorphic species could differ in size, organellar content or chromosome content. The different size sperm cells of *Plumbago zeylanica* for instance contain different numbers and types of organelles (Russell, 1984) whereas in *Zea mays* the chromosome content of the two cells differs with B-chromosomes being found in the sperm cell that fuses to an egg cell (Shi et al., 1996; Wagner et al., 1989). In tobacco, the two sperm appeared similar in organellar content but differ in size (McConchie et al., 1987b; Yu and Russell, 1993). In *Brassica* spp, the two sperm cells differ in size and in *B. oleracea* the sperm Svn contains slightly more mitochondria than Sua (McConchie et al., 1987b). In this project, serial ultra thin sections of Arabidopsis pollen grains and the resident sperm cells were obtained for TEM analysis. This study successfully revealed the organellar content of the two sperm cells and their close association with one another and to the vegetative cell nucleus. The two sperm cells were revealed to have an elongated shape with a strikingly evaginated plasma membrane contrasting with the smooth spherical cellular structure associated with isolated sperm cells. The two sperm cells appeared to be either linked or intimately associated with one another and to the vegetative cell nucleus. From the TEM micrographs it appeared that the two sperm cells may share the same double layered plasma membrane with no cytoplasmic connection being observed. Alternatively the plasma membranes of both cells could be intimately aligned with one another, remaining unfused. The observed link between the two sperm cells and vegetative nucleus (defined as the male germ unit - MGU) has been observed in all plant species studied so far and has been demonstrated to facilitate gamete delivery (Dumas et al.,

1984; Lalanne and Twell, 2002). Alteration of sperm cell shape from elongated shape to spherical after isolation is also a universal occurrence (Russell, 1991) along with the loss of the association between the sperm cells and vegetative nucleus (also observed in this project) suggesting a weak linkage between these cells. It has been claimed that the sperm cell pairs of at least two species, *Plumbago zeylanica* and *Nicotiana tabaccum* are linked by a common cross wall (Russell et al., 1996; Tian and Russell, 1998). This wall, which stained with aniline blue, was proposed to be present initially around the generative cell. The fluorescent signal from aniline blue reduced gradually during the generation of sperm cells, was detected only in the cross wall of immature sperm cells and disappeared in mature sperm cells (Russell et al., 1996). In a sperm cell fusion study in tobacco, treatment of sperm with pectinase and cellulase enzymes was found to be a crucial factor for sperm cell fusion and the separation of the two freshly isolated sperm cells. This suggests that insoluble polysaccharide materials exists in the form of a highly modified periplasm on the surface of the sperm cells which could be involved in the connection between them and also have a role in preventing the two sperm cells fusing in the pollen grain (Tian and Russell, 1998). In addition a link between sperm cells was reported in *Torenia fournieri*, a species with bicellular pollen (Chen et al., 2006) where freshly isolated sperm pairs of were observed to be surrounded by an enveloping membrane (Chen et al., 2006). Once the isolated sperm pairs were transferred into pectinase and cellulase enzymes solution, the two cells were rapidly separated (Chen et al., 2006). In another bicellular species, tobacco, the progressively rounded shape of isolated sperm cells was reported to be associated with the loss of pollen tube plasma membrane (Zhang et al., 1999). Therefore the link between the two sperm cells is likely to be the result of a number of mechanisms, polysaccharides in a periplasmic 'wall' initially surrounding sperm, interactions between sperm cell plasma membranes and a vegetative cell derived membrane that may surround the two sperm cells. In 2002, proteins proposed to be involved in the MGU connection were identified in Arabidopsis (Lalanne and Twell, 2002). Two mutants affecting organisation of the MGU, *germ unit malformed (gum)* and *MGU displaced (mud)* were isolated (Lalanne and Twell, 2002). The *gum* mutants demonstrated a vegetative cell nucleus and Svn connection via an extension of the vegetative nucleus (Lalanne and Twell, 2002). The *mud* mutants demonstrated a highly compact displaced MGU. According to the functions of these two genes demonstrated in the publication, it is possible that the GUM and MUD proteins have an important role in the male germ unit association (Lalanne and Twell, 2002). The two sperm cells observed in this project by TEM were often in opposite orientations. While one sperm cell was positioned along one axis (x), the other would always be situated along the other axis (y or z). Similar positioning of the two sperm cells was also reported in *B. campestris* and *B. oleracea* pollen

(McConchie et al., 1987b). Obtaining a complete stack of serial sections containing both sperm cells was crucial for a meaningful comparative analysis. However, obtaining such a complete series proved to be problematic due to a range of factors. Sperm cells are approximately 2 μm thick and 4μm long as observed by TEM (Figure 3.6) and therefore a minimum of 20 sections (each section 100 nm thick) were required for a complete cell reconstruction, depending on the exact orientation of the cells. As the two sperm cells were always in the opposed position, more than 40 sections were required for a complete reconstruction of the two cells (considering the minimum diameter of the sperm cell is 2 μm) (Figure 3.6). When these sections (4-5 sections per grid) are placed on the grids for EM, they have to be carefully positioned in the middle with no overlaps between the sections. There were some cases where the grid (made of thin transparent fibre) was destroyed by the electron beam resulting in at least 4-5 sections (400-500 nm) being lost from the analysis and therefore the information for that particular sperm cell was incomplete. Unfortunately, due to time and financial constraints of this project, full serial sections were only obtained for a few sperm pairs. However, despite these problems good data was obtained for these samples allowing a detailed picture of the ultrastructure of Arabidopsis sperm cell pairs to be obtained. Analysis of the micrographs suggested that in Arabidopsis the two sperm cells of a given pair are similar in their overall structure and organellar content. Both sperm cells contained a large nucleus which filled almost half of the cell (Figure 3.6). Both cells lacked microtubules and plastids as reported in *Brassica* spp (McConchie et al., 1987b). A few lipid droplets were found in some of the sperm cells analysed, along with endoplasmic reticulum, Golgi complexes and mitochondria. Typically 3-7 mitochondria were found in the sperm cells (Figure 3.6). The cellular structures of Arabidopsis sperm cells were similar to it relative, *Brassica oleracea*. Thus the sperm cells of both species demonstrated an ellipsoid shape, mitochondria and a lack of plastids (McConchie et al., 1987b).

3.3.5 Sperm cell volume and surface area measurements

The size and dimorphism of sperm cells have been reported for various species i.e. *Plumbago zeylanica* (Russell, 1984), maize (McConchie et al., 1987b; Mogensen et al., 1990), tobacco (Yu and Russell, 1993), and *Brassica* (McConchie et al., 1987b). However, sperm morphology has not been reported in detail for Arabidopsis. In this project, the sperm cells from both *B. oleracea* and *A. thaliana* were isolated and measured using USDRuler (AVsoft) software. Moreover, measurement of GFP-positive sperm cells in intact pollen using confocal microscopy with an advance imaging

software Imaris 6.1 (Bitplane) was also carried out to determine whether dimorphism occurred in Arabidopsis.

3.3.5.1 Size analysis of isolated B. oleracea and A. thaliana sperm cells using USDRuler software (AVsoft)

Sperm cells of *B. oleracea* and *A. thaliana* were isolated and observed using DIC microscopy and photographed. The spherical isolated sperm cells from both species were then measured using USDRuler (AVsoft). The mean sperm cell volume and surface area of *B. oleracea* was 63.47 μm^3 and 76.64 μm^2 respectively with a mean diameter of 4.9 μm. The volume of Brassica sperm measured in this project was very different from the volume reported by McConchie and colleagues (1987) where sperm pairs, still retained in the pollen, were measured separately using TEM micrographs (Sua = 14.81 μm^3, Svn = 17.53 μm^3) (McConchie et al., 1987b). However, it is not surprising that the figures are different for these two studies as isolated sperm cells were stored in isolation media. On isolation the shape of sperm cells change from an evaginated compact spindle shape to spherical (Russell, 1986). Thus the convoluted surface of the sperm cells will expand upon isolation and consequently the cell volume is expected to increase. On the other hand, as expected, the surface areas measured in the two studies were much more similar (McConchie et al., 1987b; Sua surface area = 46.17 μm^2, Svn = 65.10 μm^2) as the convolutions of the packed sperm cells were also included in the calculations (McConchie et al., 1987b). In Arabidopsis, as in Brassica, the two sperm cells were unable to be measured in pairs using this technique. Therefore, to be able distinguish the two cells a different technique was adopted which involved measuring sperm in intact pollen grains (for discussion see section 3.3.5.2).

3.3.5.2 Measurement of A. thaliana GFP-positive sperm pairs in vivo using confocal microscopy and image analysis software (Imaris 6.1 - Bitplane)

In this study, reconstruction of confocal image stacks permitted 3D analysis of GFP-positive sperm cell pairs resident within pollen grains. Only pollen grain-resident sperm cells were used for this technique as the size of isolated sperm cells are altered due to the different osmolarities between sperm cells and isolation media (sperm cells lack a cell wall that would normally constrain size). In other species, the size of isolated sperm cells has also been reported to be different from the *in vivo* sperm cells (Russell, 1991). A total of 100 pollen grains (200 sperm cells) were measured in pairs for comparison. The sperm within a measured sperm pair, with only a few exceptions, were of different

sizes. Average volume and surface area measurements of the GFP Arabidopsis sperm cells were recorded as 19.10 μm^3 and 40.88 μm^2 respectively. The volumes and surface area of the sperm cells were plotted into histogram to identify the size distribution and were concluded to be no different from one another within the population. However, this could possibly be because the differences between the two populations are very subtle due to the small size of the sperm. The differences between the volume of the two sperm cells of B. oleracea were only 2.7 μm^3 (McConchie et al., 1987b). Therefore, the dimorphism of the Arabidopsis sperm cells might not be able to be detected using these methods. The best way of Arabidopsis size dimorphism would be to distinguish the Sua from Svn and measure them separately. One way to distinguish the two cells is to stain the GFP-positive sperm cell pollen with DAPI and try to identify the MGU connection using fluorescent light microscopy before collecting sperm cell 3D data with confocal microscopy. However, the GFP signal in sperm cell is unstable and could be demolished during the MGU identification. The other method to distinguish the Sua from Svn is to use MGU proteins (Lalanne and Twell, 2002), MUD or GUM-GFP fusion line for sperm cell size measurement. This will allow the data collection of all three cells (Svn, Sua, and vegetative nucleus) simutaniously. This 3D data then can be analysed directly using Imaris 6.1 software. Interestingly, Arabidopsis sperm cell volumes measured from GFP sperm cells in this project appeared slightly larger than the Brassica sperm cell volumes reported by McConcchie and colleagues (McConchie et al., 1987b). This is possibly because the bright GFP signal from the sperm cells masked the convoluted nature of the cell surface making it appear smooth and thus resulting in its volume appearing larger.

A. thaliana and *B. oleracea* sperm cell isolation and purification techniques developed in this study provided good quality semi-purified sperm cells suitable for use in molecular studies which will be described in the next few chapters. Moreover, the purification of sperm cells by FACS proved to be highly successful providing pure sperm cells (justified by RT-PCR) that could be used in both transcriptomic and proteomic studies. In addition, Arabidopsis sperm cell morphology was also investigated in this study. TEM micrographs of Arabidopsis sperm cell pairs demonstrated their cellular ultrastructure and position in the pollen grain. Sperm cell volume and surface area of *A. thaliana* were measured from both isolated sperm cells and *in vivo* GFP-positive sperm cells. The *in vivo* analysis revealed size isomorphism of Arabidopsis sperm cells.

Chapter 4:

Proteome Analysis of *B. oleracea* and *A. thaliana* Sperm Cells

4.1 Introduction

In 2003, Engel and colleagues were the first to successfully construct a plant sperm cell-specific cDNA library (Engel et al., 2003). Fluorescence-activated cell sorting (FACS) was utilised to purify the sperm cells from *Zea mays* pollen. From this study approximately 1,100 transcripts were annotated in the resulting sperm EST database. Later in 2006, generative cells, the progenitors of sperm cells, were isolated from *Lilium longiflorum* and used to construct a cDNA library in which 886 ESTs were generated (Okada et al., 2006). From these transcripts many were found to be redundant, some were non-coding RNA with the most abundant class being retroelements which are generally not important for cellular function. A large number of transcripts were involved in standard cell regulatory processes i.e. DNA synthesis, chromosome structure and cellular metabolism (Engel et al., 2003; Okada et al., 2006). Despite the fact that transcriptome data has been generated for sperm cells from at least two angiosperm species, only a few genes expressed exclusively in male germ line have been characterised utilising the data from these cDNA library studies i.e. *LGC1* (Xu et al., 1999b), *gH2B* and *gH3* (Xu et al., 1999a) in lily, and *AtMGH3* (Okada et al., 2005), *GEX2* (Engel et al., 2005), and *GCS1* (Mori et al., 2006) also known as *HAP2* (von Besser et al., 2006) in Arabidopsis. The exceptionally low number of genes that, to date, have been found to be specifically expressed in sperm cells is perhaps explained by the difficulties in analysing and prioritising potential sperm-specific candidates from the large amount of transcriptome data and redundancy obtained from the library. An approach that could be used to overcome the problems associated with transcriptome studies is proteomic analysis.

For the past decade proteomics has been used extensively to characterise cell molecular composition, often in preference to transcriptomics, as only translated mRNA sequences which more accurately reflect the functional character of the cell will be revealed in the studies. However, the main challenges in proteomic studies are obtaining cells in the requisite quantity and purity for this type of analysis. Nonetheless, proteomic studies have been made possible and relatively easy for the pollen grain, the structure that sperm cells inhabit. Pollen grains are an ideal 'tissue' for proteomic studies as generally they can be isolated in large numbers and to a high degree of purity from the flower. However, in Arabidopsis, pollen collection can be a problem as the plants are small and each produces a miniscule amount of pollen. Nonetheless, pollen collection strategies have been successfully developed for this species. In this study, a vacuum collection system has been used to harvest the pollen from a large number of plants for use in sperm isolation (Johnson-Brousseau and McCormick, 2004). The technique involves suction and filtration by using a vacuum cleaner which contains three nylon filters, two large mesh filter sizes to exclude plant debris and another smaller mesh size to trap the clean pollen grains (Johnson-Brousseau and McCormick, 2004). Collection of adequate quantities of pollen grains from Arabidopsis has permitted a number of laboratories to conduct proteomic studies. Three independent research groups have reported *A. thaliana* pollen proteome data so far (Holmes-Davis et al., 2005; Noir et al., 2005; Sheoran et al., 2006). The methods used to obtain the proteins form the pollen are summarised in table 4.1 and as illustrated the amount of pollen and the protein isolation methods used were diverse - this probably explains the variation in the number of protein spots and spot resolution obtained between the studies. However, despite these differences there was no significant effect on the main classes of proteins identified in these analyses (Table 4.2). Nonetheless, as the amount of starting material and pollen disruption methods were similar in each experiment, it is likely that the differing buffers used are responsible for the variation in the protein profiles obtained (affecting subsequent fraction separation and precipitation steps). Interestingly, the amount of protein chosen to be loaded on the gel was very different in each experiment, 18 (Holmes-Davis et al., 2005), 100 (Noir et al., 2005) and 600 µg (Sheoran et al., 2006). Resolution of the gel in each experiment were affected greatly by the amount of loaded protein, however the number of protein spots selected for MS were similar in each experiment (179 spots in Holmes-Davis et al., 2005, 145 spots in Noir et al., 2005, and 150 spots in Sheoran et al., 2006). Despite the loading differences between each experiment, approximately 75-80% of the selected spots were identified in all three experiments. Holmes-Davis and colleagues identified 135 distinct proteins from 179 spots. Of these proteins, 20% were involved in metabolism, 17% energy generation, and 12% cell structure. Significantly these percentages are similar to those obtained from a pollen transcriptome study (Craigon et al., 2004), although nine proteins were identified by Holmes-Davis et al. that were not detected by the transcriptomic approach (Holmes-Davis et al., 2005). Noir *et al.* identified 121 distinct proteins resolved from 145 spots. Many of these proteins could

be categorised as being involved in energy regulation (24%), metabolism (14%), and a relatively high number were proteins of unknown function (20%) (Noir et al., 2005). The study by Sheoran and colleagues positively identified 110 from 150 spots, 66 of which were not found in the previous two studies (Sheoran et al., 2006). The protein spots identified by MS from each study were categorised and compared in table 4.2 (Holmes-Davis et al., 2005; Noir et al., 2005; Sheoran et al., 2006). Of these three individual studies, 17 proteins were common to all three, 35/26 proteins were overlapping with one or other of the two studies. As mentioned above a number of factors could explain the differences in the data obtained between the studies including plant ecotype, the amount of starting material, the protein extraction methods and buffers used and the protein spot selection (Sheoran et al., 2006). Despite the relative success of pollen proteome studies, proteomic analyses of purified plant sperm cells have yet to be reported. Nonetheless, in 1988, a comparative protein study of the pollen grain and male germ unit (MGU - two sperm cells and vegetative nucleus) enriched fraction of *Plumbago zeylanica* was carried out (Geltz and Russell, 1988). These protein fractions were subjected to two-dimensional electrophoresis. In the MGU-enriched fraction, 427 spots were revealed. In the pollen cytoplasmic-particulate and cytoplasmic water soluble proteins fractions, 515 and 285 spots were shown respectively. In total, 133 distinct spots were compared, with 18 spots being unique to the MGU fraction. 11 of the MGU-specific proteins were large being greater than 116 kDa, whereas the remaining 7 spots were smaller polypeptides of less than 80 kDa (Geltz and Russell, 1988). Unfortunately these proteins were unable to be completely characterised due to the fact that the *Plumbago zeylanica* genome has not yet been completely sequenced. Proteomic studies in animal systems have seen some recent success. Relatively large quantities of animal sperm cells can be obtained from the seminal vesicle and Dorus and colleagues exploited this fact to produce an elegant study of the Drosophila sperm proteome (Dorus et al., 2006).. The protein extracted from the sperm cells in this study were not only characterised by 2-D protein electophoresis but also identified directly using whole-sperm mass spectrometry (WSMS) The proteomic techniques successfully performed on animal sperm cells could potentially be adapted for use in plant sperm cell studies as similarities between these specialised cell types across kingdoms have been established to a certain extent (Till-Bottraud et al., 2005). The general nature of sperm cells and their function i.e. compact cells with condensed chromatin and a small amounts of cytoplasm, relatively transcriptionally inactive and particularly the fact that both plant and animal sperm lack a cell wall (Russell, 1991; Russell, 1996; Till-Bottraud et al., 2005) suggest that cross-kingdom comparative proteomic studies may yield very interesting data. Moreover, the egg-sperm interaction event across the kingdoms is relatively comparable (Berger, 2008; Till-Bottraud et al., 2005). In the Drosophila study, sample purity was critically important for the WSMS protein identification technique. Unlike conventional 2-D gel mass-spectometry, the proteins present in the sample can be identified directly by MS without the 2-DE step in which some proteins can be lost or rendered undetectable in the process (Dorus et al.,

2006; Roe and Griffin, 2006). This new WSMS technique could potentially be an excellent method for plant sperm cell proteomic studies once a purification technique for them is successfully developed. With the existing transcriptomic and proteomic resources an Arabidopsis sperm cell proteomic study would be less complicated and be developed more rapidly than previously was possible.

The protein content of sperm cells isolated from *B. oleracea* and *A. thaliana* (Col) pollen mentioned in the previous chapter will be further analysed and described in this chapter. Protein extracts from the sperm cell-enriched sample were examined first by SDS-PAGE and compared to whole pollen samples and sperm cell-excluded samples. These protein samples were then analysed further by 2-DE to obtain greater resolution and separation of individual polypeptides followed by mass spectrometry (MS) of selected proteins. Peptide Mass Fingerprints (PMF) obtained from the MS analysis were then searched against the Arabidopsis protein database. FACS sorted sperm samples were also subjected to 2-DE and MS analysis. In addition the protein 2-DE spot profiles obtained in this study will be compared to existing pollen proteome data discussed earlier in the chapter. The existing pollen proteome resource will hopefully facilitate the identification of sperm-specific proteins which could potentially be important in the fertilization process. As sperm purification techniques have been successfully performed and reported (Chapter 3), the next logical stage is to study these cells at the protein level. This preliminary sperm cell proteomic study will undoubtedly be useful for future research in the field of plant sexual reproduction. This chapter provides valuable preliminary data and identifies problems and potential resolutions for sperm cell proteomic studies. Moreover, with a few adaptions the techniques developed in this preliminary proteomic project could be utilised in cross-species studies of gametes.

Table 4.1. Comparison of pollen proteomic studies.

Three independent pollen proteomic studies (Holmes-Davis et al., 2005; Noir et al., 2005; Sheoran et al., 2006) are compared in the table. The comparison includes the quantity of pollen used in each experiment, pollen disruption methodology, buffer types, number of fractions analysed, method of protein precipitation, techniques utilised for protein quantification, amount of protein loaded on the gel, and number of protein spots analysed.

Plant species	Amount of pollen (mg)	Extraction method	Buffer	Number of fractions	Protein precipitation method	Prot. Quant. Assay	Amout of prot. Loaded on gel (µg)	No. of prot./spots analysed	Reference
A. thaliana (Col)	35	GH	1	3 (SS, SI, SDS)	Acetone	BioRad DC kit	18	135/179	Holmes-Devis et al., 2005
A. thaliana (Col)	20	V^g	2	2 (S, P)	Unspecified	Bradford	100	121/145	Noir et al., 2005
A. thaliana (Ler)	20	GH	3	2 (S, P)	TCA/DTT	BioRad prot. Assay	600	110/150	Sheoran et al., 2006

GH glass homogenizer, V^g vortexing with glass beads, SS salt-soluble, SI salt-insoluble, SDS protein fraction extracted using SDS, S supernatant, P pellet
Buffer1; 71.4mM KCl, 14.3% v/v glycerol, 37.7mM Tris pH 7.4, 0.29mM pepstatin A, 0.4mM PMSF, 28.6% ethanol, and one Complete^{TM} tablet (Roche Applied Science, Indianapolis, IN)per 2mL of solution (Holmes-Devis et al., 2005), Buffer2; 1 protease inhibitor cocktail tablet (Roche); 10 mM DTT; 50 mM, pH 8.0, Tris-Base; 10 mM EDTA; and 0.5% Chaps (Noir et al., 2005), Buffer3; 10% TCA and 1% DTT in acetone (Sheoran et al., 2006).

Table 4.2. Comparison of protein functional categories from three pollen proteomic studies

Protein spots identified in the three studies detailed in table 4.1 were categorised according to functional classes. The percentage of proteins present in each category are listed (as described in each publication) and compared below.

Functional catagories	Holmes-Devis et al., 2005 (179 spots)	Noir et al., 2005 (145 spots)	Sheoran et al., 2006 (150 sopts)
Energy regulation	18%	23%	18%
Defense/stress response	6%	6%	20%
Cytoskeleton	13%	6%	8%
Transport	20%	3%	8%
Cell growth and division	2%	3%	4%
Metabolism	20%	14%	18%
Signal transduction	7%	N/A	5%
Protein synthesis	7%	2%	N/A
Cell wall	N/A	3%	N/A
Protein processing	N/A	6%	N/A
C-compound &carbohydrate metabolism	N/A	8%	N/A
Pollen allergen	N/A	N/A	6%
Unknown	7%	26%	8%

4.2 Results

4.2.1 SDS-PAGE analysis of a *B. oleracea* sperm cell-enriched sample

Protein samples derived from 10 mg of osmotically shocked *B. oleracea* pollen, homogenised pollen, Parr pressure bomb disrupted pollen, sperm-enriched pellet, supernatant (sperm cell-excluded fraction) and pollen debris were analyzed using SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). The protein gel was stained with Coomassie R-250 (Figure 4.1a) to enable ready quantification of the proteins not only in the same sample but also between the different samples under comparison. However, Coomassie R-250 staining failed to enable visualisation of proteins in the sperm-enriched fractions and thus these gels were restained with the more sensitive silver stain to permit identification of less abundant proteins. Differences between the sperm-enriched and vegetative cell-enriched samples were able to be distinguished following silver staining (Figure 4.1b). The banding pattern in the sperm-enriched fraction was found to be largely similar to the whole pollen samples, though some differences can clearly be seen. The intensity of bands in the sperm-enriched sample is less than for the other samples as sperm cells are small compared to pollen and most of the cell contents are composed of nuclear material with little cytoplasm. Interestingly, some bands in this sperm-enriched fraction were shown to be absent from or less intense than the supernatant fraction (arrow head) and these could represent proteins that are exclusively expressed in sperm. The bands indicated by arrows in the supernatant fraction are most likely vegetative cell-derived soluble proteins as they are found to be enriched in this sample from which almost all sperm cells were removed by centrifugation or sieving. Moreover, these bands were less intense in the sperm-enriched fraction. These data clearly indicate that the sperm enrichment technique and protein analysis methodology utilised were successful in revealing potentially sperm-specific proteins. However, in order to further the study improved resolution would be required and 2-DE (two-dimensional electrophoresis) was chosen for the rest of the proteome analysis.

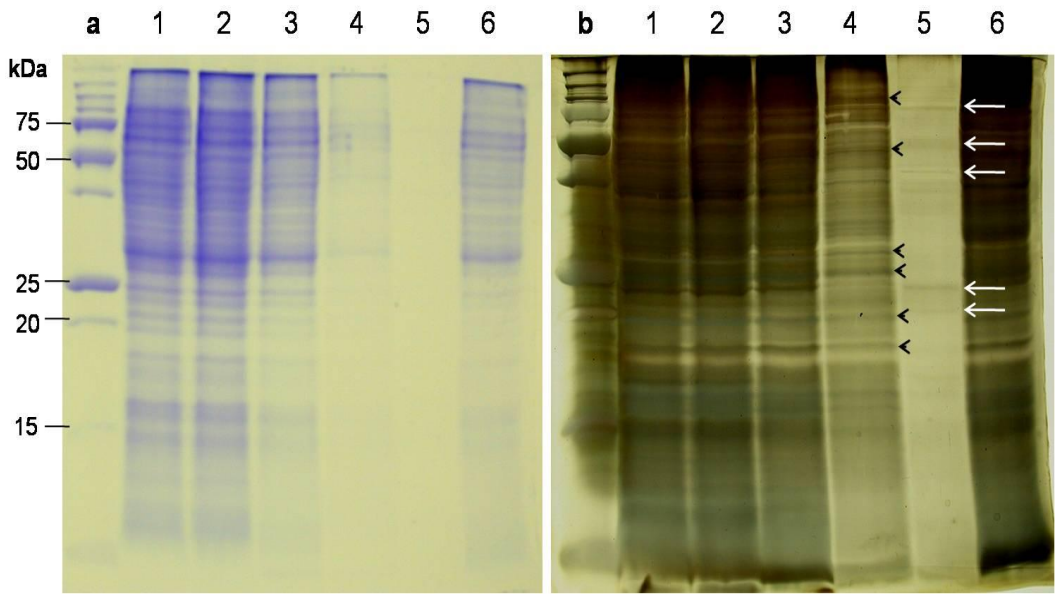


Figure 4.1. SDS-PAGE analysis of pollen and sperm cell-enriched samples from *Brassica oleraea*.

The figure represents the same SDS-PAGE gel stained initially with Coomassie R-250 (a) and then with silver stain for greater sensitivity (b). The samples represent protein extracted in SDS sample buffer from 10 mg of undisrupted *B. oleracea* pollen grains (a1, b1) and pollen disrupted by 2 different methods - homogenization (a2, b2) and pressurization/depressurisation (a3, b3), proteins from a sperm cell-enriched (a4, b4) and sperm cell-excluded fraction (a5, b5) and a pollen wall debris (a6, b6). Five distinct protein bands enriched or specific to the sperm cell-excluded sample are indicated by white arrows. Six protein bands (black arrow heads) are also indicated as potential sperm-specific proteins being present exclusively or at elevated levels in the sperm cell-enriched fraction.

4.2.2 2-DE of *A. thaliana* sperm cell-enriched samples and mass spectrometry of candidate sperm proteins

Protein extracted from the Arabidopsis sperm cell-enriched sample obtained (by *en masse* sperm isolation from 150 mg of pollen) was separated by 2-DE (Figure 4.2c, 4.3c) and compared to whole pollen (Figure 4.2a, 4.3a) and sperm-excluded samples (Figure 4.2b, 4.3b). 10 µg of protein for each sample was run on the gels stained with silver stain (Figure 4.2) and 87 µg was used for Coomassie stained gels (Figure 4.3). A total of 360 protein spots were detected for the sperm-enriched protein fraction following silver staining (Figure 4.2c). Out of these 360 spots, 12 spots were identified as having molecular weights and pIs distinct from proteins found in the sperm cell-excluded sample and were thus chosen for identification by mass spectrometry. The PMF (Peptide Mass Fingerprinting) results from MALDI-TOF MS were compared against the Swissprot database and recorded in table 4.4. In the Coomassie stained gel, a total of 90 spots were detected for the sperm-enriched fraction. Of these 90 spots, 31 were found to be also represented in the sperm-excluded sample. 27 spots which were not present in the sperm-excluded sample (Figure 4.3c red) and therefore could represent sperm proteins were chosen for MS analysis. In addition a further 20 spots present in both the sperm-enriched and sperm-excluded samples (Figure 4.3c green) were also selected for MS identification. The MS results were compared against the Swissprot database and documented in table 4.3 and the proteins were then grouped into functional categories (Figure 4.4). Out of all 36 proteins which matched proteins in the database, the majority (44%) were involved in energy processes. The other functional categories were protein processing, destination and storage (19%), metabolism (14%), defence and stress responses (11%), ubiquitination (6%), protein synthesis (3%), and chromosome remodelling (3%). However it is important to note that these proteins were obtained from a semi-purified sperm sample which could also contain contaminating proteins derived from the pollen vegetative cell.

Figure 4.2. 2-DE silver-stained gels of whole pollen, sperm cell-enriched and sperm cell-excluded samples.

Sperm cells of *A. thaliana* were released from pollen using a combination of enzymatic and pressurization/depressurisation techniques and semi-purified by filtration and centrifugation methods. Protein was extracted from the sample and analysed using 2-DE (c) in comparison with protein from whole pollen (a) and a sperm cell-excluded sample (a pollen sample from which sperm cells had been isolated/removed) (b). A total amount of 10 µg of protein was used in each sample. The gels were silver-stained to visualise protein spots. Spots present in the sperm cell-enriched sample (c) were compared with the protein spot profiles of the other two samples. Spots unique to or more intense in the sperm cell-enriched samples were selected for MS analysis. A total of 12 proteins were chosen to be identified in an initial trial and these appear as red numbered spots (c).

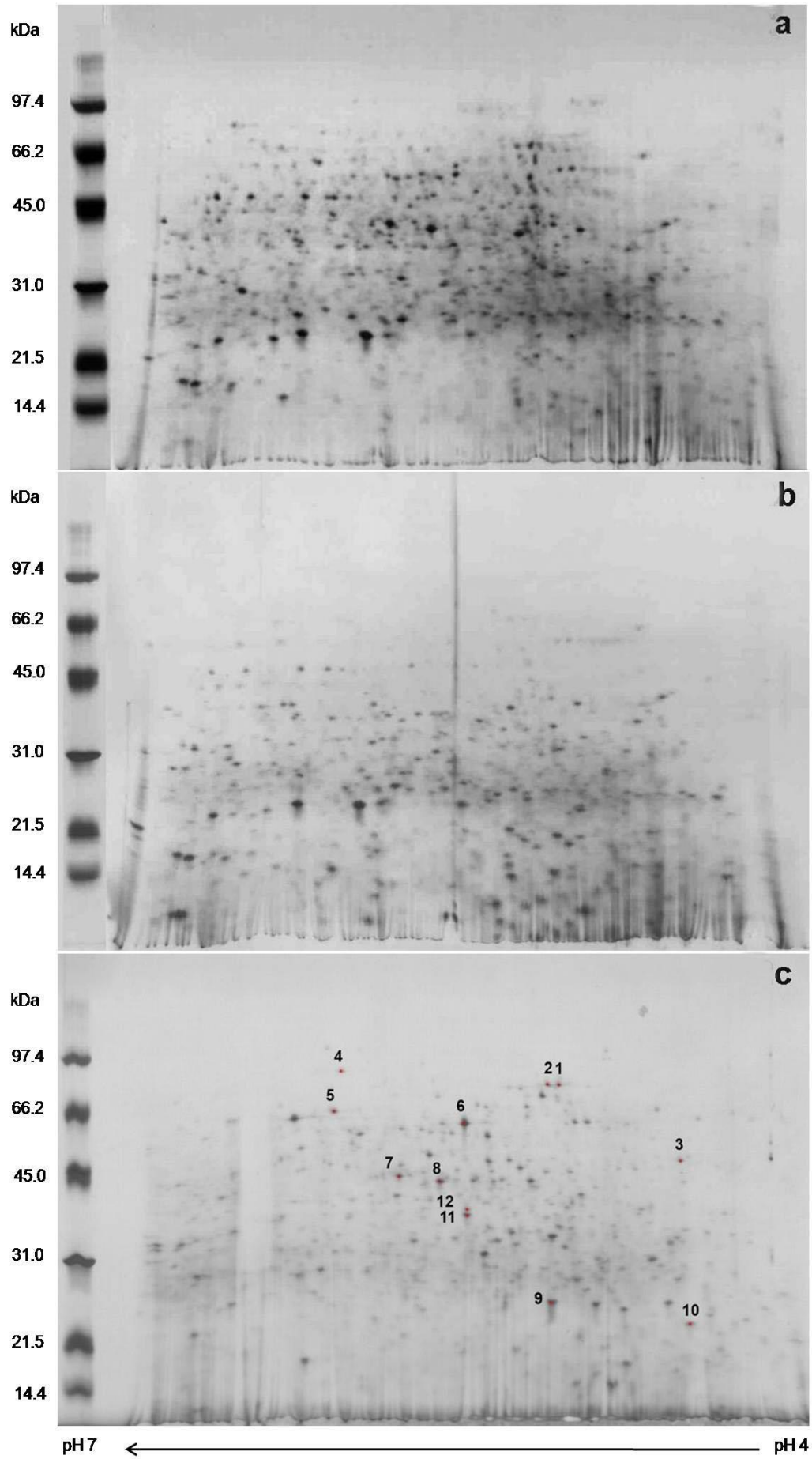

kDa
97.4
66.2
45.0
31.0
21.5
14.4
a
kDa
97.4
66.2
45.0
31.0
21.5
14.4
b
kDa
97.4
66.2
45.0
31.0
21.5
14.4
c
4
21
5
6
3
7
8
12
11
9
10
pH 7
pH 4

Figure 4.3. 2-DE Coomassie R-250 stained gels of whole pollen, sperm cell-enriched and sperm cell-excluded samples.

Sperm cells of *A. thaliana* were released from pollen using a combination of enzymatic and pressurization/depressurisation techniques and semi-purified by filtration and centrifugation methods. Protein was extracted from the sample and analysed using 2-DE (c) in comparison with protein from whole pollen (a) and a sperm cell-excluded sample (a pollen sample from which sperm cells had been isolated/removed) (b). A total amount of 87 µg of protein was used in each sample. The gels were stained with Coomassie R-250 to visualise protein spots. The spots present in the sperm cell-enriched sample (c) were compared with protein spot profiles of the other two samples. Spots unique to or more intense in the sperm cell-enriched samples were selected for mass-spec analysis (red numbered spots). However, some spots present on all 3 gels were also selected as reference proteins (green numbered spots). A total of 48 spots were chosen for identification by MS analysis (c).

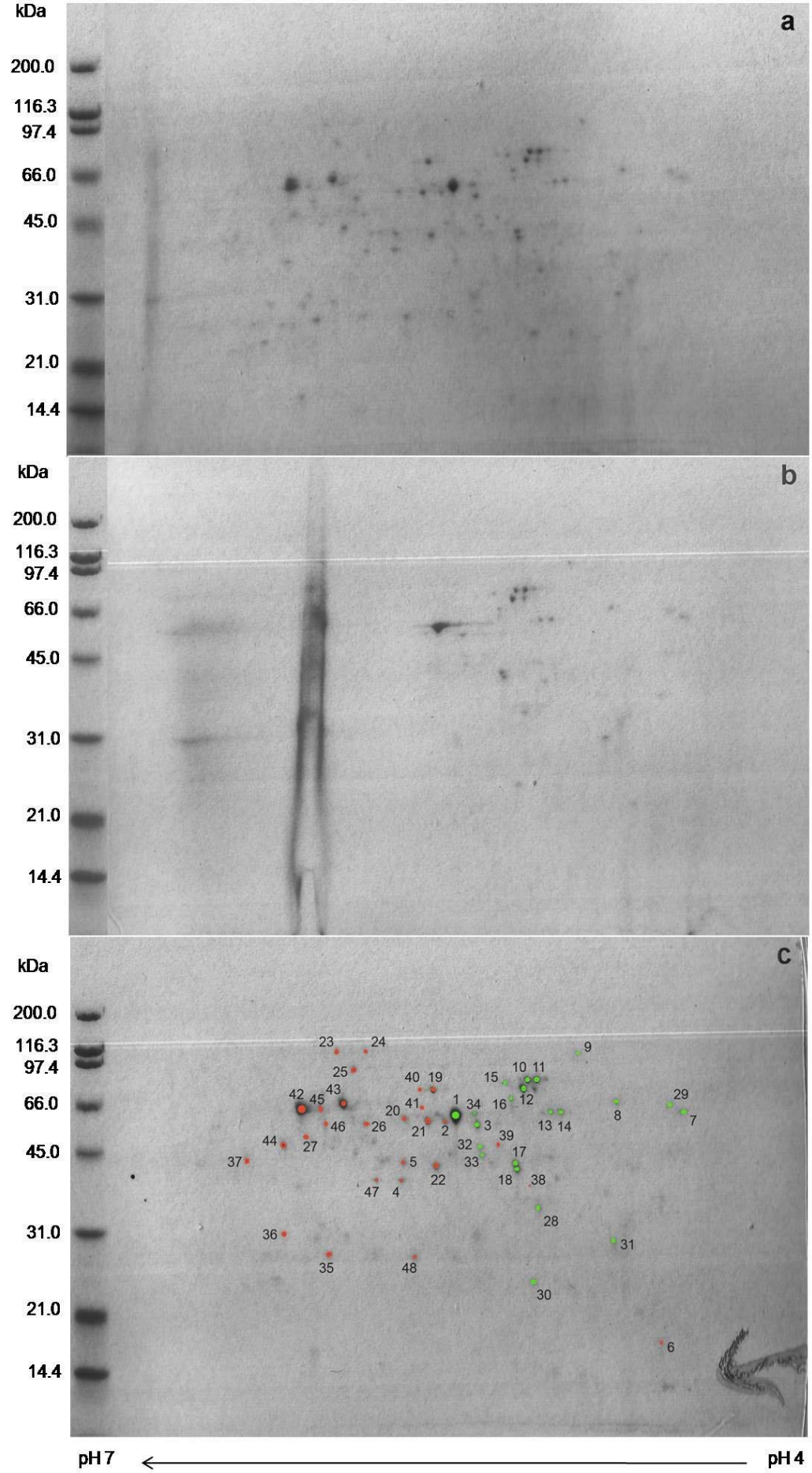

kDa
200.0
116.3
97.4
66.0
45.0
31.0
21.0
14.4
a
b
c
pH 7
pH 4

Table 4.3. *A. thaliana* proteins identified by MS from the sperm cell-enriched fraction.

Proteins identified by 2-DE and MS from the sperm cell-enriched fraction were categorised and listed according to biological function (TAIR gene ontology annotation). These proteins were compared to previously reported Arabidopsis pollen and mitochondria proteome studies for referencing and identification. Moreover, potential pollen specificity of these proteins was assessed utilising existing microarray data (Gene Atlas). Note, these proteins were not obtained from a pure sperm cell fraction - light contamination from the pollen vegetative cell is expected to be presented in the sperm cell-enriched sample used for 2DE and MS.

Spot no.	Spot colour	Gel stain	Locus Number	Description	MW (Da)	pI	Score*	Coverage**	Presented in pollen proteome	Pollen specific
Chromosome remodelling										
1	N/A	Sil	At2g33290	Histone-lysine N-methyltransferase, cytosine specific SUVH2	73657	6.2	18	14%	no	No
Defense/stress response										
9	N/A	Sil	At2g30870	Glutathione S-transferase ERD13	24215	5.5	26	15%	no	No
11	N/A	Sil	At1g75540	Putative salt tolerance-like protein	37181	6.4	14	9%	no	No
18	Green	Coo	At1g35720	Annexin D1	36296	5.2	144	49%	P2,3	No
35	Red	Coo	At4g25200	Heat shock 22 kDa protein, mitochondrial precursor	23595	6.5	15	8%	no	No
Energy										
6	N/A	Sil	At5g08670	ATP synthase subunit beta-1, mitochondrial precursor	59805	6.2	81	38%	M1,3,4	No
1	Green	Coo	At5g08670	ATP synthase subunit beta-1, mitochondrial precursor	59805	6.2	228	50%	M1,3,4	No
20	Red	Coo	At1g79440	Succinate semialdehyde dehydrogenase, mitochondrial precursor	56923	6.5	22	8%	no	No
23	Red	Coo	At2g05710	Aconitate (hydratase 2 mitochondrial precursor)	108133	7.1	11.9	14.2	P2, M2	No
24	Red	Coo	At4g26970	Aconitate (hydratase 3 mitochondrial precursor)	108427	7.1	12.5	12.2	P2	No
25	Red	Coo	At5g37510	NADH-ubiquinone oxidoreductase, mitochondrial precursor	82557	6.2	121	26%	P1, M4	No
26	Red	Coo	At5g08690	ATP synthase subunit beta-2, mitochondrial precursor	59850	6.2	141	40%	P2	No
31	Green	Coo	At1g79010	NADH dehydrogenase iron-sulfur protein 8, mitochondrial precursor	25943	5.3	50	25%	no	No
32	Green	Coo	At5g08670	ATP synthase subunit beta-1, mitochondrial precursor	59805	6.2	31	11%	P3	No
34	Green	Coo	At5g08690	ATP synthase subunit beta-2, mitochondrial precursor	59850	6.2	120	37%	no	No
42	Red	Coo	AtMg01190	ATP synthase subunit alpha, mitochondrial	55296	6.2	149	38%	M3	No

12	Green	Coo	At1g78900	Vacuolar ATP synthase catalytic subunit A	69111	5.1	220	50%	P2,3	No
13	Green	Coo	At1g76030	Vacuolar ATP synthase subunit B	54188	5	111	38%	no	No
14	Green	Coo	At1g76030	Vacuolar ATP synthase subunit B	54188	5	49	21%	no	No
28	Green	Coo	At1g76030	Vacuolar ATP synthase subunit B	54188	5	64	16%	no	No
33	Green	Coo	At1g51780	Vacuolar ATP synthase subunit C	42878	5.4	35	17%	no	No
Metabolism										
21	Red	Coo	At5g57655	Xylose isomerase	54028	5.6	86	23%	P1	No
27	Red	Coo	At2g44350	Citrate synthase 4, mitochondrial precursor	53091	6.4	58	14%	P1	No
46	Red	Coo	At1g74380	Putative glycosyltransferase 5	52094	6.8	10	5%	no	No
47	Red	Coo	At5g18200	Probable galactose-1-phosphate uridyl transferase	39380	6.2	20	8%	no	No
37	Red	Coo	At1g24180	Pyruvate dehydrogenase E1 component subunit alpha-2, mitochondrial precursor	43787	8	19	10%	M2	No
Protein processing, destination, storage										
2	Red	Coo	At3g16480	Probable mitochondrial-processing peptidase subunit alpha-2, mitochondrial precursor	54190	6	40	14%	no	No
10	Green	Coo	At5g42020	Luminal-binding protein 2 precursor	73801	5.1	65	21%	P1,3	No
11	Green	Coo	At5g42020	Luminal-binding protein 2 precursor	73801	5.1	239	48%	P1,3	No
15	Green	Coo	At5g28540	Luminal-binding (protein 1 precursor BiP1 AtBP1)	73584	5.1	11	24.8	P1,3	No
16	Green	Coo	At3g23990	Chaperonin CPN60, mitochondrial precursor	61242	5.2	22	6%	P2,3, M4	No
8	Green	Coo	At1g21750	Probable protein disulfide-isomerase 1 precursor	55852	4.8	211	47%	P1,2,3	No
43	Red	Coo	At3g02090	Probable mitochondrial-processing peptidase subunit beta, mitochondrial precursor	59180	6.3	148	39%	M4	No
Protein synthesis										
30	Green	Coo	At2g39820	Eukaryotic translation initiation factor 6	26893	4.6	14	9%	no	Yes
Ubiquitination										
5	Red	Coo	At3g16580	F-box/Kelch-repeat protein	44842	6.3	26	11%	no	Yes
6	Red	Coo	At1g75950	SKP1-like protein 1A	18017	4.5	11	7%	no	No

MW - Molecular weight; Da - Dalton; Coo - Coomassie R-250 stain; Sil - Silver stain; P - *Arabidopsis* pollen proteome studies; M - *Arabidopsis* mitochondria proteome studies;P1- Holmes-Davis et al., 2005; P2 - Noir et al., 2005; P3 - Sheoran et al., 2006; M1 - Kruft et al., 2001; M2 - Millar et al., 2001; M3 - Werhahn and Braun, 2002; M4 - Brugiere et al., 2004

** Sequence coverage (percentage of the complete protein sequence identified), *Mascot score (http://www.matrixscience.com/help/scoring_help.html), N/A; No data available

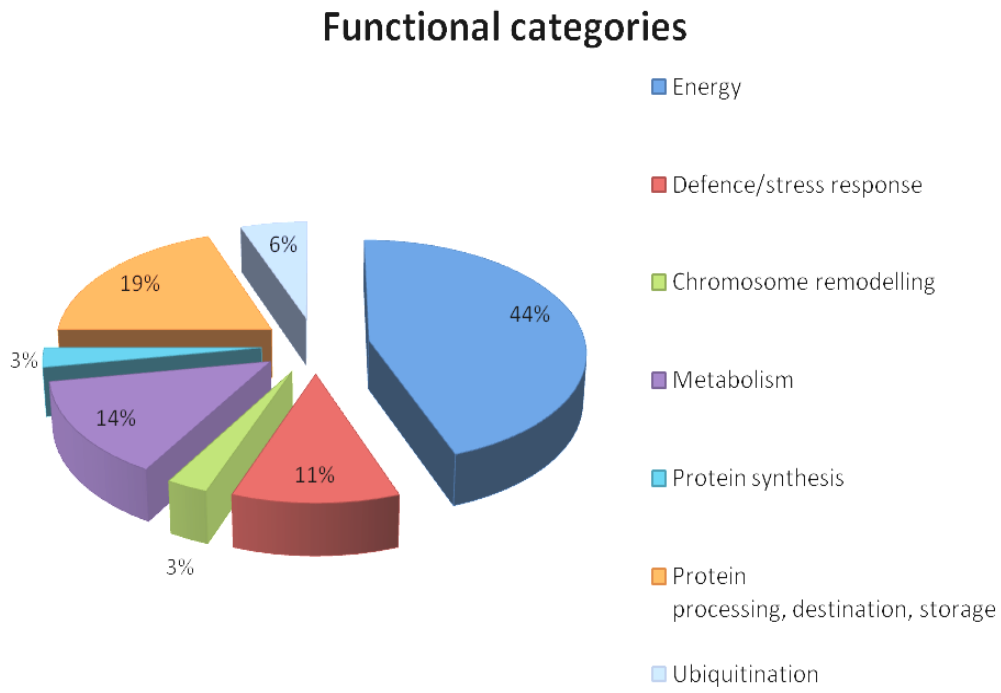


Figure 4.4. Functional categories of proteins found by 2-DE/MS to be present in an Arabidopsis sperm cell-enriched fraction (TAIR gene ontology annotation).

Protein extracted from sperm cell-enriched sample of *A. thaliana* was separated by 2-DE and stained with silver stain and Coomassie R-250. Some the protein spots revealed on the gel (107 spots) were selected to be identified by MS. The majority of the selected proteins were found to be involved in energy (44%) and protein processing/destination and storage (19%).

4.2.3 2-DE/MS analysis of FACS-purified *B. oleracea* and *A. thaliana* sperm cells

1.6×10^6 of FACS-sorted sperm cells from *B. oleracea* (obtained as described in Chapter 3) were subjected to TCA precipitation which yielded a total of 1.5 µg of protein. However, only 0.6 µg of sperm cell protein was recovered after a protein clean-up step and all of this was used for 2-DE. The resulting gel was silver-stained and a total of 21 spots were revealed (Figure 4.5). All of these spots were analysed by mass spectrometry (MS). PMF data obtained from the MS were identified utilising MASCOT tool (http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF). The search results are recorded in table 4.4. Approximately 2×10^6 *A. thaliana* FACS-purified sperm cells were precipitated using TCA and dissolved in 30µl CHAPS buffer. Unfortunately no protein was detected when quantified using EZQ protein quantification kit (Invitrogen) and further analysis was not possible due to a lack of material.

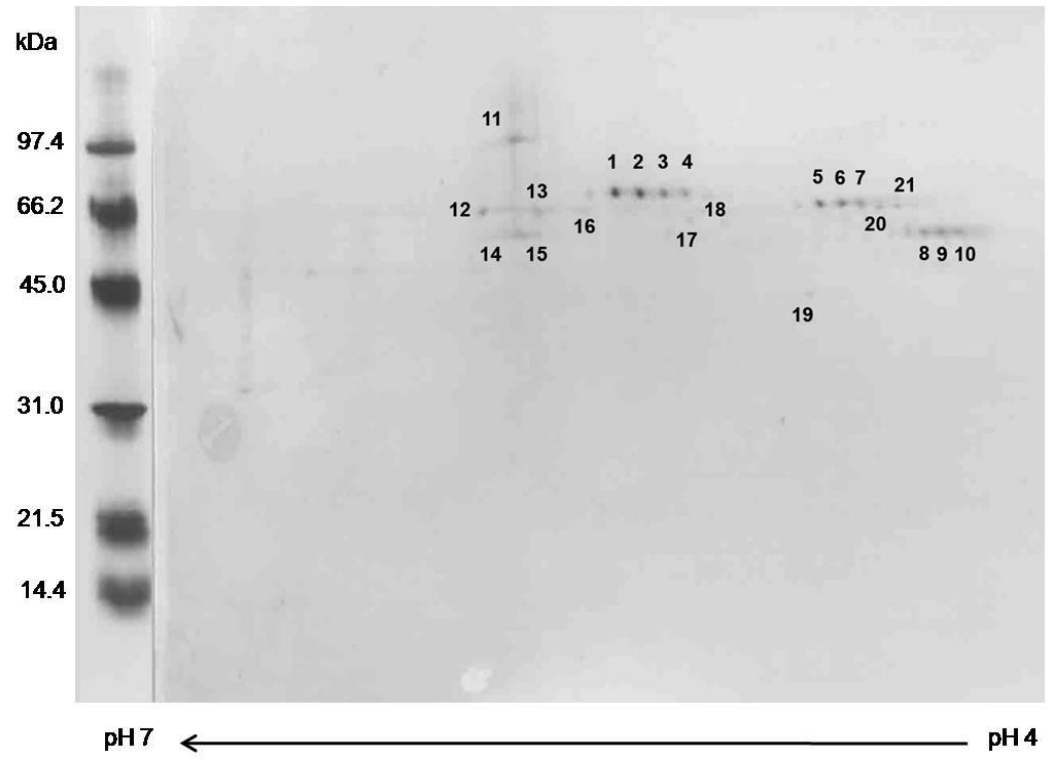


Figure 4.5. 2-DE of *Brassica oleracea* FACS-purified sperm cells.

Sperm cells of *B. oleracea* were released from pollen grains using osmotic shock coupled to pressurisation/depressurisation in a Parr pressure bomb. The cells were then semi-purified by filtration and centrifugation. Sperm cells were further purified by FACS and total protein was then extracted by TCA precipitation. A total of 0.6 µg of protein was loaded onto the gel. Protein spots were visualised by silver staining. A total of 21 spots were resolved and subjected to MS analysis for identification. Spot numbers correspond to the data presented in table 4.4.

Table 4.4. Results of protein identification in FACS purified *B. oleracea* sperm cells.

1.5 µg of protein was prepared from 1.6×10^6 FACS-sorted *B. oleracea* sperm cells by TCA precipitation. Following clean-up of the sample 0.6 µg of the purified protein sample was used for 2DE analysis. All 21 protein spots visualised on the gel were excised and subjected to MS analysis. Of the 21 protein spots, 7 were found to have some degree of match to Arabidopsis proteins when following a MASCOT search. These 7 proteins were categorised into 5 categories listed as follow (TAIR gene ontology annotation).

Spot no.	Gel stain	Locus Number	Name	Description	MW (Da)	pI	Score*	Coverage**	Presented in pollen proteome	Pollen specific
Transport										
6	B-Sil	At2g17980	SLY1_ARATH	SEC1 family transport protein SLY1	69393	6.0	9	3%	No	No
Cell growth and division										
11	B-Sil	At2g44950	BRE1A_ARATH	E3 ubiquitin-protein ligase BRE1-like 1	100740	6.5	16	3%	No	No
Metabolism										
9	B-Sil	At5g17330	DCE1_ARATH	Glutamate decarboxylase 1	57429	5.4	11	4%	No	No
18	B-Sil	At1g34120	IP5P1_ARATH	Type I inositol-1,4,5-trisphosphate 5-phosphatase 1	68368	5.7	10	2%	No	No
15	B-Sil	At2g23000	SCP10_ARATH	Serine carboxypeptidase-like 10 precursor	50323	6.4	14	5%	No	No
Signal transduction										
14	B-Sil	At2g01450	MPK17_ARATH	Mitogen-activated protein kinase 17	56028	6.5	16	4%	No	Yes
Unknown										
7	B-Sil	At1g34120	IP5P1_ARATH	Type I inositol-1,4,5-trisphosphate 5-phosphatase 1	68368	5.7	15	2%	No	No

MW; Molecular weight, Da; Dalton, B; *Brassica oleracea*, Sil; Silver stain
** Sequence coverage (percentage of the complete protein sequence identified), *Mascot score (http://www.matrixscience.com/help/scoring_help.html)

4.3 Discussion

4.3.1 SDS-PAGE of *B. oleracea* sperm cell-enriched samples

SDS-PAGE analysis of the protein profiles derived from whole pollen, sperm cell-enriched, pollen cytoplasm-enriched and pollen debris fractions was used to gauge the quality of the sperm purification procedure. Comparison of protein banding profiles highlighted differences between the sperm cell-enriched and pollen cytoplasm-enriched fractions suggesting that sperm cells have a different protein complement to vegetative cells (Figure 4.1). It is clear from the protein gel data that the sperm-enriched fraction was likely to be contaminated with components of the vegetative cell cytoplasm (compare protein bands in the sperm-enriched fraction to the first three lanes that represent the whole pollen extracts – Figure 4.1). The banding pattern in the sperm-enriched fraction is highly similar to the first three samples except that the overall concentration of the proteins is substantially reduced and, importantly, that the relative concentrations of various proteins are different, suggesting that these enriched proteins are possibly specific to sperm cells. Despite the fact that the SDS-PAGE analysis did not provide clear information about the protein complement of sperm it did demonstrate that the sperm cell fraction needed to be more highly purified for their proteins to be characterized. The SDS-PAGE was clearly not sensitive enough to distinguish differences between the sperm enriched fraction and the whole pollen samples. Thus 2D electrophoresis was selected to be utilised for the next part of the study as it allows better separation and resolution of proteins on the gel. Sperm cell-specific proteins which would be present in low concentration in the whole pollen samples would potentially become distinguishable. However many similarities between fractions would also be visualised as many housekeeping proteins will present in every sample. Therefore, comparison of each fraction following 2DE has to be carried out with great care due to the inherent complexity of the profiles obtained.

4.3.2 2-DE and mass spectrometry of *A. thaliana* sperm cell-enriched protein samples

Over 360 protein spots were observed on the 2-DE silver-stained gel of the sperm cell-enriched sample. The protein spots discovered on the sperm enriched sample were subset of protein spots on the whole pollen sample. The spot patterns of all three gels were similar however suggesting substantial overlap between the protein content of the samples. In the sperm cell-enriched fraction, despite the pattern similarity to the whole

pollen sample, the relative spot intensities were different. Spots which silver-stained with a higher intensity in the sperm-enriched sample suggest enrichment of that protein and could therefore be protein present in sperm cells. Potential sperm cell protein spots were also identified by comparing the spot pattern between the sperm cell-enriched and sperm cell-excluded fractions. Protein spots of higher intensity or only present in the sperm cell-enriched fraction were identified as sperm protein candidates. These spots were chosen to be identified using MS and initially 12 spots with different masses and pIs were chosen to be examined. Unfortunately, in an initial analysis only 1 spot was identified as having a good match (Mascot score > 60) due the levels of protein in the other spots being inadequate for MS analysis. Despite the fact that silver stain is sensitive and able to detect tiny amounts of protein it typically does not stain proteins in a quantitative manner (Candiano et al., 2004; Wheeler et al., 2000). Abundant proteins may stain to a similar degree as less abundant proteins thus complicating analyses such as MS and quantitative comparisons. Silver staining was chosen to be used in the first comparative study as a broad knowledge of the protein content of each fraction was required for effective technique development. Therefore, Coomassie stain, a less sensitive but more informative technique (with regards protein quantity) (Candiano et al., 2004; Wheeler et al., 2000) was utilised in the second analysis and for MS. 87µg of protein from each sample was run on the gels and stained with Coomassie R-250 which resulted in 90 spots being detected in the sperm cell-enriched sample. Of these 90 spots, 27 spots were found to be either unique to this fraction or highly enriched compare to the other samples and were chosen to be identified (represented in red in Figure 4.3c). Moreover, 20 spots present in all three gels (represented in green in Figure 4.3c) were also chosen to be identified as reference spots. Interestingly, almost 50% of the protein spots identified by MS in Coomassie stained gels (15 of 32 spots) were mitochondrial precursor proteins, mostly classified into the 'energy' category. However, only 7 of these mitochondrial proteins were previously identified in Arabidopsis mitochondrial proteomic studies (Brugiere et al., 2004; Kruft et al., 2001; Millar et al., 2001; Werhahn and Braun, 2002). These proteins could well be obtained from two main sources, pollen vegetative cell-derived mitochondria and sperm cell-borne mitochondria. Both pollen and sperm cells contain vast numbers of mitochondria as demonstrated by the TEM study detailed in chapter 3 and illustrated by McConchie and colleagues (McConchie et al., 1987b). It is quite possible that mitochondria were also enriched in the sperm cell preparation, being derived from the lysed vegetative cell. However, the size of the mitochondria observed in TEM micrographs from both vegetative and sperm cells were relatively small in comparison to the size of the sperm cells (~2.5 µm) and thus mitochondria derived from the lysed vegetative cell were expected to be excluded in sperm semi-purification step. Moreover, as shown by TEM analysis plant sperm cells

contain numerous mitochondria considering their small size (McConchie et al., 1987b; Russell, 1984; Wagner et al., 1989). Interestingly in tobacco, mitochondria were also discovered in sperm nuclei and were suggested to be transmitted to the embryo and endosperm during double fertilisation (Yu and Russell, 1994a), as mentioned in chapter 3. However, mitochondria were not detected in *Arabidopsis* sperm nuclei as demonstrated by the TEM study detailed in the previous chapter. Therefore, the mitochondrial proteins identified in this MS study were considered to be principally derived from mitochondria residing in the sperm cell cytoplasm. As can be seen from table 4.2 some of the proteins identified in this study were also reported in pollen proteome studies (Holmes-Davis et al., 2005; Noir et al., 2005; Sheoran et al., 2006). These proteins could still be derived from sperm cells as it would be anticipated that many proteins involved in normal cellular functions would be shared between cell types. In addition, proteins identified in our study that did not appear in the previous pollen proteome studies but were reported to be preferentially expressed in pollen according to available microarray data (Gene Atlas, table 4.2 in pollen specific column) i.e. At2g39820 (Eukaryotic translation initiation factor 6) and At3g16580 (F-box/Kelch-repeat protein) could potentially be sperm cell-specific proteins - due to the very small volume of sperm cells compared to the vegetative cell sperm proteins will be underrepresented in the whole pollen protein sample. Further, the expression levels reported in transcriptome studies for these candidates was relatively low as would be expected for sperm cell-specific gene expression. However, whether or not these proteins are specific to sperm cells remains inconclusive. In order to reveal the identity of sperm cell proteins particularly the sperm cell-specific class, more studies are required to be undertaken, in particular protein analysis from highly purified sperm cells. Nonetheless, the similarity of results obtained between this experiment and the reported pollen proteome studies is an encouraging indicator that the techniques developed for 2DE analysis of sperm cells are appropriate. However, the data also indicated that, in all likelihood, significant vegetative cell contamination was also present in the sperm-enriched sample. It is clear from these analyses that in order to perform accurate sperm cell protein identification the purification techniques need to be developed further to isolate the cells free from pollen vegetative cell contamination.

4.3.3 2-DE and mass spectrometry of *B. oleracea* and *A. thaliana* FACS-purified sperm cells

A Fluorescence-Activated Cell Sorting (FACS) technique was developed to be used for Hoechst-stained sperm cells in order to purify superior quality intact sperm cells free

from vegetative cell and pollen wall protein contamination. This FACS technique has been carried out successfully for both *B. oleracea* and *A. thaliana* in this study as described in chapter 3. Unfortunately protein isolation was only successfully carried out from *B. oleracea* FACS-purified sperm. *A. thaliana* sperm cells are considerably smaller than those from *B. oleracea* species (15.72 μm^3) (McConchie et al., 1987b) and thus total protein yield from these cells is lower. One of the major obstacles in the protein isolation process was that the sorted cells were collected in relatively large volumes of buffer. This made it very difficult to concentrate the isolated cells by centrifugation without damaging them prior to protein extraction and thus protein losses were experienced at this step. For Arabidopsis, a sperm cell pellet was not visible following the centrifugation step and consequently little, if any, protein was obtained for further analysis. However, by analysing the protein obtained from FACS purified *B. oleracea* sperm cells it was hoped the resulting information would facilitate study and understanding of *A. thaliana* sperm. 21 *B. oleracea* protein spots were chosen for MS analysis, most spots contained inadequate quantities of protein to give reasonably acceptable matches following a MASCOT search however 7 proteins were tentatively identified in the search and reported in table 4.3. Interestingly, one protein, At2g01450 (Mitogen-activated protein kinase 17) was found to be pollen specific according to pollen microarray data (Gene Atlas). This mitogen-activated protein kinase was functionally categorised as being involved in signal transduction and thought to have a role in extracellular signalling pathways (TAIR). However, yet again whether or not this protein is actually a correct match is still inconclusive as the match score was quite low. The low number of 'hits' and low matching scores obtained in this particular study could well be an effect of the fact that this is a cross-species analysis. Moreover, the other proteins identified in this experiment were expressed abundantly in other tissue types and thus may or may not play an important role in the sperm cells. Nonetheless, these housekeeping proteins were expected to be identified as they are present in almost all cell types. It is clear that, more studies need to be carried out to confirm the data obtained here and to obtain more accurate information about the protein complement of sperm cells. This preliminary experiment utilising FACS-sorted sperm cells, though useful, still requires more material to obtain a sufficient quantity of protein for a thorough MS analysis of proteins present in this specialised cell type. Despite the limitations of the study reported here it is encouraging that proteins extracted from both sperm cell-enriched fractions and FACS-sorted sperm cells could be visualised by 2-DE and identified by MS. Moreover, some proteins identified as possibly being sperm cell-specific were previously reported from a pollen proteomic study. This initial proteomic study has certainly provided valuable information for both technique development and building an initial picture of Arabidopsis sperm cell proteins. In the future, protein

extraction, precipitation, quantification and storage of *A. thaliana* FACS-sorted sperm cells will be improved thus providing higher protein yields for both 2-DE MS and whole sample mass spectrometry. Elucidating the complete set of proteins expressed in plant sperm cells is likely to be achieved in the not-too-distant future.

Chapter 5:

In Silico Cross-Species Identification of Putative Arabidopsis Sperm-Specific Transcripts

5.1 Introduction

Many techniques such as microarrays (Hennig et al., 2003; Honys and Twell, 2003), serial analysis of gene expression (SAGE) (Lee and Lee, 2003; Velculescu et al., 1995), differential display gel-based technologies (Liang and Pardee, 1992), massively parallel signature sequencing of attached amplified cDNA to microbeads (MPSS) (Brenner et al., 2000), and expressed sequence tags (EST) (Engel et al., 2003) have been used for transcription profiling analysis in various organisms and in specific tissue types (Hennig et al., 2003). Amongst these techniques, microarray is one of the most effective and commonly used methods to identify a range of transcripts in the target tissue. An overview of microarray and EST techniques utilised for pollen and sperm cell transcriptional profiling (Engel et al., 2003; Hennig et al., 2003; Honys and Twell, 2003) will be reviewed in this chapter in particular as data from these studies were used to identify Arabidopsis sperm cell transcripts in this project.

The massive data sets that have been generated by transcriptional profiling experiments in recent years have required a good management system. Therefore computational methods emerged for the purpose of organising these rapidly growing databases. Consequently, a new area of research which combines computational methods and biological data has emerged and has been termed 'Bioinformatics'. Internet-based bioinformatics tools and databases are increasing in number ranging from those for conventional sequence search and analysis to the relatively recent tools for microarray data (Kaminski, 2000). Various databases such as

GenBank (http://www.ncbi.nlm.nih.gov/ Genbank/index.html), EMBL (http://www.ebi.ac.uk/), and dbEST (http://www. ncbi.nlm.nih.gov/dbEST/) for nucleotide sequences and SWISS PROT (http://www.expasy.ch/sprot/), ExPASy (http://www.expasy.ch/), and PROSITE (http://www. expasy.ch/prosite/) for protein sequences are available for public access (Kaminski, 2000). Tools for sequence analysis are based on these DNA and protein sequence databases. These tools range from sequence alignment and motif searching to structure prediction (Kaminski, 2000). A Basic Local Alignment Search Tool (BLAST), which was heavily used in this project, is readily available on the web via the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST) (Kaminski, 2000). The sequence comparison search is reported as a probability value E – an estimate of the probability that the match happened by chance (the smaller e-values, the better the match) (Kaminski, 2000). The best match is reported at the top of the output generated and highlighted in red in the graphical view. The underlying assumption is that the matched sequences are similar because they are homologous; that is, they are derived from a common ancestral sequence or they share common functional aspects (Kaminski, 2000). Another very powerful bioinformatics tool utilised in this project was the Artemis Comparison Tool (ACT). This software allows interactive visualisation between a complete genome and sequences of interest. It can be used to perform either protein or nucleotide sequence alignments (Carver et al., 2005) and can be used by various operating systems i.e. UNIX, Windows, Macintosh, and recently can be performed online (Abbott et al., 2005). For more information on ACT features and applications please see review papers (Abbott et al., 2005; Carver et al., 2005).

5.1.1 Large scale cDNA library construction and expressed sequence tags (ESTs)

Large scale cDNA libraries have been constructed and expressed sequence tags (ESTs) from various plant species have been identified and deposited in the GeneBank database (Fernandez et al., 2004). More than 2,000,000 entries from over 100 plant species have been consigned to this database (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_ summary.html) (Fernandez et al., 2004). These single-pass cDNA sequences or ESTs are normally 300-400 bp long with an approximate error or base ambiguity rate of 3% (Boguski et al., 1993). However, despite these limitations this method was rapidly adopted by many researchers since it was started 1991 as such data is relatively easily obtained and is cost effective (Boguski, 1995). These cDNA dataset projects allow the characterisation of presumably a near full set of expressed genes present in a particular tissue type at a precise developmental stage in a specific environment (Fernandez et al., 2004). However, organ specific transcripts, which are often expressed at low levels in eukaryotic cells (Kuznetsov et al., 2002), are essential for understanding biological

process during plant development (Le et al., 2005). Unfortunately this method of gene identification has some drawbacks as these rarer transcripts are either absent or represented very infrequently in the libraries produced (Fernandez et al., 2004). Extensive subtraction and normalisation approaches such as suppression subtractive hybridization (SSH) (Diatchenko et al., 1996) and serial analysis of gene expression (SAGE) (Velculescu et al., 1995) have been utilised to obtain such rare transcripts (Sun et al., 2004). Data obtained from these studies can also be applied to large scale gene expression analysis i.e. microarray. Valuable information obtained from these transcript profiling techniques will certainly improve our understanding of gene function and important biological processes in many organisms including plant sperm cells.

5.1.2 Microarray technology

Microarray analyses have evolved rapidly and have been applied to many aspects of developmental biology. It has typically been used for gene screening, target gene identification and to measure the relative abundances of mRNAs (Stoeckert et al., 2002; van Ingen, 2002). However, it has also been employed in disease characterisation (Clarke et al., 2001; DeRisi et al., 1996), developmental biology (White et al., 1999), pathway mapping (Knight and Knight, 2001), and plant reproduction (Hennig et al., 2003). The technology is based on the binding affinity between complementary molecules of DNA or RNA (Herzel et al., 2001). The DNA molecules from a library (the target) are fixed on a solid support (glass slide or filter) and hybridised with a radioactive or fluorescently labelled mRNA probe prepared from samples of interest (Herzel et al., 2001) such as differing tissues, developmental stages, disease states, treated samples or environmentally stressed samples (Quackenbush, 2002). The hybridisation arrays are scanned and analysed to measure the fluorescence intensity of each spot by image quantification software packages (Quackenbush, 2002). Interpretation and analysis of array data are crucial steps for achieving meaningful information and computational methods are an essential component of the data mining task (Slonim, 2002). As transcriptional profiling data has amassed so have the methods used to analyse these data (Slonim, 2002). Therefore, one must choose methods wisely to suit a particular study. Moreover, array data needs to be stored in a form that is accessible to other researchers for future analyses and for the development of gene functional studies. The Web-browser interface and microarray gene expression database 'GENEVESTIGATOR' was devised for Affymetrix microarray data mining. Using this Web-based tool, the user can query the database to retrieve expression patterns for genes of interest comparing selected organs, tissues, developmental stages or

environmental conditions (Zimmermann et al., 2004) (https://www.genevestigator.ethz.ch.). GENEVESTIGATOR can present this comparative data both for individual genes or for families of genes (Zimmermann et al., 2004) in a simple graphical view that allows the user to easily understand and employ these data for further experiments. One of the highly useful tools available with GENEVESTIGATOR is Gene Atlas. This tool provides the average signal intensity values for a gene of interest in all plant organs (Zimmermann et al., 2004). Gene Atlas is commonly utilised to confirm the spatial expression pattern of a gene, for example the sperm-specific gene *HAP2* (von Besser et al., 2006) was demonstrated to be specifically expressed in the pollen (data for isolated sperm cells has not yet been included in the microarray database). Clearly GENEVESTIGATOR is a powerful resource that can facilitate the identification of 'target' genes and provide insights into their possible function. For more information on the GENEVESTIGATOR see reviews (Zimmermann et al., 2005; Zimmermann et al., 2004).

5.1.3 Transcriptional profiling of the male gametophyte

The transcript profiling techniques mentioned in section 5.1.2 have been applied by researchers working on plant reproduction. Various studies have concentrated on elucidating male gametophytic gene expression i.e. pollen and sperm cells. Approximately 20,000 to 24,000 different mRNA sequences were estimated for maize pollen grains (Honys and Twell, 2003; Willing et al., 1988). Hybridisation studies were employed to determine sporophyte-gametophyte transcript overlap and to permit the identification of pollen-specific transcripts (Honys and Twell, 2003). Before microarray analysis was introduced into the study of the pollen transcriptome very few pollen-specific genes were identified (da Costa-Nunes and Grossniklaus, 2003). Of these pollen-specific genes, only 23 were reported in Arabidopsis (da Costa-Nunes and Grossniklaus, 2003).

For Arabidopsis two commonly used microarrays are the Affymetrix Arabidopsis Genome (AG) array containing more than 8,000 probe sets and ATH1 array containing more than 22,000 probe sets (Hennig et al., 2003). Microarray analysis of transcripts from hydrated Arabidopsis pollen (ecotype Landsberg) using AG or ATH1 arrays revealed 992 pollen-expressed mRNAs with 61% of these overlapping with sporophytic gene expression (Honys and Twell, 2003). Of the remaining 39% representing pollen-specific transcripts, it was suggested that the contribution of sperm cell transcripts to this pool was likely to be negligible as these cells make up such a small fraction of the pollen grain (Honys and Twell, 2003; Xu et al., 2002). On the other hand, an

independent study on the pollen transcriptome of Arabidopsis ecotype Col-0 by Becker and colleagues showed a 90% overlap (out of 1,584 pollen transcripts) with genes expressed in vegetative tissues (siliques, seedlings, leaves, and roots) – the same Affymetrix Arabidopsis AG array was used in both experiments (Becker et al., 2003). The estimate that 10% of pollen transcripts are pollen-specific is supported by hybridisation studies for pollen cDNA libraries (Honys and Twell, 2003; Mascarenhas, 1990; Stinson et al., 1987). However, Honys and Twell commented that these estimates from the hybridisation studies did not take account of cross hybridisation between closely related gene family members and thus may well be an underestimate (Honys and Twell, 2003). The pollen grains in both experiments had undergone similar procedures and were purified by FACS. The variation between these two pollen chip experiments could be explained by i) different plant growth conditions and the ecotype used in the experiments, ii) different experimental procedures, iii) different methods used to interpret the data obtained from the chip, and iv) incorrectly annotated probe set (6.3%) of the AG array (Honys and Twell, 2004b). More recently transcripts of different developmental stages of Arabidopsis pollen were identified using the Affymetrix ATH1 (22,591 genes) arrays. Of 13,977 male gametophytic mRNAs identified (62% of total array), approximately 10% (1,355 genes) were male gametophyte specific (Honys and Twell, 2004b). Of the 13,977 male gametophytic genes, 83% (11,565 genes) were expressed in the uninucleate microspore stage, 85% (11,909 genes) in bicellular pollen, 63% (8,788 genes) in tricellular pollen, and 52% (7,235 genes) in mature pollen (Honys and Twell, 2004b). This more accurate ATH1 array experiment led to the estimate of 9,000 mature pollen-expressed genes in place of 3,500 identified in previous studies (Honys and Twell, 2004b). However, with more stringent criteria the number of pollen-specific genes was reduced from 1,355 to 800 genes (approx. 6%) in this study (Honys and Twell, 2004b).

As mentioned previously sperm cell-specific transcripts are expected to be underrepresented amongst pollen grain cDNAs (da Costa-Nunes and Grossniklaus, 2003) and therefore might not be detected in pollen microarray experiments or cDNA libraries. However in 2003, Engel *et al.* constructed a maize sperm cell library in order to identify transcripts present in the male gamete of higher plants (Engel et al., 2003). This study is considered to be of the most successful sperm cell studies as the cells used in this experiment were FACS-purified and proven to be free from vegetative cell-derived contamination. Plant sperm cell cDNA library construction has been attempted in various species prior to maize i.e. *Plumbago* (Singh et al., 2002), tobacco (Xu et al., 2002), and rice (Gou, 2000) The *Nicotiana tabaccum* sperm cell cDNA library was constructed from Percoll-purified sperm cells. Cold plaque screening was utilised to

select putative male germ cell cDNAs (Xu et al., 2002). Of 396 clones that hybridised to mature pollen cDNA, but not with cDNA probes from other tissues, 2 clones representing potential sperm cell-expressed transcripts were analysed (Xu et al., 2002). However, these two transcripts appeared to be expressed in the pollen as well as in the sperm cells (Xu et al., 2002). *Lilium longiflorum* generative cell and *Plumbago zeylanica* sperm cell cDNA libraries have also been constructed in order to identify genes that are exclusively expressed in these cells (Singh et al., 2002; Xu et al., 1999a; Xu et al., 1999b). Nearly 1,600 transcripts were identified from semi-purified generative cells of lily however only very few male gamete-specific genes were identified from this cDNA library i.e. generative cell specific gene *LGC1* (Xu et al., 1999b), *gcH2A*, and *gcH3* (Xu et al., 1999a). Despite the existence of the cDNA libraries described above, only very few plant sperm cell-specific transcripts have been identified. However, these sperm cDNA libraries and published EST data (particularly the FACS-purified maize sperm ESTs) were considered to be highly valuable resources for use in this project to facilitate identification of Arabidopsis sperm cell-expressed genes. Here a bioinformatics approach was employed for a cross-species comparative gene expression analysis utilising maize, Plumbago, rice and tobacco sperm ESTs. Pollen transcriptional profiling data was also used to select potential sperm cell-expressed gene candidates from the large number of transcripts under study. Details of these analyses will be described and discussed in this chapter.

5.2 Results

Bioinformatics approach taken to identify sperm cell-expressed gene candidates from available plant sperm ESTs data in this study is complex and for convenience is summarised by the flow chart depicted in figure 5.3 towards the end of the results section.

5.2.1 Comparative study between maize, Plumbago and rice sperm ESTs and the *Arabidopsis thaliana* genome

5,065 maize sperm ESTs (Engel et al., 2003) were searched using BLASTX against the *A. thaliana* Genbank protein database (as of Nov 2004). Sequences matching with e-values smaller than 0.001 were retained. A total of 46,210 matches was obtained. Only the best e-value for each paralogue was retained. The 2,598 remaining sequence matches were made non-redundant i.e. duplicates were removed reducing to list to 1,106 sequences. These 1,106 sequences were searched again against 97,648

Arabidopsis root and leaf ESTs to further reduce the number of genes analysed with the aim of concentrating only on potential sperm cell-expressed candidates. Out of 1,106 sequences 46 (4.16%) were found to have no match against the root and leaf ESTs (supplement CD), these were then checked against microarray data (Gene Atlas). Of the 46 sequences, 7 sequences (15.22%) (Table 5.1) were found to be specifically expressed in pollen and were thus considered as potential sperm protein candidates. 227 sequences (20.52%) that matched known root and leaf expressed genes with scores better than e^{-100} were culled out as these most likely to be general housekeeping genes which were also expressed in roots and leaves. The remaining 833 (75.32%) sequences that matched to Arabidopsis root and leaf sequences with e-values greater than 10^{-100} were subgrouped into an e^{-1}-e^{-69} group (511 sequences) for immediate study and e^{-70}-e^{-99} group (322 sequences) for future analysis, as these were considered less likely to be sperm cell-expressed genes but were retained nevertheless. Despite the fact that important sperm proteins may actually be expressed in other tissues due to scale of this analysis strict criteria had to be established such that the number of genes under study was manageable. The e^{-1}-e^{-69} group of sequences were manually searched against microarray data to search for pollen-specific transcripts utilising Gene Atlas as an analysis tool (https://iii.genevestigator.ethz.ch/at/index.php?page=geneatlas). Of 511 sequences, 85 were found to be preferentially expressed in pollen or stamens and were therefore considered as potential sperm protein candidates (microarray data from purified sperm cells is not yet available). These 85 Arabidopsis sperm protein in addition to the 7 candidates derived from maize sperm EST dataset are summarised in table 5.1.

Table 5.1. List of 92 Arabidopsis sperm protein candidates identified following a BLASTX search of maize sperm ESTs against the Arabidopsis genome.

List of 92 Arabidopsis sperm protein candidates selected from Arabidopsis genes that matched maize sperm ESTs (1,106 sequences). The search also included BLASTX comparison to root and leaf ESTs with the aim of reducing the number of candidates to those most likely to be preferentially or specifically expressed in sperm. Of 46 sequences that had no match with root or leaf ESTs, 7 sequences (highlighted in blue) were expressed preferentially in pollen as determined using Gene Atlas. Of the remaining 833 sequences, 511 sequences which matched leaf or root sequences with an e-value smaller than e^{-70} were selected as they were less likely to be root or leaf ESTs and were therefore better candidates for sperm-cell expressed genes. Of these 511 sequences 85 sequences were expressed preferentially in pollen as determined using Gene Atlas. Thus the total of 92 sequences listed below was considered to have a high

potential to be sperm cell-expressed genes and were prioritised for further investigation (pink highlight indicates identical Arabidopsis matches when blasted with Plumbago sperm ESTs listed in table 5.2).

Gi	e-value	Locus tag	Description	Size (kDa)
15218734		At1g03180	unknown protein	31.1
42571101		At2g37560	origin recognition complex subunit 2 (ORC2)	40.8
15234510		At4g04930	fatty acid desaturase family protein	38.5
15242015		At5g57000	unknown protein	21
18402437		At3g19970	unknown protein	47.1
18405582		At3g27930	unknown protein	47
18410316		At3g55600	unknown protein	26.7
42562774	0.005	At1g55915	unknown protein	47.4
18396813	0.008	At1g28490	syntaxin 61 (SYP61) / osmotic stess-sensitive mutant 1 (OSM1)	27.7
30682738	0.011	At1g12370	type II CPD photolyase PHR1 (PHR1)	56.4
30687718	0.032	At4g27750	unknown protein	33.5
15239802	0.175	At5g49150	AtGEX2	98.6
15236936	0.206	At4g10810	unknown protein	8.5
15234301	2.1e-08	At4g27850	proline-rich family protein	63.6
15226275	2.2e-09	At2g34320	unknown protein	34.2
15237976	5.3e-09	At5g46910	transcription factor jumonji (jmj) family protein	80
15235654	7.5e-17	At4g22450	unknown protein	53.1
15239141	1.1e-18	At5g05490	cohesion family protein SYN1, splice variant 1 (SYN1) [Arabidopsis thaliana]	71.23
30681535	5.6e-21	At1g10090	unknown protein	28
42569183	2.0e-24	At2g20440	RabGAP/TBC domain-containing protein	42.7
42570881	1.1e-26	At2g22740	SET domain-containing protein (SUVH6)	87.5
15241826	1.4e-26	At5g39650	unknown protein	26.5
15231874	2.4e-29	At3g07820	polygalacturonase 3 (PGA3) / pectinase	41.6
15231707	3.1e-30	At3g59550	cohesion family protein SYN3 (SYN3)	77.1
15219369	2.1e-30	At1g73070	leucine-rich repeat family protein	65.5
15233958	1.1e-31	At4g32830	protein kinase, putative	33.9
18394331	1.5e-32	At1g16170	unknown protein	10.2
15230287	2.2e-32	At3g19090	RNA-binding protein, putative	48.9
15223820	1.3e-33	At1g14560	mitochondrial substrate carrier family protein	36.3
15229482	1.3e-34	At3g23610	dual specificity protein phosphatase (DsPTP1)	22
42572743	1.4e-34	At3g60540	sec61beta family protein	8
18423731	1.1e-35	At5g55290	ATP synthase subunit H family protein	7.7
30687235	2.2e-35	At5g19310	homeotic gene regulator, putative	12
15231435	1.9e-36	At3g07270	GTP cyclohydrolase I	51.3
18379075	2.1e-38	At1g03250	unknown protein	27.7
18414376	3.1e-38	At4g15093	catalytic LigB subunit of aromatic ring-opening dioxygenase family	30
30686584	9.8e-38	At3g22290	unknown protein	33
18399023	7.3e-39	At2g19380	RNA recognition motif (RRM)-containing protein	96
15230385	1.4e-40	At3g51030	thioredoxin H-type 1 (TRX-H-1)	12.6
18423407	1.7e-40	At5g52560	UDP-N-acetylglucosamine pyrophosphorylase-related	67.8
15224165	2.5e-40	At2g18390	ADP-ribosylation factor-like protein 2 (ARL2)	21
30689216	3.7e-40	At3g28490	oxidoreductase, 2OG-Fe(II) oxygenase family protein	24.1
42572379	7.0e-41	At3g11530	vacuolar protein sorting 55 family protein / VPS55 family protein	12.4
30681950	1.3e-43	At4g11920	WD-40 repeat family protein	52.3
18416334	4.6e-43	At4g24380	unknown protein	26.3
42571723	9.9e-43	At1g32400	senescence-associated family protein	31.5
15242345	1.1e-44	At5g15600	unknown protein	12.8
15241050	1.5e-44	At5g55850	nitrate-responsive NOI protein, putative	8.3
18414504	1.8e-44	At5g04250	OTU-like cysteine protease family protein	39.2
42563852	2.4e-44	At3g08990	yippee family protein	12.5
30686081	2.3e-45	At5g16880	VHS domain-containing protein / GAT domain-containing protein	45.3
22327003	1.3e-46	At5g23130	peptidoglycan-binding LysM domain-containing	43.2

			protein	
30690730	1.4e-46	At4g36690	U2 snRNP auxiliary factor large subunit, putative	63.5
42565049	1.4e-46	At3g20190	leucine-rich repeat transmembrane protein kinase, putative	72
15227651	9.8e-46	At2g03470	myb family transcription factor / ELM2 domain-containing protein	51.8
18422940	1.0e-48	At5g48655	zinc finger (C3HC4-type RING finger) family protein	22.3
15240448	3.4e-48	At5g27150	sodium proton exchanger / Na+/H+ antiporter (NHX1)	59.5
15240905	1.4e-49	At5g01180	proton-dependent oligopeptide transport (POT) family protein	63.3
22328569	2.5e-49	At4g11560	bromo-adjacent homology (BAH) domain-containing protein	66.5
15237037	6.8e-49	At4g35560	unknown protein	101.5
15222974	2.7e-51	At1g13880	ELM2 domain-containing protein	48.4
15240750	5.5e-51	At5g07290	RNA recognition motif (RRM)-containing protein	99.9
30689114	2.0e-52	At5g23720	dual specificity protein phosphatase family protein	104.2
18410883	5.2e-52	At1g75560	zinc knuckle (CCHC-type) family protein	28.7
18402506	2.0e-53	At1g49160	protein kinase family protein	61.7
15224499	8.7e-53	At2g38400	alanine--glyoxylate aminotransferase, putative / beta-alanine-pyruvate aminotransferase, putative / AGT, putative	51.9
18409246	1.3e-55	At1g69330	zinc finger (C3HC4-type RING finger) family protein	30.4
42561646	8.4e-55	At1g03260	unknown protein	29.5
15235159	9.3e-55	At4g25590	actin-depolymerizing factor, putative	15.2
15232798	4.1e-56	At3g47440	major intrinsic family protein / MIP family protein	26.6
18403408	6.6e-56	At1g51200	zinc finger (AN1-like) family protein	18.4
15224839	6.8e-56	At2g19770	profilin 4 (PRO4) (PFN4)	14.5
18419910	5.9e-57	At4g36800	RUB1-conjugating enzyme, putative (RCE1)	12.7
30688490	1.1e-59	At4g29910	origin recognition complex subunit 5-related (ORC5)	60.6
15221783	4.1e-59	At1g54290	eukaryotic translation initiation factor SUI1, putative	12.6
18413797	2.8e-60	At5g01960	zinc finger (C3HC4-type RING finger) family protein	47.8
18391271	1.1e-61	At1g11380	unknown protein	27.9
18424892	2.7e-61	At5g64930	CPR5 protein, putative	63.1
15238015	3.2e-61	At5g47100	calcineurin B-like protein 9 (CBL9)	24.5
18391117	6.9e-62	At1g10150	unknown protein	45.5
15237901	3.0e-63	At5g17060	ADP-ribosylation factor, putative	21.8
42572053	6.8e-63	At1g69840	band 7 family protein	31.4
15235240	7.8e-63	At4g28260	unknown protein	58.2
15242766	1.0e-64	At5g03470	serine/threonine protein phosphatase 2A (PP2A) regulatory subunit B' (B'alpha)	57.5
15222887	1.5e-64	At1g13570	F-box family protein	47.6
15234910	3.3e-64	At4g10040	cytochrome c, putative	12.2
18424791	9.0e-64	At5g64130	unknown protein	12.3
15240744	4.2e-65	At5g07250	rhomboid family protein	38.4
15239697	9.9e-65	At5g54640	histone H2A	13.7
30681020	4.6e-66	At3g09800	clathrin adaptor complex small chain family protein	19.7
15223072	4.5e-68	At1g76490	3-hydroxy-3-methylglutaryl-CoA reductase 1 / HMG-CoA reductase 1 (HMG1)	63.6
15218445	6.7e-69	At1g43710	serine decarboxylase	54.1

In addition, 1,522 *Plumbago zeylanica* sperm ESTs, 61 *Oryza sativa* sperm ESTs, and 3 *Nicotiana tabaccum* sperm ESTs were searched against the Arabidopsis protein database using BLASTX and the blosum45 matrix (where the sequence alignment was less than 45% identical to one another, and with an e-values of 0.001). Of the 1,522 *Plumbago zeylanica* sperm ESTs, 519 were found to have a good match to Arabidopsis proteins. 24 rice sperm ESTs and all three tobacco sperm ESTs were also identified as having very good hits to Arabidopsis proteins (supplement CD). The 519 hits from *Plumbago zeylanica* were made non-redundant and reduced to 302 sequences. These sequences were manually searched using GeneAtlas microarray data for pollen-specific

transcripts. 46 sequences were found to be expressed highly in pollen. Out of 24 Arabidopsis proteins that matched rice sperm ESTs, 3 sequences were found to be expressed highly in pollen. 1 Arabidopsis protein that matched tobacco sperm ESTs were found to be highly expressed in pollen (all sperm candidates are shown in table 5.2 below).

Table 5.2. Arabidopsis sperm cell protein candidates obtained following *in silico* analysis of Plumbago, rice, and tobacco sperm EST data.

List of 50 Arabidopsis sperm protein candidates found to be preferentially expressed in pollen following manual analysis of 329 sequences using Gene Atlas. Of 302, 24, and 3 non-redundant sequences that matched Plumbago, rice, and tobacco sperm ESTs, 46, 3, and 1 sequences respectively were preferentially expressed in pollen as listed below (pink fill indicates identical Arabidopsis matches when blasted with maize sperm ESTs listed in table 5.1).

Locus tag	Description	Size (kDa)
Plumbago zeylanica		
At3g53750	actin 3 (ACT3)	41.8
At5g52360	actin-depolymerizing factor, putative	15.9
At5g59370	actin 4 (ACT4)	41.8
At4g25590	actin-depolymerizing factor, putative	15.2
At3g27440	uracil phosphoribosyltransferase, putative / UMP pyrophosphorylase, putative / UPRTase, putative	52.1
At4g27960	ubiquitin-conjugating enzyme E2-17 kDa 9 (UBC9)	20.2
At3g52560	ubiquitin-conjugating enzyme family protein	16.5
At4g05050	polyubiquitin (UBQ11)	25.7
At5g54640	histone H2A	13.7
At2g46170	reticulon family protein (RTNLB5)	28.7
At4g04690	F-box family protein (FBX15)	40.1
At5g04250	OTU-like cysteine protease family protein	39.2
At3g14040	exopolygalacturonase / galacturan 1,4-alpha-galacturonidase / pectinase	45.7
At3g06580	galactokinase (GAL1)	54.3
At5g26150	protein kinase family protein	78.5
At1g11250	syntaxin, putative (SYP125)	33.8
At5g38820	amino acid transporter family protein	49.7
At1g69960	serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	35
At2g02960	zinc finger (C3HC4-type RING finger) family protein	29.6
At1g55530	zinc finger (C3HC4-type RING finger) family protein	39
At3g55430	glycosyl hydrolase family 17 protein / beta-1,3-glucanase, putative	48.4
At1g55570	multi-copper oxidase type I family protein	62.2
At2g19770	profilin 4 (PRO4) (PFN4)	14.5
At3g05820	beta-fructofuranosidase, putative / invertase, putative / saccharase, putative / beta-fructosidase, putative	71.9
At4g28000	AAA-type ATPase family protein	81.2
At2g29940	ABC transporter family protein	162
At3g61230	LIM domain-containing protein	23.5
At1g34340	esterase/lipase/thioesterase family protein	59.7
At3g60540	sec61beta family protein	8.2
At1g48940	plastocyanin-like domain-containing protein	19.7
At5g48140	polygalacturonase, putative / pectinase, putative	42.5

At3g07850	exopolygalacturonase / galacturan 1,4-alpha-galacturonidase / pectinase	45.6
At1g30540	ATPase, BadF/BadG/BcrA/BcrD-type family	36.6
At1g13890	SNAP25 homologous protein, putative / synaptosomal-associated protein SNAP25-like, putative (SNAP30)	29
At1g78300	14-3-3 protein GF14 omega (GRF2)	29.1
At1g74480	RWP-RK domain-containing protein	33.9
At4g24380	unknown protein	26.4
At1g26750	unknown protein	20
At5g54120	unknown protein	22.3
At3g50910	unknown protein	49.6
At5g64400	unknown protein	14.4
At2g20700	unknown protein	19.8
At5g38760	unknown protein	7
At2g20480	unknown protein	7.1
At5g53820	unknown protein	7
At5g50930	unknown protein	27.3
Oryza sativa		
At4g05320	polyubiquitin (UBQ10) (SEN3)	52
At3g22760	CXC domain containing TSO1-like protein 1 (SOL1)	66.7
At3g18040	mitogen-activated protein kinase, putative / MAPK, putative (MPK9)	58.4
Nicotiana tabaccum		
At1g02790	exopolygalacturonase / galacturan 1,4-alpha-galacturonidase (PGA3) / pectinase	44.4

The Arabidopsis genes which matched maize, Plumbago, rice, and tobacco ESTs were divided into 14 different categories (TAIR gene ontology annotation) depending on their probable biological functions (Table 5.3 and Figure 5.1; note – for the maize analysis, of the sequences that had some degree of match to root and leaf transcripts only those sequences with match scores of e^{-0}-e^{-50} were included in this functional categorisation- the group of e^{-51}-e^{-69} were analysed later in the project). The biggest class was ‘unknown protein’ which contained 27 sequences out of the 107. The second biggest functional category included 12 sequences involved in ‘transport’ followed by ‘DNA/RNA binding’ (11 sequences), ‘signalling’, ‘ubiquitination’ and ‘miscellaneous’ (9 sequences each), ‘cytoskeleton’ and ‘carbohydrate metabolism’ (6 sequences each), ‘cell wall degradation’, ‘energy’ and ‘oxidoreductase’ (4 sequences each) and finally cell cycle, chromosome remodelling, and ion exchange at 2 sequences each (Table 5.3).

Table 5.3. Functional categorisation of Arabidopsis sperm cell-expressed gene candidates derived from maize, Plumbago, rice, and tobacco EST searches.

Sperm cell-expressed gene candidates were grouped into 14 functional categories (TAIR gene ontology annotation) to facilitate more selective candidate selection for transcriptional studies. Categories with possible roles in sperm cell production or crucial roles in sperm function were of interest (listed in table 5.4). The group that contain the most members was ‘unknown proteins’ (27 sequences). The groups that contained the least members were ‘cell cycle’, ‘chromosome remodelling’ and ‘ion exchange’ having 2 sequences each.

Locus tag	Description	Size (kDa)
Carbohydrate metabolism		
At3g55430	glycosyl hydrolase family 17 protein / beta-1,3-glucanase, putative	48.4
At3g06580	Galactokinase (GAL1)	54.3
At3g05820	beta-fructofuranosidase, putative / invertase, putative / saccharase, putative / beta-fructosidase, putative	71.9
At1g02790	exopolygalacturonase / galacturan 1,4-alpha-galacturonidase (PGA3) / pectinase	44.4
At1g55570	multi-copper oxidase type I family protein	62.2
At3g07820	polygalacturonase 3 (PGA3) / pectinase	41.6
Cell cycle		
At5g05490	cohesion family protein SYN1, splice variant 1 (SYN1)	70.2
At3g59550	cohesion family protein SYN3 (SYN3)	77.1
Cell wall degradation		
At3g14040	exopolygalacturonase / galacturan 1,4-alpha-galacturonidase / pectinase	45.7
At5g48140	polygalacturonase, putative / pectinase, putative	42.5
At3g07850	exopolygalacturonase / galacturan 1,4-alpha-galacturonidase / pectinase	45.6
At5g23130	peptidoglycan-binding LysM domain-containing protein	43.2
Chromosome remodelling		
At5g54640	histone H2A	13.7
At2g22740	SET domain-containing protein (SUVH6)	87.5
Cytoskeleton		
At3g53750	actin 3 (ACT3)	41.8
At5g59370	actin 4 (ACT4)	41.8
At5g52360	actin-depolymerizing factor, putative	15.9
At4g25590	actin-depolymerizing factor, putative	15.2
At2g19770	profilin 4 (PRO4) (PFN4)	14.5
At2g18390	ADP-ribosylation factor-like protein 2 (ARL2)	21
DNA/RNA binding		
At2g19380	RNA recognition motif (RRM)-containing protein	96
At4g36690	U2 snRNP auxiliary factor large subunit, putative	63.5
At2g37560	origin recognition complex subunit 2 (ORC2)	40.8
At4g11560	bromo-adjacent homology (BAH) domain-containing protein	66.5
At2g03470	myb family transcription factor / ELM2 domain-containing protein	51.8
At5g46910	transcription factor jumonji (jmj) family protein	80
At1g74480	RWP-RK domain-containing protein	33.9
At3g22760	CXC domain containing TSO1-like protein 1 (SOL1)	66.7
At3g61230	LIM domain-containing protein	23.5
At5g19310	homeotic gene regulator, putative	12
At3g19090	RNA-binding protein, putative	48.9
Energy		
At5g55290	ATP synthase subunit H family protein	7.7
At1g30540	ATPase, BadF/BadG/BcrA/BcrD-type family	36.6
At3g07270	GTP cyclohydrolase I	51.3
At5g52560	UDP-N-acetylglucosamine pyrophosphorylase-related	67.8
Ion exchange		
At5g27150	sodium proton exchanger / Na+/H+ antiporter (NHX1)	59.5
At1g48940	plastocyanin-like domain-containing protein	19.7
Oxidoreductase		
At3g28490	oxidoreductase, 2OG-Fe(II) oxygenase family protein	24.1
At4g15093	catalytic LigB subunit of aromatic ring-opening dioxygenase family	30
At4g04930	DES-1-LIKE (fatty acid desaturase 1-like); oxidoreductase	38.5
At3g51030	thioredoxin H-type 1 (TRX-H-1)	12.6
Signalling		
At4g32830	protein kinase, putative	33.9
At5g26150	protein kinase family protein	78.5
At3g18040	mitogen-activated protein kinase, putative / MAPK, putative (MPK9)	58.4
At1g69960	serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	35
At1g73070	leucine-rich repeat family protein	65.5
At3g20190	leucine-rich repeat transmembrane protein kinase, putative	72
At3g23610	dual specificity protein phosphatase (DsPTP1)	22
At1g78300	14-3-3 protein GF14 omega (GRF2)	29.1
At4g11920	WD-40 repeat family protein	52.3
Transport		
At5g01180	proton-dependent oligopeptide transport (POT) family protein	63.3
At2g29940	ABC transporter family protein	161.9
At5g38820	amino acid transporter family protein	49.7

At2g20440	RabGAP/TBC domain-containing protein	42.7
At3g60540	sec61beta family protein	8
At3g27440	uracil phosphoribosyltransferase, putative / UMP pyrophosphorylase, putative / UPRTase, putative	52.1
At1g13890	SNAP25 homologous protein, putative / synaptosomal-associated protein SNAP25-like, putative (SNAP30)	29
At3g11530	vacuolar protein sorting 55 family protein / VPS55 family protein	12.4
At1g14560	mitochondrial substrate carrier family protein	36.3
At1g11250	syntaxin, putative (SYP95)	33.8
At1g28490	syntaxin 61 (SYP61) / osmotic stess-sensitive mutant 1 (OSM1)	27.7
At5g16880	VHS domain-containing protein / GAT domain-containing protein	45.3
Ubiquitination		
At5g48655	zinc finger (C3HC4-type RING finger) family protein	22.3
At2g02960	zinc finger (C3HC4-type RING finger) family protein	29.6
At1g55530	zinc finger (C3HC4-type RING finger) family protein	39
At4g04690	F-box family protein (FBX15)	40.1
At3g08990	yippee family protein	12.5
At4g27960	ubiquitin-conjugating enzyme E2-17 kDa 9 (UBC9)	20.2
At3g52560	ubiquitin-conjugating enzyme family protein	16.5
At4g05050	Polyubiquitin (UBQ11)	25.7
At4g05320	Polyubiquitin (UBQ10) (SEN3)	52
Miscellaneous		
At5g04250	OTU-like cysteine protease family protein	39.2
At2g46170	reticulon family protein (RTNLB5)	28.7
At4g28000	AAA-type ATPase family protein	81.2
At5g55850	nitrate-responsive NOI protein, putative	8.3
At1g32400	senescence-associated family protein	31.5
At1g12370	type II CPD photolyase PHR1 (PHR1)	56.4
At1g34340	esterase/lipase/thioesterase family protein	59.7
At4g27850	proline-rich family protein	63.6
At5g49150	AtGEX2	98.6
Unknown		
At5g54120	unknown protein	22.3
At3g27930	unknown protein	47
At3g55600	unknown protein	26.7, 29
At4g10810	unknown protein	8.5
At1g10090	unknown protein	28
At5g39650	unknown protein	26.5
At1g16170	unknown protein	10.2
At1g03250	unknown protein	27.7
At3g22290	unknown protein	33
At5g15600	unknown protein	12.8
At4g35560	unknown protein	101.5
At4g24380	unknown protein	26.4
At1g26750	unknown protein	20
At3g50910	unknown protein	49.6
At5g64400	unknown protein	14.4
At2g20700	unknown protein	19.8
At5g38760	unknown protein	7
At2g20480	unknown protein	7.1
At5g53820	unknown protein	7
At5g57000	unknown protein	21
At3g19970	unknown protein	47.1
At1g55915	unknown protein	47.4
At4g27750	unknown protein	33.5
At4g22450	unknown protein	53.1
At5g50930	unknown protein	27.3
At2g34320	unknown protein	34.2
At1g03180	unknown protein	31.1

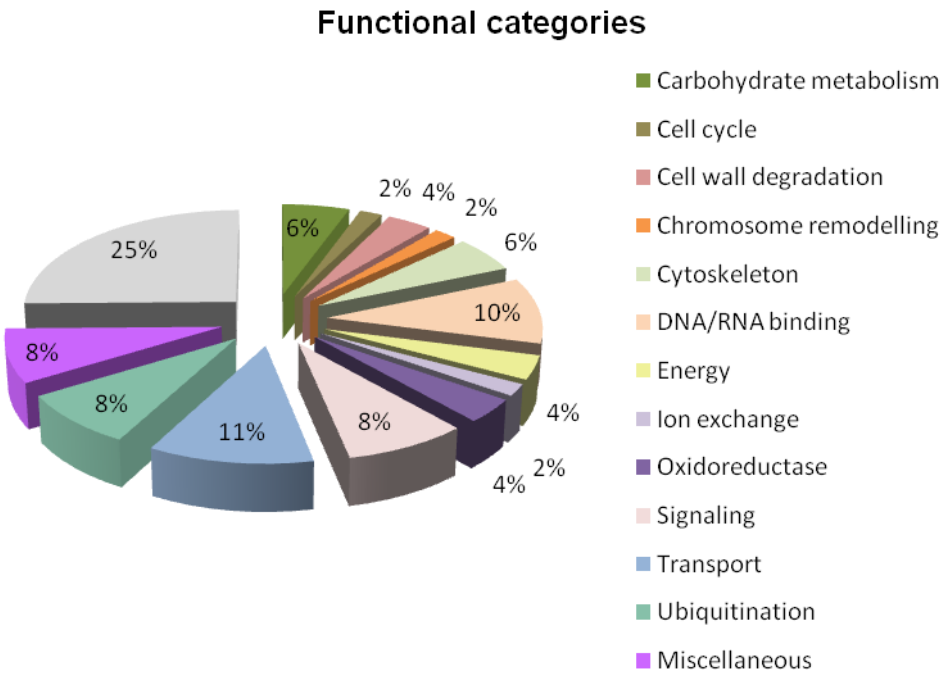


Figure 5.1. Pie chart representation of functional categories for sperm cell-expressed gene candidates.

Arabidopsis sperm cell-expressed gene candidates derived from maize, Plumbago, rice, and tobacco EST data placed in 14 functional categories (TAIR gene ontology annotation; listed in table 5.3) and displayed as a pie chart.

Genes placed in the categories ‘cell cycle’, ‘cytoskeleton’, ‘chromosome remodelling’ and ‘signalling’ (19 sequences) were chosen for the last stage of the candidate selection process as they could potentially be involved in sperm cell production and function. Moreover, transmembrane proteins (16 sequences) were also of special interest as they could potentially be involved in gamete recognition and fusion. In addition, candidate genes that were cited in the literature (18 sequences including some sequences that overlap with the groups mentioned above) as being involved in either plant or animal reproduction processes including some proteins that were present in the Drosophila sperm proteome (personal communication with Dr Timothy L Karr, University of Bath) and expressed preferentially in tricellular or mature pollen stages (Gene Atlas) were treated as potential Arabidopsis sperm cell-expressed gene candidates. All of these priority sperm cell-expressed candidate genes (Table 5.4) were investigated using T-DNA insertion knockout lines and will be reported on in detail in chapter 7. Of the proteins listed in table 5.4, those that were likely to be expressed in the pollen vegetative cell or described to have a function relating to pollen dynamics or pollen tube growth were subtracted from the list. The remaining 20 genes (Table 5.5) that could potentially be expressed in sperm cells were considered as high priority candidates for further investigated using reverse genetics and other molecular approaches. The detailed analyses of these genes will be described in chapter 6.

Table 5.4. List of candidate sperm cell-expressed genes selected from categories of interest that could potentially be involved in sperm cell functions.

Sperm cell-expressed candidates (19 sequences) categorised into the functional groups of 'cell cycle', 'chromosome remodelling', 'cytoskeleton' and signalling were selected to be investigated as they had good potential to be expressed in sperm. Moreover, proteins possessing transmembrane domains (16 sequences) were also of special interest as they could potentially be involved in the gamete fusion event. In addition, proteins that have been reported in the literature to be involved in reproductive processes (18 sequences) were also included.

Locus tag	Description	Size (kDa)
Cell cycle		
At5g05490	cohesion family protein SYN1, splice variant 1 (SYN1)	70.2
At3g59550	cohesion family protein SYN3 (SYN3)	77.1
Chromosome remodelling		
At5g54640	histone H2A	13.7
At2g22740	SET domain-containing protein (SUVH6)	87.5
Cytoskeleton		
At3g53750	actin 3 (ACT3)	41.8
At5g59370	actin 4 (ACT4)	41.8
At5g52360	actin-depolymerizing factor, putative	15.9
At4g25590	actin-depolymerizing factor, putative	15.2
At2g19770	profilin 4 (PRO4) (PFN4)	14.5
At2g18390	ADP-ribosylation factor-like protein 2 (ARL2)	21
Signaling		
At4g32830	protein kinase, putative	33.9
At5g26150	protein kinase family protein	78.5
At3g18040	mitogen-activated protein kinase, putative / MAPK, putative (MPK9)	58.4
At1g69960	Serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	35
At1g73070	leucine-rich repeat family protein	65.5
At3g20190	leucine-rich repeat transmembrane protein kinase, putative	72
At3g23610	dual specificity protein phosphatase (DsPTP1)	22
At1g78300	14-3-3 protein GF14 omega (GRF2)	29.1
At4g11920	WD-40 repeat family protein	52.3
Transmembrane		
At1g32400	senescence-associated family protein	31.5
At5g38820	Amino acid transporter family protein	49.7
At3g28490	oxidoreductase, 2OG-Fe(II) oxygenase family protein	24.1
At5g27150	sodium proton exchanger / Na+/H+ antiporter (NHX1)	59.5
At3g11530	vacuolar protein sorting 55 family protein / VPS55 family protein	12.4
At5g01180	Proton-dependent oligopeptide transport (POT) family protein	63.3
At2g46170	reticulon family protein (RTNLB5)	28.7
At2g29940	ABC transporter family protein	162
At1g48940	plastocyanin-like domain-containing protein	19.7
At3g23870	permease-related	36.3
At5g39650	unknown protein	26.5
At4g04930	fatty acid desaturase family protein	38.5
At3g20190	leucine-rich repeat transmembrane protein kinase, putative	72
At5g50930	unknown protein	27.3
At1g10090	unknown protein	28
At5g49150	AtGEX2	98.6
Genes reported to be involved in reproduction		
At1g11250	syntaxin, putative (SYP95)	33.8
At2g46170	reticulon family protein (RTNLB5)	28.7
At2g19770	profilin 4 (PRO4) (PFN4)	14.5
At4g04930	fatty acid desaturase family protein	38.5
At2g03470	myb family transcription factor / ELM2 domain-containing protein	51.8
At3g61230	LIM domain-containing protein	23.5
At2g29940	ABC transporter family protein	162

At5g04250	OTU-like cysteine protease family protein	39.2
At3g23610	dual specificity protein phosphatase (DsPTP1)	22
At2g22740	SET domain-containing protein (SUVH6)	87.5
At2g20440	RabGAP/TBC domain-containing protein	42.7
At1g69960	Serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	35
At5g54640	histone H2A	13.7
At4g27960	ubiquitin-conjugating enzyme E2-17 kDa 9 (UBC9)	20.2
At5g23130	peptidoglycan-binding LysM domain-containing protein	43.2
At4g28000	AAA-type ATPase family protein	81.2
At2g19380	RNA recognition motif (RRM)-containing protein	96
At2g18390	ADP-ribosylation factor-like protein 2 (ARL2)	21

Table 5.5. High priority Arabidopsis sperm cell-expressed candidates.

Sperm cell-expressed candidates (19 sequences) categorised into the functional groups of 'cell cycle', 'chromosome remodelling', 'cytoskeleton' and 'signalling' including transmembrane proteins (16 sequences) and proteins reported to be involved in reproduction process (18 sequences) either in plants or animals were selectively chosen using following criteria: i) have no known function involved in pollen development or pollen germination and pollen tube growth, ii) expressed preferentially in tricellular or mature pollen stages (these stages contain sperm cells). These 20 high priority candidates listed below were selected for detailed transcriptional analyses - these studies will be described in chapters 6 and 7.

Locus tag	Description	Size (kDa)
At4g32830	kinase/ protein serine/threonine kinase	33.9
At5g23130	peptidoglycan-binding LysM domain-containing protein	43.2
At1g73070	leucine-rich repeat family protein	65.5
At3g20190	leucine-rich repeat transmembrane protein kinase, putative	72
At2g29940	ABC transporter family protein	161.9
At3g23610	dual specificity protein phosphatase (DsPTP1)	22
At1g78300	14-3-3 protein GF14 omega (GRF2)	29.1
At4g11920	WD-40 repeat family protein	52.3
At4g28000	AAA-type ATPase family protein	81.2
At5g54640	histone H2A	13.7
At1g11250	syntaxin, putative (SYP95)	33.8
At2g22740	SET domain-containing protein (SUVH6)	87.5
At2g19380	RNA recognition motif (RRM)-containing protein	96
At5g39650	unknown protein	26.5
At1g10090	unknown protein	28
At5g50930	unknown protein	27.3
At1g32400	senescence-associated family protein	31.5
At3g61230	LIM domain-containing protein	23.5
At2g20440	RabGAP/TBC domain-containing protein	42.7
At1g69960	serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	35

The remaining 2,467 EST sequences of maize that were found not to match to annotated regions of the Arabidopsis genome were translated and searched again against the whole Arabidopsis genome translated into all six reading frames. The proteins were then BLASTP and 7,381 matches with e-values smaller than 0.001 were retained. The sequences with the best e-value for each match were retained and divided up to corresponding chromosomes and made into GenBank feature tables. These 215 sequences were the manually analysed further using the Artemis tool (Figure 5.2) that allows visual comparison of the selected genome sequences with the sequence of interest. Most of the sequences obtained matched retrotransposons. Only two sequences were found to potentially encode Arabidopsis sperm proteins (Table 5.6). These two sequences were analysed utilising gene identification tools. The predicted gene was then confirmed to be expressed in Arabidopsis by RT-PCR. The detail of these analyses is reported in chapter 6.

Table 5.6. Novel Arabidopsis sperm protein candidates identified in unannotated regions of the Aeabidopsis genome (obtained from maize sperm ESTs searched against the whole Arabidopsis genome)

Maize sperm ESTs that did not matched Arabidopsis proteins were blasted against the whole genome of Arabidopsis. Two potential protein sequences named AtNOV1 and AtNOV2 were identified in chromosome 2 and 4 respectively (Chm. No; chromosome number, Nt; Nucleotide).

Chro. No.	Name	Nt matched start	Nt matched stop	E-values	Matched protein Sequences (Arabidopsis)
2	AtNOV1	17231665	17231853	7.00E-10	RVRVGLYSGPPIPSWELRHGLRAKARVSCLLEWRPREAGFTEQRPPPALDGGRITGHCLPSP
4	AtNOV2	10980688	10980774	8.00E-13	FLFNNEMLQELQWFKQSYGSWFLGDYISE

Figure 5.2. Artemis tool display screen.

The Artemis tool allows visual comparison of the sequences of interest with the 'featured' Arabidopsis genome. The sequences highlighted in red are maize ESTs that match to that particular region of Arabidopsis genome. The quality of the matches is indicated by the e-value shown under the highlighted region. Both the nucleotide sequence and all 6 translated reading frames are displayed on the screen.

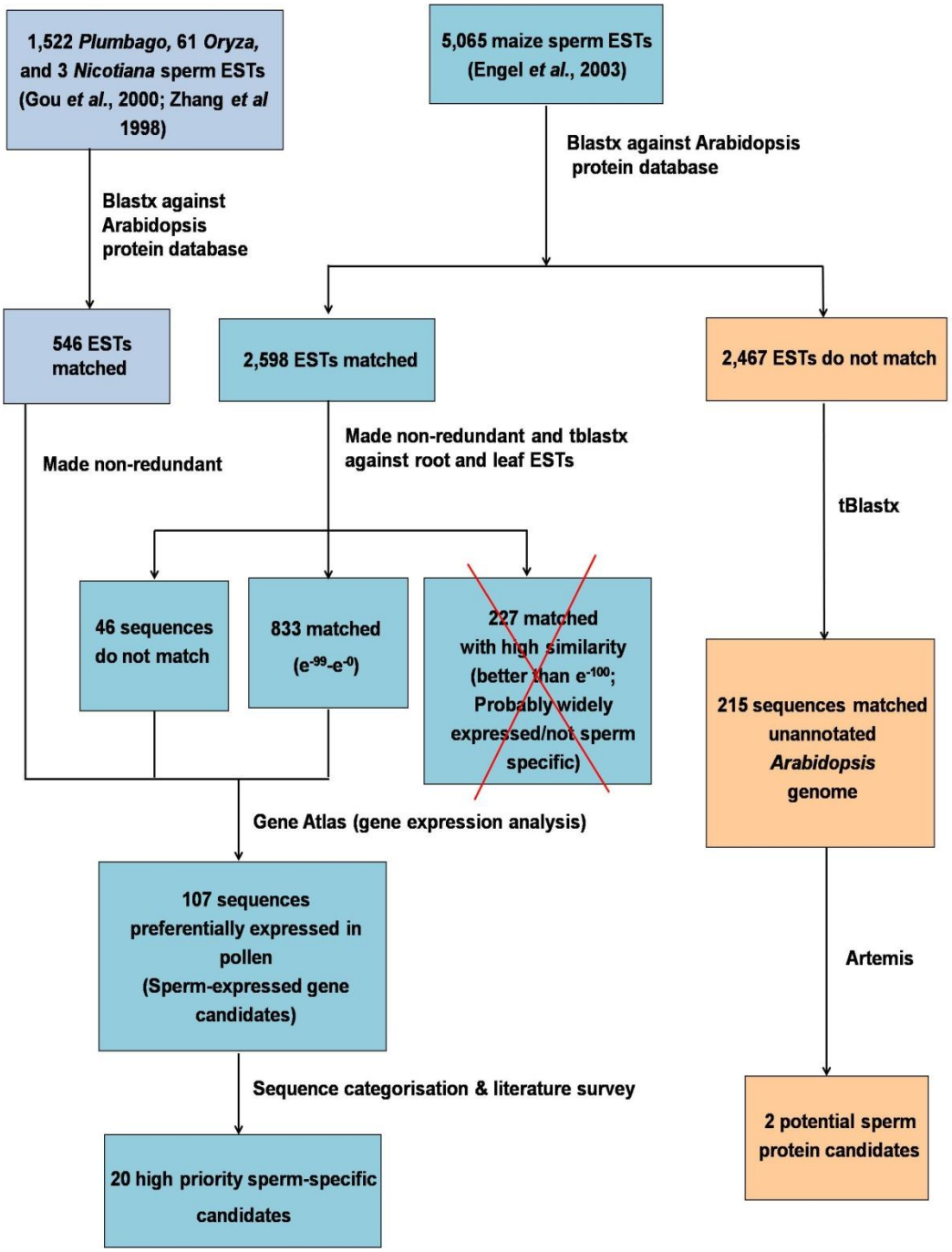


Figure 5.3. Flowchart representing the step-by-step procedures performed utilising a bioinformatics approach to identify high priority sperm cell-expressed candidates for further transcriptional analyses and T-DNA insertion mutagenesis functional studies.

5.3 Discussion

Sperm cell-expressed sequence tags (ESTs) from four plant species i.e. *Zea mays*, *Plumbago zeylanica, Oryza sativa,* and *Nicotiana tabaccum* were employed in a comparative approach to identify genes expressed in Arabidopsis sperm cells. Bioinformatics techniques were utilised to automate the large number of sequence alignments and to identify the expression pattern of each selected gene. A total of 6,654 sperm EST sequences were translated in all 6 reading frames and blasted (BLASTX) against the Arabidopsis protein database. BLASTX is very useful when homologous proteins are sought when the nucleotide coding region may have diverged, as is likely to be the case in this cross-species analysis. BLASTX compares translational products of the nucleotide query sequence to a protein database. Because BLASTX translates the query sequence in all six reading frames and provides combined significance statistics for hits to different frames, it is particularly useful when the reading frame of the query sequence is unknown or it contains errors that may lead to frame shifts or other coding errors. It is used extensively in analyzing EST sequences as it is more sensitive than nucleotide BLAST since the comparison is performed at the protein level (NCBI). Of 5,065 maize ESTs analysed by BLASTX, 2,598 sequences (51%) matched Arabidopsis proteins at a threshold E-value smaller than 0.001. Of these 2,598 sequences, 579 ESTs (22.23%) were identified as 'unknown proteins'. Some of these unknown proteins identified in this project could potentially be sperm cell-specific as their functions have never been annotated suggesting that they might not only be rare and specific to a specialised cell type but that they have a specific function limited to that cell type. These two flowering plant species, maize (monocot) and Arabidopsis (dicot), were estimated to have diverged from a common ancestor more than 150 million years ago (MYA) (Brendel et al., 2004). Maize and Arabidopsis protein and EST analysis using comparative genomic methodologies and bioinformatics suggested that approximately 90% of maize proteins have homology to Arabidopsis sequences (BLASTP e-value = $1e^{-5}$) and 68% of the maize ESTs matched Arabidopsis proteins (BLASTP e-value = $1e^{-2}$) (Brendel et al., 2004). However, when Arabidopsis ESTs were compared to the Arabidopsis proteins, 88% were matched (Brendel et al., 2004). Therefore, the adjusted estimation of protein homology between these two species was approximately 60-90% (Brendel et al., 2004). However, in this project, only 50% of maize ESTs matched annotated Arabidopsis proteins. This lower figure is possibly due to the fact that only sperm cell ESTs were used in this project whereas in Brendel et al., (2004) all available ESTs were used in the comparison. Interestingly, only 34% of *Plumbago zeylanica* ESTs (519 sequences) and 39% rice ESTs (24 sequences) matched to Arabidopsis proteins. Despite the fact that both *P. zeylanica* and Arabidopsis are dicot species, the overall homology of the proteins from these two species examined in this study seems to be less than that for maize. In addition,

the percentage of homologous proteins between rice and Arabidopsis was also lower than maize despite the fact that both rice and maize are monocots.

These Arabidopsis genes that matched to maize sperm cell ESTs were then compared to Arabidopsis root and leaf ESTs obtained from the NCBI database. This step was performed in order to reduce the number of sperm cell protein candidates to manageable levels by eliminating proteins that were also expressed in roots or leaves and thus unlikely to be sperm specific. It was postulated that many of these proteins were likely to be expressed in sperm cells however their function was unlikely to be specific for or crucial to reproductive processes. From these sequence comparisons, 46 sequences did not match any root or leaf ESTs and were selected for further analysis. The remaining sequences matched to root or leaf ESTs with different e-values. The matches with e-value smaller than e^{-100} (good match, 277 sequences) were considered as probable root or leaf ESTs and eliminated from the further studies. The rest of the matches (833 sequences) were ranked according to their e-values. The sequences that matched with values e^{-0}-10^{-69} (511 sequences) were selected for further analysis as probable sperm cell-expressed gene candidates. The expression patterns of these 511 genes from maize together with 329 genes from Plumbago, rice and tobacco were analysed using the Gene Atlas tool to identify pollen-expressed genes. This step was performed to again eliminate genes that are constitutively expressed in other tissue types and thus unlikely to be crucial or specific to plant sexual reproduction. Only genes that were expressed highly or preferentially in pollen grains were selected for further analysis. Of the 840 genes, 121 genes were selected this way. It is important to note that these genes were selected based on preferential expression in pollen rather than being exclusively expressed in pollen as some potentially important sperm genes demonstrate such a transcript signature. *GEX2*, a known sperm-expressed protein was used as a reference for identifying sperm-expressed genes. The expression pattern of *GEX2* observed utilising the Gene Atlas tool indicated relatively low expression levels in pollen (expression level = 612) but with expression also being detected in other tissue types at very low levels (less than 200). Therefore, all genes that were observed to have high expression in pollen in comparison to other tissues were retained regardless of the actual expression level. Moreover, genes that appeared to be expressed at late stages of pollen development (tricellular and mature pollen) were of particular interest when selecting sperm cell-expressed candidates for further analysis. The next step of candidate selection was functional classification of the genes (TAIR gene ontology annotation). Categories containing genes that may be involved in spermatogenesis, sperm cell delivery via pollen tube, and the gamete fusion event i.e. 'cytoskeleton', 'cell cycle', 'chromosome remodelling' and 'signal transduction' were selected. Genes encoding polypeptides that contained transmembrane domains were also chosen to be analysed further as these may be involved in male-female gamete interactions. Interestingly, two known sperm cell-expressed genes, *GEX2* (Engel et al., 2005) and *HAP2 (GCS1)* (Mori et al., 2006; von Besser et al., 2006)

were both reported to encode transmembrane proteins, with *hap2* mutants at least known to be defective in gamete fusion. Moreover, a literature search for reports on these pollen-expressed genes was carried out in order to identify probable homologues in other species with reproductive functions; for example Arabidopsis At2g29940, an ABC transporter family protein – detected in the anterior head of sperm and elongated spermatids and may play a role in regulating lipid composition in sperm (rodent) (Ban et al., 2005). However, genes that were reported to be expressed almost exclusively at early stages of pollen development (uninucleate microspore and bicellular pollen stages) were eliminated for example, At1g48940 (plastocyanin-like domain-containing protein) and At3g06580 (galactokinase [GAL1]). The genes that fell into more than one category were selected as high priority candidates to be examined in detail utilising transcriptional studies. Other genes in these categories of interest were initially only to be examined by reverse genetics methodology (Arabidopsis T-DNA insertion knockout lines). Interestingly, *GEX2* – a known Arabidopsis sperm cell-expressed gene also survived the screen to the final stage which was a good indication that the screening approach taken to identify sperm cell gene candidates was effective and reliable.

The other half of the EST sequences that had no match to annotated Arabidopsis proteins were subjected to BLAST searching against the whole Arabidopsis genome. This step was carried out as it was predicted that Arabidopsis sperm ESTs / cDNAs would be underrepresented in the databases and thus may be absent from the annotated genome sequence (Xu et al., 2002) - it seemed likely that some of the maize sperm ESTs would match to unannotated regions of the Arabidopsis genome and could represent novel sperm cell proteins. Of 7,381 matches, only 215 of the best matches were retained after redundant sequences had been eliminated. These sequence matches to the Arabidopsis genome were manually viewed using the Artemis tool (Sanger) in order to identify probable novel protein coding regions. This tool allows the visual comparison of the EST sequence matches with the 'featured' Arabidopsis genome. The results display of any analysis can be viewed in the context of the DNA sequence and its six-frame translation (Rutherford et al., 2000) and thus, the analysis of each particular sequence was straightforward. All three-frame translated EST sequences were aligned with the six translated frames of the genome. Both nucleotide and amino acid sequences were displayed on the same Artemis screen to facilitate candidate sequence selection. Most of these sequences were found to be retroelements which are reported to take up over 50% of nuclear DNA content in many plant species (Kumar and Bennetzen, 1999). However, in Arabidopsis these retrotransposons comprise approximately 15% of the genome (Mroczek and Dawe, 2003). These two sequences demonstrated a good match with their respective putative translatable regions of the Arabidopsis genome having adequate runs of amino acid sequence with no interrupting stop codons - thus these features are good indicators that novel gene regions have been revealed. Further studies of these two sequences will be described in chapter 6.

This *in silico* cross-species comparative study has provided a valuable list of numerous potential sperm cell protein candidates. Strict criteria were established in order to identify the best potential sperm cell proteins for more detailed future analyses. Although the selection criteria used to identify sperm cell proteins might have eliminated some true candidates, the remaining sequences after screening were expected to have a higher chance of encoding important sperm cell proteins that could be crucial for sperm cell function. Various bioinformatics tools were utilised in this study i.e. BLAST, Gene Atlas, and Artemis. These tools were proven to be very effective for managing and analysing the enormous amount of data. Interestingly, the known sperm-expressed gene *GEX2* (Engel et al., 2005) also came through the strict candidate selection processes suggesting that the selection methods employed were effective and that the high priority candidates identified could indeed be sperm cell-specific proteins. The 20 high priority sperm cell protein candidates selected by this strict process, including *GEX2*, were analysed in detail utilising various molecular techniques (i.e. RT-PCR, GFP translational fusion and T-DNA insertion mutagenesis) in order to confirm the expression of these candidates in sperm cells and also to characterise their function. However, the remaining candidates could also be analysed in the future utilising available information from these studied candidates to be more accurately in identification of the important sperm cell protein particularly proteins that are involved in gamete fusion event.

Chapter 6:

Characterisation of Putative Arabidopsis Sperm-Expressed Genes by Transcriptional Analysis

6.1 Introduction

Attempts to identify Arabidopsis sperm cell proteins using proteomics were demonstrated in previous chapters. Problems with these proteomic studies have been addressed and solutions have been established however, some obstacles still need to be overcome as discussed in chapter 4. A parallel approach was taken for sperm protein identification which utilised bioinformatics coupled with transcriptional analysis of selected sperm-expressed gene candidates. The bioinformatics approach and candidate selection were described and discussed in chapter 5. In this chapter the analysis of these selected candidates using transcriptional studies (reverse transcription-PCR [RT-PCR] and GFP fusion) to confirm the presence and specificity of their expression in sperm cells will be described and discussed in detail. Although some problematic aspects of transcriptome studies have been highlighted, these studies are also crucial in this project in order to identify sperm cell-specific transcripts. Moreover, as sperm cells are scarce, transcriptional studies are considered to be very practical because a much smaller amount of starting material is required compared to proteome studies as the target nucleic acids can be amplified. This feature of transcriptional analysis can be tremendously beneficial as the sperm cells of plants, especially Arabidopsis, are incredibly challenging to isolated and purify as mentioned in chapter 3. Together with proteomic approaches, transcriptional analyses should facilitate the identification of Arabidopsis sperm cell proteins.

6.1.1 Overview of techniques utilized for studying gene expression and protein localisation in Arabidopsis: Green fluorescent protein (GFP) reporter fusion and Agrobacterium mediated gene transfer

Fluorescent proteins have been utilised for live imaging of the cell. They have facilitated the discovery of protein function, localization, trafficking, cell dynamics, protein interactions, protein assemblies, or even the responses to developmental and environment cues (Ehrhardt, 2003). These commonly used fluorescent proteins were discovered in marine invertebrates: a Pacific northwest jellyfish, *Aequorea victoria* (Chiu et al., 1996; Cubitt et al., 1995). GFP is considered an effective reporter molecule for many reasons. Firstly it can be easily detected using standard long wavelength UV light when expressed at a sufficient level. Secondly, unlike commonly used reporter genes such as β-galactosidase, alkaline phosphatase, chloramphenicol acetyltransferase (CAT), and 3-glucuronidase (GUS), substrates are not required to visualise the GFP reporter (Prasher, 1995). Thirdly, GFP is a relatively small (26.9 kDa) monomeric molecule which is thus well suited to fusion to a protein of interest at either its N- or C-terminal (Pang et al., 1996; Prasher, 1995).

This engineered GFP fusion protein DNA construct requires a 'vector' for transfer into the desired plant cells or tissues. In early plant transformation studies, particle bombardment, electroporation, microinjection, virus delivery or Agrobacterium inoculation of explants were used to introduce DNA into the plant (Bent, 2000; Hansen and Wright, 1999). However, these methods are generally labour-intensive, inefficient and require tissue culture. In addition, chimeric plants can easily be obtained by using these techniques (Hansen and Wright, 1999). A novel Arabidopsis *in planta* vacuum infiltration transformation technique was developed which had no requirement for plant tissue culture or regeneration (Bechtold et al., 1993; Clough and Bent, 1998; Ye et al., 1999). This technique gives uniform (non-chimeric) transformed progeny which often contain single copy insertions and are normally hemizygous for the transgene at a given locus (Bent, 2000; Clough and Bent, 1998; Hansen and Wright, 1999). This method is simple and reliable and a reasonable number of transformed plants can be obtained from a single experiment (Clough and Bent, 1998). More recently the technique was improved on - this "floral dip" technique is less labour-intensive than the vacuum infiltration method and reduces the number of steps in the procedure (Clough and Bent, 1998). The optimum transformation efficiency determined for this technique was 0.5-3% (see Clough and Bent 1998 for detailed transformation protocol). Given its advantages this Agrobacterium transformation technique was selected to be used in this project.

6.1.2 Green fluorescent protein (GFP) - application in plant sperm cell studies

GFP fusion protein methodology is a well established procedure employed in various organisms to study protein function, localisation, interaction, and cell dynamics. Unlike conventional gene tagging systems such as the GUS reporter system (Chen and McCormick, 1996; Twell, 1992), *in situ* hybridization (Shi et al., 1996; Singh et al., 2002; Xu et al., 1999b) and immunocytochemistry (Southworth and Kwiatkowski, 1996; Southworth et al., 1999; Ye et al., 2002) which are considered to be somewhat costly, time consuming, and problematic, the GFP fusion technique has a relatively simple detection procedure and straightforward preparation protocol. Moreover the GFP reporter fusion is versatile and can be used in many forms of engineered sequences i.e. gene promoter GFP fusions, C-terminal and N-terminal translational product fusion, depending on the purpose of each individual study (Pang et al., 1996; Prasher, 1995). With all the benefits of fluorescence protein reporters, they have been widely utilised for gene expression and protein localisation studies. However, only studies involving the use of GFP in plant reproduction will be reviewed in this chapter.

The Arabidopsis *AtGEX2* gene that was first identified to be maize sperm-specific by RT-PCR was verified to be specifically expressed in Arabidopsis generative and sperm cells by a promoter-GFP fusion experiment(Engel et al., 2005). This AtGEX2::GFP transgenic plant was not only useful for gene expression and protein localisation studies but also suggested that GFP-labelled sperm could facilitate purification of the cells by fluorescence-activated cell sorting (FACS) (Engel et al., 2005). This suggestion was demonstrated to be practicable by the study reported in this research project. FACS-purified sperm cells were successfully obtained using both Hoechst 33342-stained WT sperm and stable GFP-tagged sperm cells derived from a plant line kindly provided by Dr. Scott D Russell (University of Oklahoma). The GFP-tagged sperm cells utilised in this project not only facilitated sperm isolation and purification steps but also were useful as an indicator of sperm cell viability (see detailed analysis in chapter 3). Other types of fluorescent proteins, YFP (yellow) and RFP (red) have also been used to tag generative and sperm cells thus avoiding interfering green auto-fluorescence from the pollen. The expression pattern of *DUO1*, a novel Arabidopsis R2R3 MYB gene involved in male gamete formation (Durbarry et al., 2005), was confirmed in the male germline by red fluorescent protein (Rotman et al., 2005). Translational fusion of DUO1::mRFP1 demonstrated localisation in the nucleus of generative and sperm cells (Rotman et al., 2005). The expression pattern of an Arabidopsis sperm-specific gene *HAP2* (a lily sperm-specific gene (*GCS1*) homologue), was verified by the use of a HAP2promotor::YFP construct along with RT-PCR studies. Localization of this protein within the sperm cell was demonstrated by the other YFP construct, a

HAP2protein::YFP and revealed accumulation at the plasma membrane, ER membrane, and other endomembranes i.e. golgi bodies, nuclear envelope and vesicles. This YFP-tagged protein also allowed for real-time detection of sperm cells in the pollen tube (von Besser et al., 2006).. In addition the DUO1::mRFP1 construct mentioned above was used as a sperm nuclear marker in this experiment to facilitate the *HAP2 (GCS1)* functional study (Mori et al., 2006).

As described above the GFP reporter has proven to be very useful for gene expression studies and identification of protein function and localisation. Biological processes and cell dynamics can be observed real-time and GFP-transformed cells can be detected without complicated preparation procedures. Different variants of fluorescent protein are available for more complicated analyses in which more than two proteins are involved. Therefore, together with RT-PCR analysis and reverse genetics (by Arabidopsis T-DNA insertion knockout), the GFP-translation fusion protein technique was chosen to be used in this project to identify and verify putative Arabidopsis sperm cell-expressed or perhaps sperm cell-specific genes.

6.2 Results

6.2.1 Comparative reverse transcription-PCR (RT-PCR)

The high priority Arabidopsis sperm-specific candidate genes identified by the bioinformatics approach detailed in chapter 5 were tested for sperm preferential expression by RT-PCR (reverse transcription-polymerase chain reaction). Briefly these candidates were selected following a cross-species comparative study followed by exclusion of Arabidopsis leaf and roots ESTs, identification of potential pollen-specific genes utilising Gene Atlas and manual selection of candidate genes of interest based on predicted function (see detail in chapter 5). The target sequences were subjected to RT-PCR using RNA derived from two samples, sperm cell-enriched and whole pollen samples (see section 2.1.3.2 and 2.1.4.3) for comparison, using gene specific primers (Appendix A). The sperm cell-enriched and whole pollen samples were obtained by the techniques described in chapter 3. RNA extracted from each sample was quantified by spectrometry. An equal amount of RNA was used for first strand cDNA amplification and PCR reactions were performed carefully such that the level of transcripts in both samples could be compared accurately. Moreover, confirmation that an equal amount of starting RNA was used in both samples was determined by utilising the *GAPC* gene as an RNA loading control. The level of GAPC expression in both samples was

demonstrated to be similar thus confirming that the amount of RNA template used in both samples were equal. A known sperm-expressed gene, *GEX2* (Engel et al., 2005), was also used as a positive control - more *GEX2* product was expected to be amplified in the sperm-enriched sample in comparison to the whole pollen sample. Of 20 tested genes, 7 genes were found to be preferentially expressed in the sperm cell-enriched sample (Figure 6.1). Details of the genes investigated and the RT-PCR data obtained are summarised in table 6.1.

6.2.2 Broad gene expression pattern analysis by RT-PCR of 7 genes that demonstrated preferential expression in sperm cell-enriched samples

The expression patterns of 7 genes that demonstrated preferential sperm cell expression (see previous section) were investigated by RT-PCR. RNA samples extracted from leaf, cauline leaf, flower bud, flower, and silique were used in RT-PCR reactions along with sperm cell-enriched and whole pollen RNA samples to verify the expression profile of the candidate genes (Figure 6.2). The RNA loading control, *GAPC*, was equally detected in all tissue types as expected. The sperm cell-expressed gene, *GEX2*, which was expected to be expressed in all samples containing mature pollen, was detected in pollen, sperm samples, flower and flower buds, however unexpectedly was also detected in siliques, and at low level in leaves and cauline leaves. Of 7 genes investigated, 2 genes, At3g20190 and At5g39650, demonstrated pollen specificity as the PCR products were detected only in the samples containing pollen grains and in the sperm cell-enriched sample. These RT-PCR results, from all genes investigated, corresponded to the expression patterns reported in microarray experiments (Gene Atlas, Figure 6.3). All of these sperm cell-expressed candidate genes demonstrated higher expression level in pollen than the known sperm cell-expressed genes – *GEX2* and *HAP2* (Gene Atlas). However, At4g11920 where the expression in the pollen was at 1317 demonstrated similar pattern to the *GEX2* and *HAP2* and the expression in other tissue was less than half of pollen at 569. Interestingly, in At5g39650, the expression was almost exclusively in pollen at 4730 and the other tissue was less than 100 suggesting strong potential sperm specific candidate. However, in order to be confident with the hypothesis, GFP fusion was introduced to visualise the expression of these gene *in vivo* (Figure 6.3).

Table 6.1. Summary of reverse transcription-PCR (RT-PCR) results of candidate sperm-expressed genes.

Candidate sperm-expressed genes obtained from the bioinformatics approach described in chapter 5 were investigated by RT-PCR. RNA extracted from whole pollen and sperm cell-enriched samples (see section 2.1.3.2 and 2.1.4.3) were equalised and utilised as templates. The RT-PCR products from both samples were compared. Transcripts demonstrating preferential sperm cell expression were selected for further analysis as potential sperm cell-specific genes. Of 20 genes investigated, 7 genes demonstrated elevated expression in the sperm cell-enriched fraction compared to the whole pollen fraction.

Locus tag	Description	RT-PCR
At4g32830	histone serine kinase(H3-S10 specific) / kinase/ protein serine/threonine kinase; ATAUR1 (ATAURORA1)	√
At5g23130	peptidoglycan-binding LysM domain-containing protein	×
At1g73070	leucine-rich repeat family protein	×
At3g20190	leucine-rich repeat transmembrane protein kinase, putative	√
At2g29940	ABC transporter family protein	×
At3g23610	dual specificity protein phosphatase (DsPTP1)	×
At1g78300	14-3-3 protein GF14 omega (GRF2)	×
At4g11920	WD-40 repeat family protein	√
At4g28000	AAA-type ATPase family protein	×
At5g54640	histone H2A	×
At1g11250	syntaxin, putative (SYP95)	×
At2g22740	SET domain-containing protein (SUVH6)	√
At2g19380	RNA recognition motif (RRM)-containing protein	×
At1g32400	senescence-associated family protein	×
At3g61230	LIM domain-containing protein	×
At2g20440	RabGAP/TBC domain-containing protein	×
At1g69960	serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	×
At5g39650	unknown protein	√
At1g10090	unknown protein	√
At3g50910	unknown protein	√
At5g50930	unknown protein	×

√ Genes preferentially expressed in the sperm cell-enriched sample, × Genes preferentially expressed in the whole pollen sample

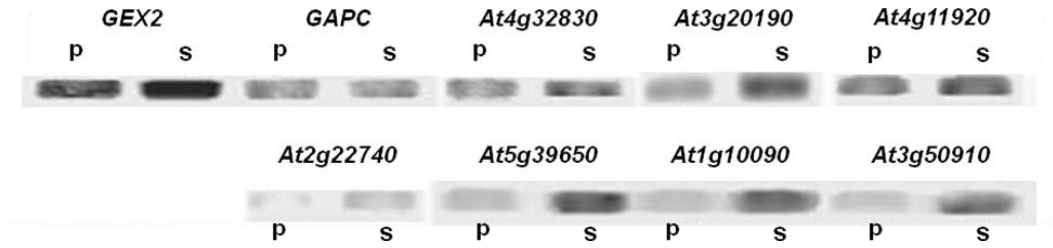


Figure 6.1. Comparative reverse transcription-PCR (RT-PCR) of candidate sperm cell-expressed genes in sperm cell-enriched and whole pollen samples.

Comparative reverse transcriptase-PCR (RT-PCR) was performed between 2 RNA samples – whole pollen and sperm cell-enriched samples (see section 2.1.3.2 and 2.1.4.3). The known sperm-expressed gene, *GEX2*, was used as a positive control where the PCR product was more abundant in the sperm cell-enriched sample than in the whole pollen sample (p). The other control, *GAPC*, was used as RNA quantification control to verify that an equal of RNA was used in both the pollen and sperm RT-PCR samples. Of 20 genes investigated by RT-PCR, 7 genes demonstrated preferential expression in the sperm cell-enriched fraction over the whole pollen sample.

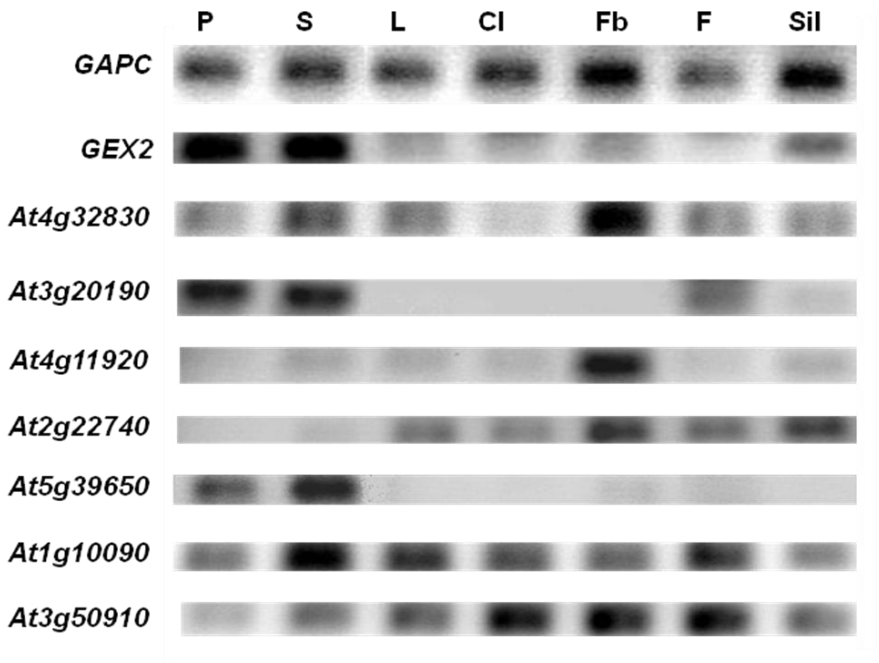


Figure 6.2. RT-PCR investigation of the expression pattern of candidate sperm cell-expressed genes across tissues types.

Putative sperm-expressed genes were investigated further by RT-PCR across different tissue types. RNA extracted from whole pollen, a sperm cell-enriched fraction, leaves, cauline leaves, flower buds, flowers and siliques was quantified and used in RT-PCR for each gene. The RNA quantification control *GAPC* and the sperm cell-expressed gene *GEX2* were used as controls. The *GEX2* transcript was expressed more highly in the sperm cell-enriched fraction in comparison to the whole pollen sample. The At4g32830 transcript was observed in all tissue types but was most abundant in flower buds. At3g20190 was expressed highly in the pollen and sperm samples and the transcript was also detected in the RNA from mature flowers. At4g11920 was expressed in all tissue types at very low level except for flower buds where transcripts were abundant. At2g22740 transcript was detected in all tissue types though the transcripts in the pollen and sperm sample were scarce in comparison to the other tissue types. At1g10090 transcript was observed at highest level in the sperm-enriched fraction and in all other tissues at similar levels. At5g39650 was expressed specifically in pollen and sperm samples. Lastly the At3g50910 transcript was detected in all tissue types. (P; pollen, S; sperm enriched, L; leaf, Cl; cauline leaf, Fb; flower bud, F; flower, Sil; siliques).

Figure 6.3. Gene expression pattern as reported from microarray data (Gene Atlas) of the genes that demonstrated preferential expression in sperm cell-enriched samples.

The expression patterns of the putative sperm-expressed gene candidates, as determined by RT-PCR, were confirmed by comparison to publicly available microarray data (Gene Atlas). The expression patterns of the RNA loading control *GAPC* and sperm cell-specific controls, *GEX2* and *HAP2*, were compared to the other genes in order to facilitate identification of sperm cell-expressed or sperm cell-specific genes. *GEX2* and *HAP2* demonstrated preferential expression in the pollen however the level of expression in the pollen was relatively low (612 for *GEX2* and 725 for *HAP2*) when compared to other genes in this study (minimum of 1317 in At4g11920). Moreover, their expression was also detected in other tissue type but at low levels (less than 200). This expression pattern was utilised as one indication for possible sperm-specific gene screening.

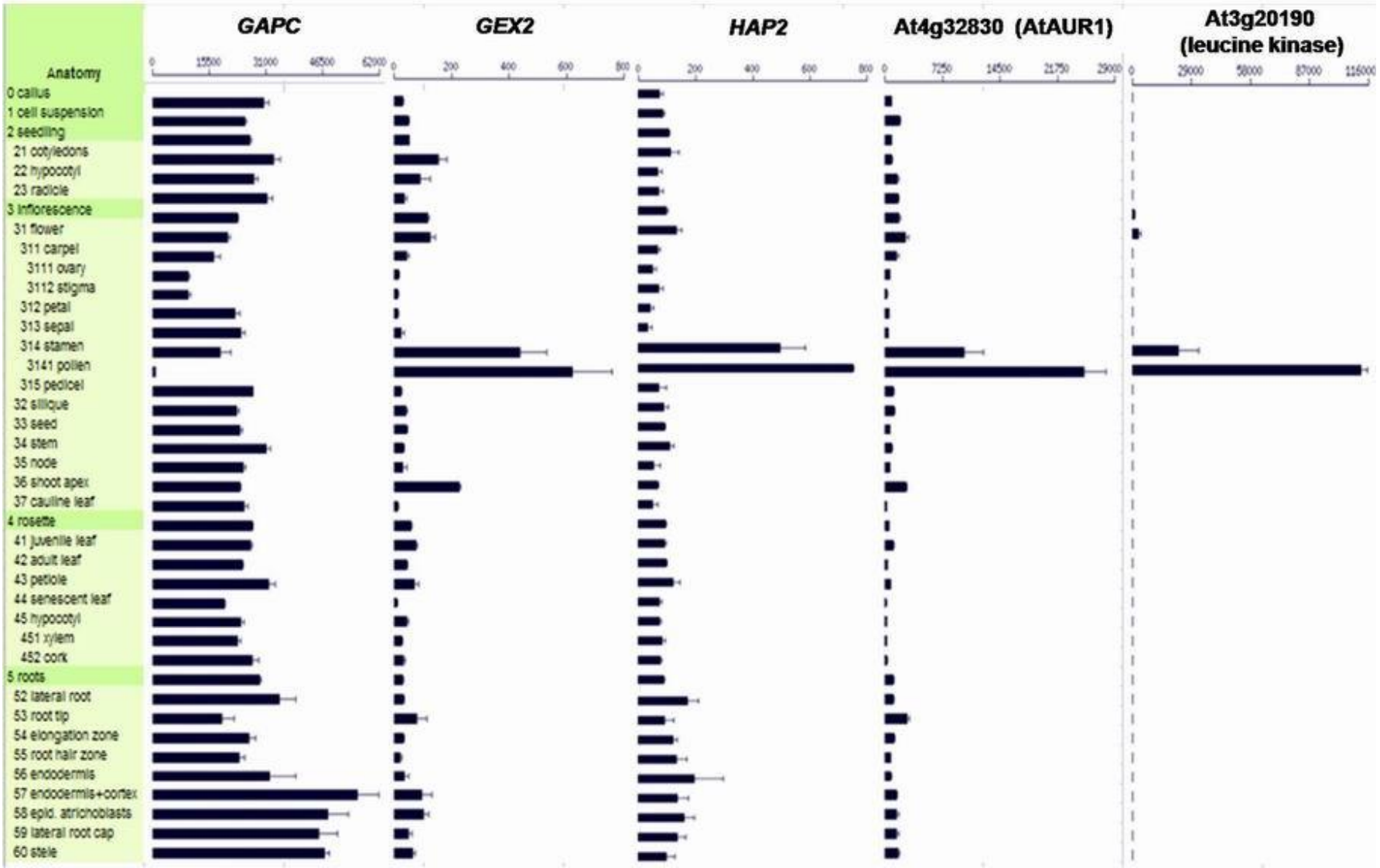
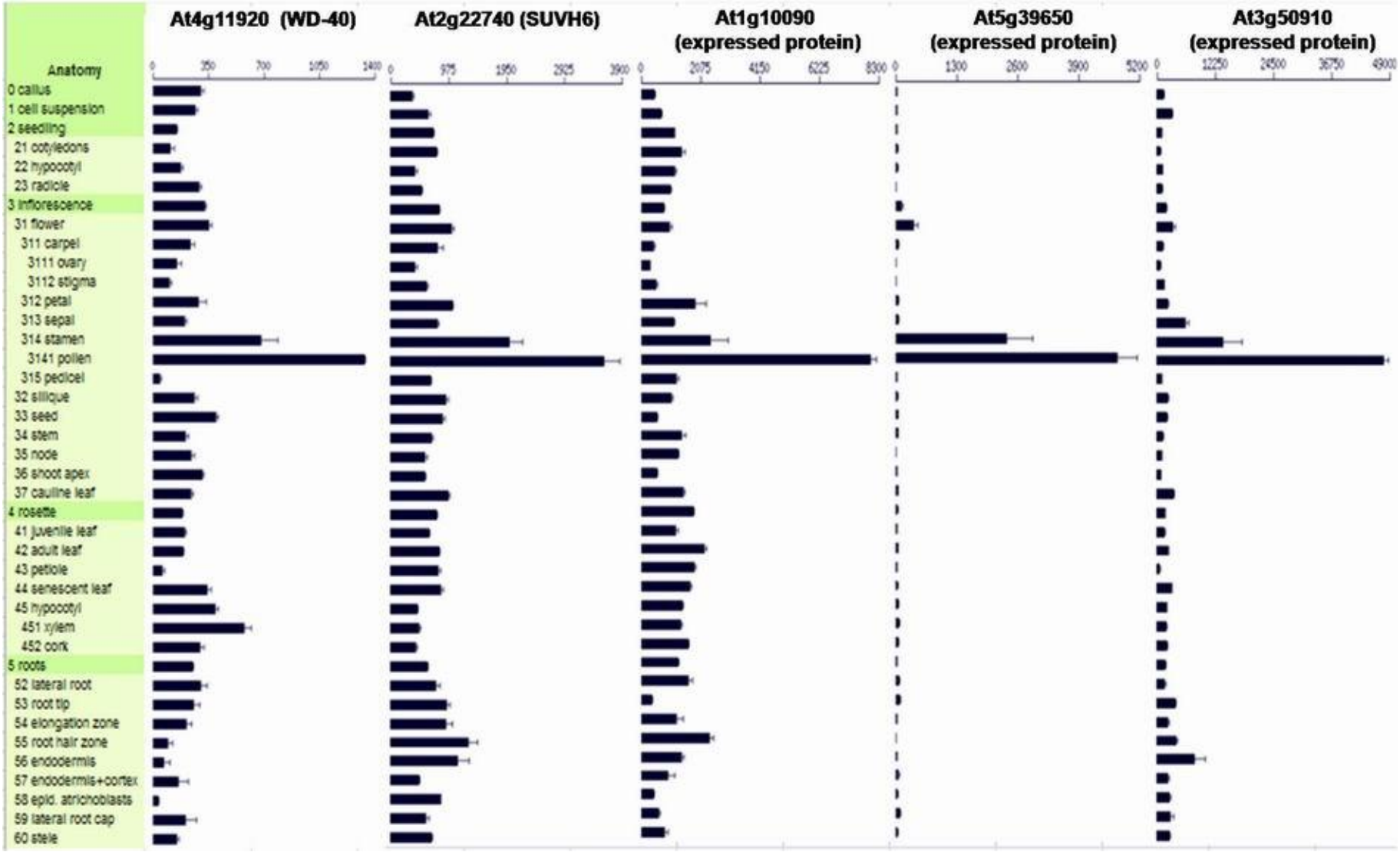

GAPC
GEX2
HAP2
At4g32830 (AtAUR1)
At3g20190
(leucine kinase)
Anatomy
0 callus
1 cell suspension
2 seedling
21 cotyledons
22 hypocotyl
23 radicle
3 inflorescence
31 flower
311 carpel
3111 ovary
3112 stigma
312 petal
313 sepal
314 stamen
3141 pollen
315 pedicel
32 silique
33 seed
34 stem
35 node
36 shoot apex
37 cauline leaf
4 rosette
41 juvenile leaf
42 adult leaf
43 petiole
44 senescent leaf
45 hypocotyl
451 xylem
452 cork
5 roots
52 lateral root
53 root tip
54 elongation zone
55 root hair zone
56 endodermis
57 endodermis+cortex
58 epid. atrichoblasts
59 lateral root cap
60 stele

Figure 6.3. Gene expression pattern as reported from microarray data (Gene Atlas) of the genes that demonstrated preferential expression in sperm cell-enriched samples (continued).


At4g11920 (WD-40)
At2g22740 (SUVH6)
At1g10090 (expressed protein)
At5g39650 (expressed protein)
At3g50910 (expressed protein)
Anatomy
0 callus
1 cell suspension
2 seedling
21 cotyledons
22 hypocotyl
23 radicle
3 inflorescence
31 flower
311 carpel
3111 ovary
3112 stigma
312 petal
313 sepal
314 stamen
3141 pollen
315 pedicel
32 silique
33 seed
34 stem
35 node
36 shoot apex
37 cauline leaf
4 rosette
41 juvenile leaf
42 adult leaf
43 petiole
44 senescent leaf
45 hypocotyl
451 xylem
452 cork
5 roots
52 lateral root
53 root tip
54 elongation zone
55 root hair zone
56 endodermis
57 endodermis+cortex
58 epid. atrichoblasts
59 lateral root cap
60 stele

6.2.3 GFP-translational fusion analyses of seven genes preferentially expressed in the sperm cell-enriched sample

Genes that demonstrated preferential expression in the sperm cell-enriched sample were studied further by GFP-translational fusion. The GFP-containing plasmid pBI-GFP having a kanamycin selection marker (Kinoshita et al., 2004) was engineered to carry the promoter and part of the coding region of the genes of interest (see Appendix D for construct map). However, the nuclear localization element (NLS) of the pBI-GFP plasmid was removed to allow an examination of the localization of the proteins of interest. These seven genes, including *GEX2* (a control indicator for successful assembly of the constructs) were amplified (Figure 6.4) using gene specific primers containing appropriate restriction enzyme sites (Appendix A) and cloned into the sequencing vector, pGEM-Teasy (Promega). The list of amplified genes and the expected product size used for cloning is detailed in table 2. The gene sequences were verified and re-cloned into the pBI-GFP vector (see construct map in Appendix C). The GFP vectors containing the genes were then transformed into *Agrobaterium tumefaciens* and then via *A. tumefaciens* into *Arabidopsis thaliana* (Col-0). Plants treated *with A. tumefaciens* were allowed to set seed and these seeds were grown on kanamycin containing MS media for selection. Transformants were able to germinate and grown on the selection media demonstrated by their green appearance amongst the seedlings without the plasmid which appeared pale green. Putative transformants were verified to contain the appropriate constructs by PCR using gene specific primers. T2 plants were checked for the presence of GFP in different tissue types. Interestingly a GFP signal could not be detected in the *GEX2* control transformants – this is possibly due to the low level of *GEX2* expression as indicated by the microarray data (Gene Atlas). The pattern of GFP expression detected in plants carrying the seven constructs were classified into 4 groups, i) sperm cell specific, ii) sperm cell expressed, and iii) pollen specific iv) pollen expressed. In the sperm cell specific class only one construct, At5g39650::GFP was identified. In the second class, sperm cell-expressed, where a GFP signal was also detected in other tissue types, At1g10090::GFP was identified. Two constructs, At4g32830::GFP and At3g20190::GFP were categorised into class 3, being pollen-specific. The other constructs were categorised into the last class where the GFP signal accumulated in pollen and other tissues. The expression profile of each gene as determined by the presence of GFP signal is described in detail in the following sections.

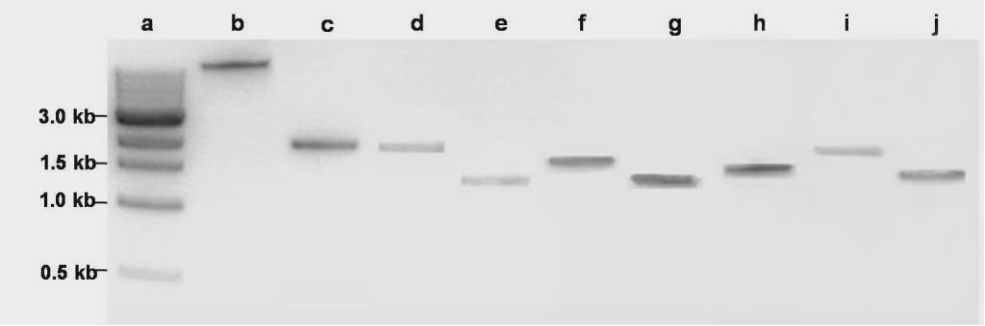


Figure 6.4. Linearised pBI-GFP plasmid and PCR products amplified from genomic DNA of the seven sperm cell-expressed candidate genes and the *GEX2* control gene.

Candidate sperm-expressed genes determined by RT-PCR to be preferentially expressed in the sperm-enriched fraction were cloned into pBI-GFP plasmid for Agrobacterium and Arabidopsis transformations. The promoter and a portion of the coding region of each gene were amplified using gene specific primers containing restriction enzyme linkers to facilitate cloning into the pBI-GFP vector. Products of digested pBI-GFP plasmid and the amplified product of each gene were demonstrated by DNA electrophoresis. (a) 1kb DNA ladder, (b) digested product of pBI-GFP, (c) 2032bp product of At3g50910, (d) 1922bp - At5g39650, (e) 1439bp - At3g20190, (f) 1886bp - *GEX2*, (g) 1462bp - At4g32830, (h) 1322bp - At4g11920, (i) 2091bp - At1g10090, (j) 1821bp - At2g22740.

Table 6.2. Summary of GFP translational fusion results for putative sperm-expressed genes.

All genes studied by translational GFP fusion are detailed by their Arabidopsis genome locus tag and gene description. The amplified PCR product size for each gene is also indicated and the GFP signal from each transformed plant line is reported as negative (no GFP detected in the transformed plants) or positive (GFP signal detected).

Locus tag	Description	Amplified product size (bp)	Pollen GFP positive
At5g49150	Gamete expressed2 (GEX2)	1886	N
At4g32830	histone serine kinase (H3-S10 specific) / kinase/ protein serine/threonine kinase ATAUR1 (ATAURORA1)	1462	Y[3]
At3g20190	leucine-rich repeat transmembrane protein kinase, putative	1439	Y[3]
At4g11920	WD-40 repeat family protein	1322	Y[4]
At2g22740	SET domain-containing protein (SUVH6)	1821	Y[4]
At5g39650	unknown protein	1922	Y[1]
At1g10090	unknown protein	2091	Y[2]
At3g50910	unknown protein	2032	Y[4]

Y: GFP positive, N: GFP negative, 1: Class1 sperm specific, 2: Class2 sperm expressed, 3: Class3 pollen specific, 4:Class4 pollen expressed

6.2.3.1 At4g32830 histone serine kinase (H3-S10 specific) / kinase/ protein serine/threonine kinase; ATAUR1 (ATAURORA1)

At4g32830::GFP expression was expected to be found in the sperm cell as RT-PCR data suggested preferential expression in the sperm cell-enriched sample. However, the GFP signal was detected specifically in pollen and not in the generative cell or sperm cells. GFP accumulated throughout all stages of pollen development in both the cytoplasm and vegetative nucleus (Figure 6.5a-c). This construct contained 1,110bp of the predicted promoter region, 12 amino acids (aa) of the first exon and 20aa of the second exon. Despite the lack of nuclear localisation signal in the construct [LOCtree (Nair and Rost, 2005); http://www.predictprotein.org/cgi/var/nair/loctree/query], this At4g32830::GFP appeared to enter the vegetative nucleus possibly by interacting with other proteins.

6.2.3.2 At3g20190 leucine-rich repeat transmembrane protein kinase, putative

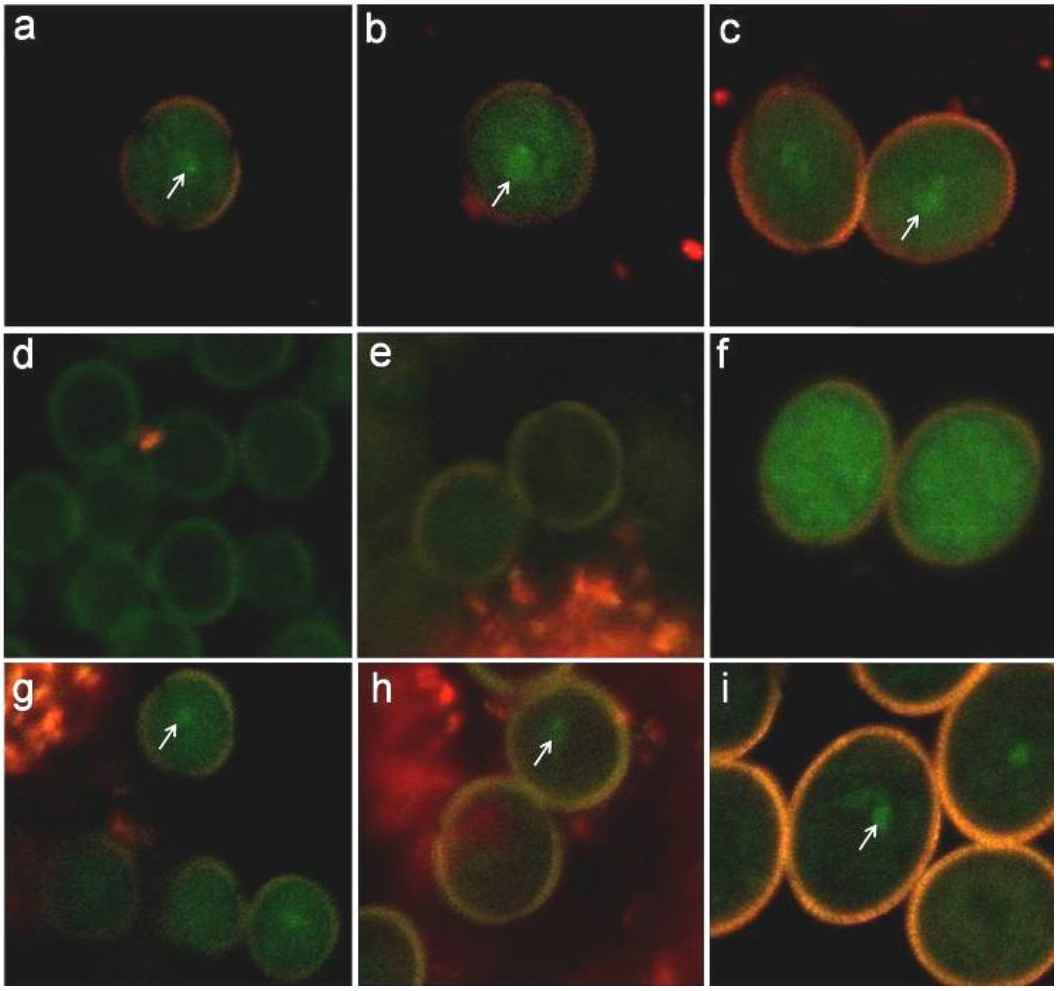
The GFP signal for this construct revealed the specificity of this gene in mature pollen. The protein was detected only in the cytoplasm of mature pollen (Figure 6.5f). Despite preferential expression of this gene transcript in the sperm cell-enriched sample as demonstrated by RT-PCR, the At3g20190::GFP accumulated in the pollen cytoplasm but not in the generative cell or sperm cells. This construct contained 1,227bp of the predicted promoter region and 67aa of the first exon. Analysis of the coding region of the construct (LOCtree) revealed no obvious secretory signal peptide sequence which could explain accumulation of GFP in the cytoplasm from a sperm expressed gene.

6.2.3.3 At5g39650 Unknown protein

At5g39650::GFP was detected exclusively in the sperm cell progenitor, the generative cell and sperm cells as predicted by the preferential expression of this gene in the sperm cell-enriched sample as revealed by RT-PCR. The signal accumulated in the nucleus of uninucleate microspore (Figure 6.5g), in the generative cell of bicellular pollen (Figure 6.5h), and in the two sperm cells of mature pollen (Figure 6.5i). This construct contained 1,497bp of the promoter region and the first 138aa of the protein which did not contain a recognisable signal peptide (LOCtree). However, this construct was predicted to contain the first two transmembrane regions of the protein out of four in total (MATDB).

Figure 6.5. Expression analysis of promoter-protein::GFP constructs for At4g32830, At3g20190, and At5g39650.

Successfully transformed plants were grown on soil. An individual plant was checked to confirm the present of the gene fused with GFP by PCR using gene specific primers. Various tissue types of the T2 plants were investigated by confocal microscopy in comparison with wild-type plant. These three constructs demonstrated expression exclusively in the male gametophyte. Three stages of pollen development, uninucleate microspore (a, d, g), bicellular pollen (b, e, h), and mature tricellular pollen (c, f, i) were investigated. At4g32830::GFP histone serine kinase (H3-S10 specific) / kinase/ protein serine/ threonine kinase (ATAURORA1) demonstrated GFP signal in all three pollen developmental stages. The expression was detected in the nucleus (a, b, c arrow) and cytoplasm of the pollen in all stages. Despite the RT-PCR result (mentioned above) suggesting preferential expression of this gene in sperm cells a GFP signal was not detected in generative cell or sperm cells. Similar to At4g32830, At3g20190::GFP leucine-rich repeat transmembrane kinase was unable to be observed in either sperm cells or the generative cell (d, e, f). This fusion protein was only detected in the cytoplasm of mature pollen. At5g39650::GFP unknown protein was the only construct detected to be expressed exclusively in generative cell (as DAPI staining could give false positive when the sample was visualised with fluorescence light microscopy, the generative cell was distinguished from the generative cell nucleus by its smaller size and position at the edge of the pollen where vegetative cell nucleus is normally positioned in the middle of the pollen) and mature sperm cells (g, h, i arrow).

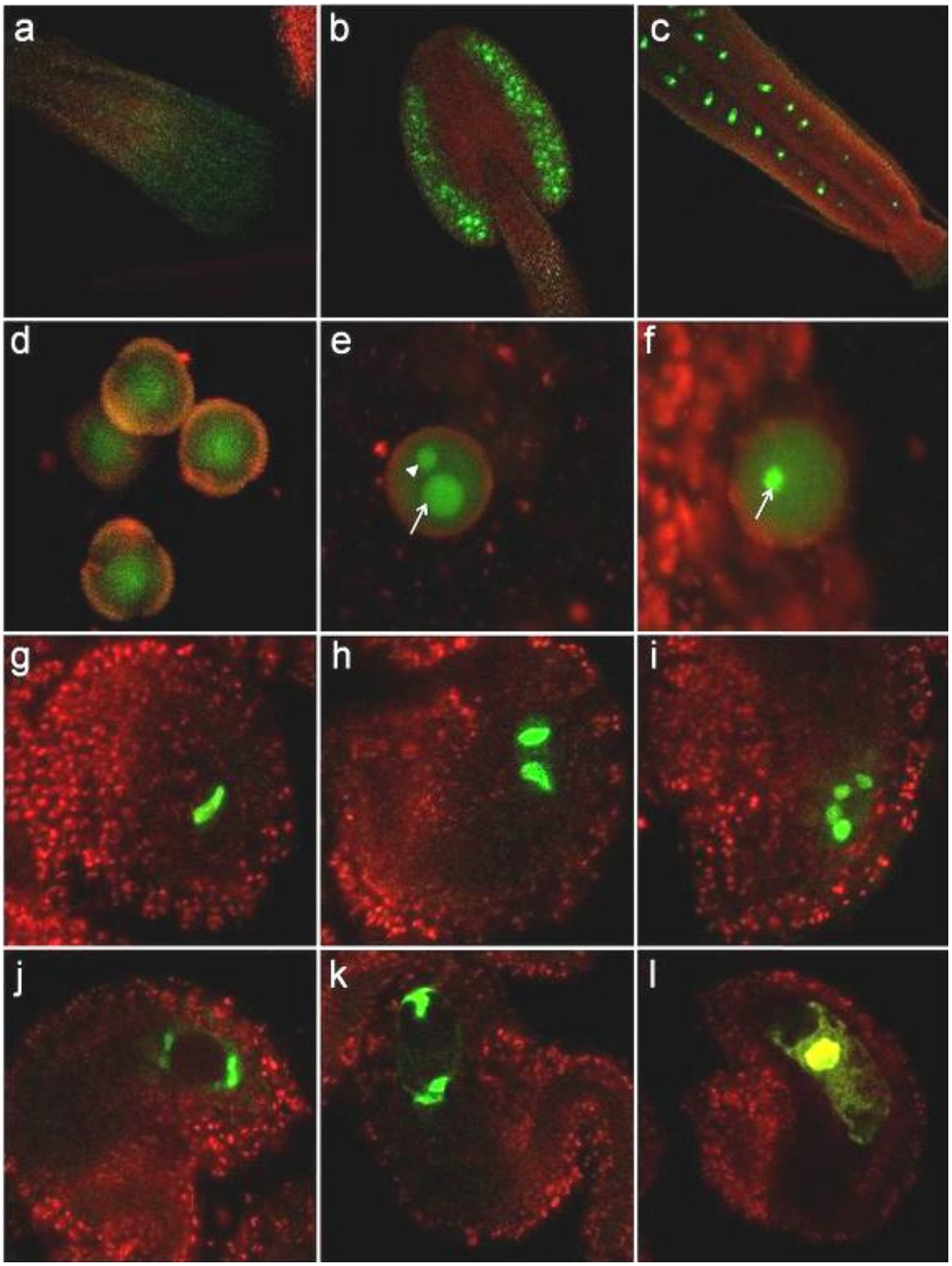

a
b
c
d
e
f
g
h
i

6.2.3.4 At2g22740 SET domain-containing protein (SUVH6)

The At2g22740::GFP signal demonstrated in this line was detected dominantly in the generative cell and pollen vegetative cell nuclei and those of the developing female gametophyte. This gene was previously reported to be localised in the nucleus and to be involved in histone methylation (Ebbs et al., 2005; Ebbs and Bender, 2006). The At2g22740::GFP in this study was first observed in the nuclei of uninucleate microspores (Figure 6.6d). In bicellular pollen (Figure 6.6e) the signal was detected in both the vegetative nucleus and generative cell nucleus. However expression appeared to switch off in the two sperm cells but remained in the vegetative cell nucleus of mature pollen grains (Figure 6.6f). In the female gametophyte the GFP signal was detected during megagametogenesis where the haploid megaspore undergoes mitosis. The GFP signal was first detected in the nucleus of the remaining megaspore (Figure 6.6g) with no expression detected before this stage (data not shown). The SUVH6::GFP protein remained present in the first (Figure 6.6h) and second mitotic divisions (Figure 6.6i). The four nuclei then migrated equally to both ends of the embryo sac and divided again (Figure 6.6j). Two nuclei from each end migrated again to the middle to form the polar nuclei (Figure 6.6k). The GFP signal began to fade away from all nuclei upon cellularisation and maturation of the female gametophyte. However the SUVH6::GFP remained in the fused polar nuclei and also appeared in the central cell cytoplasm (Figure 6.6l). The GFP signal was also detected in the petal (Figure 6.6a). This construct contained 1,585bp of the promoter region and 75aa of the first exon which contains a probable nuclear localisation signal (NLS) at amino acids 51-54. Therefore, as expected the GFP signal observed in this plant line was mainly restricted to the nucleus. However, the lack of detectable expression of this gene in the sperm cells was in contrast to RT-PCR result which suggested sperm cell expression.

Figure 6.6. Expression analysis of the At2g22740::GFP SET domain-containing protein (SUVH6) construct.

At2g22740::GFP, a SET domain-containing protein (SUVH6), was found to be expressed in petals (a), pollen (b), and unfertilised ovules (c). The signals were detected strongly in the uninucleate pollen nucleus (d, arrow), bi cellular pollen vegetative (arrow) and generative cell (arrow head) nuclei (e), and in vegetative cell nucleus of mature pollen (f, arrow). However, the signal was not detected in the two sperm cells. In the female gametophyte, the transcript was detected in the nucleus of the remaining megaspore (g). The signal persisted in the first and second mitotic division where two (h) and four nuclei (i) were produced respectively. The four nuclei migrated to both ends of the embryo sac (j) and divided again to give rise to eight nuclei. Two nuclei then migrated to the middle (k). The GFP signal began to fade in the non-migrating nuclei upon migration. The signal was detected specifically in the central cell nucleus and to a limited extent the cytoplasm of the mature female gametophyte (l).


a
b
c
d
e
f
g
h
i
j
k
l

6.2.3.5 At4g11920 WD-40 repeat family protein

At4g11920::GFP was detected in both sporophytic and gametophytic tissues. This gene was predicted to cover a wide variety of functions including adaptor/regulatory modules in signal transduction, pre-mRNA processing and cytoskeleton assembly (NCBI). The At4g11920::GFP accumulated in vascular tissue of the sepal (Figure 6.7a arrow head) and in the petal (Figure 6.7b). It was widely expressed in the carpel (Figure 6.7d) and in sporophytic tissue of the ovule (Figure 6.7e) and outer tissue layers of seed (Figure 6.7f) with no expression detected in the female gametophyte. In male gametophytic tissue, the signal was detected in the vegetative nucleus of bicellular pollen (Figure 6.7h) and that of mature pollen and in the pollen cytoplasm (Figure 6.7i) though the cytoplasmic GFP signal was at a very low level in bicellular pollen. This construct consisted of 1,014bp promoter region and 99aa of the first exon containing a probable NLS at amino acids 9-12. This GFP line demonstrated nuclear localisation as expected from the construct assembly. However, the apparent lack of gene expression in the sperm cells, as indicated by GFP, was unexpectedly demonstrated in this transgenic line.

6.2.3.6 At1g10090 unknown protein

At1g10090::GFP was observed in both sporophytic and gametophytic tissues. In sporophytic tissue it was detected in petals (Figure 6.8a), vascular tissue of sepals (Figure 6.8b), flower buds (particularly detected in the area of the sepals of young flower buds) (Figure 6.8c), anthers (Figure 6.8d), carpel (Figure 6.8e) and outer layer of embryo sac (Figure 6.8f). The GFP signal was detected in cytoplasm of uninucleate microspores (Figure 6.8g) at a very low level. Interestingly, At1g10090::GFP was detected selectively in sperm cells but not in the vegetative cell of mature pollen (Figure 6.8i). This construct contained 1,987bp promoter region and 31aa into the first exon. This region of was predicted to contain a signal peptide at amino acids 5-27 that would target the protein to the secretory pathway (MATDB). Interestingly though the GFP signal of this transformed plant line accumulated around the sperm nuclear membrane and scattered in the sperm cell cytoplasm. This pattern was detected in the sperm cells in the pollen (Figure 6.8j) and in the pollen tube (Figure 6.8k).

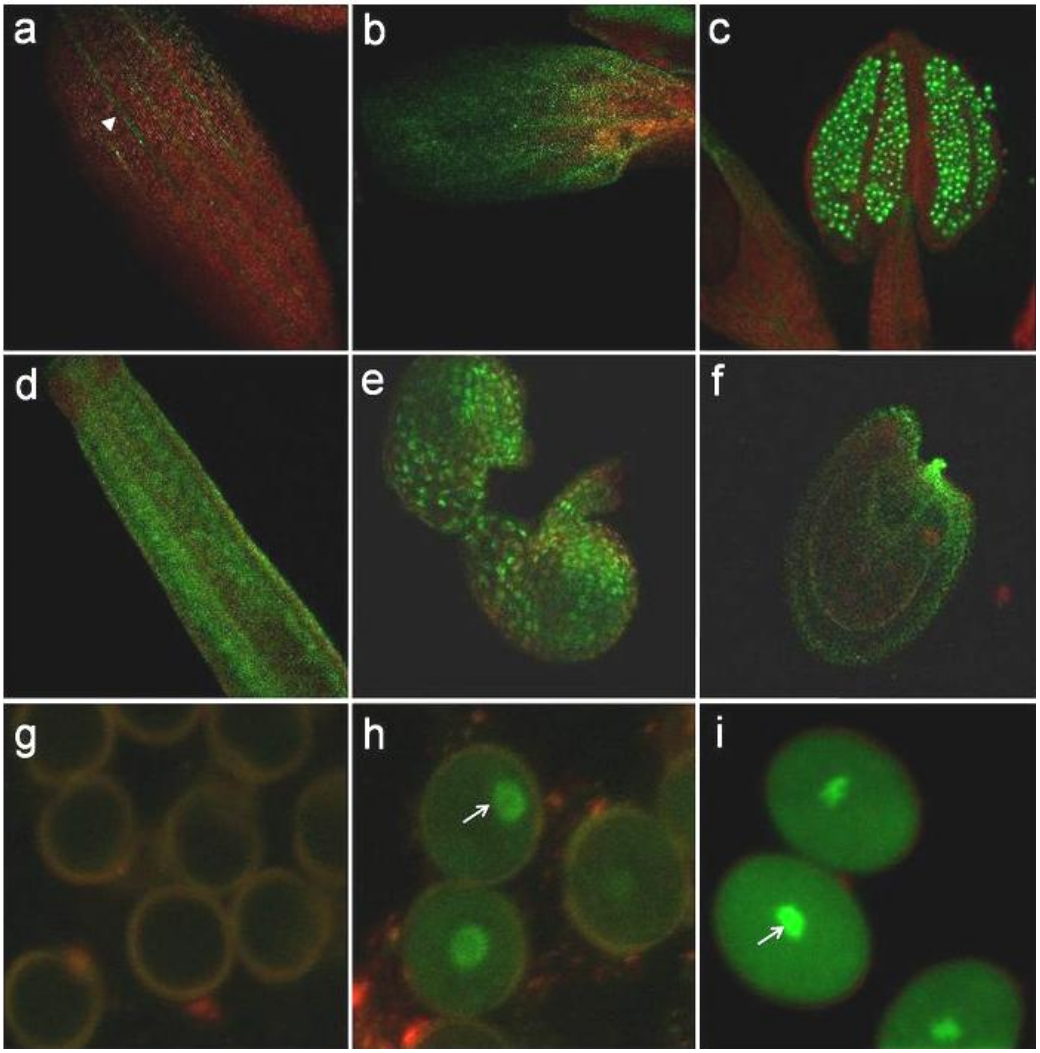
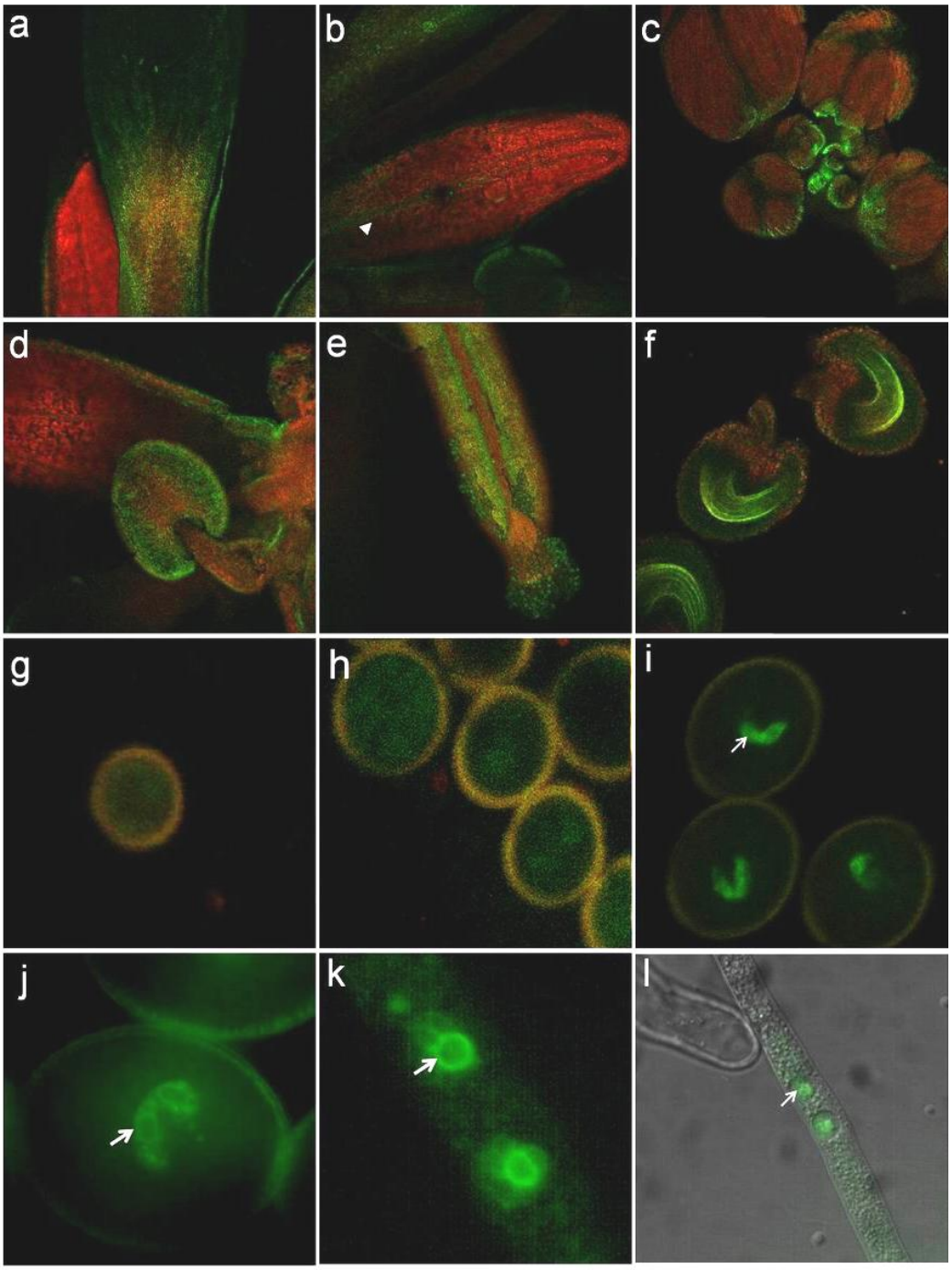


Figure 6.7. Expression analysis of the At4g11920::GFP WD-40 repeat family protein construct.

At4g11920::GFP WD-40 repeat family protein was visualised in nuclei and cytoplasm of both sporophytic (a-f) and gametophytic tissues. In sporophytic tissues, the protein was detected in the vascular tissues of sepals (a, arrow head) and petals (b), in the carpel (d), in sporophytic tissues of the unfertilised ovule (e) and outer layers of seed (f). In gametophytic tissues, the signal was unable to be detected in the embryo sac (e), but was detected in the pollen cytoplasm and vegetative nucleus (arrow) of bicellular (h) and mature pollen (i).

Figure 6.8. Expression analysis of the At1g10090::GFP (unknown protein) construct.

At1g10090::GFP signal was detected in both sporophytic and gametophytic tissues. The protein was found in petals (a), vascular tissue of sepals (b, arrow head), flower buds (c), anthers (d), the carpel (e) and outer layer of embryo sac (f). In the male gametophyte it was expressed very faintly in the cytoplasm of uninucleate microspores (g) and bicellular pollen (h). Expression was also detected in the two sperm cells of mature pollen (i, arrow). The localisation of this unknown protein was detected around the nucleus possibly in the nuclear membrane (j, k arrow) and scattered in the sperm cytoplasm (j-l).

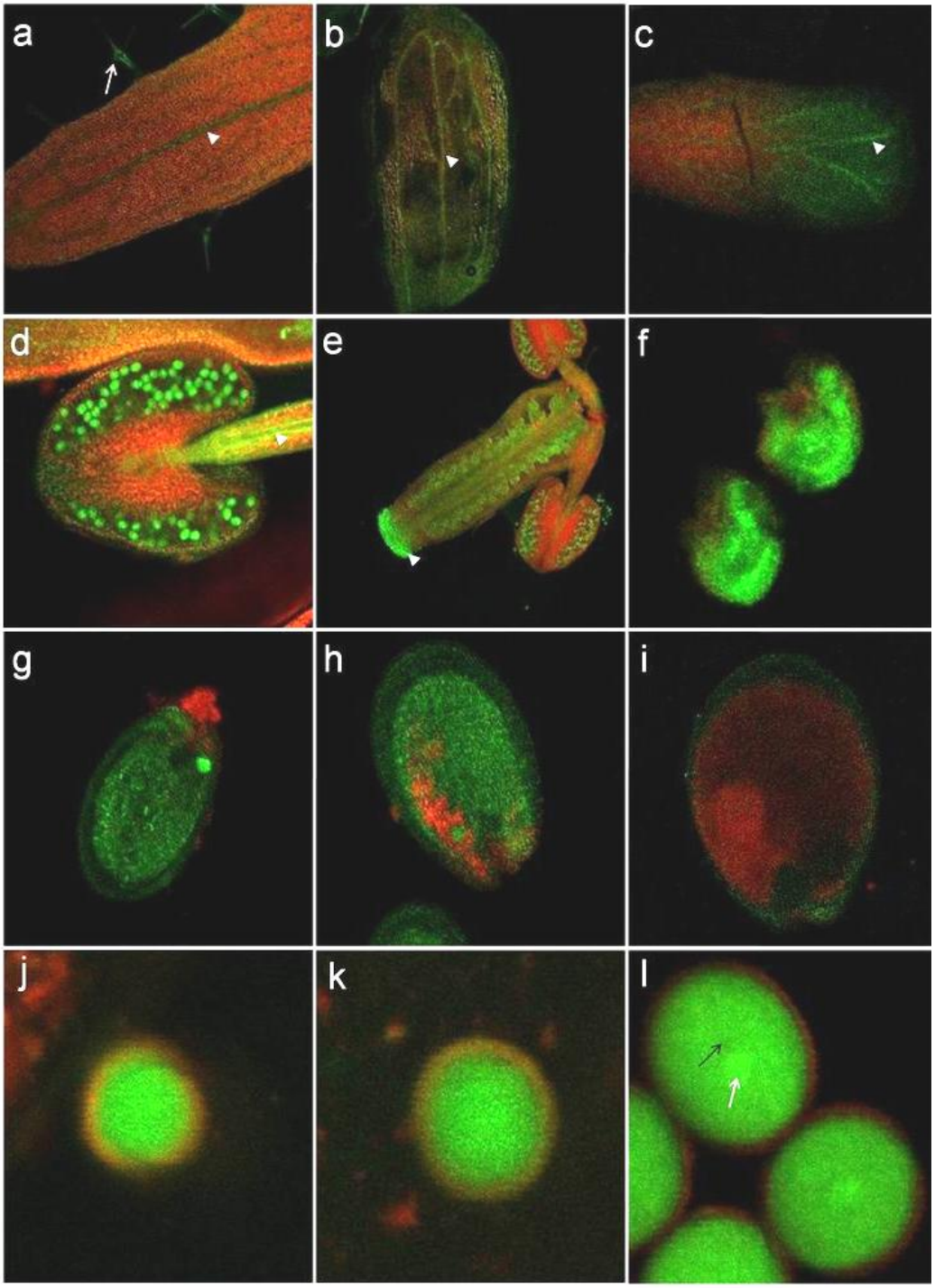

a
b
c
d
e
f
g
h
i
j
k
l

6.2.3.7 At3g50910 Unknown protein

At3g50910::GFP signal was detected in a broad range of tissue types and it accumulated extensively in all developmental stages of pollen. Expression was detected in the cytoplasm of uninucleate and bicellular stages and in both the cytoplasm and vegetative nucleus of mature pollen. This gene was also found to be expressed in the trichome and vascular tissue of leaves, sepals and petals. Moreover, expression was also detected in the stigma, embryo sac and in early seed development (Figure 6.9). This construct contained 1,553bp of the promoter region and 156aa of the first exon which contained a putative NLS at amino acids 128-131. However GFP was not detected in sperm cells possibly due to the very strong GFP signal detected in the pollen cytoplasm which may have masked weakly fluorescing sperm.

Figure 6.9. Expression analysis of the At3g50910::GFP construct.

Successfully transformed At3g50910::GFP lines were investigated by confocal microscopy in comparison to wild-type plants. The GFP fusion protein was detected throughout the plant. In leaves (a), sepals (b) and petals (c) GFP was mainly detected in the vascular tissue (arrow head) and in leaf trichomes (a - arrow). At3g50910::GFP was also detected in the filament of the anther (d - arrow head), the papillar cells of the stigma (e - arrow head) and ovule (f). The protein was also found to be expressed in the seeds, however the signal was only detected in young seeds (g [3 DAP], h [5 DAP]) and faded as the seeds matured (i [7 DAP]). GFP was detected in all pollen developmental stages mainly in cytoplasm. Nonetheless GFP signal was also detected in the vegetative nucleus of mature pollen (l – arrow). The shadow of the sperm cells can be observed in the mature pollen (l – black arrow). (j) unicellular microspore. (k) bicellular microspore.


a
b
c
d
e
f
g
h
i
j
k
l

6.2.4 RT-PCR analysis of sperm cell candidate genes utilising FACS-sorted Arabidopsis sperm cells

In order to confirm the expression of the high priority sperm cell-expressed candidate genes in sperm cells, RT-PCR (reverse transcription PCR) was performed utilizing FACS-sorted sperm cells obtained as described in chapter 3. RNA was isolated from FACS-sorted sperm cells utilising the Cell-to-Signal RNA isolation kit (ABgene). First strand cDNA was generated using the same kit with gene specific primers that had been used for the RT-PCR experiment described in section 6.2.1 (see appendix B for primer sequences). The cDNAs were then amplified again using gene specific primers. Four genes At1g10090 (unknown protein), At4g32830 (AtAUR1), At5g39650 (unknown protein) and the *GEX2* control produced the expected PCR product from transcripts derived from FACS sorted sperm cells (Figure 6.10). However the rest of the candidate gene transcripts i.e. At3g20190 (leucine-rich repeat transmembrane protein kinase), At4g11920 (WD-40 repeat family protein), At2g22740 (SUVH6) and At3g50910 (unknown protein) were unable to be amplified despite expression having been detected in sperm cell-enriched fractions by RT-PCR (see section 6.2.1). However it should be noted that the RT-PCR data obtained from FACS-sorted sperm for these genes are consistent with those obtained from the GFP translational fusion experiments described above with the exception of *AUR1* which gave a positive result from RT-PCR but could not be detected in sperm by GFP translational fusion.

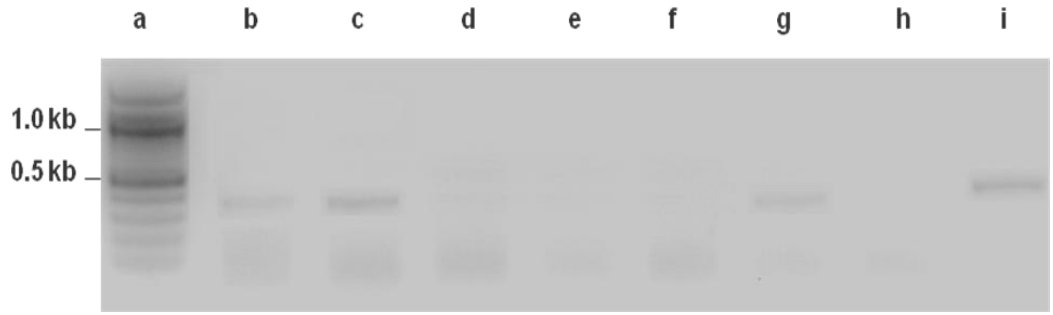


Figure 6.10. Reverse transcription-PCR (RT-PCR) analysis utilising FACS-sorted sperm cells of the seven genes selected for GFP translational fusion studies.

Reverse transcription-PCR (RT-PCR) of seven high-priority sperm cell-expressed candidates utilising RNA template from FACS-sorted sperm cells was performed in order to confirm the expression of these genes in sperm cells. Three genes At1g10090 (unknown protein) (g), At4g32830 (AtAUR1) (c), and At5g39650 (unknown protein) (i) demonstrated expression in sperm cells. The known sperm expressed gene *GEX2* was utilised as a positive contol (b). The four remaining genes i.e. At3g20190 (leucine-rich repeat transmembrane protein kinase) (f), At4g11920 (WD-40 repeat family protein) (e), At2g22740 (SUVH6) (d) and At3g50910 (unknown protein) (h) were unable to be amplified (a; 1 kb DNA ladder, b; *GEX2*, c; At4g32830, d; At2g22740, e; At4g11920, f; At3g20190, g; At1g10090, h; At3g50910, i; At5g39650).

6.2.5 Analyses of Arabidopsis unannotated gene sequences *AtNOV1* and *AtNOV2*

As sperm cell-specific transcripts are rare and might be absent from the Arabidopsis annotated gene database, maize sperm cell ESTs were also searched against unannotated regions of the Arabidopsis genome to identify potential sperm-specific candidates. Arabidopsis unannotated gene regions that matched maize sperm ESTs (obtained from bioinformatics approach described in chapter 5) were analyzed further *in silico*. Of 225 sequences that matched Arabidopsis unannotated genome regions, two sequences were selected as potential novel gene candidates (Table 6.3) and these were investigated further both *in silico* and *in vitro*. These two probable novel gene sequences (*AtNOV1* and *AtNOV2*) were analysed by gene prediction software (Softberry and Genescan) and blasted against protein database (NCBI).

Table 6.3. Maize sperm ESTs matched unannotated gene regions of *Arabidopsis thaliana*.

The *in silico* cross-species comparative study (described in chapter 5) revealed possible maize sperm cell gene homologues in unannotated gene regions of the Arabidopsis genome. These two protein sequences were identified as potential novel Arabidopsis proteins.

Name	Chro	Nt matched start	Nt matched stop	E-value	Matched protein Sequences (*Arabidopsis*)
AtNOV1	2	17231665	17231853	7.00E-10	RVRVGLYSGPPIPSWELRHGLRAKARVSCLLEWRPREAGFTEQRPPPALDGGRITGHCLPSP
AtNOV2	4	10980688	10980774	8.00E-13	FLFNNEMLQELQWFKQSYGSWFLGDYISE

Chro; Chromosome, Nt; nucleotide

Gene prediction analysis revealed multiple putative exons in the *AtNOV2* sequence region. Unfortunately, no evidence could be found to conclude that *AtNOV1* was a gene as no convincing coding region or a promoter could be identified. The *AtNOV2* sequence, including the surrounding region was predicted by Softberry (http://www.softberry.com/berry.phtml) to contain 10 coding regions encoding 275 amino acids (Figure 6.11). However, the promoter region was unable to be identified.

The *AtNOV2* region was analysed again utilising another gene prediction tool (Genescan – http://genes.mit.edu/GENSCAN.html) for confirmation and a slightly different result was obtained. Only 7 coding regions were predicted to be present in this region encoding a 177-amino acid protein (Figure 6.12). Promoter region was unable to be detected.

This predicted *AtNOV2* protein sequence (blastP) matched an 'unknown protein' of *Oryza sativa* (OSJNBa0022H21.19) with 48% identity, a score of 262 and an e-value of $2e^{-68}$. The predicted *AtNOV2* transcript was assessed by RT-PCR using RNA extracted from pollen. Primers used in the PCR were also designed to be able to test the number of exons in the gene. 3 primer pairs were designed to amplify the region between exons 1-5, 1-10, and 5-10 respectively. RT-PCR results (Figure 6.13) demonstrated pollen transcripts could indeed be amplified and that they contained at least 10 exons as predicted by the Softberry gene prediction tool.

The expression pattern of the putative *AtNOV2* gene was investigated using RT-PCR along with *GAPC* as a control. *AtNOV2* was found to be expressed in most of the tissues examined, with expression levels being quite low in pollen. Expression was readily detected in the sperm-enriched sample though at a lower level than most other tissues (Figure 6.14).

6.2.6 Reverse genetics by T-DNA insertional mutagenesis of candidate sperm genes

All 20 of the high priority gene candidates and two unannotated sequences, *AtNOV1* and *AtNOV2*, analysed in this chapter were studied further to elucidate gene function by utilising available Arabidopsis T-DNA insertion lines. Detailed study of these plant lines will be described in the next chapter though a brief overview of the T-DNA lines selected will be mentioned in the following sections.

Predicted protein(s):

>FGENESH: 1 10 exon (s) 1135 - 3437 275 aa, chain +

MSSTVWWEGAEKTRVLIAPGCGGNKPGELLTLRHPKSENGTCFLFNNEMLQEL
QWFKQSYGSWFLGDYISEDGSLYMATPVDPVFILLPIFDEARMKKGENSGKFRQ
LDEILFV EGYPGYQHLLSLAEKCMEIVCQTQEVGSMKFYRLDNSKVLAWLSCK
IYCLKNSLPE LDKNYAAQDEKQTLVDSVSIVGEYLKTEPWLKLLYDHLGLKFV
DPTMKETNMENLPTANENNMASSNSIQEKANKKPGKQTKQAKVETGSKNIRD
MFSRACKKKC

Figure 6.11. Amino acid sequence predicted from *AtNOV2* by the Softberry gene prediction tool.

The unannotated Arabidopsis sequence *AtNOV2* (that matched a maize sperm EST) and surrounding genomic region were analysed by the gene prediction tool Softberry (http://www.softberry.com/berry.phtml). A 275 amino acid long polypeptide derived from 10 coding sequences (CDS) was predicted to be translated from this sequence (the red highlighted region represents a similar amino acid prediction utilising the Genescan gene prediction tool; the blue highlighted region represents AtNOV2 sequence).

Predicted peptide sequence(s):

>12:57:13|GENSCAN_predicted_peptide_1|177_aa

RFRPDSFTSLIASLLAVVTKGENSGKFRQLDEILFVEGYPGYQHLLSLAEKCMEI
VCQTQEVGSMKFYRLDNSKVLAWLSCKIYCLKNSLPELDKNYAAQDEKQTLV
DSVSIVGEYLKTEPWLKLLYDHLGLKFVDPTMKETNMENLPTANENNMASSN
SIQEKANKKPGK QTKQAK

Figure 6.12. Amino acid sequence predicted from the *AtNOV2* genomic region by the Genescan gene prediction tool.

The unannotated Arabidopsis sequence *AtNOV2* (found to match a maize sperm EST) including its surrounding genomic region were analysed by the Genescan gene prediction tool (http://genes.mit.edu/GENSCAN.html). A 177 amino acid polypeptide derived from 7 CDS was predicted to be translated from this sequence (red highlight region represents similar amino acid prediction utilising Softberry gene prediction tool).

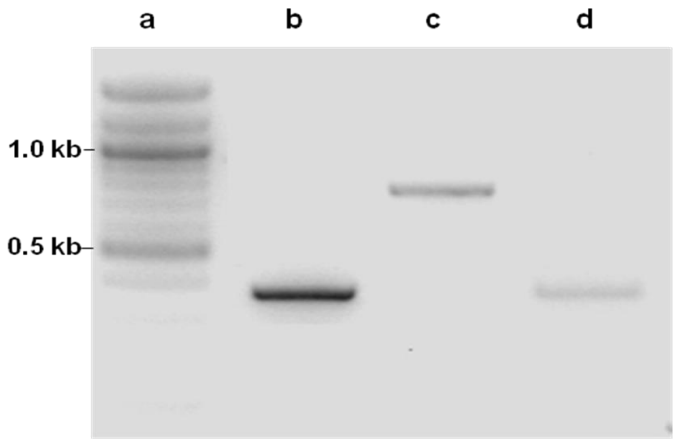


Figure 6.13. *AtNOV2* gene verification by reverse transcription-PCR (RT-PCR).
An unannotated gene region of *A. thaliana* identified as a probable sperm-expressed gene *AtNOV2* was verified by RT-PCR. Three primer pairs were designed to amplify the RNA product of this putative gene region. The three primers pairs were designed to demonstrate whether the predicted 10 exons could be transcribed. The RT-PCR result suggested that all 10 exons were active and able to produce mRNA. (a) 100bp DNA ladder. (b) 366bp product from exons 1 to 5. (c) 806bp product from exons 1 to 10. (d) 375bp product of exons 5 to 10.

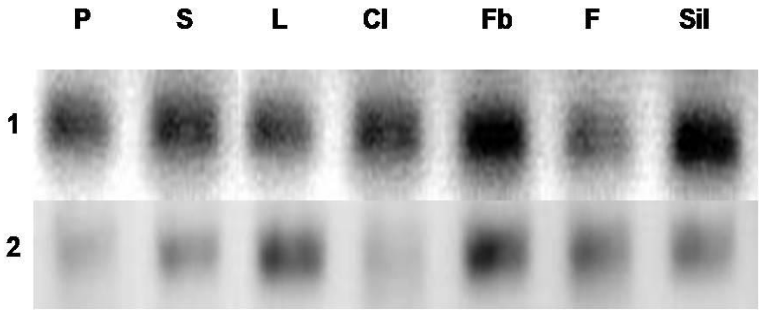


Figure 6.14. Assessment by RT-PCR of the expression pattern of the novel *A. thaliana* gene *AtNOV2*.

RT-PCR of the novel *AtNOV2* gene (2) and the control gene *GAPC* (1) was performed for various tissue types i.e. pollen, a sperm cell-enriched sample, leaves, cauline leaves, flowers, flower buds and siliques. The expression of *AtNOV2* (2) was demonstrated by the presence of a 366bp PCR product. *AtNOV2* transcripts were detected at similar levels in all tissue types tested except for the pollen,sperm fraction and cauline leaf where the expression was generally lower (P; pollen, S; sperm cell-enriched fraction, L; leaf, Cl; cauline leaf, F; flower, Fb; flower bud, Sil; silique).

6.2.6.1 Annotated genes

The 20 candidate sperm-expressed genes studied in this chapter (Table 6.1) were analysed further using reverse genetics as a tool to elucidate gene function. *Arabidopsis thaliana* plant lines with T-DNA insertions in these genes of interest were ordered from Nottingham Arabidopsis Stock Centre (NASC) (Table 6.4). However lines containing a T-DNA insertion within an open reading frame were unavailable for almost half of the candidate genes in which case insertions in 5'-, 3'- UTR, or promoter regions were utilised. Detailed analysis of these T-DNA insertion lines will be described in chapter 7.

6.2.6.2 Unannotated gene sequences – AtNOV1 and AtNOV2

The two putative gene sequences *AtNOV1* and *AtNOV2* were assessed further by examining Arabidopsis T-DNA plant lines with insertions in the predicted gene regions. The results of this analysis are summarised in table 6.4. Genes expected to be expressed in sperm cells potentially having crucial roles in reproduction could have reduced seed set phenotypes in these T-DNA insertion lines. Control plant lines in which T-DNA insertions occurred in the genes next to these 2 regions were also analysed. No phenotype was detected in the T-DNA line disrupting the *AtNOV1* region. Interestingly however a reduced seed set phenotype was found in the line containing a T-DNA in the *AtNOV2* sequence. In addition, no obvious phenotype was detected in any of the control line studied (table 6.4). These T-DNA insertion lines will be described in detail in chapter 7.

Table 6.4. Summary of T-DNA insertion plant lines investigated for each selected gene.

The high priority sperm cell-expressed gene candidates obtained from the *in silico* cross-species comparative study described in chapter 5 were investigated utilising Arabidopsis T-DNA insertion mutation lines. The available insertion lines for each gene are listed below. The location of the T-DNA insertion is also indicated - √, plant lines carrying a T-DNA in either an exon or intron; ×, plant lines carrying a T-DNA in either the promoter, 3' or 5'UTR).

Locus tag	Definition	T-DNA insertion	T-DNA in exon or intron
Annotated gene			
At4g32830	histone serine kinase (H3-S10 specific) / kinase/ protein serine/ threonine kinase ATAUR1 (ATAURORA1)	SALK_031697, SALK_112121	×
At5g23130	peptidoglycan-binding LysM domain-containing protein	SALK_111440	√
At1g73070	leucine-rich repeat family protein	SAIL_711_F10, SALK_007150, SALK_095724	√
At3g20190	leucine-rich repeat transmembrane protein kinase, putative	GABI_065H10, N612791	×
At2g29940	ABC transporter family protein	SALK_003164, SALK_003174, GK-064C11	√
At3g23610	dual specificity protein phosphatase (DsPTP1)	SALK_092811	√
At1g78300	14-3-3 protein GF14 omega (GRF2)	SAIL_904_D03	×
At4g11920	WD-40 repeat family protein	SALK_143106, SALK_073708, SALK_001978	√
At4g28000	AAA-type ATPase family protein	SALK_074465, SALK_003063, SAIL_37_F12	√
At5g54640	histone H2A	SALK_040809	√
At1g11250	syntaxin, putative (SYP95)	GK-169C03.01	×
At2g22740	SET domain-containing protein (SUVH6)	SAIL_864_E08, SAIL_1244_F04	×
At2g19380	RNA recognition motif (RRM)-containing protein	SAIL_240_D08	√
At1g32400	senescence-associated family protein	GK-505G01.01, SALK_010599	×
At3g61230	LIM domain-containing protein	SALK_024905, GT_5_109633	×
At2g20440	RabGAP/TBC domain-containing protein	SALK_006096, SALK_006098	√
At1g69960	serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	SALK_139822, SALK_013178	√
At3g50910	unknown protein	SALK_074693, SALK_132810	√
At5g39650	unknown protein	SALK_004068, SAIL_143_H01	×
At1g10090	unknown protein	SALK_131951, SALK_131877	√
At5g50930	unknown protein	SALK_009229, SALK_009951	√
Unannotated			
AtNOV1		SAIL_47_G03, SALK_137682	
At2g41310	ATRR3 ATRR3 (RESPONSE REGULATOR 3); transcription regulator	SALK_057940, SAIL_247_D08	√
At2g41330	glutaredoxin family protein	SALK_007868, SALK_014557	×
AtNOV2 (At4g20325)	unknown protein	SAIL_609_A02, SALK_015812	
At4g20320	unknown protein	SALK_127028, SALK_020074	√
At4g20330	transcription initiation factor-related	SAIL_120_A03	×

6.3 Discussion

The bioinformatics approach and gene expression pattern analysis described in chapter 5 identified a list of 20 high priority sperm cell-expressed gene candidates. These gene candidates were then checked to determine whether they were preferentially expressed in sperm cells by reverse transcription-PCR (RT-PCR). The candidates that demonstrated sperm cell preferential expression were then assessed to determine sperm specificity by GFP translational fusion studies. Not only were annotated Arabidopsis genes identified to be potentially sperm cell-expressed but some unannotated regions of the genome were also defined as potential novel Arabidopsis sperm cell-expressed genes. Each of these candidates is discussed in the following sections.

6.3.1 Identification of transcripts expressed preferentially in sperm cell- enriched fraction by reverse transcription-polymerase chain reaction (RT-PCR)

Of 20 sperm-specific candidate genes screened for expression in sperm cells by reverse transcriptase-PCR (RT-PCR) 7 genes were discovered to be potentially sperm cell-expressed. RT-PCR was performed in a comparative manner between two samples, sperm cell-enriched and whole pollen samples. Amplification of transcripts from the *GAPC* gene was used as an RNA quantification control for each sample. The known sperm expressed gene *GEX2* was used as a positive control. Transcription of a gene preferentially expressed in sperm cells was expected to be detected at a higher level in the sperm cell-enriched sample in comparison to the whole pollen sample. This sperm cell-enriched sample was prepared in a manner to eliminate most of the pollen cytoplasm and pollen wall material and to enrich for the sperm cells. However, some remnants of the pollen vegetative nucleus and some organelles of similar size to sperm cells carried through from the purification step and contaminated the sperm cell-enriched sample indicated by microscopy. The RT-PCR results of these 20 candidates demonstrated 7 genes to be preferentially expressed in the sperm cell-enriched sample (figure 6.1) and were therefore considered to be potential sperm cell-specific genes and thus subjected to further analysis GFP translational fusion experiments. The other 13 genes demonstrated preferential expression in the pollen and thus were concluded not to be sperm specific genes. However, both groups of genes were analysed by T-DNA insertional mutagenesis to identify gametophytic gene mutations and this will be described in the next chapter. The expression patterns of the 7 sperm-specific candidates, including the two control genes *GAPC* and *GEX2,* were investigated in different tissue types by RT-PCR. RNA isolated from 7 tissue types – leaf, cauline leaf,

flower, flower bud, silique, pollen and the sperm enriched fraction were used as the templates for RT-PCR. The results of these RT-PCR experiments were compared with the reported microarray data (Gene Atlas) for confirmation. All of the genes investigated demonstrated similar expression patterns between the two methods. However *GEX2* expression, which has been reported to be specific to sperm cells (Engel et al., 2005), was detected in other tissue types but at low levels according to RT-PCR and Gene Atlas. These contradictory results could be explained by different amounts of RNA template used in the RT-PCR and the different PCR conditions in the two experiments. Two genes, At3g20190 (leucine-rich repeat transmembrane protein kinase) and At5g39650 (unknown protein) were found to be exclusively expressed in the pollen as demonstrated by both RT-PCR and gene chip experiments. The expression patterns of the other 5 genes as reported by Gene Atlas and demonstrated by RT-PCR here were consistent with one another, being expressed throughout the plant but highly in pollen. These comparable results between the RT-PCR experiments and the microarray data indicated that both techniques were equally effective for tissue-specific gene identification. Thus, in the future, sperm cell-specific gene candidate selection could be greatly facilitated by screening for expression patterns similar to known sperm cell-expressed/specific genes (*GEX2* and *HAP2)* where expression in the pollen is relatively low but almost restricted to that cell type. This implication for candidate selection could be applied to the rest of the sperm cell-expressed gene candidates obtained from the cross-species comparative study described in chapter 5. A detailed analysis of the expression of the 7 sperm cell-specific candidate genes was carried out using GFP-translation fusions.

6.3.2 GFP translational fusion analysis of seven genes preferentially expressed in the sperm cell-enriched sample

GFP-tagged proteins investigated in the experiments were expected to be presented dominantly in the pollen grain or sperm cells of transformed plants. Of 8 transformed plant lines, including GEX2::GFP, 6 lines demonstrated expression of the genes in the pollen grains. Interestingly, only 2 genes were found to be expressed in sperm cells. Unexpectedly, the GFP signal was not detected in GEX2::GFP control plants which was unusual as *GEX2* had previously been shown to be expressed specifically in sperm cells (Engel et al., 2005). The GEX2::GFP construct used in this study contained the promoter region of *GEX2* designed as described in Engel *et al*, inserted in the pBI-sGFP plasmid. The gene sequence was checked by DNA sequencing and the ORF was confirmed to be in frame with the GFP reporter. However, the plasmid used in Engel et

al was pEGAD (Cutler et al., 2000), thus being different from that used in this study. As two different plasmids were used in these experiments it remains possible that pEGAD produced a better GFP signal than pBI-sGFP. In this study, the GFP signal in sperm could possibly have been too faint and masked by pollen auto fluorescence. Moreover, the expression level reported for *GEX2* (Gene Atlas) was relatively low thus supporting the speculation that the GFP signal was at an undetectable level in the present experiment.

For the other GFP lines expression was demonstrated either in the pollen vegetative nucleus or cytoplasm or both. However the GFP signal was also detected in other tissues according to the expression pattern reported by Gene Atlas and by RT-PCR. Nonetheless, the expression pattern of At4g32830 (AtAURORA1) was exclusive to the pollen whereas the RT-PCR and microarray data suggested otherwise. Again this could be because the level of expression was too low in other tissues to be readily detected using the GFP translational fusion apprach. As mentioned earlier the expression pattern reported in the microarray database (Gene Atlas) was reliable and consistent with the RT-PCR data obtained here; these GFP translational fusion experiments further confirmed the robustness of the available expression data and thus this alone could definitely be utilised for the selection of sperm cell-specific candidates. Each GFP-positive line demonstrated a unique expression pattern and thus these will be discussed individually in the following sections.

6.3.2.1 At2g22740 SET domain-containing protein (SUVH6)

Of all the transformed lines studied the GFP signal of this line appeared to be one of the strongest, indeed stronger than the signal from At4g32830::GFP (AtAURORA1) or At3g20190::GFP (leucine-rich repeat transmembrane protein kinase) where the expression levels reported by Gene Atlas were much higher. This strong signal may have been misleading as the fusion protein was nuclear localised thus concentrating GFP in the nucleus making it appear stronger than the other lines. A similar situation also occurred in At4g11920::GFP (WD-40 repeat family protein) lines where the signal was nuclear localised and appeared to be very strong contrasting to its low expression level reported by Gene Atlas.

At2g22740::GFP appeared to be expressed at a very high level in the nucleus of both female and male gametophytic cells. It was expressed at early pollen developmental stages and was restricted to the nucleus. Interestingly, this protein was present in the generative cell but was absent following division and formation of the two sperm cells.

In the female gametophyte this gene was expressed in its nuclei throughout megagametogenesis and as a result GFP expression acted as a clear marker for the process of megagametogenesis and movement of the progenitor cells. The protein was first detected in the nucleus of the remaining haploid megaspore. During the first and second mitotic divisions it was detected in both two and four nuclei stages. The GFP translational fusion clearly demonstrated that the four nuclei migrated toward both ends of an embryo sac where the final mitosis occurred. One nucleus from each end then migrated to the central region of the embryo sac upon cellularisation as described in the literature (Yadegari and Drews, 2004). Interestingly, expression of At2g22740::GFP was restricted to the central cell and central cell nucleus in the mature female gametophyte suggesting that an important function of the gene may be specific to the central cell. This gene was reported to contain a SET-domain described in the NCBI database to be identical to Histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH6, which is thought to be involved in histone methylation and is located in nucleus (Ebbs et al., 2005; Ebbs and Bender, 2006). The protein was shown to bind methylated cytosines of CG, CNG and CNN motifs but has a preference for the latter two (TAIR). The Su(var)3-9 was reported to guide CMT3 cytosine methyltransferases (MTases) to the marked sequence and to mediate non-CG methylation (Ebbs et al., 2005; Ebbs and Bender, 2006). The Arabidopsis genome encodes eight other SUVH putative H3 K9 MTases (Baumbusch et al., 2001; Ebbs and Bender, 2006). SUVH6 has similar *in vitro* H3 K9 MTase activity to SUVH4 (Ebbs and Bender, 2006; Jackson et al., 2004). Double mutation of suvh4 and suvh6 suggested that together they function in non-CG methylation (Ebbs and Bender, 2006). From the predicted function of this gene, it is speculated to be involved in chromosome packing and methylation in male gametophytic tissue i.e. pollen and generative cells (TAIR). As determined by a GFP translational fusion, At2g22740 gene expression initiates in the nucleus of the microspores and terminates at the second mitosis when the generative cell gives rise to the two sperm cells. A similar expression pattern was detected in the female gametophyte where the transcript was detected in throughout megagametogenesis but at completion remained only in the mature central cell. Despite exceedingly low expression detected in the petals, these expression patterns in the gametophytes suggested that this SUVH6 is gamete specific and functions in gamete development probably involving chromosome packing and particularly histone methylation.

6.3.2.2 At4g32830 histone serine kinase (H3-S10 specific) / kinase/ protein serine/threonine kinase; AtAUR1 (AtAURORA1)

This protein was exclusively expressed in pollen across all developmental stages in both the vegetative nucleus and cytoplasm as demonstrated by the GFP translational fusion. The protein was detected at the highest level in the binucleate microspore. However, RT-PCR had indicated expression in all tissue types. The lack of GFP signal in other tissues was possibly due to low expression levels of the gene thus resulting in levels of GFP too low to be detected by microscopical visualisation or instability of the transcript. This gene encodes a basic 339.7 kDa protein, a member of a family of Ser/Thr kinases whose activities peak during cell division (Demidov et al., 2005). In Arabidopsis, three Aurora kinases have been identified, AtAUR1, AtAUR2, and AtAUR3 (Kawabe et al., 2005; Kurihara et al., 2006). GFP fusion experiments with AtAUR1 and AtAUR2 in tobacco BY2 cells demonstrated nuclear membrane localisation in interphase with migration to the mitotic spindle and cell plates during cell division (Kawabe et al., 2005; Kurihara et al., 2006). Aurora kinases are predicted to specifically phosphorylate Ser10 of histone H3 and to co-localise with phosphorylated histone H3 during mitosis, however this function was only confirmed in AtAUR3 (Kawabe et al., 2005). As the AtAUR1 and AtAUR2 expression patterns were almost identical and detected throughout the plants being particularly high in the division zone i.e. root tip, shoot apex and pollen, the possibility that AtAUR1 is male gametophyte specific was very low. Moreover AtAUR3 also demonstrated a similar expression pattern (Gene Atlas) thus confirming the broad expression patterns of these Aurora kinases rather than indicating they are specific to the male gametophyte. However, despite expression of AtAUR1 in a range of dividing tissue types, it also demonstrated exceptionally high levels of expression in the pollen (Gene Atlas) thus suggesting the possibility that it may have a specialised function in pollen development. Based on the available information on the function and expression pattern of the Aurora kinases, AtAUR1 was concluded to be involved in cell division and to be constitutively expressed in all tissue types including sperm cells (as demonstrated by FACS-purified sperm cell RT-PCR - Figure 6.10). Despite the lack of a nuclear localisation signal (NLS) in the construct, AtAUR1::GFP was detected in the pollen vegetative nucleus as well as in the pollen cytoplasm suggested that this protein may interact with other proteins that are targeted to the nucleus. Another possibility is that this protein has undergone post translational modification permitting import into the nucleus (Jans and Hubner, 1996). More experiments are required to elucidate the function of this protein. T-DNA insertion generated mutants of this gene were investigated and demonstrated a

semi-infertility phenotype. This T-DNA insertion mutant analysis will be discussed in detail in the next chapter.

6.3.2.3 At4g11920 WD40 repeat family protein

At4g11920::GFP (WD40 repeat family protein) was found not to be specific to gametophytic tissues. GFP signal in the At4g11920::GFP line was strong in the vegetative nucleus of binucleate microspores but appeared quite faintly in the cytoplasm. The signal was detected in the both the vegetative nucleus and cytoplasm of mature pollen (Figure 6.7i). However, the transcript was not found in the generative cell nucleus or in the two sperm cells. The signal was also detected in the sporophytic tissues of the ovule that surround the embryo sac (Figure 6.7e). The expression of this gene was not only detected in gametophytic tissues but also, at a relatively low level, in other floral organs including sepals, petals, carpels and seeds (Figure 6.7a, b, d, f). At4g11920 encodes a protein containing a WD40 repeat domain which covers a wide variety of functions including adaptor/regulatory modules in signal transduction, pre-mRNA processing and cytoskeleton assembly (NCBI). This domain typically contains a GH dipeptide 11-24 residues from its N-terminus and the WD dipeptide at its C-terminus and is 40 residues long. It has been speculated that the domain coordinates interactions with other proteins and/or small ligands (NCBI). This WD40 domain-containing protein (At4g11920) is a 52 kDa basic protein (pI 9) possibly involved in the heteromeric G-protein signalling complex (TAIR). The roles of other WD40 containing proteins in plant reproduction have been reported in Arabidopsis i.e. *SLOW WALKER1 (SWA1)* (Shi et al., 2005) and *LACHESIS (LIS)* (Gross-Hardt et al., 2007). *SLOW WALKER1 (SWA1)* was identified by T-DNA insertion mutagenesis and the mutant demonstrated disruption in the progression of the mitotic division cycles of the female gametophyte (Shi et al., 2005). *SWA1* encodes a protein containing six WD40 repeats which localise to the nucleolus in interphase cells. Despite the fact that SWA1 is responsible for aspects of female gametophyte development it is found to be expressed ubiquitously throughout the plant (Shi et al., 2005). Another WD40 domain containing protein *LACHESIS (LIS)* was also found to be involved in female gametogenesis (Gross-Hardt et al., 2007). This *LACHESIS (LIS)* gene encodes a seven WD40 repeat-containing protein and the *lis* mutant plant line demonstrated a supernumerary egg cell. Moreover, RT-PCR analysis also revealed expression of *LIS* throughout the plant (Gross-Hardt et al., 2007). Interestingly, these two WD40 domain-containing proteins exhibit related functions in female gametogenesis despite their broad expression in other tissues. Moreover, a similar function was also identified in the WD40 domain-

containing protein (At4g11920) investigated in this project as the Arabidopsis T-DNA insertion lines of this gene demonstrated a semi-infertility phenotype speculated to be a result of the mutation affecting the female gametetophyte (detailed analysis can be found in chapter 7).

6.3.2.4 At3g20190 leucine-rich repeat transmembrane protein kinase, putative

Unlike the previously described GFP lines, this gene transcript was found exclusively in mature pollen cytoplasm. Both RT-PCR and microarray data (Gene Atlas) supported the expression pattern revealed by plants carrying the At3g20190::GFP construct. The GFP signal was expected to accumulate in the sperm cells as preferential expression of this gene in sperm enriched samples was demonstrated by the RT-PCR experiment. The lack of GFP signal in sperm cells could possibly be due to the strong expression level of this gene in the vegetative cell which could mask such a signal.

This gene encodes a 76 kDa protein (pI of 7.2) which contains a protein tyrosine kinase (PTK) family catalytic domain. This PTKc family is part of a larger superfamily that includes the catalytic domains of protein serine/threonine kinases, aminoglycoside phosphotransferase, choline kinase, and phosphoinositide 3-kinase (PI3K) (NCBI). PTKs catalyse the transfer of the gamma-phosphoryl group from ATP to tyrosine (tyr) residues in protein substrates (NCBI). They can be classified into receptor and non-receptor tyr kinases. PTKs play important roles in many cellular processes including, lymphocyte activation, epithelium growth and maintenance, metabolism control, organogenesis regulation, survival, proliferation, differentiation, migration, adhesion, motility, and morphogenesis (NCBI). However, evidence of gamete specific PTK expression has never been reported. Interestingly, this kinase protein was expressed exclusively in the pollen. Therefore, it may play an important role in the male gametophyte. The function of this gene was investigated utilising T-DNA insertion mutagenesis and the functional analysis will be described in the next chapter.

6.3.2.5 At1g10090 unknown protein

At1g10090::GFP was detected in the two sperm cells of mature pollen (Figure 6.8i). The signal was also observed at a low level in the uninucleate microspore and bicellular pollen cytoplasm (Figure 6.8g). The GFP signal however, was not specific to the male gametophyte but was also detected in sporophytic tissues i.e. petals (Figure 6.8a), vascular tissue of sepals (Figure 6.8b), flower buds (Figure 6.8c), anthers (Figure 6.8d),

carpels (Figure 6.8e) and the embryo sac (Figure 6.8f). This gene encodes a transmembrane protein located in the membrane or endomembrane system (TAIR). This 867.7 kDa protein is predicted to contain 3 transmembrane regions including a signal peptide targeting it to the secretory pathway (MATDB). The function of this gene is still unknown. The protein is similar to the RXW8 protein in Arabidopsis and Os12g0633600 of *Oryza sativa* (japonica cultivar-group) and contains a recognised domain of unknown function, DUF221, when blasted against the NCBI protein database. This DUF-containing family consists of hypothetical transmembrane proteins none of which have any known function (NCBI). This gene demonstrated GFP signal in sperm cells as well as other sporophytic tissues and interestingly, the localisation of GFP in the sperm cell was around the nucleus, possibly in the nuclear membrane, and scattered in the cytoplasm. More experiments are clearly required to clarify the localisation of this protein in sperm and other cell types and to identify its function. Functional analysis using a T-DNA insertion knockout line for this gene will be mentioned later in the next chapter.

6.3.2.6 At5g39650 unknown protein

This gene demonstrated expression in the nucleus of uninucleate microspores and in generative cell of bicellular pollen. In the mature pollen grain it was detected in the two sperm cells. However, due to low expression levels, the localisation of this protein at the subcellular level was not possible in At5g39650::GFP transformed plants. Weak expression of this gene was expected as the expression level obtained from microarray data (Gene Atlas) averaged 4730 whereas the other stronger sperm GFP line At1g10090::GFP was reported at 8026. In order to visualise the sublocalisation of this GFP-tagged protein in the sperm cell, overexpression of this gene, possibly by utilising other sperm cell specific promoter e.g. DUO1, could be performed in a further study.

This 26.5 kDa protein, as for At1g10090, contains a another domain of unknown function, DUF679, found in various plant proteins of unknown function (TAIR, NCBI) and is predicted to contain four transmembrane regions (MATDB). This protein is similar to 52O08_3 *Brassica napus* unknown protein (TAIR). Interestingly the At5g39650 gene was reported to be down-regulated in the *Brassica oleracea* male sterile mutant *Ms cd-1* (Kang et al., 2008). The expression of over 277 genes was suppressed in the male sterile mutant and our gene of interest was one of those categorised as a transport related gene (Kang et al., 2008). This gene is likely to be generally important for reproduction as it is conserved between the two species. The At5g39650 protein is possibly localised to the plasma membrane of sperm as it was predicted to contain

transmembrane regions and could therefore be crucial for gamete fusion. Clearly a mutagenesis approach in *Arabidopsis thaliana* is required to confirm the function of this gene. Interestingly, all T-DNA insertion lines available for this gene are predicted to have the insertion in the promoter region, none were present in exons or introns, it is possible therefore that this gene is crucial for the reproductive process and affecting both gametes, thus leading to non-transmission of a mutant allele. However, the T-DNA insertion plant lines that contain insertions in the promoter region of this gene were still investigated. Detailed analysis of these lines will be described later in chapter 7.

6.3.2.7 At3g50910 unknown protein

This gene transcript was expressed very highly in the pollen cytoplasm throughout the course of pollen development, though no signal was detected in sperm cells despite the fact that sperm expression was indicated by RT-PCR of sperm-enriched samples. A GFP signal was also detected in the vegetative nucleus of mature pollen but interestingly not in the nucleus of uninucleate microspores or bicellular pollen stages. As noted before this could be because the high GFP signal in the cytoplasm masked presence of the protein in the nucleus whereas in mature pollen the vegetative cell nucleus was more highly condensed thus giving a stronger GFP signal. RT-PCR of FACS-sorted sperm cells also demonstrated a lack of transcript in the sperm cells but this could possibly be due to instability of this transcript through the FACS process. At3g50910::GFP was also observed in various tissue types. The GFP signal was mainly detected in the vascular tissue of the leaf, sepal, petal, and anther filament. The signal was also present in female tissues, principally the stigma and ovule. In addition expression was also detected in the young seeds but disappeared as the seed developed.

This gene encodes a protein of unknown function. It has a weak similarity to A-kinase anchoring protein AKAP120 (MATDB). Although, the expression of this gene was detected in almost every tissue type, but not in sperm according to the GFP analysis, T-DNA insertion lines were still investigated to deduce function and these will be discussed in the following chapter.

6.3.3 Analyses of Arabidosis unannotated gene sequences *AtNOV1* and *AtNOV2*

Two unannotated gene regions of the *Arabidopsis thaliana* genome that matched maize sperm ESTs were analysed in detail. These two DNA sequences were speculated to be probable novel Arabidopsis genes and could possibly be sperm cell expressed genes.

The first sequence *AtNOV1*, on chromosome 2, could not be verified as a gene after analysis by various gene prediction tools. A promoter region, CDS, and poly A were not detected in this query sequence. In addition the putative encoded protein did not give a match to any plant protein in the database. On the other hand the second sequence *AtNOV2*, on chromosome 4, was identified as a gene when gene prediction tools were used. Utilising the Softberry gene prediction tool 10 coding sequences (CDS) were identified in this gene. However, this putative gene was subsequently annotated as a gene (At4g20325) in Mar 2007 (NCBI, TAIR) though was predicted to contain only 5 CDS. The prediction obtained from Softberry was considered more reliable than that from Genescan and was used in comparison with the At4g20325 gene. A comparison between the Softberry predicted amino acid sequence and the protein sequence of At4g20325 is illustrated below (Figure 6.15).

To verify the exact number of CDS RT-PCR was carried out. Primer pairs were designed to span across the different sets of introns. The first primer pair covered exons 1 to 5. The second and third pairs covered exons 5 to 10 and 1 to 10 respectively. RT-PCR demonstrated production of a transcript spanning from exon 1 to 10. The RT-PCR results confirmed that this novel gene is an active gene containing 10 exons. An update of this accession in NCBI and TAIR (27 Apr 2007) confirmed the Softberry prediction with the updated protein sequence being almost identical (2 amino acids different – due to different predicted exon-intron boundary) to the sequence predicted in this project (Figure 6.15). In addition, the expression pattern of this gene was also tested by RT-PCR in various tissue types. Transcripts were detected in all tissues tested with expression being quite low in the pollen and sperm-cell enriched fractions. It will be important to clone this transcript to confirm the predicted gene structure. A reverse genetics approach was also used to identify the function this putative gene - a T-DNA insertion knockout line in this putative gene region was investigated and will be described in detail in chapter 7.

```
Aligned sequences: 2
# 1: Softberry
# 2: At4g20325
# Matrix: EBLOSUM62
# Gap penalty: 10.0
# Extend penalty: 0.5
# Length: 277
# Identity:     275/277 (99.3%)
# Similarity:   275/277 (99.3%)
# Gaps:           2/277 ( 0.7%)
# Score: 1456.5

Softberry          1 MSSTVWWEGAEKTRVLIAP--GCGGNKPGELLTLRHPKSENGTCFLFNNE     48
                     |||||||||||||||||||  ||||||||||||||||||||||||||||||
At4g20325          1 MSSTVWWEGAEKTRVLIAPDSGCGGNKPGELLTLRHPKSENGTCFLFNNE     50

Softberry         49 MLQELQWFKQSYGSWFLGDYISEDGSLYMATPVDPVFILLPIFDEARMKK     98
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
At4g20325         51 MLQELQWFKQSYGSWFLGDYISEDGSLYMATPVDPVFILLPIFDEARMKK    100

Softberry         99 GENSGKFRQLDEILFVEGYPGYQHLLSLAEKCMEIVCQTQEVGSMKFYRL    148
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
At4g20325        101 GENSGKFRQLDEILFVEGYPGYQHLLSLAEKCMEIVCQTQEVGSMKFYRL    150

Softberry        149 DNSKVLAWLSCKIYCLKNSLPELDKNYAAQDEKQTLVDSVSIVGEYLKTE    198
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
At4g20325        151 DNSKVLAWLSCKIYCLKNSLPELDKNYAAQDEKQTLVDSVSIVGEYLKTE    200

Softberry        199 PWLKLLYDHLGLKFVDPTMKETNMENLPTANENNMASSNSIQEKANKKPG    248
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
At4g20325        201 PWLKLLYDHLGLKFVDPTMKETNMENLPTANENNMASSNSIQEKANKKPG    250

Softberry        249 KQTKQAKVETGSKNIRDMFSRACKKKC    275
                     ||||||||||||||||||||||||||||
At4g20325        251 KQTKQAKVETGSKNIRDMFSRACKKKC    277
```

Figure 6.15. Amino acid sequence alignment of AtNOV2 predicted by the Softberry gene prediction tool and At4g20325.

The AtNOV2 sequence prediction was compared to the recently annotated sequence of At4g20325 at the same locus (NCBI). An alignment demonstrated 99% similarity between the two sequences (blue highlight indicates the initial AtNOV2 sequence that corresponds to a maize sperm EST obtained from bioinformatics approached described in chapter 5.

6.3.4 General comments

These RT-PCR, promoter::GFP fusion and *in silico* gene expression analyses not only provided extensive data on the expression patterns of these high priority genes of interest but also gave valuable information to direct future studies aimed at identifying sperm cell-specific genes. Genes that demonstrated relatively high expression levels in the pollen, though expression levels not exceeding 8500 to 10000 (Gene Atlas), with either low or no expression in other tissues were hypothesised to be sperm cell-specific genes. The expression levels of sperm cell-specific genes were speculated to be low compared to other pollen expressed genes due to the small size of sperm in relation to the pollen grain (McConchie et al., 1987b). For example the sperm-specific gene *HAP2* has a reported expression level in pollen of 725 with *GEX2* having similar levels (612) whereas the pollen vegetative cell-expressed gene At3g50910 has much higher expression levels of 47569 typical of many pollen specific genes such as *PTEN1* (Gupta et al., 2002) (expression level of 64101), *TUA1* (Carpenter et al., 1992) (expression level of 124662), and *AT59* (Kulikauskas and McCormick, 1997) (expression level of 133942). Two sperm cell expressed genes – At5g39650 and At1g10090 were successfully revealed in this study, one of which (At5g39650) was a sperm-specific gene (expression level of 4730). Interestingly, these two were identified as unknown proteins with transmembrane regions and taking into account the specificity of one of them to sperm cells this may suggest it plays a specific role in fertilisation. The function of these two genes will be investigated by T-DNA insertion mutations and will be described in the next chapter. All GFP-fusion constructs were initially expected to demonstrate GFP-positive sperm cells based on comparative RT-PCR results obtained. The main reason for these unexpected results (lack of GFP signal in sperm cells) could be explained by the procedure used for sperm cell isolation and enrichment as described in chapter 3. The sperm enriched sample not only included sperm but also other pollen components i.e. some vegetative nuclei and other organelles. Some pollen vegetative nuclei were possibly concentrated with the sperm cells as they are relatively similar in size and might remain attached to the two sperm cells as the male germ unit (the male germ unit composes the connected pollen vegetative nucleus and the two sperm cells which has been reported in some species) (McConchie et al., 1987b; Russell, 1984). However, the connection between the vegetative cell nucleus and sperm cells normally disappears when sperm cells are released from the pollen grain (Russell, 1991). Therefore it seems likely that RNA from the vegetative nucleus was also present in the sperm cell-enriched sample perhaps leading to some false positive RT-PCR results. The data presented here suggest that GFP as a fusion protein acts as a reasonably good gene reporter despite the fact that its sensitivity could be improved. Therefore, a combination of different

experiments and techniques is certainly required in a gene expression study such as this to conclusively confirm the sperm cell expression of these putative sperm-expressed candidate genes.

Chapter 7:

Identification of Putative Arabidopsis Sperm Cell-Expressed Genes by Reverse Genetics: T-DNA Insertion Mutagenesis Screening

7.1 Introduction

In previous chapters a combination of proteomics, bioinformatics, transcriptional analyses and GFP fusion techniques have been utilised in order to identify Arabidopsis sperm cell-expressed genes. Another approach taken for the identification of sperm cell-expressed genes which, described in this chapter, is reverse genetics i.e. T-DNA insertion mutagenesis. T-DNA mutant screening is generally considered to be relatively straightforward and concise for the functional analysis of particular genes in comparison to other methodologies e.g. abolition of specific transcripts using RNAi or over-expression or ectopic expression of genes using the 35S promoter. Moreover, these T-DNA insertion plant lines are available from the Nottingham Arabidopsis Stock Centre (NASC) for immediate analysis. Initial screening for sperm cell-expressed genes predicted to be crucial for reproduction process was carried out on a range of mutation lines available for our sperm-expressed gene candidates previously identified by the bioinformatics approach described in chapter 5. The screen involved identifying lines that had a semi-infertility phenotype, that is, lines that had reduced seed set compared to wild-type. In parallel segregation distortion assays were performed to identify mutations in gametophytic genes. In this chapter, the phenotypic analyses of these T-DNA insertion knockout lines and segregation distortion analyses for the sperm cell-expressed gene candidates will be described and discussed. Finally a detailed analysis of the plant lines identified as having mutations likely

to be affecting the gametes will be presented and discussed. The proposed functions of these genes in plant reproduction and their involvement in sperm cell biology will also be considered in this chapter.

7.1.1 Background to T-DNA insertion mutagenesis studies and approaches to identifying gametophytic genes

Arabidopsis thaliana is an established model organism because of its small genome size, compact plant structure and short life cycle. It is also self-fertilising and produces a large number of seeds, as much as 10,000 seed per plant. Moreover, Arabidopsis is a typical flowering plant in its morphology, anatomy, growth, development and environmental responses making it an ideal model plant (Azpiroz-Leehan and Feldmann, 1997; Harris, 2004). The Arabidopsis genome project started in 1990 and was completed in 2000 (Harris, 2004). With the availability of the entire Arabidopsis genome sequence, one of the next challenges is to uncover the functions of the ~25,000 genes it encodes. A foreign traceable mutagen DNA (transfer DNA: T-DNA) was randomly integrated, via Agrobacterium, into the genome of Arabidopsis to generate gene knockout mutant plant lines (Azpiroz-Leehan and Feldmann, 1997). Many collections of T-DNA mutagenised plants e.g. SALK, SAIL, GABI-Kat, FLAG, Wisc, CSHL and RIKEN have been generated in order to create the greatest possible number of gene-knockout mutations in Arabidopsis. These collections of mutagenised plants are invaluable in gene functional studies. The details for these T-DNA lines along with additional data relevant to insertion point are available via a web interface, T-DNA Express (http://signal.SALK.edu/cgi-bin/tdnaexpress). Seeds for these T-DNA insertion lines can be obtained from the Nottingham Arabidopsis Stock Centre (NASC).

Gametophytic gene mutations can be difficult to be identify particularly for those that affect the male gametophyte as frequently homozygous plants are unable to be recovered and the phenotype of heterozygous mutant plants often appears at first glance to be wild-type (Procissi et al., 2001). However, in 1997, Feldmann and colleagues utilized a selection marker in a segregation distortion assay for the screening of gametophytic gene mutations (Azpiroz-Leehan and Feldmann, 1997). This assay has been used to differentiate sporophytic and gametophytic mutations in the T-DNA knockout lines. The assay involves segregation ratio screening of the T-DNA insertion which co-segregates with the selectable marker (Kan^R). Sporophytic genes tagged with Kan^R segregate in the normal Mendelian manner where the selectable marker segregates into the ratio $3Kan^R$: $1Kan^S$ (Figure 7.1). However, in some cases where the marker fails to transmit through to the next generation due to a mutation in a gametophytic gene

the ratio of Kan^R:Kan^S plants will distort to 1:1 when either the male or female gamete is functionally ablated (Figure 7.2) (Feldmann et al., 1997; Procissi et al., 2001). Gametophytic mutations can also be investigated further by reciprocal crosses with wild-type plants in order to determine whether T-DNA transmission was reduced through the male and/or female gametes. If the T-DNA was transmitted normally seed from reciprocal backcrosses between heterozygous and wild-type individuals would be $1Kan^R$:$1Kan^S$ (Howden et al., 1998).

7.1.2 Gametophytic genes identified by T-DNA insertion mutagenesis

Arabidopsis T-DNA insertion mutagenesis is a powerful tool to identify novel genes that are crucial for plant growth and development and to investigate the function of genes of interest. It has also proven to be effective in the identification of male and female gametophytic genes with numerous examples having been described to date. A number of important examples are covered in the following sections

7.1.2.1 Female gametophytic gene mutations

Female gametophytic mutations are relatively easy to identify as ovules carrying the mutant allele will not develop or produce seeds and thus present with a reduced seed set phenotype and consequently shorter siliques. As this project's main focus is the identification of genes expressed in the male gamete, only a few examples of female gametophytic gene mutations will be described in this chapter (other examples of female gametophytic genes have been covered in chapter1).

Development of the female gametophyte can be affected by mutations affecting the surrounding sporophytic cells of the ovule (Drews et al., 1998). Therefore, both reduced seed set screening and segregation distortion assays are required for the identification of female gametophytic mutations (Drews et al., 1998). Many such mutants have been isolated (see review in Drews, et al., 1998). Mutations affecting female gametogenesis and post fertilisation processes have been commonly isolated and reported. However mutants that demonstrate defects in the gamete fusion event have not yet been isolated. Nevertheless, Pagnussat and colleagues (2005) claimed that 18 mutant lines (UNE1-UNE18) were isolated in their study that were defective in reproduction (Pagnussat et al., 2005). The proposed function of many of these genes involved signalling between the ovule and the pollen tube (Pagnussat et al., 2005). Of these 18 mutants 12 were found to produce normal pollen tubes that were able to enter the micropyle, however the

authors suggest that failure of synergid cell death might have occurred in these mutants (Pagnussat et al., 2005). It remains possible that some of these mutants could have affected the actual fusion event between the two parental gametes. Perhaps one explanation for the difficulty in isolating female gametophytic genes that are crucial for the gamete fusion event is the possibility that genes involved are expressed in both male and female gametes. If this were true then mutant lines could not be isolated in either the homozygous or heterozygous states.

7.2.1.2 Male gametophytic gene mutations

Gametophytic mutations affecting the male gamete are comparatively rare mainly because they do not necessarily have a directly observable phenotype (Procissi et al., 2001; Yang and Sundaresan, 2000). In addition pollen grains are produced in great numbers and pollen carrying a wild-type allele can out-compete pollen carrying a mutant allele thus resulting in normal seed set. Screens for male gametophytic mutations often start with segregation distortion assays to minimize the number of candidate genes followed by phenotypic analyses in which the expected male phenotypes are carefully searched for. Despite the fact that these mutations are difficult to identify a number of mutations that affect male gametophytic genes have been discovered utilising this reverse genetics methodology. The first mutant discovered that affected pollen development was *pollen abortion factor* (Howden et al., 1998; Redei, 1964) and the second was *sidecar pollen (scp)* (Chen and McCormick, 1996). Other mutants affecting pollen development discovered by T-DNA insertion knockout screening include *limpet pollen (lip)* (Howden et al., 1998), *gemini1 (gem1)* (Park et al., 1998), *male sterile mutant (ms)* (Taylor et al., 1998), *male sterility1 (ms1)* (Wilson et al., 2001), *halfman (ham)* (Oh et al., 2003), and *duo pollen (duo)* (Durbarry et al., 2005). Another class of male gametophyte mutants affect pollen germination and include *no pollen germination1 (npg1)* (Golovkin and Reddy, 2003), *seth1, seth2* (Lalanne et al., 2004), and *sec8* (Cole et al., 2005) with a related group affecting pollen tube growth, e.g. *mad4* (Grini et al., 1999), *T-DNA transmission defect (Ttd34, 42)* (Procissi et al., 2001), *and vanguard1 (vgd1)* (Jiang et al., 2005). These collections of gametophytic mutants demonstrate the potential of reverse genetic screening for the identification of gametophytic genes in Arabidopsis. Thus, an attempt to isolate sperm-specific genes could be achieved using this powerful methodology. However, as for the female gametophytic gene mutations, none of these male mutants were defective in the final stage of reproduction, the gamete fusion event. Nonetheless, in a recent study an Arabidopsis sperm cell-specific protein *GCS1 (HAP2)* (initially identified in lily generative cell cDNA library) was revealed to have a role in the gamete fusion event

following analysis of a T-DNA insertion line (Mori et al., 2006). Therefore, attempts to characterise novel sperm-cell genes via this method are practicable.

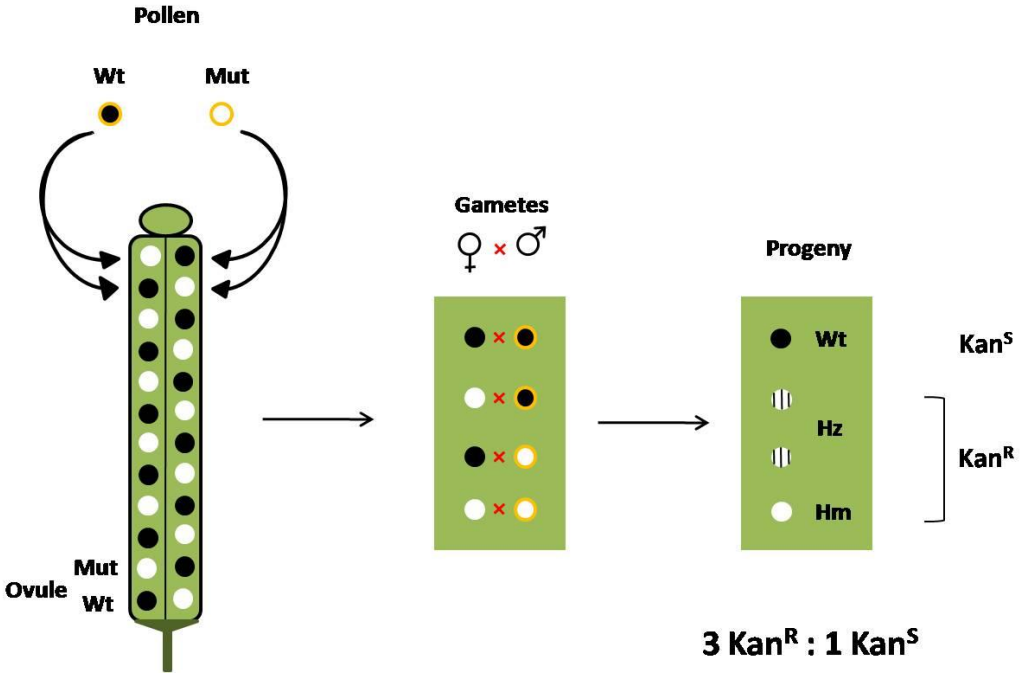


Figure 7.1. Segregation ratio for a T-DNA induced sporophytic gene mutation.

Heterozygous plants produce gametes where half carry the wild-type allele (black circle) and the other half carry the mutant allele (white circle) containing a Kan^R selection marker on the T-DNA insertion. For sporophytic gene mutations, both male (yellow line circle) and female (no line circle) gametes carrying either mutant or wild type alleles function normally. The progeny of such a self-pollinated heterozygous plant (white circle with black stripe) can be selected on media containing the antibiotic kanamycin (Kan). Progeny containing at least 1 mutant allele from either parent are kanamycin resistant (Kan^R) whereas individuals lacking a mutant allele are kanamycin sensitive (Kan^S); thus the ratio would appear $3Kan^R$:$1Kan^S$.

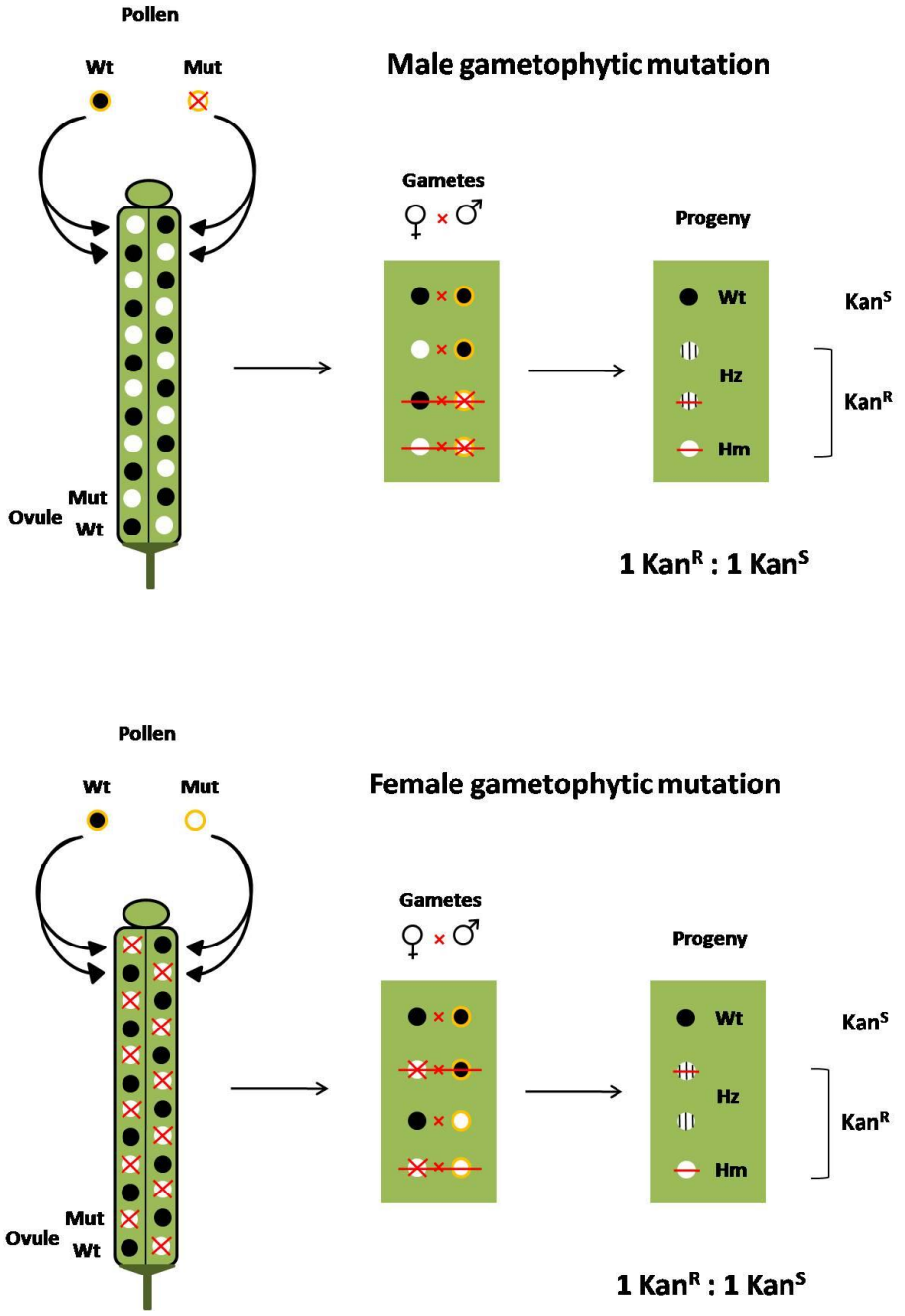


Figure 7.2. Segregation distortion for T-DNA induced gametophytic gene mutation.

A heterozygous plant produces half of its gametes carrying a wild-type allele (black circle) and half carrying a mutant allele (white circle) which carries a Kan^R selection marker. However, for a male gametophytic gene mutation, (a), the defective male gamete (crossed yellow bordered white circle) would not be able to successfully fertilise the female gamete and likewise for a mutation affecting the female gamete (crossed white circle) (b). Thus only half of the progeny could grow on the kanamycin selection plate thus producing a $1Kan^R$:$1Kan^S$ ratio.

7.2 Results

Sperm protein candidates derived from the bioinformatics approach described in chapter 5 were analysed further by phenotypic and genotypic screening of T-DNA insertion knockout lines. T-DNA insertion plant lines for the genes of interest were ordered from NASC and grown for phenotypic analysis. Segregation distortion assays were also performed in order to identify lines carrying gametophytic mutations. The first phenotypic screen was aimed at identifying plants with a semi-infertility phenotype (reduced seed set). Identifying genes that are crucial for sperm cell function, particularly in gamete fusion, was the first priority and therefore the screen was aiming for mutations affecting this step. Mutants in this category should give only 50% seed set in the heterozygous condition as half of the pollen will carry non-functional sperm cells. Moreover, homozygous plants should not be recovered as mutant sperm would fail to fuse to the female gametes resulting in no seed production. However, this initial phenotypic screen detected only female gametophytic mutations or mutations affecting both the male and female gametophytes. In any case none of the plant lines demonstrated the hypothesised phenotype. At this point, the potential effect of mutant sperm on pollen tube growth was taken into consideration as wild-type pollen could feasibly out-compete mutant pollen. Therefore a segregation distortion assay was utilised, an approach that has successfully identified male gametophytic mutations particularly those affecting pollen development, germination, and tube growth (Feldmann et al., 1997; Procissi et al., 2001). Interestingly, in this study a range of both gametophytic and sporophytic genes involved was identified with roles in gametogenesis and pre and post-fertilisation processes. A summary of the procedures carried out to identify sperm cell-expressed genes utilising T-DNA insertion mutagenesis and the results is presented in the flowchart below (Figure 7.3).

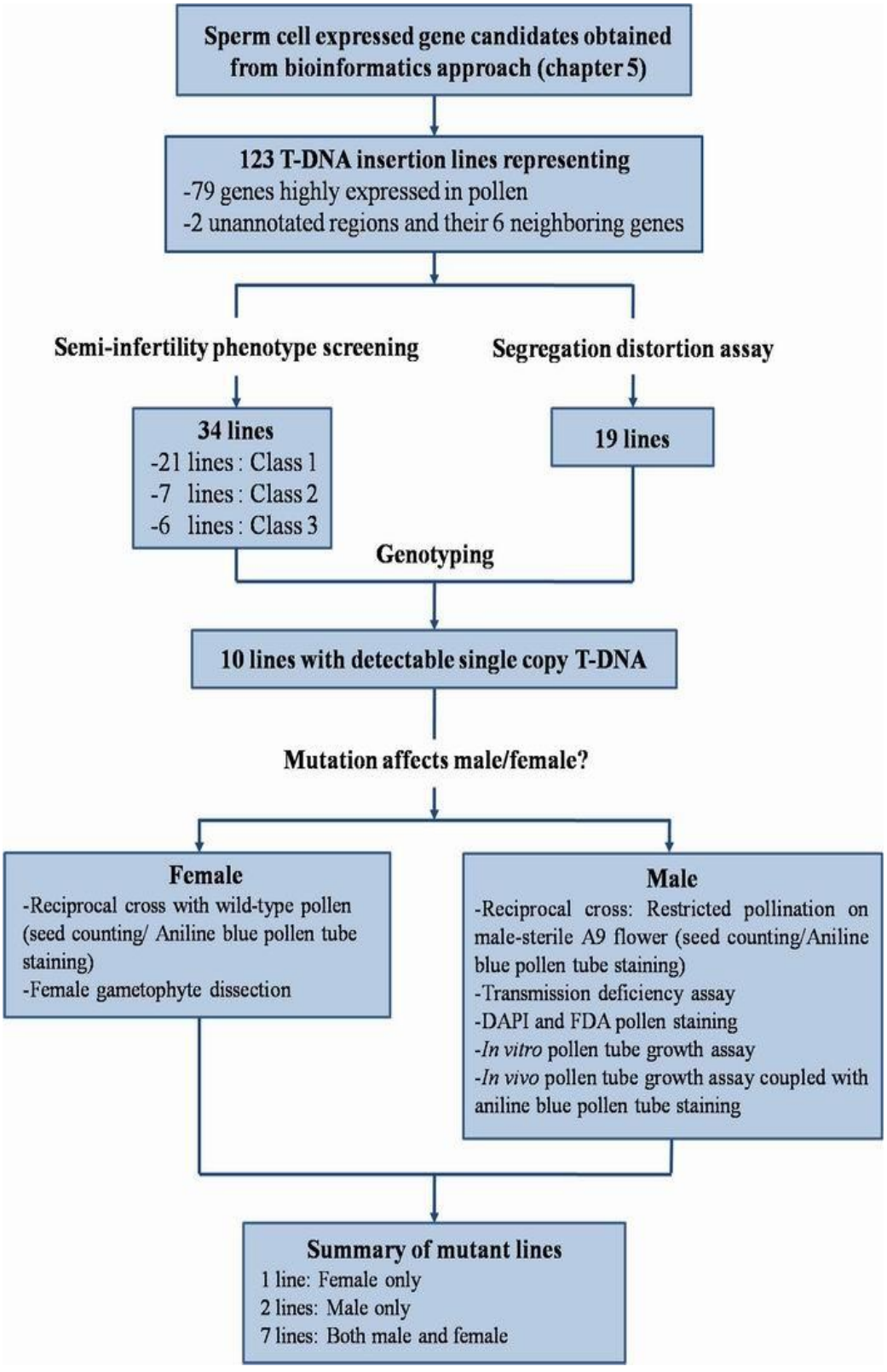


Figure 7.3. Diagram demonstrating T-DNA screening procedures to identify gametophytic gene mutations in this project.

7.2.1 Initial screen for potential sperm cell-expressed genes by semi-infertility (reduced seed set) phenotypic analysis

A. thaliana T-DNA insertion lines for the genes of interest were grown under controlled conditions. Approximately 25 plants from each line were grown for phenotypic analysis (Table 7.1). An average of 10 siliques from each plant was opened to observe seed set. Out of 123 screened lines (from 79 genes), 34 were found to have a reduced seed set phenotype. The reduced seed set phenotypes can be grouped into 3 classes. The first was identified as 'unfertilised ovules' (Figure 7.4a) as no evidence of seed development could be identified. Importantly the cause of this phenotype could be that ovules were fertilised but seeds were unable to develop from a very early stage. Thus the mutant lines in this class were analysed in detail and will be described in section 7.2.4. Most of the reduced seed set plant lines (21 lines) were classified into this category. The second class of seed set phenotype was named 'aborted seed' (Figure 7.4b). In this class the ovules were fertilised and the seeds were able initiate development but then arrested after a certain period and became brown and shrunken. Only 7 lines were categorised into this class. The third class was called 'minute silique' (Figure 7.4c) due to the small number of developing seeds in the siliques. In this class typically less than 10 seeds per silique were observed subsequently affecting the size of the siliques. This class could also be presented as a more severe form of the 'unfertilised ovule' class. Of 31 lines with the reduced seed set phenotype, 6 lines were placed into this class.

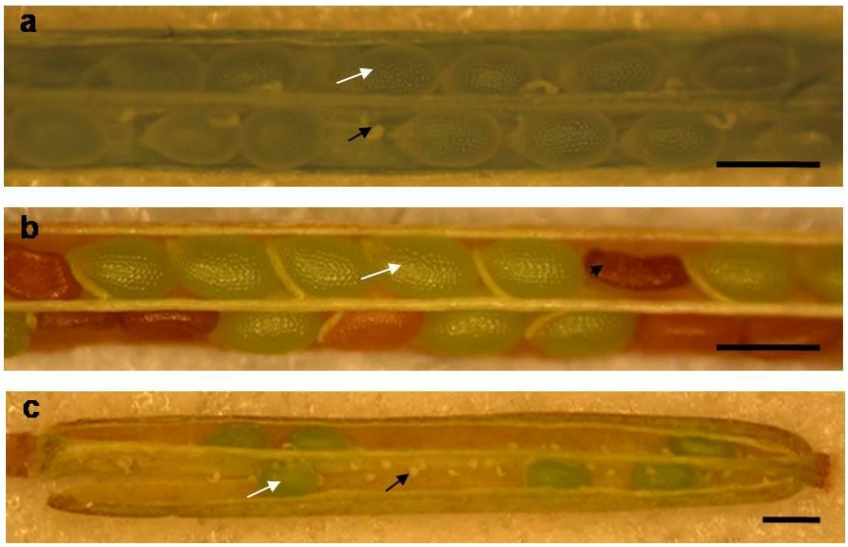


Figure 7.4. Siliques demonstrating different classes of the reduced seed set phenotype.

Reduced seed set phenotypes were demonstrated in the opened siliques of 34 plant lines out of 123 suggesting possible gametophytic mutations. The reduced seed set phenotypes can be grouped into 3 classes. a) Unfertilized ovule: this class represented the phenotype in which approximately half of the seeds were undeveloped and thought to be the result of an unfertilised ovule. These 'unfertilised' ovules appeared as white dots (black arrow) amongst fully developing green seeds (white arrow). b) Aborted seed class: this phenotype occurred after the fertilisation processes, therefore was identified as a sporophytic mutation. The brown shrunken seeds (black arrow head) amongst fully enlarged green seeds (white arrow) were aborted due to an arrest seed development. c) Minute siliques: the reduced seed set phenotype in this class represents as more severe case of the 'unfertilised ovule' class. Only a few fully developed seeds (white arrow) appeared in the opened siliques (bar = 0.5mm).

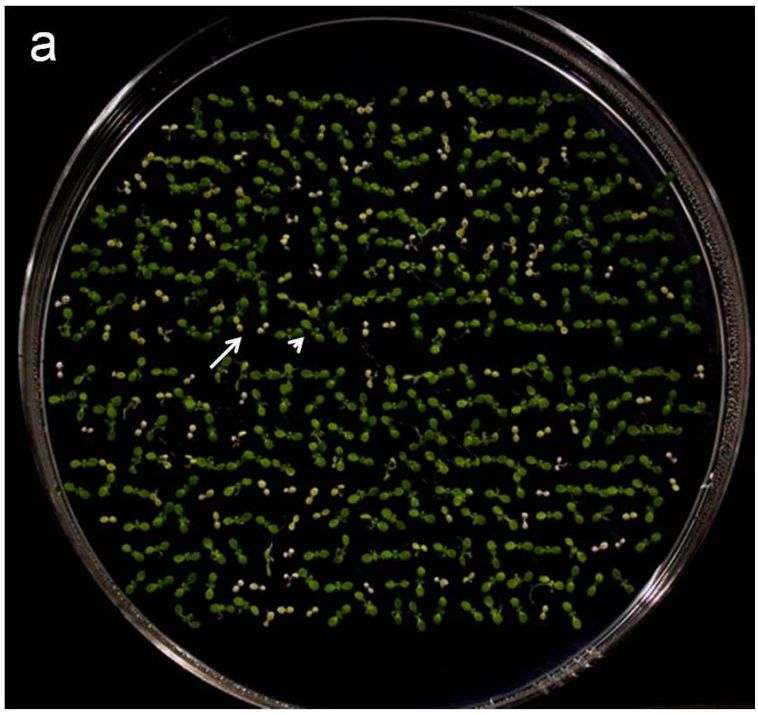
7.2.2 Identification of potential sperm cell and/or gamete expressed genes by a segregation distortion assay

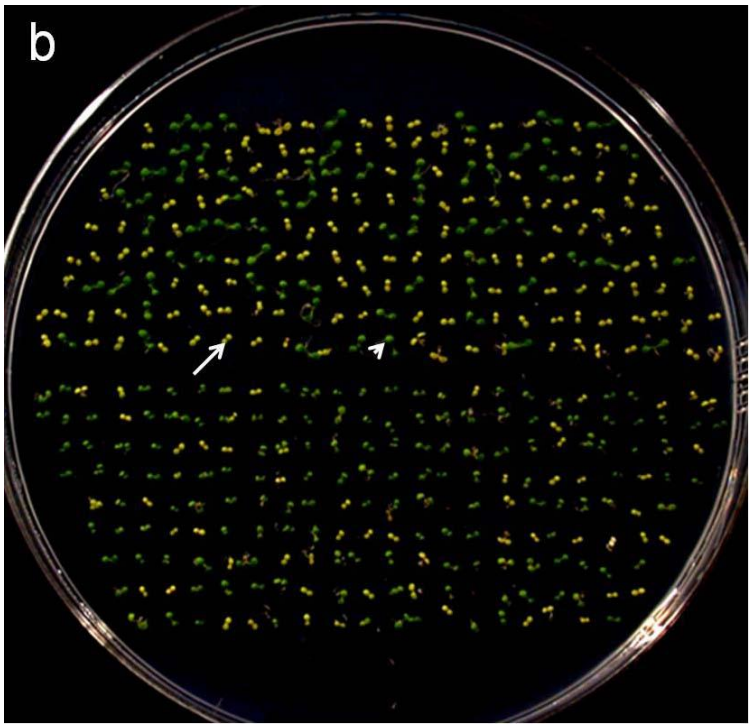
By observing the seed/silique phenotype alone it would be impossible to identify some classes of male gamtetophytic mutants, particularly mutations affecting pollen germination and tube growth as wild-type pollen could out-compete the mutant pollen leading to full seed set. Therefore an assay analysing the segregation ratio of the gene carrying the T-DNA insertion (Azpiroz-Leehan and Feldmann, 1997) was adopted in order to identify these mutants (see background of this technique in section 7.1.1).

123 T-DNA insertion lines ordered from NASC (Nottingham Arabidopsis Stock Centre) were sown on kanamycin (Kan) MS media to select a Kan^R plant for each line - heterozygous plants were required for the assay. One potential heterozygous plant per line was used for the assay as it allowed for an examination of the segregation ratio of the gene carrying the T-DNA. Due to the numbers of lines involved at this stage PCR screening of plants was not adopted to determine heterozygosity and therefore some homozygous plants would have entered the assay (both homozygous and heterozygous plants could grow on Kan selective media). The plants were allowed to set self seed. Approximately 200 seeds from each plant were grown on a Kan MS plate to score for the Kan^R to Kan^S ratio (Figure 7.5a, b). Of the123 lines screened, 20 were found to have segregation distortion in which the Kan^R to Kan^S ratio deviated from the normal 3:1 Mendelian ratio (75% Kan^R). The lines with segregation distortion where the T-DNA was expected to be inserted in gametophytically expressed genes were divided into 3 classes depending on the percentage of Kan^R plants, Kan^R > 50% (but less than 75%) possible male gametophytic gene effect, Kan^R = 50% possible female or male effects, and Kan^R < 50% possible male and female effects (Table 7.1). The lines that demonstrated > 75% Kan^R were concluded to most likely carry more than one copy of T-DNA.

Figure 7.5. Segregation distortion assay demonstrating plant lines with segregation distortion in comparison to a line demonstrating normal 3:1 Kan^R to Kan^S segregation ratio.

Segregation distortion assay demonstrating typical examples of plant lines that have a segregation ratio deviating from the normal Medelian ratio ($3Kan^R$:$1Kan^S$). Seeds collected from individual plants selected for each T-DNA insertion knockout line were grown on an MS plate containing kanamycin for selection. Plants that contained a T-DNA insertion were able to grow in the presence of kanamycin (arrow head). Plants without a T-DNA insertion were not be able to grow and appeared yellow on the plate (arrow). a) An example of plant lines (top half, N592811 [At3g23610] and bottom half, N503164 [At2g29940]) with a normal segregation ratio of $3Kan^R$:$1Kan^S$ indicating the mutated gene was probably sporophytic. b) An example of plant lines (top half, N517925 [At1g03250] and bottom half, N511269 [At5g59370]) where the segregation ratio deviated from 3:1indicating that the gene affected by the mutation was probably gametophytic.


a


b

7.2.3 Genotypic analysis of the high priority candidate lines (obtained from chapter 5) and lines that demonstrated a semi-infertility phenotype and segregation distortion

PCR genotyping of T-DNA insertion lines was carried out to confirm the presence of the T-DNA insert in the genes of interest. Genomic DNA extracted from individual plants is used as a template. The identification of wt (plant without T-DNA insertion), heterozygous (T-DNA insertion into one allele) or homozygous plants (T-DNA inserted into both alleles) is confirmed by two PCR reactions using different primer pairs.

Selected plant lines for the genotypic analysis were chosen from the following criteria; present in the priority list (Chapter 5), having a reduced seed set phenotype, and plants demonstrating segregation distortion. A total of 50 lines were selected for genotypic analysis. As mentioned above the PCR was performed using two sets of primers per individual plant, a gene specific primer pair (LP/RP) and a T-DNA left border (Lb) primer with the RP. In a wt plant, only one band from the gene specific primers can be amplified (Figure 7.6wt). In a plant homozygous (Hm) for the T-DNA, only a smaller size product derived from the Lb/RP pair can be detected (Figure 7.6Hm). Finally in heterozygous (Hz) plants, both bands derived from LP/RP and Lb/RP pairs can be detected (Figure 7.6Hz).

Of 50 lines investigated, T-DNA was detected in only 32 lines. The rest of the plant lines screened as wild-type with no T-DNA being detected. Moreover, of these 32 lines, 22 lines demonstrated inconsistency between the reproductive phenotype and the genotype suggesting that more than a single copy of T-DNA was inserted in the genome. Thus only 10 lines demonstrated consistency between the genotype, as determined by PCR, and the phenotype and these were subjected to detailed further analysis.

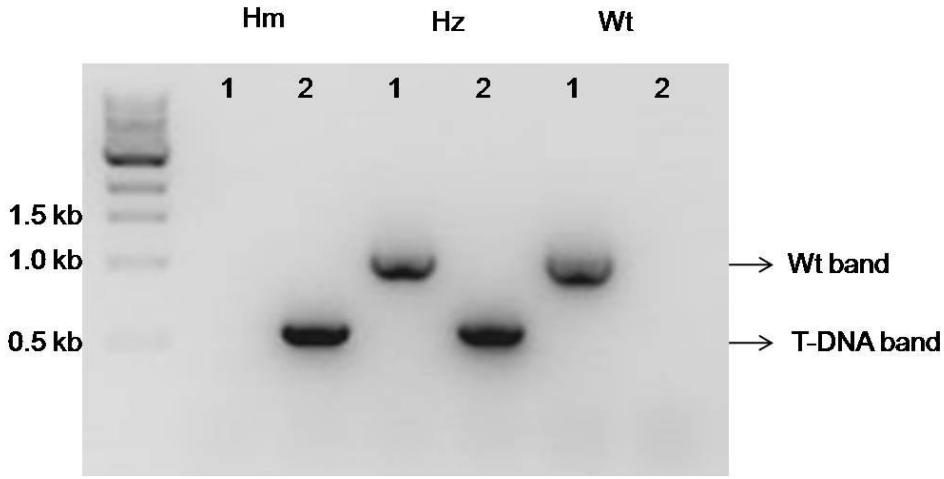


Figure 7.6. Genotypic analysis of T-DNA insertion lines by PCR.

A typical example of PCR genotyping analysis of a T-DNA insertion line (SALK_006098) with PCR products indicating wild-type, heterozygous, and homozygous plants is being presented. Plants that carried a T-DNA insertion in the genes of interest were identified using this PCR technique. In each individual plant two PCR reactions were performed. In the first reaction, a combination of gene specific primers (Appendix A) was used to amplify the wild-type allele of the gene of interest. The second reaction used a combination of the T-DNA left border primer (LP) and the gene specific right primer to identify the mutated allele of the gene. Only the larger wild-type PCR product would be found in wild-type plants. In heterozygous plants, both PCR products would be amplified. Finally in homozygous plants, both alleles have T-DNA insertions and thus only the smaller product of these mutant alleles would be amplified.

Table 7.1. T-DNA insertion lines of sperm-expressed gene candidates subjected to segregation ratio, genotypic and phenotypic analyses.

Sperm cell-expressed gene candidates identified following a bioinformatics analysis (chapter 5) were investigated utilising T-DNA insertion mutagenesis. 123 T-DNA lines representing 79 genes were analysed by i) reduced-seed set phenotype identification, and ii) a segregation distortion assay. A subset of these T-DNA lines that demonstrated the reduced-seed set phenotype, deviated segregation ratio (tested by χ^2 [$P < 0.05$ being significantly different from 3:1]), or listed in the high priority candidates were genotyped. The genotype of each line was compared with the phenotype and scored according to their correspondence. Some lines that contained a T-DNA failed to demonstrate the connection between genotype and phenotype suggesting that the phenotype was caused by disruption of other unidentified genes (identified as N in the G&P correspondence column). For some of the T-DNA lines the presence of a T-DNA insertion could not be verified (identified as N/A in the table). Not all lines were genotyped by PCR (−).

T-DNA Line	At number	Description	Phenotype	Segregation ratio (Kan^R:Kan^S)	%R	Genotyping attempted	T-DNA presence	G & P correspondence
High priority candidates (identified as described in chapter 6)								
N531697	At4g32830	histone serine kinase(H3-S10 specific) / kinase/ protein serine/threonine kinase; ATAUR1 (ATAURORA1)	Y1	S	S	Y	Y	N
N612121	At4g32830	histone serine kinase(H3-S10 specific) / kinase/ protein serine/threonine kinase; ATAUR1 (ATAURORA1)	Y1	S	S	Y	Y	N
N406238	At3g20190	leucine-rich repeat transmembrane protein kinase, putative	N	155:49	76%	N	–	–
N612791	At3g20190	leucine-rich repeat transmembrane protein kinase, putative	N	S	S	N	–	–
N643106	At4g11920	WD-40 repeat family protein	Y1	S	S	Y	N	N/A
N501978	At4g11920	WD-40 repeat family protein	N	S	S	Y	Y	N/A
N573708	At4g11920	WD-40 repeat family protein	N	160:43	79%	Y	Y	N/A
N877491	At2g22740	SET domain-containing protein (SUVH6) [Arabidopsis thaliana].	N	180:20	90%	N	–	–
N879899	At2g22740	SET domain-containing protein (SUVH6) [Arabidopsis thaliana].	N	192:8	96%	N	–	–
N504068	At5g39650	unknown protein	N	S	S	N	–	–
N806939	At5g39650	unknown protein	N	200:0	HM	N	–	–
N631951	At1g10090	unknown protein	N	S	S	Y	Y	N/A
N574693	At3g50910	unknown protein	Y3	21:29*	42%[a]	Y	Y	Y
N632810	At3g50910	unknown protein	Y1	S	S	N	–	–
N611440	At5g23130	peptidoglycan-binding LysM domain-containing protein	Y1	202:0	HM	Y	Y	N
N595724	At1g73070	leucine-rich repeat family protein	N	191:14	93%	N	–	–
N507150	At1g73070	leucine-rich repeat family protein	N	148:57	72%[b]	Y	N	N/A

N503164	At2g29940	ABC transporter family protein	N	154:49	76%	Y	N	N/A
N503174	At2g29940	ABC transporter family protein	N	S	S	N	N	N/A
N349088	At2g29940	ABC transporter family protein	N	153:50	75%	N	–	–
N592811	At3g23610	dual specificity protein phosphatase (DsPTP1)	N	165:41	80%	Y	Y	N/A
N877896	At1g78300	14-3-3 protein GF14 omega	N	145:56	72%[b]	Y	N	N/A
N503063	At4g28000	AAA-type ATPase family protein	N	152:50	75%	Y	Y	N/A
N574465	At4g28000	AAA-type ATPase family protein	N	205:0	HM	Y	Y	N/A
N801802	At4g28000	AAA-type ATPase family protein	N	173:31	85%	N	–	–
N540809	At5g54640	histone H2A	N	203:0	HM	Y	Y	N/A
N340039	At1g11250	syntaxin, putative (SYP95)	N	205:0	HM	N	–	–
N811141	At2g19380	RNA recognition motif (RRM)-containing protein	Y2	198:0	HM	Y	Y	N
N333297	At1g32400	senescence-associated family protein [Arabidopsis thaliana].	N	205:0	HM	N	–	–
N524905	At3g61230	LIM domain-containing protein [*Arabidopsis thaliana*].	N	S	S	N	–	–
N171235	At3g61230	LIM domain-containing protein [*Arabidopsis thaliana*].	N	154:44	78%	N	–	–
N506096	At2g20440	RabGAP/TBC domain-containing protein	Y1	S	S	Y	N	N/A
N506098	At2g20440	RabGAP/TBC domain-containing protein	Y1	23:22*	51%[a]	Y	Y	Y
N639822	At1g69960	serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	Y1	S	S	Y	Y	N
N513178	At1g69960	serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	N	S	S	N	–	–
N509229	At5g50930	unknown protein	N	209:0	HM	Y	Y	N/A
N509951	At5g50930	unknown protein	N	197:0	HM	N	–	–
Other pollen expressed genes								
N629264	At5g26150	protein kinase family protein	Y1	S	S	Y	Y	N
N510944	At5g26150	protein kinase family protein	N	S	S	N	–	–
N878215	At5g04250	OTU-like cysteine protease family protein	N	194:6	97%	N	–	–
N625283	At4g27750	unknown protein	N	200:0	HM	N	–	–
N839022	At4g10810	unknown protein	Y2	192:8	96%	Y	N	N/A
N638312	At4g10810	unknown protein	N	S	S	Y	N	N/A
N828582	At1g48940	plastocyanin-like domain-containing protein	Y3	190:17	92%	Y	N	N/A
N515893	At1g48940	plastocyanin-like domain-containing protein	N	200:0	HM	N	–	–
N642531	At3g55430	glycosyl hydrolase family 17 protein / beta-1,3-glucanase, putative	Y2	S	S	Y	N	N/A
N829483	At3g55430	glycosyl hydrolase family 17 protein / beta-1,3-glucanase, putative	Y1	186:15	93%	Y	Y	N
N595127	At5g53820	unknown protein	N	181:20	90%	Y	N	N/A
N527524	At5g53820	unknown protein	Y3	109:89	55%[a]	Y	Y	Y
N517925	At1g03250	unknown protein	Y1	117:89	57%[a]	Y	N	N/A
N517923	At1g03250	unknown protein	N	162:43	79%	N	–	–
N507677	At1g26750	unknown protein	Y1	139:62	69%[a]	Y	Y	Y
N603674	At3g05820	beta-fructofuranosidase, putative / invertase, putative / saccharase, putative / beta-	N	188:12	94%	N	N/A	N/A

		fructosidase, putative						
N516334	At3g05820	beta-fructofuranosidase, putative / invertase, putative / saccharase, putative / beta-fructosidase, putative	Y1	S	S	Y	N	N/A
N531941	At5g59370	actin 4 (ACT4)	Y1	172:28	86%	Y	N	N/A
N511269	At5g59370	actin 4 (ACT4)	N	147:59	71%[b]	N	–	–
N507957	At3g22760	CXC domain containing TSO1-like protein 1 (SOL1)	Y2	151:50	75%	Y	Y	N
N833185	At3g22760	CXC domain containing TSO1-like protein 1 (SOL1)	Y1	193:12	94%	Y	Y	N
N583433	At3g53750	actin 3 (ACT3)	Y1	22:28*	44%[a]	Y	Y	Y
N583426	At3g53750	actin 3 (ACT3)	N	S	S	N	–	N/A
N519248	At1g03180	unknown protein	Y2	205:0	HM	Y	Y	Y
N527250	At1g03180	unknown protein	N	202:0	HM	N	–	–
N607761	At4g04930	fatty acid desaturase family protein	N	181:22	89%	Y	Y	N/A
N612229	At5g57000	unknown protein	N	205:0	HM	Y	N	N/A
N552675	At5g57000	unknown protein	N	S	S	N	–	–
N832901	At2g46170	reticulon family protein (RTNLB5)	Y1	165:44	79%	Y	Y	N
N814621	At3g06580	galactokinase (GAL1)	Y2	186:23	89%	Y	Y	N
N581223	At3g06580	galactokinase (GAL1)	N	154:52	75%	N	–	N/A
N591193	At5g05490	cohesion family protein SYN1, splice variant 1 (SYN1)	Y3	23:28*	45%[a]	Y	Y	Y
N637095	At5g05490	cohesion family protein SYN1, splice variant 1 (SYN1)	Y1	S	S	N	–	N/A
N826961	At1g55530	zinc finger (C3HC4-type RING finger) family protein	Y3	185:18	91%	Y	N	N/A
N350709	At1g55530	zinc finger (C3HC4-type RING finger) family protein	Y2	206:0	HM	N	–	N/A
N544578	At3g51030	thioredoxin H-type 1 (TRX-H-1)	Y1	129:73	64%[a]	Y	Y	Y
N592775	At3g51030	thioredoxin H-type 1 (TRX-H-1)	N	S	S	N	–	–
N503754	At5g16880	VHS domain-containing protein / GAT domain-containing protein	N	201:0	HM	N	–	–
N503755	At5g16880	VHS domain-containing protein / GAT domain-containing protein	Y3	S	S	Y	Y	N
N564121	At3g56640	exocyst complex subunit Sec15-like family protein	N	140:61	70%[b]	Y	N	N/A
N564439	At3g18040	mitogen-activated protein kinase, putative / MAPK, putative (MPK9) [*Arabidopsis thaliana*].	N	107:99	52%[a]	Y	Y	Y
N631139	At3g23870	permease-related	N	77:131	37%[a]	Y	Y	Y
N559700	At2g34320	unknown protein	N	205:0	HM	N	–	–
N511443	At4g22450	unknown protein	N	188:14	93%	N	–	–
N561973	At1g55570	multi-copper oxidase type I family protein	N	S	S	N	–	–
N554102	At5g19310	homeotic gene regulator, putative	N	S	S	N	–	–
N514949	At1g30540	ATPase, BadF/BadG/BcrA/BcrD-type family	N	S	S	N	–	–
N519919	At5g48140	polygalacturonase, putative / pectinase, putative	N	S	S	N	–	–
N557422	At1g34340	esterase/lipase/thioesterase family protein	N	S	S	N	–	–
N505551	At3g19090	RNA-binding protein, putative	N	S	S	N	–	–

N552144	At3g52560	ubiquitin-conjugating enzyme family protein	N	S	S	N	–	–
N516767	At3g27930	unknown protein	N	S	S	N	–	–
N627773	At4g24380	unknown protein	N	156:47	77%	N	–	–
N587610	At5g15600	unknown protein	N	200:0	HM	N	–	–
N809157	At2g20700	unknown protein	N	179:20	90%	N	–	–
N844519	At3g22290	unknown protein	N	147:57	72%[b]	N	–	–
N545948	At4g35560	unknown protein	N	199:6	97%	N	–	–
N845908	At4g11560	bromo-adjacent homology (BAH) domain-containing protein	N	136:64	68%[a]	N	–	–
N645242	At3g07820	polygalacturonase 3 (PGA3) / pectinase	N	208:0	HM	N	–	–
N565623	At5g27150	sodium proton exchanger / Na+/H+ antiporter (NHX1)	N	S	S	N	–	–
N591671	At1g14560	mitochondrial substrate carrier family protein	N	S	S	N	–	–
N804520	At1g74480	RWP-RK domain-containing protein	N	148:55	73%[b]	N	–	–
N516600	At5g55290	ATP synthase subunit H family protein	Y1	202:0	HM	N	–	N/A
N800365	At4g15093	catalytic LigB subunit of aromatic ring-opening dioxygenase family	N	186:14	93%	N	–	–
N587009	At3g27440	uracil phosphoribosyltransferase, putative / UMP pyrophosphorylase, putative / UPRTase, putative	N	153:52	75%	N	–	–
N575828	At4g36690	U2 sSNP auxiliary factor large subunit, putative	N	189:11	95%	N	–	–
N542784	At5g55850	nitrate-responsive NOI protein, putative	N	137:63	69%[a]	N	–	–
N508372	At1g13890	SNAPSNAP25 homologous protein, putative / synaptosomal-associated protein SNAP25-like, putative (SNAP30)	N	202:0	HM	N	–	–
N557380	At3g07850	exopolygalacturonase / galacturan 1,4-alpha-galacturonidase / pectinase	N	202:0	HM	N	–	–
N514275	At4g04690	F-box family protein (FBX15)	N	148:52	74%[b]	N	–	–
N531570	At3g08990	yippee family protein	N	156:47	79%	N	–	–
N559199	At2g19770	profilin 4 (PRO4) (PFN4)	N	169:37	82%	N	–	–
N619629	At3g59550	cohesin family protein SYN3 (SYN3)	N	157:50	76%	N	–	–
N508661	At5g01180	proton-dependent oligopeptide transport (POT) family protein	N	187:23	89%	N	–	–
N567682	At3g28490	oxidoreductase, 2OG-Fe(II) oxygenase family protein	N	S	S	N	–	–
Unannotated 'hit' regions and neighbouring genes								
N872474	At2g41310	two-component responsive regulator / response reactor 3 (RR3)	N	185:15	92%	N	–	–
N514557	At2g41330	glutaredoxin family protein	N	200:0	HM	N	–	–
N627028	At4g20320	CTP synthase, putative / UTP--ammonia ligase, putative	N	205:0	HM	N	–	–
N520074	At4g20320	CTP synthase, putative / UTP--ammonia ligase, putative	N	157:52	75%	N	–	–
N805876	At4g20330	transcription initiation factor-related	N	161:42	79%	N	–	–
N557940	At2g41310	two-component responsive regulator / response reactor 3 (RR3)	N	157:44	78%	N	–	–
N507868	At2g41330	glutaredoxin family protein	N	S	S	N	–	–
N627028	At4g20325	unannotated (later identified as unknown protein similar to Os04g0476400 [Oryza sativa]	Y1	154:50	76%	Y	Y	Y
N515812		unannotated	N	S	S	N	–	–

N637682	unannotated	N	157:43	79%	N	−	−
N870498	unannotated	N	169:36	82%	N	−	−

S: silenced kanamycin selection marker, G: genotype, P: phenotype, R: resistant seedling, Y: yes, N: no, Y1: class 1 phenotype (unfertilised ovule), Y2: class 2 phenotype (aborted seed), Y3: class 3 phenotype (minute silique), N/A: unable to be identified due to PCR problems, −: unable to be identified as not genotyped, HM: homozygous, *: performed by PCR due to silented Kan marker, a: significantly different from 3:1 (χ^2, $P < 0.05$), b :not significantly different from 3:1 (χ^2, $P > 0.05$), Pink fill: lines that demonstrated a reduced seed set phenotype and were chosen for further analyses, Blue fill: lines that demonstrated segregation distortion and were chosen for further analyses, Green fill: genes selected for GFP fusion studies (chapter 6), Yellow highlight; lines that demonstrated segregation distortion.

7.2.4 Detailed phenotypic analysis of single copy T-DNA insertion knockout lines that demonstrated semi-infertility phenotype and/or segregation distortion

T-DNA insertion knockout lines that had reduced seed set were analysed in further detail. Moreover according to segregation distortion, phenotypic and genotypic analysis, these lines contained a single copy of T-DNA that co-segregated with the genes of interest. The list of these plant lines (10 lines) which represent a subset of those detailed in table 7.1 is summarised in table 7.2. Their gene expression patterns in different tissue types (Figure 7.7) were also examined using available microarray data (Gene Atlas). Most of these genes were found to be expressed highly in pollen in comparison to other tissue types. The number of seeds present in the siliques of these lines were counted and subjected to statistical analysis. Heterozygous plants of each line were reciprocally crossed with wild-type to determine which parent was responsible for the phenotype (parental effect), either male, female or both. The seed set from the crosses was recorded and moreover, aniline blue staining of pollen tubes in each reciprocal cross was also carried out. Furthermore, both the pollen and ovules of each plant were examined in detail by DAPI, FDA pollen staining, *in vitro* pollen tube growth and ovule microscopic analyses - the details and the results of each technique will be described later in this chapter. These detailed analyses were performed in order to identify putative gene function and the effect of the mutations on the gametes.

Table 7.2. List of plant lines carrying detectable single copy T-DNA insertions in the gene of interest that demonstrated a semi-infertility phenotype and/or segregation distortion.

The listed T-DNA insertion lines were selected for detailed phenotypic analysis as they demonstrated a reduced seed set phenotype and/or segregation distortion. Moreover, according to the segregation assays and phenotypic/genotypic analyses, these plant lines carried a single copy of T-DNA which facilitates gene identification and mutant characterisation. The maternal affect of the mutated genes was investigated by scoring for the semi-infertility phenotype (reduced seed set phenotype – see table 7.1) and reciprocal backcross with wt pollen (see section 7.2.4.3) and summarised in the 'Female effect' column. The paternal affect of the mutant genes was investigated by the transmission efficiency (TE) assay of seeds obtained from a cross between mut pollen and wt female (tested by χ^2 [$P < 0.05$ being significantly different from 1:1]). The TE results from each line is summarised in the 'Male effect' column below.

T-DNA Line	At number	Description	Fem effect	Male effect	Seg ratio	%R	Kan^R:Kan^S (for TEmale)	%R	TE (male)
N506098	At2g20440	RabGAP/TBC domain-containing protein	Y	Y	1.04:1	51%	22:32*	41%[b]	69%
N507677	At1g26750	unknown protein	Y	Y	2.22:1	69%	18:32*	36%[a]	56%
N527524	At5g53820	unknown protein	Y	Y	1.22:1	55%	19:35*	35%[a]	54%
N544578	At3g51030	thioredoxin H-type 1 (TRX-H-1)	Y	Y	1.78:1	64%	88:112	44%[b]	79%
N564439	At3g18040	mitogen-activated protein kinase, putative / MAPK, putative (MPK9)	N	Y	1.08:1	52%	85:113	43%[a]	75%
N574693	At3g50910	unknown protein	Y	Y	0.72:1	42%	20:30*	40%[b]	67%
N583433	At3g53750	actin 3 (ACT3)	Y	Y	0.79:1	44%	17:31*	35%[a]	55%
N591193	At5g05490	cohesin family protein SYN1, splice variant 1 (SYN1)	Y	Y	0.82:1	45%	N/A	N/A	N/A
N631139	At3g23870	permease-related	N	Y	0.59:1	37%	72:133	35%[a]	54%
N826057	At4g20325	unknown protein similar to Os04g0476400 [Oryza sativa]	Y	N	3.17:1	76%	102:98	51%[b]	104%

Fem: female, Seg: segregation, Y: yes, N: no, N/A: unable to be performed due to anther phenotype, R: resistant seedling, TE: transmission efficiency, *: prformed by PCR due to silented Kan marker, a: significantly different from 1:1 (χ^2, $P < 0.05$), b: not significantly different from 1:1 (χ^2, $P > 0.05$), Blue text indicates lines with mutation affecting male only.

Figure 7.7. Expression pattern of genes under investigation (data from Gene Atlas, At4g20325 is unavailable).

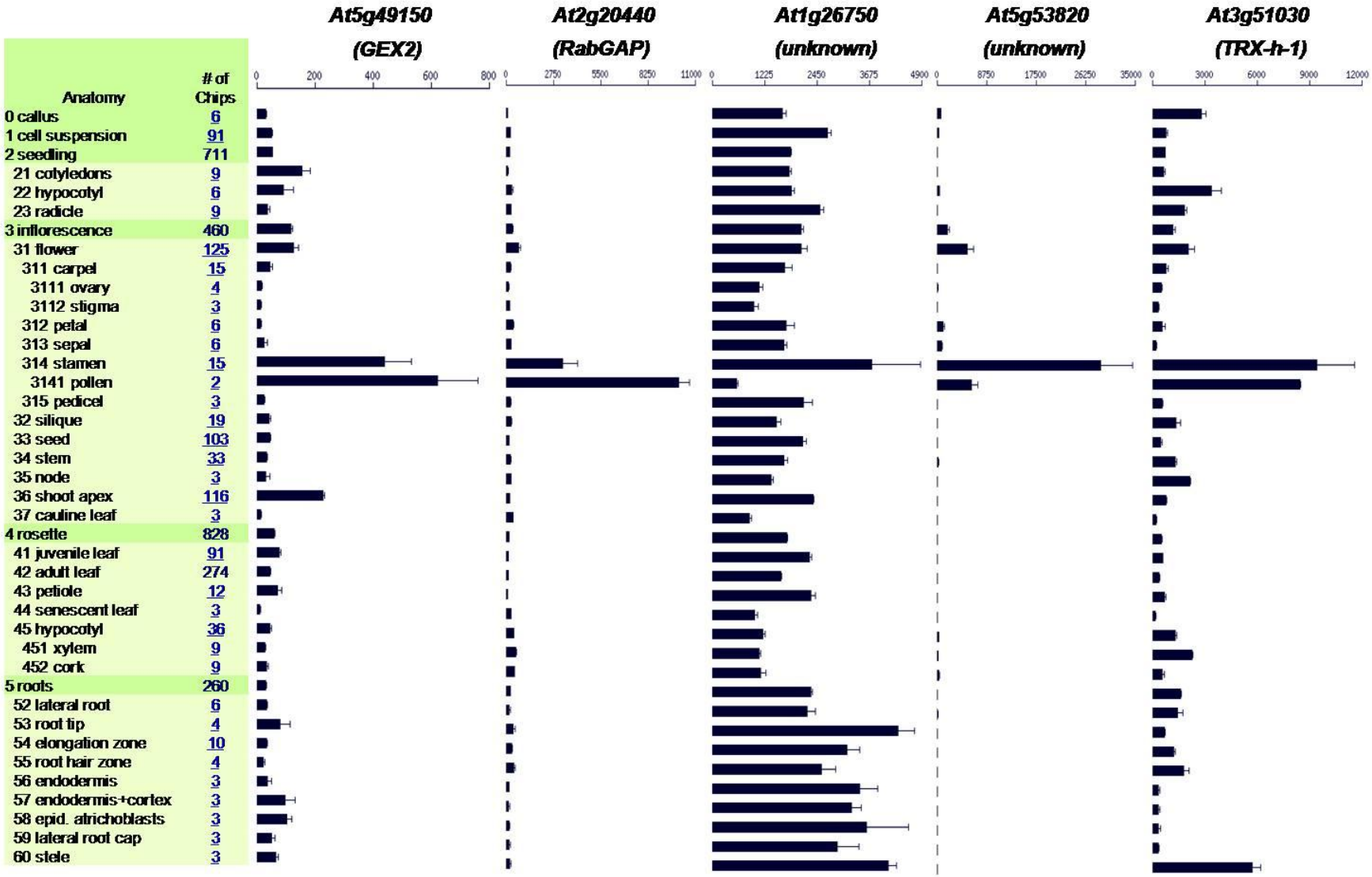
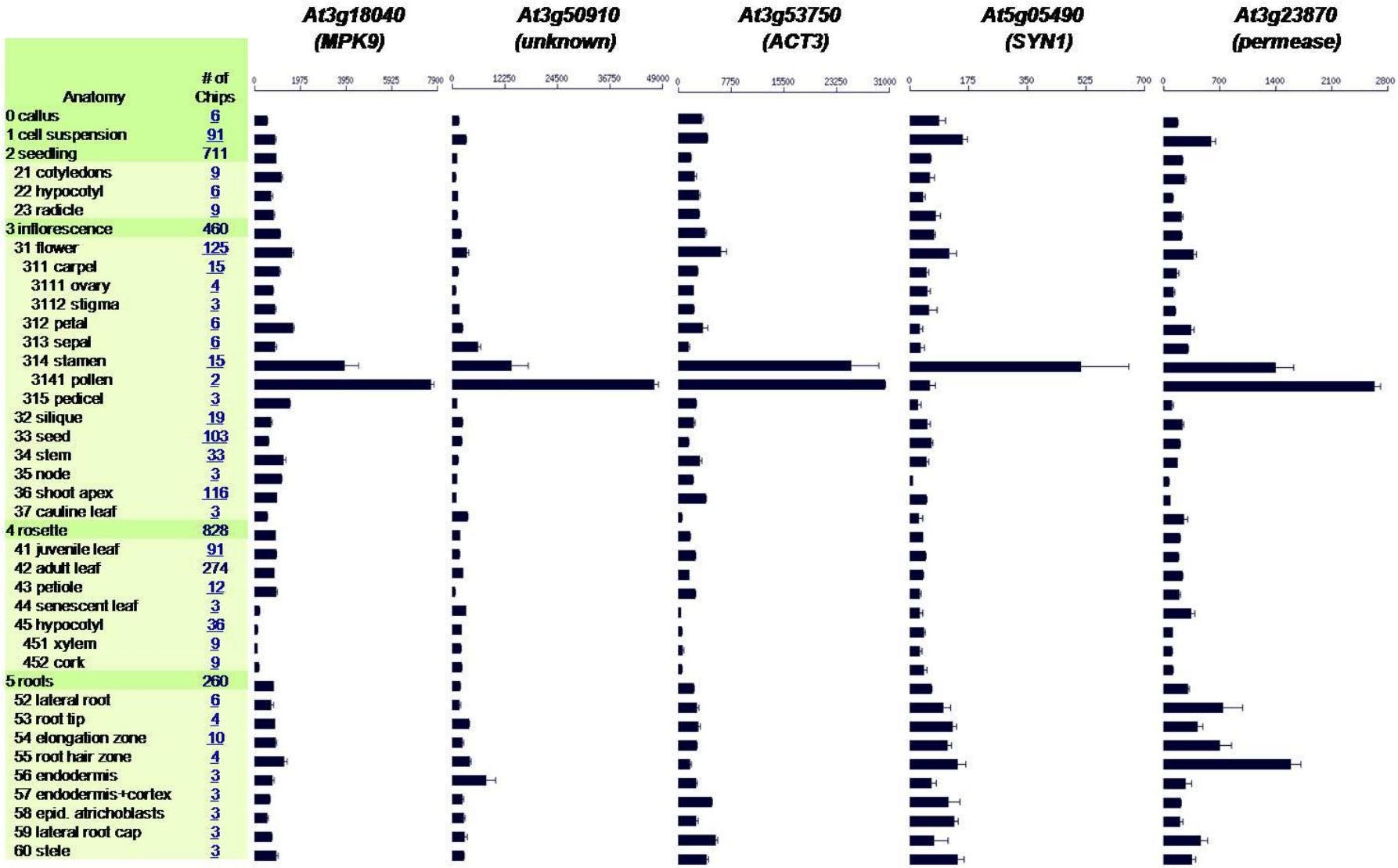

At5g49150
(GEX2)
At2g20440
(RabGAP)
At1g26750
(unknown)
At5g53820
(unknown)
At3g51030
(TRX-h-1)
Anatomy
# of Chips
0 callus 6
1 cell suspension 91
2 seedling 711
21 cotyledons 9
22 hypocotyl 6
23 radicle 9
3 inflorescence 460
31 flower 125
311 carpel 15
3111 ovary 4
3112 stigma 3
312 petal 6
313 sepal 6
314 stamen 15
3141 pollen 2
315 pedicel 3
32 silique 19
33 seed 103
34 stem 33
35 node 3
36 shoot apex 116
37 cauline leaf 3
4 rosette 828
41 juvenile leaf 91
42 adult leaf 274
43 petiole 12
44 senescent leaf 3
45 hypocotyl 36
451 xylem 9
452 cork 9
5 roots 260
52 lateral root 6
53 root tip 4
54 elongation zone 10
55 root hair zone 4
56 endodermis 3
57 endodermis+cortex 3
58 epid. atrichoblasts 3
59 lateral root cap 3
60 stele 3
0 200 400 600 800
0 2750 5500 8250 11000
0 1225 2450 3675 4900
0 8750 17500 26250 35000
0 3000 6000 9000 12000

Figure 7.7. Expression pattern of genes under investigation (continued).


At3g18040 (MPK9)
At3g50910 (unknown)
At3g53750 (ACT3)
At5g05490 (SYN1)
At3g23870 (permease)
Anatomy
# of Chips
0 callus 6
1 cell suspension 91
2 seedling 711
21 cotyledons 9
22 hypocotyl 6
23 radicle 9
3 inflorescence 460
31 flower 125
311 carpel 15
3111 ovary 4
3112 stigma 3
312 petal 6
313 sepal 6
314 stamen 15
3141 pollen 2
315 pedicel 3
32 silique 19
33 seed 103
34 stem 33
35 node 3
36 shoot apex 116
37 cauline leaf 3
4 rosette 828
41 juvenile leaf 91
42 adult leaf 274
43 petiole 12
44 senescent leaf 3
45 hypocotyl 36
451 xylem 9
452 cork 9
5 roots 260
52 lateral root 6
53 root tip 4
54 elongation zone 10
55 root hair zone 4
56 endodermis 3
57 endodermis+cortex 3
58 epid. atrichoblasts 3
59 lateral root cap 3
60 stele 3
0 1975 3950 5925 7900
0 12250 24500 36750 49000
0 7750 15500 23250 31000
0 175 350 525 700
0 700 1400 2100 2800

7.2.4.1 Seed set counting

Of these 10 T-DNA lines summarised in table 7.2, N564439 (SALK_064439 [At3g18040 – MPK9]) and N631139 (SALK_131139 [At3g23870 - permease related]) did not show a reduced seed set phenotype indicating a male gametophytic defect. The unfertilised ovule class was prevalent amongst these lines i.e. N506098 (SALK_006098 [At2g20440 - RabGAP]), N507677 (SALK_007677 [At1g26750 – unknown protein]), N544578 (SALK_044578 [At3g51030 - TRX-H-1]), N583433 (SALK_083433 [At3g53750 - ACT3]), and N826057 (SAIL_609_A02 [At4g20325 - unknown protein]). The minute silique class was demonstrated in N591193 (SALK_091193 [At5g05490 - SYN1]), N527524 (SALK_027524 [At5g53820 – unknown protein]), and N574693 (SALK_074693 [At3g50910 – unknown protein]). Viable seed from these reduced seed set siliques (n = 20; χ^2 being significantly different from wt, $P < 0.001$) were counted and scored in table 7.3. A histogram comparing the percentage of viable seed in each line is presented in figure 7.8 Interestingly, this seed set phenotype was observed only in heterozygous plants where homozygous plants appear as wild-type in most of these screened plant lines (N506098 [At2g20440 - RabGAP], N507677 [At1g26750 – unknown protein], N544578 [At3g51030 - TRX-H-1], N583433 [At3g53750 - ACT3], and N826057 [At4g20325 – unknown protein]).

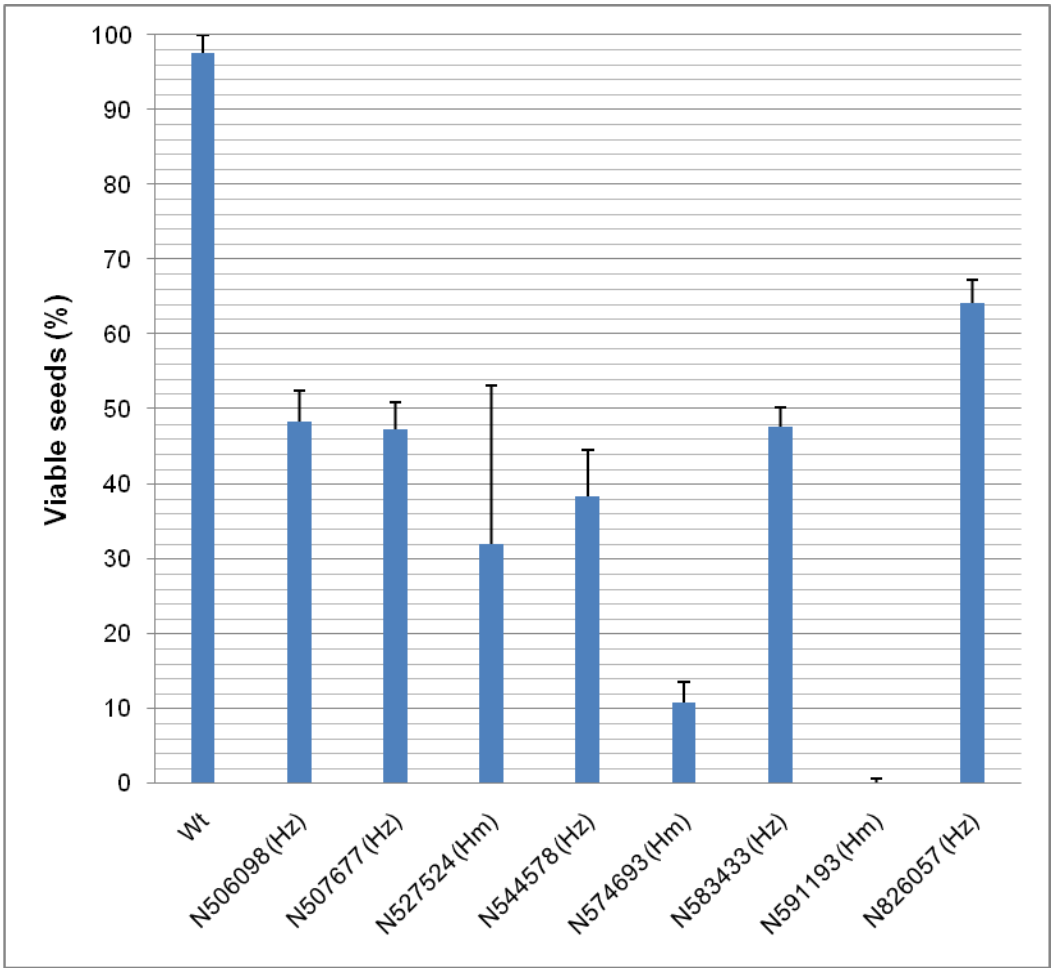


Figure 7.8. Percentages of viable seed from T-DNA lines having a reduced seed set phenotype.

Seed set data from the lines that demonstrated a reduced seed set phenotype were obtained (n = 20 siliques; χ^2 being significantly different from wt, $P < 0.001$). Histograms demonstrating the percentages of viable seed were generated and compared to wild-type (Wt). Only heterozygous (Hz) plants from the lines N506098 (At2g20440 - RabGAP), N507677 (At1g26750 – unknown protein), N544578 (At3g51030 - TRX-H-1), N583433 (At3g53750 - ACT3), and N826057 (At4g20325 – unknown protein) demonstrated the reduced seed set phenotype with homozygous lines having full seed set. For lines N591193 (At5g05490 - SYN1), N527524 (At5g53820 – unknown protein), and N574693 (At3g50910 – unknown protein), only homozygous (Hm) plants demonstrated the reduced seed set phenotype (bar: standard deviation).

7.2.4.2 Identification of a paternal effect in T-DNA insertion lines by transmission efficiency assays

Using the same principle as the segregation distortion assay, transmission efficiency assays were also performed with the aid of the Kan selectable marker of the T-DNA. Unlike segregation distortion assays where sporophytic and gametophytic mutations were identified, this transmission efficiency assay allows further separation between male and female gamete mutations. A mutation affecting the male gamete was elucidated by crossing pollen from a heterozygous mutant plant onto emasculated wild-type flowers. Seeds collected from this cross were grown on a Kan MS plate and scored for Kan^R. For a normal sporophytic gene mutation 50% of the seedlings would demonstrate Kan^R as half of the pollen carrying the mutant allele would fertilise wild-type ovules. On the other hand, if the male gametophyte (pollen) is in some way defective due to the mutation, the 50% of pollen carrying the wild-type allele would out-compete pollen carrying the mutant allele resulting in less than 50% of the seedling progeny containing the Kan^R marker. Of the 10 lines tested 9 lines demonstrated less than 50% Kan^R seedlings indicating the mutation affected the male gametophyte (Table 7.2). However, one line, N826057 (At4g20325 – unknown protein) showed 1:1 transmission ratio through male gamete indicating that the pollen was not defective. Nonetheless, all of the other lines demonstrated transmission efficiencies of approximately 70-90% suggesting only slight male effect.

7.2.4.3 Characterisation of the parental contribution to the mutant phenotypes by reciprocal crosses with wild-type

To confirm the parental affect of the mutation, reciprocal crosses of mutant (mut) with wild-type (wt) pollen and male-sterile A9 Arabidopsis with mut pollen were performed. Identification of a female defect in the mutant was demonstrated by observing reduced seed set when crosses were performed by using the mutant as a mother and wild-type as a father. Of 10 lines investigated, 8 lines showed a female defect (summarised in table 7.2) as the number of developing seed in these backcrosses corresponded to number of developing seeds in mut selfing (see section 7.2.4.1). A male defect was investigated by crossing male-sterile A9 with mutant pollen. A reduced seed set phenotype was absent from the siliques of such crosses in all lines. This result indicated that the mutation was not affecting male gametes, at least not at the stage of gamete fusion. However, these crosses alone were unable to rule out a paternal defect due to the non-limiting quantities of pollen on the stigma. Restricted pollinations coupled to aniline blue staining and *in vitro* pollen tube growth assays were then used to investigate possible male effects.

Seed set data derived from flowers that had undergone restricted pollination (n = 20) is reported in figure 7.9. Four lines N506098 (At2g20440 – RabGAP), N527524 (At5g53820 – unknown protein), N574693 (At3g50910 – unknown protein), and N631139 (At3g23870 – permease related) demonstrated a small effect of the mutation to the male side as shown by slight reduction in seed number (χ^2 significantly different from wt, $P < 0.001$)

7.2.4.4 Identification of paternal effects by aniline blue staining of self-pollinated mutant lines and reciprocal crosses with male-sterile A9 Arabidopsis

Plant lines that demonstrated a reduced seed set phenotype were allowed to self-pollinate. Aniline blue staining was performed on two day-old siliques (two days after pollination – 2 DAP) to observe the pollen tubes. All lines demonstrating the phenotype showed normal tube growth (Figure 7.10a). However, approximately 50% of the ovules failed to attract the tube to the micropyle resulting in no seed development (Figure 7.10b). Reciprocal crosses coupled with aniline blue staining were then performed to assess whether a paternal effect was contributing to the reduced seed set phenotype.

7.2.4.5 Assessment of male gametophytic defects by pollen viability tests and DAPI pollen staining

Pollen viability test was carried out by FDA staining (n = 300) (Figure 7.11c, d). All the lines screened were not significantly different from wild-type (χ^2, $P > 0.05$). DAPI staining (n = 300) (Figure 7.11a, b) was used to investigate the pollen grain and sperm cells morphologies. Once again all the tested lines appeared to be normal compared to wild-type.

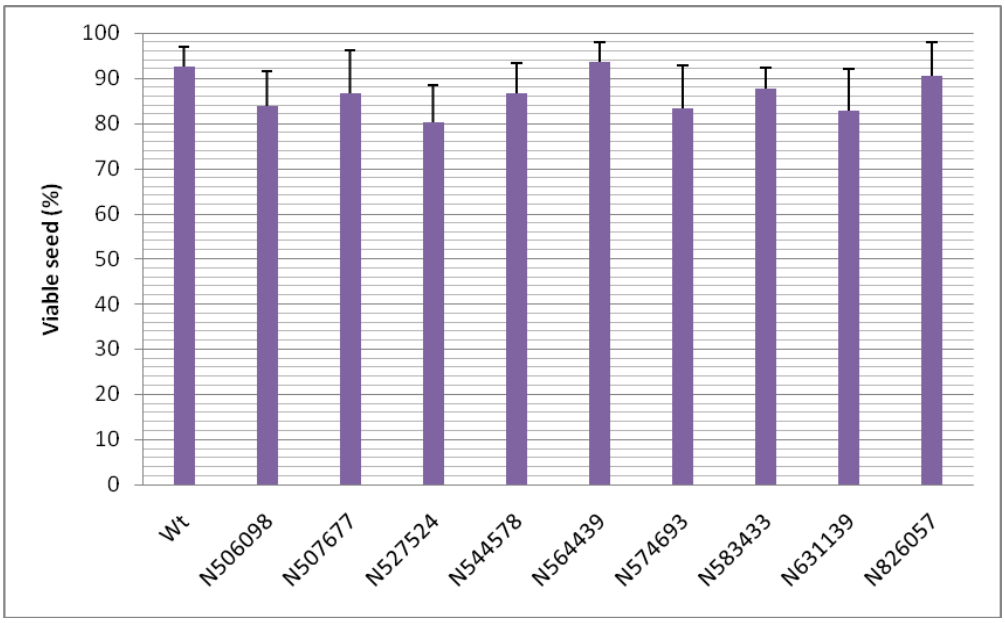


Figure 7.9. Percentages of viable seed obtained from restricted pollinations of mutant pollen on male-sterile-A9 flowers of the lines that demonstrated a male gametophytic defect.

Crossing of pollen derived from the mutant plant lines on A9 flowers revealed the paternal affect of the mutated gene. Restricted pollination was performed with care taken to ensure a roughly equal number of pollen grains in each cross. Approximately 70 grains were pollinated on each stigma. The number of seeds produced from each cross were counted and compared to a wild-type control. Four lines N506098 (At2g20440 – RabGAP), N527524 (At5g53820 – unknown protein), N574693 (At3g50910 – unknown protein), and N631139 (At3g23870 – permease related) demonstrated a small effect of the mutation to the male side as shown by slight reduction in seed number (χ^2 significantly different from wt, $P < 0.001$) N591193 (At5g05490 - SYN1) was unable to be performed due to the anther phenotype. Bar: standard deviation

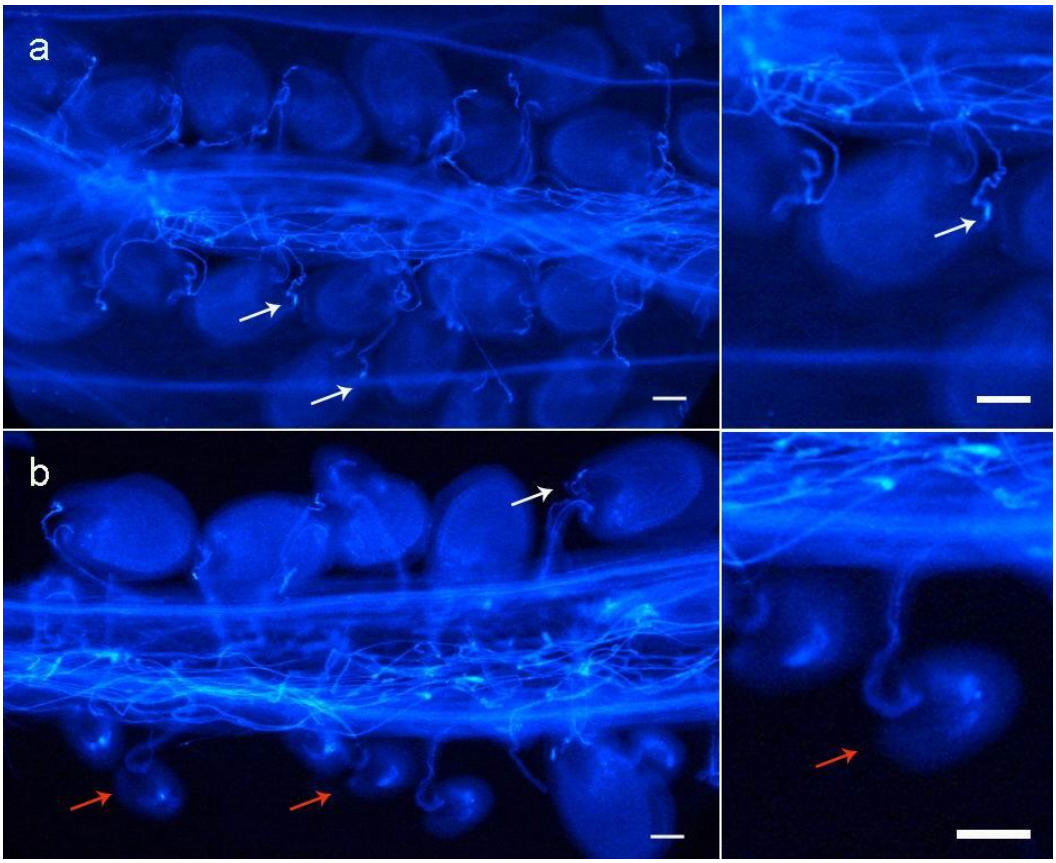


Figure 7.10. *In vivo* pollen tube growth assay coupled to aniline blue staining.

Observations of *in vivo* pollen tube growth were performed for every line that demonstrated a reduced seed set phenotype. Mutant plants and a wild-type control were allowed to self pollinate and the siliques were collected two day after pollination. The siliques were fixed and stained with aniline blue and observed by fluorescence light microscopy. a) Wild-type silique demonstrating full seed set with pollen tubes successfully entering the micropyle (arrow). b) In mutant siliques (SALK_006098 was used as an example) approximately half of the ovules were unfertilised as the pollen tube failed to navigate to them (red arrow, bar 50 μm).

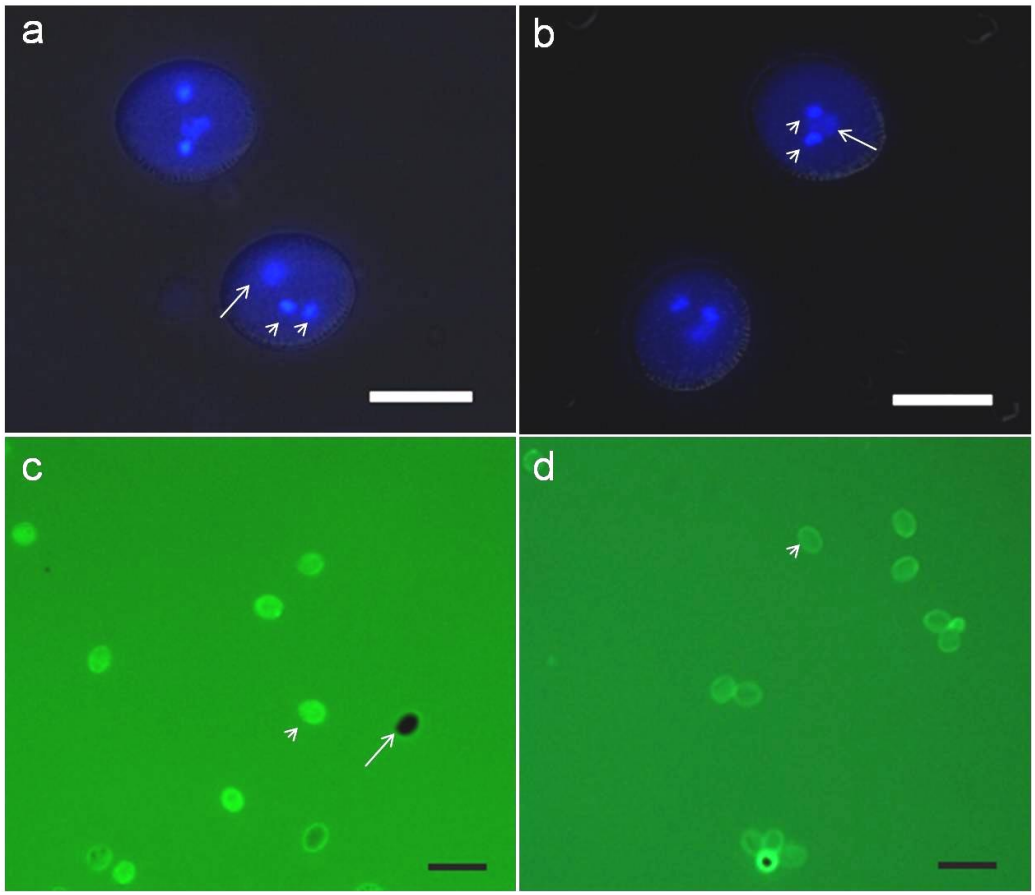


Figure 7.11. Assessment of 'mutant' pollen morphology and viability by DAPI and FDA staining.

Morphology of mature pollen grains was investigated for each T-DNA insertion line by DAPI staining (a, b). The two sperm cells (arrow head) fluorescenced brightly and were situated near the middle of the grain in the pollen from both wild-type (a) and T-DNA insertion line mutants (b). The less condensed pollen vegetative nucleus (arrow) appeared close to the two sperm. A pollen viability test was performed using fluorescein diacetate (FDA) (c, d). Viable pollen grains (arrow head) fluoresce green when observed under microscope equipped with UV (with FITC filter) where non-viable pollen remained dark (arrow). Both pollen samples from wild-type (c) and mutants (d) demonstrated the green product generated in the cytoplasm of viable pollen (a and b bars = 20 µm, c and d bars = 50 µm).

7.2.4.6 Assessment of male gametophyte defects by in vitro and in vivo pollen germination and elongation assays

Pollen germination and pollen tube growth assays were performed *in vitro* and *in vivo* to assess pollen germination ability in all the mutant lines except N591193 (At5g05490 - SYN1; unable to be performed due to anther phenotype) in comparison to wild-type. *In vitro* pollen germination was performed in a humid environment in the presence of a carpel to stimulate germination (Johnson-Brousseau and McCormick, 2004). The length of pollen tubes was measured after 6 hours incubation (n = 50). Pollen germination and tube length in almost every line studied were found to be similar to wild-type (approximately 80% of the pollen germinates at around 450 µm in length; paired-difference t-test, $P>0.05$) (Figure 7.12a, b) except N527524 (At5g53820 – unknown protein) where approximately 15% of the pollen was misshapen and failed to germinate (detail in section 7.2.5.4). However, following *in vivo* assays for tube growth and aniline blue staining of pollen tubes slight differences were found in the pollinations between mutant lines as the pollen donor and male-sterile A9 in all lines. Some A9 ovules were left unfertilized (white arrow) when restricted pollination was performed using pollen from heterozygous mutant plants from all lines except N826057 (At4g20325 – unknown protein) (Figure 7.12c, d) whilst in pollinations using wt pollen all ovules were fertilised and developing properly – this data suggests a fairly trivial problem on the male side for these lines.

7.2.4.7 Morphological assessment of female gametophytic defects

Following seed set counting morphological analyses were carried out of female gametophytes obtained from the lines that demonstrated reduced seed set in order to investigate the cause of the mutant phenotype. In all 8 lines demonstrating a reduced seed set phenotype, some of the unfertilised excised ovules were found to have an undeveloped embryo sac lacking the egg cell, synergid cells, and central cell. The numbers of defective ovules in each line appeared to correspond broadly to the number of affected seeds observed previously - see section 7.2.4.1 (Figure 7.13). In most lines these abnormal embryo sacs were found to be contained in normally shaped ovules (sporophytic tissues). However, in N527524 (SALK_027524 [At5g53820 – unknown protein]) and N574693 (SALK_074693 [At3g50910 – unknown protein]) the ovule structure as well as the embryo sac appeared to be aberrant suggesting that the mutation was sporophytic in nature (Figure 7.13).

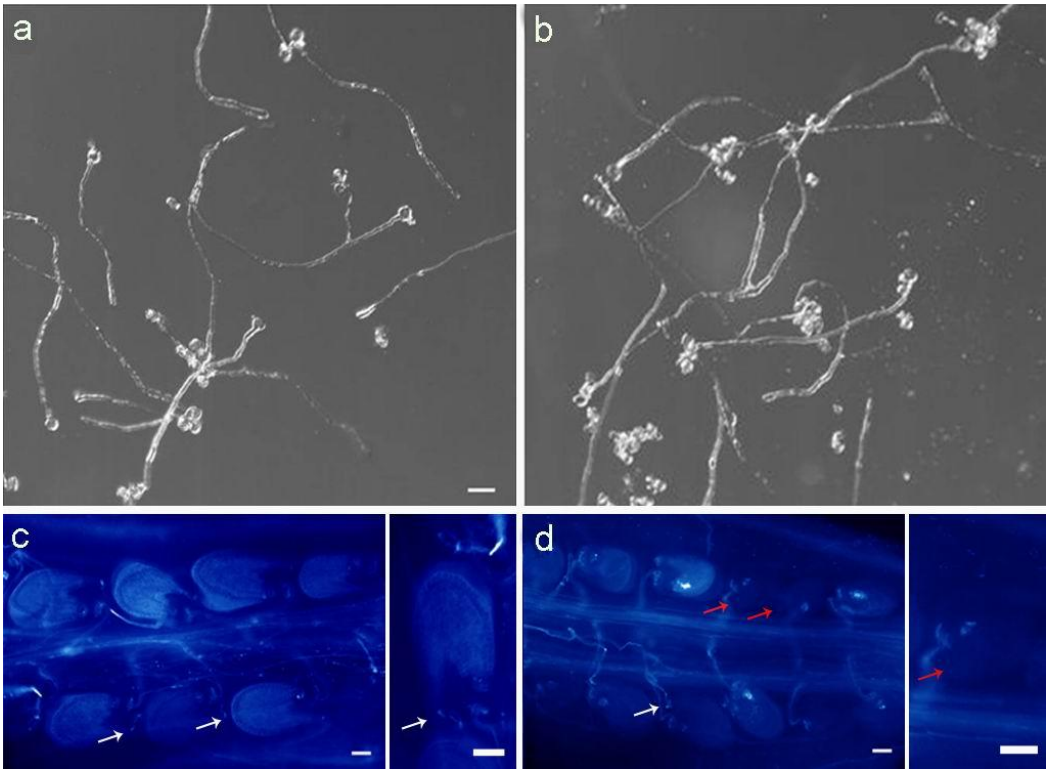


Figure 7.12. *In vivo* and *in vitro* pollen germination and pollen tube growth assays for mutant lines.

In vitro pollen germination assays of one example of mutant pollen (b) were performed in comparison with wild-type pollen (a) (line N564439 is used here to illustrate typical results for the lines under study). The pollen grains were germinated on glass slides coated with GM agar media for 5 hours under controlled conditions. Germination rates for both wild-type and mutant lines were not significantly different (paired-difference t-test, $P>0.05$). *In vivo* pollen tube growth was performed on A9-barnase male sterile flowers. A small quantity of pollen from each of the mutants (Hz) and wild-type was transferred onto A9 stigmas to assess pollen germination and growth and competition effects. Wild-type pollen (c) was able to fertilise all the ovules in an A9 silique (arrow), whereas the pollen from N564439 (d) failed to fertilise all the ovules (unfertilised ovules – red arrow; fertilised ovules – white arrow). Bar = 50µm.

Figure 7.13. Morphological examination of excised ovules from mutant T-DNA insertion lines.

Ovules derived from each plant line demonstrating the reduced seed set phenotype were excised from carpels and observed with DIC microscopy (×40). Most of the mutant lines demonstrated aberrant embryo sacs inside fully developed ovules (a). However, SALK_027524 and SALK_074693 lines demonstrated aberrant embryo sacs inside poorly developed sporophytic ovule tissues. Figures 'b' and 'c' (magnified from 'b') display a normal female gametophyte present in the same siliques as the mutants shown in 'a'. However, in SALK_091193, all embryo sacs were undeveloped and in this case a wt ovule is presented in 'b' and 'c' for comparison. White arrows - synergid cells; black arrow - central cell; black arrow head - egg cell; bar = 20µm.

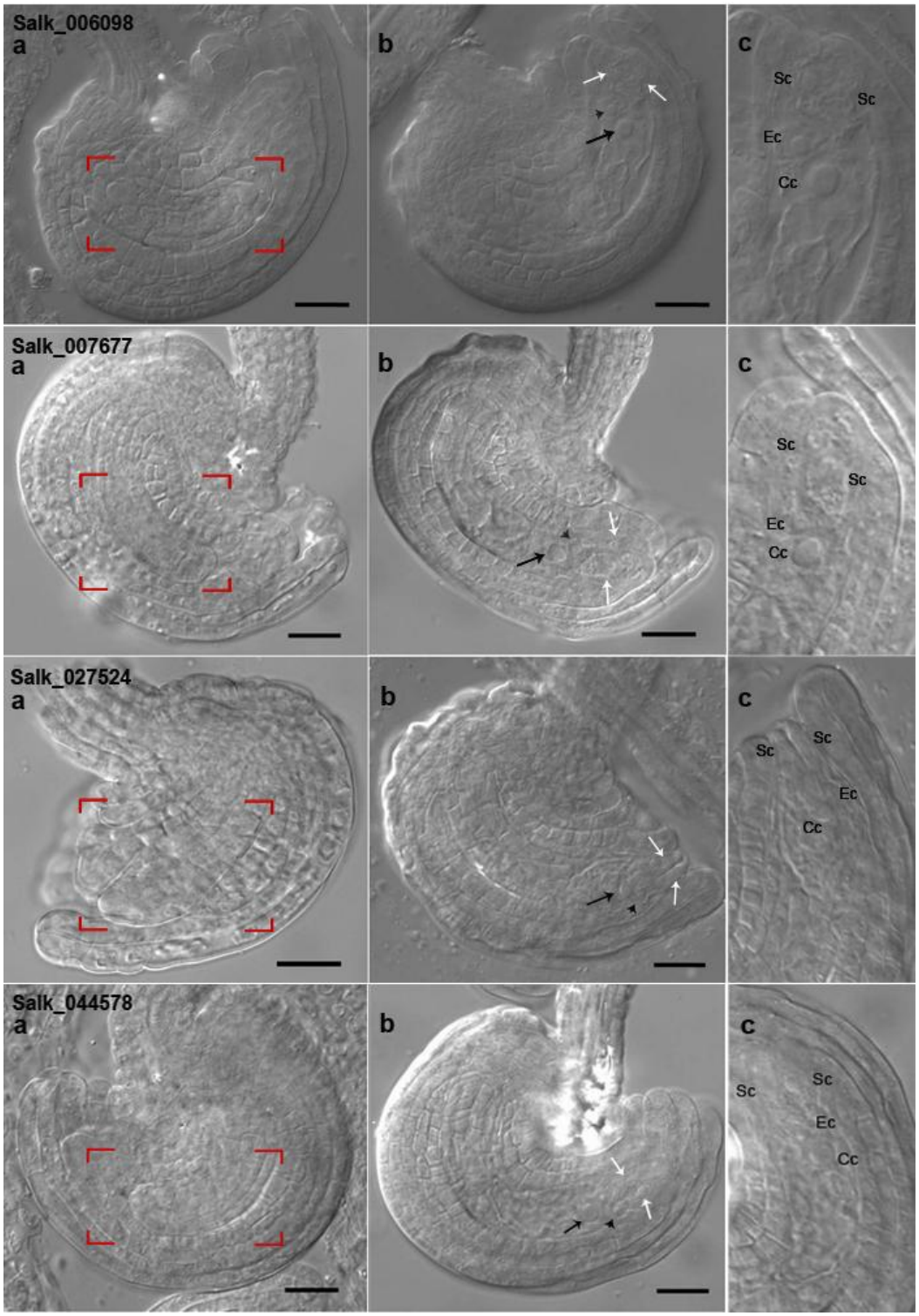
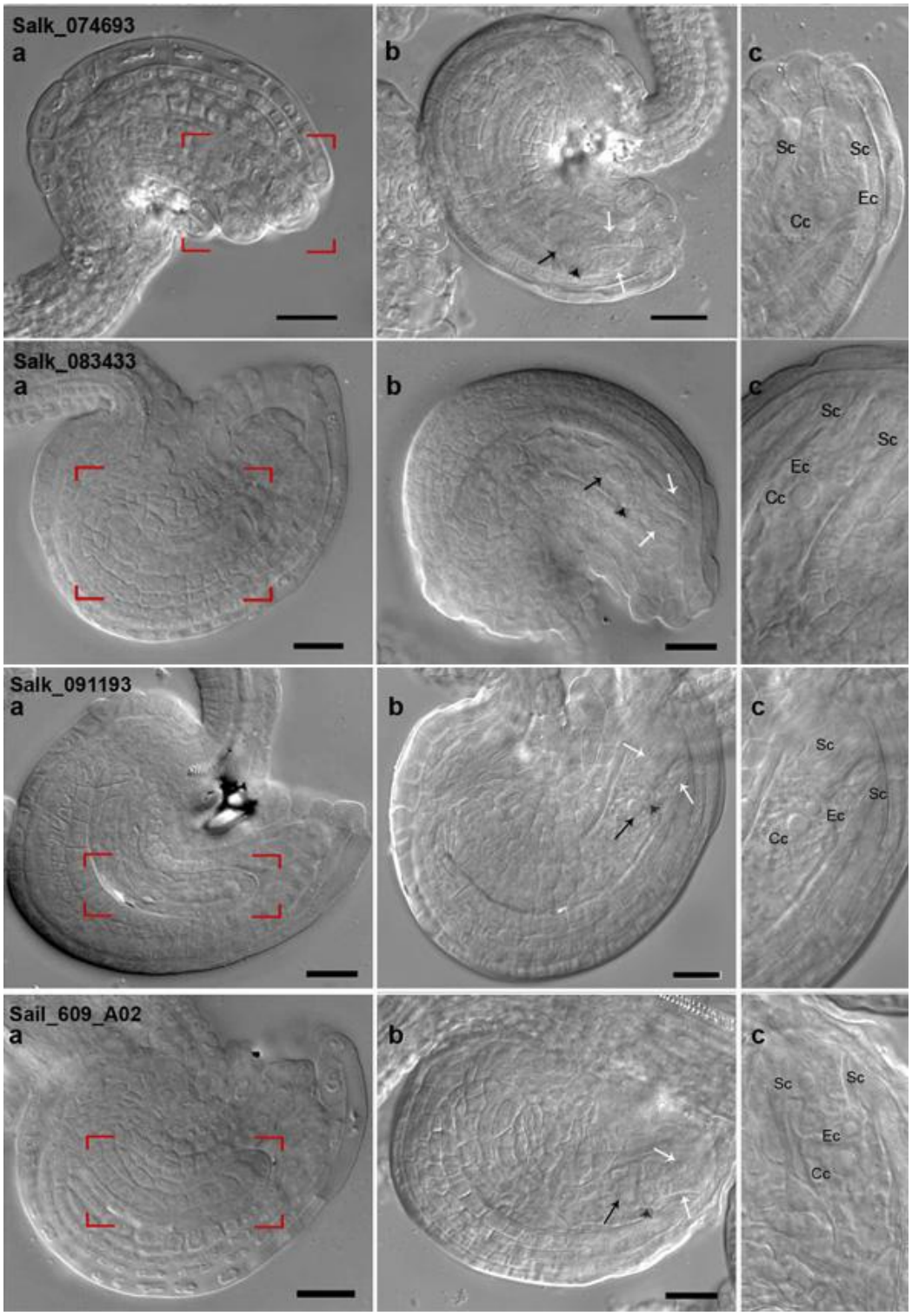

Salk_006098
a
b
c
Sc
Sc
Ec
Cc
Salk_007677
a
b
c
Sc
Sc
Ec
Cc
Salk_027524
a
b
c
Sc
Sc
Ec
Cc
Salk_044578
a
b
c
Sc
Sc
Ec
Cc

Figure 7.13. Morphological examination of excised ovules from mutant T-DNA insertion lines (continued).


Salk_074693
a
b
c
Sc
Sc
Ec
Cc
Salk_083433
a
b
c
Sc
Sc
Ec
Cc
Salk_091193
a
b
c
Sc
Sc
Ec
Cc
Sail_609_A02
a
b
c
Sc
Sc
Ec
Cc

7.2.5 Detailed phenotypic analyses of other selected T-DNA insertion lines demonstrating mutant reproductive phenotypes

All selected lines list in table 7.2 and lines with T-DNA insertions in the high priority gene candidates (Table 7.3) were investigated in detail and the results are summarised in the following sections.

7.2.5.1 High priority sperm cell-expressed candidates (continued from chapter 6)

A total of 37 T-DNA lines representing 20 high priorities candidate sperm cell-expressed genes (mentioned in chapter 5 and 6) were investigated (Table 7.3). The results obtained can be categorised in three groups; **i)** lines that demonstrated a reduced seed set phenotype and contained a single T-DNA insertion - of 37 lines, 10 lines representing 7 genes demonstrated a reduced seed set phenotype (yellow highlight). However, only 2 insertion lines, N506098 (At2g20440) and N574693 (At3g50910) were analysed further (see section 7.2.5.2 and 7.2.5.6 respectively) as these contained a single T-DNA copy (pink highlight); **ii)** lines that had a reduced seed set phenotype but where the mutant phenotype did not consistently cosegregate with the genotype - the remaining 8 lines from 'i' could not be studied in a straightforward manner as more than one copy of T-DNA was detected indicated by inconsistency between phenotype and genotype. This inconsistency suggested that the phenotype was caused by mutation of another unknown T-DNA-containing gene (these lines were not studied further), **iii)** lines that demonstrated segregation distortion – a segregation distortion assay detected two lines N507150 (At1g73070) and N877896 (At1g78300) to have gametophytic mutations (green highlight). Unfortunately PCR failed to detect a T-DNA insertion in both of these lines despite multiple attempts. Therefore these two lines required more studies and will not be described in this report

Two genes known to be expressed in sperm cells (At1g10090 and At5g39650) as demonstrated by GFP fusion (described in chapter 6), were found to have normal phenotypes in their T-DNA lines (blue highlight). Unfortunately the segregation distortion assay was unable to be performed due to the silencing of Kan resistance marker. RT-PCR was then performed in the N631951 line (At1g10090) to confirm the gene knockout. Moreover, this gene transcript was detected by RT-PCR (Figure 7.14) suggesting that this T-DNA line failed to knock out the At1g10090 gene. Unfortunately as At5g39650 was identified to be expressed specifically in the sperm cells late in this study, only phenotypic analysis and segregation distortion assay were able to be performed for the T-DNA insertion lines N504068 and N806939. Further studies

essential for confirming the function of these two sperm cell-expressed genes are still ongoing and will be reported in the near future.

Table 7.3. Summary of T-DNA insertion analyses of high priority sperm-expressed gene candidates.

Lines carrying a T-DNA in the high priority sperm cell-expressed gene candidates were investigated. Plants were scored for the reduced seed set phenotype and segregation distortion. Genotyping was then carried out on lines that demonstrated either a reduced seed set phenotype (yellow highlight) or segregation distortion (tested by χ^2 [$P < 0.05$ being significantly different from 3:1]; green highlight). T-DNA lines for two genes that demonstrated expression in the sperm cells (blue fill - see detail in chapter 6) did not present an obvious phenotype and could not be screened for segregation distortion as the Kan^R selection marker was silent in these lines. Only 2 lines demonstrated a phenotype that corresponded to the genotype and these were thus selected for further detailed analyses (pink fill).

T-DNA Line	At number	Description	P	%R	G	T-DNA presence	G & P corresp-ondence
N531697	At4g32830	ATAURORA1	Y1	S	Y	Y	N
N612121	At4g32830	ATAURORA1	Y1	S	Y	Y	N
N406238	At3g20190	leucine-rich repeat transmembrane kinase	N	76%	N	–	–
N612791	At3g20190	leucine-rich repeat transmembrane kinase	N	S	N	–	–
N643106	At4g11920	WD-40 repeat family protein	Y1	S	Y	N	N/A
N501978	At4g11920	WD-40 repeat family protein	N	S	Y	Y	N/A
N573708	At4g11920	WD-40 repeat family protein	N	79%	Y	Y	N/A
N877491	At2g22740	SUVH6	N	90%	N	–	–
N879899	At2g22740	SUVH6	N	96%	N	–	–
N504068	At5g39650	unknown protein	N	S	N	–	–
N806939	At5g39650	unknown protein	N	HM	N	–	–
N631951	At1g10090	unknown protein	N	S	Y	Y	N/A
N574693	At3g50910	unknown protein	Y3	S	Y	Y	Y
N632810	At3g50910	unknown protein	Y1	S	N	–	–
N611440	At5g23130	peptidoglycan-binding LysM domain-containing protein	Y1	HM	Y	Y	N
N595724	At1g73070	leucine-rich repeat family protein	N	93%	N	–	–
N507150	At1g73070	leucine-rich repeat family protein	N	72%[b]	Y	N	N/A
N503164	At2g29940	ABC transporter family protein	N	76%	Y	N	N/A
N503174	At2g29940	ABC transporter family protein	N	S	N	N	N/A
N349088	At2g29940	ABC transporter family protein	N	75%	N	–	–
N592811	At3g23610	DsPTP1	N	80%	Y	Y	N/A
N877896	At1g78300	14-3-3 protein GF14 omega	N	72%[b]	Y	N	N/A
N503063	At4g28000	AAA-type ATPase family protein	N	75%	Y	Y	N/A
N574465	At4g28000	AAA-type ATPase family protein	N	HM	Y	Y	N/A
N801802	At4g28000	AAA-type ATPase family protein	N	85%	N	–	–
N540809	At5g54640	histone H2A	N	HM	Y	Y	N/A
N340039	At1g11250	SYP95	N	HM	N	–	–
N811141	At2g19380	RNA recognition motif (RRM)-containing protein	Y2	HM	Y	Y	N
N333297	At1g32400	senescence-associated family protein	N	HM	N	–	–
N524905	At3g61230	LIM domain-containing protein	N	S	N	–	–
N171235	At3g61230	LIM domain-containing protein	N	78%	N	–	–

N506096	At2g20440	RabGAP/TBC domain-containing protein	Y1	S	Y	N	N/A
N506098	At2g20440	RabGAP/TBC domain-containing protein	Y1	S	Y	Y	Y
N639822	At1g69960	PP2A5	Y1	S	Y	Y	N
N513178	At1g69960	PP2A5	N	S	N	–	–
N509229	At5g50930	unknown protein	N	HM	Y	Y	N/A
N509951	At5g50930	unknown protein	N	HM	N	–	–

S: silenced kanamycin selection marker, G: genotype, P: phenotype, R: resistant seedling, Y: yes, N: no, Y1: class 1 phenotype (unfertilised ovule), Y2: class 2 phenotype (aborted seed), Y3: class 3 phenotype (minute silique), N/A: unable to be identified due to PCR problems, –: unable to be identified as not genotyped, HM: homozygous, b: not significantly different from 1:1 (χ^2, $P > 0.05$).

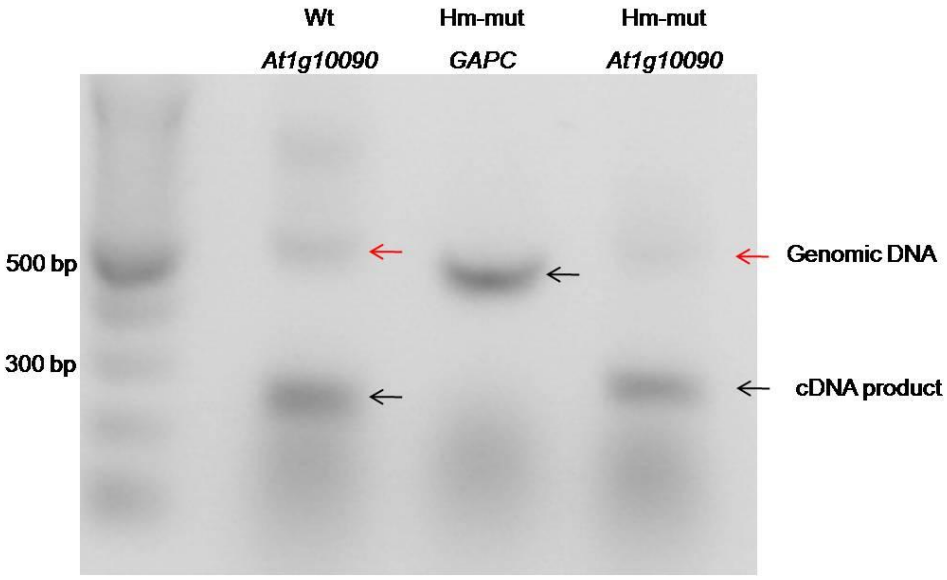


Figure 7.14. RT-PCR analysis of SALK_131951 (At1g10090 - unknown protein).

RT-PCR was performed using primers spaning exon 3-5 and RNA template derived from flowers of SALK_131951 (At1g10090 - unknown protein) homozygous for the T-DNA insertion in order to determine whether gene function was disrupted. The T-DNA insertion in this line was reported in the seventh exon. The result demonstrated that the At1g10090 gene was capable of producing transcripts as the expected 285 bp RT-PCR product (arrow) was detected indicated that either this line produces truncated At1g10090 protein or the T-DNA does not disrupt the gene. *GAPC* was included as an RNA positive control. The upper bands represent PCR amplification of genomic DNA remaining in the RNA template.

7.2.5.2 SALK_006098 (At2g20440 – RabGAP/TBC domain-containing protein)

SALK_006098 – an insertion line in exon 7 of At2g20440 (RabGAP/TBC domain-containing protein) demonstrated a reduced seed set phenotype (50%) but only in the heterozygous condition. In all other aspects the mutant was similar to wild-type in both the homozygous and heterozygous conditions. The numbers of seeds produced remained at 50% when crossed with wt pollen suggesting the fertility problem resided on the female side. Dissection and microscopical observation of female tissues confirmed this as undeveloped embryo sacs lacking the egg cell, central cell, and synergids were observed (Figure 7.13 SALK_006098a). Aniline blue staining demonstrated that the defective embryo sacs were also unable to attract pollen tubes (Figure 7.10 b). Pollen viability tests and *in vitro* germination assays demonstrated normal male gametophyte development. However, the transmission of this gene by pollen was slightly defective as shown by the less than 1:1 transmission ratio (Table 7.2) and thus suggested a slight problem with pollen tube growth. Moreover, restricted pollinations on male-sterile A9 plants coupled with aniline blue staining (Figure 7.12) and seed counting (Figure 7.8) supported the conclusion that there was a slight problem on the paternal side. The proposed function of this gene will be discussed later in this chapter.

7.2.5.3 SALK_007677 (At1g26750 – unknown protein)

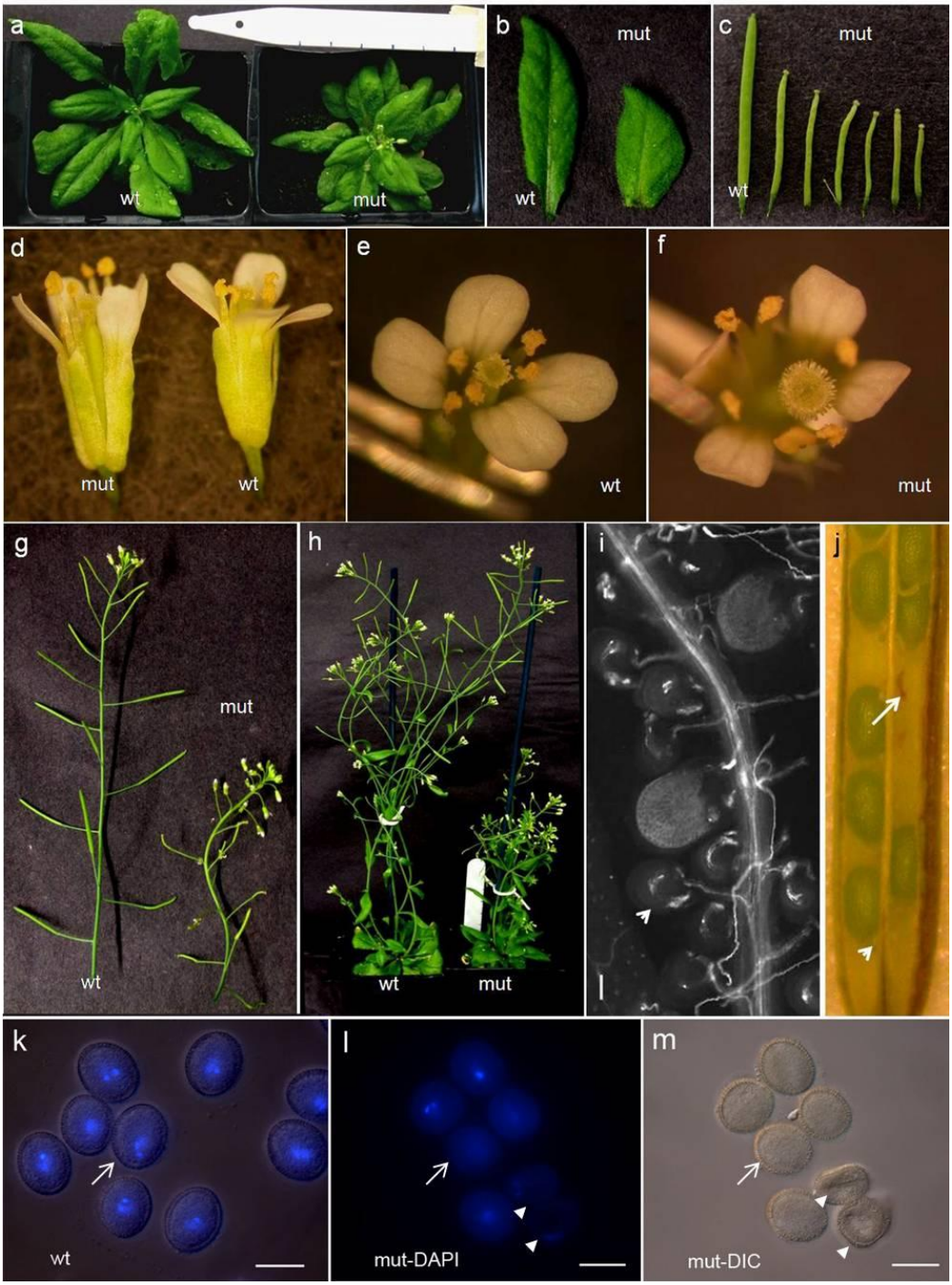
Similar to SALK_006098 described above this insertion line of At1g26750 (unknown protein) demonstrated 50% seed set but only in plants heterozygous for the T-DNA insertion (Table 7.2). Again in all other respects the mutant looked similar to wt plants. Taken together, reciprocal crosses with male-sterile A9 plants coupled with aniline blue staining (Figure 7.12), seed counting (Figure 7.8) and transmission efficiency assays suggested that both male and female gametophytes were affected by the mutation (Table 7.2). Pollen tube staining with aniline blue of the self-pollinated mutant demonstrated normal pollen tube growth along the transmitting tract but failure of tubes to enter the micropyle of aberrant ovules (Figure 7.10 b). Microscopical analysis of excised ovules revealed the absence of important female structures i.e. egg cell, central cell, and synergids (Figure 7.13 SALK_007677a) and also confirmed the problem on the female side. Microarray data from Gene Atlas demonstrated high levels of expression of this gene in stamens (Figure 7.7 – Gene Atlas) as well as in other tissues.

7.2.5.4 SALK_027524 (At5g53820 – unknown protein)

The T-DNA insertion for this line was reported to be sited in the promoter region of the gene (At5g53820). This gene is identified as an unknown protein but is similar to a pollen coat protein of *B. oleracea* (TAIR). Microarray analysis of this gene demonstrated expression to be specific to the stamen and pollen (Figure 7.9 – Gene Atlas). Homozygous plants of this line appeared to be smaller than wild-type (Figure 7.15 a, b). Rosette leaves were closely packed, rounded and shortened in comparison to wild-type. Inflorescences were shorter (Figure 7.15g, h) and the position of anthers was higher than that of the stigma in the mutant whereas anthers of wt plants were generally slightly lower than the stigma (Figure 7.15d, e, f). Moreover 15% of the pollen released from the anther was misshapen (Figure 7.15 l, m), however DAPI staining demonstrated that these pollen grains still contained a vegetative nucleus and two sperm cells. Pollen viability was tested positive with FDA (no different than wt). Reciprocal crosses and transmission efficiency assays demonstrated transmission defects on both male and female sides. Pollen tube staining with aniline blue also confirmed the female effect on the reduced seed set phenotype with no tubes detected entering ovules that failed to produce seeds (Figure 7.15i). Seed set analysis demonstrated a wide range of viable seed numbers in different sized siliques (Figure 7.15c). Both unfertilized ovules and seeds aborted at an early stage were found in the siliques (Figure 7.15j). Microscopical examination of excised ovules revealed many undeveloped ovules with serious imperfections (Figure 7.13 SALK_027524a) - all important structures of the ovule i.e. egg cell, central cell, and synergids were absent. Plants heterozygous for the T-DNA insertion appeared similar to wild-type.

Figure 7.15. N527524 (At5g53820 – unknown protein) phenotypic analysis.

Homozygous plants for this line had a clearly observable gross phenotype. Plants were smaller than wild-type (a, h) and rosette leaves were closely packed, rounded and short in comparison to wild-type (b). Inflorescences were also markedly shorter (g, h). The position of the anthers was higher than the stigma (d-mut, f) whereas anthers of wild-type plants (d-wt, e) were generally slightly lower than the stigma. Approximately 15% of pollen grains released from stamen were misshapen but still contained a vegetative nucleus and two sperm cells (k; wt pollen, l; mut pollen DAPI, m; mut pollen DIC). Self pollinations, reciprocal crosses and transmission efficiency assays demonstrated reproductive problems on both the male and female sides. Pollen tube staining with aniline blue confirmed the female effect on the reduced seed set phenotype. In crosses with wild-type pollen the same percentage of unfertilized ovules (i arrow head) was found as for self pollinated plants. None of the pollen tubes appeared to enter ovules that failed to develop into seeds (i arrow head). Seed set analysis (j) demonstrated a wide range in the number of viable seeds in different sized siliques (c). Both unfertilized (j arrow head) and early aborted seeds (j arrow) were found in the siliques (bar = 50 μm).


a
wt
mut
b
mut
wt
c
mut
wt
d
mut
wt
e
wt
f
mut
g
mut
wt
h
wt
mut
i
l
j
k
wt
l
mut-DAPI
m
mut-DIC

7.2.5.5 SALK_044578 (At3g51030 – thioredoxin H-type 1 [TRX-H-1])

The T-DNA in this insertion line is located in a 3'UTR region of the gene At3g51030. This gene encodes a thioredoxin H-type 1 protein. Microarray analysis demonstrated high levels of expression of this gene in stamen and pollen (see Figure 7.7). Only heterozygous plants in this line showed a reduced seed set phenotype. Other than the reduced seed set the mutant plant looked similar to wt plants. Reciprocal crosses between the heterozygous mutant (mut-hz) and wild-type pollen and male-sterile A9 plants with mut-hz pollen demonstrated a problem with the female gametophyte. Pollen tube staining with aniline blue also confirmed a problem with female function as the pollen tubes were failed to enter approximately half the ovules. Microscopical examination of ovules revealed the absence of important female structures i.e. egg cell, central cell, and synergids (Figure 7.12 SALK_044578a). Transmission efficiency assays and restricted pollinations also demonstrated transmission problem via the male gamete.

7.2.5.6 SALK_074693 (At3g50910 – unknown protein)

Only homozygous (hm) plants in this line appeared significantly smaller than wild-type (wt) (Figure 7.16a, h) while heterozygous (hz) plants appeared normal. Rosette leaves of the hm mutant were wrinkled and shortened in comparison to wt (Figure 7.16b). Inflorescences were shorter (Figure 7.16g, h) and the position of an anther was below the stigma (Figure 7.16d, e, and f) thus making self-pollination less effective. Nonetheless, pollen was released normally from anthers of the mutant plant. DAPI stained pollen showed the presence of normal sperm cells and the vegetative nucleus positioned in the middle of the grain. Pollen viability was tested positive with FDA (no different than wt). Seed set analysis revealed a wide range of viable seed numbers in different size siliques (Figure 7.16c). Unfertilized ovules, early and late aborted seeds were found in the siliques (Figure 7.16j). Sporophytic tissues surrounding the embryo sac appeared undersized and structures of the female gametophyte i.e. egg cell, central cell, and synergid cells were absent from the embryo sac (Figure 7.13 SALK_074693a). Reciprocal crosses, seed set counting, aniline blue staining, and transmission efficiency assay demonstrated transmission problems on both parental sides. The T-DNA in this line was reported to be inserted in the first exon of At3g50910. This gene is identified as an unknown protein with an unknown function. Microarray data from Gene Atlas revealed high expression of this gene in pollen with much lower expression of the gene in other tissues (Figure 7.7). This mutation considered as sporophytic gene mutation

affecting development. The possible function of this gene will be discussed in the discussion part of this chapter.

7.2.5.7 SALK_083433 (At3g53750 – actin3 [ACT3])

Interestingly only heterozygous plants in this line demonstrated the reduced seed set phenotype. The overall morphology of the mutant plants look similar to wt. Reciprocal crosses between heterozygous mutant plants (mut-hz) with wild-type pollen and male sterile A9 with mut-hz pollen demonstrated reduced seed set when the mutant line acted as the mother. Ovules lacked important female gametophytic structures i.e. egg cell, synergids, central cell (Figure 7.12 SALK_083433a). The transmission efficiency assay and restricted pollination on male-sterile A9 plants coupled with aniline blue staining also suggested a slight problem on the male side. The T-DNA in this insertion line was reported to be inserted in the third exon of the actin3 gene (At3g53750). Microarray analysis demonstrated high expression levels of this gene in mature pollen (Figure 7.7)

7.2.5.8 SALK_091193 (At5g05490 – cohesin family protein [SYN1])

Only homozygous plants exhibited a reduced seed set phenotype (Figure 7.17f, g, j) and mutants looked similar to wild-type. However the stamens of the mutant were short and much lower than the level of the stigma (Figure 7.17a, b, c). Moreover, pollen was seldom released from these stamens (Figure 7.17e) and additionally seeds were rarely found in the siliques of mutants (only one seed was found in twenty siliques screened in this study). DAPI and FDA staining was not performed as no pollen was produced for testing. Reciprocal crosses with wt also demonstrated transmission deficiency on the female side (Figure 7.17h, i). Pollen tube staining with aniline blue also confirmed that wt tubes were unable to locate ovules confirming a serious female defect. Analysis of ovules revealed the absence of the egg cell, central cell and synergids (Figure 7.13 SALK_091193a). The T-DNA in this insertion line was reported to be located in the sixth exon of At5g05490. This gene product is described as a cohesin family protein (SYN1) essential for meiosis (TAIR). Microarray analysis demonstrated high expression levels of this gene in stamens relative to other tissues (Figure 7.7).

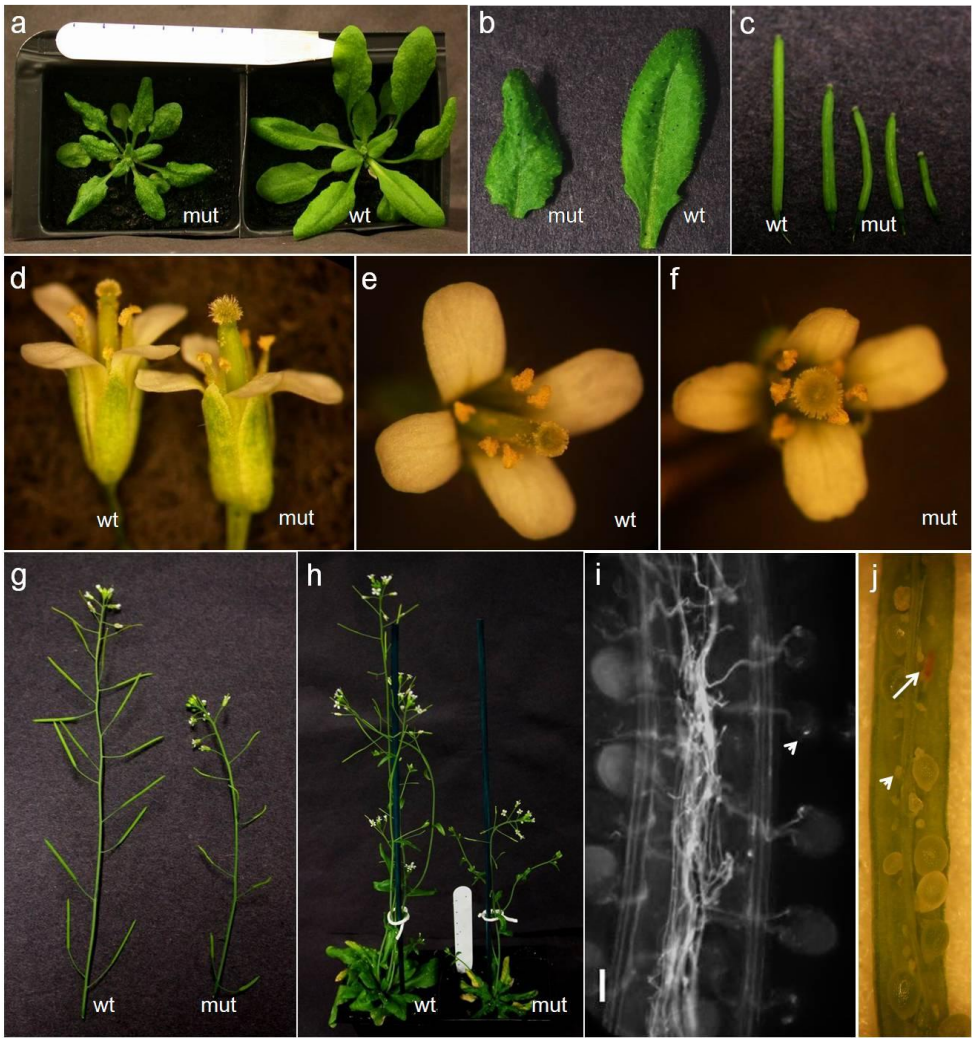


Figure 7.16. Phenotypic analysis of T-DNA line N574693 (At3g50910 – unknown protein).

Homozygous plants of this line had an obvious whole plant phenotype. The size of the plant was appeared smaller than wild-type (a, h). Rosette leaves were wrinkled and short in comparison to wild-type (b). Inflorescences were shorter (g, h) and the positioning of anthers was unusual being much shorter than stigma (d-mut, f) where as anthers of wild-type (d-wt, e) plants was only slightly lower than stigma. Self pollination and reciprocal crossing demonstrated transmission deficiency on both male and female. Pollen tube staining with aniline blue confirmed the female effect on the reduced seed set phenotype as tubes were not attracted to ovules that failed to set seed. In crosses with wild-type as pollen donor, the same percentage of unfertilized ovules (i arrow head) was found as in self pollinated plants. None of the pollen tubes appeared to enter the unfertilized ovules (i arrow head). Seed set analysis (j) demonstrated a wide range of viable seed number in different size siliques (c). Both unfertilized (j arrow head) and early aborted seeds (j arrow) were found in the silique (bar = 50 µm).

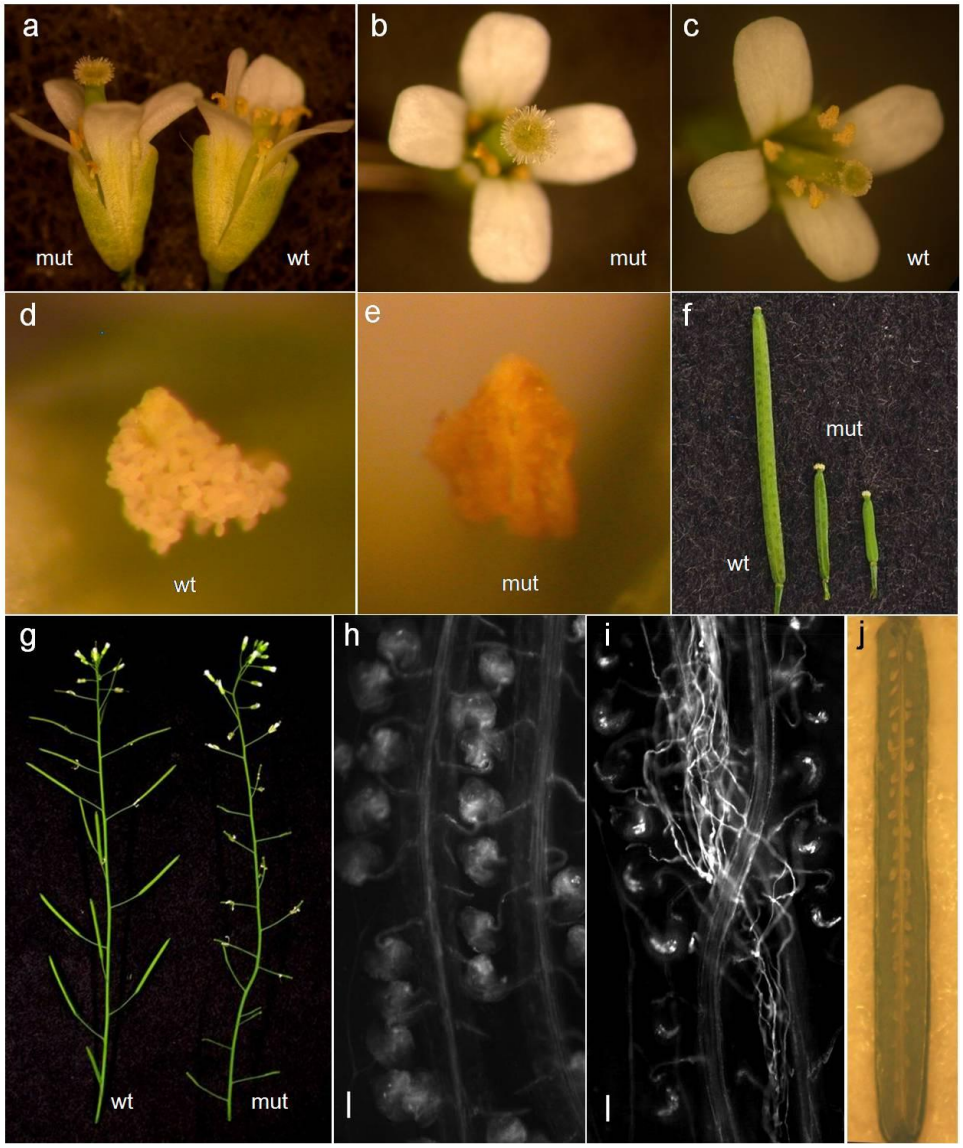


Figure 7.17. Phenotypic analysis of N591193 (At5g05490 – cohesin family protein [SYN1]).

Only homozygous plants of this line gave the reduced seed set phenotype. The overall appearance of the mutant plant was similar to wild-type plant however, stamens of the mutant (a-mut, b) were much lower than the stigma in comparison to wild-type (a-wt, c). Moreover, no pollen was produced from these stamens (e). Siliques were shorter than wild-type (f, g) and no seeds developed within them (j). No evidence of the pollen was found on the stigma of mutant plants and no pollen tubes could be observed in the transmitting tract (h, bar = 50µm) when stained with aniline blue. A Crosse between mut-hm with wild-type pollen demonstrated the defect female side as wt pollen tubes were unable to find the ovules of the mutant (i) confirming the severe problem on the female.

7.2.5.9 SAIL_609_A02 (At4g20325 – unknown protein)

The appearance of the mutant plant in this line was similar to wt. However, the number of seeds in siliques was reduced by 65% but only in plants heterozygous for the T-DNA insertion (data not shown). Reciprocal crosses between mut-hz plants with wild-type pollen and male sterile A9 with pollen derived from mut-hz plants demonstrated transmission deficiency on the female side only. Pollen tube staining with aniline blue also confirmed a problem with female reproductive structures with tubes seen to be failing to locate unfertilised ovules (data not shown). Microscopical examination of excised ovules revealed the absence of important female structures i.e. the egg cell, central cell and synergid cells (Figure 7.13 SAIL_609_A02a). This insertion line was initially chosen as it was located in an unannotated putative sperm cell-expressed gene region. This region was identified from bioinformatics approach mentioned in chapter 5. However, this region was recently annotated as a gene (At4g20325 - unknown protein) similar to Os04g0476400 [Oryza sativa] (NCBI). The T-DNA is located in an intron of the predicted gene (TAIR).

7.2.5.10 SALK_064439 (At3g18040 – mitogen-activated protein kinase 9 [MPK9])

The T-DNA insertion line for *MPK9* resulted in a clear male gametophytic defect. Seed set from this lines was identified as normal and no other morphological differences were detected in comparison to wild-type. However, segregation ratios were distorted from $3Kan^R$:$1Kan^S$ to $1.08Kan^R$:$1Kan^S$ suggesting gametophytic mutation (Table 7.1). Moreover, the results were confirmed by reciprocal crosses with wild-type in which transmission efficiency was measured when mutant plants acted as the pollen donor (paternal side) (Table 7.2). DAPI and FDA demonstrated that the pollen grains of this mutant line are normal and viable (Figure 7.11). Pollen germination rate and tube growth length of this mutant line (*in vitro*) were similar to wt (paired-difference t-test, $P>0.05$) (Figure 7.12). The T-DNA for SALK_064439 was inserted in the second exon of At3g18040. The product of this gene is predicted to be involved in signal transduction and located in mitochondria (TAIR). This gene was identified to be expressed in every tissue but particularly highly in pollen (Gene Atlas – see Figure 7.7).

7.2.5.11 SALK_131193 (At3g23870 – permease-related protein)

T-DNA disruption of this gene also resulted in a male gametophytic mutant phenotype that demonstrated similar trends of segregation ratio (Table 7.1) and transmission efficiency (Table 7.2) as SALK_064439 (insertion in *MPK9***)** described above. DAPI and FDA also demonstrated that the pollen gains of this mutant line are normal and viable (Figure 7.11). Pollen germination rate and tube growth length of this mutant line (*in vitro*) were similar to wt as SALK_064439 mentioned above (Figure 7.12). The SALK_131193 T-DNA was inserted in the third exon of At3g23870 (permease-related) gene of unknown function (TAIR). This gene was also found to be expressed in most plant tissues but highly in pollen (Gene Atlas – Figure 7.7).

7.2.6 Validation of gene knockouts in the investigated T-DNA lines by RT-PCR

Transcript production in T-DNA lines of the nine genes under investigation was investigated by RT-PCR (the RT-PCR was not performed in SALK_091193 (At5g05490 – cohesin family protein [SYN1]) as it has already been confirmed to be a knockout line). Although these genes were known to be highly expressed in pollen expression was also detected in leaves (GeneAtlas – see Figure 7.7) and therefore RNA was extracted from leaves for RT-PCR. The transcripts from these genes were verified to be expressed in leaves by RT-PCR using RNA extracted from wild-type leaves as a template (Figure 7,18b; see primer sequences and position of T-DNA in Appendix A). RT-PCR products of the expected sizes were amplified for all genes tested. RNA was extracted from hm plants from each line and RT-PCR was performed utilising *GAPC* as a positive control (Figure 7.18 a). Half of the plant lines tested, SALK_006098 (At2g20440 – Rab/GAP), SALK_074693 (At3g50910 – unknown protein), SALK_064439 (At3g18040 – MPK9), SAIL_609_A02 (At4g20325 – unknown protein), confirmed clear gene knockouts as no transcript was detected by RT-PCR. Insertion lines reported to disrupt an exon or intron region of the genes were found to result in a gene knockout except for SALK_083433 (At3g53750 – actin3 [ACT3]) where a transcript was still detected despite the T-DNA being located in an exon. However, T-DNA insertions in promoter, 5’ or 3’UTR regions failed to disrupt the gene and transcripts were still detected. These RT-PCR results will be discussed in detail in section 7.3.5.

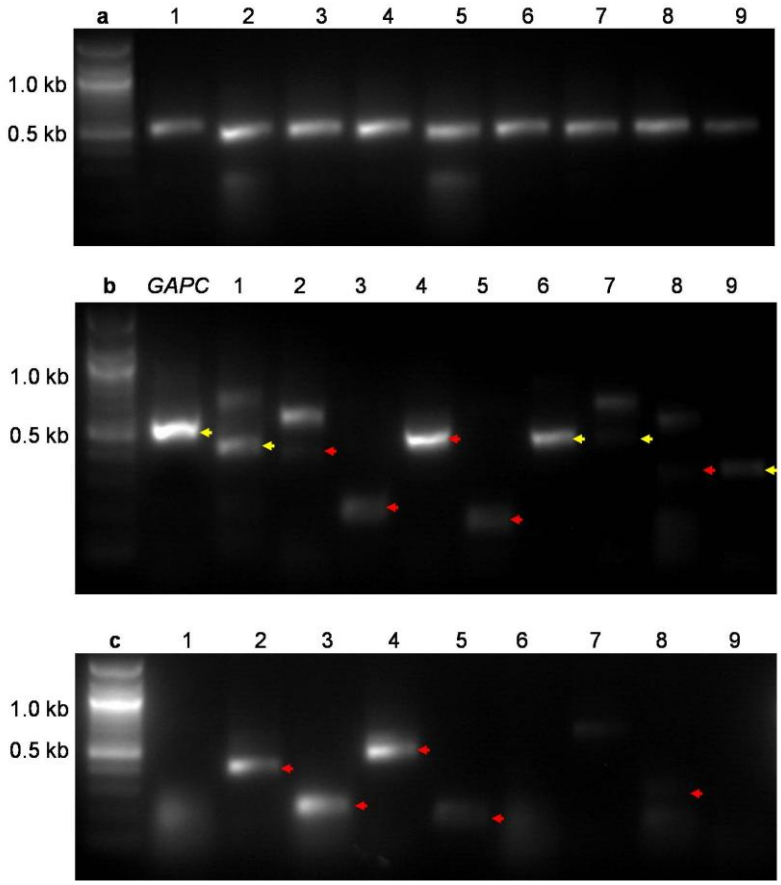


Figure 7.18. RT-PCR demonstrating transcript production in T-DNA insertion homozygous lines.

Transcript production the nine T-DNA lines (except SALK_091193 (SYN1) as it has been reported to be a knockout) was investigated by RT-PCR. RT-PCR primers for each line were tested to be effective using RNA extracted from wt leaves (b; arrow - expected RT-PCR products; yellow arrow - the products that only presence in the wt; red arrow - the products presence in both wt and hm). The RNAs extracted from hm plants from each line were tested utilising *GAPC* primers (a). Only four lines with T-DNA inserted in an exon (SALK_006098 [At2g20440 – Rab/GAP], SALK_074693 [At3g50910 – unknown protein], SALK_064439 [At3g18040 – MPK9], SAIL_609_A02 [At4g20325 – unknown protein]) demonstrated gene disruption as the RT-PCR products were unable to be amplified (1 - SALK_006098 [At2g20440 – Rab/GAP], 2 - SALK_027524 [At5g53820 – unknown protein], 3 – SALK_007677 [At1g26750 – unknown protein], 4 – SALK_083433 [At3g53750 – ACT3], 5 – SALK_044578 [At3g51030 – TRX-h-1], 6 – SALK_074693 [At3g50910 – unknown protein], 7 – SALK_064439 [At3g18040 – MPK9], 8 – SALK-131139 [At3g23870 – permease], 9 - SAIL_609_A02 [At4g20325 – unknown protein]).

7.3 Discussion

Several gametophytic genes were identified by a reverse genetic approach utilising Arabidopsis T-DNA insertion lines. In this study, a semi-infertility phenotype (reduced-seed set phenotype) was initially analysed in order to identify putative sperm cell-defective mutants. However this initial screen failed to identify mutants that had their effects exclusively in sperm cells. Nonetheless, this screen identified several interesting female or male/female gametophytic mutations which will be discussed in detail later in this section. Segregation distortion assays were then utilised to facilitate the identification of further gametophytic genes which the initial screen was unable to identify, particularly mutations affecting the pollen tube. This revised strategy was in part due to the recent report in the literature that the sperm cell-specific gene *HAP2* (*GCS1*) also has an influence on pollen tube growth (Mori et al., 2006; von Besser et al., 2006). Thus, mutants with defects in pollen tube function could also potentially have defects in sperm cell functions. Mutants identified by both these techniques were then subjected to further detailed examination to determine the biological basis of the mutant phenotype.

7.3.1 Initial screen for potential sperm cell-expressed genes by a semi-infertility (reduced-seed set) phenotypic analysis

Due to the large number of plant lines screened for the semi-infertility phenotype, approximately 25 plants per line, where homozygous and heterozygous plants were expected to be included in the population, were grown and initially investigated without genotyping. Of 123 T-DNA insertion lines investigated, 27.6% (34 lines) were found to have a reduced seed set phenotype (Table 7.1). These reduced seed set phenotypes can be divided into 3 classes; ‘unfertilised ovule’, ‘aborted seed’ and ‘minute silique’. Plants in the first class, ‘unfertilised ovule’, were potentially gametophytic mutants because approximately 50% of the ‘seed’ appeared as small white masses similar to undeveloped ovules as described by Howden et al. (Howden et al., 1998). On the other hand, fully developed ovules that were left unfertilised by defective sperm cells would also demonstrate a similar phenotype (Mori et al., 2006). Thus the plant lines that demonstrated this ‘class 1’ phenotype were of special interest. Detailed analyses to confirm the basis for these mutant phenotypes were carried out and will be discussed later in section 7.3.5. The second class, ‘aborted seed’ phenotype was concluded to be sporophytic in nature with the mutation acting post-fertilization on seed development (Boisson et al., 2001) and thus will not be considered further here. The third class ‘minute silique’, a more severe manifestation of the first class, were also considered as potential gametophytic mutations - the known gametophytic gene *SYN1* was also

included in this class (Peirson et al., 1997). This third class of mutant lines was also analysed in detail and will be discussed later in this section.

7.3.2 Identification of potential sperm cell and/or gamete expressed genes by segregation distortion assay

This assay required the seed progeny from a selfed heterozygous plant carrying a single T-DNA insertion in the gene of interest. Therefore, individual kanamycin resistant plants were selected from the mixed F2 seed populations provided by NASC (Nottingham Arabidopsis stock centre). Some of the plants selected for the assay were homozygous for the T-DNA insertion as Kan^R plants were selected at random without genotyping (due to the considerably large number of plant lines investigated), and therefore segregation data could not be generated for these lines..The segregation distortion assay identified at least 19 lines (15%) with segregation ratios of Kan^R being less than 75% (Table 7.1). This could be considered a reasonably large percentage as a previous study reported that only 9% of plants demonstrated segregation distortion (Howden et al., 1998). However, the large proportion of genes that demonstrated segregation distortion in this study is not surprising as these genes are in fact a subset of the genes that matched to maize sperm cell ESTs (see detail in chapter 5). Therefore, most of these sperm gene candidates were naturally expected to be gametophytic genes. Unfortunately, in some lines the Kan^R marker was silent (Table 7.1) and thus the segregation data could not be generated as all seedlings were kanamycin sensitive (Kan^S). For some of these lines where the Kan^R marker was silenced PCR genotyping was performed to identify the genotype of each progeny plant permitting the ratio to be calculated. These lines selected for genotyping were the lines that demonstrated a reduced seed set phenotype, carried a single copy of T-DNA and the phenotype corresponded to the genotype (N506098, N527524, N574693, and N583433; Table. 7.2). Of 19 lines found to have a distorted segregation ratio only 3 lines demonstrated a segregation ratio of approximately 1:1 suggesting that either male or female gametophytes were affected. However for the majority of lines displaying segregation distortion, the ratio deviated from 1:1. The T-DNA lines that demonstrated the Kan^R between 75-50% were expected to be incompletely penetrant male or female gametophytic mutations. On the other hand, where the Kan^R was less than 50% the mutation would be expected to affect both the male and female gametophytes. Of 19 lines, 8 lines carried potential exclusive female gametophytic mutations as the Kan^R was between 75-50%. However, only 2 lines, N564439 (At3g18040 – mitogen-activated protein kinase 9 [MPK9]) (52%) and N631139 (At3g23870 – permease-

related) (37%) were selected to be investigated as these had segregation ratios that deviated markedly from 3:1 (1.08:1 and 0.59:1 respectively). The other 6 lines were reserved for analysis in the future as the segregation ratio did not indicate a severe gametophytic effect and were thus unlikely to be mutations affecting sperm-expressed genes crucial for the gamete fusion event. Detail analysis of these two lines, N564439 (52%) and N631139 (37%) will be described later in section 7.3.5.

7.3.3 Genotypic analysis

Genotypic analysis was performed on 47 lines selected from the high priority gene list detailed in chapter 5, lines with clear phenotypes and lines that demonstrated segregation distortion (Table 7.1). However, PCR genotyping was only able to positively identify a T-DNA in the genes of interest for 31 of these lines. Failure to detect T-DNA in many of these lines by PCR could possibly be due to corrupted T-DNA flanking sequences as occasionally T-DNA insertion leads to disruption of genomic regions close to the insertion point. Consequently, the primer designing tool (iSect Tools – SALK) would be unable to generate a primer to be used in this scenario. Thus new right primers were manually designed to match different flanking region sequences for each line. In addition, instead of using the T-DNA left border primer a right border primer was used with a gene specific left primer to amplify the mutant allele. Unfortunately the mutant allele PCR product was unable to be amplified in any cases. The other reason for these negative results could simply be because the T-DNA was not inserted in the expected gene. These mutations could be analysed further by sequencing the T-DNA flanking region using the appropriate primer pair to recover the unidentified sequence. Alternatively RACE PCR could possibly be used to amplify the gene product using random oligo anchor primers with T-DNA border primers. The other situations where multiple copies of T-DNA were detected in the mutant lines demonstrating reduced seed set phenotype, the single copy T-DNA has to be isolated by backcrossing the plants containing phenotype with the wild-type plant. This should eliminate the excessive copy of T-DNA by segregation. In this study, only 24% (8 lines) of 33 lines that demonstrated semi-infertility phenotype could easily be investigated further. For 16 of these lines a T-DNA insertion could not be detected in the gene of interest. The remaining 9 lines contained multiple T-DNA copies and therefore these lines were regarded as less of a priority here due to the time that would be required for backcrossing to generate single copy lines - thus more straightforward lines were investigated in this project due to the time constraints. This investigation demonstrated a great problem with the available T-DNA insertion lines (SALK and SAIL) as only

24% (8 lines) of those of interest could be characterised. Not only did 27% (9 lines) of these 33 lines carry multiple copies of T-DNA but in the remaining 48% (16 lines) the position of the T-DNA could not be identified and together this proved a big obstacle to this study.

7.3.4 Pollen viability, pollen tube growth and transmission efficiency assays

Pollen viability and pollen tube growth assays demonstrated normal pollen development and viability in comparison to wild-type in all lines investigated except for SALK_027524 (At5g53820 – unknown protein) where 15% of the pollen was misshapen. Normal pollen and pollen tube growth in these lines suggested that the mutation did not dramatically affect pollen development and function despite the fact that problems with male function were detected in other experiments. This discrepancy is probably because subtle changes in the rate of pollen tube germination and growth were difficult to be determined *in vitro* due to the problematic nature of *in vitro* growth of Arabidopsis pollen tubes. The limitations of this assay have also been documented by Boavida and McCormick (2007) who highlighted that many factors can greatly influence the assay. Various pollen germination mediaave been tested for the best results and different laboratory and growth room conditions i.e. temperature, humidity, ventilation, and light also seem to greatly affect the germination and growth rate of the pollen tube (Boavida and McCormick, 2007). As all the lines investigated were expected to have only slight problems with pollen tube growth (as indicated by the relatively small deviations from a 3:1 segregation ratio), the *in vitro* pollen tube assay was deemed unlikely to be precise enough to substantiate the segregation data. Therefore, these lines were investigated further using a transmission efficiency assay to define the problem on the male. This assay allows a precise determination of the transmission efficiency of mutated gene via the male gamete. All lines investigated except N826057 (SAIL_609_A02) demonstrated a slight problem on the male side as the transmission ratio deviated from 1:1 (Kan^R less than 50%; Table 7.2). This slight transmission efficiency reduction via male gamete was also reported in at least two female gamete mutants, *slower walker1 (swa1)* (Shi et al., 2005) and *lachesis (lis)* (Gross-Hardt et al., 2007).

7.3.5 Detailed analysis of T-DNA insertion knockout lines

7.3.5.1 High priority sperm cell-expressed candidates (continue from chapter 6)

Of 37 lines investigated, 10 lines representing 7 genes demonstrated a reduced seed set phenotype. Only 2 insertion lines, SALK_006098 (At2g20440 – RabGAP/TBC domain-containing protein) and SALK_074693 (At3g50910 – unknown protein) lines were able to be analysed further as these had single copy T-DNA insertions. SALK_074693 (At3g50910 – unknown protein) appeared to have a wild-type phenotype and thus, RT-PCR was performed to verify the knockout of this gene. The result demonstrated that an At1g10090 transcript was still produced in this line indicating that the T-DNA insertion failed to completely disrupt the function of the gene. The study of more insertion lines will be required for further analysis of this gene's function. A detailed analysis of SALK_006098 (At2g20440 – RabGAP/TBC domain-containing protein) is described below.

7.3.5.2 SALK_006098 (At2g20440 – RabGAP/TBC domain-containing protein)

The SALK_006098 insertion line demonstrated an approximately 50% reduced seed set phenotype in the heterozygous condition. This line however had two potential insertion sites in the genome; one in the gene of interest At2g20440 (SALK_006098.17.05.x) and the other in At1g08180 (SALK_006098.46.75.x) according to the accession data in TAIR (http://www.arabidopsis.org/) and T-DNA express (http://signal.SALK.edu /cgi-bin/tdnaexpress). At1g08180 is listed as an unknown protein (NCBI) of 12.9 kDa and is constitutively expressed throughout the plant with high levels of transcript in hypocotyl and xylem tissue (Gene Atlas). Taking into account the semi-infertility phenotype of the plant line together with expression pattern of both genes, the observed phenotype was likely to be the result of an insertion in the At2g20440 gene. Importantly RT-PCR analysis of homozygous plants confirmed that the At2g20440 gene was knocked out in this line. At2g20440 encodes a RabGAP/TBC domain-containing protein similar to At4g28550 (TAIR). In this study SALK_006098 demonstrated a probable dominant negative phenotype as heterozygous plants had a reduced seed set phenotype whereas homozygous plants had a wild-type phenotype. The insertion site is in the 7th exon of the gene and thus it is quite possible that a truncated version of the protein is synthesised that results in competition between wild-type and mutant proteins.

Rab GTPases are reported to regulate diverse functions of the cell i.e. regulation of secretory and endocytic pathways and cell signalling (TAIR). Rabs activate specific effector proteins that regulate vesicle tethering and fusion, cargo sorting and cytoskeleton dependent organelle transport (TAIR). The Rab small G protein family are present in all eukaryotic genomes (TAIR) with 57 Rabs predicted in Arabidopsis. They typically localise near organellar membranes and have a general function in regulating vesicle-mediated trafficking (Richardson and Zon 1995; Bernard 2003). Nearly all Rabs possess a TBC domain (Tre17 Bub2 Cdc16) (Richardson and Zon 1995) that is involved in regulating GTPase activity (Bernard 2003). At2g20440 (RabGAP/TBC domain-containing protein) was not only shown to be expressed exclusively in pollen (Gene Atlas) but the SALK_006098 insertion line also demonstrated a slight problem on the male side. The data from segregation ratio analysis, the transmission efficiency assay and seed set from crosses to wt suggested that this gene could possibly have a role in pollen tube growth. Moreover, this gene was also concluded to be involved in female gametophyte development as revealed by the reduced seed set phenotype in reciprocal crosses to wt and ovule morphology. It was also speculated that At2g20440 was expressed in both sporophytic and gametophytic tissue because the phenotype was present only in heterozygotes in a dominant-negative-like fashion. A dominant-negative sporophytic mutation though should affect all ovules and here only 50% of ovules were affected in plants heterozygous for the mutation. Clearly the nature and relationship of genotype and phenotype is complex for this T-DNA line and more experiments are required to confirm this gene function in reproduction. Additional T-DNA insertion plant lines for both genes will be investigated.

7.3.5.3 SALK_027524 (At5g53820 – unknown protein)

The T-DNA in this insertion line was reported to be inserted in a promoter region of the gene At5g53820 as mentioned in the results section. This plant line had a semi-sterile dwarf phenotype in homozygous plants. The mutation did not only affect fertility resulting in reduced seed set but also the overall appearance of the plant structure - therefore this mutation could be acting both sporophytically and gametophytically during vegetative plant development and also during male and female gametogenesis. Examination of excised ovules from homozygous plants revealed undeveloped embryo sacs lacking the egg cell, central cell and synergids thus accounting for the reduced seed set phenotype. Moreover, 15% of the mature pollen was misshapen and unable to germinate however they still contained a vegetative nucleus and two sperm cells.

The At5g53820 gene encodes a small basic (pI=10) unknown protein containing 67 amino acids similar to a *Brassica oleracea* pollen coat-like protein and ABA-inducible protein (TAIR, TIGR, and MATDB). This ABA-inducible protein was described to have a late embryogenesis abundant (LEA) protein domain (TAIR). This domain is normally present in proteins responsible for ABA responses (EBI). It was also predicted to have a signal peptide and to be targeted to chloroplasts (MATDB). Interestingly, ABA is one of the plant hormones that controls plant growth and development (Finkelstein, 1994; Gray, 2004) having a role in seed dormancy (Gray, 2004) and glucose responses during early seedling development (Arenas-Huertero et al., 2000). In most publications reporting on ABA responses in mutant lines, phenotypes were mainly mentioned involving seed germination rate and stomata responses. Interestingly At5g53820 has also been reported to be up-regulated during shoot development (Che et al., 2006) though the expression pattern of this gene investigated using the Gene Atlas tool indicates that this gene is only expressed in anthers and pollen (Figure 7.9) which is perplexing as the phenotype noted in this study affects the overall structure of the plant. This could possibly be explained if the T-DNA insertion in this line (promoter region) did not knock down the gene expression (transcripts were still detected in this T-DNA line by RT-PCR) but altered the temporal and spatial expression levels of the gene. Importantly this T-DNA is also reported as being located in the promoter region of a neighbouring gene (At5g53810) which is a putative O-methyltransferase, reported in Gene Atlas and AtGenExpress to be expressed almost exclusively in flower organs, particularly petals. In order to confirm the function of At5g53820, more T-DNA lines with insertions in this gene or the genes nearby will be required to be investigated. Alternatively an RNAi experiment could be carried out for this gene in order to confirm the phenotype.

7.3.5.4 SALK_ 091193 (At5g05490 – cohesin family protein [SYN1])

Homozygous plants of this line exhibited an infertility phenotype affecting both male and female gametophytes. The T-DNA in this insertion line was reported to be located in the sixth exon of the At5g05490 gene. This gene product is described as a cohesin family protein (SYN1) and reported to be highly expressed in stamens. It has been found to be essential for correct male and female meiosis (Bai et al., 1999). Studies have revealed oversized microspores containing variable amounts of DNA in the polyads (Bai et al., 1999). Microsporocytes in the *syn1* mutant demonstrated defects in chromosome condensation and chromosome pairing (Bai et al., 1999; Peirson et al., 1997). As this mutant has previously been analysed thoroughly and documented to be

involved in the early stages of gametophyte development, the analysis of this gene was not performed in great detail in this study.

7.3.5.5 SALK_074693 (At3g50910 – unknown protein)

This insertion line exhibited reduced seed set and a dwarfism phenotype and it was therefore feasible that the gene mutation acted both sporophytically and gametophytically. The T-DNA in this insertion line was reported to be located in the first exon region of At3g50910. This gene is encodes a 49.6kDa protein of unknown function with a pI of 4.6 (TAIR). Given the mutant phenotype it is clear that the gene may have a general role not only in plant growth and development but also in gametogenesis. Microarray data obtained from Gene Atlas revealed very high levels of expression in pollen though the gene is also expressed elsewhere. Importantly the phenotype for this line was only apparent in homozygous plants. Pollen viability tests and germination assays confirmed that pollen was phenotypically normal. On the other hand examination of excised ovules demonstrated a problem in female gametogenesis. Defective ovules were unable to develop normally and produce normal gamete structures i.e. egg cell, synergid cells, central cell and antipodal cells. However, some ovules in the homozygous plant did develop normally. *In vivo* pollen tube growth assays also confirmed the presence of a female defect - Aniline blue staining of reciprocal crosses of mutant homozygous (mut-hm) plants with wild-type pollen and male sterile A9 with mut-hm pollen demonstrated the abnormal ovules in the mut-hm were failing to attract pollen tubes. Pollen tubes are attracted and guided to the ovule by the sporophytic tissues of the ovule and from the the female gametophyte itself (Johnson and Preuss, 2002; Ray et al., 1997; Rotman et al., 2003; Shimizu, 2000; Wilhelmi and Preuss, 1996). The pollen tubes germinated and elongated normally through the stigma and transmitting tract, however tubes were unable to turn towards the funiculus of ovules that contained undeveloped embryo sacs. The lack of embryo sacs in some ovules was most likely to be a consequence of abnormally developed sporophytic tissue of the ovule surrounding the embryo sac. As mentioned earlier, signals from sporophytic tissue could affect gametogenesis, therefore this mutation is expected to be sporophytic. In addition, not every ovule in the siliques was abnormal suggesting that although the gene function may be totally abolished the phenotype was not fully penetrant Moreover, as the phenotype was only apparent in plants homozygous for the insertion this demonstrated that the mutant allele was fully recovered in the heterozygous condition, which would not be expected if this were a gametophytic gene - only one allele can be expressed in each gamete therefore some mutant ovules would

be expected to be seen in heterozygous plants even if the mutation was not fully penetrant.

7.3.5.6 SALK_ 007677 (At1g26750 – unknown protein)

The T-DNA in this insertion line was reported to be inserted in a 5'UTR region of At1g26750. This gene was predicted to encode a protein of 195 amino acids, pI=10, with a signal peptide for targeting to mitochondria. It was identified as an unknown protein with an unknown function. Protein or nucleotide sequences of this gene have no match to any known protein or gene when blastp/n was performed. Interestingly, the segregation distortion assay suggested a gametophytic function for this gene. However, homozygous plants were identified thus suggested that this gene is not a gametophytic lethal gene that is crucial for reproduction. In addition the 50% reduced seed set phenotype was found only in heterozygous plants. Microscopical examination of excised ovules revealed undeveloped embryo sacs suggesting a probable function in female gametogenesis for this gene. Unlike N574693 discussed earlier, the undeveloped embryo sac was surrounded by normally developed ovule tissue. However, it remains possible that the sporophytic signals that control female gametogenesis were affected. Together with the segregation distortion data, these results suggest that At1g26750 acts either purely gametophytically or is functional in both sporophytic and gametophytic tissues. In addition, pollen tubes were unable to locate the aberrant ovules and enter the micropyle suggesting that the signal released from the ovule to attract the pollen tube was absent. As only heterozygous plants were found to have the reduced seed set phenotype, this gene can't possibly be exclusively expressed in gametophytic tissue. Therefore, it is crucial that southern blot is carried out in order to confirm the number of T-DNAs in this insertion line. The insertion in this line occurs in the 5'UTR region of the gene and thus protein production could be completely abolished by this mutation even though transcripts are still present (as detected by RT-PCR). However, mutations in 5'UTR regions of genes can also cause an increase or decrease in translation efficiency (Signori et al., 2001) or disrupt protein translation. In the case of a complete functional knockout of At1g26750 the phenotype can be explained by a complex inhibitory model (Figure 7.19). In the model, the gene product is hypothesised to be expressed in both gametophytic and sporophytic tissues and to interact with another protein involved in female gametogenesis. These two proteins (A and B) interact in various combinations AA, BB and AB dimers, and each dimer has its own unique function. BB promotes normal female gametophyte development while AB acts as a BB inhibitor. AA inhibits the inhibitory effect of AB on BB. B protein can also function as

an inhibitor for the AA dimer. In the heterozygous mutant situation, production of A protein is reduced and consequently more free B protein is available to inhibit AA production therefore promoting AB dimer. The AB dimer then inhibits BB production thus resulting in the reduced seed set phenotype. On the other hand, no A protein is produced in homozygous plants thus only the BB dimer can be formed in the plant resulting in normal seed set as wild-type. The model for this pathway in the tissues is demonstrated in figure 7.20. These two proteins are produced in both sporophytic and gametophytic tissues and can move from one to the other and form dimers in the gametophytic tissue. In male gamete, a slight disadvantage on the mutant in the pollen tube growth competition was identified. Therefore, this implies that this proposed inhibitory pathway also occurs in the growing pollen tube. However, the impact of the mutation is more severe in the female than in the male tissues.

Nonetheless, more experiments are required in order to verify this proposed hypothesis. The study of more T-DNA insertion lines is crucial to confirm the function of the gene as SALK 007677 could possibly disrupt the adjacent At1g26740 (structural constituent of ribosome) gene in the 5'UTR region. Moreover, backcrossing of the mutant with wild-type for a few generations should also be utilised to be certain that a single locus T-DNA insertion is responsible for the observed phenotype.

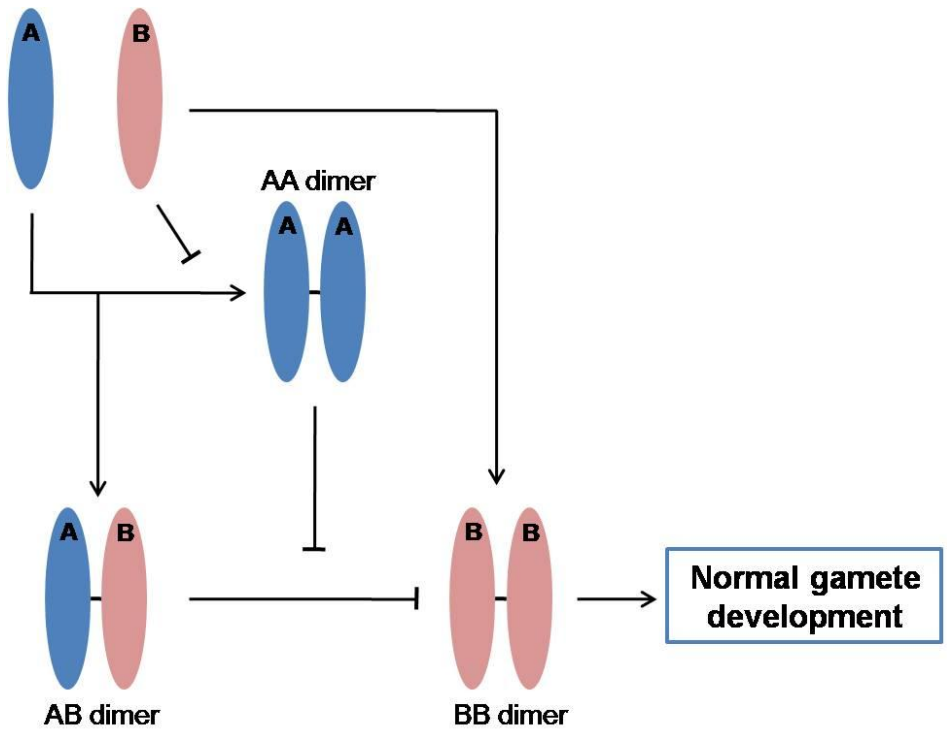


Figure 7.19. Proposed model for the function of the protein product of At1g26750.

This gene product (A) is hypothesised to interact with an unknown protein (B) and to function in female gametogenesis. These two unknown proteins (A and B) interact in various combinations AA, BB and AB dimers, and each dimer has its own unique function. BB promotes normal female gametophyte development while AB acts as a BB inhibitor. AA inhibits the inhibitory effect of AB on BB. B protein can also function as an inhibitor for the AA dimer. In the heterozygous situation, production of A protein is reduced consequently in more free B protein in which inhibit AA production therefore promoting AB dimer. AB dimer then inhibits BB production thus resulting in reduced seed set phenotype. However, in homozygous condition, no A protein is produced thus only BB dimer can be formed in the plant resulting in normal seed set as wild-type.

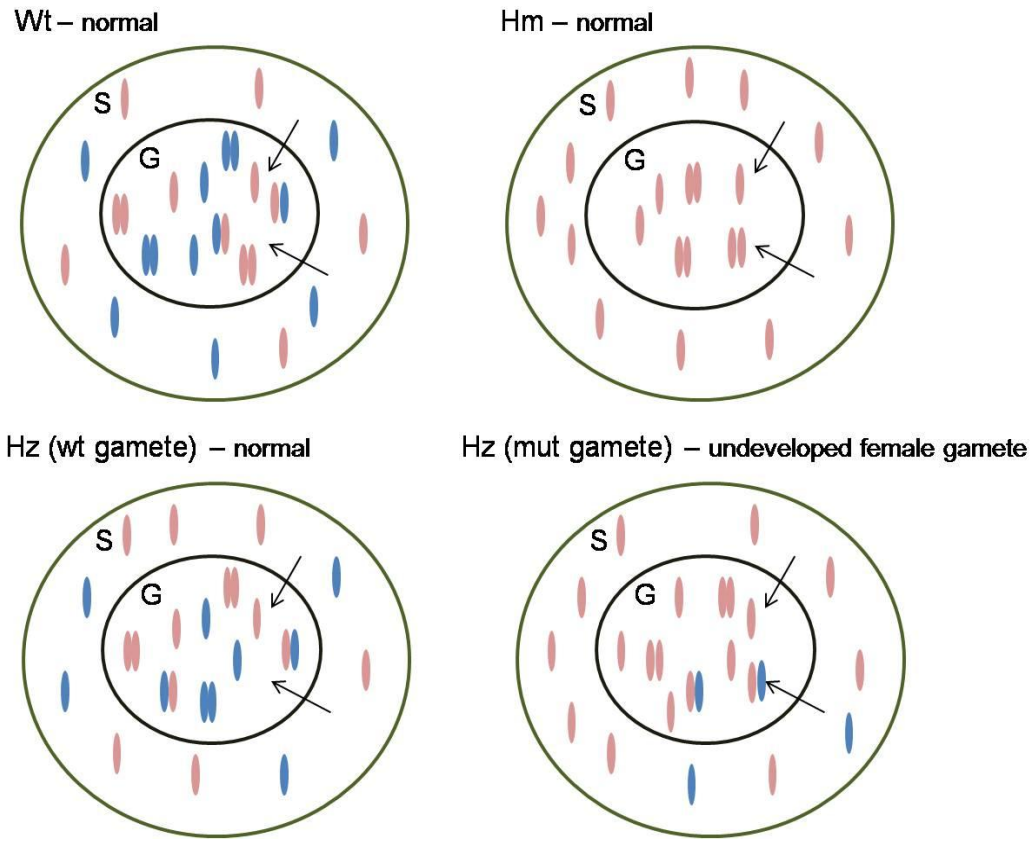


Figure 7.20. Proposed model for the protein product of At1g26750 in sporophytic and gametophytic tissues.

These two proteins are produced in both sporophytic and gametophytic tissues. However they were transported to form dimers in the central gametophytic tissue. In wild-type plants all three forms of dimers are balanced thus providing normal BB function leading to normal gamete development. In the homozygous mutant condition only B and BB are produced therefore the AB dimer is unable to inhibit BB function resulting in normal gamete development. However, in heterozygous mutants, excess B production promotes the formation of the AB dimer therefore inhibiting BB and thus leading to reduced seed set phenotype.

7.3.5.7 SALK_083433 (At3g53750 – ACTIN3 [ACT3])

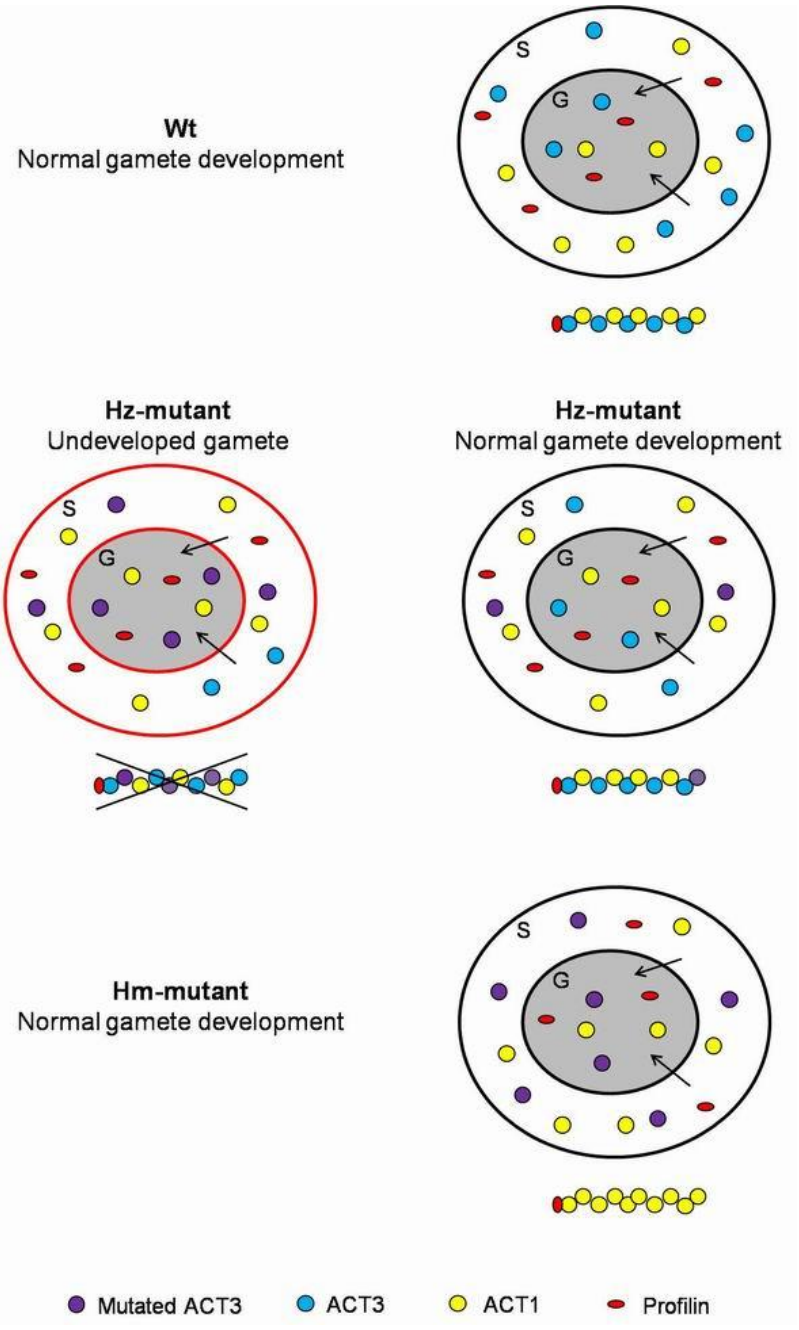
This insertion line demonstrated a 50% reduced seed set phenotype again only in plants heterozygous for the mutation. The T-DNA in this insertion line was expected to disrupt the third exon of the At3g53750 ACTIN3 (ACT3) gene. The protein produced in this mutant was predicted to be disrupted/truncated from close to amino acid 303 of 377 according to the insertion line information (TAIR). 3D analysis of the structure of actin has demonstrated that actin binds profilin at Glu-364 on the actin polypetide (Nyman et al., 2002) and this region is expected to be absent in this T-DNA insertion line. Microarray data obtained from the Gene Atlas tool (Genevestigator) demonstrated high expression of this gene in mature pollen. Moreover, it is also thought to be expressed during pollen germination according to data deposited on the TAIR database and to be involved in cytoskeleton organisation and biogenesis (TAIR).

ACT1 and *ACT3* encode nearly identical proteins (42 kDa) with only one amino acid difference in the conserved region and they are also members of the same actin subclass (An et al., 1996). The expression patterns of these two genes are also very similar being very highly expressed in mature pollen and pollen tubes as demonstrated by RNA-gel blot hybridization, RT-PCR and GUS-gene fusion (An et al., 1996). The 8 actin genes found in Arabidopsis has been divided into 5 subclasses based on phylogeny and 2 subclasses – vegetative and reproductive, based on their expression patterns (An et al., 1996; Kandasamy et al., 1999). The reproductive actins which are expressed strongly in pollen are composed of three subclasses: subclass 3 – ACT11, subclass 4 – ACT1 and ACT3, and subclass 5 – ACT4 and ACT12 (Kandasamy et al., 1999). ACT11 in particular, was found to be expressed in ovules and embryonic tissues (Kandasamy et al., 1999); McDowell et al., 1996; Meagher et al., 1999b). These actins are the major components of the cytoskeleton which are involved in cell division, cell wall deposition, organelle movement, nuclear positioning, transport, cytoplasmic streaming, tip growth and polysome distribution (Kandasamy et al., 1999). The major function of the reproductive actins that is of interest here is their involvement in pollen germination and tube growth (Steer and Steer, 1989; Steer, 1990; Mascarenhas, 1993). More importantly, they were also shown to be involved in positioning of sperm cells and the vegetative cell nucleus during pollen tube growth and to facilitate sperm discharge from the tube (Perdue and Parthasarathy, 1985; Pierson et al., 1986; Lancelle et al., 1987). However, ACT1-GUS and ACT3-GUS not only demonstrated preferential expression in pollen and ovules but also around meristematic tissues which supports studies describing a role of actin in cell division (An et al., 1996). Actins are also important in defining the position and formation of cell division planes (An et al., 1996).

The reduced seed set phenotype obtained from *ACT3* mutation is consistent with the proposed reproductive function for actin mentioned above. The phenotype, segregation ratio data and the results from the transmission assay suggested a problem on both male and female sides. The pollen tube growth rate of the heterozygous mutant was slightly affected as demonstrated in restricted pollinations coupled with aniline blue staining, where some ovules were left unfertilised. Moreover, examination of excised ovules suggested the mutation also affected the female gametophyte. As the reduced seed set phenotype was present only in plants heterozygous for the mutation a proposed mechanism explaining this phenomenon in this line is presented in figure 7.21 In wild-type plants, ACT3, ACT1, and profilin are produced in both gametophytic and sporophytic tissues. These three proteins are transported into the gametophytic tissue to polymerise the F-actin filament. Profilin promotes ATP exchange for polymerisation in both actins. In the heterozygous condition, ACT3 is mutated in the profilin binding region This mutated ACT3 generated in both sporophytic and gametophytic tissues is proposed to have high affinity binding to the wild-type ACT3 and to become incorporated into the ACT3-ACT1 filament consequently generating a non functional actin filament. In a plant homozygous for the mutation only ACT1 can form F-actin as mut-ACT3 is unable to interact directly with ACT1 without wt-ACT3.

Figure 7.21. Proposed model for ACT3-ACT1 polymerisation and the SALK_083433 phenotype.

SALK_083433 which carries a T-DNA in an exon of the act3 gene (At3g53750) demonstrated a reduced seed set phenotype in the heterozygous condition. The mutation was predicted to disrupt the profilin binding region of ACT3 and consequently alter its ability to bind ACT1. This mutated ACT3 (mut-ACT3) has a high affinity for wild-type ACT3 which therefore interferes with the F-actin chain polymerisation. This severely aberrant F-actin chain is unable to function, leading to undeveloped gametes. In the 50% of normal female gametophytes in plants heterozygous for the mutation, mut-ACT3 is produced only in sporophytic tissue thus the actin filament is generated mostly form wt-ACT3 and ACT1 with few mut-ACT3 molecules being present which cannot disrupt the actin filament function. In the homozygous mutant condition, only mut-ACT3 and ACT1 are produced. These two proteins are unable to form an F-actin chain thus ACT1 begins to polymerise an ACT1 chain in place of an ACT3ACT1 chain thus masking the reduced seed set phenotype (Wt; wild-type, Hm; homozygous, Hz; heterozygous, S; sporophytic tissue, G; gametophytic tissue).


Wt
Normal gamete development
S
G
Hz-mutant
Undeveloped gamete
Hz-mutant
Normal gamete development
Hm-mutant
Normal gamete development
Mutated ACT3
ACT3
ACT1
Profilin

7.3.5.8 SALK_044578 (At3g51030 – thioredoxin H-type 1)

This T-DNA line demonstrated approximately 50% reduced seed set again only in the heterozygous condition. The phenotype of the plant, segregation ratio, and transmission efficiency of the T-DNA via male gamete suggested that the mutation affected both male and female gametes. The T-DNA in this insertion line was reported to be inserted in the 3' UTR region of the At3g51030 gene which is predicted to encode a cytosolic thioredoxin H-type 1 protein (12.6 kDa, pI = 5.7). Microarray analysis (Gene Atlas) revealed high expression levels of this gene in stamens and pollen.

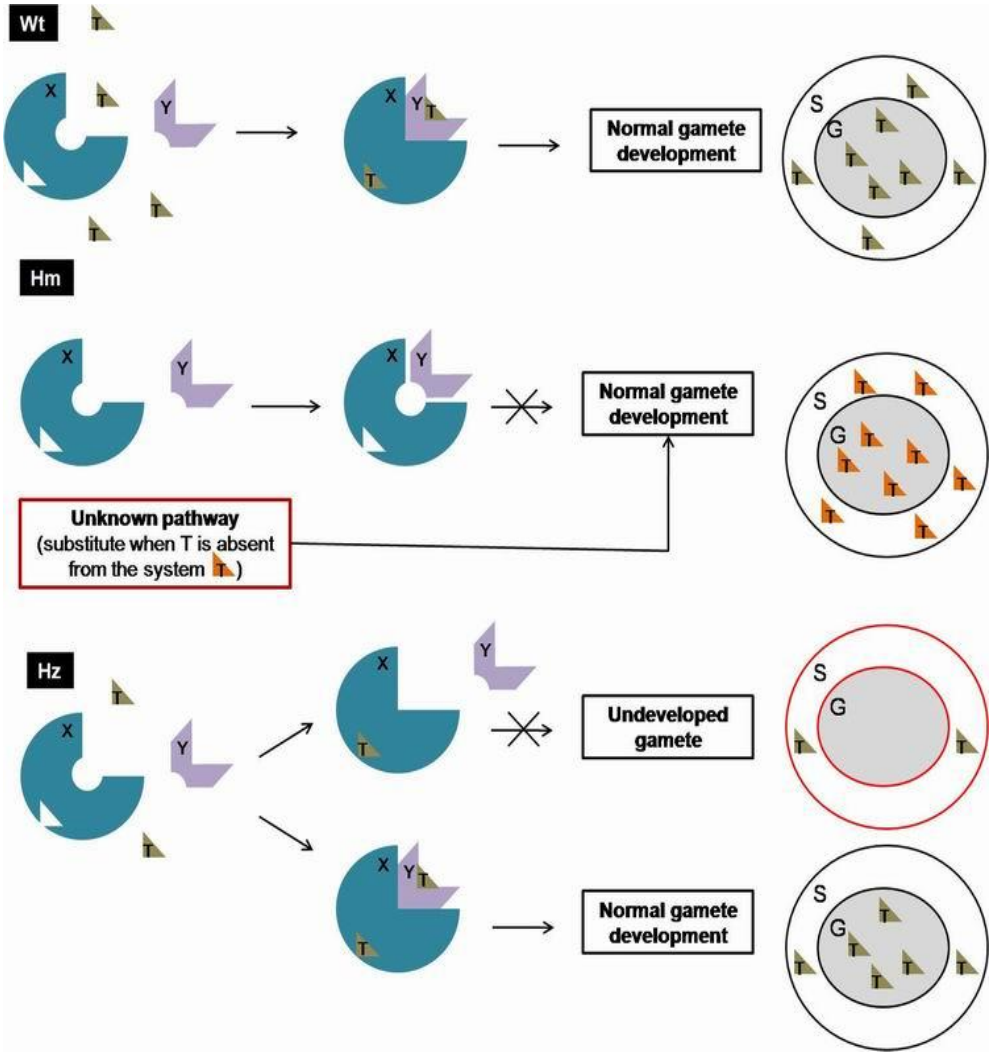
Thioredoxins are small (105-120 aa) proteins containing a conserved sequence Trp-Cys-Gly-Pro-Cys or Cys-Pro-Pro-Cys (Rivera-Madrid et al., 1995). Two thioredoxin systems are known in higher plants. Thioredoxin m and f, which are reduced by a ferredoxin-dependent thioredoxin reductase, regulate dark/light-related cycles and operate in the chloroplasts. The second system, thioredoxin h, is reduced by an NADP-dependent thioredoxin reductase (Rivera-Madrid et al., 1995). Thioredoxin is ubiquitously expressed and likely to have a general role in plant function (Bower et al., 1996; Reichheld et al., 2002). In its reduced form it can function as a hydrogen donor or as a regulatory factor for various target proteins like metabolic enzymes, redox proteins, transcription factors or MAP kinases (Reichheld et al., 2002). There are 8 thioredoxin H-type proteins identified in Arabidopsis, each of which targets different protein subsets (Peterson et al., 2005). AtTRXh1, AtTRXh2, and AtTRXh4 were found to be expressed at a high level in the flower compared to other organs (Reichheld et al., 2002). AtTRXh1 was proposed to be the only thioredoxin in Arabidopsis that is capable of plasmodesmal translocation (Peterson et al., 2005). THL-1 and THL-2 in *Brassica oleracea* were discovered to be involved in self-incompatibility and ubiquitously expressed in many organs but highly in floral tissue (Bower et al., 1996). THL-1 inhibits SRK autophosphorylation in the absent of the ligand (SCR from pollen coat) (Cabrillac et al., 2001). However, the self-incompatibility mechanism does not exist in Arabidopsis. Nevertheless, AtTRXh could still have an important function in the floral organs and possibly be involved in reproduction.

Similar to mutations in 5'UTRs, 3'UTR mutations can trigger conformational changes in the mRNA resulting in alteration of translation efficiency which could be decreased, increased or completely abolished (Kuersten and Goodwin, 2003; Mazumder et al., 2003). In the case where translation is completely terminated and given that *TRXh1* gene is expressed in both sporophytic and gametophytic tissues, in plants heterozygous for the mutation half of the protein is produced ovules will still produce protein and develop normally. However, the other half contained mutant gametes the production of

TRXh1 is very low and unable to initiate down stream mechanism resulting in undeveloped gametophyte. On the other hand, in the homozygous condition no protein is produced but due to redundancy of thioredoxin genes in Arabidopsis, the mutant phenotype is consequently 'rescued' by another trx gene. Moreover, the presence of TRXh1 protein in wild-type plants might also act to suppress the expression or function of the complementary TRX. This proposed trx gene function can be tested by investigating other T-DNA insertion lines of the other trx genes and through the use of lines carrying multiple trx knockouts. Another possible model that could explain this complex phenotype is summarised in Figure 7.22 where TRXh1 binds to and causes a conformational change in two target proteins, X and Y, that promotes their interaction which in turn can promote normal gametophyte development. However, TRXh1 has preferential binding to the X protein over the Y protein. In the homozygous condition X and Y can loosely interact but are unable to function and regulate downstream events leading to normal gamete development. However in the homozygous mutant condition, another pathway controlling identical downstream developmental processes can substitute the function of this pathway when AtTRXh1 is absent from the system. Thus in this scenario normal gamete development is occurs in the homozygous condition. However, in the heterozygous condition where only half of TRXh1 is produced and is therefore limiting, only X can interact with TRXh1 which undergoes a conformational change but results in either a loss or no interaction with Y leading to the reduced seed set phenotype (Figure7.22).

Figure 7.22. Protein interaction model explaining the mutant phenotype of SALK_044578 (At3g51030 – thioredoxin H-type 1).

Mutated At3g51030 – AtTRXh1 demonstrated approximately 50% reduced seed set in plants heterozygous for the T-DNA insertion but normal seed set in homozygous plants. As the phenotype is only present in hz condition, this gene was hypothesised to be expressed in both gametophytic and sporophytic tissues. The phenotype can be explained by a protein interaction model. AtTRXh1 interacts with X and Y proteins (with a preference for X over Y) and induces a conformational change to enhance their interaction and bring about activation of the XY complex which in turn regulates downstream events leading to normal gamete development. X and Y can also form a complex when AtTRXh1 is absent (also occurs in Hm), however, this complex is non-functional and cannot regulate normal gamete development. Importantly an alternative pathway controlling the same developmental process can substitute the function of XY when AtTRXh1 is entirely absent from the system. Therefore, normal gamete development is identified in the homozygous condition. In heterozygous plants (Hz), only half of the AtTRXh1 is produced and binds preferentially to X causing its conformational change. This X-AtTRXh1 is unable to interact with Y thus consequently leads to undeveloped gamete (Wt; wild-type, Hm; homozygous, Hz; heterozygous, T; AtTRXh1, X; X protein, Y; Y protein).


Wt
X
Y
T
Normal gamete development
S
G
Hm
Unknown pathway
(substitute when T is absent from the system T)
Hz
Undeveloped gamete

7.3.5.9 SAIL_609_A02 (At4g20325 – unknown protein)

This initially unannoted region was predicted to be probable gene according to results obtained from the *in silico* comparative study described in chapter 5. Indeed this region was later annotated as At4g20325 on the Genbank (NCBI) database. This T-DNA insertion line carries a T-DNA in an intron of the gene. This gene is predicted to contain an RNase H domain. The RNase H enzymes have functions in the hydrolysis of RNA when annealed to complementary DNA and are present in all living organisms (NCBI). This gene is ubiquitously expressed in all tissue types in Arabidopsis according to the microarray data (Gene Atlas). The insertion line demonstrated a reduced seed set phenotype in plants homozygous for the T-DNA insertion. According to segregation and transmission assays this gene was expected to act sporophytically and to be indirectly involved in female gametophyte development.

7.3.5.10 Potential male gametophytic gene mutant line – SALK_064439 (At3g18040 – mitogen-activated protein kinase 9 [MPK9])

The phenotype of plants both homozygous and heterozygous for the T-DNA insertion appeared as wild-type. However, segregation and transmission ratios were lower than the normal 3:1 and 1:1 respectively suggesting that the mutation affected the male gametophyte. Moreover, reciprocal crosses with male-sterile A9 also demonstrated a problem on the male side. The T-DNA in this line is inserted in the second exon of At3g18040 – mitogen-activated protein kinase 9 (MPK9). This gene encodes a 58.4 kDa basic protein with a pI of 8.6 (TAIR). Mitogen-activated-protein kinases (MAPKs) play an important role in plant signal transduction and are involved in many regulating pathways e.g. stress responses, plant hormone responses, the cell cycle and various developmental processes (Wrzaczek and Hirt, 2001). MAP kinase cascades consist of three kinase modules MAP kinase kinase kinase (MAPKKK), MAP kinase kinase (MKK) and MAP kinase (MPK) (Alzwiy and Morris, 2007; Wrzaczek and Hirt, 2001). These three kinases work in correlations and sequentially phosphorylate the next kinases. The last kinases – MPKs are responsible for phospholylation of cytosolic and nuclear proteins (Alzwiy and Morris, 2007). In Arabidopsis, 24 MPKs have been identified and subdivided into three main sub families PERKα, PERKβ, and PERKγ (Wrzaczek and Hirt, 2001). MAPKs have been detected in pollen and hypothesised to be involved in pollen development and the regulation of germination (Heberle-Bors et al., 2001). The expression of the tobacco MAP kinase, ntf4, was discovered to be restricted to pollen and seeds (Heberle-Bors et al., 2001; Voronin et al., 2001). MAP kinase was found to be transcribed during pollen development in the mid bicellular

stage. The translated protein was discovered to be stored in an inactive state in mature pollen grains and activated upon hydration (Wilson et al., 1997). However, the authors also hypothesised that the MAP kinase is activated during the mid bicellular stage and deactivated when the pollen is fully mature and reactivated upon hydration (Wilson et al., 1997). Therefore, it is possible that the same MAP kinase cascade is responsible for both pollen dehydration and rehydration during maturation and germination. However, there is no evidence in our mutant line supporting this hypothesis as all pollen grains appeared normally developed and hydrated normally in all experiments i.e FDA and DAPI staining and. *in vitro*/*in vivo* pollen germination assays,. The data obtained here is in broad agreement with an antisense experiment to reduce the production of MAP kinase in tobacco which resulted in a slight acceleration of the timing of first pollen mitosis but no change to the germination frequency and the function of the pollen (Heberle-Bors et al., 2001). Nonetheless in this study not only were some of the ovules left unfertilised in the restrict pollination experiments with male-sterile A9 but some pollen tubes were also unable to find their targets thus indicating that this MAPK might also be involved in pollen tube growth and/or signalling between the male and female gametophytes at a later stage. In summary, studies on this mutant line suggested that the MPK9 gene has an important function in the male gametophyte possibly linked to pollen tube growth and perhaps guidance. Further experiments are required to investigate the function of this gene. As there are many MPKs present in Arabidopsis the investigation might be problematic due to redundancy issues. Therefore, multiple gene knockouts are required in the study. Moreover, other T-DNA lines in the same gene might also be useful for the investigation.

7.3.5.11 Potential male gametophytic gene mutant line – SALK_131193 (At3g23870 – permease-related)

The T-DNA insertion in this line was located in the third exon of At3g23870, a permease-related gene encoding a 36 kDa basic protein (pI 9) (TAIR). This protein is predicted to contain 9 transmembrane regions (MIPS). Permeases have a general function in the cell to transport different molecules across membranes (either in or out of the cell) depending on the types of permeases e.g. amino acid permease, auxin permease, and glycerol permease in or out of the cell (TAIR). However, this permease-related gene was not placed in any particular class. The phenotype of the plant was very subtle and provided no clues as to the specific function of this putative permease. Only segregation distortion and transmission efficiency assays revealed a problem with the male gametophyte in this mutant line, however, no obvious pollen phenotype has been

identified. To elucidate and confirm the role of this gene in reproduction more data needs to be obtained, in particular other mutant lines for this gene should be studied to validate the observed phenotype.

7.3.5.12 Other T-DNA lines demonstrating a reduced seed set phenotype

Unfortunately, over 50% of the T-DNA insertion lines investigated either carried multiple T-DNA copies or demonstrated a reduced seed set phenotype but were unable to be investigated further due to an inability to positively confirm the presence of the T-DNA by PCR. For the lines that carried multiple copies of T-DNA, backcrossing of these lines with wild-type could be performed in order to isolate single copy T-DNA lines for each gene of interest. In some of the lines where a T-DNA product could not be amplified by PCR, but are worthy of further investigation, other alternative T-DNA lines could be sought if available or alternatively T-DNA sequencing could be performed again in order to identify the exact position of the T-DNA. Lines worthy of further investigation included N503755, N506096, N516334, N517925, N531697, N531941, N643106, N592811, N612121, N629264, N826961, and N828582 (see Table 7.1) as they demonstrated semi-infertility phenotype and/or segregation distortion and are highly expressed in pollen and thus could possibly be important gametophytic genes. Moreover, for some of the candidate genes, the T-DNA insertions in either exons or introns were unable to be identified i.e. At3g20190, At1g11250, At5g04250, At2g22740, At5g39650, and At3g61230. This could possibly be because these genes have a crucial reproductive function and disruption makes transmission of the mutant allele rare but nevertheless highlights these genes as appealing for further studies. For quite a few of the lines studied which did not show a semi-infertility phenotype, the T-DNA insertions were reported to be located in the promoter, 5'UTR or 3'UTR regions. For these lines it would be informative to carry out RT-PCR to determine if the genes were indeed knocked out.. RNAi methodologies could also be utilised to generate the knockout lines and identify the function of these genes if deemed important enough.

Interestingly, despite the objective of this chapter being to identify mutant lines of sperm cell-expressed genes or at least male gametophytic genes and the fact that the these candidates originated from a comparative study to maize sperm ESTs, several female gametophyte mutations were identified in place of male gamete mutations. This intriguing result could simply be because female gametophytic mutations are more easily observed whilst male gametophytic mutations could be easily overlooked due to the large quantities of pollen and the nature of male gamete delivery involving pollen tube growth where in most cases only the fittest can be succeed in fertilising the female.

Segregation ratio analysis of plants heterozygous for a T-DNA mutation and the absence of homozygous mutant plants in population could be used as an indicator for such mutations. However, unfortunately genotyping of the plants in every insertion line was unable to be performed due to the large number of the candidate genes and the significant PCR problems reported earlier. Only two male specific gametophytic mutation lines were detected in comparison to 7 female mutants. However, interestingly, most of these mutations affecting the female gametophyte also had a slight effect on the male gametophyte. This occurrence suggests that dual function of the gamete expressed genes in both male and female is not uncommon.

This T-DNA insertion mutagenesis study is still ongoing as many other interesting mutant lines with a reduced seed set phenotype and unusual segregation ratios have not yet been studied in detail due to the T-DNA insertion location problems mentioned earlier. These obstacles are not impossible to overcome however and it is hoped that these studies will be completed in the near future.

Chapter 8:

General Discussion and Conclusions

8.1 Introduction

Despite extensive studies on plant sperm cells, relatively little is known about this specialised cell type. Due to the fact that plant sperm cells are embedded in another cell, the vegetative cell of the pollen grain, the isolation of sperm cells is problematic. However sperm cell isolation has been successfully developed in various plant species i.e. *Plumbago zeylanica* (Russell, 1986), *Nicotiana tabacum* (Cao et al., 1996), *Brassica oleracea,* and *Zea mays* (Matthys-Rochon et al., 1987). The morphology of sperm cells has also been investigated by light and electron microscopy (McConchie et al., 1987b; Southworth et al., 1988; Southworth et al., 1997; van Aelst et al., 1990; Yu and Russell, 1994b). Sperm cells have been found to typically contain a large nucleus with little cytoplasm and a few organelles. Not only is there an urgent need for more knowledge of the egg-sperm fusion event but also the second fertilisation event between the central cell and remaining sperm. In some plant species such as *Plumbago zeylanica* and *Zea mays* where dimorphic sperm cells are recognized, the question of preferential fusion has been raised. Preferential fusion partners have been discovered in both species however, the reason behind these preferences is still ambiguous (Russell, 1984b; Southworth et al., 1997; Wagner et al., 1989). In addition, molecular analyses of plant sperm cells have also been carried out for several decades. Once more, despite extensive sperm cell research and highly advanced molecular biology techniques only one extensive set of transcript data has been generated from a FACS-purified maize sperm cell population (Engel et al., 2003). However, this one study has provided a vast amount of valuable information on the sperm cell transcriptome. This research has been taken forward in this study and by others in an attempt to identify some sperm-specific gene homologues in the model plant Arabidopsis. To date, three Arabidopsis sperm cell-expressed genes – *GEX1, GEX2* and *GEX3* have been derived from this maize EST library and none of these have proved to be sperm cell-specific (Alandete-Saez et al., 2008; Engel et al., 2005). Therefore, one of the main goals for this project is to identify sperm cell-specific genes

in Arabidopsis utilising the existing sperm cell EST dataset generated from FACS-purified maize sperm cells. Along side this approach, a proteomic study was undertaken to identify proteins in the sperm cell. Proteomic studies are valuable as they have significant advantages over transcriptomic studies and indeed have been successfully carried out for the sperm cells of the fruit fly, *Drosophila*, using new proteomic technique, whole-sperm mass spectrometry (WSMS) (Dorus et al., 2006). Together with these two approaches, various molecular techniques i.e. RT-PCR, promoter-GFP fusion and T-DNA insertion gene knockout have been utilised in order to identify sperm cell-specific proteins which may have crucial roles in the plant sexual reproduction process.

8.2 Sperm cell morphological studies: sperm cell measurement and transmission electron microscopy (TEM)

Plant sperm cell morphology has been studied by electron microscopy (EM) in various species i.e. *Zea mays* (McConchie et al., 1987a; Southworth et al., 1988), *Plumbago zeylanica* (Russell, 1984; Russell and Cass, 1981; Southworth et al., 1997), and *Brassica oleracea* (McConchie et al., 1987b). However such a morphological analysis has never been carried out in Arabidopsis. In this study, Arabidopsis sperm cell pairs were analysed utilising transmission electron microscopy (TEM). Analysis of the two cells revealed they were indistinguishable from one another at the cellular level as for *Brassica oleracea* (McConchie et al., 1987b). Both sperm cells contained similar numbers and type of organelles in contrast to well established dimorphic species – *Plumbago zeylanica* where one of the sperm cells (Sua) contains more plastids and the other (Svn) contains more mitochondria. The plasma membrane of the two sperm cells appeared similar in the TEM micrograph. However, freeze fracture analysis of the membrane is a more suitable technique in order to identify the molecules on the plasma membrane of the cells as demonstrated in various plant sperm cell morphological studies (Southworth et al., 1988; Southworth et al., 1994; Southworth et al., 1997; van Aelst et al., 1990). In this study, volume measurements of Arabidopsis sperm using GFP-labelled sperm cells demonstrated no size differences between the two sperm cells. However, since the sperm cells of well known dimorphic sperm species (i.e. *P. zeylanica* and *Zea mays*) demonstrated preferential fusion (Russell, 1985; Shi et al., 1996), it is still possible that the sperm cells of Arabidopsis is dimorphic in other criteria other than size and could well preferentially fuse with their partners, the egg and central cell. The basis of sperm cell dimorphism has not yet been confirmed in plants - whether it is derived from an unequal division of the generative cell or later during sperm cell maturation. Evidence supporting both theories was demonstrated in *Plumbago zeylanica* where an unequal division of the generative cell was identified along with increasing numbers of mitochondria in the Svn sperm during maturation (Russell and Strout, 2005; Russell

et al., 1996). Nonetheless, this could occur differently in other plant species. As there are two different fusion partners for each of the sperm cells, the possibility that sperm dimorphism is universal is not unrealistic. Interestingly, sperm cell dimorphism has been found in almost all plant species that have been intensively studied so far (Chen et al., 2006; McConchie et al., 1987b; Mogensen et al., 1990; Qiu et al., 2004; Russell, 1984; Yu and Russell, 1994b). More importantly, the two sperm fusion partners, the egg and central cell, are clearly morphologically different and have different functions in the reproduction process. Therefore, the recognition molecules on the egg and central cell surfaces are likely to be different. Moreover, both egg and central cells may require different mRNAs or proteins that could be carried in each sperm cell for immediate use after fusion. However, until more experiments are carried out, it remains to be seen whether the two sperm cells of Arabidopsis are different in aspects and whether Arabidopsis sperm cells have specified fusion partners.

8.3 Experiments performed to identify Arabidopsis sperm cell-expressed proteins

8.3.1 Pollen disruption and sperm cell isolation method development for *B. oleracea* and *A. thaliana*

In order to obtain sperm cells for molecular studies, sperm cell isolation is the first crucial step to be developed. As noted above and elsewhere, plant sperm cell isolation has proven to be problematic due to the fact that sperm cells reside within the vegetative cell of the pollen (Russell, 1991). Sperm cells can readily be disrupted by slight environmental variations i.e. osmotic and pH changes (Russell, 1991). Therefore, the most crucial factor for sperm isolation is to formulate an isolation buffer that is capable of bursting the pollen grain but can preserve the released sperm cells. Other methods such as pH shock, physical disruption and enzymatic treatment have been used alongside osmotic shock to facilitate sperm isolation (Chen et al., 2005; Matthys-Rochon et al., 1988; Mo and Yang, 1992; Roeckel et al., 1988; Russell, 1986; Shivanna et al., 1988; Vergne and Dumas, 1988; Wilms et al., 1986). In species that produce relatively large pollen grains i.e. maize, osmotic shock alone is enough to bring about pollen disruption (Roeckel et al., 1988; Yang and Zhou, 1989). However, in the model plant Arabidopsis where the pollen grain is small by comparison with a relatively strong pollen wall, other disruption techniques need to be used to facilitate sperm release. In this study, physical disruption by the Parr pressure bomb was used together with osmotic shock to burst the pollen grain of *B. oleracea*. However, in *A. thaliana* these two methods alone were not efficient enough to disrupt the majority of the grains and

therefore enzymatic treatment was introduced to facilitate sperm release. Following the pollen disruption step these sperm cells required purification in order to be used in further studies (Russell, 1991). One of the most widely used purification technique in cell biology is the Percoll gradient (Russell, 1991). This technique can be adapted to separate a range of cell types by their size and density (Lessley and Garner, 1983). However, in this study filtration and centrifugation was found to be more efficient to enrich and preserve the isolated sperm cells. One of the main aims of this study was to identify proteins that are expressed in this cell type using proteomic techniques. Thus the sperm cell population utilised in this study required absolute purity which was impossible to obtain by crude purification methods such as filtration or Percoll gradient centrifugation. Thus, florescence activated cell sorting (FACS) was applied as a further purification step. The sample used for FACS required careful preparation with factors such as the concentration of cells,, the buffer for the cell sample, cell labelling, the criteria for sorting and the precise calibration of the machine being crucial for successful sorting. In this study GFP-labelled sperm cells, generously provided by Dr. Scott D. Russell (University of Oklahoma), were proven to be exceptionally useful for sperm cell viability examination for the entire processes of sperm cell isolation, semi-purification, and FACS-purification. These GFP tagged cells were not only demonstrated to be viable throughout the whole process but also a few hours after the FACS step. The concentration of sorted sperm cells after FACS was found to be an issue as the cells were sorted into a large volume of FACS sheath buffer and sperm isolation buffer. Therefore for future studies sorted cells will be collected directly into a buffer that is appropriate for subsequent experiments to avoid losing cells via centrifugation. A purity assessment was performed for the FACS-purified sperm cell population by RT-PCR and proven to be free from pollen cytoplasmic contamination. These highly purified Arabidopsis sperm cells can be used in whole range of sperm molecular studies such as 2DE, mass spectrometry, WSMS, cDNA library construction and microarray studies, as they are free from pollen cytoplasmic contamination.

8.3.2 Proteomic analysis of *B. oleracea* and *A. thaliana* sperm cells

A semi-purified sperm cell population, a sperm cell-excluded sample and a sample of disrupted pollen grains were separated by two dimension protein electrophoresis (2DE). The protein spots in each gel were compared and selected for mass spectrometry (MS) protein identification. A relatively large amount of total protein was required for efficient protein spot identification. Unfortunately a limited number of FACS-sorted sperm cells from both Brassica and Arabidopsis were available in this preliminary study

and were insufficient to provide a whole protein profile of the sperm cells. The insufficient amount of protein obtained from these sorted cells could be a result of the protein precipitation protocol was ineffective for Arabidopsis as the cells were diluted in too large a volume of buffer. Accordingly, other cell disruption and protein precipitation procedures which result in complete disruption of the cells and precipitation of minute amounts of protein may be required for future studies. Alternatively, sperm cells could be sorted directly into cell lysis buffer to avoid protein loss due to handling processes such as centrifugation and pipetting. Nonetheless, this study was tremendously useful for technique development for further sperm cells proteomics studies in the future. This study is the first to prove that FACS-sorted sperm cells could be successfully obtained from Arabidopsis. Moreover, the sorted sperm cells were able to be used for 2DE and MS analyses as was demonstrated for *B. oleracea* sperm. In this study, the *B. oleracea* sorted sperm sample demonstrated more diversity in the protein profile than the semi-purified sperm sample which was dominated by mitochondrial proteins (40%). This indicated that i) the protein profile obtained from FACS-sorted cells almost certainly represented proteins from sperm cells and ii) the semi-purified sperm sample was highly contaminated with mitochondria from the pollen, thus confirming that FACS is crucial for successful plant sperm cell proteomic studies.

8.3.3 A bioinformatics approach for the identification of Arabidopsis sperm cell protein candidates: comparative cross-species study of sperm cell ESTs

In parallel with the proteomic approach a bioinformatics approach was also adopted for sperm cell protein identification. An EST library derived from FACS-purified maize sperm cells was successfully obtained in 2003 (Engel et al., 2003). Therefore, this valuable dataset along with 1,522 *P. zeylanica*, 61 *O. sativa*, and 3 *N. tabaccum* sperm ESTs were utilised in a comparative cross-species study alongside the proteomics study to identify Arabidopsis specific proteins. Approximately 50% (2,598 sequences) of maize sperm ESTs matched to Arabidopsis proteins as expected for a cross-species comparison (Brendel et al., 2004). The microarray expression profile reported on Genevestigator was found to be tremendously useful for sperm cell-expressed candidate selection as it presents information for most of the genes expressed in Arabidopsis (22k) and provides expression data for almost every tissue type. Moreover, the data is very accurate as demonstrated in this study where the RT-PCR results of each sperm cell-expressed candidate gene in different tissue types was compared to the database. However, if the Gene Atlas tool allowed batch analysis it would be much more

powerful as the process of candidate selection had to be carried out manually. Because of the cumbersome nature of this process the expression pattern of only 250 genes was analysed using this tool. In the future the number of genes could be expanded in order identify more candidates for sperm-specific genes as in this study only two genes were found to have very similar patterns to known sperm-specific genes *GCS1/HAP2* and the sperm cell-expressed gene *GEX2*. Importantly these two candidates were proven to be sperm cell-specific genes by GFP fusion analysis.

8.3.4 Characterisation of putative Arabidopsis sperm cell transcripts selected from bioinformatics approach by RT-PCR and GFP fusion

High priority sperm cell-expressed gene candidates derived from the cross-species bioinformatics study were selected and primarily investigated by RT-PCR. Approximately 35% (7 genes) of the high priority sperm cell-expressed gene candidates investigated (20 genes) were preferentially expressed in the sperm cells which suggested that the process of candidate selection was effective. However, the GFP reporter was detected in sperm cells of only two constructs (10%) At1g10090::GFP and At5g39650::GFP. The promGEX2::GFP positive control could not be detected in this GFP experiment and suggested that these two genes were likely to be expressed at a relatively high level in comparison to the sperm cell-expressed gene *GEX2*. These results corresponded well with the information obtained from Gene Atlas and also provided a gauge to the sensitivity of our GFP gene fusion system which could be useful in future studies. Interestingly, both genes, including *GEX2*, were identified as 'unknown proteins' suggesting that many plant sperm cell-specific proteins possibly have unique functions that are different to other known proteins. The other five GFP constructs where the GFP signal could not be detected in the sperm cells might have failed because the signal in the sperm cells was very low and was masked by the GFP signal from the surrounding pollen grain. These constructs highlighted problems in using a green fluorescent marker in plant tissues and particularly in pollen as it autofluoresces green under UV light. This drawback can be easily overcome in the future by using other fluorescent proteins such as RFP and YFP. Moreover, as demonstrated in literature and in this study, sperm cell specific proteins are likely to be expressed at very low levels compared to other pollen cytoplasmic proteins and thus the signal emitted from the fusion protein could be very low (At5g39650::GFP) or unable to be detected (promGEX2::GFP) as demonstrated in this study. Therefore other methods for the study of gene expression might be more useful where the detection level can be control i.e. GUS. Nonetheless, two novel sperm cell-expressed genes were

successfully identified in this study which raises the likelihood that there are many more genes specifically expressed in sperm cells that are yet to be found. In addition, as the GFP signal was detected only in mature sperm cells this confirmed that the sperm are truly active cells and are not just passive carriers of preformed generative cell mRNA cargos destined for use in the zygote and endosperm. However the function of these sperm cell-expressed genes still needs to be elucidated.

8.3.5 Identification and assessment of novel Arabidopsis genes that matched maize sperm ESTs

The other half of the maize ESTs (2,467 sequences) that did not match known Arabidopsis proteins could still potentially have Arabidopsis sperm cell-expressed protein homologues. This is possible not only because the initial blast search was a cross-species analysis but also because sperm cell-specific proteins of Arabidopsis might also have been missed from the annotation of the genome and that many sperm-specific transcripts may not have been identified in cDNA libraries due to their scarcity. Two sequences (Atnov1 and Atnov2) were found to match to regions that might potentially be novel Arabidopsis genes. Only one sequence though, Atnov2, was predicted to be a possible gene utilising the gene prediction tools, Softberry and Genescan. RT-PCR confirmed the prediction as the Atnov2 transcript was detected. Later in 2007 this gene was annotated and deposited on the GenBank database as At4g20325, thus confirming that the techniques used to identify sperm cell-expressed genes in this study were indeed effective.

8.3.6 Gene functional analysis utilising Arabidopsis T-DNA insertion mutagenesis

8.3.6.1 High priority sperm cell expressed gene candidates

Plant sperm cell-expressed gene identification and functional analysis can be performed utilising available Arabidopsis T-DNA mutagenesis plant lines. Therefore, mutant phenotypes were screened for in plant lines carring T-DNA insertions in our genes of interest. Interestingly, for approximately half of the seven high priority sperm cell-expressed gene candidates, including the sperm cell-specific gene At5g39650, T-DNA insertion lines were unavailable (in exon or intron regions). This could potentially be because these genes are crucial for the reproductive process on both the male and female sides and thus mutant alleles would not be transmitted to the next generation.

However, T-DNA insertion lines in the promoter or untranslated regions of these genes were analysed and no phenotype was detected suggesting that these available T-DNA insertions did not disrupt the genes. Additional studies are required to elucidate the function of these genes in sperm, particularly for the sperm cell-specific gene At5g39650 identified in this study. RNAi could be adopted to knock down the gene. Overexpression of the gene could also be utilised to aid in identifying the gene function.

8.3.6.2 Screening of candidates derived from the bioinformatics cross-species study

Sperm-expressed gene candidates obtained from the bioinformatics approach described in chapter 5 were analysed utilising available Arabidopsis T-DNA insertion plant lines. These Arabidopsis sperm cell-expressed gene candidates were derived from a cross-species comparative study of maize, *P. zeylanica*, rice, and tobacco sperm ESTs with the Arabidopsis genome. Candidates were then 'filtered' utilising microarray expression pattern data available online (Gene Atlas) for the genes that expressed preferentially in the sperm cell carrier – the pollen grains.

8.3.6.3 Identification of potential sperm cell and/or gamete expressed genes by semi-infertility phenotype screening

An initial screen for potential sperm cell-expressed genes was performed by semi-infertility (reduced-seed set) phenotypic analysis. This screen revealed 3 classes of semi-infertility phenotypes, i) 'unfertilised ovule' (21 lines), ii) 'aborted seed' (7 lines), and iii) 'minute silique' (6 lines). However, when genotypic screening was performed, more than 50% of all the plant lines contained more than one copy of T-DNA. Moreover, the genotypic analysis in some of the lines could not be completed as the T-DNA insertion could not be located utilising PCR. These two problems with the T-DNA insertion plant collection could probably be overcome but would require time and thus the experiments will be performed in the future. Importantly, not all classes of male gametophytic mutations would be detected utilising this semi-infertility phenotype screen due to the abundance of wt pollen in heterozygotes that would mask the presence of mutant pollen. Indeed in this study, none of the male gametophytic mutants were detected utilising the semi-infertility phenotype screening alone. This highlights the interesting point that the resources invested for the production of vast numbers of pollen grains is not only used to guarantee successful reproduction but also to assure the fitness of the progeny through pollen competition effects. This was demonstrated in the

restricted pollination experiments where limiting quantities of pollen applied to stigmas can affect the number of normal seeds produced in the siliques. Thus the detection of male gametophytic mutations can be more difficult than those affecting the female which are normally easily detected by screening for reduced seed set. Therefore segregation distortion assays was adopted to screen for gametophytic mutations, particularly those 'hidden' ones affecting the male gametophyte.

8.3.6.4 Identification of potential sperm cell and/or gametophyte-expressed genes by segregation distortion assay

A segregation distortion assay was utilised for the detection of gametophytic gene mutations. As the tested genes were selected using criteria to identify those expressed in the male gametophyte a relatively high percentage (65%) of the single copy T-DNA plant lines screened (19 out of 29 lines) demonstrated segregation distortion. This suggested that the criteria used for candidate selection were effective. Of these 19 lines 9 lines were probable male gametophytic mutations as no semi-infertility phenotype was detected. Interestingly, though the candidate selection was directed at identifying male gametophytic gene mutations, approximately half of the mutant lines affected both male and female gametophytes and many genes were expressed in both sporophytic and gametophytic tissues. These results suggested that producing and maintaining haploid cells probably uses many of the same genes and these genes could possibly be expressed in both gametophytes. This could be because the sporophytic and gametophytic tissues or the two gametes regularly work together in order to ensure successful reproduction e.g. the timing of the maturation of the two gametes has to be synchronised (Friedman, 1999; Tian et al., 2005).

This T-DNA mutagenesis screening together with transcriptional analysis and GFP fusion experiments in this study demonstrated that mutations affecting the gamete fusion event are rare as only two potential candidates were detected by GFP without the support of T-DNA knockout lines. A number of theories can be raised to explain these observations, i) only very few genes control the actual gamete fusion event, ii) as the fusion event is crucial for the success of a species, multiple genes are involved to maintain normal gamete function (redundancy) and thus multiple gene knockouts are required for a phenotype to be detected, iii) there are mechanisms to prevent defective male gametes reaching the female to promote successful reproduction as demonstrated in *hap2* mutant pollen where tube growth is also affected by the mutation (von Besser et al., 2006), iv) mutant alleles of genes that are involved in gamete fusion cannot be recovered (thus T-DNA lines are unavailable) - this also means that these mutations

affect both gametes. Indeed, as demonstrated in this study, most of the T-DNA lines showed transmission problems via both gametes and the At2g22740::GFP construct also revealed expression in both male and female suggesting that the expression of gametophytic genes in both gametophytes is common.

8.4 Future work

The main aim of this project was to identify Arabidopsis sperm cell proteins utilising a wide range of molecular and bioinformatics techniques including proteomics, transcriptional studies, and T-DNA insertion mutagenesis. This project has successfully provided techniques for Arabidopsis sperm isolation and FACS purification. The isolated sperm cells were proven to be pure and can be used in future transcriptomic studies. Unfortunately the amount of protein extracted from the purified Arabidopsis sperm cells was inadequate for proteomic analysis. In addition the 2DE and MS experiments for *Brassica oleracea* using FACS-sorted cells, although successful to a degree, demonstrated a similar problem with protein being limiting. With sufficient time and resources proteomics should be successful for both Arabidopsis and Brassica sperm. However, as demonstrated by RT-PCR of the FACS-sorted sperm cells their RNA could be extracted and used for transcriptional studies. Therefore microarray experiments to reveal the full set of Arabidopsis sperm cell transcripts form FACS-sorted cells could also be performed with a need for much less material.

The list of genes obtained from the bioinformatics approach was also very valuable as it was confirmed to contain sperm cell-expressed genes - although only one gene candidate from this list demonstrated sperm cell specificity. Therefore it is possible that the processes used for selecting the correct candidate genes might be insufficiently directed. However, as a sperm cell-specific gene has been identified in this project and a few others are known the level and pattern of their gene expression (as viewed in Gene Atlas) could be accurately applied in order to precisely identify other candidates for further analysis. The list of T-DNA insertion lines that still require careful investigation will be continued. The problem in identifying T-DNA insertions in these genes of interest by PCR will be tackled in order to confirm the position of the knockout gene, particularly those that demonstrated gametophytic gene mutations. The T-DNA lines where multiple copies of T-DNA were detected, again particularly those lines a reduced seed set phenotype was detected, will be back crossed to isolate single copy T-DNA lines. This valuable information on technique development and novel sperm-specific genes and candidates generated from this project will surely be useful for future plant sperm cells studies. Therefore this study will have positive effects for studies on plant sexual reproduction in the future.

References

Abbott, J. C., Aanensen, D. M., Rutherford, K., Butcher, S. and Spratt, B. G. (2005). WebACT-an online companion for the Artemis Comparison Tool, vol. 21 (ed., pp. 3665-3666: Oxford Univ Press.

Ainsworth, C. (2005). Cell biology: the secret life of sperm. *Nature* **436**, 770-771.

Alandete-Saez, M., Ron, M. and McCormick, S. (2008). GEX3, Expressed in the Male Gametophyte and in the Egg Cell of Arabidopsis thaliana, Is Essential for Micropylar Pollen Tube Guidance and Plays a Role during Early Embryogenesis. *Molecular Plant*.

Alzwiy, I. A. and Morris, P. C. (2007). A mutation in the Arabidopsis MAP kinase kinase 9 gene results in enhanced seedling stress tolerance. *Plant Science* **173**, 302-308.

An, Y. Q., Huang, S., McDowell, J. M., McKinney, E. C. and Meagher, R. B. (1996). Conserved expression of the Arabidopsis ACT1 and ACT 3 actin subclass in organ primordia and mature pollen. *The Plant Cell* **8**, 15.

Arenas-Huertero, F., Arroyo, A., Zhou, L., Sheen, J. and Leσn, P. (2000). Analysis of Arabidopsis glucose insensitive mutants, gin5 and gin6, reveals a central role of the plant hormone ABA in the regulation of plant vegetative development by sugar. *Genes & Development* **14**, 2085-2096.

Aslam, I., Robins, A., Dowell, K. and Fishel, S. (1998). Isolation, purification and assessment of viability of spermatogenic cells from testicular biopsies of azoospermic men. *Human Reproduction* **13**, 639-645.

Ayele, M., Haas, B. J., Kumar, N., Wu, H., Xiao, Y., Van Aken, S., Utterback, T. R., Wortman, J. R., White, O. R. and Town, C. D. (2005). Whole genome shotgun sequencing of Brassica oleracea and its application to gene discovery and annotation in Arabidopsis, vol. 15 (ed., pp. 487-495: Cold Spring Harbor Lab.

Azpiroz-Leehan, R. and Feldmann, K. A. (1997). T-DNA insertion mutagenesis in Arabidopsis: going back and forth. *Trends Genet* **13**, 152-6.

Babula, D., Kaczmarek, M., Barakat, A., Delseny, M., Quiros, C. F. and Sadowski, J. (2003). Chromosomal mapping of Brassica oleracea based on ESTs from Arabidopsis thaliana: complexity of the comparative map. *Molecular Genetics and Genomics* **268**, 656-665.

Bai, X., Peirson, B. N., Dong, F., Xue, C. and Makaroff, C. A. (1999). Isolation and Characterization of SYN1, a RAD21-like Gene Essential for Meiosis in Arabidopsis. *The Plant Cell Online* **11**, 417-430.

Balhorn, R. (1982). A model for the structure of chromatin in mammalian sperm. *The Journal of Cell Biology* **93**, 298-305.

Ban, N., Sasaki, M., Sakai, H., Ueda, K. and Inagaki, N. (2005). Cloning of ABCA17, a novel rodent sperm-specific ABC (ATP-binding cassette) transporter that regulates intracellular lipid metabolism. *Biochemical Journal* **389**, 577-585.

Banks, J. A. (1999). GAMETOPHYTE DEVELOPMENT IN FERNS. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**, 163-186.

Barrell, P. J. and Grossniklaus, U. (2005). Confocal microscopy of whole ovules for analysis of reproductive development: the elongate1 mutant affects meiosis II. *Plant Journal* **43**, 309-320.

Baumbusch, L. O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., Aalen, R. B. and Journals, O. (2001). The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Research* **29**, 4319-4333.

Bechtold, N., Ellis, J. and Pelletier, G. (1993). In planta Agrobacterium mediated gene transfer by infiltration of adult Arabidopsis thaliana plants. *Comptes rendus de l'Acad̃mie des sciences. S̃rie 3, Sciences de la vie* **316**, 1194-1199.

Becker, J. D., Boavida, L. C., Carneiro, J., Haury, M. and Feijo, J. A. (2003). Transcriptional Profiling of Arabidopsis Tissues Reveals the Unique Characteristics of the Pollen Transcriptome. *Plant Physiology*, 103028241.

Bellis, M. A., Baker, R. R. and Gage, M. J. G. (1990). Variation in Rat Ejaculates Consistent with the Kamikaze-Sperm Hypothesis. *Journal of Mammalogy* **71**, 479-480.

Bent, A. F. (2000). Arabidopsis in planta transformation. Uses, mechanisms, and prospects for transformation of other species. *Plant Physiol* **124**, 1.

Berger, F. (2008). Double-fertilization, from myths to reality. *Sexual Plant Reproduction* **21**, 3-5.

Boavida, L. C. and McCormick, S. (2007). Temperature as a determinant factor for increased and reproducible in vitro pollen germination in Arabidopsis thaliana. *The Plant Journal* **52**, 570-582.

Boavida, L. C., Vieira, A. M., Becker, J. D. and Feijo, J. A. (2005). Gametophyte interaction and sexual reproduction: how plants make a zygote. *Int J Dev Biol* **49**, 615-32.

Boguski, M. S. (1995). The turning point in genome research. *Trends in Biochemical Sciences* **20**, 295-296.

Boguski, M. S., Lowe, T. M. J. and Tolstoshev, C. M. (1993). dbEST— database for"expressed sequence tags". *Nature Genetics* **4**, 332-333.

Boisson, M., Gomord, V., Audran, C., Berger, N., Dubreucq, B., Granier, F., Lerouge, P., Faye, L., Caboche, M. and Lepiniec, L. (2001). Arabidopsis glucosidase I mutants reveal a critical role of N-glycan trimming in seed development. *The EMBO Journal* **20**, 1010-1019.

Bottino, D., Mogilner, A., Roberts, T., Stewart, M. and Oster, G. (2002). How nematode sperm crawl. *Journal of Cell Science* **115**, 367-384.

Bower, M. S., Matias, D. D., Fernandes-Carvalho, E., Mazzurco, M., Gu, T., Rothstein, S. J. and Goring, D. R. (1996). Two members of the thioredoxin-h family interact with the kinase domain of a Brassica S locus receptor kinase. *The Plant Cell* **8**, 1641.

Brawley, S. H. (1991). The fast block against polyspermy in fucoid algae is an electrical block. *Dev Biol* **144**, 94-106.

Brendel, V., Kurtz, S. and Walbot, V. (2004). Comparative genomics of Arabidopsis and maize: prospects and limitations. *feedback*.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M. and Ewan, M. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* **18**, 630-634.

Brown, R. C. and Lemmon, B. E. (2007). The Pleiomorphic Plant MTOC: An Evolutionary Perspective. *Journal of Integrative Plant Biology* **49**, 1142-1153.

Brugiere, S., Kowalski, S., Ferro, M., Seigneurin-Berny, D., Miras, S., Salvi, D., Ravanel, S., d'Herin, P., Garin, J., Bourguignon, J. et al. (2004). The hydrophobic proteome of mitochondrial membranes from Arabidopsis cell suspensions. *Phytochemistry* **65**, 1693-1707.

Cabrillac, D., Cock, J. M., Dumas, C. and Gaude, T. (2001). The S-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins. *Nature* **410**, 220-3.

Callow, J. A. (1985). Sexual recognition and fertilization in brown algae. *J Cell Sci Suppl* **2**, 219-32.

Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G. M., Carnemolla, B., Orecchia, P., Zardi, L. and Righetti, P. G. (2004). Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**, 1327-33.

Cao, Y., Reece, A. and Russell, S. D. (1996). Isolation of viable sperm cells from tobacco (Nicotiana tabacum). *Zygote* **4**, 81-4.

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

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. and Parkhill, J. (2005). ACT: the Artemis comparison tool, vol. 21 (ed., pp. 3422-3423: Oxford Univ Press.

Cass, D. D. (1973). An ultrastructural and Nomarski-interference study of the sperms of barley. *Revue canadienne de botanique* **51**, 601-605.

Catt, J. W., Vithanage, H. I., Callow, J. A., Callow, M. E. and Evans, L. V. (1983). Fertilization in brown algae. V. Further investigations of lectins as surface probes. *Exp Cell Res* **147**, 127-33.

Che, P., Lall, S., Nettleton, D. and Howell, S. H. (2006). Gene Expression Programs during Shoot, Root, and Callus Development in Arabidopsis Tissue Culture. *Plant Physiology* **141**, 620.

Chen, S.-H., Yang, Y.-H., Liao, J.-P. and Tian, H.-Q. (2005). Isolation of egg cells and zygotes from Torenia fournieri. *Zhiwu Shengli yu Fenzi Shengwuxue Xuebao* **31**, 383-388.

Chen, S. H., Liao, J. P., Kuang, A. X. and Tian, H. Q. (2006). Isolation of two populations of sperm cells from the pollen tube of Torenia fournieri. *Plant Cell Reports* **25**, 1138-1142.

Chen, Y. C. S. and McCormick, S. (1996). sidecar pollen, an Arabidopsis thaliana male gametophytic mutant with aberrant cell divisions during pollen development. *Development* **122**, 3243-3253.

Chen, Z., Tan, J. L. H., Ingouff, M., Sundaresan, V. and Berger, F. (2008). Chromatin assembly factor 1 regulates the cell cycle but not cell fate during male gametogenesis in Arabidopsis thaliana. *Development* **135**, 65.

Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. and Sheen, J. (1996). Engineered GFP as a vital reporter in plants. *Curr. Biol* **6**, 325-330.

Christensen, C. A., Subramanian, S. and Drews, G. N. (1998). Identification of Gametophytic Mutations Affecting Female Gametophyte Development inArabidopsis. *Developmental Biology* **202**, 136-151.

Clarke, P. A., te Poele, R., Wooster, R. and Workman, P. (2001). Gene expression microarray analysis in cancer biology, pharmacology, and drug development: progress and potential. *Biochemical Pharmacology* **62**, 1311-1336.

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

Cole, R. A., Synek, L., Zarsky, V. and Fowler, J. E. (2005). SEC8, a Subunit of the Putative Arabidopsis Exocyst Complex, Facilitates Pollen Germination and Competitive Pollen Tube Growth. *Plant Physiology*.

Craigon, D. J., James, N., Okyere, J., Higgins, J., Jotham, J. and May, S. (2004). NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Research* **32**, D575.

Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A. and Tsien, R. Y. (1995). Understanding, improving and using green fluorescent proteins. *Trends in Biochemical Sciences* **20**, 448-455.

Cutler, S. R., Ehrhardt, D. W., Griffitts, J. S. and Somerville, C. R. (2000). Random GFP:: cDNA fusions enable visualization of subcellular structures in cells of Arabidopsis at a high frequency. *Proceedings of the National Academy of Sciences* **97**, 3718.

da Costa-Nunes, J. A. and Grossniklaus, U. (2003). Unveiling the gene-expression profile of pollen. *Genome Biol* **5**, 205.

Demidov, D., Van Damme, D., Geelen, D., Blattner, F. R. and Houben, A. (2005). Identification and Dynamics of Two Classes of Aurora-Like Kinases in Arabidopsis and Other Plants W in Box. *Plant Cell* **17**, 836-848.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A. and Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* **14**, 457-60.

Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N. and Sverdlov, E. D. (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci US A* **93**, 6025-6030.

Digonnet, C. (1997). First evidence of a calcium transient in flowering plants at fertilization, vol. 124 (ed., pp. 2867-2874.
Dorus, S., Busby, S. A., Gerike, U., Shabanowitz, J., Hunt, D. F. and Karr, T. L. (2006). Genomic and functional evolution of the Drosophila melanogaster sperm proteome. *Nature Genetics* **38**, 1440-1445.
Drews, G. N., Lee, D. and Christensen, C. A. (1998). Genetic analysis of female gametophyte development and function. *The Plant Cell* **10**, 5.
Dumas, C., Knox, R. B., McConchie, C. A. and Russell, S. D. (1984). Emerging physiological concepts in fertilization. *What's New Plant Physiol* **15**, 17-20.
Dunphy, G. B. and Nolan, R. A. (1979). Effects of Physical Factors on Protoplasts of Entomophthora egressa. *Mycologia* **71**, 589-602.
Dupuis, I., Roeckel, P., Matthys-Rochon, E. and Dumas, C. (1987). Procedure to Isolate Viable Sperm Cells from Corn (Zea mays L.) Pollen Grains 1. *Plant Physiology* **85**, 876-878.
Durbarry, A., Vizir, I. and Twell, D. (2005). Male Germ Line Development in Arabidopsis. duo pollen Mutants Reveal Gametophytic Regulators of Generative Cell Cycle Progression 1 [w]. *Plant Physiology* **137**, 297-307.
Ebbs, M. L., Bartee, L. and Bender, J. (2005). H3 Lysine 9 Methylation Is Maintained on a Transcribed Inverted Repeat by Combined Action of SUVH6 and SUVH4 Methyltransferases. *Molecular and Cellular Biology* **25**, 10507-10515.
Ebbs, M. L. and Bender, J. (2006). Locus-Specific Control of DNA Methylation by the Arabidopsis SUVH5 Histone Methyltransferase. *The Plant Cell Online* **18**, 1166.
Ehrhardt, D. (2003). GFP technology for live cell imaging. *Current Opinion in Plant Biology* **6**, 622-628.
Engel, M. L., Chaboud, A., Dumas, C. and McCormick, S. (2003). Sperm cells of Zea mays have a complex complement of mRNAs. *Plant Journal* **34**, 697-707.
Engel, M. L., Holmes-Davis, R. and McCormick, S. (2005). Green Sperm. Identification of Male Gamete Promoters in Arabidopsis 1 [w]. *Plant Physiology* **138**, 2124-2133.
Escobar-Restrepo, J. M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, J., Yang, W. C. and Grossniklaus, U. (2007). The FERONIA Receptor-like Kinase Mediates Male-Female Interactions During Pollen Tube Reception. *Science* **317**, 656.
Evans, J. P. and Kopf, G. S. (1998). Molecular mechanisms of sperm-egg interactions and egg activation. *Andrologia* **30**, 297-307.
Faure, J. E., Digonnet, C. and Dumas, C. (1994). An in Vitro System for Adhesion and Fusion of Maize Gametes. *Science* **263**, 1598.
Faure, J. E., Rusche, M. L., Thomas, A., Keim, P., Dumas, C., Mogensen, H. L., Rougier, M. and Chaboud, A. (2003). Double fertilization in maize: the two male gametes from a pollen grain have the ability to fuse with egg cells. *The Plant Journal* **33**, 1051-1062.
Feldmann, K. A., Coury, D. A. and Christianson, M. L. (1997). Exceptional Segregation of a Selectable Marker (Kan (R)) in Arabidopsis Identifies Genes Important for Gametophytic Growth and Development. *Genetics* **147**, 1411-1422.
Fernandez, P., Paniego, N., Lew, S., Hopp, H. E. and Heinz, R. A. (2004). Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project. *feedback*.
Finkelstein, R. R. (1994). Mutations at two new Arabidopsis ABA response loci are similar to the abi 3 mutations. *The Plant Journal* **5**, 765-771.
Franklin-Tong, V. E. (2002). The difficult question of sex: the mating game. *Current Opinion in Plant Biology* **5**, 14-18.
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.
Gaff, D. F. (1971). The Use of Non-permeating Pigments for Testing the Survival of Cells. *Journal of Experimental Botany* **22**, 756.
Geltz, N. R. and Russell, S. D. (1988). Two-Dimensional Electrophoretic Studies of the Proteins and Polypeptides in Mature Pollen Grains and the Male Germ Unit of Plumbago-Zeylanica. *Plant Physiology* **88**, 764-769.
Gilbert, S. F. (2000). Developmental biology: Sinauer Associates Sunderland, Mass.

Golovkin, M. and Reddy, A. S. N. (2003). A calmodulin-binding protein from Arabidopsis has an essential role in pollen germination. *Proceedings of the National Academy of Sciences* **100**, 10558.
Gou, X., Yuan, T., Wei, X. and Russell, S. D. (2005). Expressed gene products of the dimorphic sperm cells of Plumbago zeylanica. *Acta Biologica Cracoviensia* **47 suppl. 1**, 18.
Gou, X. P., Xu, Y., Tang, L., Yan, F., Chen, F. (2000). Expressed Sequence Tags from subtracted rice sperm cells cDNA library, (ed.: Unplublished.
Gray, W. M. (2004). Hormonal regulation of plant growth and development. *PLoS Biol* **2**, e311.
Grini, P. E., Schnittger, A., Schwarz, H., Zimmermann, I., Schwab, B., Jurgens, G. and Hulskamp, M. (1999). Isolation of Ethyl Methanesulfonate-Induced Gametophytic Mutants in Arabidopsis thaliana by a Segregation Distortion Assay Using the Multimarker Chromosome 1. *Genetics* **151**, 849-863.
Gross-Hardt, R., Kagi, C., Baumann, N., Moore, J. M., Baskar, R., Gagliano, W. B., Jurgens, G. and Grossniklaus, U. (2007). LACHESIS Restricts Gametic Cell Fate in the Female Gametophyte of Arabidopsis. *PLoS Biol* **5**, e47.
Gupta, R., Ting, J. T. L., Sokolov, L. N., Johnson, S. A. and Luan, S. (2002). A Tumor Suppressor Homolog, AtPTEN1, Is Essential for Pollen Development in Arabidopsis. *The Plant Cell Online* **14**, 2495.
Hansen, G. and Wright, M. S. (1999). Recent advances in the transformation of plants. *Trends in Plant Science* **4**, 226-231.
Harcourt, A. H. (1991). Sperm Competition and the Evolution of Nonfertilizing Sperm in Mammals. *Evolution* **45**, 314-328.
Harris, S. (2004). Arabidopsis—map makers of the plant kingdom, (ed. http://www.nsf.gov/about/history/nsf0050/pdf/arabidopsis.pdf.
Heberle-Bors, E., Voronin, V., Touraev, A., Sanchez Testillano, P., Risue◦o, M. C. and Wilson, C. (2001). MAP kinase signaling during pollen development. *Sexual Plant Reproduction* **14**, 15-19.
Hennig, L., Menges, M., Murray, J. A. H. and Gruissem, W. (2003). Arabidopsis transcript profiling on Affymetrix GeneChip arrays. *Plant Molecular Biology* **53**, 457-465.
Herzel, H., Beule, D., Kielbasa, S., Korbel, J., Sers, C., Malik, A., Eickhoff, H., Lehrach, H. and Schuchhardt, J. (2001). Extracting information from cDNA arrays. *Chaos* **11**, 98–107.
Heslop-Harrison, J., Heslop-Harrison, Y. and Shivanna, K. R. (1984). The evaluation of pollen quality, and a further appraisal of the fluorochromatic (FCR) test procedure. *TAG Theoretical and Applied Genetics* **67**, 367-375.
Higashiyama, T., Kuroiwa, H., Kawano, S. and Kuroiwa, T. (2000). Explosive Discharge of Pollen Tube Contents in Torenia fournieri, vol. 122 (ed., pp. 11-14: Am Soc Plant Biol.
Higashiyama, T., Kuroiwa, H. and Kuroiwa, T. (2003). Pollen-tube guidance: beacons from the female gametophyte. *Curr Opin Plant Biol* **6**, 36-41.
Higashiyama, T., Yabe, S., Sasaki, N., Nishimura, Y., Miyagishima, S., Kuroiwa, H. and Kuroiwa, T. (2001). Pollen Tube Attraction by the Synergid Cell, vol. 293 (ed., pp. 1480-1483.
Hill, K. A. (1985). Hartsoeker's homonculus: a corrective note. *J Hist Behav Sci* **21**, 178-9.
Holmes-Davis, R., Tanaka, C. K., Vensel, W. H., Hurkman, W. J. and McCormick, S. (2005). Proteome mapping of mature pollen of Arabidopsis thaliana. *Proteomics* **5**, 4864-4884.
Honys, D. and Twell, D. (2003). Comparative analysis of the Arabidopsis pollen transcriptome. *Plant Physiol* **132**, 640-652.
Honys, D. and Twell, D. (2004a). Male gametogenesis. New York: Marcel Dekker Inc.
Honys, D. and Twell, D. (2004b). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. *Genome Biology* **5**, R85.
Howden, R., Park, S. K., Moore, J. M., Orme, J., Grossniklaus, U. and Twell, D. (1998). Selection of T-DNA-Tagged Male and Female Gametophytic Mutants by Segregation Distortion in Arabidopsis. *Genetics* **149**, 621-631.
Huang, B. Q. and Russell, S. D. (1994). Fertilization in Nicotiana tabacum: Cytoskeletal modifications in the embryo sac during synergid degeneration. *Planta* **194**, 200-214.

Huck, N., Moore, J. M., Federer, M. and Grossniklaus, U. (2003). The Arabidopsis mutant feronia disrupts the female gametophytic control of pollen tube reception. *Development* **130**, 2149-2159.
Iwakawa, H., Shinmyo, A. and Sekine, M. (2006). Arabidopsis CDKA; 1, a cdc2 homologue, controls proliferation of generative cells in male gametogenesis. *Plant J* **45**, 819-831.
Jackson, J. P., Johnson, L., Jasencakova, Z., Zhang, X., PerezBurgos, L., Singh, P. B., Cheng, X., Schubert, I., Jenuwein, T. and Jacobsen, S. E. (2004). Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. *Chromosoma* **112**, 308-315.
Jans, D. A. and Hubner, S. (1996). Regulation of protein transport to the nucleus: central role of phosphorylation, vol. 76 (ed., pp. 651-685: Am Physiological Soc.
Jiang, L., Yang, S. L., Xie, L. F., Puah, C. S., Zhang, X. Q., Yang, W. C., Sundaresan, V. and Ye, D. (2005). VANGUARD1 encodes a pectin methylesterase that enhances pollen tube growth in the Arabidopsis style and transmitting tract. *Plant Cell* **17**, 584-596.
Johnson-Brousseau, S. A. and McCormick, S. (2004). A compendium of methods useful for characterizing Arabidopsis pollen mutants and gametophytically-expressed genes. *The Plant Journal* **39**, 761.
Johnson, M. A. and Preuss, D. (2002). Plotting a Course Multiple Signals Guide Pollen Tubes to Their Targets. *Developmental Cell* **2**, 273-281.
Johnston, D. S., Wooters, J. O. E., Kopf, G. S., Qiu, Y. and Roberts, K. P. (2005). Analysis of the Human Sperm Proteome. *Annals of the New York Academy of Sciences* **1061**, 190-202.
Kaji, K. and Kudo, A. (2004). The mechanism of sperm-oocyte fusion in mammals. *Reproduction* **127**, 423-429.
Kaminski, N. (2000). Bioinformatics A User's Perspective, vol. 23 (ed., pp. 705-711: Am Thoracic Soc.
Kandasamy, M. K., McKinney, E. C. and Meagher, R. B. (1999). Technical advance: the late pollen-specific actins in angiosperms. *Plant J* **18**.
Kang, J., Zhang, G., Bonnema, G., Fang, Z. and Wang, X. (2008). Global analysis of gene expression in flower buds of Ms-cd1 Brassica oleracea conferring male sterility by using an Arabidopsis microarray. *Plant Molecular Biology* **66**, 177-192.
Karl, D. M. (1980). Cellular nucleotide measurements and applications in microbial ecology. *Microbiology and Molecular Biology Reviews* **44**, 739-796.
Karr, T. L. (2007). Fruit flies and the sperm proteome. *Human Molecular Genetics* **16**, R124.
Kawabe, A., Matsunaga, S., Nakagawa, K., Kurihara, D., Yoneda, A., Hasezawa, S., Uchiyama, S. and Fukui, K. (2005). Characterization of plant Aurora kinases during mitosis. *Plant Molecular Biology* **58**, 1-13.
Ke-Feng, F., Ting-Ting, Y. and Meng-Xiang, S. (2005a). Mosaicism in the organization of Con A binding sites on the membrane surface of female cells of Nicotiana tabacum. *New Phytologist* **167**, 743-750.
Ke-Feng, F., Ting-Ting, Y. and Meng-Xiang, S. (2005b). Rapid report Mosaicism in the organization of Con A binding sites on the membrane surface of female cells of Nicotiana tabacum. *New Phytologist* **167**, 743.
Khalequzzaman, M. and Haq, N. (2005). Isolation and in vitro fusion of egg and sperm cells in Oryza sativa. *Plant Physiology and Biochemistry* **43**, 69-75.
Kim, S., Dong, J. and Lord, E. M. (2004). Pollen tube guidance: the role of adhesion and chemotropic molecules. *Curr Top Dev Biol* **61**, 61-79.
Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., Fischer, R. L. and Kakutani, T. (2004). One-Way Control of FWA Imprinting in Arabidopsis Endosperm by DNA Methylation, vol. 303 (ed., pp. 521-523: American Association for the Advancement of Science.
Knight, H. and Knight, M. R. (2001). Abiotic stress signalling pathways: specificity and cross-talk. *Trends in Plant Science* **6**, 262-267.
Kranz, E. and Kumlehn, J. (1999). Angiosperm fertilisation, embryo and endosperm development in vitro. *Plant Science* **142**, 183-197.
Kranz, E. and Lorz, H. (1993). In Vitro Fertilization with Isolated, Single Gametes Results in Zygotic Embryogenesis and Fertile Maize Plants. *The Plant Cell Online* **5**, 739-746.

Kranz, E. and Scholten, S. (2008). In vitro fertilization: analysis of early post-fertilization development using cytological and molecular techniques. *Sexual Plant Reproduction* **21**, 67-77.
Kranz, E., Wiegen, P. and Lorz, H. (1995). Early cytological events after induction of cell division in egg cells and zygote development following in vitro fertilization with angiosperm gametes. *The Plant Journal* **8**, 9-23.
Kruft, V., Eubel, H., Jansch, L., Werhahn, W. and Braun, H. P. (2001). Proteomic Approach to Identify Novel Mitochondrial Proteins in Arabidopsis. *Plant Physiology* **127**, 1694-1710.
Kuersten, S. and Goodwin, E. B. (2003). The power of the 3' UTR: translational control and development. *Nature Reviews Genetics* **4**, 626-637.
Kulikauskas, R. and McCormick, S. (1997). Identification of the tobacco and Arabidopsis homologues of the pollen-expressed LAT59 gene of tomato. *Plant Molecular Biology* **34**, 809-814.
Kumar, A. and Bennetzen, J. L. (1999). PLANT RETROTRANSPOSONS. *Annual Review of Genetics* **33**, 479-532.
Kura, T. and Nakashima, Y. (2000). Conditions for the evolution of soldier sperm classes. *Evolution* **54**, 72-80.
Kurihara, D., Matsunaga, S., Kawabe, A., Fujimoto, S., Noda, M., Uchiyama, S. and Fukui, K. (2006). Aurora kinase is required for chromosome segregation in tobacco BY-2 cells. *The Plant Journal* **48**, 572-580.
Kuznetsov, V. A., Knott, G. D. and Bonner, R. F. (2002). General Statistics of Stochastic Process of Gene Expression in Eukaryotic Cells. *Genetics* **161**, 1321-1332.
LaFountain, K. L. and Mascarenhas, J. P. (1972). Isolation of vegetative and generative nuclei from pollen tubes. *Exp Cell Res* **73**, 233-6.
Lalanne, E., Honys, D., Johnson, A., Borner, G. H. H., Lilley, K. S., Dupree, P., Grossniklaus, U. and Twell, D. (2004). SETH1 and SETH2, two components of the glycosylphosphatidylinositol anchor biosynthetic pathway, are required for pollen germination and tube growth in Arabidopsis. *Plant Cell* **16**, 229-240.
Lalanne, E. and Twell, D. (2002). Genetic Control of Male Germ Unit Organization in Arabidopsis. *Plant Physiology* **129**, 865.
Lavitrano, M., Camaioni, A., Fazio, V. M., Dolci, S., Farace, M. G. and Spadafora, C. (1989). Sperm cells as vectors for introducing foreign DNA into eggs: genetic transformation of mice. *Cell* **57**, 717-723.
Le, Q., Gutierrez-Marcos, J. F., Costa, L. M., Meyer, S., Dickinson, H. G., Lorz, H., Kranz, E. and Scholten, S. (2005). Construction and screening of subtracted cDNA libraries from limited populations of plant cells: a comparative analysis of gene expression between maize egg cells and central cells. *Plant Journal* **44**, 167-178.
Lee, J. Y. and Lee, D. H. (2003). Use of serial analysis of gene expression technology to reveal changes in gene expression in Arabidopsis pollen undergoing cold stress. *Plant Physiol* **132**, 517-529.
Lessley, B. A. and Garner, D. L. (1983). Isolation of motile spermatozoa by density gradient centrifugation in Percoll®. *Gamete Research* **7**, 49-61.
Liang, P. and Pardee, A. B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967-971.
Liu, C. and Meinke, D. W. (1998). The titan mutants of Arabidopsis are disrupted in mitosis and cell cycle control during seed development. *The Plant Journal* **16**, 21-31.
Loken, M. R. and Herzenberg, L. A. (1975). Analysis of cell populations with a fluorescence-activated cell sorter. *Ann. NY Acad. Sci* **254**, 163.
Lord, E. M. and Russell, S. D. (2002). The mechanisms of pollination and fertilization in plants. *Annual Review of Cell and Developmental Biology* **18**, 81-105.
Marton, M. L. and Dresselhaus, T. (2008). A comparison of early molecular fertilization mechanisms in animals and flowering plants. *Sexual Plant Reproduction* **21**, 37-52.
Mascarenhas, J. P. (1990). Gene Activity During Pollen Development. *Annual Review of Plant Physiology and Plant Molecular Biology* **41**, 317-338.

Mathilde, G., Ghislaine, G., Daniel, V. and Georges, P. (2003). The Arabidopsis MEI1 gene encodes a protein with five BRCT domains that is involved in meiosis-specific DNA repair events independent of SPO11-induced DSBs. *The Plant Journal* **35**, 465-475.
Matthys-Rochon, E., Detchepare, S., Wagner, V., Roeckel, P. and Dumas, C. (1988). Isolation and characterization of viable sperm cells from tricellular pollen grains. *Sexual reproduction in higher plants. Springer, Berlin Heidelberg New York*, 245–250.
Matthys-Rochon, E., Vergne, P., Detchepare, S. and Dumas, C. (1987). Male Germ Unit Isolation from Three Tricellular Pollen Species: Brassica oleracea, Zea mays, and Triticum aestivum 1, vol. 83 (ed., pp. 464-466: Am Soc Plant Biol.
Mays-Hoopes, L. L., Bolen, J., Riggs, A. D. and Singer-Sam, J. (1995). Preparation of spermatogonia, spermatocytes, and round spermatids for analysis of gene expression using fluorescence-activated cell sorting. *Biology of Reproduction* **53**, 1003-1011.
Mazumder, B., Seshadri, V. and Fox, P. L. (2003). Translational control by the 3'-UTR: the ends specify the means. *Trends in Biochemical Sciences* **28**, 91-98.
McConchie, C. A., Hough, T. and Knox, R. B. (1987a). Ultrastructural analysis of the sperm cells of mature pollen of maize, Zea mays. *Protoplasma* **139**, 9-19.
McConchie, C. A., Russell, S. D., Dumas, C., Tuohy, M. and Knox, R. B. (1987b). Quantitative Cytology of the Sperm Cells of Brassica campestris and Brassica oleracea. *Planta* **170**, 446-452.
McLeskey, S. B., Dowds, C., Carballada, R., White, R. R. and Saling, P. M. (1998). Molecules involved in mammalian sperm-egg interaction. *Int Rev Cytol* **177**, 57-113.
Millar, A. H., Sweetlove, L. J., Giege, P. and Leaver, C. J. (2001). Analysis of the Arabidopsis Mitochondrial Proteome. *Plant Physiology* **127**, 1711-1727.
Miyado, K., Yamada, G., Yamada, S., Hasuwa, H., Nakamura, Y., Ryu, F., Suzuki, K., Kosai, K., Inoue, K. and Ogura, A. (2000). Requirement of CD9 on the egg plasma membrane for fertilization. *Science* **287**, 321-4.
Mo, Y. S. and Yang, H. Y. (1992). Isolation and fusion of sperm cells in several bicellular pollen species. *Acta Bot Sin* **34**, 688-697.
Mogensen, H. L., Wagner, V. T. and Dumas, C. (1990). Quantitative, three-dimensional ultrastructure of isolated corn (Zea mays) sperm cells. *Protoplasma* **153**, 136-140.
Mori, T., Kuroiwa, H., Higashiyama, T. and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat. Cell Biol* **8**, 64–71.
Morrow, E. H. (2004). How the sperm lost its tail: the evolution of aflagellate sperm. *Biological Reviews* **79**, 795-814.
Mroczek, R. J. and Dawe, R. K. (2003). Distribution of Retroelements in Centromeres and Neocentromeres of Maize. *Genetics* **165**, 809-819.
Murata, T., Tanahashi, T., Nishiyama, T., Yamaguchi, K. and Hasebe, M. (2007). How do Plants Organize Microtubules Without a Centrosome? *Journal of Integrative Plant Biology* **49**, 1154-1163.
Murgia, M., Huang, B. Q., Tucker, S. C. and Musgrave, M. E. (1993). Embryo sac lacking antipodal cells in Arabidopsis thaliana (Brassicaceae). *Am. J. Bot* **80**, 824-838.
Myles, D. G. (1978). The fine structure of fertilization in the fern Marsilea vestita, vol. 30 (ed., pp. 265-281.
Myles, D. G. and Primakoff, P. (1997). Why did the sperm cross the cumulus? To get to the oocyte. Functions of the sperm surface proteins PH-20 and fertilin in arriving at, and fusing with, the egg. *Biol Reprod* **56**, 320-7.
Nair, R. and Rost, B. (2005). Mimicking Cellular Sorting Improves Prediction of Subcellular Localization. *Journal of Molecular Biology* **348**, 85-100.
Noir, S., Brautigam, A., Colby, T., Schmidt, J. and Panstruga, R. (2005). A reference map of the Arabidopsis thaliana mature pollen proteome. *Biochemical and Biophysical Research Communications* **337**, 1257-1266.
Nyman, T., Page, R., Schutt, C. E., Karlsson, R. and Lindberg, U. (2002). A Cross-linked Profilin-Actin Heterodimer Interferes with Elongation at the Fast-growing End of F-actin. *Journal of Biological Chemistry* **277**, 15828-15833.
Oh, S. A., Park, S. K., Jang, I., Howden, R., Moore, J. M., Grossniklaus, U. and Twell, D. (2003). halfman, an Arabidopsis male gametophytic mutant associated with a 150 kb

chromosomal deletion adjacent to an introduced Ds transposable element. *Sexual Plant Reproduction* **16**, 99-102.
Okada, T., Bhalla, P. L. and Singh, M. B. (2006). Expressed sequence tag analysis of Lilium longiflorum generative cells. *Plant and Cell Physiology* **47**, 698-705.
Okada, T., Endo, M., Singh, M. B. and Bhalla, P. L. (2005). Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. *Plant Journal* **44**, 557-568.
Okada, T., Singh, M. B. and Bhalla, P. L. (2007). Transcriptome profiling of Lilium longiflorum generative cells by cDNA microarray. *Plant Cell Reports* **26**, 1045-1052.
Pagnussat, G. C., Yu, H. J., Ngo, Q. A., Rajani, S., Mayalagu, S., Johnson, C. S., Capron, A., Xie, L. F., Ye, D. and Sundaresan, V. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development* **132**, 603-614.
Palanivelu, R., Brass, L., Edlund, A. F. and Preuss, D. (2003). Pollen Tube Growth and Guidance Is Regulated by POP2, an Arabidopsis Gene that Controls GABA Levels. *Cell* **114**, 47-59.
Pang, S. Z., DeBoer, D. L., Wan, Y., Ye, G., Layton, J. G., Neher, M. K., Armstrong, C. L., Fry, J. E., Hinchee, M. A. W. and Fromm, M. E. (1996). An improved green fluorescent protein gene as a vital marker in plants. *Plant Physiol* **112**, 893-900.
Park, S. K., Howden, R. and Twell, D. (1998). The Arabidopsis thaliana gametophytic mutation gemini pollen1 disrupts microspore polarity, division asymmetry and pollen cell fate. *Development* **125**, 3789-3799.
Park, S. Y., Jauh, G. Y., Mollet, J. C., Eckard, K. J., Nothnagel, E. A., Walling, L. L. and Lord, E. M. (2000). A Lipid Transfer-like Protein Is Necessary for Lily Pollen Tube Adhesion to an in Vitro Stylar Matrix. *The Plant Cell Online* **12**, 151-164.
Peirson, B. N., Bowling, S. E. and Makaroff, C. A. (1997). A defect in synapsis causes male sterility in a T-DNA-tagged Arabidopsis thaliana mutant. *The Plant Journal* **11**, 659-669.
Peterson, F. C., Lytle, B. L., Sampath, S., Vinarov, D., Tyler, E., Shahan, M., Markley, J. L. and Volkman, B. F. (2005). Solution structure of thioredoxin h1 from Arabidopsis thaliana. *Protein Science* **14**, 2195.
Pettitt, J. M. (1977). Detection in primitive gymnosperms of proteins and glycoproteins of possible significance in reproduction. *Nature* **266**, 530-532.
Pittoggi, C., Beraldi, R., Sciamanna, I., Barberi, L., Giordano, R., Magnano, A. R., Torosantucci, L., Pescarmona, E. and Spadafora, C. (2006). Generation of biologically active retro-genes upon interaction of mouse spermatozoa with exogenous DNA. *MOLECULAR REPRODUCTION AND DEVELOPMENT* **73**, 1239.
Prasher, D. C. (1995). Using GFP to see the light. *Trends in Genetics* **11**, 320-323.
Procissi, A., de Laissardiere, S., Ferault, M., Vezon, D., Pelletier, G. and Bonhomme, S. (2001). Five Gametophytic Mutations Affecting Pollen Development and Pollen Tube Growth in Arabidopsis thaliana. *Genetics* **158**, 1773-1783.
Qiu, Y. L., Yang, Y. H., Zhang, S. Q. and Tian, H. Q. (2004). Isolation of two populations of sperm cells from the pollen tube of tobacco. *Acta Botanica Sinica* **46**, 719-723.
Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics* **32**, 496-501.
Quiros, C. F., Grellet, F., Sadowski, J., Suzuki, T., Li, G. and Wroblewski, T. (2001). Arabidopsis and Brassica Comparative Genomics Sequence, Structure and Gene Content in the ABI1-Rps2-Ck1 Chromosomal Segment and Related Regions. *Genetics* **157**, 1321-1330.
Ray, S. M., Park, S. S. and Ray, A. (1997). Pollen tube guidance by the female gametophyte. *Development* **124**, 2489-98.
Redei, G. P. (1964). A pollen abortion factor. *Arabidopsis Inf. Serv.* **1**, 10.
Reichheld, J. P., Mestres-Ortega, D., Laloi, C. and Meyer, Y. (2002). The multigenic family of thioredoxin h in Arabidopsis thaliana: specific expression and stress response. *Plant Physiology et Biochemistry* **40**, 685-690.
Renzaglia, K. S., Johnson, T. H., Gates, H. D. and Whittier, D. P. (2001). Architecture of the sperm cell of Psilotum. *American Journal of Botany* **88**, 1151-1163.

Renzaglia, K. S., Rasch, E. M. and Pike, L. M. (1995). Estimates of Nuclear DNA Content in Bryophyte Sperm Cells: Phylogenetic Considerations. *American Journal of Botany* **82**, 18-25.
Rivera-Madrid, R., Mestres, D., Marinho, P., Jacquot, J. P., Decottignies, P., Miginiac-Maslow, M. and Meyer, Y. (1995). Evidence for five divergent thioredoxin h sequences in Arabidopsis thaliana. *Proc Natl Acad Sci US A* **92**, 5620-5624.
Roe, M. R. and Griffin, T. J. (2006). Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes. *Proteomics* **6**, 4678–4687.
Roeckel, P., Dupuis, I., Detchepare, S., Matthys-Rochon, E. and Dumas, C. (1988). Isolation and viability of sperm cells from corn (Zea mays) and kale (Brassica oleracea) pollen grains. *Plant sperm cells as tools for biotechnology. Pudoc, Wageningen*, 105-109.
Rooney, A. P., Zhang, J. and Nei, M. (2000). An Unusual Form of Purifying Selection in a Sperm Protein. *Molecular Biology and Evolution* **17**, 278-283.
Rotman, B. and Papermaster, B. W. (1966). Membrane Properties of Living Mammalian Cells as Studied by Enzymatic Hydrolysis of Fluorogenic Esters. *Proceedings of the National Academy of Sciences* **55**, 134-141.
Rotman, N., Durbarry, A., Wardle, A., Yang, W. C., Chaboud, A., Faure, J. E., Berger, F. and Twell, D. (2005). A Novel Class of MYB Factors Controls Sperm-Cell Formation in Plants. *Current Biology* **15**, 244-248.
Rotman, N., Rozier, F., Boavida, L., Dumas, C., Berger, F. and Faure, J. E. (2003). Female Control of Male Gamete Delivery during Fertilization in Arabidopsis thaliana. *Current Biology* **13**, 432-436.
Rubinstein, E., Ziyyat, A., Wolf, J. P., Le Naour, F. and Boucheix, C. (2006). The molecular players of sperm–egg fusion in mammals. *Seminars in Cell and Developmental Biology* **17**, 254-263.
Russell, S. D. (1983). Fertilization in Plumbago zeylanica: Gametic Fusion and Fate of the Male Cytoplasm. *American Journal of Botany* **70**, 416-434.
Russell, S. D. (1984). Ultrastructure of the sperm of Plumbago zeylanica. II: Quantitative cytology and three-dimensional organization. *Planta* **162**, 385-391.
Russell, S. D. (1985). Preferential Fertilization in Plumbago: Ultrastructural Evidence for Gamete-Level Recognition in an Angiosperm. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 6129-6132.
Russell, S. D. (1986). Isolation of Sperm Cells from the Pollen of Plumbago-Zeylanica. *Plant Physiology* **81**, 317-319.
Russell, S. D. (1991). Isolation and Characterization of Sperm Cells in Flowering Plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **42**, 189-204.
Russell, S. D. (1996). Attraction and transport of male gametes for fertilization. *Sexual Plant Reproduction* **9**, 337-342.
Russell, S. D. (2003). Plant sexuality, cell expression and preferential fertilization. *Bol. Soc. Argent. Bot* **38**, 349-356.
Russell, S. D. and Cass, D. D. (1981). Ultrastructure of the sperms ofPlumbago zeylanica. *Protoplasma* **107**, 85-107.
Russell, S. D. and Strout, G. W. (2005). Microgametogenesis in Plumbago zeylanica (Plumbaginaceae). 2. Quantitative cell and organelle dynamics of the male reproductive cell lineage. *Sexual Plant Reproduction* **18**, 113-130.
Russell, S. D., Strout, G. W., Stramski, A. K., Mislan, T. W., Thompson, R. A. and Schoemann, L. M. (1996). Microgametogenesis in Plumbago zeylanica (Plumbaginaceae). 1. Descriptive cytology and threedimensional organization. *American Journal of Botany* **83**, 1435-1453.
Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. and Barrell, B. (2000). Artemis: sequence visualization and annotation, vol. 16 (ed., pp. 944-945: Oxford Univ Press.
Sanchez, A. M., Bosch, M., Bots, M., Nieuwland, J., Feron, R. and Mariani, C. (2004). Pistil factors controlling pollination. *Plant Cell* **16**, S98-S106.
Santelices, B. (2002). RECENT ADVANCES IN FERTILIZATION ECOLOGY OF MACROALGAE 1. *J Phycol* **38**, 4-10.

Satoh, M., Hosoi, S. and Sato, S. (1990). Chinese hamster ovary cells continuously secrete a cysteine endopeptidase. *In Vitro Cellular & Developmental Biology-Plant* **26**, 1101-1104.
Schwitzguebel, J. P. and Siegenthaler, P. A. (1984). Purification of Peroxisomes and Mitochondria from Spinach Leaf by Percoll Gradient Centrifugation 1. *Plant Physiology* **75**, 670-674.
Scott, R. J. (2007). Polyspermy in apomictic Crataegus: yes and no. *New Phytol* **173**, 227-230.
Sheoran, I. S., Sproule, K. A., Olson, D. J. H., Ross, A. R. S. and Sawhney, V. K. (2006). Proteome profile and functional classification of proteins in Arabidopsis thaliana (Landsberg erecta) mature pollen. *Sexual Plant Reproduction* **19**, 185-196.
Shi, D. Q., Liu, J., Xiang, Y. H., Ye, D., Sundaresan, V. and 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, L., Zhu, T., Mogensen, H. L. and Keim, P. (1996). Sperm Identification in Maize by Fluorescence in Situ Hybridization. *The Plant Cell Online* **8**, 815-821.
Shimizu, K. K. (2000). Attractive and repulsive interactions between female and male gametophytes in Arabidopsis pollen tube guidance, vol. 127 (ed., pp. 4511-4518.
Shimizu, K. K. and Okada, K. (2000). Attractive and repulsive interactions between female and male gametophytes in Arabidopsis pollen tube guidance. *Development* **127**, 4511-8.
Shivanna, K. R., Xu, H., Taylor, P. and Knox, R. B. (1988). Isolation of Sperms from the Pollen Tubes of Flowering Plants during Fertilization 1. *Plant Physiology* **87**, 647-650.
Signori, E., Bagni, C., Papa, S., Primerano, B., Rinaldi, M., Amaldi, F. and Fazio, V. M. (2001). A somatic mutation in the 5'UTR of BRCA1 gene in sporadic breast cancer causes down-modulation of translation efficiency. *Oncogene* **20**, 4596-600.
Singh, M. B., Xu, H., Bhalla, P. L., Zhang, Z., Swoboda, I. and Russell, S. D. (2002). Developmental expression of polyubiquitin genes and distribution of ubiquitinated proteins in generative and sperm cells. *Sexual Plant Reproduction* **14**, 325-329.
Slonim, D. K. (2002). From patterns to pathways: gene expression data analysis comes of age. *Nature Genetics* **32**, 502-508.
Snook, R. R. (1997). Is the Production of Multiple Sperm Types Adaptive? *Evolution* **51**, 797-808.
Southworth, D. and Cresti, M. (1997). Comparison of flagellated and nonflagellated sperm in plants. *American Journal of Botany* **84**, 1301-1301.
Southworth, D. and Kwiatkowski, S. (1996). Arabinogalactan proteins at the cell surface of Brassica sperm and Lilium sperm and generative cells. *Sexual Plant Reproduction* **9**, 269-272.
Southworth, D., Kwiatkowski, S., Smith, A. R., Sharpless, H., Merwin, J. and Marusich, M. F. (1999). Antibodies to flowering-plant sperm. *Protoplasma* **208**, 115-122.
Southworth, D. and Morningstar, P. A. (1992). Isolation of generative cells from pollen of Phoenix dactylifera. *Sexual Plant Reproduction* **5**, 270-274.
Southworth, D., Platt-Aloia, K. A. and Thomson, W. W. (1988). Freeze fracture of sperm and vegetative cells in Zea mays pollen. *J. Ultrastruct. Mol. Struc. Res.* **101**, 165-172.
Southworth, D., Salvatici, P. and Cresti, M. (1994). Freeze Fracture of Membranes at the Interface between Vegetative and Generative Cells in Amaryllis Pollen. *International Journal of Plant Sciences* **155**, 538-544.
Southworth, D., Strout, G. and Russell, S. D. (1997). Freeze-fracture of sperm of Plumbago zeylanica L. in pollen and in vitro. *Sexual Plant Reproduction* **10**, 217-226.
Stinson, J. R., Eisenberg, A. J., Willing, R. P., Pe, M. E., Hanson, D. D. and Mascarenhas, J. P. (1987). Genes expressed in the male gametophyte of flowering plants and their isolation. *Plant Physiol* **83**, 442-447.
Stoeckert, C. J., Causton, H. C. and Ball, C. A. (2002). Microarray databases: standards and ontologies. *Nature Genetics* **32**, 469-473.
Sun, M., Zhou, G., Lee, S., Chen, J., Shi, R. Z. and Wang, S. M. (2004). SAGE is far more sensitive than EST for detecting low-abundance transcripts. *feedback*.
Swanson, R., Edlund, A. F. and Preuss, D. (2004). SPECIES SPECIFICITY IN POLLEN-PISTIL INTERACTIONS. *Annual Review of Genetics* **38**, 793-818.

Taylor, P. E., Glover, J. A., Lavithis, M., Craig, S., Singh, M. B., Knox, R. B., Dennis, E. S. and Chaudhury, A. M. (1998). Genetic control of male fertility in Arabidopsis thaliana: structural analyses of postmeiotic developmental mutants. *Planta* **205**, 492-505.
Tian, H. Q. and Russell, S. D. (1998). The fusion of sperm cells and the function of male germ unit (MGU) of tobacco (Nicotiana tabacum L.). *Sexual Plant Reproduction* **11**, 171-176.
Tian, H. Q., Yuan, T. and Russell, S. D. (2005). Relationship between double fertilization and the cell cycle in male and female gametes of tobacco. *Sexual Plant Reproduction* **17**, 243-252.
Tian, H. Q., Zhang, Z. J. and Russell, S. D. (2001). Sperm dimorphism in Nicotiana tabacum L. *Sexual Plant Reproduction* **14**, 123-125.
Till-Bottraud, I., Joly, D., Lachaise, D. and Snook, R. R. (2005). Pollen and sperm heteromorphism: convergence across kingdoms? *Journal of Evolutionary Biology* **18**, 1-18.
Twell, D. (1992). Use of a nuclear-targeted b-glucuronidase fusion protein to demonstrate vegetative cell-specific gene expression in developing pollen. *Plant J* **2**, 887-892.
van Aelst, A. C., Theunis, C. H. and van Went, J. L. (1990). Freeze-fracture studies on isolated sperm cells ofSpinacia oleracea L. *Protoplasma* **153**, 204-207.
Van der Maas, H. M., de Jong, E. R., Van Aelst, A. C., Verhoeven, H. A., Krens, F. A. and Van Went, J. L. (1994). Cytological characterization of isolated sperm cells of perennial ryegrass (Lolium perenne L.). *Protoplasma* **178**, 48-56.
van Ingen, C. (2002). Biology's century: just the beginning for microarrays. *Nature Genetics* **32**, 463-463.
Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995). Serial Analysis of Gene Expression. *Science* **270**, 484.
Vergne, P. and Dumas, C. (1988). Isolation of Viable Wheat Male Gametophytes of Different Stages of Development and Variations in Their Protein Patterns 1, vol. 88 (ed., pp. 969-972: Am Soc Plant Biol.
von Besser, K., Frank, A. C., Johnson, M. A. and Preuss, D. (2006). Arabidopsis HAP2 (GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. *Development* **133**, 4761.
Voronin, V., Touraev, A., Kieft, H., van Lammeren, A. A. M., Heberle-Bors, E. and Wilson, C. (2001). Temporal and tissue-specific expression of the tobacco ntf4 MAP kinase. *Plant Molecular Biology* **45**, 679-689.
Wagner, V. T., Dumas, C. and Mogensen, H. L. (1989). Morphometric analysis of isolated Zea mays sperm. *Journal of Cell Science* **93**, 179-184.
Wang, Y. Y., Kuang, A., Russell, S. D. and Tian, H. Q. (2006). In vitro fertilization as a tool for investigating sexual reproduction of angiosperms. *Sexual Plant Reproduction* **19**, 103-115.
Wassarman, P. M. (1999). Mammalian Fertilization: Molecular Aspects of Gamete Adhesion, Exocytosis, and Fusion. *Cell* **96**, 175-183.
Wassarman, P. M., Jovine, L. and Litscher, E. S. (2001). A profile of fertilization in mammals. *Nature Cell Biology* **3**, E59-E64.
Wassarman, P. M., Jovine, L., Litscher, E. S., Qi, H. and Williams, Z. (2004). Egg–sperm interactions at fertilization in mammals. *European Journal of Obstetrics and Gynecology* **115**, 57-60.
Wassarman, P. M., Jovine, L., Qi, H. Y., Williams, Z., Darie, C. and Litscher, E. S. (2005). Recent aspects of mammalian fertilization research. *Molecular and Cellular Endocrinology* **234**, 95-103.
Werhahn, W. and Braun, H. P. (2002). Biochemical dissection of the mitochondrial proteome from Arabidopsis thaliana by three-dimensional gel electrophoresis. *Electrophoresis* **23**, 640-6.
Weterings, K. and Russell, S. D. (2004a). Experimental Analysis of the Fertilization Process. *The Plant Cell Online* **16**, 107-118.
Weterings, K. and Russell, S. D. (2004b). Experimental analysis of the fertilization process. *Plant Cell* **16**, S107-S118.
Wheeler, C. H., Dunn, M. J., Rheumatology, H., London, U. K. and Pharmacia, A. (2000). A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry. *Electrophoresis* **21**, 3666-3672.

White, K. P., Rifkin, S. A., Hurban, P. and Hogness, D. S. (1999). Microarray Analysis of Drosophila Development During Metamorphosis. *Science* **286**, 2179.
Wilhelmi, L. K. and Preuss, D. (1996). Self-sterility in Arabidopsis due to defective pollen tube guidance. *Science* **274**, 1535-7.
Willing, R. P., Bashe, D. and Mascarenhas, J. P. (1988). An analysis of the quantity and diversity of messenger RNAs from pollen and shoots of Zea mays. *TAG Theoretical and Applied Genetics* **75**, 751-753.
Wilms, H. J., Leferink-Ten Klooster, H. B. and Van Aelst, A. C. (1986). Isolation of Spinach Spinacia-Oleracea Sperm Cells 1. Ultrastructure and Three-Dimensional Construction in the Mature Pollen Grain. In *Mulcahy, D. L., G.B. Mulcahy and E. Ottaviano*, pp. 307-312.
Wilson, C., Voronin, V., Touraev, A., Vicente, O. and Heberle-Bors, E. (1997). A developmentally regulated MAP kinase activated by hydration in tobacco pollen. *Plant Cell* **9**, 2093-2100.
Wilson, Z. A., Morroll, S. M., Dawson, J., Swarup, R. and Tighe, P. J. (2001). The Arabidopsis MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. *The Plant Journal* **28**, 27-39.
Wong, J. L. and Wessel, G. M. (2006). Defending the zygote: search for the ancestral animal block to polyspermy. *Curr. Top. Dev. Biol* **72**, 1–151.
Wouters-Tyrou, D., Martinage, A., Chevaillier, P. and Sautire, P. (1998). Nuclear basic proteins in spermiogenesis. *Biochimie* **80**, 117-128.
Wrzaczek, M. and Hirt, H. (2001). Plant MAP kinase pathways: how many and what for. *Biol Cell* **93**, 81-7.
Wu, S. M., Baxendale, V., Chen, Y., Lap-Yin Pang, A., Stitely, T., Munson, P. J., Yiu-Kwong Leung, M., Ravindranath, N., Dym, M. and Rennert, O. M. (2004). Analysis of mouse germ-cell transcriptome at different stages of spermatogenesis by SAGE: Biological significance. *Genomics* **84**, 971-981.
Xu, H., Swoboda, I., Bhalla, P. L. and Singh, M. B. (1999a). Male gametic cell-specific expression of H2A and H3 histone genes. *Plant Molecular Biology* **39**, 607-614.
Xu, H., Swoboda, I., Bhalla, P. L. and Singh, M. B. (1999b). Male gametic cell-specific gene expression in flowering plants. *Proc Natl Acad Sci U S A.* **96**, 2554-8.
Xu, H., Weterings, K., Vriezen, W., Feron, R., Xue, Y., Derksen, J. and Mariani, C. (2002). Isolation and characterization of male-germ-cell transcripts in Nicotiana tabacum. *Sexual Plant Reproduction* **14**, 339-346.
Xu, H. P. and Tsao, T. H. (1997). Detection and immunolocalization of glycoproteins of the plasma membrane of maize sperm cells. *Protoplasma* **198**, 125-129.
Xu, X. Z. S. and Sternberg, P. W. (2003). A C. elegans Sperm TRP Protein Required for Sperm-Egg Interactions during Fertilization. *Cell* **114**, 285-297.
Yadegari, R. and Drews, G. N. (2004). Female Gametophyte Development. *The Plant Cell Online* **16**, 133-141.
Yang, H. Y. and Zhou, C. (1989). Isolation of viable sperms from pollen of Brassica napus, Zea mays and Secale cereale. *Chin J Bot* **1**, 80-84.
Yang, W. C. and Sundaresan, V. (2000). Genetics of gametophyte biogenesis in Arabidopsis. *Current Opinion in Plant Biology* **3**, 53-57.
Ye, G. N., Stone, D., Pang, S. Z., Creely, W., Gonzalez, K. and Hinchee, M. (1999). Arabidopsis ovule is the target for Agrobacterium in planta vacuum infiltration transformation. *The Plant Journal* **19**, 249-257.
Ye, X. L., Yeung, E. C. and Zee, S. Y. (2002). Sperm movement during double fertilization of a flowering plant, Phaius tankervilliae. *Planta* **215**, 60-66.
Yu, H. S., Hu, S. Y. and Russell, S. D. (1992). Sperm Cells in Pollen Tubes of Nicotiana-Tabacum-L - 3-Dimensional Reconstruction, Cytoplasmic Diminution, and Quantitative Cytology. *Protoplasma* **168**, 172-183.
Yu, H. S. and Russell, S. D. (1993). Three-dimensional ultrastructure of generative cell mitosis in the pollen tube of Nicotiana tabacum. *Eur J Cell Biol.* **61**, 338-48.
Yu, H. S. and Russell, S. D. (1994a). Occurrence of Mitochondria in the Nuclei of Tobacco Sperm Cells. *Plant Cell* **6**, 1477-1484.

Yu, H. S. and Russell, S. D. (1994b). Quantitative Differences in Sperm Cells and Organelles of Tobacco (Nicotiana-Tabacum-L) Grown under Differing Environmental-Conditions. *Theoretical and Applied Genetics* **89**, 814-817.
Zhang, G., Williams, C. M., Campenot, M. K., McGann, L. E. and Cass, D. D. (1992a). Improvement of Longevity and Viability of Sperm Cells Isolated from Pollen of Zea mays L. 1. *Plant Physiology* **100**, 47-53.
Zhang, G. C., Campenot, M. K., McGann, L. E. and Cass, D. D. (1992b). Flow Cytometric Characteristics of Sperm Cells Isolated from Pollen of Zea-Mays L. *Plant Physiology* **99**, 54-59.
Zhang, Z., Tian, H. Q. and Russell, S. D. (1999). Localization of myosin on sperm-cell-associated membranes of tobacco (Nicotiana tabacum L.). *Protoplasma* **208**, 123-128.
Zhang, Z. J. and Russell, S. D. (1999). Sperm cell surface characteristics of Plumbago zeylanica L-in relation to transport in the embryo sac. *Planta* **208**, 539-544.
Zhang, Z. J., Xu, H. L., Singh, M. B. and Russell, S. D. (1998). Isolation and collection of two populations of viable sperm cells from the pollen of Plumbago zeylanica. *Zygote* **6**, 295-298.
Zimmermann, P., Hennig, L. and Gruissem, W. (2005). Gene-expression analysis and network discovery using Genevestigator. *Trends in Plant Science* **10**, 407-409.
Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. (2004). GENEVESTIGATOR. Arabidopsis Microarray Database and Analysis Toolbox. *Plant Physiology* **136**, 2621.

Appendices

Appendix A: Primer sequences

1 RT-PCR primers

Locus Tag	Description	Primer sequence	Product Size (bp)	Genomic DNA product size (bp)
At1g10090	unknown protein	LP-CCTTAATCCACGCTCACGAT RP-TGAATGAAGCCATCACGGTA	285	557
At1g11250	syntaxin, putative (SYP95)	LP-CGAGAATGTGAAGGACGACA RP-GCTCTTGAATCGCTTTTTGG	461	547
At1g32400	senescence-associated family protein	LP-TGTTCGAGGAGTGTCTGCTG RP-CCATAAGGGCAAGCAAGAAG	244	692
At1g69960	serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	LP-CCGGAGATATCGATCGTCAG RP-TTGACGGCTTTCATGATTCC	347	775
At1g73070	leucine-rich repeat family protein	LP-AGTGGCACCATTTGTGTTGA RP-GACTTGGCCATCGTTCAGTT	291	681
At1g78300	14-3-3 protein GF14 omega (GRF2)	LP-TCGAGCAGAAGGAAGAGAGC RP-TTCAATGCAAGACCAAGACG	331	508
At2g18390	ADP-ribosylation factor-like protein 2 (ARL2)	LP-CTGGGAAGACGACGATTGTT RP-CTTGCCAATTTCATCAGGTG	311	567
At2g19380	RNA recognition motif (RRM)-containing protein	LP-CATAATCAATGCATCACCGA RP-AACAAAGTCTGCTCGCTGGT	371	715
At2g20440	RabGAP/TBC domain-containing protein	LP-CGGTGCATGGAAAGAAGAAT RP-GACACCCATTGATGTTGCTG	442	701
At2g22740	SET domain-containing protein (SUVH6)	LP-CGATCGCAGACGAAGATCTT RP-CTAGCGGCTGAGTTTGATCC	595	823
At2g29940	ABC transporter family protein	LP-TGATGGAAACGTGGTTCAAA RP-AGAAGCAACGTCGATTTTCC	399	670
At3g20190	leucine-rich repeat transmembrane protein kinase, putative	LP-CAATGACCTCGAGGGACCTA RP-TCTGATCCGGTACTGCTCCT	392	491
At3g23610	dual specificity protein phosphatase (DsPTP1)	LP-CGACAATGTCCCTTCCCTTA RP-GCCAACAAAGCAATGAACAA	274	477
At3g61230	LIM domain-containing protein	LP-TGCAAGACCCATTTTGAACA RP-TTGGGTTCGGTTTTGTCTTC	401	580
At4g11920	WD-40 repeat family protein	LP-AAGAAGATGGGGCTGGTTCT RP-ACAGTTTCATCGACCCCAAG	397	816
At4g28000	AAA-type ATPase family protein	LP-ACAAGTCAGCCTCCACGACT RP-AGTCTTGAGCCGAGAACCAA	299	504
At4g32830	kinase/ protein serine/threonine kinase	LP-GCGAAAGACGAGCTGCTACT RP-AGAATCCCAAGGCTCCAGAT	265	522
At5g23130	peptidoglycan-binding LysM domain-containing protein	LP-CCGAATACCATTACCTGGAA RP-CAACCTCAAAGACTGGAAGG	132	713
At5g39650	unknown protein	LP-ACCATCACCATCACCACCTT RP-CAAAACGCTCATCACTGCAT	499	499
At5g50930	unknown protein	LP-GATGATCGAGGAAGCTGGAC RP-CTAAATCTGCCACACATGCC	447	959
At5g49150	GEX2 (GENAMETE EXPRESSED2)	LP-TGGATCAAGTCACCAAGCTG RP-GGCTGTGACGGTACAGGATT	369	466

At5g54640	Histone H2A	LP-GGTAAATACGCCGAACGTGT RP-CAAGCTGAATGTGACGAGGA	148	495
At3g50910	unknown protein	LP-CTCAGCTTTCTTCCGTGGAG RP-GTTTCCAACCGTACGAGCAT	516	914
At1g14420	AT59 (Arabidopsis homolog of tomato LAT59); lyase/ pectate lyase	LP-ACGGTGGGTTGATCAGAGAC RP-TCCTTCTGATTGCCAATTCC	506	603
AT1G64740	TUA1 TUA1 (ALPHA-1 TUBULIN)	LP-GCTGGTATTCAGGTCGGAAA RP-TCAACAATTTCCCTTCCAA	314	986
AT5G39400	pollen specific phosphatase, putative / phosphatase and tensin, putative (PTEN1)	LP-AAGTCAAGTCCGTGCTCGAT RP-TAACCTCGGGAGGTCCTTTT	417	614
	AtNOV2 exon1	GAAGGAGCTGAGAAGACAC		
	AtNOV2 exon5 sense	GTTGTCACTTGCAGAGAAG		
	AtNOV2 exon5 antisense	CTTCTCTGCAAGTGACAAC		
	AtNOV2 exon10 antisense	TCCTGGCTTCTTATTCGC		

2 GFP construct primers

Locus Tag	Description	Primer sequence	Product Size (bp)	RE
At2g22740	SET domain-containing protein (SUVH6)	LP-CCACTGTCGACACCACTAGGCCATCAT	1821	SalI
		RP-CCATTGGATCCAGCAATCTTCACTTCCAT		BamHI
At1g10090	unknown protein	LP-CAAAAGTCGACATGTCGTATTCCGTAGAGAC	2091	SalI
		RP-CAATAGGATCCTGGCTGCTTTCGTAG		BamHI
At3g20190	leucine-rich repeat transmembrane protein kinase, putative	LP-ATCTGGTCGACCTGCTACTTGCATTGCC	1439	SalI
		RP-CTGCTGGATCCCGATGCATTAACCAAAGTATC		BamHI
At4g11920	WD-40 repeat family protein	LP-CATATGTCGACGAATATCAGGTATGATTTCTCTC	1322	SalI
		RP-TTCTCGGATCCCACCGGACCAAAAAG		BamHI
At4g32830	kinase/ protein serine/threonine kinase	LP-ATTACGTCGACGATACTCGAAGCTGGTAA	1462	SalI
		RP-GGCTTGGATCCGTCGAAATCGCTTAAAGT		BamHI
At5g39650	expressed protein	LP-ATTTAGTCGACGAACTCTAACTCTATTCAAGCG	1922	SalI
		RP-AGTCCGGATCCCGTCACGAAACCGTAGTAGAT		BamHI
At5g49150	GEX2 (GENAMETE EXPRESSED2)	LP-GGTCTGTCGACTTACATCGGATGGATTCAC	1886	SalI
		RP-CTAACTGATCAGTACATTAACCCTTCACAACAAGA		BclI
At3g50910	unknown protein	LP-CCGTGGTCGACAAGATAAGTTCAAGGGTC	2032	SalI
		RP-TTATAGGATCCAACCGTACGAGCATGAAC		BamHI
	GFP antisense	CGGACACGCTGAACTTGT		
	GFP middle sense	CCATCTTCTTCAAGGACGACG		
	GFP middle antisense	CGTCGTCCTTGAAGAAGATGG		
	pBI b4 RE sense	GCTCACTCATTAGGCACC		

3 T-DNA insertion line primers

Locus Tag	Description	T-DNA line	Primer sequence
		Lba1	TGGTTCACGTAGTGGGCCATCG
			GCCTTTTCAGAAATGGATAAAT
		LB1	AGCCTTGCTTCC
At4g32830	histone serine kinase(H3-S10 specific) / kinase/ protein serine/threonine kinase; ATAUR1 (ATAURORA1)	SALK_031697.47.35.x	LP-TTGAGTATGCTGCTAGAGGCG
			RP-TTTCCTTCTTCCGTTAATGCC
At4g32830	histone serine kinase(H3-S10 specific) / kinase/ protein serine/threonine kinase; ATAUR1 (ATAURORA1)	SALK_112121.37.70.n	LP-TTCATTGTGCCAGTATGTTGC
			RP-TTTCCTTCTTCCGTTAATGCC
At4g11920	WD-40 repeat family protein	SALK_143106.49.40.x	LP-TTCAACACAACACAATTTGGTG
			RP-CACACGCAGGTGTGTAATTTG
At4g11920	WD-40 repeat family protein	SALK_001978.45.25.x	LP-CCTTTTCACTCTGGGTCAAGG
			RP-GACTTCTTGCTTCTGGTGGTG
At4g11920	WD-40 repeat family protein	SALK_073708.44.35.x	LP-AATGGAAGGTCATCGGCTAAG
			RP-GGAGAAGGGAAGACATTCCAG
At1g10090	unknown protein	SALK_131951.50.30.x	LP-TTAATGCAAGGTCTCGACGAC
			RP-GGAGGCTGCTAGGCAGTTG
At3g50910	unknown protein	SALK_074693.54.75.x	LP-ATCTTCCTCAGCTTTCTTCCG
			RP-GCTGATTCACCTGGGTCATAG
At5g23130	peptidoglycan-binding LysM domain-containing protein	SALK_111440.51.35.x	LP-GCCATTAAGTATGGCGTTGAG
			RP-TGTAGACTGCCATCCCAAAAC
At1g73070	leucine-rich repeat family protein	SALK_007150.56.00.x	LP-AATTGGCGTAACTGAACGATG
			RP-GCTTGCTGCAGTTTCCAATAG
At2g29940	ABC transporter family protein	SALK_003164.53.50.x	LP-CGAATAATACAGCCCAACACG

			RP-CTTCCTCATCTTGCTCCACTG
At3g23610	dual specificity protein phosphatase (DsPTP1)	SALK_092811.54.50.x	LP-TCTACCAACCAAATGACCAATG
			RP-CATATTCCAAATTCGTTTCCATG
At1g78300	14-3-3 protein GF14 omega	SAIL_904_D03	LP-TCATTAGTGGAGAGTGGTGGG
			RP-CCTTGATCATGCAGCTTCTTC
At4g28000	AAA-type ATPase family protein	SALK_003063.54.75.x	LP-TGAGTTGACAATGGACAAAATG
			RP-AGAACCTCGGCTATGTGGTTC
At4g28000	AAA-type ATPase family protein	SALK_074465.55.50.x	LP-GATTGAGAAGCGACTTGCTTG
			RP-TCCAGGAGAGGAGGAAGAGAG
At5g54640	Histone H2A	SALK_040809.18.50.x	LP-CACACCACATTGACTGGTCTG
			RP-ATTTTTGCATGTCAAATTCGG
			LP-
At2g19380	RNA recognition motif (RRM)-containing protein	SAIL_240_D08	AAAAATAAGGGCTCTTTTTCATC
			RP-GCAGTGCACATACTCATGTG
At2g20440	RabGAP/TBC domain-containing protein	SALK_006096.39.60.x	LP-AAAATGGGAACCCAATAGACG
			RP-TGCGTTGTTTCGGTTTTCTAG
At2g20440	RabGAP/TBC domain-containing protein	SALK_006098.17.05.x	LP-AAAATGGGAACCCAATAGACG
			RP-TACCCAAACAGCAGATAACGC
At1g69960	serine/threonine protein phosphatase PP2A-5 catalytic subunit (PP2A5)	SALK_139822.45.55.x	LP-ATTGAATCCAAACACAAACCG
			RP-TGCTTCCATACATTAGCATTTCC
At5g50930	unknown protein	SALK_009229.54.50.x	LP-GGTTATTGGACCAGAAGACAGC
			RP-CAGACGAAGCTTCATCGAATG
At5g26150	protein kinase family protein	SALK_129264.44.55.x	LP-TAAGGCATCCTCACATGGTTC
			RP-AAGATTTGGCCTTGTAGGTCG
			LP-
At4g10810	unknown protein	SAIL_869_C07	TCACTGCACAAGTTGTAAAAACG
			RP-AGAAACACCAAATCTGGCATG
At4g10810	unknown protein	SALK_138312.17.60.x	LP-AGAAACACCAAATCTGGCATG
			RP-
			TCACTGCACAAGTTGTAAAAACG
At1g48940	plastocyanin-like domain-containing protein	SAIL_657_G08	LP-AATGGAGATGAATCGATGCAC
			RP-AAAACCGTATCATCGTGGTTG
At3g55430	glycosyl hydrolase family 17 protein / beta-1,3-glucanase, putative	SALK_142531.34.70.x	LP-AACGTAGAATGTTGCGTTTGC
			RP-ATAACAAAAGATGCGTGCGTC
At3g55430	glycosyl hydrolase family 17 protein / beta-1,3-glucanase, putative	SAIL_674_C09	LP-TCCTCCTACAAACACGTGTCC
			RP-AAATGTTAACCGAAACCCACC
At5g53820	unknown protein	SALK_095127.26.55.x	LP-TATTTGTTATGGTGGCCATGC
			RP-ATTGCCTAAATGGCTTCGATC
At5g53820	unknown protein	SALK_027524.27.95.x	LP-GACGAGAAATCGAAGCATACG
			RP-AACCTGAGTTTGGCCTTTAGC
At1g03250	unknown protein	SALK_017925.46.35.x	LP-GGTTTTTGAATGATGTGGGTG
			RP-TTTGGCTGTCCGATACGATAG
At1g26750	unknown protein	SALK_007677.48.20.x	LP-GCAAATGGAAAAACAATGGAC
			RP-AACACAGGAGGTGGTGATCTG
At3g05820	beta-fructofuranosidase, putative / invertase, putative / saccharase, putative / beta-fructosidase, putative	SALK_016334.30.30.x	LP-TGTGTTGTGGTTCCAGAGTTG
			RP-TTGGTTCTGTTTTGGTGTTCC
At5g59370	actin 4 (ACT4)	SALK_031941.55.50.x	LP-GACCACATCCATTTAACCGAG
			RP-CTTGACCTAGCAGGTCGTGAC
At3g22760	CXC domain containing TSO1-like protein 1 (SOL1)	SALK_007957.55.75.x	LP-AAAGGCTTGAATATCTCTGAACG
			RP-GGCTTTACACAAAACCGTACG
At3g22760	CXC domain containing TSO1-like protein 1 (SOL1)	SAIL_742_H03	LP-TGATTAGCAATATTCAGCCAGC
			RP-CTTTATGAGAAACCGCGTGAG
At3g53750	actin 3 (ACT3)	SALK_083433.55.00.x	LP-AAGCTTGATGACGTGATTTGC
			RP-GATGAAGATCCTCACTGAGCG
At1g03180	unknown protein	SALK_019248.54.00.x	LP-ACAAAGCCCTAACCCAGCTAG
			RP-TGCCTTACTGTGTCATTGCAG
At4g04930	fatty acid desaturase family protein	SALK_107761.42.15.x	LP-AGGTTGTCGAAGAAACAAACG
			RP-TGATGAACTCCCAGTAGCCAG
At5g57000	unknown protein	SALK_112229.51.75.x	LP-TTCACACTTGCTACAAATCCAC
			RP-ATATGTTGAAAAAGGCGAGGG
At2g46170	reticulon family protein (RTNLB5)	SAIL_736_E02	LP-GCACTATCCTGTGGAGCGTAG
			RP-TGGATCGATTCAAGAAAATGC
At3g06580	galactokinase (GAL1)	SAIL_315_F01	LP-TTGAACCCTAAACCCTATGGC
			RP-TATCCTGACGAATAGCCATCG
At5g05490	cohesion family protein SYN1, splice variant 1 (SYN1)	ALK_091193.52.65.x	LP-CTTCTTAAGGATGGCCGCTAC
			RP-CCACTTTTATGGGCAATGAAG
At1g55530	zinc finger (C3HC4-type RING finger) family protein	SAIL_628_E10	LP-TTGTTGTGCTCCATTTCCTTC
			RP-AACAACGAGCGATAACAATGG
At3g51030	thioredoxin H-type 1 (TRX-H-1)	SALK_044578.56.00.x	LP-AGGGTTTTCAGCTTGTATCCG
			RP-GCTCCATTCTTTGCTGATTTG
At5g16880	VHS domain-containing protein / GAT domain-containing protein	SALK_003755.54.25.x	LP-TGAGCAACAAGATCAGTGCTG
			RP-CACATTCAAGGCTTCGAAAAG
At3g56640	exocyst complex subunit Sec15-like family protein	SALK_064121.55.50.x	LP-GGTTGGAAGACGGAAAGAATC
			RP-ACGGAAATTGTCACTTGCAAG

At3g18040	mitogen-activated protein kinase, putative / MAPK, putative (MPK9) [*Arabidopsis thaliana*].	SALK_064439.41.10.x	LP-TGGACCTCGTAGGTTTTGTTG RP-CCAAAGTCGCAAATCTTCAAC
At3g23870	permease-related	SALK_131139.51.80.x	LP-ACCGCTGTGTTGAATGTATCC RP-TGGATACTTGAAGGAGCCATG
At4g20325	unannotated (later identified as unknown protein similar to Os04g0476400 [Oryza sativa]	SALK_127028.37.50.n	LP-CAGCAACGGAGACAGAGAGTC RP-GGGAGGCACTATGAGGTTAGG

4 T-DNA line RT-PCR

Locus Tag	Description	T-DNA line	Primer sequence	Product Size (bp)	Genomic DNA product size (bp)	T-DNA region	RT region
At2g20440	Rab/GAP	N506098	LP-CGGTGCATGGAAAGAAGAAT RP-GACACCCATTGATGTTGCTG	442	701	exon7	e5-8
At1g26750	unknown protein	N507677	LP-ACGAGAGCCTGTTTGGCTTA RP-ATGCCATGCTCCTTCAATTC	228	690	5'UTR	e1-3
At5g53820	unknown protein	N527524	LP-CAGTCAGAGCATGAGCTTCAA RP-AGAAATAACTGCATGGCTACAAA	413	601	promoter	e1-3
At3g51030	TRX-h-1	N544578	LP-TTGAGACATGGAACGAGCAG RP-CCTGTATCGCCCAATCACTT	201	689	3'UTR	e1-3
At3g18040	MPK9	N564439	LP-TCATCCTCAGATGCTGCAAG RP-GATTCCCCGTACTGCAAAAA	408	694	exon2	e8-11
At3g50910	unknown protein	N574693	LP-CTCAGCTTTCTTCCGTGGAG RP-GTTTCCAACCGTACGAGCAT	516	914	exon1	e1-2
At3g53750	ACT3	N583433	LP-CAAGTATTGTTGGCCGTCCT RP-GCTCAGCTGTTGTGGTGAAA	522	605	exon3	e3-4
At3g23870	permease	N631139	LP-CAGGTGAAGGAGGGTATGGA RP-GCCACCACAATCACAATCAC	352	604	exon3	e2-5
At4g20325	unknown protein	N826057	LP-GAAGGAGCTGAGAAGACAC RP-CTTCTCTGCAAGTGACAAC	366	989	intron1	e1-5
At1g10090	unknown protein	N631951	LP-CCTTAATCCACGCTCACGAT RP-TGAATGAAGCCATCACGGTA	285	557	exon7	e3-5

Appendix B: Media, solution, and buffer recipes

All reagents were supplied by Sigma-Aldrich otherwise stated.

1 Sperm cell isolation

1.1 Sperm cell isolation

1.1.1 Brassica oleracea sperm isolation buffer

12.5% sucrose, 5% sorbitol, 0.1 g/l KNO_3 0.36 g/l $CaCl_2$, 1.95g/l MES, adjust to pH6.3.

1.1.2 Arabidopsis thaliana sperm isolation buffer M1 (BK pollen germination medium)

100 ppm H_3BO_3, 300 ppm $Ca(NO_3)_2.4H_2O$, 200 ppm $MgSO_4.7H_2O$, 100 ppm KNO_3, 10% sucrose.

1.1.3 Arabidopsis thaliana sperm isolation buffer M2 (Fei and Nelson, 2003)

1 M sucrose, 1 mM H_3BO_3, 2 mM $CaCl_2$.

1.1.4 Arabidopsis thaliana sperm isolation buffer M3 *(Zhang et al., 1992)*

0.44 M sucrose, 2 mM MES, 0.1% bovine serum albumin (BSA), adjust to pH 6.7.

1.1.5 Arabidopsis thaliana sperm isolation buffer M4 *(Li et al., 1999)*

0.01% H_3BO_3, 1 mM $MgSO_4$, 2 mM $CaCl_2$, 18% sucrose, adjust to pH 6.5.

1.1.6 Arabidopsis thaliana sperm isolation buffer M5 *(Adapted from Mo and Yang 1991)*

8.5% sucrose, 5% sorbitol, 0.1 g/l KNO_3, 0.36 g/l $CaCl_2$.2H2O, 1.95 g/l MES, adjust to pH6.7.

1.1.7 Arabidopsis thaliana sperm isolation buffer M6 *(Mo and Yang, 1991)*

12.5% sucrose, 0.1 g/l KNO_3, 0.36 g/l $CaCl_2$.2H2O, 0.3% potassium dextran sulphate (PDS), 0.6% BSA, 0.3% polyvinylpyrrolidon (PVP).

1.1.8 Arabidopsis thaliana sperm isolation buffer M7 *(Fan et al., 2001)*

5mM MES (pH5.8), 1 mM KCl, 10 mM CaCl2, 8 mM MgSO4, 1.5 mM H3BO3, 16.6% sucrose, 3.65% sorbitol, 10 µg/ml myoinositol.

1.2 Pollen germination media

0.8% agar, 15% sucrose, 80 ppm boric acid, 10 ppm myoinositol, and 0.36 mg/ml $CaCl_2$, and adjust to pH 5.8-6.0.

2 <u>Proteomic studies</u>

2.1 SDS PAGE

2.1.1 12% resolving gel

3 ml 40% acrylamide (Biorad), 1.70 ml 2% bis-acrylamide (Biorad), 2.6 ml H_2O, 2.5 ml 1.5M Tris-HCl pH 8.8, 100 µl 10% SDS, 50 µl 10% APS, and 5 µl TEMED (BDH).

2.1.2 5% stacking gel

485 µl 40% acrylamide (Biorad), 250µl 2% bis-acrylamide (Biorad), 3,575 ml H_2O, 0.625 ml 0.5M Tris-HCl pH 6.8, 50 µl 10% SDS, 25 µl 10% APS (Biorad), and 5 µl TEMED (BDH).

2.1.3 TGS buffer

3.03g Tris, 14.4g glycine, 1g/l SDS, add ddH_20 to 1 liter, and adjust to pH 8.3.

2.1.4 SDS sample buffer

0.06M Tris-HCl, 5% glycerol, 2% SDS, 4% β-mercaptoethanol (BDH), and 0.0025% bromophenol blue.

2.1.5 Coomassie staining

250ml methanol (Fisher scientific), 100ml acetic acid (Fisher scientific), and 1g Coomassie R-250.

2.1.6 Coomassie destaining solution

100ml acetic acid (Fisher scientific), 250ml methanol (Fisher scientific), and 650 ml ddH_2O.

2.2 Two-dimensional protein gel electrophoresis

2.2.1 CHAPS cell lysis buffer

7 M Urea, 2 M Thiourea, 4% CHAPS, 36 mM DTT, 2.5 mM TCEP, 1 x Protease inhibitor mix (Amersham)

2.2.2 Rehydration buffer

7 M Urea, 2M Thiourea, 4% CHAPS, 0.002% bromphenolblue, (18 mM DTT and 0.5% IPG 4-7 were added as final concentration to the rehydration solution [rehydration buffer + sample]).

2.2.3 SDS equilibration solution

50 mM Tris (pH8.8), 6M urea, 30% glycerol (87%), 2% SDS, 1% bromophenol blue.

3 Transcriptional analysis

3.1 LB (Luria-Bertani) medium

10g Bacto-Tryptone, 5g yeast extract, 10g NaCl, 15g agar (optional), and ddH_2O to 1litre.

3.2 2×YT medium

16g tryptone, 10g yeast extract, 5g NaCl, and 15g agar (optional), and ddH_2O to 1litre.

3.3 Floral dipping solution

5% sucrose, 10mM MgCl2, and 0.03% Silwet L-77 (Lehle Seeds).

3.4 Plant MS growth medium

8g/l agar, 4.4g/l Murashige and Skoog (MS) basal medium with Gamborg's vitamins, and 1% sucrose, adjust to pH 5.8, and ddH_2O to 1litre.

4 Analysis of T-DNA insertion mutagenesis plant lines

4.1 Arabidopsis DNA extraction buffer

0.14M D-Sorbitol, 0.8M NaCl, 0.8% CTAB, 0.1% N-Lauroylsarcosine, 0.22M Tris-HCl pH 8.0, and 0.022M EDTA.

4.2 Plant tissue fixative

10% acetic acid, 30% chloroform, and 60% ethanol.

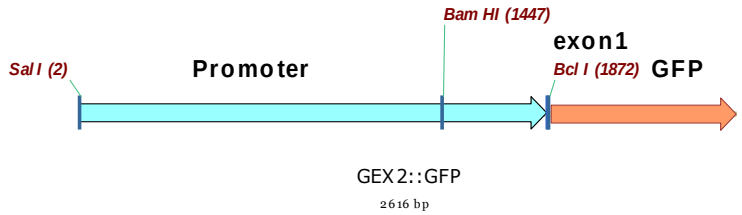
4.3 Plant tissue mounting solution

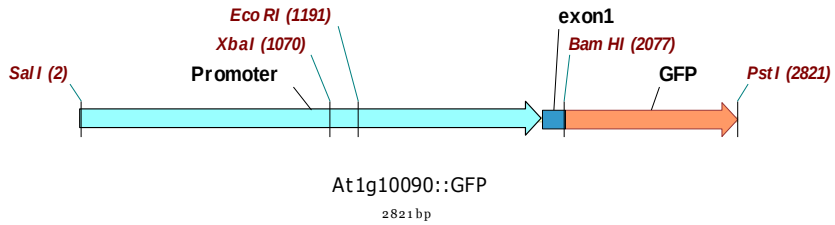
50% glycerol

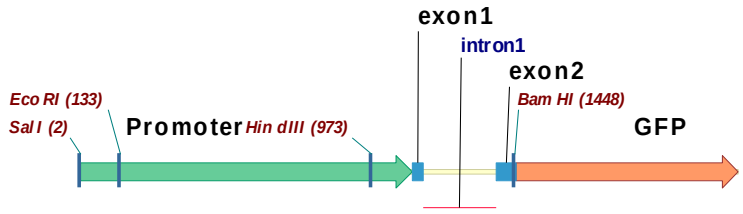
4.4 Adapted Hoyer's solution (Liu C.M. and Meinke D.W. [1998])

24 g chloral hydrate, 4 ml glycerol, 12ml ddH_2O, and dissolve by stirring for 3-5 hrs, or overnight.

Appendix C: GFP construct maps

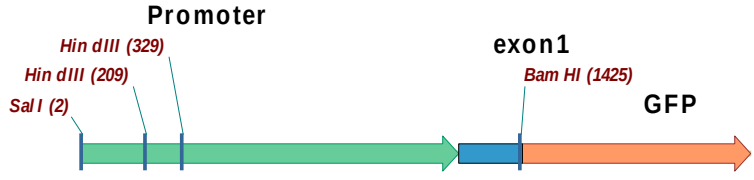






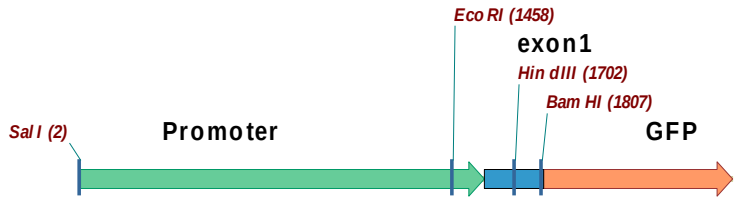
At4g32830::GFP

2192 bp



At3g20190::GFP

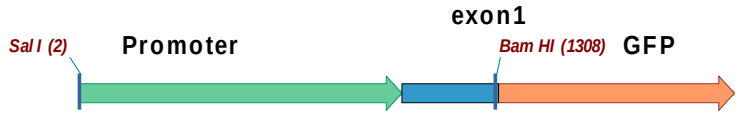
2169 bp



At2g22740::GFP

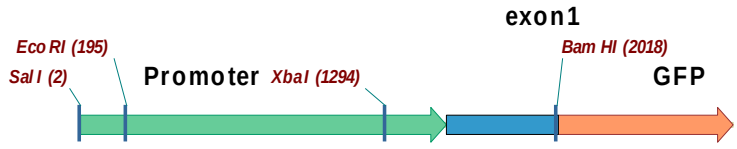
2551 bp

Wd40



At4g11920::GFP

2052 bp



At3g50910::GFP

2762 bp

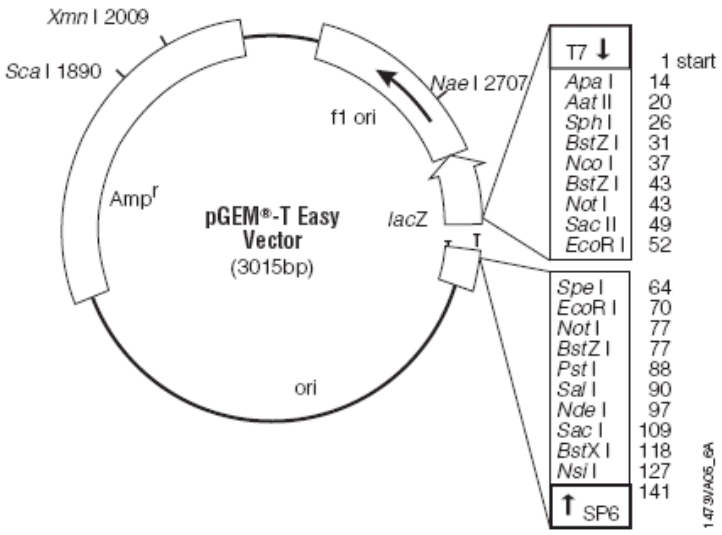
At5g39650::GFP

2652 bp

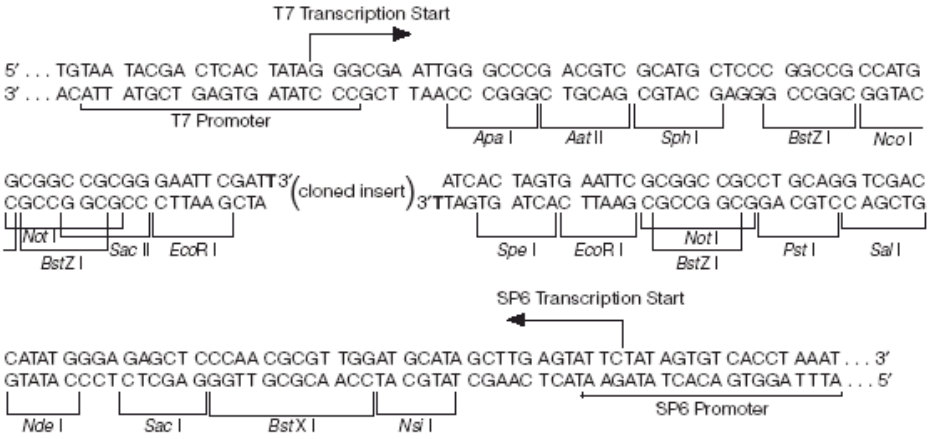
Appendix D: Vector maps

1 pGEM-T Easy

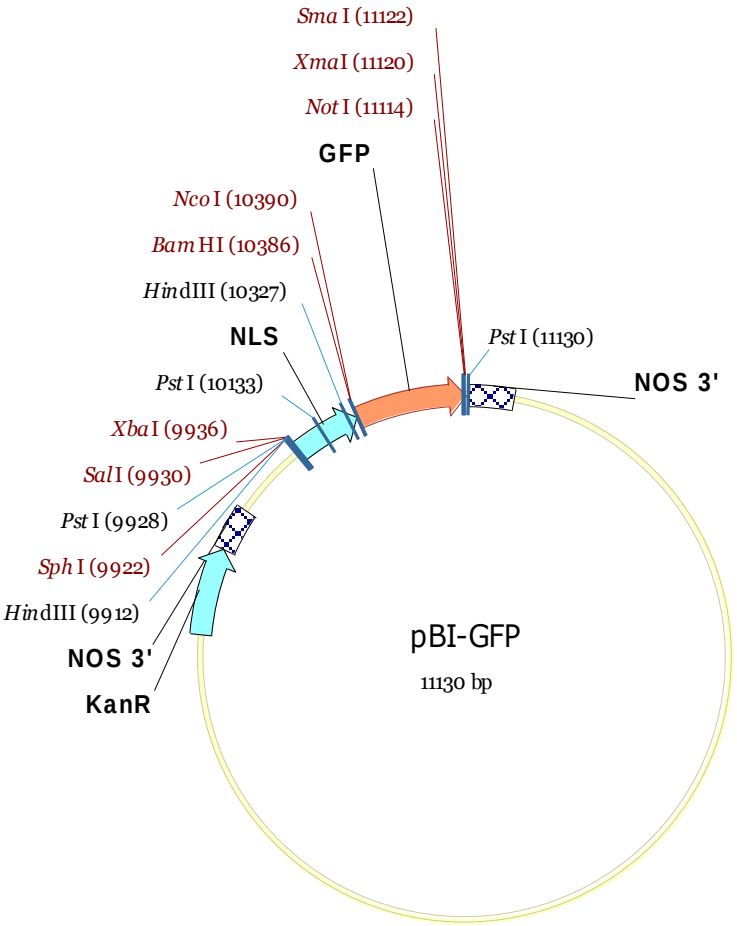
pGEM®-T Easy Vector Map and Sequence Reference Points



pGEM®-T Easy Vector



2 pBI-GFP




```
NNNNNNNNNN NNNNNNNNNN NNNNNNCCGT CAATATTNCC TTCCCTCCCT CAATCGGTTG AATGTCGCCC TTTGTCTTTG GCCCAATACG CAAACCGCCT
NNNNNNNNNN NNNNNNNNNN NNNNNNGGCA GTTATAANGG AAGGGAGGGA GTTAGCCAAC TTACAGCGGG AAACAGAAAC CGGGTTATGC GTTTGGCGGA

CTCCCCGCGC GTTGGCCGAT TCATTAATGC AGCTGGCACG ACAGGTTTCC CGACTGGAAA GCGGGCAGTG AGCGCAACGC AATTAATGTG AGTTAGCTCA
GAGGGGCGCG CAACCGGCTA AGTAATTACG TCGACCGTGC TGTCCAAAGG GCTGACCTTT CGCCCGTCAC TCGCGTTGCG TTAATTACAC TCAATCGAGT

CTCATTAGGC ACCCCAGGCT TTACACTTTA TGCTTCCGGC TCGTATGTTG TGTGGAATTG TGAGCGGATA ACAATTTCAC ACAGGAAACA GCTATGACCA
GAGTAATCCG TGGGGTCCGA AATGTGAAAT ACGAAGGCCG AGCATACAAC ACACCTTAAC ACTCGCCTAT TGTTAAAGTG TGTCCTTTGT CGATACTGGT

          SphI          SalI
         ~~~~~~~       ~~~~~~~
   HindIII        PstI          XbaI
   ~~~~~~        ~~~~~~        ~~~~~~~
TGATTACGCC AAGCTTGCAT GCCTGCAGGT CGACTCTAGA ATGCAGAGCA TTATGGACTC GTCTGCTGTT AATGCGACGG AAGCTACTGA ACAAAATGAT
ACTAATGCGG TTCGAACGTA CGGACGTCCA GCTGAGATCT TACGTCTCGT AATACCTGAG CAGACGACAA TTACGCTGCC TTCGATGACT TGTTTTACTA

GGCAGCAGAC AAGATGTTCT GGAGTTCGAC CTTAACAAAA CTCCTCAGCA GAAACCCTCC AAAAGGAAAA GGAAGTTCAT GCCCAAGGTG GTCGTGGAAG
CCGTCGTCTG TTCTACAAGA CCTCAAGCTG GAATTGTTTT GAGGAGTCGT CTTTGGGAGG TTTTCCTTTT CCTTCAAGTA CGGGTTCCAC CAGCACCTTC

                    PstI
                   ~~~~~~~
GCAAACCTAA AAGAAAGCCA CGCAAACCTG CAGAACTTCC CAAAGTGGTC GTGGAAGGCA AACCTAAAAG GAAGCCACGC AAAGCTGCAA CTCAGGAAAA
CGTTTGGATT TTCTTTCGGT GCGTTTGGAC GTCTTGAAGG GTTTCACCAG CACCTTCCGT TTGGATTTTC CTTCGGTGCG TTTCGACGTT GAGTCCTTTT

AGTGAAATCT AAAGAAACCG GGAGTGCCAA AAAGAAAAAT TTGAAAGAAT CAGCAACTAA AAAGCCAGCC AATGTTGGAG ATATGAGCAA CAAAAGCCCT
TCACTTTAGA TTTCTTTGGC CCTCACGGTT TTTCTTTTTA AACTTTCTTA GTCGTTGATT TTTCGGTCGG TTACAACCTC TATACTCGTT GTTTTCGGGA

                                                                                  NcoI
                                                                                 ~~~~~~~
                   HindIII                                                    BamHI
                   ~~~~~~~                                                   ~~~~~~~
GAAGTCACAC TCAAAAGTTG CAGAAAAGCT TTGAATTTTG ACTTGGAGAA TCCTGGAGAT GCGAGGCAAG GTGACTCTGA GTCTGGATCC ATGGTGAGCA
CTTCAGTGTG AGTTTTCAAC GTCTTTTCGA AACTTAAAAC TGAACCTCTT AGGACCTCTA CGCTCCGTTC CACTGAGACT CAGACCTAGG TACCACTCGT

AGGGCGAGGA GCTGTTCACC GGGGTGGTGC CCATCCTGGT CGAGCTGGAC GGCGACGTGA ACGGCCACAA GTTCAGCGTG TCCGGCGAGG GCGAGGGCGA
TCCCGCTCCT CGACAAGTGG CCCCACCACG GGTAGGACCA GCTCGACCTG CCGCTGCACT TGCCGGTGTT CAAGTCGCAC AGGCCGCTCC CGCTCCCGCT

TGCCACCTAC GGCAAGCTGA CCCTGAAGTT CATCTGCACC ACCGGCAAGC TGCCCGTGCC CTGGCCCACC CTCGTGACCA CCTTCACCTA CGGCGTGCAG
ACGGTGGATG CCGTTCGACT GGGACTTCAA GTAGACGTGG TGGCCGTTCG ACGGGCACGG GACCGGGTGG GAGCACTGGT GGAAGTGGAT GCCGCACGTC

TGCTTCAGCC GCTACCCCGA CCACATGAAG CAGCACGACT TCTTCAAGTC CGCCATGCCC GAAGGCTACG TCCAGGAGCG CACCATCTTC TTCAAGGACG
ACGAAGTCGG CGATGGGGCT GGTGTACTTC GTCGTGCTGA AGAAGTTCAG GCGGTACGGG CTTCCGATGC AGGTCCTCGC GTGGTAGAAG AAGTTCCTGC

ACGGCAACTA CAAGACCCGC GCCGAGGTGA AGTTCGAGGG CGACACCCTG GTGAACCGCA TCGAGCTGAA GGGCATCGAC TTCAAGGAGG ACGGCAACAT
TGCCGTTGAT GTTCTGGGCG CGGCTCCACT TCAAGCTCCC GCTGTGGGAC CACTTGGCGT AGCTCGACTT CCCGTAGCTG AAGTTCCTCC TGCCGTTGTA
```

```
CCTGGGGCAC AAGCTGGAGT ACAACTACAA CAGCCACAAC GTCTATATCA TGGCCGACAA GCAGAAGAAC GGCATCAAGG TGAACTTCAA GATCCGCCAC
GGACCCCGTG TTCGACCTCA TGTTGATGTT GTCGGTGTTG CAGATATAGT ACCGGCTGTT CGTCTTCTTG CCGTAGTTCC ACTTGAAGTT CTAGGCGGTG

AACATCGAGG ACGGCAGCGT GCAGCTCGCC GACCACTACC AGCAGAACAC CCCCATCGGC GACGGCCCCG TGCTGCTGCC CGACAACCAC TACCTGAGCA
TTGTAGCTCC TGCCGTCGCA CGTCGAGCGG CTGGTGATGG TCGTCTTGTG GGGGTAGCCG CTGCCGGGGC ACGACGACGG GCTGTTGGTG ATGGACTCGT

CCCAGTCCGC CCTGAGCAAA GACCCCAACG AGAAGCGCGA TCACATGGTC CTGCTGGAGT TCGTGACCGC CGCCGGGATC ACTCACGGCA TGGACGAGCT
GGGTCAGGCG GGACTCGTTT CTGGGGTTGC TCTTCGCGCT AGTGTACCAG GACGACCTCA AGCACTGGCG GCGGCCCTAG TGAGTGCCGT ACCTGCTCGA

           SmaI
          ~~~~~~~
           XmaI
          ~~~~~~~
      NotI         PstI
    ~~~~~~~~      ~~~~~~
GTACAAGTAA AGCGGCCGCC CGGGCTGCAG
CATGTTCATT TCGCCGGCGG GCCCGACGTC
```