The Plant Journal (2001) **28**(1), 27–39

The *Arabidopsis MALE STERILITY1* (*MS1*) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors

Zoe A. Wilson[1,*], Shaun M. Morroll[2], Janet Dawson[1] Ranjan Swarup[1] and Paddy J. Tighe[3]
[1]*Plant Science Division, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leics., UK,*
[2]*Compton Laboratory, Compton, Newbury, Berkshire, UK, and*
[3]*Division of Immunology, School of Clinical Laboratory Sciences, University of Nottingham, Queen's Medical Centre, Nottingham, UK*

Received 1 March 2001; revised 19 June 2001; accepted 22 June 2001.
[*]For correspondence (fax +44 115 951 6334; e-mail Zoe.Wilson@nottingham.ac.uk).

Summary

We report here the molecular characterisation of the *Arabidopsis MALE STERILITY1* gene, which is a critical sporophytic controlling factor for anther and pollen development. Homozygous *ms1* mutants do not produce viable pollen, but are otherwise phenotypically normal. Degeneration of pollen occurs soon after microspore release from the tetrads, at which time the tapetum also appears abnormally vacuolated. The *MS1* gene is expressed at low levels in anthers from closed buds, with expression in the tapetum at the stage of microspore release. No expression is seen in open flowers. The deduced MS1 protein sequence shows strong homology to the PHD-finger motif found in known transcription factors from humans, yeast and higher plants. Six alleles of *ms1* have been identified; all result in premature termination of the MS1 protein and loss of the PHD-finger motif. *MS1* is likely to play a key role in regulating transcription during specific stages of male gametogenesis and anther development. As such, *MS1* provides a valuable tool for the manipulation of male sterility in higher plants.

Keywords: sporogenesis, male-sterility, Arabidopsis, transcription factor, PHD-finger, anther development.

Introduction

The process of floral development and reproduction in higher plants is a complex developmental strategy involving a diverse range of gene interactions (Goldberg *et al.*, 1993; Gorman and McCormick, 1997). As such male gametogenesis is an area of major biological importance in plant developmental biology, but is also of significant commercial interest for the control of crop fertility. For example, manipulation of pollen development is crucial for F1 hybrid seed production, to reduce labour costs to industry and for the production of low-environmental impact genetically engineered crops. Male sterile (*ms*) mutants have been reported in many species of higher plants as the result of both spontaneous and induced mutations (Kaul, 1998). The isolation of mutants defective at specific stages during the process of male gametogenesis has been used as a tool to investigate the processes of anther and pollen development(Chaudhury, 1993; Dawson *et al.*, 1993b; Goldberg *et al.*, 1993; Peirson *et al.*, 1996; Preuss *et al.*, 1993; Regan and Moffatt, 1990; Ross *et al.*, 1997; Sanders *et al.*, 1999; Taylor *et al.*, 1998). Male sterile mutants have been identified that show a range of different phenotypes, including structural aberrations such as short filaments (Mulligan *et al.*, 1994) and lack of dehiscence (Dawson *et al.*, 1999), through to functional defects associated with gametogenesis and specifically meiosis (Dawson *et al.*, 1993a; Glover *et al.*, 1998; He *et al.*, 1996; Peirson *et al.*, 1996; Ross *et al.*, 1997; Sanders *et al.*, 1999) leading to a lack of pollen. Mutants have been identified by a variety of approaches and characterised based on the stage at which pollen development fails (Chaudhury, 1993; Taylor *et al.*, 1998).

It has been proposed that about 3500 genes may be expressed within the anther Sanders *et al.*, 1999, however, the numbers of genes that are specifically involved in

gametophyte development and result in the loss of anther and pollen development are not known. Equivalent work in maize and tomato indicates that at least 60 genes have been identified that have a specific effect on male fertility (Horner and Palmer, 1995). However, to date very little is known about the molecular process of gametogenesis. A number of genes have been identified based on anther-specific expression patterns (Rubinelli *et al.*, 1998; Scott *et al.*, 1991; Ursin *et al.,* 1989), however, although possible functions have been assigned based on protein homology

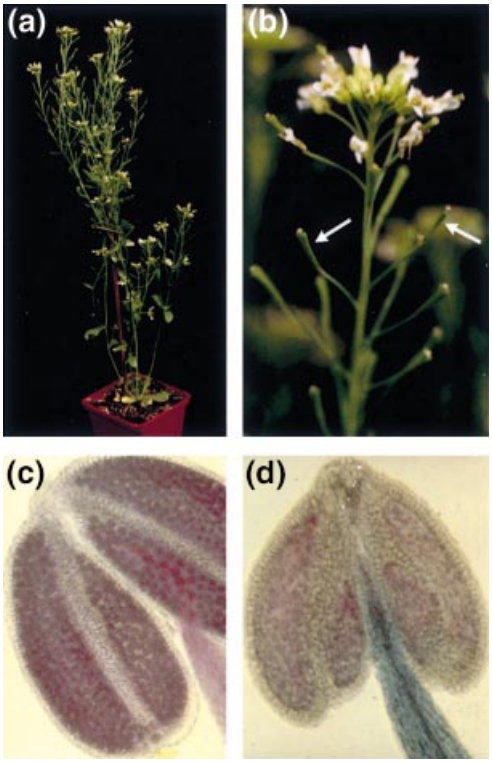
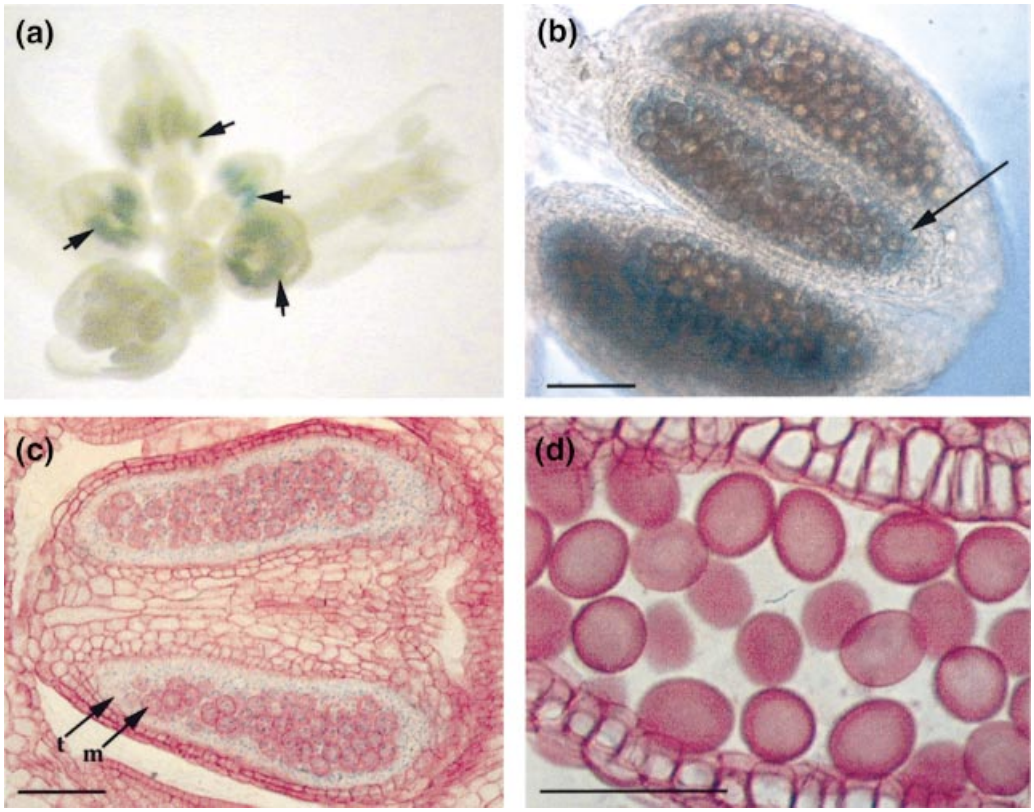


Figure 1. Phenotype of the *ms1* mutant.
(a) *ms1* mutant plant. The plant appears as wild type except that silique development is abnormal since no viable pollen is produced and self-fertilization does not occur.
(b) Detail of *ms1* mutant inflorescence, the *ms1* flowers do not self-fertilise and silique elongation does not occur (arrows).
(c) Ler wt anther, viable pollen is clearly seen stained purple (Alexander, 1969).
(d) *ms1* anther, no viable pollen (Alexander, 1969) is seen.

Figure 7. *MS1*:*GUS* promoter fusions.
(a) Whole mounts of entire Ler inflorescence showing specific GUS staining in closed buds, around the stage of microspore release (arrows).
(b) Anther squashes of Ler buds showing *GUS* expression from the tapetal tissues (arrow).
(c) Sections through Ler anthers showing *GUS* expression from the tapetum (t). GUS staining is also seen within the locule and the developing microspores (m). This may be due to low levels of expression within the gametophytic tissue, or to release of the GUS protein during tapetal degeneration and deposition on the microspores during pollen maturation.
(d) Section through Ler anther after complete tapetal breakdown. No GUS staining is seen at this stage. Bar = 50 microns.



data, it has been difficult to determine a defined role for these proteins in male sporogenesis. Only a few genes that have a specific phenotype associated with loss of male fertility have been characterised.

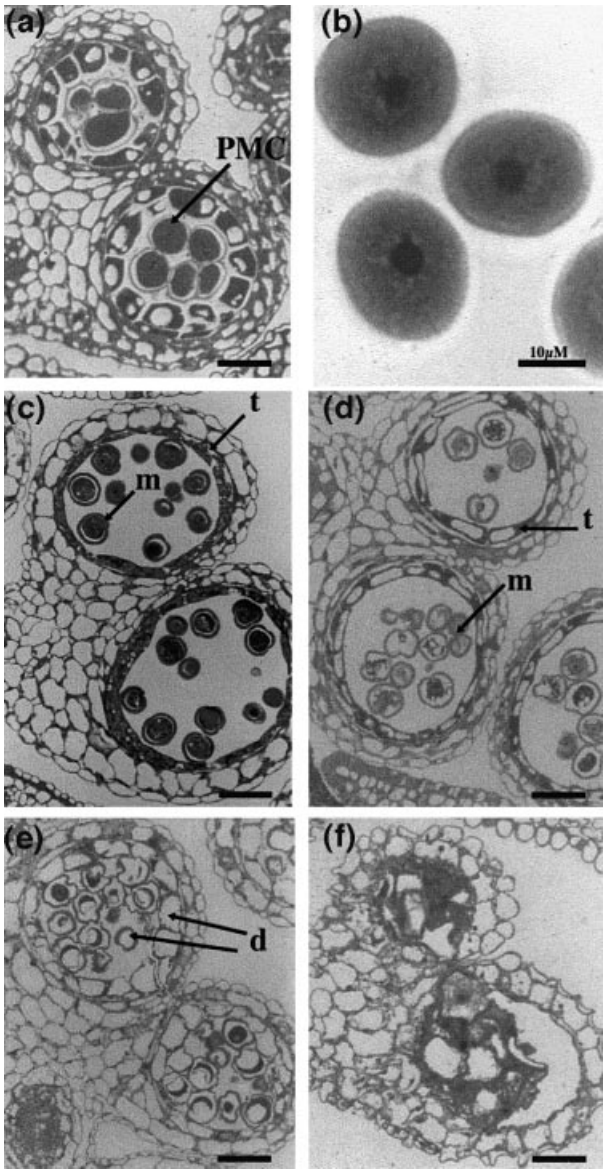


Figure 2. Cross-sections through anthers of *ms1* mutant and Ler (wt) plants.
(a) *ms1* pollen mother cells (PMCs) in meiosis, development is as seen in the wt with PMCs surrounded by a layer of callose.
(b) *ms1* isolated microspores after release from tetrad. At this stage they appear similar to wt (Dawson *et al.*, 1993a,b).
(c) Ler wt anther after microspore (m) release and prior to tapetal degeneration. Tapetal cells (t) have a granular appearance. Pollen mitosis is thought to occur at this point (Regan and Moffatt, 1990).
(d) *ms1* anthers at equivalent stage to (c). In the *ms1* mutant the microspores (m) enlarge and develop a granular, vacuolated cytoplasm and the tapetum (t) becomes abnormally vacuolated.
(e) *ms1* anther at a slightly later stage. The tapetal cells and microspores enlarge further and the cytoplasm degenerates (d).
(f) Later stage *ms1* anther. The *ms1* microspores and the tapetum degenerate to form a mass of undifferentiated cells. Ultimately, the *ms1* anther locule is completely empty. Bar = 40 microns, unless otherwise stated.

A number of genes associated with meiosis have been cloned, these include the *Arabidopsis MS5* (Glover *et al.*, 1998)/*pollenless3* (Sanders *et al.*, 1999) and *ASK1* (SKP1-LIKE1) genes (Yang *et al.*, 1999a), both of which are associated specifically with male meiosis. The *AtDMC1* gene, which is involved in bivalent formation/stabilisation at meiosis in both male and female meiocytes (Couteau *et al.*, 1999; Klimyuk and Jones, 1997), and the MADS box transcription factor *sporocyteless*/*NOZZLE* gene (Schiefthaler *et al.*, 1999), which is also expressed during both micro and megasporogenesis (Yang *et al.*, 1999b) have also been characterised.

The *MS2* gene (Aarts *et al.*, 1993), which has homology to a fatty acyl reductase involved in pollen wall formation, is thought to function at the stage of microspore release from the tetrads. Another stage of pollen development that has been characterised is that of pollen wall formation, for example the *pop1* mutation (Preuss *et al.*, 1993) which disrupts pollen–pistil interactions by eliminating extracellular pollen coat (tryphine) formation.

We report here the cloning of an essential gene required for *Arabidopsis* male gametogenesis, the *MALE STERILITY1* (*MS1*) gene. This gene is sporophytic in nature, appears to function around the stage of microspore release and is critical for viable pollen development. The MS1 protein has homology to the PHD-finger class of transcription regulators. As such, this is the first report of a transcriptional regulator that specifically affects pollen development without influencing other aspects of sporogenesis and floral development. *MS1* will therefore be a critical tool to provide further insights into pollen developmental processes.

Results

Phenotype of the ms1 *mutant*

The *ms1–1* mutant was generated by EMS mutagenesis of *Arabidopsis thaliana* Landsberg *erecta* (L*er*) seed (Van der Veen and Wirtz, 1968). It is a recessive mutation and appears as wild type (wt) in the heterozygous state. Homozygous *ms1* mutants fail to produce viable pollen (Figure 1), with pollen degeneration occurring soon after microspore release from the tetrads. Young buds show characteristic agglutination of immature pollen, which appears viable using FDA staining (Dawson *et al.*, 1993a) but cannot be induced to form pollen tubes (data not shown); gentle pressure on the buds at this stage results in the degenerating pollen being exuded in as a 'sausage-shaped' mass. Complete degeneration of the immature pollen then follows (Figure 2e,f), so that no pollen viable, or inviable, is present in the mature locule (Figure 1d). The

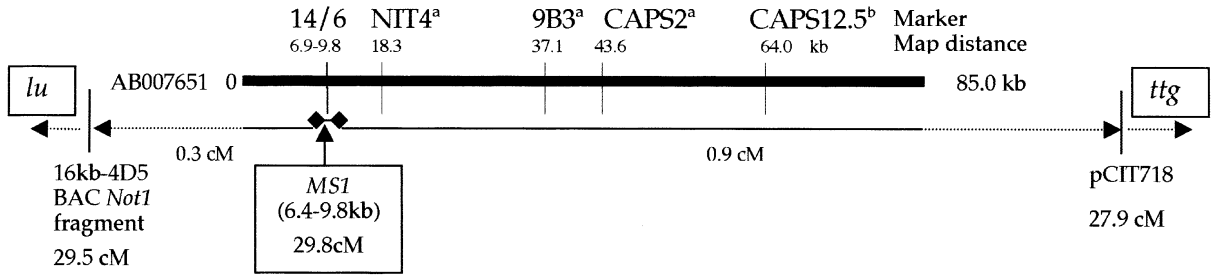


Figure 3. Molecular, genetic map of the *ms1* region.
The map corresponds to the sequenced AB007651 (MWD9) clone. The *MS1* gene lies in the region 6.4–9.8kb region of this clone. Molecular markers were mapped using Ler/Sn(5)-1 recombinant lines (Thorlby *et al.*, 1997) [a]markers cosegregate with *ms1*, [b]marker maps 0.14 cM from *ms1* towards *ttg*.

Table 1. Primers designed to amplify across the AB007651 (MWD9) clone

Primer name	Forward primer sequence 5′-3′	Reverse primer sequence 5′-3′	Region of AB007651 amplified (bp)	Size PCR product (bp)
1	TTGGTGACTTTATTGGCAGCTGAAGCG	CTCGACGTTCTCCGCCGTGG	313–5213	4900
2	CAACGGGCCATCCCTGCGAA	CGCAGCCGGAGACTCGTCGAA	38372–45067	6696
3	CCACGGCGCCTGTGTTTCCC	TGAAAAAGCGTGTCATGGTGATGCAA	44500–53575	9076
4	TGGTTGTATGGACTTGGGGTGATGGA	TCAGTGGCCAGTGGGCAGAACAA	53001–62064	9064
5	CCCACTGGCCACTGAGCTGGAGG	TCGCCCGGAACAGGACGACTTT	62050–69353	7304
6	TGGTGACTTTATTGGCAGCTGAAGCG	TGTCCAACCACGCAAACACACACG	314–9789	9475
7	TCTTTCGCCTTTTGGTGCACGACC	CGCTTCAAACGTGCAATTTTCGCC	15838–24194	8356
8	TTCTCGCCGACACGCAGCTTTTCT	TCGGAAGGCCGCTTGTGAACTTTG	24023–35511	11488
9	GTGAGGACAATCAGGGGCAGTGGC	AGCAATCTAAGCGCCTTCGCCGAG	35140–43786	8646
10	CTCCAAAGATGACCGTCGCGGAAA	CCCAACAGGACCAGCAAGCAGAGG	65962–76300	10338
11	GCAAAGCTCTTGGGAGGCCCTGAG	AGAAAATCCGTTTGCGGAAGGCGA	73316–82790	9474
13	TGCTCCATTTCCGGTTTACCA	GTCGCTCTCCTCTTCCCCTGA	3348–6451	2251
14/6R	GCCATTTCATCGACCCTTGTG	CGTGTGTGTTTGCGTGGTTGGACA	6920–9789	2870
15	ATCACCCGTTGCAAAGAATCAA	TATCATTTGTTGGCCGTGCGT	11153–12583	1451
16	TGACTCAAAAGGCAAGTGAAGCAG	GCCGATTTCTTACGCGCTCTT	13684–16482	2799
17	GCGCTAGCGGTGGTGATTACA	TTCACTTGCTCGTCGAAACCG	22656–24213	1558
18	CCCTAGCCGTTGGATGTTACAGA	TCACGGTTGGGAACAGAGACA	27034–29444	2411
19/8R	GGCCGTCTCACTCTCTCGCTA	CAAAGTTCACAAGCGGCCTTCCGA	33417–35511	2095
20	GCCTCAGTGTTCAAAGAATATGGG	GTATCGGTCTCGTTCGTGCG	37624–39540	1917
21	GCGCCTAGCAAAAACTGACGA	TGTGTTGAGGATGAGAAGGATGG	39375–40905	1531
22	CATCTCGTGGAAAACAGGATCG	CCAGAATGCTTAGCACTTGTAGTTGG	40526–42779	2254

early stages of callose production, pollen meiosis and tetrad formation occur normally, followed by callose dissolution and microspore release (Figure 2a,b). Initiation of exine synthesis appears normal, however, immediately after release from the callose wall, the microspores develop an unusually granular, vacuolated cytoplasm (Figure 2d) compared with Ler wt (Figure 2c). At this stage the tapetum becomes abnormally vacuolated and cytoplasmic degeneration begins. The microspores appear to adhere to each other, collapse and the cytoplasm degenerates (Figure 2e,f). In mature flowers the microspores and tapetal cells have undergone complete degeneration resulting in an empty locule. In all other aspects of plant development the *ms1* mutant is completely normal (Figure 1); female fertility is unaffected and the mutation can be maintained by crossing using wt or heterozygous pollen.

We have isolated a series of five alleles of *ms1*, using EMS (*ms1–2, ms1–3* and *ms1–4*), X-rays (*ms1–5*) (Dawson *et al.* 1993a) and gamma-rays (*ms1–6*) (Kalantidis *et al.*, 1994). All allelic lines exhibit the same phenotype, of complete male sterility regardless of growth under a range of different environmental growth conditions (data not shown).

Mapping and localisation of the MS1 *gene*

We have previously mapped the *ms1* gene to 29.8 ± 0.8 cM on the top-arm of chromosome 5 and identified a marker, NIT4 (Bartel and Fink, 1994), which co-segregates with *ms1* (Thorlby *et al.*, 1997). This region corresponds to the sequenced clone AB007651 (MWD9) (Nakamura *et al.*,. 1997). Two PCR primers, CAPS12.5 and

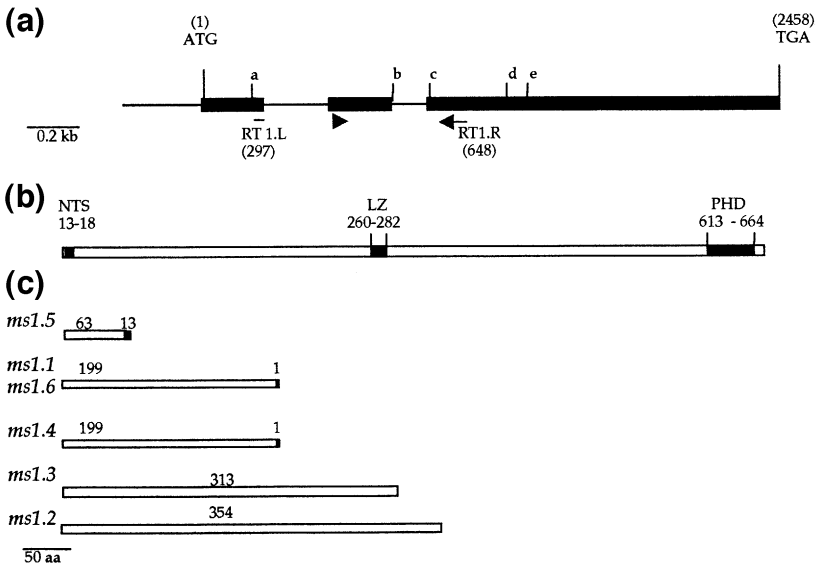


Figure 4. Structure of the *MS1* gene and protein.
(a) Structure of *MS1* gene. Dark boxes represent exons, open boxes represent untranslated 5′ and 3′ sequences; *ms1* mutation sites a: *ms1–5* (T^{215} deleted), b:*ms1–1* and *ms1–6* (G^{882} to A), c:*ms1–4* (G^{1036} to A), d: *ms1–3* (C^{1380} to T), e: *ms1–2* (C^{1500} to T). RT1.L and RT1.R correspond to the RT–PCR primer positions.
(b) MS1 peptide sequence, NTS = nuclear targeting sequence, LZ = leucine zipper, PHD = plant homeodomain motif. The numbers show the amino acid numbers of the region in the protein.
(c) The truncated proteins produced by the *ms1* mutant alleles, unfilled boxes represents the equivalent amino acid sequence to wt, dark boxes show novel amino acids; values represent the relative numbers of each.

CAPS2, were designed using the AB007651 sequence to generate polymorphisms to confirm the location of *ms1.* Recombinants that had been previously generated between the Ler line carrying the mutations *ms1* combined with *lu*, or *ttg*, and the alternative ecotype Sn(5)-1 (Thorlby *et al.,* 1997) were used for fine-mapping. Amplification and digestion with a range of restriction enzymes was carried out on DNA from the parental, Ler and Sn(5)-1 lines, and recombinants on either side of *ms1* (Thorlby *et al*, 1997). Thirteen recombinants on the *ttg* side of *ms1* were analysed. These had breakpoints on average every 0.07 cM between the pCIT718 marker, which maps 0.9 cM from *ms1*, and *ms1.* Three *lu* recombinants with breakpoints on average every 0.1 cM between an RFLP marker derived from the 16kb *Not1* fragment from the BAC clone 4D5 (16kb-4D5 marker maps 0.3 cM from *ms1* towards *lu*) and *ms1* (Thorby *et al.*, 1997) were also used.

A *Xba1* polymorphism, corresponding to approximately 64 kb in the AB007651 clone, was identified with the CAPS12.5 primers and was found to map 0.14 cM from *ms1*, towards *ttg* (Figure 3). A *Sal1* polymorphism was identified in the CAPS2 PCR product, corresponding to 39.4kb in the AB007651 clone, which co-segregated with *ms1* (Figure 3).

Three markers, NIT4, 9B3 (Thorby *et al.*, 1997) and CAPS2, that co-segregate with *ms1* have been identified and these markers map to 18.3, 37.1 and 43.6 kb, respectively, on the AB007651 clone (Figure 3). Additionally a marker 16kb-BAC 4D5 was found to map 0.3 cM away from *ms1* towards *lu.* This marker was generated from the 4D5 BAC that overlaps with the AB007651 clone. Therefore, it was hypothesised that *ms1* would lie in the first part of the AB007651 clone. The CAPS12.5 marker maps 0.14 cM from *ms1,* which equates to position 64 kb on the AB007651 clone, and delimits the *ttg* side of the mapped region. The *MS1* gene was therefore predicted to lie between 0 and 64kb on the AB007651 clone.

The AB007651 clone was analysed for ORFs (GeneFinder; ACeDB) in this region that might correspond to *MS1 and* 27 putative ORFs were identified. No prediction of which ORF corresponded to *MS1* could be made based on DNA sequence or protein homologies. Primers (Table 1) were therefore designed to amplify, by long-PCR across the AB007651 clone and by targeting specific ORFs within the clone, using the Ler wt and allelic *ms1* mutants. Given that two of the allelic lines were generated by X-ray and gamma irradiation, it was expected that these would contain deletions and that polymorphisms in the PCR products associated with the *ms1* gene would be detected.

Polymorphisms associated with the *ms1* mutation were not seen with any of the PCR products, except for the 2870 bp product generated with primers 14/6R. This corresponded to the 6920–9789 bp region of AB007651. A *Rsa1* restriction site at position 883 was lost in the 14/6R PCR product from the *ms1–1* and *ms1–6* mutants when compared with Ler wt. This region was sequenced and shown to correspond to a G to A transition, which resulted in the loss of the *Rsa1* restriction site (Figure 4a).

This mutation is associated with a splice site in a potential ORF. To confirm that this was the *ms1* gene, the region corresponding to this ORF was amplified and directly sequenced from Ler and all of the allelic lines. Sequencing revealed aberrations in this gene in all of the *ms1* alleles, resulting in splice defects or truncated polypeptides (Figures 4a,c and 5).

RT–PCR

Three alleles, *ms1–1, ms1–6* and *ms1–4* showed mutations associated with the splicing of the second intron. RT–PCR was conducted to confirm this. The forward RT–PCR primer was designed to span the first splice site and the reverse to lie in the third exon (Figure 4a). The Ler wild type produced the expected RT–PCR product of 351 bp, whilst the splice mutants *ms1–1 and ms1–6* gave products of 506 bp (Figure 6a) due to the lack of splicing caused by the G to A transition at the splice donor site. Direct fluorescent sequencing of all the allelic and Ler RT–PCR products showed the expected lack of splicing in the *ms1–1* and *ms1–6* mutants. No difference in RT–PCR product size could be detected by electrophoresis for the *ms1–4* mutant, however, sequencing of this RT–PCR product showed a 1-bp deletion from the spliced message, resulting in a frameshift mutation and a premature truncation of the MS1 protein (Figure 4c). This corresponds to the predicted mutation, observed from genomic sequencing, of a G to A transition at the splice acceptor site (Figure 4a).

```
-300  aaagttgttaatgttacaagtcatgcacatatcaagacctaacatacaatccaatggtaa
-240  caattgttaattcacagtttaattaacctttctctttctctcactcacttttccttttca
-180  tcaccacaaagcactaattgtgtgaccatcatcttaatgactactatatctctaaaccct
-120  taacacacttcctcaacACTTTTCGGTCTCTTCTTATAACAAACCCCATCATAATATATA
-60   TCCAAATCTATCATATTCCTATTCCCTCTCTTCTCAAGATCAAACCAAATTTCTGATTCG

1     ATGGCGAATCTGATTCGAACAGACCATCAGCAACATATTCCAAAGAAGAGGAAGAGAGGG
1     M  A  N  L  I  R  T  D  H  Q  Q  H  I  P  K  K  R  K  R  G

61    GAAAGTAGAGTTTTTAGGCTGAAGACGTTCGGAGAGTCTGGACATCCAGCTGAGATGAAC
21    E  S  R  V  F  R  L  K  T  F  G  E  S  G  H  P  A  E  M  N

121   GAGTTGTCTTTTCGAGATAACCTCGCTAAACTACTTGAGTTTGGTCACTTTGAGAGCTCC
41    E  L  S  F  R  D  N  L  A  K  L  L  E  F  G  H  F  E  S  S

181   GGTCTAATGGGAAGTTGGTCTTTTCAGCTCGAGATTCAACGAAATCCAAATCCTCTCTAT
61    G  L  M  G  S  W  S  F  Q  L  E  I  Q  R  N  P  N  P  L  Y

241   GTTCTTCTCTTTGTCGTAGAAGAGCCCATCGAAGCCTCTCTCAATCTCCGTTGCAACCAT
81    V  L  L  F  V  V  E  E  P  I  E  A  S  L  N  L  R  C  N  H

301   TGCCAATATGTTGgtgacgtcctctctctcttttcccttctctctctctgtatacaatttt
101   C  Q  Y  V  G

361   ttcttttgcgtatctatctatatatgcatgtgaatacattcgtaagctgaaaatttagta
421   agagttgacctagctagctcaacatgtaagcatttttcttgcatatatatctatatacaa
481   gcttacacgtacttatgaacaaaaatagcaataataatttaatatatggttaaaagtgat
541   gcatttatgtaaagatcaaaaatatatattatacaatggtttatacggaaccttgcagGT
105

601   TGGGGAAATCAAATGATATGCAACAAGAAGTACCATTTCGTGATCCCCTCAAAGGAAACA
106   W  G  N  Q  M  I  C  N  K  K  Y  H  F  V  I  P  S  K  E  T

661   ATGGCGGCTTTTTTAAAACTGGAAGGTGGAGGCTACGCTTTTCCCGAAAAGGAAAGTTTC
126   M  A  A  F  L  K  L  E  G  G  G  Y  A  F  P  E  K  E  S  F

721   TCCCATCTTGTGGAGCTTCAAGGCCATGTCCTTCACGGCTTCTTTCACTCCAACGGATTT
146   S  H  L  V  E  L  Q  G  H  V  L  H  G  F  F  H  S  N  G  F

781   GGTCACTTGCTCTCTCTCAACGGCATTGAAACCGGCTCCGACTTAACCGGTCATCAGGTC
166   G  H  L  L  S  L  N  G  I  E  T  G  S  D  L  T  G  H  Q  V

841   ATGGATTTGTGGGACCGGCTCTGCACCGGTTTAAAGGCCAGgtactaattagactttata
186   M  D  L  W  D  R  L  C  T  G  L  K  A  R

901   tatttgtatatcatataaaccattcaacacaataaccattcaacacaattctgttagttt
961   tcatatcataatggattattcctagctattaatttggatttacacataatgagtcatttt
1021  gtttttgttcataaagGAAAATAGGGTTGAATGACGCGTCGCACAAAAAAGGAATGGAGT
199                   K  I  G  L  N  D  A  S  H  K  K  G  M  E  L

1081  TGAGGCTGCTGCATGGGGTAGCAAAAGGAGAGCCCTGGTTCGGTCGTTGGGGCTACCGGT
215     R  L  L  H  G  V  A  K  G  E  P  W  F  G  R  W  G  Y  R  F

1141  TCGGGTCAGGGACATACGGAGTGACTCAAAAGATTTACGAGAAGGCACTTGAGTCGGTCC
235     G  S  G  T  Y  G  V  T  Q  K  I  Y  E  K  A  L  E  S  V  R
```

Figure 5. Sequence of the Landsberg *erecta* (Ler) *MS1* locus.The cDNA sequence is shown in capitals; untranscribed regions and the introns are in lowercase. Nucleotide numbering starts from the start of translation. The protein sequence is given underneath the nucleotide sequence. The putative CAAT and TATA boxes are underlined. The asterisk denotes the termination codon.

```
1201  GTAACATACCCTTGTGCTTGCTTAACCATCACCTAACCAGCCTTAACCGAGAAACTCCAA
255     N  I  P  L  C  L  L  N  H  H  L  T  S  L  N  R  E  T  P  I

1261  TCCTCTTGTCAAAGTACCAAAGTTTATCCACCGAGCCATTGATCACTCTCAGTGACCTCT
275     L  L  S  K  Y  Q  S  L  S  T  E  P  L  I  T  L  S  D  L  F

1321  TCAGGTTCATGCTTCATCTCCATTCACGTCTTCCAAGAGATAACTACATGAGTAACTCCC
295     R  F  M  L  H  L  H  S  R  L  P  R  D  N  Y  M  S  N  S  R

1381  GAAACCAAATCATCTCCATTGATAGTACCAACTGCAGATGGTCTCAAAAACGGATCCAAA
315     N  Q  I  I  S  I  D  S  T  N  C  R  W  S  Q  K  R  I  Q  M

1441  TGGCTATCAAAGTGGTCATAGAGTCACTGAAAAGAGTCGAATACCGATGGATATCGAGAC
335     A  I  K  V  V  I  E  S  L  K  R  V  E  Y  R  W  I  S  R  Q

1501  AAGAAGTGAGGGATGCAGCTAGAAATTACATTGGGGACACTGGTTTGCTTGATTTTGTGT
355     E  V  R  D  A  A  R  N  Y  I  G  D  T  G  L  L  D  F  V  L

1561  TGAAGTCGCTTGGGAACCAGGTGGTTGGAAACTATTTGGTCCGACGTAGTCTAAATCCGG
375     K  S  L  G  N  Q  V  V  G  N  Y  L  V  R  R  S  L  N  P  V

1621  TGAAGAAAGTGCTAGAGTATTCCTTGGAGGATATATCAAATTTGTTACCAAGTGGTAACA
395     K  K  V  L  E  Y  S  L  E  D  I  S  N  L  L  P  S  G  N  N

1681  ATGAACTCATAACCCTTCAAAACCAAAACTCAATGGGGAAGATGGCGACAAACGGTCACA
415     E  L  I  T  L  Q  N  Q  N  S  M  G  K  M  A  T  N  G  H  N

1741  ATAAGATCACAAGAGGTCAAGTTATGAAAGACATGTTTTATTTTTACAAACACATTCTCA
435     K  I  T  R  G  Q  V  M  K  D  M  F  Y  F  Y  K  H  I  L  M

1801  TGGACTACAAGGGAGTGTTAGGCCCCATAGGAGGTATATTGAACCAAATCGGAATGGCTT
455     D  Y  K  G  V  L  G  P  I  G  G  I  L  N  Q  I  G  M  A  S

1861  CAAGAGCAATCCTCGACGCTAAGTACTTCATCAAAGAGTATCACTACATTAGAGATACAT
475     R  A  I  L  D  A  K  Y  F  I  K  E  Y  H  Y  I  R  D  T  S

1921  CGGCGAAAACGTTACACTTAGATCGAGGGGAAGAATTAGGAATATTCTGCACGATCGCGT
495     A  K  T  L  H  L  D  R  G  E  E  L  G  I  F  C  T  I  A  W

1981  GGAAATGTCATCATCATAACAACGAGATAAAAGTTCCTCCACAAGAATGCATTGTAGTGA
515     K  C  H  H  H  N  N  E  I  K  V  P  P  Q  E  C  I  V  V  K

2041  AGAAAGATGCAACATTGAGTGAAGTGTACGGAGAGGCAGAAAGAGTGTTTAGAGATATCT
535     K  D  A  T  L  S  E  V  Y  G  E  A  E  R  V  F  R  D  I  Y

2101  ATTGGGAACTAAGAGACGTCGTGGTGGAGTCAGTGGTGGGTGGTCAAATAGAGATCACAA
555     W  E  L  R  D  V  V  V  E  S  V  V  G  G  Q  I  E  I  T  R

2161  GGGTCGATGAAATGGCCTTGAATGGGAATAAGGGATTGGTGTTAGAAGGGAACGTAGGAA
575     V  D  E  M  A  L  N  G  N  K  G  L  V  L  E  G  N  V  G  M

2221  TGATGATGAACATTGAAGTGACGAAATGTTATGAAGATGATGATAAAAAGAAGGATAAGA
595     M  M  N  I  E  V  T  K  C  Y  E  D  D  D  K  K  K  D  K  R

2281  GAATAGAGTGTGAGTGTGGAGCAACGGAAGAAGATGGAGAGAGAATGGTGTGTTGTGATA
615     I  E  C  E  C  G  A  T  E  E  D  G  E  R  M  V  C  C  D  I

2341  TTTGTGAAGTATGGCAACACACAAGGTGTGTTGGTGTTCAACACAATGAGGAAGTGCCTC
635     C  E  V  W  Q  H  T  R  C  V  G  V  Q  H  N  E  E  V  P  R

2401  GCATTTTTCTTTGTCAAAGTTGTGATCAACATCTTATTCCTCTCTCTTTTTTACCCTAAA
655     I  F  L  C  Q  S  C  D  Q  H  L  I  P  L  S  F  L  P  *

2461  TCATTCATGTTATATTAAACCTTCTTTTTTTCTATCATTACGTCATAATACATAAATTAT
2521  ACTCCTCGATCAAcgtgaattctaatgatccctcctgcgtctatatcaccaacatgttga
2581  atcacaaagtcatatccttgcctcctctggagcctgactgttctaataatatcattatca
2641  acatgtcgctctcctcttcccctgatgactcatgtgtcgtcatctactccaaccgtgatt
2701  ccaagttttacttgacaacaaccttaaggcttgtctccaagaaaaggcagttaaatttta
2761  acagaaaataattcacaacacaaggcggagaaagaagtgtgtagtgttttaaataattta
2821  ggct
```

Since six independent *ms1* lines all show mutations in this gene and differences in RT–PCR products were seen in the putative splice mutants, we therefore conclude that this ORF corresponds to the *MS1* gene. The *MS1* gene lies on the reverse strand of clone AB07561 (Figures 3 and 5) and was not previously identified from sequence annotation.

RACE-PCR

RACE-PCR was utilised to identify the transcription start site of the *MS1* cDNA. This was shown to lie 103 bp upstream of the translation start (Figure 5). Sequencing of the Ler wt *MS1* cDNA supported the computer predictions of a 2019-bp translated message, containing three exons with a 73-bp untranslated 3′ region prior to the polyadenylation tail (Figure 5).

MS1 *expression patterns*

Wild type plants were analysed for the expression patterns of the *MS1* gene. No signal could be detected by northern hybridisations (data not shown). RT–PCR using 45 cycles showed reproducible *MS1* products with low levels of expression seen in closed flower buds, but with no expression seen in open flowers (Figure 6b). No expression was seen in leaves or stem tissues. Control RT–PCR reactions using *Arabidopsis* actin-2 specific primers showed similar levels of high expression in cDNA prepared from leaves, stem, buds and open flowers (Figure 6c).

MS1:GUS promoter fusions showed similar results to those seen from the RT–PCR expression studies. Levels of reporter gene expression were low as expected, given the low levels of expression as detected by RT–PCR. The reporter gene expression patterns reflected the phenotypic observations of the mutant plants, with GUS staining seen around the stages of microspore release from the tetrads (Figure 7a–c). Expression was not seen in younger anthers that did not show clearly defined microspores, or in anthers once the tapetal tissue had degenerated (Figure 7a,d). Maximal expression was seen from the tapetal tissues (Figure 7b,c) although GUS staining was also apparent within the locule and the pollen grains (Figure 7c). GUS staining was evident in the tapetal tissues at a late stage where tapetal cell walls were not clearly defined due to the onset of tapetal breakdown. No GUS staining was see in older anthers where complete tapetal degeneration had occurred (Figure 7d). GUS staining was not seen in vegetative tissues.

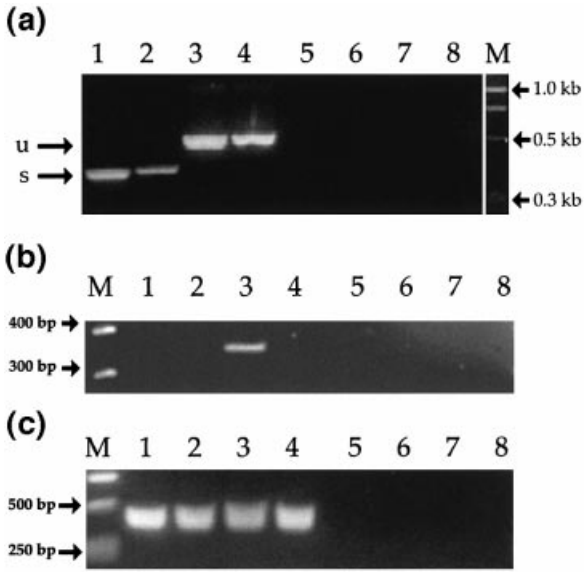


Figure 6. RT–PCR analysis of *MS1* gene expression.
(a) *MS1* RT–PCRs of wt and *ms1* splice mutants. Lanes 1–4 floral cDNA and 5–8 floral RNA controls; 1 and 5 = Ler, 2 and 6 = *ms1–2,* 3 and 7 = *ms1–1,* 4 and 8 = *ms1–6.* u = unspliced *ms1* transcript (506 bp); s = spliced *MS1* transcript (351 bp).
(b) RT–PCRs showing tissue specific expression of *MS1* in Ler wt tissues. Expression is only seen in cDNA prepared from closed buds. Lanes 1 = Leaf cDNA, 2 = Stem cDNA, 3 = Bud cDNA, 4 = Open flower cDNA, 5 = Leaf RNA control, 6 = Stem RNA control, 7 = Bud RNA control, 8 = Open flower RNA control.
(c) Control actin-2 RT–PCRs showing expression in all Ler wt tissues. Lanes 1 = Leaf cDNA, 2 = Stem cDNA, 3 = Bud cDNA, 4 = Open flower cDNA, 5 = Leaf RNA control, 6 = Stem RNA control, 7 = Bud RNA control, 8 = Open flower RNA control. M = marker.

MS1 *sequence homology*

The *MS1* genomic sequence shows no direct homology to any previously identified genes involved in floral development or male sporogenesis. However, limited homology was detected to an *Arabidopsis* hypothetical protein (AC006069; TREMBL:Q9ZUA9) from chromosome 2 (Lin *et al.*, 1999), which has a PHD-finger motif. Homology (29%; Expect value of 4e-84) is seen throughout the protein and extends beyond the PHD-finger motif. The Q9ZUA9 protein has been identified as part of the genomic sequencing effort and has to date no defined expression pattern or function.

The MS1 protein comprises of 672 amino acids with a putative nuclear targeting sequence (Figure 4b) at the N terminal region. Motif analysis of the deduced MS1 protein shows significant homology to the PHD-finger motif (Asaland *et al.*, 1995; Figure 8). This relatively rarely reported motif is strongly conserved in a number of homeodomain proteins. It is cysteine-rich with regular spacing between cysteine and histidine residues, implying a possible metal binding function. PHD-finger motifs (C_4HC_3 zinc-finger like motifs) have been reported in a diverse range of organisms ranging from humans

(HRX), yeast, *Drosophila* (trithorax) and *C. elegans*, to plants including *Arabidopsis*, alfalfa and maize. The homologous proteins have diverse functions, but all are involved in transcriptional regulation. This motif is thought to be specifically involved in chromatin-mediated gene regulation (Asaland *et al.*, 1995).

Discussion

Mapping and localisation of the MS1 *gene*

The *ms1* mutation was mapped using Sn(5)-1/Ler recombinants with breakpoints on either side of *ms1.* The *MS1* gene was localised to the 0–64 kb region of the sequenced clone AB007651 (MWD9). PCR analyses, combined with restriction digests, were conducted on Ler wt and the *ms1* allelic lines, to identify polymorphisms associated with the mutations. Two of the mutants, *ms1–5* and *ms1–6*, were generated by irradiation using X-rays and gamma-rays, respectively. It was therefore expected that these mutants would be deletions, or contain significant chromosomal rearrangements and that size, or restriction site, polymorphisms would be readily seen.

Deletions were detected for the *ms1–6* and *ms1–1* mutants, which lost a *Rsa1* restriction site in the PCR product generated by primers 14/6R. The product from these primers corresponded to an ORF in the 6920–9789 bp region of the reverse strand of AB007651. Analysis of the remaining five *ms1* alleles revealed aberrations involving single base pairs associated with this ORF in all of the mutants. One, *ms1–5* contained a single base deletion resulting in a premature termination of the MS1 polypeptide. The other mutations were transitions; *ms1–1* and *ms1–6* were associated with the splice donor site of the second intron; *ms1–4* with the splice acceptor site of the second intron, and *ms1–3 and ms1–2* caused premature termination of the polypeptide in the final third exon. RT–PCR and sequencing of the three splice mutants (*ms1–1, ms1–6* and *ms1–4*) confirmed the expected loss of splicing in these mutants. In all six of the allelic mutants only single base changes were seen, one which was a single base deletion, the other five single base transitions.

Given that six independent *ms1* lines all show mutations in this gene and differences in RT–PCR products were seen in the putative splice mutants, we conclude that this ORF corresponds to the *MS1* gene.

Phenotypic analysis and expression pattern of MS1

The *ms1* mutation results in complete male sterility and the lack of viable pollen under a range of environmental conditions. In all other aspects of development the plant appears as normal Ler wt. Female fertility is unaffected and the mutant can be rescued using viable pollen from an alternative male parent. The *ms1* mutation is a recessive, Mendelian inherited gene, which does not influence fertility in the heterozygous state.

The *MS1* gene is expressed at low levels with a highly localised temporal and spatial expression pattern, emanating from the tapetal tissue in closed buds around the stage of microspore release. No expression is seen in very young buds, prior to the stage of clearly identifiable microspores, in older anthers once the tapetum has degenerated, or in open flowers. No expression was seen in the leaves or stem tissue by RT–PCR, or in the *MS1* promoter:*GUS* fusions. The expression levels of the *MS1* mRNA appear to be extremely low, as well as transient, since expression could not be detected by northern hybridisations and 45 cycles of RT–PCR were needed to detect the *MS1* message.

The *MS1* gene is active during the late stages of microspore development. In the *ms1* mutants meiosis and microspore release from the tetrad proceed normally. However, soon after this the tapetum and microspores become highly vacuolated and the developing microspores are extremely prone to lysis. Around this stage, when the microspores become clearly defined in the anthers *GUS* expression could be detected in *MS1*:*GUS* fusion transgenic plants (Figure 7). GUS staining was maximal from the tapetal tissues, however, some staining was also apparent around the microspores. Thin sections (5 μm) did not show an altered pattern of *GUS* expression around the microspores, suggesting that GUS staining may be present both on the surface and within the microspores. Given that the *MS1* mutation is sporophytically inherited it seems unlikely that *MS1* is being expressed within the developing microspores. However, one can not rule out the possibility of low levels of expression in the immature pollen grains, as seen in some other anther-specific genes, for example the *APG* gene from *A. thaliana* and *Brassica napus* (Roberts *et al.*, 1993). Another alternative is that the GUS protein is being released from the tapetum during tapetal degeneration and is being deposited onto the developing pollen grains during wall development. Once complete degeneration of the tapetum has occurred no GUS expression is seen in the anthers.

In the *ms1* mutants premature degeneration of the tapetal cytoplasmic occurs and by maturation the anther locule is completely empty. The presence of *MS1:GUS* expression in the tapetal tissues and possibly within the microspores suggests a possible role for the MS1 protein in the regulation of gene expression both within the maternal anther tissue and the maturing microspores.

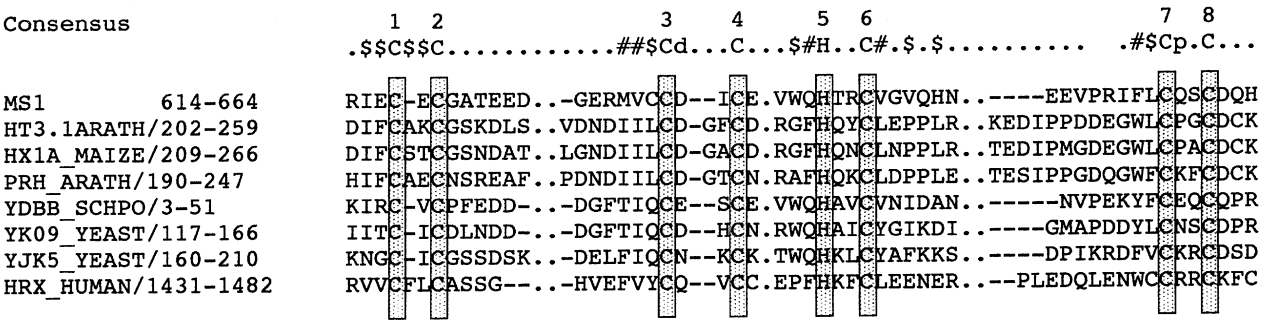


Figure 8. PHD-finger motif of MS1. Alignment of the MS1 protein and a number of other PHD-finger containing regulatory proteins (aligned by Pfam, Sanger Centre, http://www.sanger.ac.uk/software/pfam). HT3.1 (Schindler *et al.*, 1993), HX1A (Bellman and Werr, 1992), PRH (Plesch *et al.*, 1997), YDBB; Acc. No. Q10362Y, YK09; Acc. No. P36124, YJK5 (Rasmussen, 1995); HRX, Acc. No. Q03164. Consensus symbols (Asaland *et al.*, 1995) are: #, strongly conserved hydrophobicity; $, semi-conserved hydrophobicity.

The MS1 *locus encodes a PHD-finger transcription factor*

The *MS1* gene (Figure 5) shows no significant homology to any previously identified genes involved in pollen or floral development. Limited protein homology, throughout the protein, has been found to an *Arabidopsis* hypothetical protein (AC006069; Q9ZUA9) from chromosome 2 (Lin *et al.*, 1999). No known developmental function or expression patterns have been established for this gene, although it is classed as a PHD-finger transcription factor. The MS1 protein shows general high homology to this family of cysteine-rich transcription factors, the PHD-finger family. The PHD-finger contains a Cys4-His-Cys3 motif, which is closely related to the RING finger motif (Cys3-His-Cys4), in which the spacing between cysteine/histidine residues is variable, but is thought to bind two zinc ions (Asaland *et al.*, 1995). The term plant homeodomain (PHD) protein motif was coined by Schindler *et al.* (1993) for the *Arabidopsis HAT3.1* gene. The HAT3.1 protein was isolated based on its ability to bind to a light-inducible *cab-E* promoter. It has been shown to interact with DNA of greater than 100 bp and that this binding is dependent on the PHD-finger motif, indicating a role for this motif in either protein-DNA binding or protein-protein binding (Schindler *et al.*, 1993). PHD-fingers have been identified in several other plant proteins including the *Arabidopsis prha* gene (Plesch *et al.*, 1997) and the maize *Zmhox1a* gene (Bellman and Werr, 1992). The PHD-finger motifs in these genes are highly conserved, whilst the homeodomains are very different, for example the homeodomains of Zmhox1a and HAT3.1 proteins have little similarity. No defined homeodomain has been identified in the MS1 protein.

The PHD-finger motif has often been identified in animal proteins that have been implicated in tumorgenesis, recombination, DNA repair and specifically chromatin-mediated transcriptional regulation, for instance the polycomb and trithorax proteins (Asaland *et al.*, 1995; Breen, 1999). In these proteins it seems likely that the PHD-finger is a site for protein–protein interactions for the combining of a group of proteins into a large complex, or for the recognition of specific targets associated with chromatin structure and regulation. It has been speculated that the PHD-finger motif may have a function in recognising specifically modified histones (Asaland *et al.*, 1995).

In all of the *ms1* allelic mutations the PHD-finger motif has been lost. This supports the functional importance of this region. In addition to the PHD-finger domain, the MS1 protein contains a putative leucine zipper (LZ) motif (Figure 4b) implying that MS1 may form dimeric complexes (Liu *et al.*, 1999). LZ motifs are also found in both the Zmhox1a (Bellman and Werr, 1992) and PRHA proteins (Plesch *et al.*, 1997), but are absent from HAT3.1 (Schindler *et al.*, 1993). The LZ region remains transcribed in the truncated MS1–2 and MS1–3 polypeptides, but no phenotypic differences are seen in these mutants, implying that this region may be of secondary importance compared with the PHD-finger motif.

The protein motif homology, the presence of a nuclear targeting sequence and the predicted instability of the MS1 protein is supportive evidence of the role of this protein in trancriptional regulation. *MS1* gene expression is seen at low levels in immature anthers, with expression predominantly from the tapetum. At the stage of maximal *MS1:GUS* transgene expression, no defined cell walls are seen in the tapetal tissues, implying that degeneration of the tapetal cell layer is in progress. No expression is seen once complete tapetal degeneration has occurred. Low levels of expression are also seen within the developing microspores during tapetal breakdown. This may be due to *MS1:GUS* expression within the microspores, or due to the release of the GUS protein during tapetal tissue degeneration and subsequent deposition on the immature pollen grains. The *MS1* gene is likely to be a critical factor associated with the regulation of gene

expression during the later stages of anther and pollen development. The MS1 protein may, by altering the chromatin structure, transcriptionally activate other genes in the anther associated with the late stages of pollen wall development and maturation. Analysis of MS1 binding factors will provide vital information about these late acting genes involved in male sporogenesis. As such the *MS1* gene provides a valuable tool to analyse the molecular processes of anther and pollen development and will also provide a means to facilitate the future control of crop fertility, for example for F1 hybrid seed production and for the production of low-environmental impact genetically engineered crops.

Experimental procedures

Plant growth conditions and microscopy

Plants were grown as described by Dawson *et al.* (1993a). Seeds were obtained from the Nottingham *Arabidopsis* Stock Centre (The University of Nottingham, UK) and DNA stocks from the *Arabidopsis* Biological Resource Center (The Ohio State University, USA).

Flower buds were selected from a range of developmental stages and fixed, sectioned and stained as described by Dawson *et al.* (1993a).

Localisation of the MS1 *gene*

Primers CAPS12.5 ($^{5'}$ATT GCA TAG GCG CAG TTT CAC$^{3'}$ and $^{5'}$TTT TCC CCC AAA CGA AAT CCA C$^{3'}$) and CAPS2 ($^{5'}$CAA CGG GCC ATC CCT GCG AA'^{3} and $^{5'}$CGC AGC CGG AGA CTC GTC GAA'^{3}) were used to amplify DNA from parental, Ler and Sn(5)-1, and recombinant lines (Thorlby *et al.*, 1997). Three recombinants between the 16kb-4D5 BAC marker and *ms1* (0.3 cM), and 13 recombinants between *ms1* and pCIT718 (0.9 cM) were used (Thorlby *et al.*, 1997). CAPS12.5 PCRs were conducted for 35 cycles of 94°C, 30 sec; 58°C, 45 sec; 72°C, 2min. CAPS2 long-PCRs, using the Extensor DNA polymerase enzyme according to the manufacturers' recommendations (AB gene House, Epson, Surrey, UK) were carried out for 35 cycles of 94°C, 30 sec; 68°C, 9 min. The PCR products were screened using a range of different restriction enzymes to identify polymorphisms between the parental lines, Ler and Sn(5)-1. The CAPS12.5 primers gave a 1.5kb band from both the Ler and Sn(5)-1 lines, however, when digested with *Xba1* the Sn(5)-1 CAPS12.5 product gave two bands of 1.1 and 0.4kb, whilst Ler gave a single 1.5kb band. This polymorphism maps to approximately 64 kb in the AB007651 clone. Eleven of the *ttg* recombinants gave the Sn(5)-1 phenotype with the CAPS12.5 marker, whilst the remaining two *ttg* recombinants exhibited the Ler phenotype, indicating they had breakpoints beyond CAPS12.5 and close to *ms1.* As expected, in all three of the *lu* recombinants the CAPS12.5 marker co-segregated with *ms1* (Sn(5)-1 phenotype). The CAPS12.5 marker was therefore calculated as mapping 0.14 cM from *ms1*, towards *ttg* (Figure 3).

The CAPS2 primers gave a 6.7-kb band from both Ler and the Sn(5)-1 lines. When digested with *Sal1* the Sn(5)-1product (and Columbia; AB007651 sequence) gave a 5.641-kb and a 1.055-kb band, whilst the Ler product was uncut (6.696 kb). This polymorphism corresponded to position 39.425 kb in the AB007651 clone. Recombinants on both sides of *ms1* showed co-segregation (Sn(5)-1 phenotype) of the CAPS2 marker with *ms1* (Figure 3).

Twenty-one primer pairs (Table 1) were designed to amplify putative ORFs in the AB007651 clone that might correspond to *ms1*, or to long-PCR (Extensor DNA polymerase; ABgene) across defined regions of the AB007651 clone. Each long-PCR was designed to span a region of approximately 5–10kb.

Undigested fragments were size separated by agarose gel electrophoresis (Sambrook *et al.*, 1989) to look for deletions; samples were also digested with selected restriction enzymes, including *Bgl1, Dra1*, *EcoR1, HaeIII*, *HindIII, Rsa1, Sac1*, *Sal1, Sma1*, *Sty1*, *Xba1*, based on the AB007651 sequence, to identify potential polymorphisms associated with the *ms1* mutations.

RNA analyses

RNA was isolated from *Arabidopsis* tissues using the RNeasy kit (Qiagen, Crawley, W. Sussex, UK). Floral RNA was isolated from entire flowers either prior to opening (bud stage) or at the open flower stage. RNA samples were DNase treated and then purified using RNeasy spin columns (Qiagen). RT–PCR was conducted using oligodT primed first strand cDNA ($^{Reverse\text{-}iT^{TM}}$, ABgene) and RNA-specific primers spanning the first splice site (RT-1.L $^{5'}$CCA TTG CCA ATA TGT TGG TTG$^{3'}$ and RT-1.R $^{5'}$CAG CCT CAA CTC CAT TCC TT$^{3'}$; Figure 4a). Amplification was conducted using 1 μl of the cDNA template, or equivalent RT-control which lacked MMuLV reverse transcriptase, and HotStar Taq DNA polymerase (Qiagen) for one cycle of 15 min at 94°C, followed by 45 cycles of 94°C, 30 sec; 55°C, 30 sec; 72°C, 90 sec.

Analysis of *MS1* expression was conducted initially by northern hybridisations using radiolabelled *MS1* probes and filters containing 25 μg total RNA (Sambrook *et al.*, 1989). However, no detectable signal was seen. RT–PCR, as previously described, was therefore conducted using RNA isolated from buds, open flowers, leaves and the inflorescence stem. Control PCRs were conducted to check the integrity of the cDNA using *Arabidopsis* actin-2 primers (Act2F: $^{5'}$TGCTGACCGTATGAGCAAAG$^{3'}$; Act2R: $^{5'}$CAGCATCATCACAAGCATCC$^{3'}$) which generate a 419-bp fragment. Actin-2 RT–PCRs were conducted using 1 μl of the cDNA template, or equivalent RT-control which lacked MMuLV reverse transcriptase. PCRs were carried out using HotStar Taq DNA polymerase (Qiagen) for one cycle of 15 min at 94°C, followed by 35 cycles of 94°C, 30 sec; 50°C, 30 sec; 72°C, 60 sec.

RACE-PCR

RACE-PCR was conducted to obtain full-length 5′ and 3′ ends of the wt *MS1* message. Total RNA (5 μg) was isolated as previously described and 5′ and 3′ RACE-PCR conducted using the GeneRacer^{TM} kit (Invitrogen, Paisley, UK) according to the manufacturer's instructions. The mRNA was reverse-transcribed using the GeneRacer TM oligo^{TM} primer (Invitrogen) and then diluted 2-fold with sterile water. Gene specific 5′ (MS1RACE.R5′: $^{5'}$GCC GTT GAG AGA GAG CAA GTG ACC AA$^{3'}$) and 3′ (MS1RACE.L3′: $^{5'}$TCG AGA CAA GAA GTG AGG GAT GCA GCT A$^{3'}$) primers were designed based on the wt Ler *MS1* cDNA sequence. Amplification was carried out using Extensor DNA polymerase (ABgene) and 1 μl of the 2x diluted cDNA. Amplification was initially conducted at a high annealing temperature to exploit the high melting temperature of the GeneRacer^{TM} primers and to allow specific amplification products to accumulate. The annealing temperature was then lowered to maximise exponential amplification of the desired gene-specific template. The conditions used were one cycle of 94°C, 2 min; five cycles of 94°C, 30 sec and 70°C, 3 min; ten cycles of 94°C 30 sec; 70°C, 30 sec and 72°C, 3 min; thirty

cycles of 94°C, 30 sec; 68°C, 30 sec and 72°C, 3 min; one cycle of 72°C for 10 min

The 5′ and 3′ RACE PCRs gave defined bands, but with a background smear. Nested PCR was therefore conducted using 1 μl of the initial RACE-PCR product, the nested GeneRacer™ primers and nested Gene-specific primers (5′-nested-652 $^{5'}$AAA AAG CCG CCA TTG TTT CCT$^{3'}$ and 3′-nested-4991 $^{5'}$AGA CGT CGT GGT GGA GTC AGT G$^{3'}$). Amplification was for one cycle of 94°C 2 min; twentyfive cycles of 94°C, 30 sec; 56°C, 30 sec; 68°C, 90 sec and one cycle of 68°C for 10 min. Strong bands were obtained. The PCRs were then directly sequenced to identify the transcription start and polyadenylation sites.

Sequencing

PCR products (genomic and RT–PCR products) were gel, or column, purified (Qiaquick PCR clean-up, Qiagen) and sequenced using big dye terminator kit (Applied Biosystems, North Warrington, UK). Primers were designed based on the sequence database information and the determined genomic sequence. Sequencing products were analysed using an ABI310 Genetic analyser (Applied Biosystems).

Promoter fusions

A 2.9kb region upstream of the MS1 gene was PCR amplified using primers –2985XhoI: $^{5'}$CAA TGA GAC CCT CTC TCA TCT TGC$^{3'}$ and the reverse primer –31Xba1: $^{5'}$CGA ATC AGA AAT TTG GTT TGA TCT TG$^{3'}$. The PCR product was cloned upstream of the *uidA* gene between the *XhoI* and *XbaI* site of the MOG402 based Binary vector MOG*IAA2:GUS* replacing the *IAA2* promoter to create MOG *MS1:GUS*. The MOG *MS1:GUS* construct was introduced into *Agrobacterium tumifaciens* strainC58 (pGV3850) by electroporation (Wen-Jun and Forde, 1989). Transformation of Ler *Arabidopsis* was performed by the floral dip method (Clough and Bent, 1998). The transformed plants were selected on MS plates containing 50 mg l^{-1} kanamycin.

Beta-glucuronidase activity was visualised by staining whole inflorescences, leaves, and stem tissues from hemizygous transformants, overnight in X-Gluc solution (Willemsen *et al.*, 1998). Tissues were then cleared in 95% (v/v) ethanol, incubated overnight in FAA fixative (ethanol 50% v/v), acetic acid 5.0% (v/v), formaldehyde 3.7% (v/v)) and embedded in Technovit 7100 resin (Kulzer Histo-Technik, Armonk, NY, USA). Sections (5–15 μm) were then stained for 20 min in ruthenium red (0.05% (w/v), pH 9), mounted and examined.

Acknowledgements

We would like to thank M. J. Bennett for useful discussions and assistance during manuscript preparation. Special thanks to S. Walsh and S. May for bioinformatics support, and E. Okan for technical assistance. We thank K. Pyke for help with the figure preparations and discussions. Thanks also to B. J. Mulligan for useful discussions during the project. This work was supported by the BBSRC and The University of Nottingham, UK.

References

Aarts, M.G.M., Dirkse, W.G., Stiekema, W.J. and Pereira, A. (1993). Transposon tagging of a male sterility gene in *Arabidopsis*. *Nature*, **363**, 715–717.

Alexander, M.P. (1969). Diferential staining of aborted and nonaborted pollen. *Stain Technol.* **44**, 117–122.

Asaland, R., Gibson, T.J. and Stewart, A.F. (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biotechnol.* **20**, 56–59.

Bartel, B. and Fink, G.R. (1994). Differential regulation of an auxin-producing gene family in. *Arabidopsis Thaliana. Proc. Natl Acad. Sci. USA*, **91**, 6649–6653.

Bellman, R. and Werr, W. (1992). Zmhox1a, the product of a novel maize homeobox gene, interacts with the *Shrunken* 26bp *feedback* control element. *EMBO J.* **11**, 3367–3374.

Breen, T.R. (1999). Mutant alleles of the Drosophila *trithorax* gene produce common and unusual homeotic and other developmental phenotypes. *Genetics*, **152**, 319–344.

Chaudhury, A. (1993). Nuclear genes controling male fertility. *Plant Cell*, **5**, 1277–1283.

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

Couteau, F., Betzile, F., Horlow, C., Grandjean, O., Verzon, D. and Doutriaux, M.-P. (1999) Random chromosome segregation without meiotic arrest in both male and female meiocytes of a *dmc1* mutant of *Arabidopsis*. *The Plant Cell*, **11**, 1623–1634.

Dawson, J., Sözen, E., Vizir, I., Van Waeyenberge, S., Wilson, Z.A. and Mulligan, B.J. (1999). Characterization and genetic mapping of a mutation (*ms35)* which prevents anther dehiscence in *Arabidopsis thaliana* by affecting secondary wall thickening in the endothecium. *New Phytologist*, **144**, 213–222.

Dawson, J., Wilson, Z.A., Aarts, M.G.M., Braithwaite, A.F., Briarty, L.G. and Mulligan, B.J. (1993a). Microspore and pollen development in six male-sterile mutants of *Arabidopsis thaliana*. *Can. J. Bot.* **71**, 629–638.

Dawson, J., Wilson, Z.A., Briarty, L.G. and Mulligan, B.J. (1993b). Development of anthers and pollen in male sterile mutants of *Arabidopsis thaliana*. In: *Arabidopsis: an Atlas of Morphology* (Bowman, J., ed.). New York: Springer Verlag, pp. 282–295.

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

Goldberg, R., Beals, T. and Sanders, P. (1993). Anther development: basic principles and practical applications. *Plant Cell*, **5**, 1217–1229.

Gorman, S.W. and McCormick, S. (1997). Male sterility in tomato. *Crit Rev. Plant Sci.* **16**, 31–53.

He, C.P., Tirlapur, U., Cresti, M., Peja, M., Crone, D.E. and Mascarenhas, J.P. (1996). An *Arabidopsis* mutant showing aberrations in male meiosis. *Sexual Plant Reproduction*, **9**, 54–57.

Horner, I. and Palmer, R. (1995). Mechanisms of genic male sterility. *Crop Sci.* **35**, 1527–1535.

Kalantidis, K., Dawson, J., Wilson, Z.A., Briarty, L.G. and Mulligan, B.J. (1994). Isolation and characterization of twelve male sterile mutants of *Arabidopsis thaliana*. In *Proceedings of the ISPMB (Amsterdam)*, pp. 750.

Kaul, M.L.H. (1988). Male Sterility in Higher Plants, *1st Edn*, **Vol. 10** (Berlin, Heidelberg. Springer-Verlag).

Klimyuk, V.I. and Jones, J.D.G. (1997) AtDMC1, the *Arabidopsis* homologue of the yeast DMC1 gene: Characterisation, transposon-induced allelic variation and meiosis-associated expression. *Plant J.* **11**, 1–14.

Lin, X., Kaul, S., Rounsley, S.D. *et al.* (1999). Sequence and

analysis of chromosome II of *Arabidopsis thaliana*. *Nature*, **402**, 761–768.

Liu, L.S., White, M.J. and MacRae, T.H. (1999). Transcription factors and their genes in higher plants – Functional domains, evolution and regulation. *Eur. J. Biochem.* **262**, 247–257.

Mulligan, B.J., Wilson, Z.A., Dawson, J., Kalantidis, K., Vizir, I., Briarty, L.G., Thorlby, G., Morroll, S. and Shlumukov, L. (1994). The use of male sterile mutants of *Arabidopsis* to identify genes essential for male gametophyte development. *Flowering Newsletter*, **17**, 12–20.

Nakamura, Y., Sato, S., Kaneko, T., Kotani, H., Asamizu, E., Miyajima, N. and Tataba, S. (1997). Structural analysis of *Arabidopsis thaliana* Chromosome 5. III. Sequence features of the regions of 1,191,918 bp covered by seventeen physically assigned P1 clones. *DNA Res.* **4**, 401–414.

Peirson, B.N., Owen, H.A., Feldmann, K.A. and Makaroff, C.A. (1996). Characterization of three male-sterile mutants of *Arabidopsis thaliana* exhibiting alterations in meiosis. *Sexual Plant Reproduction*, **9**, 1–16.

Plesch, G., Störman, K., Torres, J.T., Walden, R. and Somssich, I.E. (1997). Developmental and auxin-induced expression of the *Arabidopsis prha* homeobox gene. *Plant J.* **12**, 635–647.

Preuss, D., Lemieux, B., Yen, G. and Davies, R. (1993). A conditional sterile mutation eliminates surface components from *Arabidopsis* pollen and disrupts cell signalling during fertilisation. *Genes Dev*, **7**, 974–985.

Rasmussen, S.W. (1995). A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and CSD3 genes, a TCP-1-related gene, an open reading frame similar to the DAL80 gene, and a tRNA (Arg). *Yeast*, **11**, 873–883.

Regan, S.M. and Moffatt, B.A. (1990). Cytochemical analysis of pollen development in wild-type *Arabidopsis* and a male-sterile mutant. *Plant Cell*, **2**, 877–889.

Roberts, M.R., Foster, G.D., Blundell, R.P., Robinson, S.W., Kumar, A., Draper, J. and Scott, R. (1993) Gametophytic and sporophytic expression of an anther-specific *Arabidopsis thaliana* gene. *Plant J.* **3**, 111–120.

Ross, K.J., Fransz, P., Armstrong, S.J., Vizir, I., Mulligan, B.J., Franklin, F.C.H. and Jones, G.H. (1997) Cytological characterisation of four meiotic mutants of *Arabidopsis* isolated from T-DNA transformed lines. *Chromsome Res.* **5**, 551–559.

Rubinelli, P., Hu, Y. and Ma, H. (1998). Identification, sequence analysis and expression studies of novel anther-specific genes of *Arabidopsis thaliana*. *Plant Mol Biology*, **37**, 607–619.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. 2nd edn: Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press.

Accession numbers: AJ344209 and AJ344210.

Sanders, P.M., Bui, A.Q., Weterings, K., McIntire, K.N., Hsu, Y.C., Lee, P.Y., Truong, M.T., Beals, T.P. and Goldberg, R.B. (1999). Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. *Sexual Plant Reproduction*, **11**, 297–322.

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

Schindler, U., Beckmann, H. and Cashmore, A.R. (1993). Hat3.1, a novel *Arabidopsis* homeodomain protein containing a conserved cysteine-rich region. *Plant J.* **4**, 137–150.

Scott, R., Dagless, E., Hodge, R., Paul, W., Soufleri, I. and Draper, J. (1991). Patterns of gene expression in developing anthers of *Brassica napus*. *Plant Mol Biology*, **17**, 195–207.

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.

Thorlby, G.J., Shlumukov, L., Vizir, I., Yang, C.-Y., Mulligan, B.J. and Wilson, Z.A. (1997). Fine-scale molecular genetic (RFLP) and physical mapping of a 8.9cM region on the top arm of *Arabidopsis* chromosome 5 encompassing the male sterility gene *ms1*. *Plant J.* **12**, 471–479.

Ursin, V.M., Yamaguchi, J. and McCormick, S. (1989). Gametophytic and sporophytic expression of anther-specific genes in developing tomato anthers. *Plant Cell*, **1**, 727–736.

Van der Veen, J.H. and Wirtz, P. (1968). EMS-induced genic male sterility in *Arabidopsis thaliana*: a model selection experiment. *Euphytica*, **17**, 371–377.

Wen-Jun, S. and de Forde, B.G. (1989). Efficient transformation of *Agrobacterium* spp by high voltage electroporation. *Nucl Acids Research*, **17**, 8385.

Willemsen, V., Wolkenfelt, H., Vrieze, G., Weisbeek, P. and Sheres, B. (1998). The *Hobbit* gene is required for formation of the root meristem in the *Arabidopsis* embryo. *Development*, **125**, 521–531.

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

Yang, W.C.Y.e.D., Xu, J. and Sundaresan, V. (1999b). The SPOROCYTELESS gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes &. Development*, **13**, 2108–2117.