An *En/Spm* based transposable element system for gene isolation in *Arabidopsis thaliana*

M.G.M. Aarts

Promotor: dr. ir. M. Koornneef
Persoonlijk hoogleraar bij de Vakgroep Erfelijkheidsleer
Co-promotor: dr. A. Pereira
Senior onderzoeker afdeling Moleculaire Biologie, CPRO-DLO

PN08201, 2194

An *En/Spm* based transposable element system for gene isolation in *Arabidopsis thaliana*

M.G.M. Aarts

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van de Landbouwuniversiteit Wageningen,
dr. C.M. Karssen,
in het openbaar te verdedigen
op maandag 16 december 1996
des namiddags te vier uur in de Aula

ISN 9313916

B. SCHIERBEEK
LANDBOUWUNIVERSITEIT
WAGENINGEN

ISBN 90-5485-622-X

The research described in this thesis was supported by the European Community BRIDGE/BIOTECH project on Development and use of *Arabidopsis thaliana* as a tool for isolating genes of agronomic importance (grant no. BIOT-CT90-0207).

NN08201, 2194

Stellingen

1. In vergelijking met *Ac-Ds* systemen is de continue transpositie van het 'in cis two-element' *En/Spm-I/dSpm* systeem een groot praktisch voordeel voor transposon-tagging in Arabidopsis.

2. De mutatiefrequentie van een gen in een doelgericht transposonmutagenese-experiment, wordt behalve door de genetische afstand tussen doelwitgen en transposondonorlocus, door nog onbekende factoren bepaald.

3. Het maken van onderscheid tussen *Enhancer/Inhibitor* versus *Suppressor-mutator/defective Suppressor-mutator* transposons, wordt niet meer gerechtvaardigd door wetenschappelijke argumenten.

4. Mannelijke steriliteit leidt tot een niet te missen fenotype in Arabidopsis.

5. Het is wetenschappelijk onverantwoord de functie van een eiwit te voorspellen louter op basis van overeenkomsten tussen de aminozuurvolgorde van het eiwit en die van een ander eiwit.

6. Het nut van wetenschappelijk onderzoek aan het modelgewas Arabidopsis, dringt slechts langzaam door tot de plantenveredeling.

7. Het financieren van grootschalige DNA-volgordebepaling van een organisme heeft geen zin als niet ook op grote schaal genfunctie-analyse-onderzoek gefinancierd wordt.

8. Het huidige natuurbeleid is weinig op natuur maar voornamelijk op menselijke wensen en voorkeuren afgestemd.

9. Het invoeren van nieuwe spellingregels leidt eerder tot een generatiekloof dan tot een uniforme spelling.

10. De Amsterdam Arena doet haar naam eer aan.

    (Van Dale Groot Woordenboek der Nederlandse Taal: "arena (v.(m.): -'s) [Lat.], 1 met zand bestrooid middendeel in een klassiek amfitheater, in een circus enz.;...")

11. Het verdedigen van stellingen bij een wetenschappelijke promotie lijkt een goed gebruik maar dient een onduidelijk doel.

12. Op de korte termijn is biologische landbouw voor de belastingbetaler geen goed alternatief voor reguliere landbouw.

Stellingen behorende bij het proefschrift getiteld "An *En/Spm* based transposable element system for gene isolation in *Arabidopsis thaliana*", door M.G.M. Aarts, in het openbaar te verdedigen op 16 december 1996, te Wageningen.

Inhoud

Voor mijn lieve ouders

General introduction

The concept of gene tagging

Molecular plant genetic research depends largely on the study of genotypic variants for the functional analysis of genes. Such variants are either available as part of the genetic variation within a plant species, or they can be generated as mutants, by applying ionizing radiation or DNA-damaging chemical agents to plant parts. A prerequisite for the functional analysis of a gene that determines a certain phenotype, is the isolation of the gene. As plants contain several ten thousands of genes, dispersed over at least hundreds, but often thousands of millions of DNA base pairs, this seems like looking for a needle in a haystack. One efficient way to isolate a gene for which a phenotype is known, is by map based cloning. This technique uses genetic markers closely flanking the gene of interest to subclone part of the genomic region surrounding the gene in more amenable host organisms (Tanksley et al., 1995). When the position of the target gene on the physical map is determined, the appropriate DNA fragments can be used to transform mutant plants for the complementation of the mutant phenotype. Eventually this results in one DNA clone containing little more than the gene of interest, which DNA sequence can be determined.

Map based cloning relies on the continuously accumulating amount of physical genome mapping data that are generated for increasingly more plant species. It has already led to successful gene cloning in Arabidopsis (e.g. Arondel et al., 1992), tomato (Martin et al., 1993) and rice (Song et al., 1995). However, the need for physical mapping data is a disadvantage when the gene of interest is located on a part of a chromosome for which such data are not available. Then it is often easier to use insertional mutagenesis or gene tagging as the way to clone the gene of interest. The concept of gene tagging is based on the phenomenon that the insertion of a piece of DNA into a gene often causes a mutation that blocks the activity of the gene. When such a mutation is present in a homozygous situation, it can lead to a visible aberrant phenotype. If the DNA insert is well characterized in terms of DNA sequence, then the insert can act as a molecular tag of the mutated gene. With molecular biological techniques it is relatively easy to recover the DNA flanking the tag, which can be subsequently used to obtain a functional copy of the tagged gene.

In principle any way of generating genomic insertions with characterized DNA fragments can be used for gene tagging in a (plant) species. Various tagging systems have been developed in plants. Inserts that are used are: a) the T-DNA molecule, transferred into the plant genome by *Agrobacterium tumefaciens* (Feldmann et al., 1989); b) retrotransposons, such as the Tnt1 element from tobacco, which is also functional in the heterologous host Arabidopsis (Grandbastien, 1992; Lucas et al., 1995); c) transposable elements.

Plant transposable elements

The first gene tagging experiments in higher organisms were carried out in the fruit fly,

Drosophila melanogaster, using well studied transposable elements as tags (Bingham et al., 1981). Transposable elements or transposons are genetic elements with the ability to 'transpose', i.e. move through the genome of an organism by an excision and insertion mechanism (Saedler and Nevers, 1985). Often transposons themselves bear the genes that encode for the transposase proteins required for transposition. These transposases specifically recognize the ends of a transposable element, they cut the element out of the genome and insert it at another position. Transposable elements that are able to move on their own account are called autonomous elements. This in contrast to non-autonomous elements, which are often derived from autonomous elements by deletion or mutation of the transposase genes. In the absence of transposase proteins, non-autonomous elements are immobile, but they can be activated to transpose in the presence of an autonomous element.

Plant transposable elements have been discovered and studied in detail by the maize geneticist Barbara McClintock (McClintock, 1948), who could attribute several cases of genetic instability to the existence of mobile genetic elements. Later it turned out that transposable elements are present in many plant species (Walbot, 1992), but still the maize elements are the best characterized. The *Activator* (*Ac*) element was the first autonomous transposable element described (McClintock, 1951). This element can activate a family of non-autonomous *Dissociation* (*Ds*) elements, named after the frequently observed chromosome breakages associated with *Ds* insertions. *Suppressor-mutator* (*Spm*), another autonomous element described by McClintock (1954), is named after its two functions: suppression of gene expression and gene-mutation. Non-autonomous *Spm* elements are called *defective Spm* (*dSpm*) (Fedoroff, 1989). Previous to the discovery of *Spm*, the autonomous *Enhancer* (*En*) element was described (Peterson, 1953), which can activate non-autonomous *Inhibitor* (*I*) elements (Peterson, 1960). After isolation and DNA sequencing of both the *En* and *Spm* elements from maize (Masson et al., 1987; Pereira et al., 1985 and 1986), they turned out to be virtually identical. Nonetheless both names are still used.

Soon after the molecular characterization of *Ac* (Fedoroff et al., 1983) and *En/Spm* (Masson et al., 1987; Pereira et al., 1985 and 1986) their importance for molecular genetic analyses was demonstrated by the isolation of transposon tagged genes in the anthocyanin biosynthesis pathway of maize (Fedoroff et al., 1984; O'Reilly et al., 1985). In maize, *Ac* and *En/Spm* elements have mutations frequencies of around 1 in 10^{-5} to 10^{-6} (Walbot, 1992), which makes them attractive tools for insertional mutagenesis. Despite the fact that transposable elements are endogenous to many plant species, the transposon based gene tagging technique has been developed for only a few species. Besides maize, endogenous transposon tagging systems are now established for snapdragon (*Antirrhinum majus*) and Petunia using the *Tam*, respectively the *dTph* transposable element families (Coen et al., 1989; Gerats et al, 1990). The success of transposon tagging encouraged the development of heterologous tagging systems, using maize transposons *Ac-Ds* and *En/Spm-I/dSpm* in species lacking well characterized endogenous transposons, such as tobacco, tomato and *Arabidopsis thaliana* (reviewed by Balcells et al., 1991; Haring et al., 1991). These heterologous tagging systems are now so well established, that it remains to be seen if newly isolated endogenous transposons, like the recently isolated active *Tag1* element from

Arabidopsis (Tsay et al., 1993), will still be able to replace them.

Arabidopsis as the species of choice for heterologous transposon tagging.

Thale cress *(Arabidopsis thaliana)* or Arabidopsis for short, has through the years become the model plant species in (molecular) genetic research for physiological, biochemical and developmental processes. The small size of the plant and the short life cycle make it ideal for the generation and characterization of mutant phenotypes, and the small genome of only around 100 million base pairs (Mb) makes molecular genome analysis relatively easy compared to many crop species (Meyerowitz, 1989). Especially the combination of molecular biology and genetics has fuelled the Arabidopsis research field to emerge at the frontier of plant research (Meyerowitz and Somerville, 1994).

With an increasing interest in Arabidopsis as a genetic model, the demand for gene isolation tools increased, and therefore soon after their isolation, maize transposable elements were introduced into Arabidopsis. Two types of Arabidopsis heterologous transposon tagging systems have been studied, either based on one or on two elements. In a one element system, an autonomous element is used, while in a two element system, a non-autonomous element is used in combination with a transposase source. Heterologous systems have the practical advantage over endogenous systems that they can be accommodated with several selectable marker genes. In general all heterologous systems are introduced via *Agrobacterium tumefaciens* transformation using T-DNA vectors. These vectors carry antibiotic resistance genes as selectable markers for their presence in the plant genome. The transposable element is often cloned between a promoter and another selectable marker gene, blocking expression of this gene. When the element excises, the activity of the selectable marker gene is restored, which can be monitored in a phenotypic excision essay (Baker et al., 1987). The transposition activity of heterologous transposable element systems is often determined by the germinal excision frequency. This is the fraction of seedlings with an excised transposon among the total number of seedlings in the progeny of a plant.

Initially heterologous systems based on *Ac-Ds* have been developed. When the *Ac* element was first introduced into Arabidopsis tissue it was found to transpose at a very low frequency (Van Sluys et al., 1987). A germinal excision frequency of 0.2-0.5 %, as determined by Schmidt and Willmitzer (1989) and Dean et al. (1992), is insufficient for efficient gene tagging. The transposition frequency could be increased by about 10-fold when a 560 bp methylation sensitive CpG-rich NaeI fragment was deleted from the 5' untranslated leader of the *Ac* transposase gene (Lawson et al., 1994). This element has since been proven to be useful for the isolation of tagged mutants (Bhatt et al., 1996).

Higher germinal excision frequencies were obtained when *Ds* elements were mobilized by stable transposase sources (Bancroft et al., 1992; Grevelding et al., 1992; Honma et al., 1993; Long et al., 1993b; Swinburne et al., 1992). Frequencies of over 30% could be achieved when expression of the *Ac* transposase gene was controlled by the CaMV 35S promoter. This high excision frequency was not as advantageous as expected, as it resulted in predominantly early transpositions and only few different *Ds* inserts transmitted to the next

generation (Long et al., 1993b). Despite this drawback, most of the *Ac-Ds* based two element systems published so far have been used for the generation of tagged mutants and analysis of the genes involved (Bancroft et al., 1993; Long et al., 1993a). *Ds* elements have been further successfully modified, such as by incorporation of a read-out promoter inducing (dominant) overexpressor mutants upon *Ds* insertion (Wilson et al., 1996) or by construction of *Ds* gene-trap elements that are able to detect the activity of nearby genes by activation of a marker gene present on the transposon (Fedoroff and Smith, 1993; Springer et al., 1995; Sundaresan et al., 1995; Tsugeki et al., 1995).

In contrast to the *Ac-Ds* elements, the *En/Spm-I/dSpm* elements have not been widely used in heterologous tagging experiments, which was one of the reasons to start the research described in this thesis. Based on experiments with tobacco using an unaltered maize *En/Spm* element (Pereira and Saedler, 1989), a one element *En/Spm* system was used in Arabidopsis (Cardon et al., 1993a), which surprisingly gave a much higher transposition frequency when compared to *En* transposition in tobacco or potato (Cardon et al., 1993b; Frey et al., 1989) or to *Ac* transposition in Arabidopsis. A reason for the efficient transposition was found in the level of transcription and the relative abundance of the two *En/Spm* transposase mRNAs, which in Arabidopsis resemble the situation in maize much more than either *En/Spm* in Solanaceous species or *Ac* in Arabidopsis. Important for gene tagging purposes were the occurrence of independent *En/Spm* transpositions, and an average germinal excision frequency of 7.5%, which remained constant over a number of generations. Cardon et al. (1993a) also described the use of a three element system, consisting of a non-autonomous *dSpm* element, which is activated to transpose in the presence of two transposase genes, *tnpA* and *tnpD*, each controlled by a CaMV 35S promoter. Transposition of this *dSpm* element in Arabidopsis was demonstrated, but the system has not been characterized in detail.

Arabidopsis as a source of genes of agronomic importance

For a long time, Arabidopsis was considered to be just a model species: very useful to answer fundamental research questions, but with apparently no agricultural importance. This concept has gradually changed and at present Arabidopsis research is an important source of molecular genetic and biochemical data that are used for plant breeding purposes. For example, the isolation of fatty acid biosynthesis genes has greatly enhanced the knowledge of the fatty acid biosynthesis pathway in plants. It has resulted in the isolation of homologous genes from *Brassica napus* or other species, which are now used to modify the fatty acid composition and content of oil seed rape (Töpfer et al., 1995). Another example is the characterization of several genes involved in flower morphology (Weigel and Meyerowitz, 1993). Overexpression of some of these genes induces early flowering, which has major implications for tree breeding (Mandel and Yanofsky, 1995; Weigel and Nilsson, 1995). Another way to induce early flowering is by overexpression of a flowering time gene isolated from Arabidopsis (Igeno et al., 1996). Furthermore, the identification and isolation of disease resistance and pathogenesis related genes from Arabidopsis (Crute et al., 1994; Kunkel, 1996), has opened new possibilities for the engineering of disease resistant crop varieties

(Staskawicz et al., 1995). In retrospect, much of the Arabidopsis research, that originally started with the purpose of answering fundamental research questions, is now producing spin-offs useful for crop plant improvement.

Two plant characters interesting for transposon tagging are male fertility and epicuticular wax biosynthesis. Mutations of genes in these processes often give obvious phenotypes, such as male sterility and wax layer deficiency, that are attractive markers to screen for in transposon mutagenesis experiments. In addition, the genes controlling fertility or wax biosynthesis are of agronomic importance. Cytoplasmic or nuclear male sterility is frequently used for the production of F_1 hybrid seeds yielding superior crops. Mariani et al. (1989) reported the first example to achieve an artificial male sterility system using recombinant DNA techniques. Male fertility genes can similarly be used for the design of a controlled male sterility system. In Arabidopsis, so far over 30 male sterile (ms) mutations have been described (Chaudhury, 1993; Mulligan et al., 1994).

Wax biosynthesis genes determine the production of epicuticular wax, a protective hydrophobic layer covering nearly the entire superterrestrial surface of a plant. This layer has been found to be important for drought tolerance, insect and microbial resistance and pollen-pistil signalling; all agronomically important traits. The wax biochemical pathway has been studied in detail for several decades now and a number of biochemical steps have been defined (Kolattukudy, 1976; Post-Beittenmiller, 1996; Von Wettstein-Knowles, 1994). Wax components are all derived from long chain fatty acids, which are elongated, reduced or decarbonylated. As these substances can be further modified to be used as food additives or cosmetic compounds, the wax biosynthesis genes have potentially interesting industrial applications. In Arabidopsis there are at least 21 different *eceriferum* or *cer* genetic loci involved in wax biosynthesis (Hannoufa et al., 1993). For some of these genes, a specific biochemical function has already been proposed (Lemieux et al., 1994).

Scope of this thesis

The main aim of the research described in this thesis was to develop, study and apply an efficient transposon tagging system for the generation of tagged mutants and to provide further insights in the possibilities for future applications of such a transposon tagging system in studying plant gene functions. The choice to use the maize *En/Spm-I/dSpm* system in Arabidopsis does not stand on its own, but was partially instigated by serious attempts of others trying to reach similar targets. At the time when this research started, the first experiments with the *Ac-Ds* element based transposon systems had demonstrated the possibilities of heterologous transposition in Arabidopsis (Masterson et al., 1989; Schmidt and Willmitzer, 1989; Van Sluys et al., 1987), while at the same time the T-DNA, transferred by *Agrobacterium tumefaciens* into plants, appeared suitable as a molecular tag (Feldmann, 1991; Koncz et al., 1990).

When this research was initiated, several questions had to be answered. First of all, would the *En/Spm-I/dSpm* system be active in Arabidopsis? And, if it was active, would it remain active over many generations? Furthermore, how would it compare to other transposon

tagging systems in Arabidopsis? These questions are dealt with in chapter 1. With an active transposon system, it was possible to determine transposition frequencies and the distribution of transposed elements over the Arabidopsis genome. Both issues are also described in chapter 1. An important reason for using a transposon tagging system is to obtain tagged mutants. Was this at all possible with the *En/Spm-I/dSpm* system? This question is answered in chapter 2, which reports on the isolation of the first *I/dSpm* tagged mutant, an insertion in the *MALE STERILITY 2* (*MS2*) gene. The further characterization of the *MS2* gene, to understand its role in male gametogenesis, is described in chapter 3. The isolation of another important gene, the *CER1* gene involved in wax alkane decarbonylation, is described in chapter 4.

When it was established that the *En/Spm-I/dSpm* system could be used successfully for the cloning of Arabidopsis genes, it became interesting to determine the most efficient strategy to obtain not only many different mutants but also mutants at specific genes for which the approximate location was already known. In chapter 5, random and targeted tagging strategies are discussed, and suggestions are made for the creation of a 'mutation machine'. An interesting aspect of the biology of the *En/Spm-I/dSpm* transposable elements is the influence which *En/Spm* transposase expression has on the expression of some *I/dSpm* tagged genes, as is described in chapter 6. This complex transposon related phenomenon was before only reported for endogenous transposable elements (Schwarz-Sommer et al., 1987), but it can also be studied in a genetically much simpler species like Arabidopsis. The last chapter of this thesis describes the use of *I/dSpm* elements as gene traps. This application of transposable elements has turned out to be a major tool in *Drosophila* and mouse molecular genetic research, but its development in plants has just started (Topping and Lindsey, 1995).

Parts of the work described in this thesis have been or will be published elsewhere:

Aarts, M.G.M., Corzaan, P., Stiekema, W.J. and Pereira, A. (1995). A two-element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **247**, 555-564.

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.

Aarts, M.G.M., Keijzer, C.J., Stiekema, W.J. and Pereira, A. (1995). Molecular characterization of the *CER1* gene of Arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* **7**, 2115-2127.

Aarts, M.G.M., Hodge, R., Kalantidis, K., Florack, D., Wilson, Z.A., Mulligan, B.J., Stiekema, W.J., Scott, R. and Pereira, A. The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. Submitted to *Plant J.*

Pereira, A. and Aarts, M.G.M. Transposon tagging with the *En-I* transposon system. In *Arabidopsis protocols*. J.M. Martinez-Zapater and J. Salinas (eds), series on Methods in Molecular Biology. Humana Press, Totowa, NJ, USA. In press.

Other publications by the author related to the topic of this thesis:

Pereira, A., Aarts, M., van Agtmaal, S., Stiekema, W.J. and Jacobsen, E. (1991). *Waxy* variegation in transgenic potato. *Maydica* **36**, 323-327.

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

Chapter 1

A two element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana*

Mark G.M. Aarts, Paul Corzaan, Willem J. Stiekema and Andy Pereira[1]

(Published in *Molecular and General Genetics* (1995) **247**, 555-564)

Summary

The *Enhancer-Inhibitor (En-I)*, also known as *Suppressor-mutator (Spm-dSpm)*, transposable element system of maize was modified and introduced into *Arabidopsis thaliana* by *Agrobacterium tumefaciens* transformation. A stable *En/Spm* transposase source under control of the CaMV 35S promoter mediated frequent transposition of *I/dSpm* elements. Transposition occurred continuously throughout plant development in at least seven consecutive plant generations after transformation. New insertions were found at both linked and unlinked positions relative to a transposon donor site. The independent transposition frequency was defined as a parameter for transposition, which quantified the rate of unique insertion events from 4.4 to 29.2% in different populations. An increase as well as a decrease in *I/dSpm* element copy number was seen on individual plant level, but not on population level after several plant generations. The continuous frequent transposition as observed for this transposon system makes it an attractive tool to use for gene tagging in Arabidopsis.

[1]Unless stated otherwise, the authors of the chapters in this thesis are affiliated with the Department of Molecular Biology, DLO-Centre for Plant Breeding and Reproduction Research (CPRO-DLO), Postbus 16, 6700 AA Wageningen, The Netherlands.

Introduction

Due to the well characterized endogenous transposons, transposon tagging is the system of choice when isolating a gene correlated with a mutant phenotype from *Zea mays* (maize) (Fedoroff et al., 1984) and *Antirrhinum majus* (snapdragon) (Martin et al., 1985). Maize (*Ac-Ds*, *En/Spm*) and snapdragon (*Tam3*) transposons have also been transferred to other plant species for heterologous transposon tagging (Haring et al., 1991). Successful tagging of the *PH6* gene in *Petunia hybrida* with the maize *Ac* element was reported by Chuck et al. (1993), while Bancroft et al. (1993) tagged the *DRL1* gene using an *Ac-Ds* system in Arabidopsis.

The autonomous *Enhancer (En)* or *Suppressor-mutator (Spm)* transposable element, originally identified in maize (Peterson (1953) and McClintock (1954)), has been molecularly characterized (Masson et al., 1987; Pereira et al., 1986) and also introduced into several heterologous hosts (Cardon et al., 1993a; Frey et al., 1989; Masson and Fedoroff, 1989; Pereira and Saedler, 1989). We recently reported the use of a maize *En/Spm* derived system to tag and clone the *MS2* gene from Arabidopsis (Aarts et al., 1993; see chapter 2). Non-autonomous derivatives of *En/Spm* are termed *Inhibitor (I)* or *defective Spm (dSpm)* elements. The T-DNA construct we used to introduce an *En/Spm-I/dSpm* '<u>in cis</u> two element system' into plants, contains an unmarked *I/dSpm* element inserted in the open reading frame of an NPTII marker gene, to monitor excision phenotypically. The same construct contains an *En/Spm* element as transposase source, of which both 5' and 3' ends have been removed, creating a stable *En/Spm*, and replaced by CaMV 35S promoter (5') and terminator (3') sequences. They control transcription of the two mRNAs that are needed to produce the TNPA and TNPD transposase proteins (Frey et al, 1990; Masson et al., 1991).

This chapter gives a detailed description of the *En/Spm-I/dSpm* '<u>in cis</u> two element tagging system in Arabidopsis. Four basic aspects concerning the transposition behaviour of this system, have been studied: 1) The frequency of transposition. 2) Possible limitation of activity of the transposon system after a number of generations. 3) Transpositional loss of *I/dSpm* elements. 4) Transposition to linked sites. These aspects were examined by phenotypic and molecular genetic means in seven subsequent generations. The prospects of this transposon system for efficient gene tagging in Arabidopsis will be discussed.

Results

The '<u>in cis</u> two element system'

The *I/dSpm* element transposon as well as the *En/Spm* transposase genes have been cloned on one T-DNA construct cwEnN::I (Figure 1). In contrast to most phenotypic assay constructs in which the transposon is inserted in the 5' untranslated leader of a marker gene (Baker et al., 1987), we cloned the *I/dSpm* element directly downstream of the ATG startcodon of a neomycin phosphotransferase (NPTII) gene. This seemed necessary as *I/dSpm* element insertion in the untranslated leader of the NPTII gene did not completely block expression of kanamycin resistance (Cardon et al., 1993a; Pereira and Saedler, 1989). The

reading frame of the NPTII gene is restored to display kanamycin resistance when *I/dSpm* element excision events add 1+3n basepairs or delete 2+3n basepairs. Excision of an *I/dSpm* element is mediated by transposase proteins expressed from a stable *En/Spm* element under control of the CaMV 35S promoter.

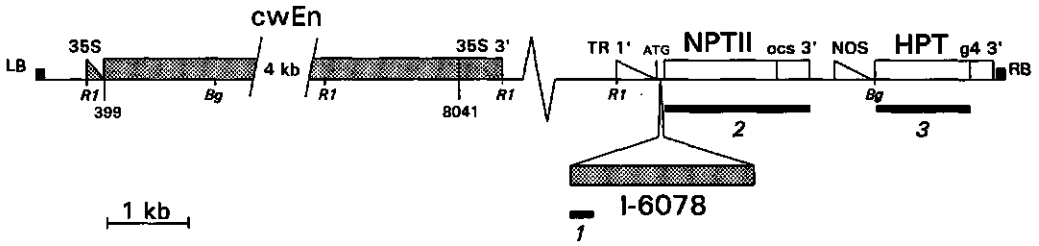

Figure 1: Schematic drawing of the T-DNA construct cwEnN::I, containing the '*En/Spm-I/dSpm* in cis two element' transposon system.
A stable wings-clipped *En-1* element expressing transposase genes *tnpA* and *tnpD*, is controlled by CaMV 35S promoter and terminator sequences. A 2.2-kb *I/dSpm* element (*I-6078*), cloned in the open reading frame of a NPTII gene, in opposite orientation relative to the *En/Spm* transposase, can give kanamycin resistance upon excision. Hygromycin resistance as conferred by a HPT gene is used to select for transformation. Promoter sequences are depicted as triangles, pointing in the direction of the transcription unit they control. 399 and 8041 are the positions of the BssHII sites in *En-1* used for cloning the cwEn. Fragments used as probes for DNA blot hybridizations are indicated with horizontal filled bars numbered 1 to 3. Probe 1 is a 0.27-kb fragment of the *En/Spm-I/dSpm* left border (up to SalI site). Probe 2 and 3 are the full coding plus terminator sequences of the NPTII and HPT genes respectively. The positions of EcoRI, BglII and HindIII restriction sites are given below the construct as R1, Bg and H. ATG marks the translation start codon of the NPTII markergene. LB and RB are the left respectively the right border of the T-DNA construct.

Phenotypic analysis of I/dSpm element excision

T-DNA cwEnN::I was introduced into Arabidopsis by *Agrobacterium* mediated transformation (Valvekens et al., 1988) and progeny from four transgenic shoots out of one transformed callus (1A to 1D), displaying active transposition, was further analysed. T_2 seeds obtained from these *in vitro* grown T_1 plants were germinated and segregation of T_2 plants for hygromycin resistance, showed that the original callus was hemizygous for two unlinked T-DNA loci (data not shown). DNA blot analysis revealed that at these *En/Spm* transposase loci respectively 5 and 2 right border sequences of T-DNA were integrated, which will be further referred to as the TEn5 and TEn2 loci respectively. Another part of the T_2 seedlings was screened for excision on kanamycin containing medium (Table 1). The ratios of kanamycin resistant to variegated to sensitive plants varied substantially between the four lines. Variegation for kanamycin resistance, visible as a few green spots on white cotyledons, and caused by excision events in the T_2 plant itself, showed that the excision rates between progenies 1A, 1C and 1D was similar. Only progeny from tetraploid T_1 plant 1B contained significantly more variegated T_2 plants, probably due to the two times higher T-DNA (and thus NPTII::I) copy number per cell.

Table 1: Assaying T_2 progeny of primary transformants carrying the cwEnN::I T-DNA for excision. The percentages kanamycin resistant (K^r), variegated (K^v) and sensitive (K^s) plants of the total number of plants (n) are given. Plant 1B is tetraploid, the others are diploid.

plantnr.	K^r (%)	K^v (%)	K^s (%)	n
1A	13	17	70	76
1B	0	32	68	19
1C	2	17	81	48
1D	21	14	65	57

Molecular analysis of I/dSpm element excision

The frequency of kanamycin variegated T_3 plants (descending from kanamycin variegated T_2 plants) was in general much lower than in the previous generation (Table 2). A reason for this reduction could be that only part of the excisions result in a functional NPTII gene, and that after two generations, there is a substantial reduction in the number of T-DNA inserts still containing an *I/dSpm* element in the NPTII gene. This was confirmed after analyzing the T_3 generation by hybridization of EcoRI/BglII digested DNA from randomly chosen single T_3 plants, descending from different T_2 plants, with a NPTII specific probe. Both the original T-DNA fragment was detected (5.3 kb) as well as an empty donor site fragment of 3.1 kb (Figure 2). In addition, hybridizing fragments other than 5.3 kb or 3.1 kb (lanes 2 and 10, figure 2) were detected, which might be caused by aberrant excision or somatic DNA rearrangements associated with transposition. Based on the intensity of the 5.3 and 3.1-kb DNA bands, the fraction of unexcised *I/dSpm* elements, capable of giving kanamycin resistance, was much smaller than the fraction of empty donor sites explaining that effective use of the phenotypic assay to monitor excision was limited to the first generation after transformation.

Genetic positions of T-DNAs

The TEn2 T-DNA locus was found to comprise a single complete T-DNA copy, while the TEn5 locus has 2 to 4 complete T-DNA copies. The T-DNA borders plus some flanking plant DNA sequences were isolated by plasmid-rescue (Koncz et al., 1990) and the flanking DNA was used as an RFLP probe on a population of Recombinant Inbred Lines (RILs) (Lister and Dean, 1993). The TEn5 locus was mapped to chromosome 2, approximately 2 cM from g17288 between markers m220 and g17288 (Hauge et al, 1993). The TEn2 locus could not be mapped with RILs as its flanking DNA did not detect a RFLP between the Landsberg *erecta* and Columbia ecotypes used to make the RIL population. By classical genetic analysis using a *I/dSpm* element tagged *cer1* mutant (Aarts et al., 1995b; see chapter 4), the TEn2 locus was found to reside on the top of chromosome 1, very closely linked to *cer1*.

Table 2: Assaying T_3 progeny of kanamycin variegated T_2 plants carrying the cwEnN::I T-DNA for excision. The percentages kanamycin resistant (K^r), variegated (K^v) and sensitive (K^s) plants of the total number of plants (n) are given.

T_2 plantnr.	K^r (%)	K^v (%)	K^s (%)	n
1A-H13	0	3	97	412
1A-K20	54	7	39	139
1A-9	38	15	47	143
1A-54	0	8	92	75
1B-6	2	3	95	113
1B-13	0	5	95	104
1C-1	68	4	28	195
1C-3	1	7	92	156
1C-8	0.5	0.5	99	197
1D-8	0	4	96	165
1D-10	0.6	3	96.4	179
1D-33	0.6	6	93.4	174
1D-51	0.6	6	93.4	180

Reinsertion of excised I/dSpm elements

For analysis of reinsertion, the progeny of three T_2 plants (1A-H2, 1A-H12, 1A-H36; see Figure 3 for a pedigree of all lines and plants described), hemizygous for one *En/Spm* transposase locus were studied. Plants 1A-H2 and 1A-H36 carried the TEn5 locus, whereas plant 1A-H12 carried the TEn2 locus. Individual transposed *I/dSpm* elements could be identified as DNA fragments of specific length by blot analysis of DNAs from randomly chosen plants using *Hind*III, which does not cut within the *I/dSpm* elements, as restriction enzyme and an *I/dSpm* element specific probe (Figure 4). To check whether all bands represented *I/dSpm* elements transposed from the T-DNAs, the same blot was hybridized with the HPT probe (see Figure 1) to discriminate between T-DNA with or without an *I/dSpm* element. Only progeny of 1A-H12 contained an *I/dSpm* element inserted in the T-DNA (data not shown). Most plants with transposase genes (Figure 4) displayed a background smear of hybridization apart from distinct bands, which is due to somatic transposition. From the observation of a number of plants which lacked T-DNA but contained transposed *I/dSpm* elements, it was concluded that most of the *I/dSpm* elements had transposed to sites genetically unlinked to the T-DNA. A high frequency of transposition was suggested by the large number of DNA fragments accommodating *I/dSpm* elements and by the variation in DNA fragment number and pattern between the three T_2 lines and T_3 plants derived from each T_2 line (Figure 4).

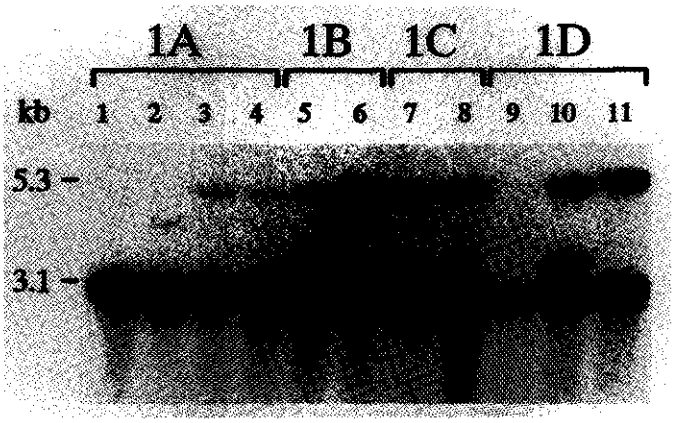

Figure 2: DNA blot of T_3 plants propagated by single seed descent from the four primary transformed T_1 shoots 1A to 1D.
DNAs were digested with EcoRI and BglII and hybridized with the NPTII specific probe (Figure 1). This probe discriminates between T-DNA with (5.3-kb band) and without *I/dSpm* element insert (3.1-kb band). The difference in intensity between these two bands reflects the ratio of T-DNA with unexcised versus excised *I/dSpm* elements. Lanes 2 and 10 show bands of aberrant fragment lengths.

Quantification of transposition

As a way to quantify transposition in every generation, the independent transposition frequency (itf) was determined. The itf is defined as the frequency of unique *I/dSpm* element inserts relative to the total number of *I/dSpm* element homologous DNA fragments that appear on a DNA blot made from a population of plants containing *En/Spm* transposase, all descending from one plant or cross. A minimal population size of eight progeny plants has been chosen, as there is only a small chance that, after Mendelian segregation, a plant is found in a selfed progeny that has a unique *I/dSpm* element insert (e.g. the chance of one plant in a progeny of eight having a unique element insert, while the parent was hemizygous for this insert, is 3.6×10^{-4}). Only bands that had a higher intensity than about half times the intensity of hemizygous bands were counted and HindIII digests were used, which best revealed discriminatory bands.

The itf in progeny of the male sterile plant carrying a mutation caused by an *I/dSpm* element insertion in the *MS2* gene of Arabidopsis (Aarts et al., 1993; chapter 2) was first determined. This mutant appeared in the T_3 generation derived from plant 1A-H2. As *ms2* mutants confer a low level of self-fertilization at the end of flowering, several generations were obtained by selfing in the course of examining the nature of the mutation. In 18 T_4 progeny of the T_3 male sterile mutant, hemizygous for TEn5, DNA blot analysis revealed 10 unique inserts on a total of 127 inserts. The itf for this population was 7.8% (10/127 x 100%). Among the T_4 progeny was a variegated male sterile plant with a large sector of fertile flowers (Aarts et al., 1993). Both male sterile and fertile flowers were allowed to self-pollinate as well as out-crossed with Landsberg *erecta*. A blot containing DNA from all four progenies was hybridized with the *I/dSpm* element specific probe (Figure 5). When the progenies derived from outcross and self-pollination were regarded separately, similar itfs

were found (9.4% (10 in 106) and 12.4% (18 in 145) respectively), not significantly different from the T_4 generation itf.

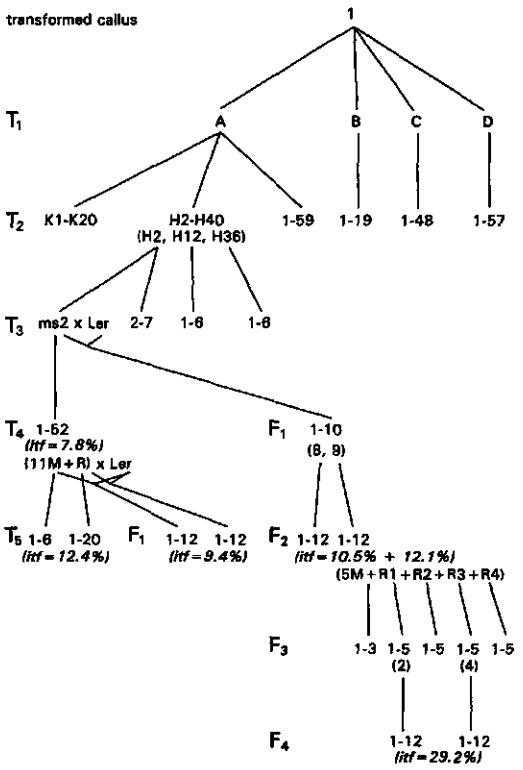

Figure 3: Pedigree of plants used for analysis of transposition.
Transformation of Arabidopsis with construct cwEnN::I yielded callus 1, producing four transformed shoots 1A, 1B, 1C and 1D (T_1). Self-fertilization of the primary transformants resulted in T_2. Some T_2 plants were selected on hygromycin (H) or kanamycin medium (K). T_3 progeny from self-fertilized plants 1A-H2, 1A-H12 and 1A-H36 were used for further analysis. A *ms2* male sterile plant (ms2) was found amongst 1A-H2 progeny. The number of *ms2* progeny plants used to study continuous transposition is given for every generation. Plants selected from a progeny to produce the next generation are given between brackets. M+R and M+R1+R2+R3+R4 indicate that revertant (R) sectors were found on the male sterile (M) plants indicated. The independent transposition frequency (*itf*) is given for every progeny tested.

The high number of elements in the T_4 and T_5 progeny increased the probability that elements of approximately the same restriction fragment length could not be distinguished. In order to obtain plants with less or (preferably) only one *I/dSpm* element, the original male sterile T_3 mutant was crossed to Landsberg *erecta* (Figure 3) and subsequent progenies were made and analysed for the transposition rate. The F_2s derived from two F_1 plants hemizygous for *En/Spm* transposase showed similar itfs (10.5% (2 in 19) and 12.1% (4 in 33) respectively) as found for the T_4 and T_5 progenies.

One male sterile F_2 plant (*ms2::I/dSpm,ms2::I/dSpm*) with only two inserts and

hemizygous for *En/Spm* transposase was chosen and the F_3 progeny from fertile flowers (resulting after excision of the *I/dSpm* element from the *ms2::I/dSpm* locus) were screened for plants with only one *I/dSpm* element. Two F_3 plants (Figure 6A) were chosen to give the F_4 generation (Figure 6B), which allowed the quantification of the activity of *En/Spm* transposase on one single *I/dSpm* element insert. The itf for this generation was 29.2% (7 in 24), significantly higher than for previous generations. The activity of the TEn2 T-DNA locus was not extensively studied, but in progenies of plant 1A-H12, this transposase source was found to be a factor 3 to 5 less capable of activating *I/dSpm* elements than the TEn5 T-DNA locus.

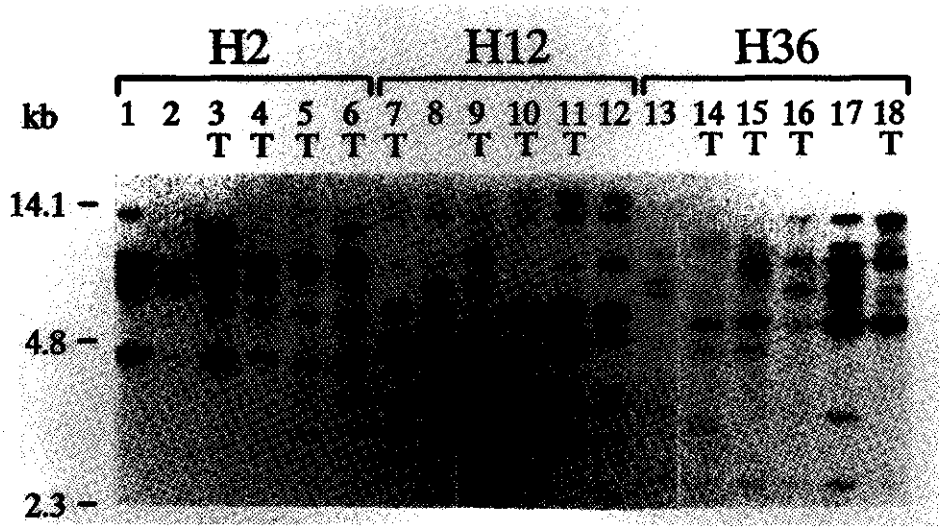

Figure 4: DNA blot of T_3 plants from three different 1A T_2 plants demonstrating reinsertions of excised *I/dSpm* elements.
DNAs were digested with both EcoRI and BglII. The blot was hybridized with the *I/dSpm* element specific probe (see Figure 1). The three progenies shown descended from plants 1A-H2 (lanes 1 to 6), 1A-H12 (lanes 7 to 12) and 1A-H36 (lanes 13 to 18). Presence of *En/Spm* transposase genes is depicted by T above every lane.

Timing of excision

Early transposition in the parental plant at or before the bolting stage, generally gives a large sector of the plant containing the insert, which will be inherited by a large number of progeny plants if present in the germline. Later transpositions will give accordingly smaller sectors and thus less progeny containing the insert, while very late transpositions in single flowers or even gametes, will give progeny containing unique inserts. The F_3 progeny shown in figure 6A demonstrated timing of excision when only two elements were involved. Most of the fertile revertant sectors found on the parental F_2 plant were small (only one or a few flowers) and thus arose after late excision. The progeny of the four revertant sectors did not have any inserts in common, apart from the two *I/dSpm* element inserts already present in

the parental plant (lane 1, Figure 6A). Apparently early transposition had not occurred in the F_2 plant or it was not stable enough to be transmitted to the germline. Within every revertant sector, progeny plants had one or more new inserts in common (late transposition) as well as unique inserts (very late transposition). Late transpositions were not restricted to the germline of a plant, as shown by the smear of hybridization visible in some lanes (Figure 6A) caused by late somatic transposition in leaf tissue of the F_3 progeny plant from which DNA was isolated. The instability of the *ant* mutant (chapter 5), which gives very small, single cell, as well as large multicellular somatic excision sectors adds to the property of continuous transposition.

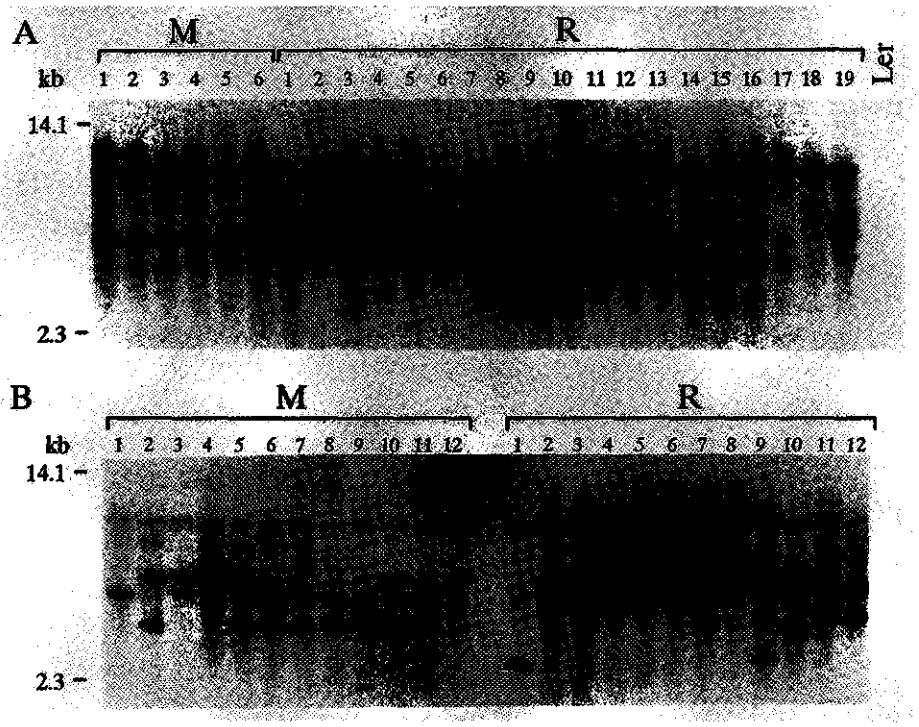

Figure 5: DNA blots of progeny from a revertant (R) and a male sterile mutant part of male sterile variegated plant ms2-11, demonstrating several unique transposed *I/dSpm* element inserts.
A. DNA from the (selfed) T_5 progeny, digested with HindIII
B. DNA from the F_1 progeny made by crossing with Landsberg *erecta*, digested with HindIII.
Both blots were hybridized with the *I/dSpm* element specific probe (see Figure 1). Unique *I/dSpm* element inserts which were used for determining the itf are indicated (o).

The "burst" of late transposition which occurred in the R4 sector (Figure 6A) is remarkable. The two original *I/dSpm* element inserts in the parental plant amplified to about 10 inserts in four of the five R4 progeny plants shown. Cross-fertilization as the source of these multiple inserts was ruled out as the plant was kept enclosed and as the progeny segregated 3:1 for fertility to male sterility as well as for hygromycin resistance to sensitivity. This

transposition burst happened most likely in only a part of the revertant sector, as one of the progeny plants contained only the two parental *I/dSpm* element inserts. Similar bursts of transposition were observed in other progenies (data not shown), indicating that such events are not uncommon.

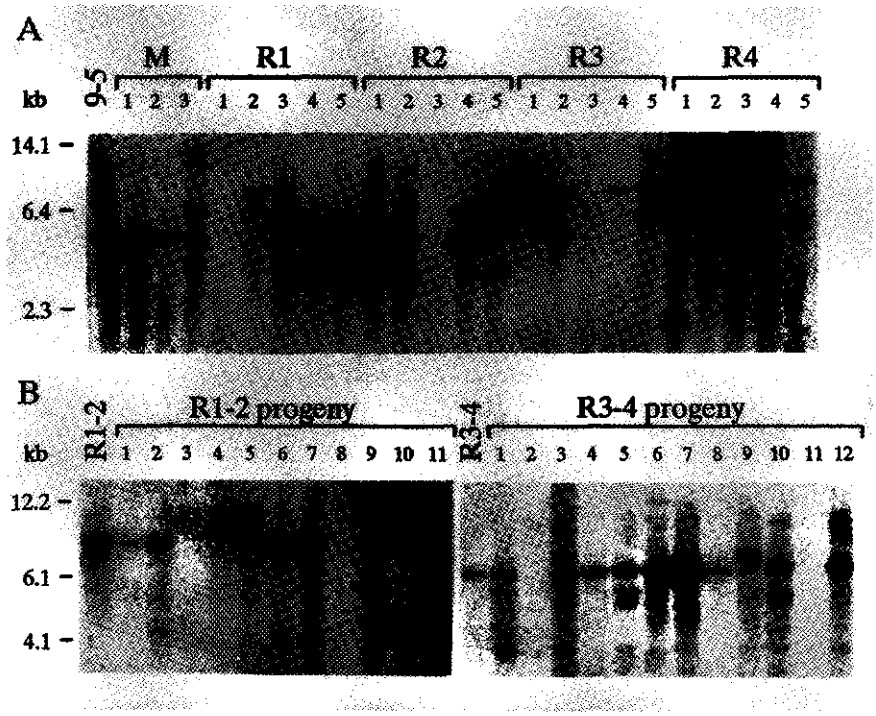

Figure 6: DNA blot progenies from plants with two, respectively one *I/dSpm* element insert.
A. F_3 progeny originating from four fertile revertant sectors (R1 to R4) and the remaining male sterile (M) rest of F_2 plant 9-5 (see Figure 3). DNA was digested with HindIII and hybridized with the *I/dSpm* element specific probe (see Figure 1). This probe detects several *I/dSpm* elements present in only one of each sector progeny. A transposition burst caused amplification of the transposon copy number in progeny of sector R4.
B. F_4 progeny originated from F_3 plants R1-2 and R3-4, which contained only one *I/dSpm* element insert, showing both loss of this element by Mendelian segregation or reinsertion failure (plants R1-2-3, 4, 8 and R3-4-2 and 11), as well as amplification of the number of elements.

DNA analysis of excision and insertion sites

To examine whether the activity of this *En/Spm-I/dSpm* system could be attributed to a change in the transposition mechanism when compared to its original host maize (reflected in altered excision and insertion footprints), some excision and insertion sites were analysed.

Two T-DNA empty donor sites from one plant were cloned after PCR and sequenced. In one footprint five nucleotides were deleted, thus restoring the reading frame of the NPTII gene. This was in accordance with the kanamycin resistant phenotype of the plant from which the excision site was cloned. The other empty donor site was the result of an excision leaving

the 3 bp target site duplication behind but altering one basepair.
To characterize reinsertion sites, DNA flanking 17 *I/dSpm* elements was cloned and sequenced. The insertion of 12 *I/dSpm* elements generated a three base pair target site duplication characteristic for the *En/Spm* transposon family (data not shown, but sequences available on request). In four cases the target site duplication had a one basepair mismatch, while one insertion revealed no target site duplication. The deviations at the target sites are significant, as other regions of the *I/dSpm* elements sequenced in the same reactions showed random deviations of about only 0.5%. None of the sequenced DNAs flanking *I/dSpm* elements showed any homology to sequences present in the databases.

Genetic mapping of I/dSpm elements

Most *I/dSpm* element inserts from which flanking DNA was isolated detected single copy loci when used as RFLP probe on ecotypes Landsberg *erecta* or Columbia DNA. In about half the cases they did not detect a RFLP and could not be positioned on a genetic map. Seven *I/dSpm* elements (out of 12 cloned *I/dSpm* elements giving major bands in figure 6A) were mapped using the RIL population generated by Lister and Dean (1993) (Figure 7), as well as four additional chimaeric somatic insertions (out of seven cloned) not visible by DNA blot analysis. *I/dSpm* elements revealing major bands on DNA blots were mostly found to be transmitted to the next generation, whereas somatic insertions were not. Of the two original *I/dSpm* elements giving the mapped inserts, only the *I/dSpm* element inserted at the *ms2* locus could be mapped. Physical or genetic linkage was found for *I/dSpm18* and *I/dSpm51*, both coming from the R1 sector, and for *I/dSpm41G* and *I/dSpm27* (mapped after YAC library hybridization (Schmidt et al., 1995)) coming from the same R2 progeny plant. *I/dSpm6* and *I/dSpm60*, though clustered, are coming from different sectors M and R4.

Two examples of tight clustering of *I/dSpm* elements were found. In a full sibling of the original *ms2* mutant, the *I/dSpm* element had excised from the *ms2::I/dSpm* allele and a new insert was found which was inseparable from the male sterile phenotype in a segregating population of 44 plants (data not shown). Transposition to a very close site was found in one progeny plant of the tagged *ms2* mutant (Aarts et al., 1993), where a new insertion within the 4.4-kb HindIII fragment containing the *MS2* gene was detected by DNA blot analysis (data not shown).

Discussion

The isolation of an *I/dSpm* element tagged male sterile mutant in progeny of a particular *Arabidopsis thaliana* transformant, demonstrated that the '*En/Spm-I/dSpm* in cis two element system' can be used for heterologous gene tagging (Aarts et al., 1993; see chapter 2). The system consists of two *En/Spm* transposase lines (TEn2 and TEn5) and many lines with several transposed *I/dSpm* elements either with or without *En/Spm* transposase. In this chapter different aspects of the transposition behaviour of this system have been described,

which will be beneficial for those who are planning to use this system for random or targeted gene isolation.

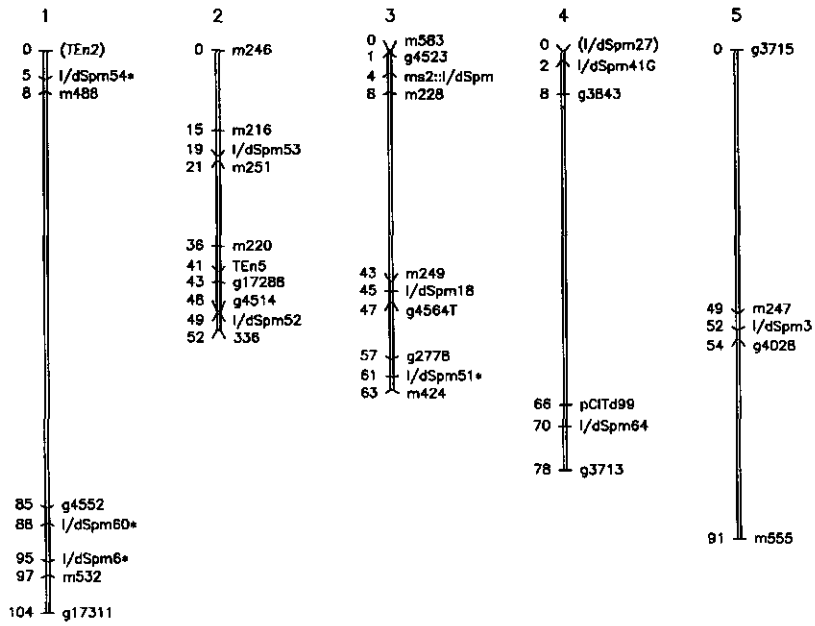

Figure 7: Genetic map of *Arabidopsis* showing the positions of transposed *I/dSpm* elements (I/dSpm) and flanking markers. *I/dSpm* elements with * were not visible as RFLP fragment after DNA blot analysis and are thought to be not transmitted to progeny. TEn2 and TEn5 are transposase T-DNA loci. Markers between brackets were not mapped by RFLP mapping. TEn2 was mapped with classical markers and *I/dSpm27* was mapped on the physical map after YAC hybridization (Schmidt et al., 1995).

The system differs in a few characteristics with other transposon systems used in Arabidopsis: 1) The *En/Spm* transposase genes and the *I/dSpm* element are located on one T-DNA. Thus only one transformant is needed and new positions of *I/dSpm* elements are generated by self-pollination. Presence of *En/Spm* transposase genes can be selected for by germinating seeds on hygromycin containing medium (20 mg/l). 2) The *I/dSpm* element is cloned in the open reading frame instead of in the untranslated leader of an NPTII excision marker gene to abolish background activity. Excision from this position leads to kanamycin resistance in about a third of the events, which is only visible as a phenotypic assay in the T_2 generation due to the high frequency of excision. 3) Both *En/Spm* transposase genes *tnpA* and *tnpD* are expressed from a stable *En-1* element under control of one strong CaMV 35S promoter.

The transposition behaviour of this system in Arabidopsis showed that *I/dSpm* elements can be activated to transpose in the presence of the *En/Spm* transposase source. Correct expression of TNPA and TNPD proteins from wild-type maize *En-1* in Arabidopsis was already shown by Cardon et al. (1993a). Apparently Arabidopsis is able to perform the complex differential splicing which is needed to obtain both the *tnpA* and the *tnpD* mRNAs in the right relative amounts, even if the endogenous *En-1* promoter is replaced by the CaMV

35S promoter. A non-autonomous *I/dSpm* element can be activated in the presence of *35S-tnpA* and *35S-tnpD* (Cardon et al., 1993a), and the stable *En/Spm* described here adds to the list of *En/Spm* transposase sources that can be used to activate *I/dSpm* elements.

The transposition frequency was one of the aspects of transposition we studied. The independent transposition frequency described here is another way of examining the potential use of a transposon system for gene tagging than by determining a germinal excision frequency by means of a phenotypic assay (Schmidt and Willmitzer, 1989). The itf was found to be roughly around 10% for populations which carried several *I/dSpm* elements, but could increase to 29.2% for progenies of plants containing only one *I/dSpm* element.

The two *En/Spm* transposase loci TEn2 and TEn5 differ in the frequency of *I/dSpm* element transposition (in progeny as well as somatically), with the TEn5 locus being the most active. Autoregulation of transcription by binding of the TNPA protein to repetitive motifs in the 5' end of the element as described by Fedoroff (1989) and Raina et al. (1993) is abolished, as most of the 5' end of the *En/Spm* transposase element has been removed. The difference in transposition rate between TEn2 and TEn5 might therefore be caused by a positive gene dosage effect, but a position effect of genomic sequences flanking the T-DNAs on the level of expression cannot be ruled out.

The activity of the transposase source was studied over many generations and appeared not to be limited to a few generations. On plant level late and often independent transpositions events were found. The timing of transposition affects the number of independent insertions that are found in the following generation. Our data show that transpositions occur throughout plant development as late as in single gametes, which is in accordance with the activity of the CaMV 35S promoter. In tobacco this promoter is active throughout plant development, starting several days after pollination (Benfey and Chua, 1990; Scofield et al., 1993). The continuous activity for this *En/Spm-I/dSpm* system differs from 35S-transposase based *Ac-Ds* transposon systems in Arabidopsis or tobacco. Scofield et al. (1993) proposed that *Ac*-transposase expression in tobacco can exceed an inhibitory level prevalently leading to early transpositions, which might explain the few unique *Ds* element inserts found in progeny of Arabidopsis plants containing a 35S *Ac-Ds* system (Long et al., 1993b). Inhibition of late transposition does not occur in this *En/Spm-I/dSpm* system, and as the autoregulation system is not functional any more, timing of transposition follows the timing of activity of the 35S promoter.

The strong amplification of the number of transposable elements in some progeny plants ("transposition bursts"; Figure 6A), is a phenomenon also known to occur with *Ac* in maize and tomato (Osborne et al., 1991) where it results from transposition of an element from a replicated to an unreplicated part of a DNA molecule during mitosis (Greenblatt and Brink, 1961). *En/Spm* exhibits this transposition related phenomenon in maize less frequently than *Ac*. Dash and Peterson (1994) showed that *En* preferably transposed from replicated to replicated DNA. In Arabidopsis, despite the occasional occurrence of transposition bursts, the number of *I/dSpm* elements stayed relatively stable over at least three subsequent generations without selection (T_3 to T_5) suggesting a maintenance of the number of transposons as seen in maize.

To characterize the transpositional mechanism, we analysed excision and insertion sites of *I/dSpm* elements. All the excision footprints obtained after excision of the *I/dSpm* element from the NPTII::I T-DNA, as well as from the *ms2::I/dSpm* locus (Aarts et al., 1993 and unpublished results) showed that excisions occurred according to the transposition model proposed by Saedler and Nevers (1985). Of 17 sequenced DNAs flanking transposed *I/dSpm* elements, four had one basepair mismatches when comparing the three basepair target site duplications directly flanking the insert, while one insert was not flanked by a three basepair duplication. Such alterations were not observed among 12 *En/Spm* or *I/dSpm* element flanking sequences coming from transposition in maize, tobacco and potato (Cardon et al., 1991; Frey et al., 1989; Frey et al., 1990; Gierl et al., 1985; Menssen et al., 1990; Pereira and Saedler, 1989), but analysis on a larger number of insertion sequences from both Arabidopsis and maize is needed to conclude that there are differences in the *En/Spm-I/dSpm* transposition mechanism between these species.

The availability of a population of RILs (Lister and Dean, 1993) greatly facilitated positioning *I/dSpm* elements on a genetic map by using transposon-flanking DNA as RFLP probes. The lack of RFLPs however caused only a relatively low number of *I/dSpm* elements to be mapped, and although both clustered and unlinked transpositions were found, there was no indication of a strong preference for transposition to nearby loci. The observation that *ms2::I/dSpm* along with some of the other *I/dSpm* elements, seemed to be located at similar, distal parts of the chromosomes, suggests a tendency of *I/dSpm* elements to transpose to new chromosomal positions of similar constitution (i.e. nuclear architecture or replication time). *I/dSpm* element transposition in Arabidopsis resembles *En* transposition in maize, which was found to transpose to linked sites in about 25% of the excision events (Peterson, 1970). For tobacco, linked transposition was found after 44% of *dSpm* transpositions (Cardon et al., 1993b). Transposition to linked and unlinked sites has also been shown for *Ac* or *Ds* in tomato (Belzile and Yoder, 1992; Healy et al., 1993, Knapp et al., 1994; Osborne et al., 1991). A both random and linked distribution of transposed *I/dSpm* elements has a practical consequence for gene tagging. By outcrossing to wild type and subsequent self-fertilization, the resulting F_2 can contain plants with one or a few stable *I/dSpm* element inserts. By segregating the mutation away from the transposase source, unstable mutants can be stabilized and as the number of *I/dSpm* elements will be reduced, DNA flanking the mutating *I/dSpm* element insert can be isolated easier.

One of the major aims of developing a heterologous transposon system for *Arabidopsis* is to apply transposon mutagenesis. "Targeted tagging" can be facilitated by using male sterile plants containing both *I/dSpm* elements and *En/Spm* transposase, that are crossed (on a large scale) with the target mutant. By determining the itf it is possible to estimate the minimal size of a population that has to be screened to find one plant having a transposon insertion in a target gene (assuming a similar mutation frequency for every gene). Based on the assumption that transposon carrying plants are used with about 10 randomly distributed *I/dSpm* elements and an itf of 15%, the number of F_1 plants to be screened will be around 10^5, comparable with T-DNA mutagenesis (Feldmann, 1991). This figure accounts for

tagging a gene when no *I/dSpm* elements are in close linkage (within 5 cM). To target a gene of unknown location, a strategy with *En/Spm*-plants, which have 10-20 randomly distributed *I/dSpm* elements, that will be crossed to the target mutant, will give good chances for success. From experiments with *En/Spm* in maize it is known that the mutation frequency can increase 100-fold when the transposon is tightly linked to the target locus (Cone, 1994). In Arabidopsis it can be expected that the mutation frequency will similarly increase when an *I/dSpm* element is genetically linked to the target locus. If the target gene is mapped, it is worth the effort to get a transposon closely linked to the target locus before making F_1 populations for screening. The mapped *I/dSpm* elements shown (Figure 7) will enable targeted tagging of closely located genes. The *En/Spm-I/dSpm* system described here showed abundant as well as late transpositions and may be an attractive alternative to *Ac-Ds* based transposon systems for targeted gene tagging.

Material and methods

Binary vector plasmid cwEnN::I

A chimaeric neomycin phosphotransferase II (NPTII) gene was inserted into binary vector pGDW3.1 (Wing et al., 1989), containing a chimaeric nopaline synthase (*nos*) promoter-hygromycin phosphotransferase (HPT) gene for selection during transformation. The NPTII gene was modified from Baker et al. (1987), by introduction of a ClaI linker to generate a new translation start (GCGATGG) 5' to a unique BamHI site present in the untranslated leader between the T_R1' promoter and the NPTII codogenic region. Insertion of a 2.2-kb *I/dSpm* element with flanking DNA (*I-6078*; Pereira and Saedler, 1989) in the BamHI site and exchanging the original downstream NPTII sequences with sequences downstream of the BamHI site in plasmid pBCK1 (after Kaulen et al., 1986), altered the original N-terminus. Upon excision (e.g. with a two bp deletion) the correct reading frame would be MDPACPKTLEVDPAGSG after which it follows the normal NPTII reading frame from lysine (pos. 12) onwards. A 7.6-kb BssHII Klenow treated fragment from *En-1* (Pereira et al., 1986), was cloned between the Cauliflower Mosaic Virus (CaMV) 35S promoter-terminator cassette, originating from pDH51 (Pietrzak et al., 1986) in pBR322. This wings-clipped *En-1* (cw*En*) under control of the 35S promoter was cloned into pBHNI to produce plasmid cwEnN::I (Figure 1), which was mobilized into the *Agrobacterium* strain pGV3101(pMP90RK) (Koncz and Schell, 1986).

Plant material

For plant transformation, root explants of ecotype Landsberg *erecta* were infected with *Agrobacterium* containing cwEnN::I, according to Valvekens et al. (1988). Transformed calli were selected on medium containing 20 mg/l hygromycin. Emerging T_1 shoots were grown without selection and allowed to self-fertilize and set seed *in vitro*. Growth conditions were 12 to 16 hrs light at 22°C in a growth chamber.

Seeds of T_2 and subsequent generations were sterilized if needed by treatment with ethanol 70% (2 min), 50% commercial bleach (5% sodium hypochlorite, 5 min) and five subsequent washes with sterile water, all in Eppendorf tubes. For selection on antibiotic resistance, the seeds were sown on GM medium (Valvekens et al., 1988) or half strength MS medium (Murashige and Skoog, 1962) plus the required antibiotic (kanamycin 100 mg/l or hygromycin 20 mg/l), solidified with 0.8% purified agar. Using no sugar in the medium permitted sowing unsterilized seeds on hygromycin containing medium. Segregation for antibiotic resistance was scored 5 to 10 days after germination. When appropriate, seedlings were transferred to soil (compost : vermiculite : sand = 4 : 1 : 1) and grown in a climate controlled greenhouse at 20-22°C with additional light (16 hrs.). To prevent cross-fertilization and dispersal of seeds, plants were grown in Aracon containers (Beta Developments, Gent, Belgium). For the phenotypic excision assay seeds were only germinated on GM medium containing 100 mg/l kanamycin. Resistance, variegation or sensitivity was scored 5 to 10 days after germination. Variegated and sensitive seedlings were transferred to GM without antibiotic, before transferring to soil.

Ploidy number was determined by counting chloroplasts in stomatal guard cells of the lower epidermis of young leaves as described by Detrez et al. (1989). At least 25 stomata of different T_2 and T_3 plants were counted.

DNA analysis

Genomic DNA was isolated according to Dellaporta et al. (1983) or Shure et al. (1983). PCR analysis to amplify DNA of empty donor sites of the cwEnN::I T-DNA was performed with 50 to 200 ng of genomic DNA. The primer sequences are 5' CTT-ACG-TCA-CGT-CTT-GCG-CA 3' (*1'* promoter) and 5' CCA-GTC-ATA-GCC-GAA-TAG-CC 3' (*NPTII* gene). For normal PCR 30 cycles of denaturation (94°C, 30 sec), annealing (55-60°C, 1 min) and extension (72°C, 2 min) were carried out. PCR products of interest were cloned in Bluescript SK^+ as a blunt ended fragment after Klenow treatment and gel elution. Inverse PCR was performed on 200 ng HinfI digested genomic DNA, Klenow treated and self-ligated. The self-ligated DNA molecules were linearized by digestion with SalI. Two rounds of 25 PCR cycles (annealing 55°C, extension 3 min 72°C) were applied, using the nested primers 103 to 106 described by Masson et al. (1991). 2-5 μl of the first round was used for the second round of PCR. Cloning of IPCR fragments was as for PCR fragments. To determine the DNA sequence, a normal PCR was performed on most cloned IPCR product using one primer (5' GAC-ACT-CCT-TAG-ATC-TTT-TCT-TGT-AGT-G 3') annealing to both left and right TIR, after which the PCR product was cloned as described previously. Double stranded supercoiled plasmid was used for determining the DNA sequence on an automated DNA sequencer (Applied Biosystems). One strand and if needed both strands were sequenced.

For DNA blot analysis 200 to 400 ng DNA was used per lane, separated on a 0.8% agarose gel in Tris-acetate running buffer and, after electrophoresis, alkali-denatured and transferred onto Gene Screen Plus or Hybond N^+ nylon membrane by vacuum blotting. The blots were prehybridized and hybridized following the procedure recommended by the

membrane manufacturer. [^{32}P] random prime labelled DNA fragments were used as probes (Figure 1). A 0.27-kb fragment containing the left border of *En-1* (up to a SalI restriction site) was used to detect *I/dSpm* elements. The T-DNA right border probe was a 1.1-kb fragment containing the HPT gene. A fragment of 1.8-kb containing the complete NPTII gene plus octopine synthase (*ocs*) terminator was used to detect excision fragments (see figure 1 for probe locations on the T-DNA). After hybridization the membranes were washed with 2xSSC, 1% SDS at 65°C for 2x30 minutes and autoradiographed to X-ray films (Fuji or Kodak) at 80°C using intensifying screens.

Mapping T-DNA and I/dSpm elements

For mapping T-DNA and *I/dSpm* elements, genomic DNA flanking the inserts was used as RFLP probe on a population of Recombinant Inbred Lines (RILs) made by Lister and Dean (1993). Routinely five restriction enzymes were used (BglII, DraI, EcoRI, EcoRV and HindIII) to test for RFLPs between the ecotypes Landsberg *erecta* and Columbia. Probes giving a RFLP were hybridized to a blot containing the DNAs of 76 RILs digested with the appropriate enzyme. Recombination frequencies and map distances were calculated with the JoinMap program designed by Stam (1993), using mapping data from RFLP probes provided by Clare Lister and Caroline Dean (Cambridge Lab, John Innes Institute, Norwich, UK).

Acknowledgements

We thank Maarten Koornneef for helpful suggestions on handling Arabidopsis, for his aid and advice in mapping transposons and for critically reading the manuscript. Renate Schmidt was very helpful in mapping one of the *I/dSpm* elements by YAC hybridization. Clare Lister and Caroline Dean were so kind to distribute seeds through the Nottingham Arabidopsis Stock Centre and mapping data of their set of recombinant inbred lines prior to publication. DNA isolation of the RILs was done in cooperation with Ton Peeters, Ben Scheres, Pieter Vos and Casper Vroemen. Prof. Heinz Saedler is acknowledged for the availability of the *En-1* plasmid clones originally developed at the Max-Planck-Institute für Züchtungsforschung (Cologne). Gerrit Stunnenberg cared for most of the plants in the greenhouse, for which we are grateful. This work was supported by a grant (BIOT-CT90-0207) from the European Community under the BRIDGE T-project on 'Development and use of *Arabidopsis thaliana* as a tool for isolating genes of agronomic importance.'

Chapter 2

Transposon tagging of a male sterility gene in Arabidopsis

Mark G.M. Aarts, Wim G. Dirkse, Willem J. Stiekema and Andy Pereira

(Published in *Nature* (1993) **363**, 715-717)

Summary

Transformation of the well studied maize transposable elements into other plant species is expected to enable the employment of transposon tagging methodology for the isolation of interesting genes in the heterologous host. We fulfill this by describing here the isolation of a transposon tagged male sterile mutant in *Arabidopsis thaliana* using the maize *Enhancer/Suppressor-mutator* (*En/Spm*) - *Inhibitor/defectiveSuppressor-mutator* (*I/dSpm*) transposable element system introduced into Arabidopsis. The mutant lacks pollen, preventing normal self-fertilization, a characteristic important for production of hybrid seed in many crop plants. We identified an *Enhancer* transposase mediated *I/dSpm* element insertion responsible for the male sterile phenotype, and isolated the corresponding gene named *MALE STERILITY 2*. Critical evidence that the *I/dSpm* element containing gene is involved in the male sterile phenotype, is provided by the DNA sequences of new excision derived alleles from independent stable fertile and male sterile progeny of the original mutant.

Introduction

The maize transposable element system consisting of the autonomous *Enhancer/Suppressor-mutator* (*En/Spm*) (McClintock, 1954; Pereira et al., 1985; Peterson, 1953) and non-autonomous *Inhibitor/defective Suppressor-Mutator* (*I/dSpm*) elements has been used for transposon tagging of a number of genes in maize (reviewed by Walbot et al., 1992) and was shown to transpose when introduced by transformation into heterologous plant species (Baker et al., 1986; Frey et al., 1989; Masson and Fedoroff, 1989; Pereira and Saedler, 1989). We constructed a transformation vector containing 3 units, namely: 1) a non-mobile *En* transposase source (Masson and Fedoroff, 1989), under control of the strong Cauliflower Mosaic Virus 35S promoter (CaMV 35S); 2) a mobile *I/dSpm* element (Schwarz-Sommer et al., 1985a) as insertion mutagen; 3) a hygromycin phosphotransferase gene (Van den Elzen et al., 1985) conferring resistance to the antibiotic hygromycin. When this 'in cis two-element *En/Spm-I/dSpm* system' was introduced into *Arabidopsis thaliana, I/dSpm* elements started to transpose and inserted at different positions of the genome. Lines containing many transposed *I/dSpm* elements were screened for mutants. One of the mutants found was showing a male sterile phenotype. This chapter describes the isolation of the mutant and the corresponding *MALE STERILITY 2* gene.

Results and discussion

A male sterile plant was found among ten unselected third generation (T_3) progeny of a self-fertilized primary transformant (T_1) with frequently transposing *I/dSpm* elements. The plant was coded PA for Pollen Absent. As the male sterile mutation can be complemented by *ms1* (data not shown), the only other male sterile mutant without pleiotropic effects described in Arabidopsis (Van der Veen and Wirtz, 1968), we called this novel mutated gene *ms2*.

Some seeds (52) were produced after self-fertilization, indicating that the mutant PA is leaky. Among the PA progeny plants grown from this seed (PA1-PA52), 47 displayed a male sterile phenotype proving they were *bona fide* self-fertilized progeny. Blot hybridization analysis with an *I/dSpm* element specific probe indicated that also the five fertile plants arose from self-fertilization as they contained at least one homozygous *I/dSpm* element hybridizing fragment (data not shown). Several sterile plants had a few fertile flowers and one plant (PA11) had a completely fertile side branch (Figure 1A). This observation of fertile sectors in otherwise sterile plants suggested that the mutation was due to insertion of a transposable element and that subsequent excisions restored the activity of the gene involved in male fertility.

Genetic analysis, showing a 3 : 1 segregation of fertile : male sterile plants in progeny of self-fertilized stable heterozygous *MS2,ms2::I/dSpm* plants, and the recovery of homozygous male steriles, indicate that the mutant acts sporophytically, rather than gametophytically. In cross-sections of full grown male sterile anthers no pollen could be detected (Figure 1B and C). In younger anthers the tapetal layer appears affected, leading to pollen abortion shortly after release from tetrads. Careful examination of a close-contained mutant plant without

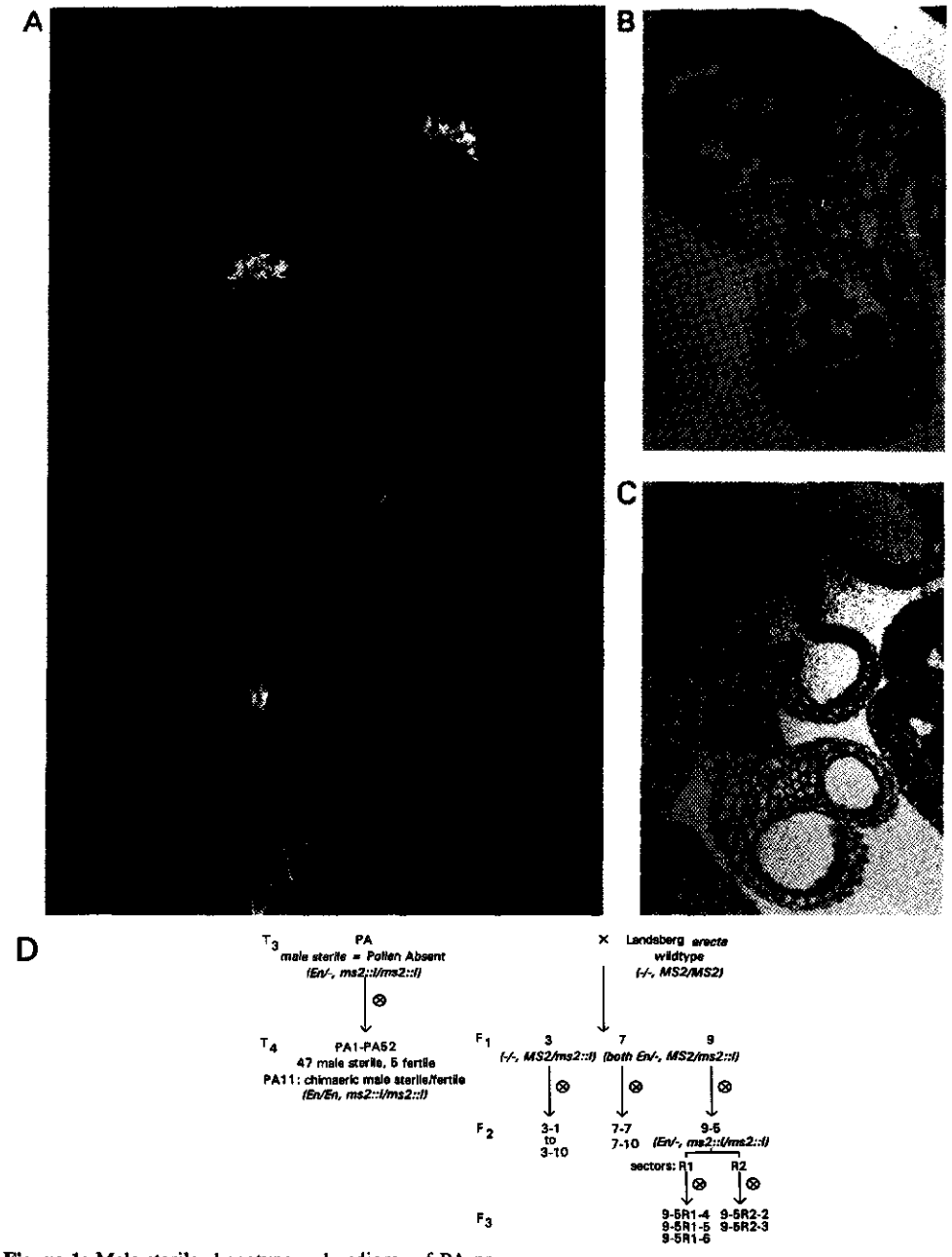

Figure 1: Male sterile phenotype and pedigree of PA progeny.
A. The phenotype of a genetically unstable male sterile mutant leading to a chimaeric male sterile/fertile phenotype. As Arabidopsis is a strictly self-fertilizing species, male sterility is characterized by short, empty siliques due to absence of self-pollination (right branch). In one side branch (left) of this otherwise male sterile plant, early excision of an *I/dSpm* element from the mutated gene has resulted in a reversion from mutant male sterile to wild-type male fertile phenotype (*ms2::I/dSpm* reverts to *MS2*). Fertile flowers of the revertant sector result in thick, elongated and seed filled siliques. The plant shown (PA11) is unique among progeny of PA in displaying such a large revertant sector. In other plants reversion was restricted to single flowers, phenotypically visible as an occasional completely fertilized silique on a male sterile plant.
B. Cross-section of a wild-type anther shortly before pollen release.
C. Cross-section of a male sterile anther of similar stage as in *B*.
D. Pedigree of PA progeny used for subsequent molecular genetic analysis. For the genotypes, first the presence (En) or absence (-) of *En*-transposase is given, followed by the *MS2* genotype. R1 and R2 are the revertant sectors of plant F_2 9-5.

En/Spm transposase, revealed that flowers developing late in plant growth yield a few seeds, which explained the leaky phenotype. This phenomenon has been reported before for male sterile Arabidopsis (Van der Veen and Wirtz, 1968).

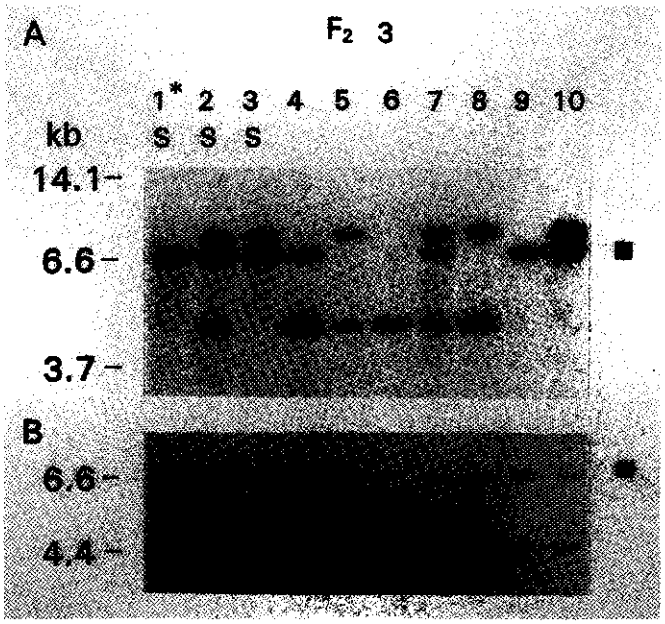

Figure 2: Segregation analysis of a specific *I/dSpm* element with the male sterile phenotype.
A. Segregating F_2 progeny lacking *En/Spm* transposase from the cross PA x Landsberg *erecta* (see figure 1*D*) was used for hybridization with an *I/dSpm* element specific probe. DNA was digested with HindIII, which does not cut inside the *I/dSpm* element. The F_2 progeny lacks *En/Spm* transposase genes, therefore all detected fragments denote stable *I/dSpm* element inserts. A 6.6-kb *I/dSpm* element containing fragment (■) appears homozygous only in plants displaying a male sterile (S) phenotype. Plant F_2 3-1 (*) has only the *ms2* accompanying *I/dSpm* element and was therefore used for the isolation of the transposon flanking DNA, ultimately yielding a probe for the *MS2* gene.
B. Homozygosity of plant F_2 3-1 for the *ms2::I/dSpom* allele was confirmed by rehybridizing the blot with a *MS2* specific probe. In total 4 different F_2 families (43 plants) lacking *En/Spm* transposase were analysed (data not shown). Every male sterile plant showed a homozygous *ms2::I/dSpm* allele.

To identify the *I/dSpm* element insertion responsible for the male sterile phenotype, DNA blot analysis was performed on progeny of PA and on other families originating from the same primary transformant, but not segregating for the mutant phenotype. This revealed the presence of a unique *I/dSpm* homologous band in the PA progeny (data not shown). As PA was hemizygous for *En/Spm* transposase, an outcross to an untransformed plant resulted in F_1 plants (Figure 1D) half of which lacked *En/Spm* transposase, as seen by hygromycin sensitivity. All 10 F_1 plants tested revealed the specific *I/dSpm* homologous band, suggesting homozygosity for this *I/dSpm* element in the parent. From 4 F_1 plants without transposase the F_2 progeny, segregating 3:1 for fertility to male sterility, revealed the same specific

I/dSpm homologous band, homozygous in all plants showing the male sterile phenotype. Male sterile plant F_2 3-1 contained a single homozygous *I/dSpm* element as shown by the unique DNA fragment hybridizing to the *I/dSpm* probe (Figure 2A). DNA of this plant was used to isolate the transposon flanking DNA by the Inverse Polymerase Chain Reaction (IPCR) (Masson and Fedoroff, 1989; Ochman et al., 1988). The IPCR fragment was cloned, sequenced and used as a probe to isolate clones from a flower specific cDNA library (Weigel et al., 1992). DNA sequence information revealed that the *I/dSpm* element was inserted in the 3' end of the codogenic region of the cloned *MALE STERILITY 2* (*MS2*) gene. Rehybridization of the F_2 progeny blot with an *MS2* gene probe, confirmed that the specific *I/dSpm* element was homozygous in all male sterile plants (Figure 2B).

We identified plants hemizygous for *En/Spm* transposase (*En/-* in figure 1D) which displayed chimaerism for male sterility versus fertility, either in somatic sectors or in progeny. The excision products from the selfed progeny which lacked transposase, were isolated and sequenced (Table 1). In all male sterile plants either the inserted *I/dSpm* element is still present, or excision has led to frameshift mutations in the reading frame. The fertile plants contain a "revertant" allele which is either wild-type or an "in frame" excision footprint restoring the proper reading frame. In maize 1 out of 10 excision events leads to an "in frame" excision without leaving additional nucleotides (Schwarz-Sommer et al., 1985b), yet these infrequent events are present here in every plant with a revertant phenotype (which is dominant). For every plant the excision derivative alleles, defined by their nucleotide sequence, can explain the phenotype observed. These independently derived "new" alleles (minimal 6), which determine the male sterile or fertile phenotype, demonstrate conclusively that this *I/dSpm* element tagged gene is responsible for the male sterile mutation. The *MS2* cDNA sequence (Figure 3) reveals a short region of 77% identity to a DNA sequence located about 2 kb upstream of the wheat mitochondria 26S ribosomal RNA *rrn26* gene at positions 650-814 of the published sequence (Spencer et al., 1992). No function has been ascribed to this mitochondrial DNA segment which contains an open reading frame with 81.7% identity to the *MS2* protein (55 amino acids). The homology in the mitochondrial DNA segment is precisely flanked by intron-exon boundaries. The relevance of mitochondria and respiration to pollen fertility has been well established (Hanson, 1991). Also rearrangements in the mitochondrial genome have been shown responsible for "cytoplasmic" male sterility. Thus homology associated with the *MS2* gene now provides a clue to its biochemical function.

These results establish tagging of a novel male sterility gene (*MS2*) by correlation of phenotype with genotype at the DNA sequence level and subscribe the further employment of transposon tagging with heterologous transposons. This is the first isolated gene determining a nuclear male sterile mutant (monogenic recessive) phenotype. Such sporogeneous male sterile mutants with impaired microgametogenesis have been described in many plants (Kaul, 1988). Further characterization of the *MS2* gene and homologous genes from other species will increase our knowledge on pollen development (Bedinger, 1992) and provide methods for the production of male sterile crop plants useful in hybrid seed production (Scott et al., 1991).

Table 1: Nucleotide sequences of the *ms2::I/dSpm* allele and *En/Spm* transposase mediated excision derivatives in relation to plant phenotypes.
The wild-type Landsberg *erecta* (Ler) trinucleotide target site sequence (AAA) is duplicated upon insertion of the *I/dSpm* element (*I*) (see F_2 3-1). Codons are separated by gaps. Nucleotides altered in excision derivative alleles compared to the wild-type sequence are underlined. Introduction of a stopcodon is denoted (*). The plants used are: F_3 plants from two self-fertilized fertile revertant flowers of F_2 plant 9-5 (F_3 9-5R); self-fertilized progeny of the original mutant (PA); and two F_2 plants (F_2 7-7, 7-10) derived from the cross PAxLer. All plants lack *En/Spm* transposase and contain stable excision derivative alleles that originated in the parental plant (genotypic notation as in figure 1D). Fertile plants with the wild-type sequence contain at least one other homozygous *I/dSpm* element ruling out cross-pollination.

DNA origin	allele sequences	genotype	phenotype
Landsberg *erecta* (wild type)	ACA AAC ACA AAC	*MS2,MS2*	fertile
F_2 3-1	ACA AA I A AAC ACA AA I A AAC	*ms2::I/dSpm,ms2::I/dSpm*	male sterile
F_3 9-5R1-4	ACA AAA C ACA AAA C	*ms2,ms2*	male sterile
F_3 9-5R1-5	ACA AAA C ACA AA I A AAC	*ms2,ms2::I/dSpm*	male sterile
F_3 9-5R1-6	ATA AAC ACA AAA C	*MS2,ms2*	fertile
F_3 9-5R2-2	TTA AAC ACA AA I A AAC	*MS2,ms2::I/dSpm*	fertile
F_3 9-5R2-3	TTA AAC ACA AAA C	*MS2,ms2*	fertile
PA6	ACA AAC ACA AA I A AAC	*MS2,ms2::I/dSpm*	fertile
PA13	ACA AAT TAA* C ACA AAT TAA* C	*ms2,ms2*	male sterile
PA45	ACA AAT TAA* AC ACA AA I A AAC	*ms2,ms2::I/dSpm*	male sterile
PA46	ACA AAC ACA AA I A AAC	*MS2,ms2::I/dSpm*	fertile
F_2 7-7	ACA AAT TTA AC ACA AAT TTA AC	*ms2,ms2*	male sterile
F_2 7-10	ACA AA I A AAC ACA AA I A AAC	*ms2::I/dSpm,ms2::I/dSpm*	male sterile

Material and methods

Identification and phenotypic analysis of the I/dSpm tagged ms2 mutant

A primary transformant (T_1) containing the '<u>in cis</u> two element *En/Spm-I/dSpm*' T-DNA construct (five T-DNA copies at one locus) was self-fertilized and its progeny selected for the presence of T-DNA by virtue of hygromycin resistance (see chapter 1). A male sterile plant (PA) was found among 10 non-selectively grown T_3 progeny (each containing 5-10

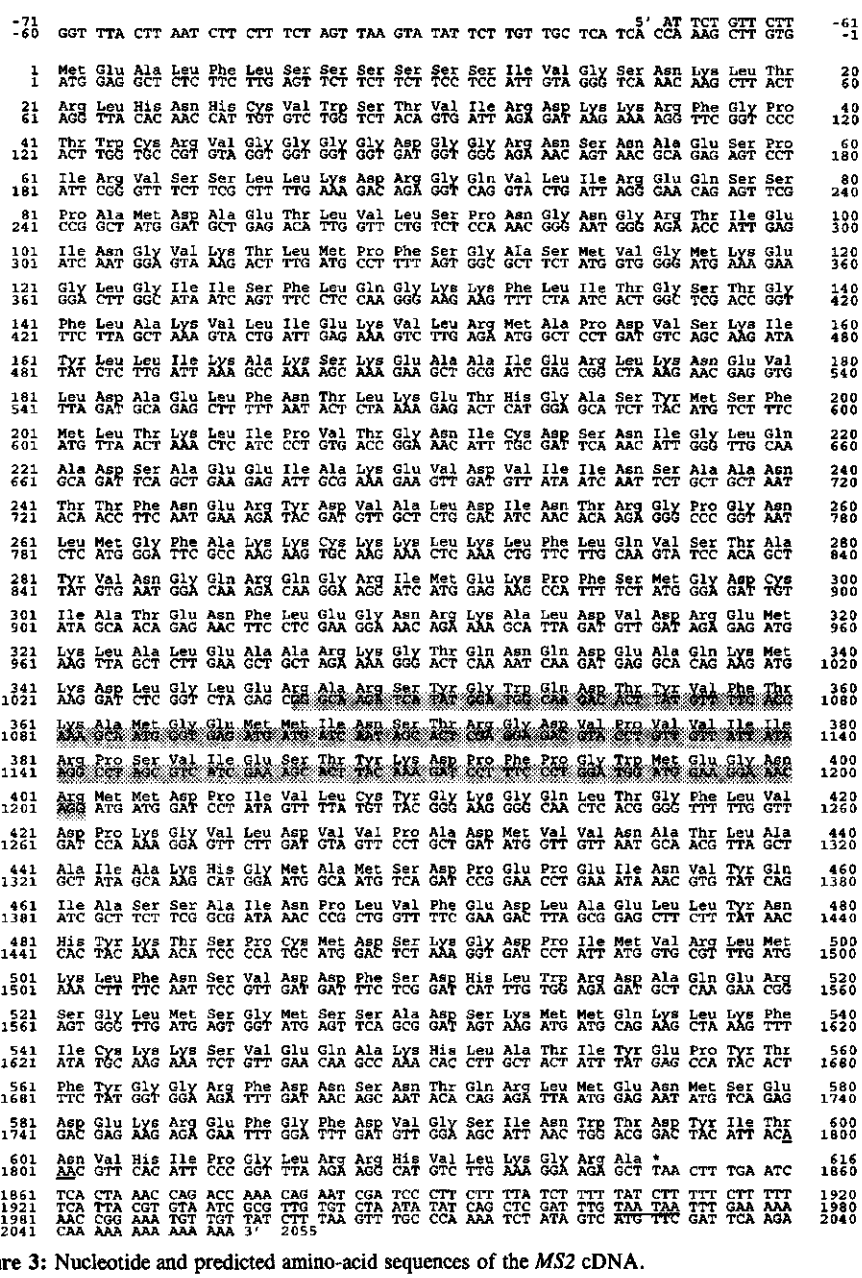

Figure 3: Nucleotide and predicted amino-acid sequences of the *MS2* cDNA.
Nucleotides are numbered beginning with the first position of the ATG initiation triplet. A nonsense codon upstream (-36) indicates that this is the translation start. The polyadenylation signal sequence (position 1967, underlined) precedes the poly(A) tail. The position of the *I/dSpm* element insertion (AAA) at position 1800 is underlined. The last three amino-acids (Gly-Arg-Ala) contain a putative C-terminal microbody targeting signal (CMTS) suggesting an essential role for this part of the protein, which is absent in male sterile mutant alleles due to excision footprints causing frameshifts (Table 1). The region of homology of the *MS2* cDNA to the wheat mitochondrial DNA is shaded (1040-1203).

transposed *I/dSpm* elements) of one of these T_2 plants. One inflorescence of PA was outcrossed to the untransformed ecotype Landsberg *erecta*. The rest of the plant was permitted to self-fertilize yielding only 52 seeds. PA11 was grown from one of these seeds.

For microscopic examination and comparison with Landsberg *erecta*, whole inflorescence tops were embedded in Technovit and sectioned. 5 mm thick sections were stained with Toluidine Blue and examined by light microscope.

DNA analysis

To establish linkage of an *I/dSpm* element to the *ms2* phenotype, F_2 progenies from the cross PA x Ler, lacking *En/Spm* transposase, were grown in the greenhouse. Their phenotypes were determined and DNA was isolated from each plant (Dellaporta et al., 1983). HindIII restricted DNA separated by 0.8% agarose gel electrophoresis was transferred to GeneScreen Plus nylon membranes following the manufacturer's protocol. The DNA blot was hybridized with a radiolabelled 0.27-kb SalI fragment specific to the *I/dSpm* element (see chapter 1). Homozygosity of the 6.6-kb *ms2::I/dSpm* containing fragment was determined by intensity comparison with other bands in the same lane and confirmed later by rehybridizing the blot with an *MS2* specific probe. Rehybridizing is the only way to conclusively determine homozygosity of the *I/dSpm* homologous fragment in plant F_2 3-1 (Figure 2B).

The cDNA clone (largest out of three sequenced) was isolated by screening an Arabidopsis ecotype Landsberg *erecta* flower specific cDNA library (frequency 1/12000). Both strands of the cDNA were sequenced. Subclones of specific restriction fragments and specific oligonucleotide primers were used to obtain the complete sequence. DNA sequence analysis was performed using the Genetics Computer Group (GCG) software package (Devereux et al., 1984).

(Inverse) PCR analysis

DNA flanking the *I/dSpm* elements was obtained by Inverse PCR (see chapter 1). Primers MS25 (5'-CAG-CAA-TAC-ACA-GAG-ATT-AAT-GGA-G-3') and MS24 (5'-GAC-ACA-ACG-CGA-TTA-CAC-GTA-ATG-3'), respectively 5' and 3' of the *I/dSpm* insertion site, were designed based on the sequence of the IPCR fragment from plant F_2 3-1 (*ms2::I/dSpm,ms2::I/dSpm*) and used to amplify DNA sequences from mutant *ms2* or revertant *MS2* alleles after excision of the *I/dSpm* element.

IPCR and PCR products were blunt end cloned in Bluescript SK^+ (Stratagene). Double stranded supercoiled plasmid was used for automated sequencing using the Taq Dye Deoxy Terminator Cycle Sequencing Kit (Applied Biosystems). Per plant at least two cloned products of a PCR were sequenced. If the first two sequences were identical, two cloned products of at least one additional PCR were sequenced. The presence of *ms2::I/dSpm* alleles was determined by hybridizing Southern blots with a *MS2* specific probe.

Acknowledgements

We thank Dr. J. Dangl from the EC Arabidopsis T-Project DNA center, Max Delbrück Laboratorium, Köln for distributing the flower cDNA library deposited by Drs. D. Weigel and E. Meyerowitz and the genomic library deposited by Dr. D. Coates. The Centrale Fotodienst TFDL-DLO is acknowledged for preparing the photographs. We thank Maarten Koornneef for critical reading of the manuscript. This work was supported by a grant (BIOT-CT90-0207) from the EC Arabidopsis BRIDGE Program.

Chapter 3

The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes

Mark G.M. Aarts, Rachel Hodge[1], Kriton Kalantidis[2], Dion Florack, Zoe A. Wilson[2], Bernard J. Mulligan[2], Willem J. Stiekema, Rod Scott[1] and Andy Pereira

[1]Department of Botany, University of Leicester, University Road, Leicester LE1 7RH, UK; [2]Department of Life Science, University of Nottingham, University Park, Nottingham NG7 2RD, UK

(submitted for publication)

Summary

The *MS2* gene from *Arabidopsis thaliana* is essential for proper pollen formation. The gene is expressed in the tapetum at the time of microspore release from tetrads. In tobacco the *MS2* promoter is active in the tapetum at a comparable stage. Genes with similarity to *MS2* are found in Arabidopsis and other species. The similarity of the MS2 protein to a wax fatty acid reductase from jojoba, and to reductases from other species, suggests that the MS2 protein acts as a reductase in an elongation/condensation biosynthesis pathway. A possible function of the MS2 protein in the formation of pollen wall substances is discussed.

Introduction

Male gametogenesis in plants is a complex developmental process that requires the temporally and spatially controlled activity of several thousand specific genes (Goldberg et al., 1993). The widely appreciated model plant *Arabidopsis thaliana* is an attractive species to use for studying such developmental processes, and there are a number of approaches available to do so.

The genetic approach involves the generation and characterization of mutants disturbed in male gametogenesis. A number of male sterile mutants are already known (Chaudhury, 1993; Chaudhury et al., 1994; Dawson et al., 1993; Mulligan et al., 1994; Regan and Moffat, 1990; Van der Veen and Wirtz, 1968). The phenotype of these sporophytic mutants varies from structural defects, such as altered stamen length or non-dehiscent anthers, to functional defects affecting the gametogenesis process itself and leading to absent or non-functional pollen.

Another approach involves the isolation of anther specific genes to identify the corresponding plant phenotype or protein function. The small anthers of Arabidopsis are not very attractive for mRNA isolations, but fortunately anther-specific genes isolated from "large anther" Brassica species like *Brassica napus* and *B. campestris* are often very similar to their Arabidopsis counterparts (Hird et al., 1993; Paul et al., 1992; Xu et al., 1995). With these intermediate sources, the corresponding Arabidopsis genes can be isolated and the advantages of Arabidopsis as a model species can be further exploited.

In this paper we show a combination of genetic and reverse genetic approaches for the further characterization of the *MALE STERILITY 2* gene from Arabidopsis and its homologue from *Brassica napus*. The Arabidopsis gene was isolated after transposon tagging (Aarts et al., 1993; see chapter 2) and simultaneously, the homologue from *B. napus* was isolated from a cDNA library. The MS2 amino acid sequence is compared to related proteins and a possible function of *MS2* in male gametogenesis is discussed.

Results

The male sterile ms2 phenotype

The male sterile phenotype of *ms2* plants is evident by a lack of seed set after self-fertilization. Upon closer examination, the anthers of the mutant are brown and shrivelled and contain no pollen. The plants flower much longer and as a result they are often taller than comparable wild-type plants, a general characteristic of (male) sterile plants. The *ms2* mutation does not completely abolish functional pollen production. Genetically unstable *ms2::1/dSpm* plants as well as plants with stable *ms2* mutant excision alleles produce approximately 1% of wild-type seed set. The allele can thus be propagated in a homozygous state, with all progeny produced this way again expressing the male sterile phenotype. The majority of *ms2* seeds are formed on the side shoots emerging late in development or high up the inflorescence upon aging of the plant. A variation in the level of self-fertilization was

also observed between *ms2* plants sown at different times of the year. Especially in winter the level of self-fertilization increased, sometimes making *ms2* plants appear hardly sterile. This leaky phenotype points at a certain redundancy of the *MS2* gene (function) in male gametogenesis.

Microscopic examination of ms2 pollen development

Pollen development in *ms2* mutants was studied by light microscopy in thin sections of mutant flower buds. As a comparison, similar sections of wild-type Landsberg *erecta* were studied and the key stages in microsporogenesis of wild-type and *ms2* plants are shown (Figure 1). Microspore mother cells (MMCs) and tapetal cells are formed from sporogenous tissue in the anther. During meiosis, MMCs divide to form a tetrad encapsulated by a thick callose layer (Figure 1A). The microspore wall starts to develop whilst still within the tetrad. At this stage, large vacuoles are formed in the tapetal cells, which remain until the callose layer is dissolved after meiosis and the microspores are released from the tetrads (Figure 1B). At this stage, the tapetum becomes more densely stained and the cell walls between the individual tapetal cells become less and less distinct. Microspores mature steadily, changing into a three lobed shape with increasingly thicker exine wall deposits (Figure 1C). During microspore maturation, the tapetal cells start to degenerate, they become less densely stained and are again visible as individual cells (Figure 1G and H). Shortly before maturation, microspores divide mitotically to become binuclear gametophytes. Upon release, the pollen grains are nearly round and very densely stained, with a thick pollen wall showing the typical exine structure. When the mature pollen is ready to be released from the anther locules, the tapetum has completely disappeared (Figure 1I).

Development of *ms2* pollen resembles the wild type until shortly after MMC meiosis. After tetrads have been formed, the tapetal cells are more vacuolated than in the wild type (Figure 1D). At the time of release from the tetrads, young microspores are very irregularly shaped compared to the more round shape of wild-type microspores. The tapetal cells are densely stained at this stage, with more vacuoles than the wild type (Figure 1E). The microspores eventually become round, but there is no sign of pollen wall synthesis (Figure 1F). From this stage on, the microspores and tapetal cells have more vacuoles than their wild-type counterparts. Microspore wall development does not proceed and the microspores become more densely stained and show much more internal structures than in the wild type (Figure 1J). The tapetal cells enlarge and expand into the anther locule. After strong vacuolation, the microspores and tapetal cells are eventually degraded (Figure 1K). In the mature flower, the anther locules are empty, apart from a very thin layer of tapetal and microspore remains, occasionally seen deposited against the locule side (Figure 1L). For figure 1, flower buds early in *ms2* plant development were chosen, at a stage in which the leaky phenotype is often not visible, which explains why no normal pollen grains were observed in the studied anthers at anthesis (from 3-4 flowers). When searching more flowers with anthers at anthesis, an occasional pollen grain was found. These were compared to wild-type pollen grains of similar developmental stage by transmission electron microscopy (Figure 2). Apart from aberrant

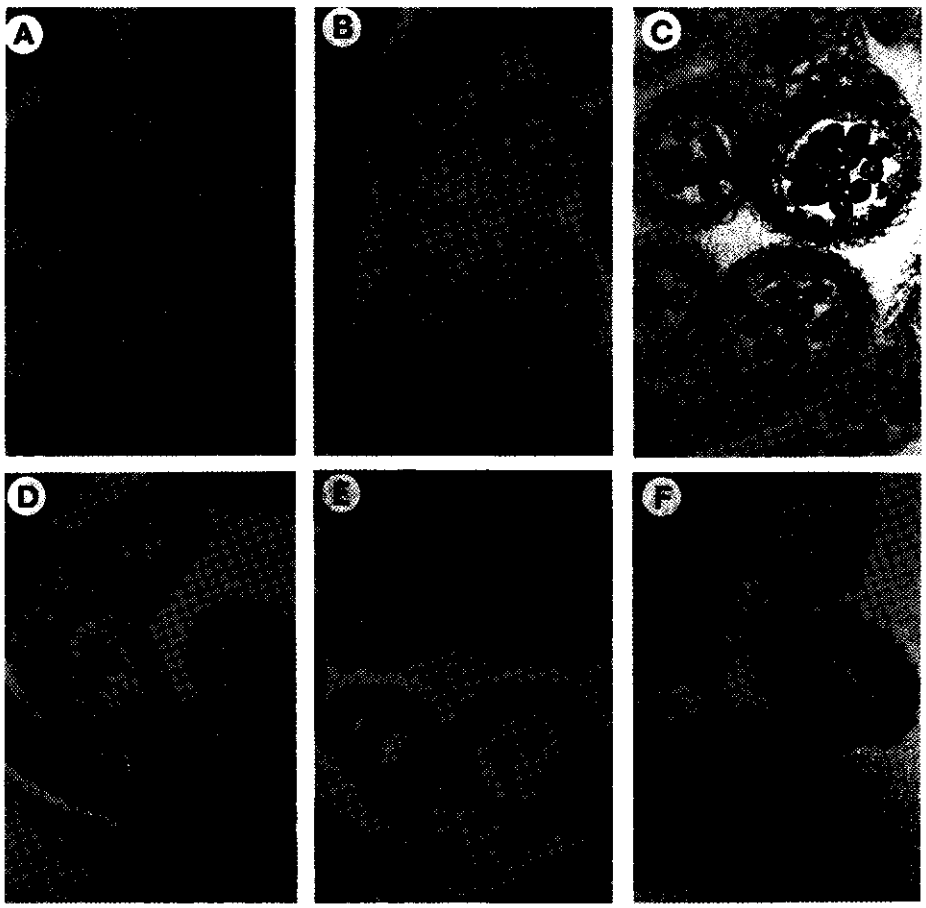

Figure 1: Cytological studies of subsequent stages in microspore development in wild-type Arabidopsis (*A, B, C, G, H, I*) and in male sterile *ms2* (*D, E, F, J, K, L*) flower buds

A. Wild-type (w.t.) anthers around the time of microspore release from tetrads. The left anther is still in tetrad stage, in the right anther young microspores are released.

B. W.t. anther with microspores developing first layers of exine.

C. W.t. anther with three-lobed microspores.

D. ms2 anthers around the time of microspore release from tetrads. Note the strongly vacuolated tapetal cells at tetrad stage (right) and the irregularly shaped released microspores (left).

E. ms2 anther with young irregularly shaped microspores.

F. ms2 anther with round and collapsed microspores. Note the large nuclei and other structures in the microspores and large vacuoles in tapetal cells.

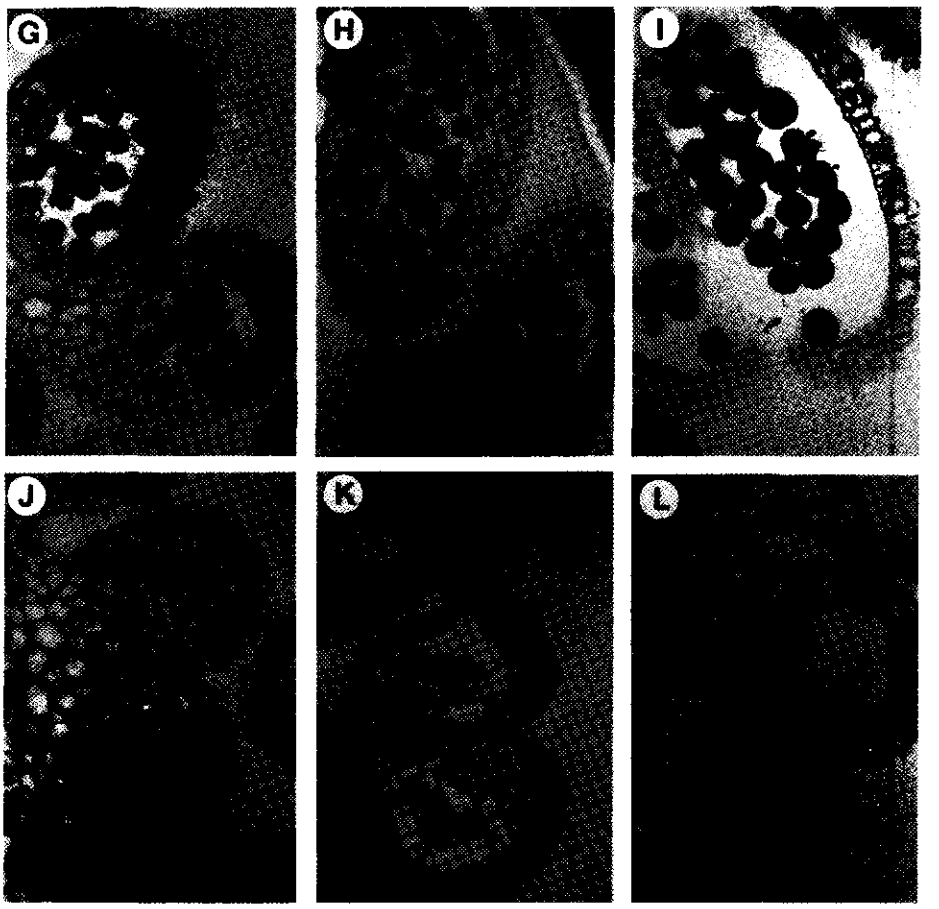

Figure 1(continued)
G. W.t. anther with uninuclear microspores.
H. W.t. anther with binuclear microspores.
I. Open w.t. anther releasing mature pollen grains.
J. *ms2* anther with degrading microspores. Most microspores have collapsed, the few remaining microspores have not formed any exine wall. Note the large tapetal cells.
K. *ms2* anther with most microspores degraded. Tapetal cells have enlarged and are highly vacuolated.
L. Empty *ms2* anther at the time of anthesis with hardly any microspore and tapetal remains left.

cytoplasmic organization due to degradation of the *ms2* pollen grain, the complete absence of exine is remarkable. Instead a thin layer of unknown substance has been formed. The lack of exine on *ms2* pollen grains renders them very sensitive to acetolysis (Figure 3).

Expression of MS2

MS2 cDNA and genomic DNA clones have been isolated using the tagged *ms2::I/dSpm* mutant (Aarts et al., 1993). Preliminary genetic mapping in an F_2 population segregating for *ms2* and the phenotypic markers *hy2*, *gl1* and *tt5*, positioned the *MS2* locus on chromosome 3. The locus was accurately mapped using the cDNA clone as a RFLP probe on DNA blots of recombinant inbred lines (RILs) generated from a Landsberg *erecta* x Columbia cross (Lister and Dean, 1993). With these RILs *MS2* was positioned about 5 cM from RFLP marker g4523 and 3 cM from m228 (Figure 4).

Northern blot analysis of wild-type plants confirmed the expectation that *MS2* was only expressed in flower buds, and not in seedling, stem, leaf, root or mature flower tissue (data not shown). The site of RNA transcription was specified by mRNA in situ hybridization (Figure 5). *MS2* transcription is initiated in wild-type tapetal cells at or very shortly after young microspores are released from the tetrads at a stage when hardly any exine wall has been formed. *MS2* expression is confined to a short period in microspore development and the mRNA disappears upon thickening of the exine wall. This transcription period corresponds very well with the appearance of the mutant phenotype (Figure 1). RNA blot hybridization determined that expression of *MS2* in *ms1* flower buds is similar to that in wild-type flower buds (data not shown). Transcription of the *MS2* gene is thus not affected by *MS1* gene expression. This agrees well with cytological studies, which place the effect of the *ms1* mutation after the first exine has been deposited (Dawson et al., 1993).

MS2 promoter analysis

With the *MS2* cDNA as a probe, a genomic lambda clone containing the *MS2* gene was isolated from a Landsberg *erecta* library. The DNA sequence of a 1077 bp HindIII fragment including the 5' end of the cDNA and around 950 bp of promoter DNA was determined (Figure 6A). A putative TATA box sequence (TATAAAA) was found 58 bp upstream of the cDNA sequence start. The promoter region of *MS2* was compared to promoter sequences of several other anther-specific genes present in the DNA databases. Regions around 250 bp upstream of the TATA-box showed similarity to a conserved box in the promoters for the anther epidermis specific *sf2* and *sf18* genes from sunflower (Domon and Steinmetz, 1994; Evrard et al., 1991). A similar, though weaker, homologous sequence was found in the A6 promoter from Arabidopsis (Hird et al., 1993) and the TA29 promoter from tobacco (Seurinck et al., 1990) (Figure 6B), located respectively 102 and 181 bp upstream of their putative TATA-boxes.

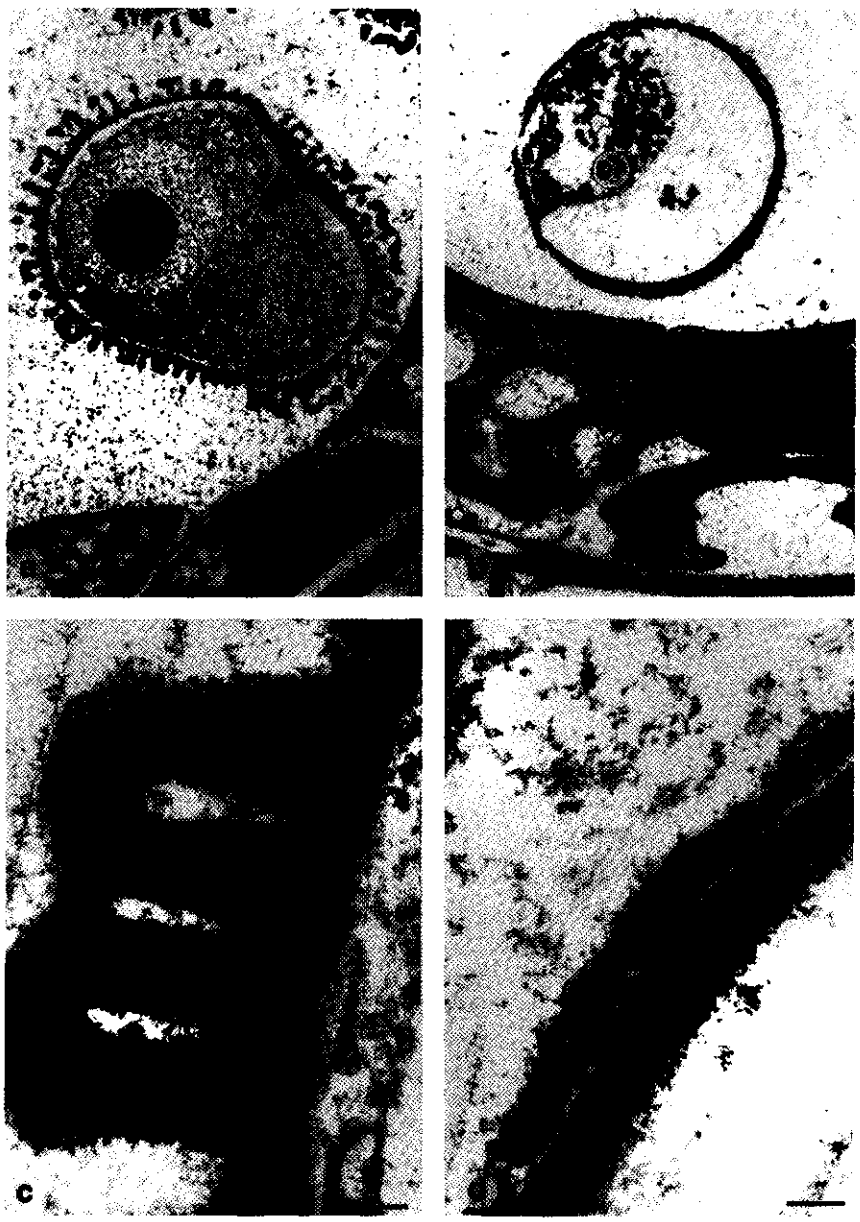

Figure 2: Transmission Electron Microscopy of wild-type and *ms2* pollen grains.
A. Wild-type anther locule showing a young microspore with a well-developed exine layer. Bar = 2 μm.
B. *ms2* anther locule showing a developing pollen grain which lacks exine deposition. Bar = 2 μm.
C. Close-up of wild-type exine layer. Bar = 0.2 μm.
D. Close-up of *B*. showing the aberrant *ms2* pollen wall. Bar = 0.2 μm.

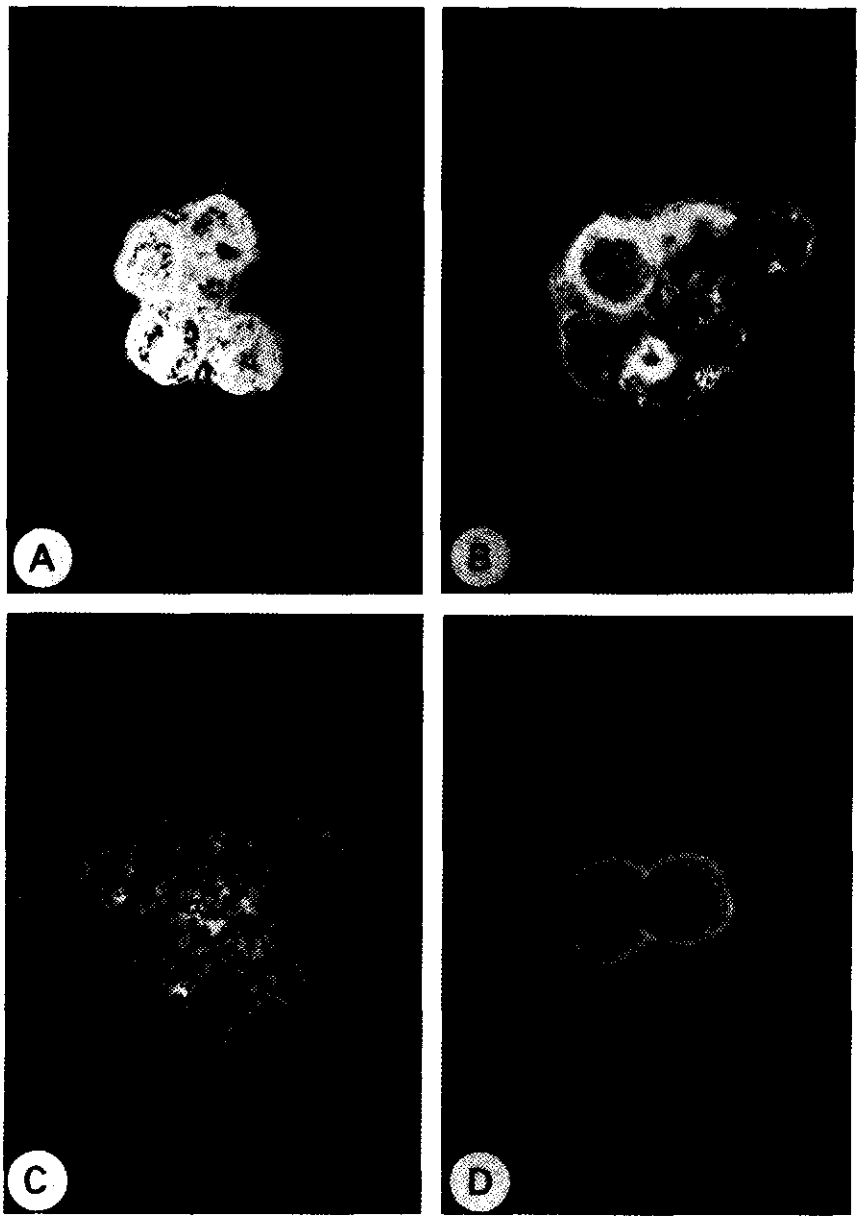

Figure 3: Increased sensitivity of *ms2* pollen grains to acetolysis, compared to wild-type pollen grains.
A. *ms2* pollen grain prior to acetolysis treatment, showing irregular shape and lack of exine development.
B. Same pollen grain as in *A*. after 5 mins. in acetolysis treatment.
C. Same pollen grain as in *A*. after 15 mins. in acetolysis treatment.
D. Wild-type pollen grain after 15 mins. in acetolysis treatment.

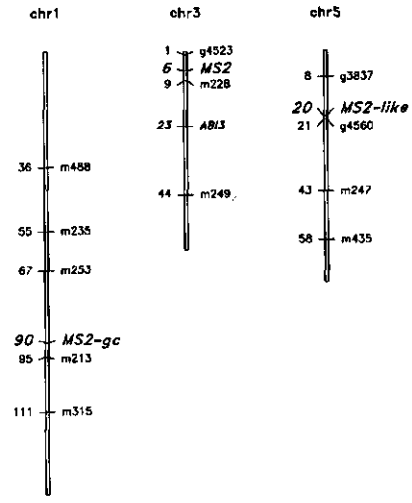

Figure 4: Linkage map showing the genetic positions of the *MS2* cDNA, the weakly hybridizing *MS2-gc* genomic DNA copy and the *MS2-like* cDNA relative to RFLP markers in a population of recombinant inbred lines (Lister and Dean, 1993).

Heterologous activity of the *MS2* promoter was tested using a *MS2* promoter-*β-glucuronidase* (GUS) gene fusion transformed into tobacco. Flowers of two transformants were examined after staining with X-Gluc. GUS activity was observed in the tapetum of anthers at the time of microspore release from tetrads in 1.0 to 1.5 cm long flower buds, comparable to stages 2 to 4 as described by Koltunow et al. (1990). In all other stages of flower development the *MS2* promoter was inactive, except for mature pollen, which segregated about 1:1 for blue and colourless pollen. The blue pollen clearly originated from GUS-expression in pollen grains and not from diffusion of the protein from maternal tissue, as pollen were collected by dipping the dehisced anthers in the GUS assay buffer before adding substrate. The unexpected ectopic pollen expression may be artifactual, as was also found for PR-1a and A6 promoter-GUS fusions in tobacco (Hird et al., 1993; Uknes et al., 1993).

MS2 homologues

After long exposure of an autoradiogram of genomic Arabidopsis DNA probed with the *MS2* cDNA, an extra, weakly hybridizing *MS2* genomic DNA copy (*MS2-gc*) besides the *MS2* gene was detected, which has been mapped to chromosome 1 (Figure 4). In addition an Arabidopsis partial cDNA sequence with similarity to the *MS2* cDNA was found in the Expressed Sequence Tag (EST) database. This *MS2-like* cDNA was mapped to the top half of chromosome 5 between g3837 and g4560 (Figure 4). This is also the region where the *male sterility1* (*ms1*) locus has been mapped, but *MS2-like* clearly maps above RFLP marker g4560, while *ms1* maps below this marker (Hauge et al., 1993; Z.A. Wilson, unpublished results). *MS2-gc* does not map in the vicinity of any other known *ms* locus. Over all, *MS2* seems to be part of a small gene family in Arabidopsis, with at least two transcriptionally active members.

Expression of the *MS2-like* cDNA was tested by semi-quantitative RT-PCR analysis on wild-type total RNA. The highest level of transcription was found in open flowers, and an approximately two-fold lower level in leaves and flower buds. The gene is expressed 10 to 20-fold less in stem than in open flowers (data not shown). Expression of *MS2* was used as a control for tissue specificity of the RT-PCR.

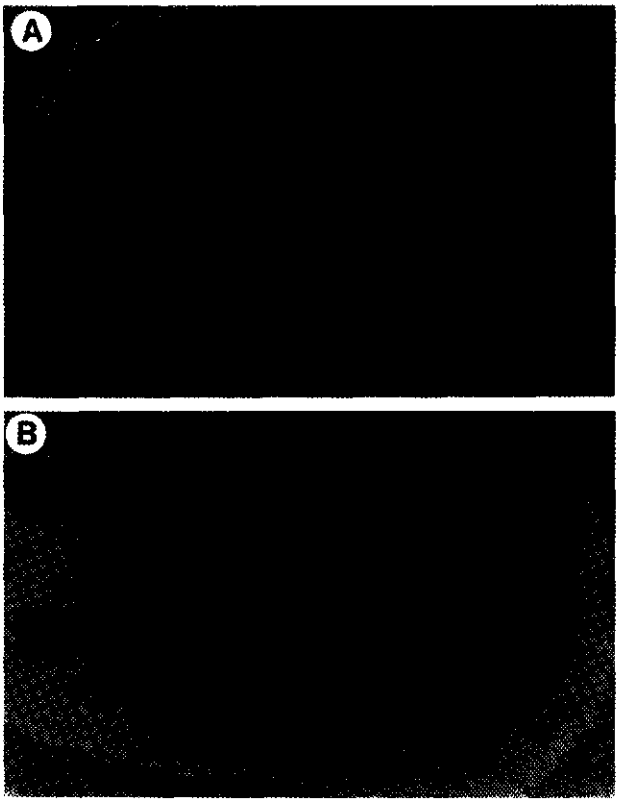

Figure 5: In situ RNA hybridization of wild-type Arabidopsis flower buds with a *MS2* specific RNA probe.
A. Wild-type flower shortly after microspore meiosis, with microspores still within tetrads. No detection of hybridizing probe.
B. Wild-type flower shortly after release of microspores from tetrads. The *MS2* mRNA is detected only in the tapetum. No signal was detected in similar staged flower buds using a sense probe.

The *MS2* cDNA probe hybridizes to genomic DNA of *B. napus* and *Raphanus sativus*, two other members of the Brassicae family (data not shown). At least one of the copies in *B. napus* is an active gene, which was isolated by cold plaque screening (Hodge et al., 1992) as a low abundant anther specific cDNA (clone C103). The *MS2Bnap* cDNA encodes one open reading frame of 616 amino acids with 90.4% overall identity to the predicted MS2 protein sequence (Aarts et al., 1993). This cDNA is a true homologue of *MS2*, as was

confirmed by complementation of the Arabidopsis *ms2* mutation after transgenic expression of the *MS2Bnap* cDNA under control of the tapetum specific A9 promoter (data not shown).

When the sequence databases were searched, several other *MS2* homologues were found. The best similarity (41.4% amino acid identity) was found to a 493 amino acid protein encoded by the jojoba (*Simmondsia chinensis*) fatty acyl reductase (JJFAR) gene (Metz et al., 1994). The similarity of this shorter protein sequence starts at position 118 of the MS2 amino acid sequence, but extends throughout the JJFAR sequence (Figure 7). JJFAR is involved in the formation of seed wax esters, by reducing wax fatty acids to the corresponding wax alcohols (Metz et al., 1994). The complete sequence of the Arabidopsis *MS2-like* cDNA was determined and as with the JJFAR protein, the predicted a.a. sequence started at pos. 118 of the MS2 a.a. sequence. This sequence has 39.8% overall identity with the MS2 sequence. A maize EST with unknown function was found, for which the predicted open reading frame similarly started at pos. 116 of the MS2 sequence (Figure 7). Weak but significant homology was found to the HETM protein from *Anabaena* (Black and Wolk, 1994) and the Saf3 domain of the SafA protein from *Myxococcus xanthus*. The *hetM* gene is proposed to be part of a set of genes required for polyketide synthesis (Black and Wolk, 1994). The *safA* gene encodes a peptide synthetase involved in the biosynthesis of saframycin Mx1, a heterocyclic quinone antibiotic (Pospiech et al., 1995). These genes have in common that they encode reducing proteins in the condensation-elongation steps required for fatty acid, polyketide or cyclic peptide synthesis.

We previously reported homology of MS2 to an open reading frame present in the wheat mitochondrial DNA upstream of the 26S ribosomal RNA *rrn26* gene (Aarts et al., 1993). This region of homology could be extended using the database sequence, by assuming and combining putative plant mRNA splice acceptor-donor sites in the mitochondrial genome (See Material and Methods). Homology starts at the 5' end of the DNA sequence entry and is composed by joining four exonic sequences, with AG-GT sequences present at five out of seven exon/intron or intron/exon borders. The predicted amino acid sequence of the open reading frame covers positions 238-403 of the MS2 protein sequence (Figure 7).

Discussion

Little is understood about the function of genes which are known to be involved in male gametogenesis and which give a male sterile phenotype when inactivated. One example of a gene whose function has been elucidated is the adenine phosphoribosyl-transferase (APRT) gene of Arabidopsis. Mutation of this gene causes pleiotropic effects in the whole plant, including male sterility (Regan and Moffat, 1990). Mutation of the *MS2* gene does not show any pleiotropic effects and to our knowledge it remains the only cloned and characterized Arabidopsis gene that is exclusively expressed in microsporogenesis.

Mutant *ms2* plants show the general phenomena observed with other male sterile Arabidopsis mutants, such as longer inflorescences, more flowers and extended flowering

A

```
   1 GATCTAAGACAAAAACGTGGCCATTTGCTAATTGTTGTTTTTGTTGTAGCAATAACCTTA  60
  61 GTCAAAGGATTTTGTTTATTGCGGACCCAAGTTGGTTGGTCGGCTCTTGCTTAAACCACA 120
 121 TTTGGAATTTGTTGTTCTGGAGTCTGGAGATCATTGAAACACAACCAAGAAGATAGCGCA 180
 181 CTGGTTTTAAAGTCGTATGTGTAGTTCTTTGTTCACCACGAGTTTAAGGTTCTCTTTCAT 240
 241 GTCTCATTGTTCTAAATATTCATCTTCGGTTGCATGTTTAACTTCATAGTCCAGTTTATA 300
 301 TTTTCCATCTAGATGATTGGGAACATTTTGCTTACTTTTATGATCTTAAACAGATGAACG 360
 361 GTCTCATGTTAACAACATAGTACTCTTGACTTCATGATAATTTCATATCATCTAATGACT 420
 421 AAATTCTTTGCAGAGTTTAATGGTGTTGATTGTTGAAACAAGAGCAGATTGGTCAATCAC 480
 481 TACAGAAAAAAAAAAAGTTGGTAACATGTAAGTTTAACGTTATTTAATAAAGGAGGATCTA 540
 541 AGTTTTCTACAAAAGCTATAATTTTTATGATGACCATATAATCCTCAAACCCTTCAAGAT 600
 601 GTGATGTGAATTATCTAAATCCCAACACGAAGAAATGAGATTTTTTTAAAGTTAGCTATTT 660
 661 ATCCTTAGTTGATTTCTTAATTATAGGGTAATGGCAATATTTTTTGGAACTGATAATACG 720
 721 TTTCTTTTTTTTTTTCTGAATTCTAGATGATCACGTGTAGGAAACTGATAAAATGTTGGAA 780
 781 AGAATTCGTAAGGCAATCTTTTATTTCACTTGATTTTTAAAATATTTATTTGCCTATAAA 840
 841 ACAGAGGAAGTTTTTCATCATCTTTTGTCCTTAGAACTAACCAATCTTTCATTCCTCTTA 900
 901 TAAAAACAAAACCTACTTTACTTGTCTCTTAACGATAACAAAATAACAAATAATTAATTC 960
 961 TGTTCTTGGTTTACTTAATCTTCTTTCTAGTTAAGTATATTCTTGTTGCTCATCACCAAA1020
1021 GGTATGCTTTCTAGGTTAAGTATATTACAAGTCACCAATTTCTTAACCAACAAGCTT   1077
```

B

```
NTTA2      1265 CATAAC--TGAAATC-AGGGTGAGACAAAATTTTG 1296
                ||||||    ||||  |   |||||    |||
HASF2ASP    922 CATAACAAAACAATCGAAATTGAGATTTTATTAAA  956
                | | |  ||||     || |||||||||  |||||
MS2         616 TAAATCCCAACACGAAGAAATGAGATTTT-TTAAA  649
                | | || ||||     ||||||||||   |||||
HASF18     1420 CATAACC-AACAACCGAAAATGAGATTTA-TTAAA 1452
                ||||  |  ||   |  ||||||  ||    |||
ATA6GLUA    685 CATAGC--ACAAAAGAACTATGAGAGTTTCAAAAA  717

consensus       CATAACCNAACAANCAAAAATGAGATTTTATTAAA
```

Figure 6: DNA sequence with conserved homology of the *MS2* promoter.
A. DNA sequence of a 1077 bp long HindIII *MS2* promoter fragment. The putative TATA-box (pos. 899) has been underlined. Part of the genomic DNA corresponding to the cDNA sequence (Aarts et al., 1993) has been double underlined, and thus positions 1022 to 1074 define the first intron. A conserved DNA box (see *B*) is shaded.
B. Comparison of a short DNA motif in the *MS2* promoter with a conserved DNA box in the promoters of the *sf2* and *sf18* genes from sunflower (HASF2ASP and HASF18), the TA29 promoter from tobacco (NTTA29) and the A6 promoter from Arabidopsis (ATA6GLUA). A consensus sequence is shown below.

time (Chaudhury, 1993; Henzel et al., 1994). A seasonally influenced, low frequency of self-fertilization, as we observed for stable and unstable *ms2::1/dSpm* mutants and for *ms2* mutants, has previously been observed for other Arabidopsis male steriles (Van der Veen and Wirtz, 1968). The stable *ms2::1/dSpm* insertion allele can be considered a null-allele, and therefore, the leaky male sterile phenotype indicates that pollen formation is not strictly dependent of *MS2* expression. Even in the absence of functional MS2 protein, a few pollen grains are formed, although with a poorly constructed pollen wall. Some pollination occurs, so apparently pollen grains survive the unfavourable environmental conditions, reach the stigma, germinate and cause fertilization. It can easily be envisioned that the number of survivors depends on the environment, thus explaining the observed variation in leakiness of the mutant depending on environmental conditions.

Pollen development is arrested shortly after release of microspores from tetrads in young *ms2* buds. Similar phenotypes are known for other species, like the *ms9* mutant of tomato (Rick, 1948) or the *ms2* mutant of maize (Albertsen and Philips, 1981), both of which are

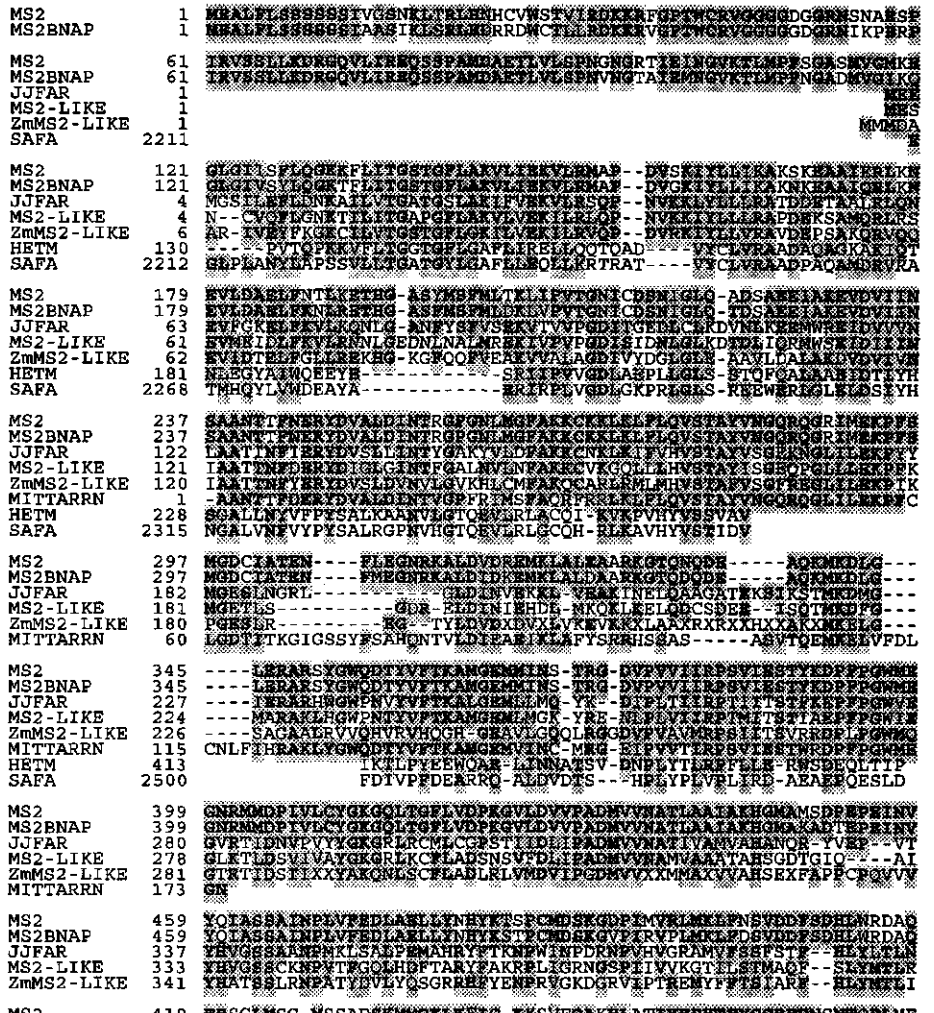

Figure 7: Comparision of the predicted amino acid sequence of MS2 with similar amino acid sequences. The MS2 sequence is compared to the amino acid sequences of the *MS2* homologue from *Brassica napus* (MS2BNAP), the jojoba fatty acid reductase (JJFAR), an Arabidopsis *MS2* homologue (MS2-LIKE) and the partial sequence of a maize *MS2* homologue (ZmMS2-LIKE). Some conserved sequences were found in the HetM protein of *Anabaena* sp. strain PCC7120 (HETM) and the SafA protein from *Myxococcus xanthus* (SAFA). The homologous region derived from a wheat mitochondrial DNA region (MITTARRN) was constructed by combining plant mRNA splice donor-acceptor sites. Similar amino acid residues are shown as shaded boxes, identical residues are indicated in bold face.

severely disturbed in the formation of the pollen wall. When the phenotypes of all male sterile Arabidopsis mutants are considered, the *ms2* mutant falls between the meiotic mutants *mei1* (He et al., 1996), 6492, 7219 and 7593 (Peirson et al., 1996), and post-meiotic mutants *ms1*, *ms7*, *ms8* and *ms36* (Chaudhury, 1993; Mulligan et al., 1994; Van der Veen and Wirtz, 1968).

The expression of *MS2* is confined to the tapetum at the time of microspore release, as was confirmed by RNA in situ hybridization. This timing of *ms2* promoter activity is conserved in tobacco transformed with a *MS2* promoter-*β-glucuronidase* gene fusion. Anther specific GUS staining starts in flowers of stage 2 (Koltunow et al., 1990), coinciding with the time of microspore release from tetrads, and continues until stage 4, which ends before the first mitotic microspore division. The action of the *MS2* promoter in tobacco resembles the activity of the rice Osg6B promoter in tobacco (Tsuchiya et al., 1994), both acting slightly later than the A9 promoter from Arabidopsis (Paul et al., 1992) and the TA29 promoter from tobacco (Mariani et al., 1989). The expression of GUS in pollen of the two *MS2-GUS* tobacco plants was unexpected based on the *MS2* in situ hybridization results in Arabidopsis, but the 1:1 segregation of wild-type and GUS tobacco pollen showed that staining was indeed caused by a pollen-expressed *GUS* gene. A similar phenomenon of unexpected GUS activity in pollen was observed by Uknes et al. (1993) in transgenic PR-1a promoter-*GUS* tobacco and by Hird et al. (1993) in transgenic A6 promoter-*GUS* containing tobacco, who suggested that the differences in promoter expression may be due to differences in pollen development between Arabidopsis and tobacco. Alternatively the differences in pollen GUS expression between tobacco and Arabidopsis may be caused by the presence of pollen specific enhancer sequences within the GUS reading frame, that are recognized by tobacco and not by Arabidopsis.

When searching the *MS2* promoter DNA sequence for regions conserved between anther specific promoters, sequences in the *MS2* promoter were found to share similarity to a conserved sequence box in the promoters of the *sf2* and *sf18* genes from sunflower (Domon and Steinmetz, 1994; Evrard et al., 1991). Similar, but less conserved, sequences are found in the tapetum specific Arabidopsis A6 promoter and the tobacco TA29 promoter (Hird et al., 1993; Seurinck et al., 1990), but not in the Osg6B promoter from rice (Tsuchiya et al., 1994). This sequence box may be required for directing expression to the anther in general, in contrast to the tapetum specific "anther box" found in Petunia genes (Van Tunen et al., 1994).

In search of a putative function for the MS2 protein, we have screened the databases for homologues with a known function. This revealed the wheat mitochondrial sequence that has been reported before (Aarts et al., 1993). When introns are artificially spliced out, according to the 5'-GT, AG-3' splice acceptor-donor sites rule for plant pre-mRNA splicing (Brown, 1986), the homology extends even beyond the region previously reported. The extended amino acid sequence shows a remarkably high degree of conservation compared to the MS2 sequence. This suggests either an evolutionary advantage to maintain this sequence as a functional gene in the mitochondrial genome or a recent integration of a nuclear DNA copy with no function in the mitochondrion. There are several arguments for the latter explanation.

First of all the conservation of nuclear splice donor/acceptor sites makes mitochondrial expression doubtful. Introns in mitochondrial genes are normally bordered by mitochondrial group II splice sites, which differ from nuclear DNA splice sites. Although the reading frame shows a long stretch of similarity with MS2, it ends abruptly and no further homology to the 3' part of the gene was found in over 4 kb of downstream sequence. Considering that plant mitochondrial genes evolve extremely slowly compared to animal mitochondrial genes (Gray, 1990), the wheat mitochondrial DNA copy is probably not a functional mitochondrial gene, but more likely a recent incorporation of nuclear DNA in the mitochondria.

The only protein similar to MS2 with a known function was the fatty acid reductase from jojoba (JJFAR). This protein acts together with a condensing enzyme, the acyl-CoA:fatty alcohol acyltransferase, in synthesizing the storage wax esters in developing jojoba seed (Metz et al., 1994; Shockey et al., 1995). Although the JJFAR protein sequence lacks a 117 amino acid N-terminal domain present in the MS2 protein sequence, the similarity between the two sequences extends over the entire JJFAR length. Also the ESTs from Arabidopsis and maize with similarity to *MS2* encode proteins without the extra N-terminal domain. All these similar proteins share a conserved region also found in a number of reductase proteins that act together with condensing enzymes in either fatty acid, polyketide or cyclic peptide synthesis (Black and Wolk, 1994; Chirala et al., 1989). This conserved region (positions 133-150 of MS2; Figure 7) has a proposed role in NAD(P)H binding, as it contains a motif [I,V,F]-X-[I,L,V]-T-G-X-T-G-F-L-[G,A] very similar to the consensus sequence [I,V,L]-X-[I,L,V]-X-G-X-G-X-X-G for ADP binding (Eggink et al., 1990; Wierenga et al., 1986).

The function of the *MS2-like* and *MS2-gc* genes remain unclear as yet. The *MS2-like* gene is expressed in leaves, so analogous to the JJFAR protein it may be a fatty acid reductase needed for wax biosynthesis. None of the 9 mapped epicuticular wax or *eceriferum* (*cer*) mutants described so far, is positioned in the vicinity of *MS2-like* on chromosome 5. However, *cer5* maps close to *MS2-gc* on chromosome 1 (Koornneef et al., 1983). For this gene no function in the wax biosynthesis pathway has been proposed (Lemieux et al., 1994) and the possibility that *CER5* is an *MS2* homologue may be considered.

The similarity to a fatty acid reductase and the presence of a conserved NAD(P)H binding site are strong evidence that MS2 is involved in fatty acid reduction. The JJFAR protein is localized in the endoplasmatic reticulum, but both MS2 and its *B. napus* homologue have a C-terminal microbody targeting signal and they are therefore probably located in microbodies of tapetal cells at the time of exine wall deposition on the young microspores. The subcellular localization in tapetal microbodies for the production of exine components is in line with a general requirement of plant microbodies for fatty acid processing (Beevers, 1979). The major component of exine is sporopollenin. Although the composition and assembly of this polymerized macromolecule is still largely unknown, there are clues that long carbon chain molecules are involved in polymerization (Southworth, 1990). Reduction of fatty acyl groups to fatty alcohol groups may be one of the steps in the formation of sporopollenin. The absence of exine wall deposition in *ms2* mutant microspore development further supports the idea that fatty alcohols may be the precursors from which sporopollenin polymerization proceeds (Scott, 1994).

Material and methods

Cytological analysis

Flower buds of Landsberg *erecta* (wild type) and stable *ms2::I/dSpm* plants were fixed in 5% (v/v) glutaraldehyde (in 0.1 M phosphate buffer, pH 7.2) for 2 hrs. After rinsing with phosphate buffer, the tissue was dehydrated in ethanol series and finally embedded in Technovit 7100 (Kulzer Histo-Tec). 5-10 μm thick sections were cut, mounted and stained with 1% toluidine blue. Transmission electron microscopic analysis was as described by Worrall et al., (1992). The acetolysis analysis was as described by Erdtman (1960).

In situ hybridization

Young inflorescences of wild-type Landsberg *erecta* were fixed for 3 hrs in 1% (v/v) glutaraldehyde in 0.02 M HEPES buffer (pH 7.2). Fixed inflorescences were dehydrated in an ethanol series, cleared by dipping in 1:1 ethanol/Histoclear for 2 hrs and then 3 changes of 100% Histoclear for 1 hr each. Cleared samples were embedded in paraffin wax for sectioning and preparation of slides (Cox and Goldberg, 1988). Slides were prehybridized and hybridized based on Bradley et al. (1993), using a digoxigenin (DIG)-labelled RNA probe derived from a 505 bp 5' ScaI *MS2* cDNA fragment (Aarts et al., 1993). Hybridizations were carried out at 50°C. After hybridization slides were washed for 3 hrs (2X SSC, 50% (v/v) formamide at 50°C), rinsed twice with 0.5 M NaCl, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, treated with RNase A (20 μg/ml) for 30 min at 37°C, and again rinsed twice. Slides were then washed again for 30 min (2X SSC, 50% (v/v) formamide at 50°C) followed by 2 min with 1X SSC at room temperature. Slides were stored overnight at 4°C in PBS (pH 7.2). Immunolocalization of the hybridized probe was according to Bradley et al. (1993).

Genetic mapping of MS2, MS2-like and MS2-gc

All *MS2* homologous probes were mapped on a population of Columbia x Landsberg *erecta* Recombinant Inbred Lines (RILs) (Lister and Dean, 1993), kindly obtained through the Nottingham Arabidopsis Stock Centre (Nottingham, UK). DNA from 50 to 100 RILs digested with BglII were used for mapping. Mapping data were processed using the JoinMap program (Stam, 1993). An 1.0-kb XhoI-XhoI fragment representing the 3' end of the *MS2* cDNA was used for mapping *MS2* and *MS2-gc*. The complete 1.6-kb partial cDNA insert from cDNA clone 32C3T7 (Newman et al., 1993) was used as probe to map *MS2-like*.

MS2 promoter analysis

A HindIII genomic DNA fragment known to reside upstream of the *MS2* cDNA sequence, was subcloned in SK^+ and the DNA sequence was determined. This fragment was used to

construct a *MS2* promoter-*GUS* gene fusion in a pBIN19 derived binary vector. *Nicotiana tabacum* SamsonNN was transformed with *Agrobacterium tumefaciens* strain LBA4404, harbouring the *MS2* promoter-*GUS* fusion gene. Transformants were selected for resistance to 100 mg/l kanamycin. Two transformed shoots (6301 and 6302) were obtained, transferred to the greenhouse and assayed for GUS expression.

GUS assay

Flowers and flower buds were collected from transformants 6301 and 6302 containing the *MS2 promoter-GUS* T-DNA. According to length and appearance, the flowers and buds were arranged in different developmental stages (Koltunow et al., 1990). Anthers from buds of stages -1 to +12, either complete or in half, were incubated in GUS assay buffer (0.05 M phosphate buffer pH 7.2, 2 mM $K_3[Fe(CN)_6]$, 1mM EDTA, 0.1% Triton) with 1mM X-Gluc (5-bromo-4-chloro-3-indoyl-ß-D-glucuronide) for 1-2 hrs. After staining the tissue was cleared with 100% ethanol and examined microscopically.

Analysis of MS2 homologues

The *MS2* homologue from *Brassica napus* was isolated as an anther specific mRNA (Hodge et al., 1992). GenBank and Genesequence databases were searched for other similar sequences using BLAST programs (Altschul et al., 1990). Entries identified as homologues were a partial cDNA clone from Arabidopsis (EST clone no. 32C3T7; acc. no. T04771) as *MS2-like*, a partial cDNA clone from maize (EST clone no. 6c02e02; acc. no. T18413) as *ZmMS2-like*, a protein sequence from jojoba representing a fatty acyl reductase (JJFAR; acc. no. R26898; Metz et al., 1994) and a mitochondrial DNA fragment from wheat (MITTARNG; acc. no. Z11889). A few stretches of significant homology were found to the HETM protein from *Anabaena* sp. strain PCC7120 (acc. no. P37693; Black and Wolk, 1994) and the SafA protein from the myxobacterium *Myxococcus xanthus* (acc. no. U24657; Pospiech et al., 1995). The partial cDNAs for *MS2-like* and *ZmMS2-like* from Arabidopsis and maize, were requested and additional DNA sequence was determined to extend the reading frame and correct for the frame shifts found in the database entries. The *ZmMS2-like* cDNA was found to contain a deletion of 184 bp when compared to RT-PCR derived cDNA sequence from maize seedling RNA, which was corrected in figure 7. The amino acid sequence of the wheat mitochondrial *MS2* homologue was composed by combining DNA sequences 1-28, 125-225, 349-552 and 621-811. All introns were bordered by 5'-GT and 3'-AG sequences.

Acknowledgements

We thank Maarten Koornneef for critically reading the manuscript, Stefan C. Hyman for preparing sections for electron microscopy, Christiane Leonards-Schippers for mapping the *MS2-gc* DNA copy, the Nottingham Arabidopsis Stock Centre for providing the RIL lines, the Maize cDNA project (University of Arizona, Tuscon, USA) and the Arabidopsis Biological Resource Center (Columbus OH, USA) for providing the cDNA clones,

Gerrit Stunnenberg and Gerard Scholten for taking care of the plants in the greenhouse, and the Centrale Fotodienst (SC-DLO) for their help in preparing the figures. RH was supported by Nickerson Biocem Ltd (Cambridge, UK) and MA was supported by European Union BIOTECH grant BIOT-CT90-0207.

Chapter 4

Molecular characterization of the *CER1* gene of *Arabidopsis thaliana* involved in epicuticular wax biosynthesis and pollen fertility

Mark G.M. Aarts, Christian J. Keijzer[1], Willem J. Stiekema and Andy Pereira

[1]Department of Plant Cytology and Morphology, Wageningen Agricultural University, Arboretumlaan 4, 6703 BD Wageningen, The Netherlands.

(Published in *The Plant Cell* (1995) 7, 2115-2127)

Summary

The aerial parts of plants are coated with an epicuticular wax layer, important as a first line of defense against external influences. In *Arabidopsis thaliana* the *ECERIFERUM* (*CER*) genes effect different steps of the wax biosynthesis pathway. In this chapter we describe the isolation of the *CER1* gene that encodes a novel protein involved in the conversion of long chain aldehydes to alkanes, a key step in wax biosynthesis. *CER1* was cloned after gene tagging with the heterologous maize *En/Spm-I/dSpm* transposable element system. Mutant *cer1* plants display glossy green stems and fruits and are conditionally male sterile. The similarity of the CER1 protein to a group of integral membrane enzymes, which process highly hydrophobic molecules, points to a function of the CER1 protein as a decarbonylase.

Introduction

Waxes are found in a wide variety of living organisms as a mixture of long chain fatty acid derived substances (Kolattukudy, 1976). In plants, these components are specifically found as an epicuticular layer that covers leaves and young stems and which is often visualized by a characteristic glaucous appearance (Kolattukudy, 1975). The primary function of epicuticular wax deposition is to reduce water loss through the epidermis (Hall and Jones, 1961), a feature contributing to drought tolerance. In addition this outer layer has a major function in the interaction with herbivorous insects and plant pathogenic fungi (Eigenbrode and Espelie, 1995; Podila et al., 1993; Städler, 1986; Thompson, 1963). A rather unexpected function of wax has been described recently by Preuss et al. (1993), who found pollen wax composition and structure as important factors for a proper pollen-pistil interaction.

The wax composition is determined by the various biochemical steps of wax biosynthesis. Plant wax layer mutants, which unravel these biochemical steps, are available (Bianchi et al., 1985; Kolattukudy, 1980; Von Wettstein-Knowles, 1979) and have led to a basic outline of the wax biosynthetic pathway (Bianchi et al., 1985; Lemieux et al., 1994; Von Wettstein-Knowles, 1979; Von Wettstein-Knowles, 1994). Starting with hexadecanoic acid, long chain fatty acids with an even carbon number (in general C_{20} to C_{32}) are produced by elongation. These fatty acids are reduced to fatty aldehydes and primary alcohols, or reduced and decarbonylated to yield alkanes with an uneven carbon number (Cheesbrough and Kolattukudy, 1984). The latter can be further converted to secondary alcohols and ketones (Kolattukudy, 1980).

A large number of genetic loci influencing wax deposition, has been identified in maize, barley, Brassica spp. and *Arabidopsis thaliana* (Baker, 1974; Bianchi et al., 1985; Koornneef et al., 1989; McNevin et al., 1993; Von Wettstein-Knowles, 1979). Mutants with altered wax production or wax composition are in general characterized by a bright green phenotype. In Arabidopsis there is an additional effect on fertility for some mutants, and the 21 different genetic wax layer or *ECERIFERUM* (*CER*) loci identified in Arabidopsis have been grouped into four classes based on fertility and visual degree of glossiness of the mutants (Koornneef et al., 1989; McNevin et al., 1993). Isolation of these *cer* genes will contribute to the understanding of pollen-pistil interactions as well as plant-herbivorous insect interactions. Furthermore it may provide tools for the manipulation of wax composition in crop species to generate a source of broad host range resistance against herbivorous insects. A highly active *En/Spm-I/dSpm* transposon tagging system, as described in chapter 1, has been used to generate transposon mutagenized populations which were screened for *cer* mutants.

In this chapter we report the identification and phenotypic analysis of an *I/dSpm* transposon-tagged *cer1* mutant and the subsequent isolation and characterization of the *CER1* gene involved in wax biosynthesis and pollen fertility. Stem wax of *cer1* mutants has been analysed previously (Hannoufa et al., 1993; Lemieux et al., 1994; McNevin et al., 1993) and found to be especially rich in aldehydes but lacking alkanes, suggesting the CER1 protein is involved in the conversion of aldehydes to alkanes. We present evidence for the function of the CER1 protein as an aldehyde decarbonylase.

Results

Phenotypic and genetic analysis of a transposon induced cer mutant

To isolate genes involved in epicuticular wax biosynthesis, we used a *I/dSpm* transposon tagging approach to generate and screen Arabidopsis lines containing the *En/Spm-I/dSpm* transposon tagging system (see Material and methods; Aarts et al., 1995a). Among the screened lines, we selected one line with multiple *I/dSpm* elements and the TEn2 transposase T-DNA (Aarts et al., 1995a), which revealed some bright green semi-sterile mutants among normal wild type plants. Phenotypically these mutants strongly resembled a known class of *cer* mutants with a glossy stem and reduced fertility (Koornneef et al., 1989). Complementation tests with the *cer1*, *cer3*, *cer6*, and *cer10* mutants in this class revealed that the transposon induced mutant was allelic to *cer1-1*. No clear phenotypic differences were found between our *cer1* mutants (which we called *cer1-m*) and *cer1-1* mutants. Both displayed a strong glossy stem and fruit phenotype, without any visible sign of wax production. Wax production on other organs of the plant was apparently not altered, but it must be noted that epicuticular wax production in wild type plants is mainly confined to stem and fruit surface and not visually detectable on other parts of the plant.

Mutation of *cer1* not only alters the wax deposition on stem and fruit, but has a pleiotropic effect on plant fertility. A similar effect has been observed for *cer6-2* or *pop1* (defective in pollen-pistil interactions), described as a conditionally male sterile *cer* mutant, but was never characterized for *cer1* mutants. Therefore, we decided to examine the cause of sterility in much the same way as was performed for *cer6-2* (Preuss et al., 1993). Although normal amounts of pollen are produced by the *cer1-1* and *cer1-m* mutants, they were completely self-sterile under dry (30-40% relative humidity) conditions. Crossing *cer1-m* mutant flowers with wild type Landsberg *erecta* pollen led to a normal seed set. Self-fertility was increased to the wild type level by growing the plants under high humidity conditions (90-100% relative humidity), indicating that, as with *cer6*, the self-sterility is a environmentally controlled form of male sterility (Preuss et al., 1993). We tested germinaton of *cer1-m* and Landsberg *erecta* wild type pollen on mutant and wild type stigma papillae (see Material and methods) and found that mutant pollen does not germinate on either wild type or *cer1-m* pistils, whereas the pollinations with wild type pollen were normal, yielding normal seed set. Pollen from *cer1-m* mutants germinates similarly to wild type pollen *in vitro* (data not shown). Similar to in the *cer6-2* mutant, the deficiency of *cer1-m* pollen grains to germinate *in vivo* coincided with the inability to rehydrate on the stigma surface. Rehydration of wild type pollen grains occurred within minutes after deposition on the stigma papillae, during which time the pollen grain has swollen to about twice its original size (Preuss et al., 1993, M.G.M. Aarts, unpublished observations). The accumulation of callose on the stigma surface in response to pollination, which was observed for *cer6-2* pollen (Preuss et al., 1993), was also seen when fully receptive wild type or *cer1-m* stigmas were pollinated with *cer1-m* pollen, but not when pollinated with *Brassica napus* or *Petunia hybrida* pollen (data not shown).

In terms of pollen-pistil interaction the *cer1* mutants are very similar to *cer6-2* mutants. For the latter mutant, conditional male sterility is explained based on an alteration of the tryphine layer covering the pollen grain (Preuss et al., 1993). When examined in a scanning electron microscope, exines of mature, shed pollen of wild type, *cer1-m* and *cer1-1* all appeared to contain tryphine, which generally was found covering the entire pollen grain (Figure 1A). Using transmission electron microscopy, the amount of lipid in the tryphine of the three genotypes examined was comparable. However, lipid droplets in the tryphine of *cer1-m* and *cer1-1* pollen (Figs. 1C and D) were more numerous and considerably smaller than in tryphine of wild type pollen (Figure 1B).

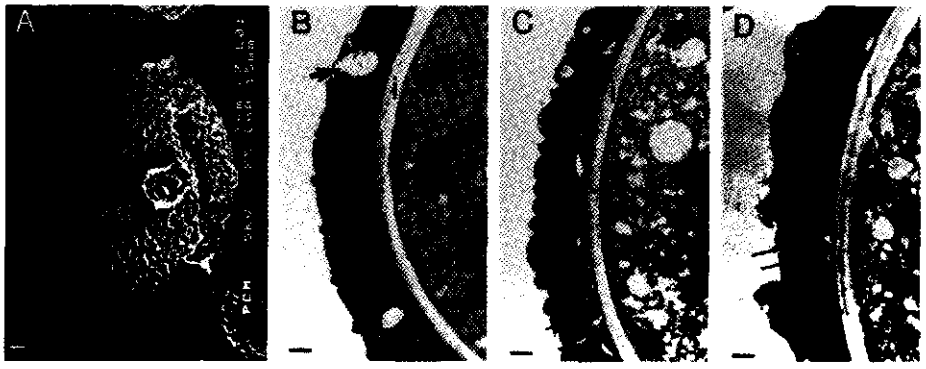

Figure 1: Tryphine on the pollen grains of wild type plants and *cer1-1* and *cer1-m* mutants
A. Scanning electron microscopy of a *cer1-1* pollen grain with tryphine covering its entire surface (Bar = 1 μm).
B. Transmission electron microscopy showing a detail of a cross-section through a Landsberg *erecta* wild type pollen grain. Tryphine can be observed filling the exine (Bar = 500 nm).
C. Transmission electron microscopy showing a comparable detail of a cross-section through a *cer1-1* pollen grain (Bar = 500 nm).
D. Transmission electron microscopy showing a comparable detail of a cross-section through a *cer1-m* pollen grain (Bar = 500 nm).
Tryphine in *cer1-1* and *cer1-m* appears more granulated than in wild type, with more but smaller lipid droplets and gas inclusions. Exine (e), intine (i), gas inclusions (thick arrows) and lipid droplets (thin arrows) are indicated in *B*, *C* and *D*. Gas inclusions are the white spherical or irregularly shaped spaces within the exine cavities. Lipid droplets are spherical or oval shaped grey inclusions in the exine, which do not contrast very much with the rest of the tryphine, but can be distinguished due to the amorphous structure of their contents.

The cer1-m mutant is tagged by an I/dSpm element

The *cer1-m* mutant was found in a line with transposing *I/dSpm* elements and was presumably caused by insertion of an *I/dSpm* element. To determine whether *cer1* was tagged, large offspring populations from mutants were screened for progeny that had reverted to the wild type phenotype. This is a phenomenon typical of transposon-induced mutations.

Germinal reversions were found at a frequency between 1 in 50 to 1 in 300 in four independent progenies, indicating that the unstable mutation was indeed due to a transposon insertion in the *CER1* gene.

DNA blot analysis of segregating progenies (see Material and methods) revealed one *I/dSpm* insert (*I/dSpm89*) cosegregating with the *cer1-m* mutant phenotype. The flanking DNA of this *I/dSpm89* insert was amplified by inverse PCR (IPCR) and cloned. Based on the DNA sequence, primers were designed for PCR amplification of wild-type and revertant *I/dSpm89* excision alleles. Three independently derived germinal revertant plants all contained an excision allele, demonstrating that the *cer1-m* mutant was indeed tagged by the *I/dSpm89* insertion, creating a *cer1::I/dSpm89* allele. Excision of *I/dSpm* elements normally create short base pair deletions and additions (Aarts et al., 1993), but in these three cases, the DNA sequences of the revertant alleles were identical to the wild type DNA sequence, suggesting insertion of *I/dSpm89* at a vital position of the gene.

The availability of a transposon tagged *cer1* mutant allowed an easy screen for determining if the *CER1* gene acts cell-autonomous. The presence of clearly cell specific somatic reversion sectors would mean a cell-autonomous expression. Somatic reversions were observed on the stem of about 20% of the mutant plants. Such variegated plants showed one or more very small glaucous grey sectors of wax deposition on a bright green background (Figure 2A). The distinct boundaries of the wild type revertant sectors on the stem suggested a cell-autonomous expression of *CER1*. Scanning electron micrographs of these sectors revealed individual epidermal cells with overlaying wax deposition shown as the rod-like wax structures typical for Arabidopsis (Koornneef et al., 1989) (Figure 2B). Upon higher magnification, it was clear that the effect of reversion was not completely cell-autonomous, as some wax structures had formed on epidermal cells flanking revertant cells (Figure 2C). Either the *CER1* gene product or the wax components generated by this gene product, had diffused a short distance away from the producing cell.

The TEn2 *En*-transposase, present in the *cer1-m* mutants, induces in general a low frequency of transposition and therefore we combined the *cer1::I/dSpm89* allele with TEn5, a different, more active, *En/Spm* transposase locus (Aarts et al., 1995a; See Material and methods), which increased the number of variegated mutants and the size and number of wild-type sectors per plant about fivefold. To determine the effect of *CER1* expression in different layers, we searched for large excision sectors, that extended over whole inflorescences. One such sector was found, and to our surprise, this sector now had wild-type stem wax, and displayed a concomitant reversion to wild-type fertility. To test whether the reversion had occurred in the L2 layer, 22 offspring descending from the reverted inflorescence were sown. All offspring were found to be *cer1* mutant, suggesting that the excision was L1 rather than L2 layer specific.

Cloning of CER1

With the *I/dSpm89* flanking genomic DNA as a probe, a homologous cDNA clone as well as a 17-kb long genomic clone was isolated from the respective DNA libraries. To confirm

that the cDNA clone originated from the *CER1* locus, part of the insert DNA was used as a probe and hybridized to a blot of *cer1::I/dSpm89* mutant and revertant plants (Figure 3). All mutants were homozygous for a fragment containing the *I/dSpm89* insert, although excision could be observed, and all revertants were either hemizygous for the *I/dSpm89* insert or they lacked the insert.

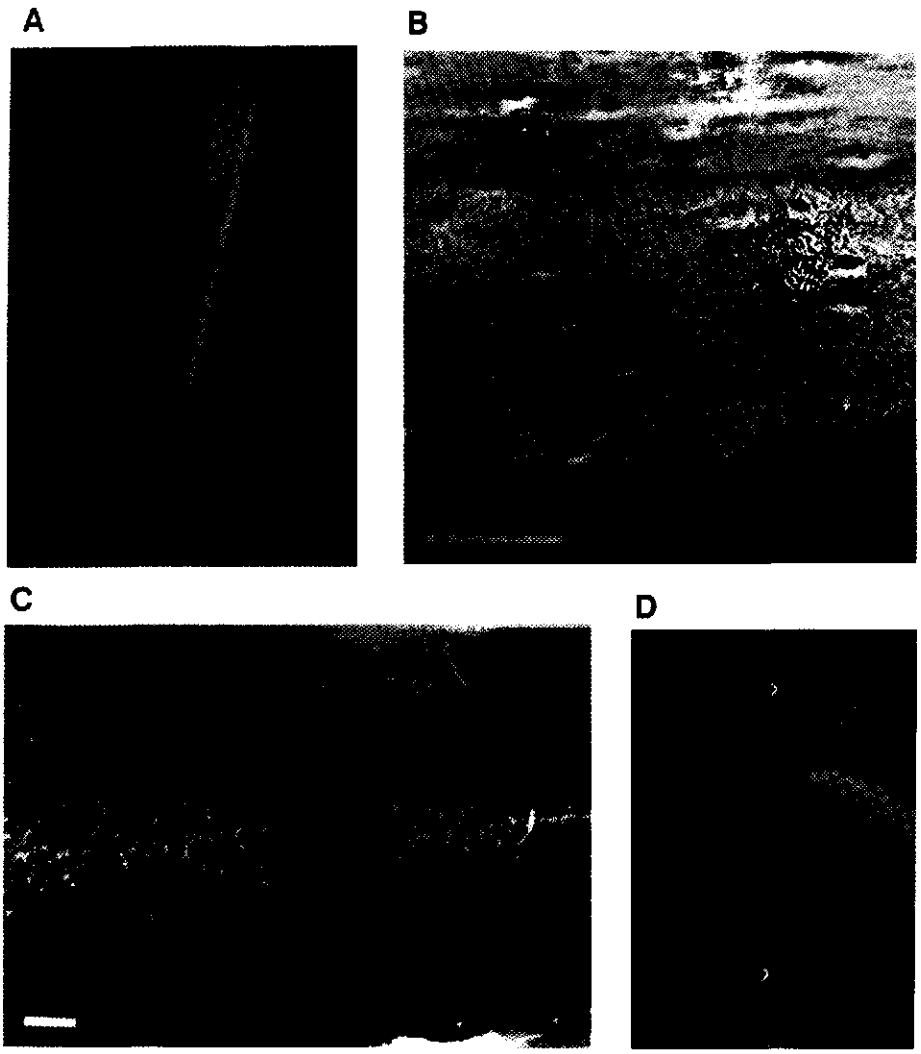

Figure 2: Phenotype of a transposon tagged *cer1* mutant.
A. Stem of a *cer1::I/dSpm* mutant. The arrowhead indicates a glaucous wild type somatic excision sector.
B. Scanning electron microscopy of a wild type excision sector similar to the one shown in *A* (Bar = 100 μm).
C. One-cell-wide excision sector with wild type epicuticular wax production. The presence of the typical wild-type Arabidopsis epicuticular wax rods is not limited to the surface directly above the epidermal cells carrying the excision, but some wax structures are also formed on neighboring cells (Bar = 10 μm).
D. Stem of a wild type (*CER1,cer1::I/dSpm89*) plant with a *cer1* mutant sector (arrowhead).

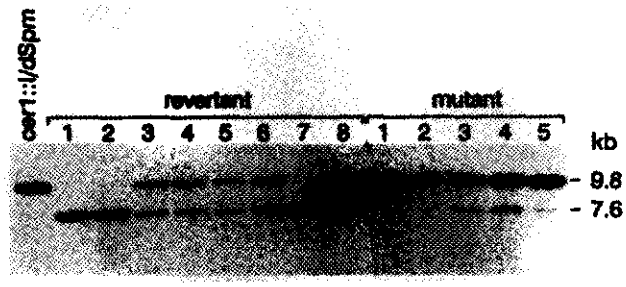

Figure 3: All wild-type revertant offspring of a homozygous *cer1::I/dSpm89* mutant plant have at least one *CER1* excision allele.
EcoRI-digested DNAs of the *cer1::I/dSpm* parental plant, the wild-type revertant, and *cer1* mutant offspring, were probed with part of the *CER1* cDNA. The 9.8-kb EcoRI fragment present in the parental plant represents the *cer1::I/dSpm89* allele with the transposon insert. Excision of the 2.2-kb *I/dSpm89* element results in a 7.6-kb *CER1* allele EcoRI fragment, which is present in all revertants. All mutants show somatic excision of *I/dSpm89* from their two *cer1::I/dSpm89* insertion alleles leading to a small amount of the 7.6-kb EcoRI excision fragment in these lanes.

Conclusive proof that the isolated gene was indeed the *CER1* gene involved in epicuticular wax formation was obtained from the analysis of a plant with wild-type phenotype except for a small mutant *cer* sector (Figure 2D). In the course of *cer1::I/dSpm89* analysis three such plants were found in various progenies. The mutant sector in one of these plants, hemizygous for *I/dSpm89*, ended in a small leaf from which DNA was isolated for PCR analysis. Combinations of an *I/dSpm* specific terminal primer and different *CER1* specific primers (see Material and methods), were used for PCRs with DNA from the *cer* sector and from the wild type rosette leaves of the same plant (Figure 4A). Two *cer* sector specific DNA fragments were amplified for two primer combinations. The new *cer* sector-specific *I/dSpm* insertion was positioned within the coding region of the cloned gene, 1.0 kb upstream of the *I/dSpm89* insert (Figure 4B). As a new insertion of an *I/dSpm* element into the cloned gene, resulted again in a mutant *cer* phenotype we conclude that the cloned gene is indeed *CER1* involved in epicuticular wax biosynthesis.

Analysis of the CER1 cDNA

Epicuticular waxes are mainly found on the stem and fruit epidermis of Arabidopsis and the isolated *CER1* gene should be expressed in these organs. *CER1* transcription was therefore tested by RNA gel blot hybridization and as expected, the *CER1* transcript was found in wild type stem and fruit tissue. Additional strong expression was detected in Arabidopsis flowers, in which expression of the *CER1* gene could be expected based on the male sterile phenotype of the mutant. Arabidopsis has little wax formation on the leaves, explaining the low level of leaf transcript (Figure 5A). Transcription of *CER1* gene was blocked in *cer1::I/dSpm89* mutant flowers, while the transcription in *cer1-1* flowers was not affected (Figure 5C). The mutant phenotype of the chemically induced *cer1-1* mutant is probably due to a minor rearrangement such as a point mutation. In flowers of the F_1 hybrid between *cer1::I/dSpm*

and *cer1-1*, the transcription level of the gene was the intermediate of the two parents (Figure 5C).

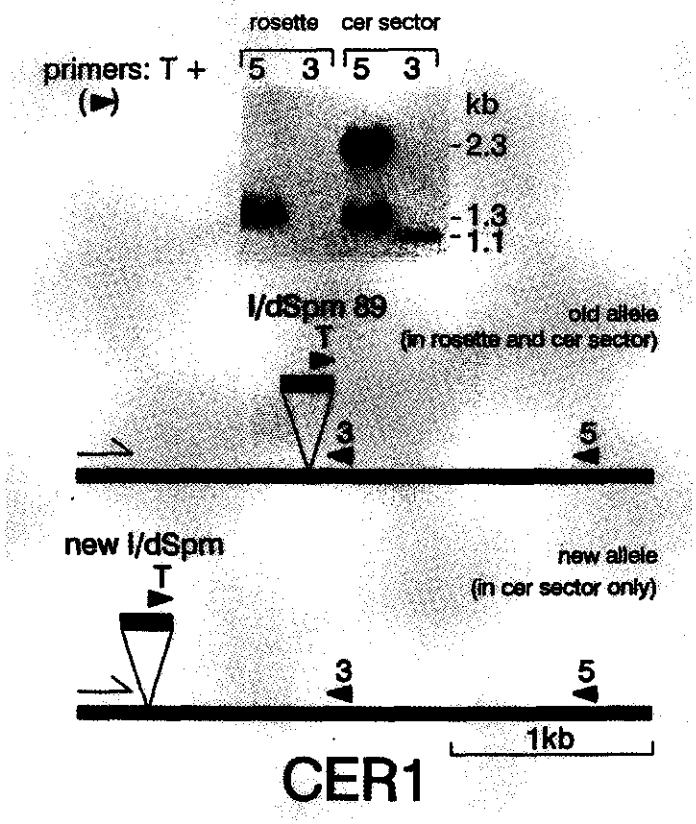

Figure 4: A *cer* mutant sector on a wild type *CER1cer1::I/dSpm89* plant (See figure 2D) is caused by a new insertion of an *I/dSpm* element in the *CER1* gene.
A. Gel blot hybridization of the *CER1* cDNA probe to DNA from four PCR reactions performed with two primer combination T and 5, and T and 3 on DNA from rosette leaves (rosette) or from a cauline leaf on which a *cer* sector (see figure 2D) ended (cer). Primer T is a terminal inverted repeat primer specific to *I/dSpm* elements, and primers 3 and 5 are specific to the *CER1* gene with their hybridization sites 1.2 kb apart. DNA polymerase extends from these primers towards the 5' end of the gene. With rosette and cer DNA, the combination of primers T and 5 amplified a 1.3-kb fragment derived from the *cer1::I/dSpm89* (old) allele. Only with cer DNA, the same primer combination amplified an extra 2.3-kb fragment derived from a new mutant *cer1* allele, replacing the wild type rosette *CER1* allele by insertion of another *I/dSpm* element. This new allele was confirmed with the primer combination T and 3, that amplified an 1.1-kb fragment only with cer DNA. The expected common fragment of 0.1 kb derived from the *I/dSpm89* insertion was too small to be seen on the DNA gel blot.
B. Schematic representation of the position of the primers (arrowheads) used in *A*. and of the insertions of *I/dSpm89* (old) and another *I/dSpm* element (new) in the *CER1* gene. The old allele is present in both rosette and cer samples and the new allele specific to the cer sample has replaced the wild type *CER1* allele present in the rosette sample. The arrows indicate the direction of the *CER1* reading frame.

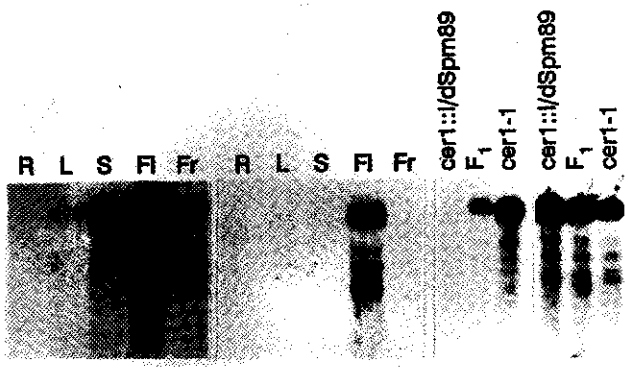

Figure 5: Transcription of the *CER1* and *CER1-like* genes in *Arabidopsis*.
A. RNA gel blot of 10 μg root (R), leaf (L), stem (St), flower (Fl) and fruit (Fl) total RNA, hybridized with the 822-bp 3' part of the *CER1* cDNA. There is no transcription in root, little in leaf, and strong transcription in stem, flower and fruit.
B. Duplicate RNA gel blot of *A*. hybridized with the ATTS1001 (EST) insert *CER1-like* probe, detects strong transcription in flower tissue only. The detected RNA is of a length similar to the one in *A*.
C. RNA gel blot of total flower RNA from *cer1::I/dSpm89*, an F_1 between *cer1::I/dSpm89* and the chemically induced *cer1-1* mutant, and the *cer1-1* parent, hybridized with the *CER1* probe used in *A*. *CER1* transcription is almost completely blocked by the homozygous *I/dSpm89* insertion, but only half that in the hemizygote and not at all in the chemically induced *cer1-1* mutant.
D. Same RNA gel blot as shown in *C*, stripped and rehybridized with the *CER1-like* probe used for *B*. There is no difference in transcription between the three tested plants showing that transcription of the *CER1-like* gene is in no way affected by transcription of the *CER1* gene.

The 2109-bp long *CER1* cDNA contained an open reading frame of 625 amino acids (Figure 6A). Part of the corresponding genomic DNA sequence was determined, and an in frame stopcodon was found 33 bp upstream of the ATG start codon, indicating that the cDNA clone comprised the complete open reading frame. A putative TATA transcription initiation sequence was present 72 bp upstream of the ATG start codon in the genomic DNA sequence. The predicted protein has an apparent molecular mass of 72.3 kD and a pI of 8.23. Analysis of the amino acid sequence with the PC\Gene computer package classifies the protein as an integral membrane protein. Two putative transmembrane helices are predicted stretching from amino acid positions 178 to 213 and 325 to 350 (Figure 6A), and additional membrane associated helices, cover amino acid positions 7-27, 45-65, 99-119 and 126-146. Two possible Asn glycosylation sites are found at positions 258 and 456. Insertion of *I/dSpm89* disrupts the reading frame from Thr (amino acid pos. 272) onwards (Figure 6A).

CER1 homologues are present in other species

Wax production is common to many plant species, and genes involved in wax biosynthesis may well be conserved among species. This was confirmed as database searches, carried out with the *CER1* cDNA and predicted amino acid sequences, revealed significant homologies

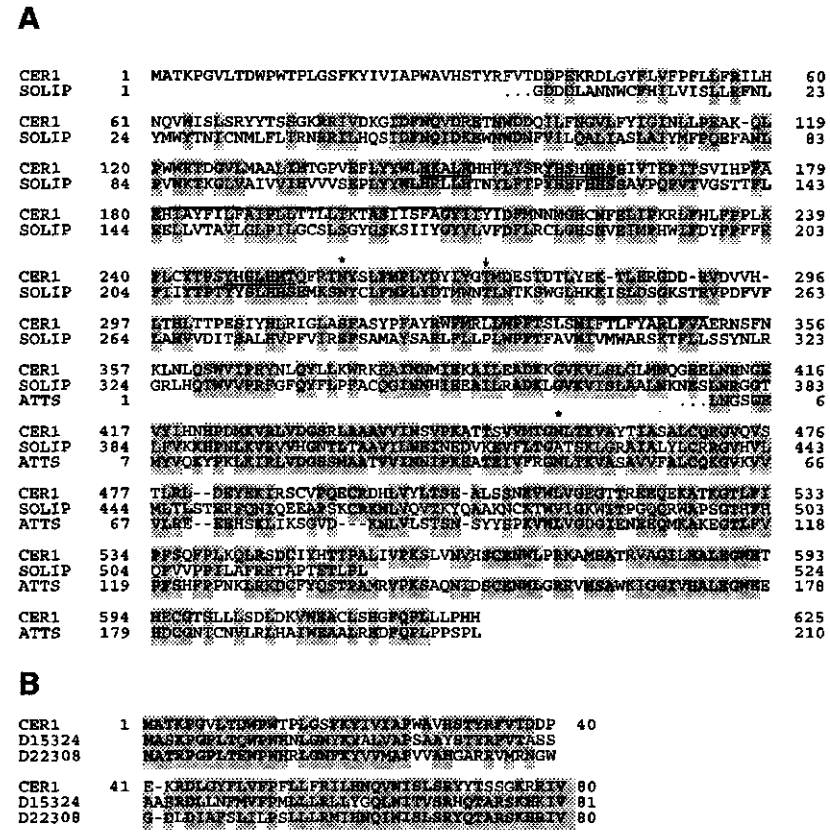

Figure 6: *CER1* deduced amino acid sequence and comparison to homologous amino acid sequences.
A. Amino acid sequence deduced from the *CER1* cDNA (CER1), compared to the homologous amino acid sequences derived from the partial cDNA sequence of SOLIPTRB (SOLIPT) from *Senecio odorus*, and ATTS1001 (AT1001) from Arabidopsis. Two putative membrane spanning sequences in the CER1 amino acid sequence are overlined. A histidine rich motif is underlined. Putative glycosylated asparagine residues in the CER1 amino acid sequence are indicated (*) as well as the site of the target site duplication caused by insertion of *1/dSpm89* (↓), that disrupts the reading frame in *cer1::1/dSpm89* mutants. Dots indicate that the SOLIPTRB and ATTS1001 cDNAs are partial and that the N terminal ends of the complete amino acid sequences are missing.
B. N terminal stretch of about 80 amino acids with 37.5% identity overall between the CER1 protein and two amino acid sequences deduced from two rice cDNA clones (D15324 and D22308). Shaded boxes indicate similar amino acid residues, identical amino acid residues are indicated in bold face. Similar residues are grouped as follows: (V,L,I,M), (S,T), (Q,N,E,D), (K,R), (G,A) and (F,W,Y).

with cDNA and expressed sequence tag (EST) sequences from both dicot and monocot species (see Material and methods). The predicted amino acid sequence of the EST ATTS1001 cDNA isolated from flower buds of Arabidopsis, showed 53.8% identity to the C-terminal region (210 amino acids) of the predicted *CER1* amino acid sequence (Figure 6A). Additionally, a *Brassica campestris* flower bud EST was found with 49.1% predicted amino acid sequence identity (117 amino acids), a potato epidermal EST with 67.4% amino acid identity (46 amino acids), and a *Senecio odorus* epidermal cDNA with 31.3% amino acid identity (513 amino acids) (Figure 6A). This family of related sequences could be extended

to monocot species maize and rice. A maize vegetative meristem EST showed 52.7% amino acid identity over 110 amino acids. The homology of two rice callus cDNAs sequenced from their 5' end, started exactly at the N terminus of the predicted *CER1* amino acid sequence extending over the entire length of sequenced cDNA, representing about 80 amino acids (37.5% overall identity; Figure 6B). Interestingly, the predicted amino acid sequence of another rice cDNA with a short stretch of amino acid similarity to the predicted CER1 amino acid sequence, showed additional homology in this region to the C-5 sterol desaturase protein of yeast encoded by the *ERG3* gene (Arthington et al., 1991). These two short stretches of homology are conserved between CER1, SOLIPTRB, the rice EST, and ERG3, and a part of it is also found in the maize EST (Figure 7). Each stretch of homology reveals a short motif with the consensus sequence: Tyr-His-Ser/Thr-X-His-His (where X stands for any amino acid).

A CER1-like gene is closely linked to the CER1 gene, but transcribed differently

Part of the genomic sequence upstream of the *CER1* gene was determined and when used in a DNA data base search, DNA sequences between 1.0 and 1.7 kb upstream of the *CER1* start codon were surprisingly found to be identical with the ATTS1001 EST sequence (data not shown). The DNA sequence of ATTS1001 shows only 62.2% identity with the *CER1* cDNA (compared to the 53.8% amino acid identity found previously) and when used as a probe it does not cross-hybridize to the *CER1* cDNA sequence. From these data we concluded that the ATTS1001 EST clone is part of a transcribed *CER1* homologous *CER1-like* gene located directly 5' to *CER1* and oriented in the same direction.

To confirm that this gene is transcribed in Arabidopsis and to examine any correlation between *CER1* and *CER1-like* expression we tested *CER1-like* transcription by Northern hybridization using the ATTS1001 EST insert probe (Figs. 5B and D). In contrast to the strong transcription of *CER1* in stem, flower and fruit, transcription of the *CER1-like* gene was restricted to flowers, but with a similar mRNA size and transcription level as the *CER1* gene (Figure 5B). Transcription of the *CER1-like* gene was not affected by the *cer1* mutation, as was seen by the similar level of *CER1-like* transcription in the flowers of either *cer1::I/dSpm89*, *cer1-1* or the F_1 (Figure 5D), demonstrating the difference in transcriptional regulation between *CER1* and *CER1-like*. The transcription of both the *CER1* and the *CER1-like* gene in chemically or radiation induced *cer2* to *cer9* flowers and stems was like in the wild type (data not shown), suggesting that none of these loci are transcriptional regulators.

Discussion

The CER1 protein has a function in wax alkane biosynthesis

We have cloned and characterized the *CER1* gene from Arabidopsis, involved in epicuticular wax biosynthesis and pollen fertility. *CER1* is a cell-autonomously expressed gene, mainly

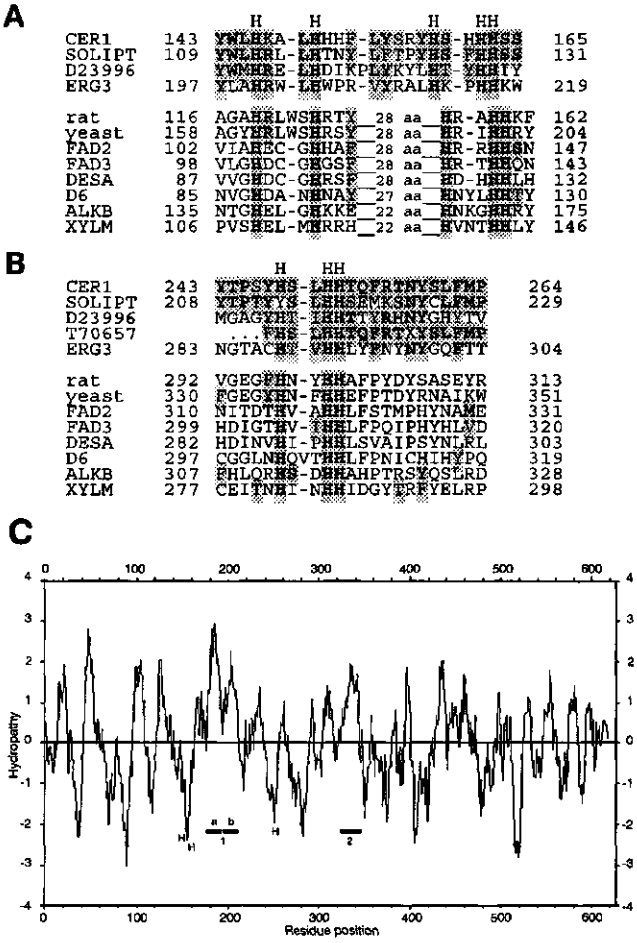

Figure 7: Comparison of the histidine rich regions from the CER1 and other amino acid sequences.
A. Comparison of the $HX_{3,4}H$ and first $HX_{2,3}HH$ motifs present in CER1 homologous sequences found after searching the sequence data bases with TBLASTN (top) and motifs present in a number of membrane-bound fatty acid desaturases, alkane hydroxylase and xylene monooxygenase (bottom; Shanklin et al., 1994). The CER1 homologous sequences are SOLIPTRB (SOLIPT) cDNA from *S. odorus*, a rice EST (D23996) and the *ERG3* or *SYR1* gene from yeast, encoding for the sterol C-5 desaturase. The additional integral membrane sequences shown are the stearoyl-CoA desaturase from rat (Thiede et al., 1986) and yeast (Stukey et al., 1990), the Δ12 and Δ9 fatty acid desaturases from Arabidopsis (FAD2, FAD3; Okuley et al., 1994; Arondel et al., 1992), the Δ12 fatty acid desaturase from *Synechocystis* (DESA; Wada et al., 1990), the Δ6 fatty acid desaturase from *Synechocystis* (D6; Reddy et al., 1993), the alkane hydroxylase gene from *Pseudomonas oleovorans* (ALKB; Kok et al., 1989) and the xylene monooxygenase from *Pseudomonas putida* (XYLM; Suzuki et al., 1991) are shown. Similar and identical residues are indicated as in figure 6.
B. Comparison of the second $HX_{2,3}HH$ motif present in the same sequences as in *A*. Homology to only the second HX_2HH motifs was found in a partially sequenced EST clone from maize (T70657) that showed further homology to CER1. The first and second $HX_{2,3}HH$ motif are separated by a transmembrane sequence crossing the membrane twice.
C. Hydrophobicity plot of the predicted CER1 amino acid sequence according to Kyte and Doolittle (Devereux et al., 1984), indicating the location of putative transmembrane sequences (horizontal bars labelled 1 and 2) and the HX_3H or HX_2HH motifs (H). Transmembrane sequence 1 is long enough to span the membrane twice (a and b).

active in stem, flower and fruit and at a low level in leaf. The *cer1* mutant is one among four of the *cer* mutants with a drastically changed epicuticular wax phenotype, for which a biochemical function has been proposed to the corresponding wild type gene. *CER2* and *CER6* are thought to encode components of fatty acid elongation, and *CER4* is suggested to be involved in fatty aldehyde reduction (Hannoufa et al., 1993; Jenks et al., 1995; Lemieux et al., 1994). Biochemical studies (Hannoufa et al., 1993; Lemieux et al., 1994; McNevin et al., 1993) have shown that *cer1* mutants are blocked in the conversion of stem wax C_{30} aldehydes (triacontanal) to C_{29} alkanes (nonacosane) and that they also lack the secondary alcohols (14- and 15-nonacosanol) and ketones (15-nonacosanone) derived thereof. Alkanes, secondary alcohols and ketones comprise around 65% of total wax in wild-type Landsberg *erecta* (Lemieux et al., 1994). The conversion of aldehydes to alkanes is moderated by aldehyde decarbonylases (Cheesbrough and Kolattukudy, 1984) and the CER1 protein may be an enzymatic component in this biochemical step to produce long chain alkanes.

CER1 protein contains an iron binding motif

Because no genes encoding fatty aldehyde decarbonylases have been cloned previously, this biochemical function has not been described at the sequence level. The homologies we observed between *CER1* and other cDNA or EST sequences in the sequence data bases describe a family of related proteins, but they are not very elucidative in defining a biochemical function for these proteins. The *S. odorus* and potato partial cDNA sequences encoding for CER1 homologous amino acid sequences, are described to be epidermis specific, which corresponds with a function for CER1 in epicuticular wax biosynthesis. The only homology we found to a functional protein, is to the C-5 sterol desaturase from yeast. This homology is actually confined to a very short sequence, centred around three histidine rich motifs ordered as: HX_3H + HX_2HH + HX_2HH (where X stands for any amino acid). These, or the related $HX_{3-4}H$ + $HX_{2-3}HH$ + $HX_{2-3}HH$ motifs, were recently found to be conserved among a number of integral membrane fatty acid desaturases from mammals, fungi, insects, higher plants and cyanobacteria, and bacterial membrane alkane hydroxylase and xylene monooxygenase (Figure 7A and B). All eight conserved histidine residues are essential for the enzymatic function, as was tested with complementation studies in yeast using site specifically mutagenized rat stearoyl-CoA desaturase genes (Shanklin et al., 1994). Apart from the histidine rich motifs, there is no apparent homology to other regions of these proteins (Figure 7A and B), suggesting that the reported proteins with these motifs (including the C-5 sterol desaturase and the protein encoded by the partial cDNA clone from rice) are likely to share some biochemical properties, rather than performing the same biochemical function. The presence of closely spaced histidine residues is typical for metal binding motifs, and based on the requirement of fatty acid desaturases for iron, the histidine motifs are strongly suggested to be involved in the binding of iron ions (Okuley et al., 1994). The spacing separating the first two histidine motifs from the third motif, which is reported to be conserved among membrane desaturases and also among the alkane hydroxylase and the xylene monooxygenase, is similar albeit slightly smaller in the CER1, SOLIPTRB, D23996

and ERG3 proteins (Figure 7A and B). This distance appears to be conserved within families of related proteins, as it is also smaller in the alkane hydroxylase and the xylene monooxygenase compared to the desaturases. A reason why there is some variation in the spacing between the first two and the third histidine motifs among protein families is that not the spacing between the motifs but rather the structure of this protein part is important for function of the proteins. All proteins with the histidine rich motifs described so far were found to contain long hydrophobic domains between the first two and the third histidine motifs (Shanklin et al., 1994). These domains are able to span a membrane twice. The same is found for the CER1, the SOLIPTRB and the ERG3 protein, in which both HX_2HH motifs are separated by a predicted transmembrane sequence that can be divided into two parts long enough to satisfy this requirement (Figure 7C).

Based on all these characteristics, the CER1 protein closely resembles a class of structurally and perhaps evolutionary related integral membrane enzymes that share the preference for highly hydrophobic substrates, the presence of metal binding histidine rich motifs and the need for electron donors to perform their catalytic function.

The CER1 protein as part of an aldehyde decarbonylase

The CER1 protein has been proposed to function in the decarbonylation of aldehydes to alkanes (Hannoufa et al., 1993; Lemieux et al., 1994). So far two plant aldehyde decarbonylases (from pea (*Pisum sativum*) and a green colonial alga (*Botryococcus braunii*)) have been studied in some detail (Cheesbrough and Kolattukudy, 1984; Dennis and Kolattukudy, 1992). Both are integral membrane proteins, with the pea decarbonylase suggested to be located in the cuticular cell membrane and the alga decarbonylase in the microsomal membranes. Both use highly hydrophobic fatty aldehydes as substrate and need metal ions for their function. The metal identity is only known for the purified decarbonylase from *B. braunii*, which interacts with cobalt present in a Co-porphyrin or corrin structure (Dennis and Kolattukudy, 1992). The partially purified decarbonylase from pea, is merely known to depend on metal ions, as the activity was severely inhibited in the presence of metal ion chelators (Cheesbrough and Kolattukudy, 1984). All enzymatic proteins with the histidine rich motifs identified so far, are proposed to have non-heme iron-containing active sites (Shanklin et al., 1994). This is in contrast to the observed interaction of *B. braunii* decarbonylase with porphyric cobalt. However, to assume all plant decarbonylases to interact with heme-bound metals based on the characterization of only one (lower) plant decarbonylase would be a bit premature and it can still be envisaged that decarbonylases from higher plants use non-heme Fe rather than heme Co. The similarities between the observed biochemical properties of the partially purified aldehyde decarbonylase from pea as an integral membrane protein processing highly hydrophobic molecules in the presence of metal ions, and the properties ascribed to the CER1 protein on basis of the homology to some functional integral membrane enzymes, is strong supporting evidence for the CER1 protein acting as the catalytic iron containing part of a fatty aldehyde decarbonylase enzyme involved in the Arabidopsis epicuticular wax and pollen tryphine alkane biosynthesis.

The aldehyde decarbonylase from pea is thought to be present in the extracellular or cuticular matrix of the epidermis (Cheesbrough and Kolattukudy, 1984). This is also the case for certain non-specific lipid transfer proteins suggested to be involved in cuticle formation (Pyee et al., 1994; Sterk et al., 1991; Thoma et al, 1993). In the absence of a signal sequence, the CER1 protein may be transported post-translationally to the cell membrane, where the putative membrane-spanning helices allow it to be anchored, with the histidine motifs probably located on the intracellular side. Wax biosynthesis at the site of wax deposition outside the cell can explain the diffusion in wax structures from the surface of stem epidermal cells with a functional *CER1* gene to cells without the gene, as we observed in a wax layer variegated Arabidopsis plant (Figure 2C).

The CER1 protein is needed for pollen fertility

Additional to stem wax alkane synthesis, the *CER1* gene has a essential function in pollen development. Mutant *cer1* plants have a stem wax and male sterility phenotype very similar to that of *cer3* and *cer6* mutants. As in *cer1* mutants, both stem and leaf wax alkane content of *cer3* and *cer6* mutants is low (Jenks et al., 1995; Lemieux et al., 1994). For *cer6-2* mutants the pollen lipid content has been examined biochemically and was found to be equally low in alkanes and wax components derived thereof (Preuss et al., 1993). The similarity in phenotype and the concomitant absence of stem alkanes in *cer1* and *cer6* mutants leads us to conclude that the CER1 protein performs a similar function in pollen tryphine synthesis as it does in stem wax synthesis. The low pollen alkane content of the *cer6-2* mutants was accompanied by the virtual absence of a tryphine layer deposited on the mature pollen grain (Preuss et al., 1993). Conditional male sterility of *cer1::I/dSpm89* and *cer1-1* mutants was not linked to a reduction of the tryphine deposition on the pollen grains. Although the lipid in the tryphine of both *cer1* mutants is more dispersed than in tryphine of the wild type pollen grains, the total amount of (visible) lipids in the tryphine was similar for all three genotypes. As a whole, the entire tryphine consistently shows a more granular appearance in the two *cer1* mutants. This might have to do with a further state of tapetal cytoplasm degeneration prior to tryphine deposition. Also the capillary process of tryphine transfer to the exine layer of pollen grains, might influence the final texture of the tapetal remnants. As with both *cer1* mutants, the conditionally male sterile *cer6-1* mutant also does not show a reduction in tryphine deposition (Preuss et al., 1993; C.J. Keijzer, unpublished results), in contrast to the *cer6-2* mutant. It must be mentioned that a correct observation of this phenomenon requires a strictly timed fixation: shortly before anthesis, tryphine can merely be found inside the tapetal membrane sacs and as late as a few hours before anther dehiscence it is transferred to the exines of the pollen grains. Analysing still closed anthers (sometimes even found in open flowers) might lead to the incorrect conclusion that pollen grains of that given plant lack tryphine.

Based on the electron microscope analyses, pollen sterility in *cer1* mutants is clearly not a mechanistic problem associated with lack of tryphine as might be with the *cer6-2* mutant. Tryphine deposition is a passive process following degradation of tapetal cells. The

differences in tryphine composition and ultrastructure must therefore be a direct consequence of differences in tapetum between wild type and *cer1* or *cer6* mutants. Lack of tapetal *CER1* or *CER6* activity results in a different constitution of tapetal remains, and depending on the severity of tapetal alterations caused by *cer1* or *cer6* mutations, the tryphine composition and ultrastructure show less or more visible differences with wild type tryphine. Our observations with *cer1* mutants emphasize the hypothesis proposed by Preuss et al. (1993) that long chain lipid molecules, in particular alkanes, are needed in the tryphine layer for proper pollen-pistil signalling. These lipids are produced in the tapetum and as they are probably not needed for proper tapetal functioning, there absence will not have an effect on tapetal appearance, but is only visible as an altered tryphine ultrastructure in the *cer1* and *cer6* mutants. A secondary, may be stress induced, effect of the altered tryphine is the accumulation of callose in stigma papillae in response to *cer1* and *cer6* pollen grains. No callose accumulation was observed when pollinating wild type or *cer1* pistils with non-specific *B. napus* or *Petunia hybrida* pollen and the response thus is not a general mechanism to prevent pollination by other species.

We used a cell layer chimaera to study expression of *CER1* in flowers. One *cer1::I/dSpm89* plant with a fertile revertant inflorescence sector containing phenotypically normal stem wax was found. In plant development, the reproductive cells, including the tapetum, descend from the L2 layer (Goldberg et al., 1993) and the epidermis descends from the L1 layer. We therefore tested offspring from the sector, which was all *cer1* mutant showing that the *cer1::I/dSpm89* to *CER1* reversion did not occur in the L2 layer, but only in the L1 layer. Unexpectedly *CER1* expression in the L1 layer alone was enough to restore a fertile phenotype, inferring that there is some way of transport of the CER1 protein or its products to the tapetum or pollen wall.

The CER1 gene is conserved among plant species

Genes performing steps in such a general plant biosynthetic pathway as for wax synthesis are expected to be found in many plant species. This is indeed the case for the *CER1* gene, as we found homologues from a related Brassicae species, but also from the unrelated Compositae *S. odorus* and from even more distantly related monocot species rice and maize. By comparing the derived amino acid sequences it is clear that all members of the *CER1* gene family we have found, encode for equally divergent homologous proteins. The level of identity between these proteins (between 30 to 70%) is in the same range as the homology found among fatty acid desaturases from different species, irrespective of the position of the double bond they introduce (Iba et al., 1993; Okuley et al., 1994; Sakamoto et al., 1994). Based on this level of identity, the CER1-homologous proteins are presumably all performing a similar function. At least one additional *CER1* family member is present in Arabidopsis. The *CER1-like* gene, as we termed it, is physically (and genetically) linked to *CER1*, with a physical distance of only 1 kb. This *CER1* gene cluster most likely originated from an ancient gene duplication. Since then both genes have evolved to such an extent that the

ATTS1001 partial cDNA clone derived from the *CER1-like* gene is not recognized by the *CER1* cDNA probe upon DNA gel blot hybridization. Also their expression patterns have changed. The *CER1-like* gene is only transcribed in flowers and is therefore not involved in stem wax biosynthesis. A function of the *CER1-like* gene in epicuticular wax biosynthesis on flower petals or sepals is unlikely given the visual absence of wax structures on flower organs, contrasting the high expression of the *CER1-like* gene. By analogy to the range of fatty acid desaturases that use different unsaturated fatty acids as substrate, the CER1-like protein may have a similar function as the CER1 protein in pollen development, for instance in the production of alkenes in tryphine. Alternatively it may be expressed in another part of the flower. The lipid transfer protein *LTP1* is known to be expressed in stigma papillae where it can function in the secretion of stigma lipids (Thoma et al., 1994) and the CER1-like protein may be involved in the production of these lipids.

Conclusions

The isolation of *CER1* is a first step toward the understanding of the biosynthesis of epicuticular wax components. The CER1 protein is part of an integral cell-membrane enzyme which we deduce to be an aldehyde decarbonylase. The presence of histidine rich motifs $HX_{3-4}H$ or $HX_{2-3}HH$, described as iron binding sites in fatty acid desaturases, may help to identify domains of the protein that interact with cofactors, which will provide further knowledge on the biochemical properties of decarbonylases. The identification of *CER1* homologues allows their testing in the manipulation of wax alkane contents in other species. This can be important for influencing drought resistance or insect resistance, for instance in rice, for which drought tolerance is associated with high wax lines, especially rich in C_{29}, C_{33} and C_{35} alkanes (Haque et al., 1992; O'Toole and Cruz, 1983) and for which a higher ratio of long chain to medium long chain aldehydes or alkanes promotes resistance to brown planthoppers (Woodhead and Padgham, 1988). The *CER1* gene is the first gene isolated that is responsible for pollen-pistil interaction in the self-compatible species *Arabidopsis thaliana*, and with it the role of lipids in pollen-pistil signalling can be further analysed. Meanwhile the isolation of other wax biosynthesis genes by transposon tagging in Arabidopsis continues, and we expect it will provide us with more tools to study and modify the wax biosynthesis pathway.

Material and methods

En/Spm-I/dSpm transposon plants and cer mutants

All experiments were carried out with the Landsberg *erecta* ecotype of *Arabidopsis thaliana*, which was also the genetic background of the chemically or physically induced *eceriferum* (*cer*) mutants tested for phenotypic complementation (all mutants provided by M. Koornneef, Wageningen Agricultural University, The Netherlands). For screening, 25 *En/Spm-I/dSpm* transposon tagging lines with 12 plants each, were grown individually in the greenhouse and

examined for *cer* mutations. All lines were obtained after two generations of self-pollination, starting with one plant containing the TEn2 *En/Spm* transposase T-DNA locus along with several transposed *I/dSpm* elements (Aarts et al., 1995a). The original *cer1::I/dSpm89* transposon tagged mutant was found in line H12.1.6.2, containing about 15 different *I/dSpm* elements and homozygous for the TEn2 T-DNA. TEn5 is another *En/Spm* transposase line containing a different, more active T-DNA locus and no other *I/dSpm* elements. This line was crossed to a *cer1::I/dSpm89* plant and *cer1* F_2 plants were screened for excision sectors. All plants grown for progeny were kept in Aracon containers (BetaTech, Gent, Belgium) to prevent cross-pollination. Fertility of *cer* mutants was conditioned by keeping the plants enclosed in a plastic bag to increase relative humidity (Koornneef et al., 1989).

Identification of a cer1-m cosegregating I/dSpm element and isolation of flanking genomic DNAs

The original *cer1-m* mutant was back-crossed to Landsberg *erecta* wild type for two generations. Genomic DNA was isolated from second backcross offspring plants, and tested for the presence of *I/dSpm* elements. All plants were allowed to self and their progeny was tested for segregation of the *cer1* phenotype to confirm linkage of an *I/dSpm* element with the *cer1* phenotype in the second backcross offspring generation. Genomic DNA from plants containing the *cer1*-linked *I/dSpm89* element and a few other unlinked *I/dSpm* elements was used to obtain DNA flanking both sides of *I/dSpm89* after *I/dSpm* specific inverse PCR (IPCR; Masson et al., 1991). Additional PCR amplification using primer T (5'-GAC ACT CCT TAG ATC TTT TCT TGT AGT G-3') fitting both terminal inverted repeats (TIR) of *I/dSpm* enabled the isolation of fragments with minimal transposon DNA. Based on *I/dSpm89* flanking sequences, primers 2 and 3 (5'-GGA GCA TGA GAA TTG CAG ATA CC-3' and 5'-GGC GTC GTC AGG TGA GTT AAG TGC-3') were designed which amplified a 189-bp wild-type DNA fragment covering the *I/dSpm89* insertion site.

cDNA and genomic library screening

An amplified cDNA lambda library representing different Arabidopsis tissues (Newman et al., 1994) and a Landsberg *erecta* genomic library obtained through the Arabidopsis Biological Resource Center (Ohio State University, Columbus, OH, USA) and the European DNA Resource Centre (Max-Delbrück Laboratory, Köln, Germany), were screened with the *I/dSpm89* IPCR fragment probe. The DNA insert of the genomic clone was subcloned as EcoRI fragments.

DNA and RNA analysis

DNA and RNA gel blots were standardly hybridized at 65°C overnight and washed twice at 65°C with 2xSSC, 1% SDS or (more stringently) with 0.1xSSC, 1% SDS. DNA sequences

were determined using an ABI Sequencer. *CER1* cDNA SalI and SalI-XbaI fragments were subcloned in pBluescript SK^+ and sequenced. The double-stranded DNA sequence was completed using cDNA specific primers 1 (5'-GGC CTC CGG CAA TAG GTT GAT G-3'), 4 (5'-GGT GCT TAG TCT GGG TCT CAT G-3'), 5 (5'-CAC AGG AGT GGA CAT TCA CCA GAG-3') and 6 (5'-CGC ATG AGT GTG GCA CAT CCC-3') (Isogen Bioscience, Amsterdam). The same primers as well as primers 2 and 3 flanking *I/dSpm89* were used to test for a new insertion in *CER1* causing the mutant *cer1* sector in combination with the *I/dSpm* TIR primer (T in figure 4). PCR conditions for primers 1 to 6 and T are 5 min at 94°C, followed by 30 cycles of 94°C (30 sec), 55°C (30 sec) and 72°C (3 min.). In addition to the cDNA sequence, a single strand of genomic DNA sequence of 1656 bp upstream of the *CER1* start codon was determined. The *CER1* cDNA sequence data and the determined *CER1* genomic DNA sequence data will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers D64155 and D64156.

The cDNA sequence and the predicted amino acid sequence were analysed using the PC\Gene computer package (IntelliGenetics, Geneva, Switzerland).

Data base searches

GenBank and EST data bases were searched for *CER1* homologues using BLAST programs (Altschul et al., 1990). GenBank accession numbers of the reported homologous sequences are: L33792 for the *Senecio odorus* SOLIPTRB partial cDNA; T22420 and Z18418 for two Arabidopsis ESTs with nearly 100% identity to the *CER1* cDNA; Z25487 for the Arabidopsis ATTS1001 EST; L35835 for the *Brassica campestris* EST; R27543 for the potato EST; T70657 for the maize EST; D15324 and D22308 for two rice ESTs with N-terminal homology to CER1; D40658 and D23996 for two rice ESTs with internal homology. The cDNA clones corresponding with the four rice ESTs have been kindly obtained from Dr. Yoshiaki Nagamura of the Rice Genome Research Program (STAFF Institute, Ibaraki, Japan). The 5' ends of both D15324 and D22308 have been resequenced to correct for frameshifts and other occasional misreadings found in the original data base sequence.

Microscopy

Germination of *cer1-m*, *cer1-1* and wild type pollen was examined by means of bright field and fluorescent microscopy after staining with aniline blue according to Preuss et al. (1993). Revertant excision sectors and mutant insertion sectors on four to five week old stems were examined with a preparation microscope. For a more detailed observation, fresh 1 cm long stem parts were excised and mounted on stubs, closing the excision wounds with rapidly drying cyano-acrylate glue. Subsequently they were transferred to a JEOL 5200 scanning electron microscope, partially dehydrated for 5 min in the vacuum of this microscope and finally photographed at 15 kV. Closing the wounds prevented excessive water loss from the specimens, which would have disturbed the vacuum of the microscope; the (gentle) dehydration step was introduced to better expose the epidermis cell boundaries.

For scanning and transmission electron microscopy of tryphine on pollen grains, freshly opened anthers were fixed in 3% glutardialdehyde in 0.7 M cacodylate buffer for 15 min, several times rinsed in the buffer, postfixed in 1% osmium tetroxide in the same buffer for 15 min and stepwise dehydrated into ethanol 100%, all treatments at room temperature. For scanning electron microscopy these specimens were critical point dried via carbon dioxide, mounted on stubs, coated with platinum and observed in a JEOL 6300 field-emission scanning electron microscope at 5 kV. For transmission electron microscopy the 100% ethanol saturated specimens were stepwise infiltrated with resin (Spurr, 1969), polymerized, ultrathin sectioned, stained with lead citrate and uranyl acetate, and observed in a JEOL 1200 TEM at 80 kV.

Acknowledgements

We thank Maarten Koornneef for sharing his experience with *cer* mutants and for critically reading the manuscript. We thank Christiane Leonards-Schippers for helpful discussions, Jan-Peter Nap for reading the manuscript critically, and Gerrit Stunnenberg and Gerard Scholten for taking care of the plants in the greenhouse. We also thank the members of our ESSA DNA sequencing team for their support in DNA sequencing, and Bauke Ylstra for his advice on the analysis of pollen germination. Our colleagues at the Centrum voor Fotografie en Beeldverwerking SC-DLO are gratefully acknowledged for making the photographs of *cer1* and *CER1* sectors and for their help in preparing the other figures. This work was supported by the European Union BRIDGE/BIOTECH project (grant nr. BIOT-CT90-0207) on 'Development and use of *Arabidopsis thaliana* as a tool for isolating genes of agronomic importance'.

Chapter 5

Random and targeted tagging with *En/Spm-I/dSpm*: towards a mutation machine

Mark G.M. Aarts, Christiane Leonards-Schippers, Krista Kooi, Anne Mulder, Bas te Lintel Hekkert, Hiroyoshi Kubo[1], Anton J.M. Peeters[1], Maarten Koornneef[1] and Andy Pereira.

[1]Department of Genetics, Wageningen Agricultural University, Dreijenlaan 2, 6703 HA Wageningen, The Netherlands.

Summary

The *En/Spm-I/dSpm* two element system has previously been used to generate tagged mutants at random. A preferred characteristic would be to use the system for locus specific mutagenesis or targeted transposon tagging. Increasing the number of *I/dSpm* elements in a population - one potential way to get a higher specific tagging frequency - has been tested, and found to give only a slight increase in the number of tagged mutants. Two more tagged mutants obtained this way are described. A targeted transposon tagging experiment was conducted for which a large F_1 progeny of a cross between a (transposon-free) target mutant and a transposon donor, was screened. Our results show that, to be successful, this approach requires a very large F_1 population when the target locus is not closely linked to the transposon donor locus.

Introduction

After a heterologous *En/Spm-I/dSpm* transposable element system was introduced in Arabidopsis, lines with transposed elements were generated and screened for tagged mutants. This random transposon mutagenesis strategy resulted in the isolation of the *MS2* and *CER1* genes as described previously (Aarts et al., 1993 and 1995b). Random tagging may be beneficial when the objective is to generate tagged mutants of any kind, because in general mutants with various (major) phenotypes are expected to be recovered from screening different transposon containing lines or populations. Random tagging is predominantly used with high copy transposable element systems, such as *dTph1* in Petunia, *Mu* in maize and *Tam3* in *Antirrhinum majus*, that give high mutation frequencies (Gerats et al., 1990; Walbot, 1992). Cloning a tagged gene can sometimes be difficult in these systems due to the large number of elements. Low copy systems like the *En/Spm-I/dSpm* in Arabidopsis, have the advantage that cloning a tagged gene is relatively easy, but the chances of finding the required mutant by random tagging will be lower. To shed light on the spectrum and frequency of tagged mutants that can be expected with the *En/Spm-I/dSpm* element system in random tagging experiments, several large plant populations have been screened for obvious major phenotypes. The characterization of the mutants that were isolated is reported in this chapter.

In contrast to the variety of mutants that can be obtained by random transposon mutagenesis, the research interest is often in tagging only one specific gene or only genes involved in one specific biochemical or developmental pathway. To increase the chance of tagging such genes with a low copy transposon system, an alternative strategy can be used that is based on the tendency of transposable elements to transpose preferentially to nearby sites rather then to distant loci. For instance when the gene of interest has been genetically mapped, it is often most convenient to use a donor plant with a transposable element closely linked to the target gene and screen progeny for insertions in the gene of interest. This strategy is referred to as targeted or directed transposon tagging. In maize, targeted tagging has been used with *Ac/Ds* and *En/Spm-I/dSpm* to generate mutants at approximately 100 times higher frequencies than in a random tagging strategy (Cone, 1994). With the development of heterologous transposon systems in Arabidopsis, the prospect of using targeted transposon tagging for making specific Arabidopsis mutants has become very attractive. The recent cloning of the *FAE1* gene after tagging with *Ac* is the first example of a successful targeted transposon tagging experiment in Arabidopsis (James et al., 1995). In this chapter we discuss the use of the *En/Spm-I/dSpm* system for targeted tagging.

Results

Random tagging: the isolation of two new I/dSpm tagged mutants

The tagged mutants *ms2::I/dSpm* (Aarts et al., 1993) and *cer1::I/dSpm* (Aarts et al., 1995b), were found after screening a relatively small number of progeny derived from well

characterized parental plants. To get insight in the frequency of such random mutations on a larger scale, seven large populations of 4000 M_1 seeds each were sown at high density to produce seven times eight subsets (#1 to #8) of M_2 seed. The seeds were derived from plants H22-3, H25-1, H26-1, H27-2, H32-2, H33-1 and H37-6 (Aarts et al., 1995a), which were selected for having over eight different *I/dSpm* elements. After two subsequent generations of self-pollination, the frequency of homozygous independent *I/dSpm* insertions in the M_2 populations is expected to be considerably higher compared to that in the M_1, which should result in an increased number of tagged mutants. As a test, about 1000 seeds from each of the #1 subsets of the bulked populations were sown at high density in the greenhouse and screened for obvious mutant phenotypes (e.g. variants in size, leaf colour, trichomes, flower development or general appearance). Although aberrant plants were found in all subpopulations (small, stunted, narrow leaves, three cotyledons, yellowish etc), only two clear mutants were recovered. In the subpopulation derived from H32-2, some *glabra* (i.e. trichome-less) mutants were detected. When resowing progeny of the original H32-2 plant, this *gl-m* mutation was found to be segregating in a 3:1 ratio (wild type to *gl-m*). Another obvious phenotype was observed in subpopulation H37-6. About 5-10% of the plants in this population had a variegated leaf anthocyanin pigmentation pattern, suggesting that this *anthocyaninless* or *anl-m* mutation was transposon induced. Other *I/dSpm* tagged mutants which were found by large scale progeny screening, but which will not be described here, are an *abscisic acid insensitive 3 (abi3)*, a *leafy cotyledon 1 (lec1)* (M. Koornneef, H. Kubo and A.J.M. Peeters, unpublished results), a *leafy (lfy)*, a *wilty* (*wil*), and a *sterile apetala* (*sap*) mutant (M. Byzova, unpublished results).

The glabra2 tagged mutant

Trichome-less or *glabra* mutants belong to a group of trichome mutants which also includes mutants disturbed in trichome branching, trichome expansion and the patterning of trichome formation on the epidermis (Hülskamp et al., 1994). The *glabra* mutants found in H32-2 offspring have a clearly reduced number of leaf and stem trichomes. Remaining trichomes are reduced in size (Figure 1). Especially the trichomes near the main axis of the leaf are affected, with nearly normal trichomes on the edge of the leaf. In progeny of *gl-m* mutant plants, revertants with a normal trichome phenotype and number were found. A subset of both revertant and *gl-m* plants was analysed on DNA gel blot for the presence of *I/dSpm* elements. One element was hemizygous or absent in revertant plants and homozygous in nearly all *gl-m* plants, and thus likely to be inserted in the *GL* gene (data not shown). To confirm the identity of this candidate *gl::I/dSpm* element, F_2 plants of a cross between a mutant plant and the wild type, were tested for the presence of *I/dSpm* elements. All *gl-m* F_2 mutants were homozygous for the candidate 6-kb HindIII *I/dSpm* containing fragment (Figure 2). Some of the *gl-m* F_2 plants contained the TEn5 *En/Spm* transposase locus, which induced somatic reversion evidenced by both wild type and mutant leaves and occasionally a few chimaeric leaves. None of the mutants, either with or without TEn5, was without trichomes and this phenotype must be considered to be specific for the mutation.

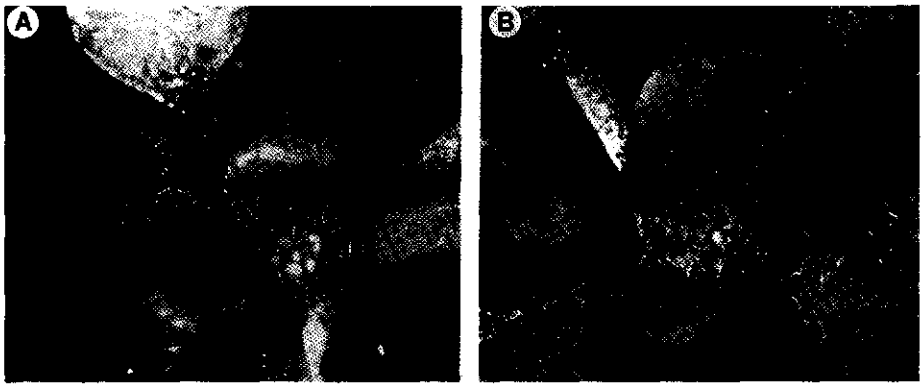

Figure 1: Phenotype of the *I/dSpm* tagged *glabra2* (*gl2*) mutant.
The *gl2::I/dSpm* mutant (*A*) has a reduced number of leaf trichomes and shows reduced branching of remaining trichomes compared to a wild-type revertant plant (*B*).

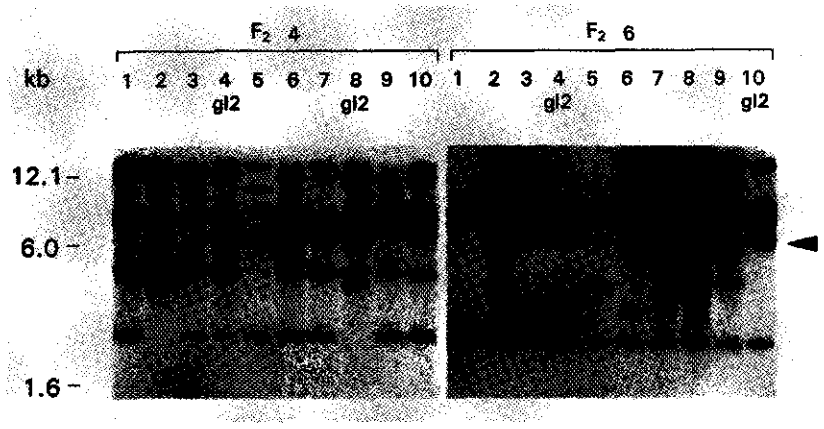

Figure 2: Cosegregation of a 6-kb HindIII fragment along with the *gl2::I/dSpm* mutation.
A blot with HindIII digested DNAs from two F_2 families (*gl2::I/dSpm* x Ler) was hybridized with the *I/dSpm* specific I-LJ probe. Only plants with a homozygous 6-kb HindIII fragment (arrowhead) display the reduced trichome phenotype caused by the *gl2::I/dSpm* mutation. Homozygosity of each *I/dSpm* specific fragment was determined by band intensity comparison with other fragments in the same lane. F_2 plant 6-4, which does not contain the TEn5 *En/Spm* transposase locus any more, clearly shows the difference between the homozygous *gl2::I/dSpm* band, and two other hemizygous *I/dSpm* bands.

The phenotype of the *gl-m* mutant resembled the description of two known Arabidopsis trichome mutants, *glabra2-1* and *glabra3-1*, both of which have a reduced number of trichomes (Hülskamp et al., 1994; Koornneef, 1981). Allelism tests revealed that the *gl-m* mutation was allelic to two *gl2* mutations (*gl2-1* and CS3125) present in the Arabidopsis Biological Resource Center (Columbus, OH, USA) (Scholl and Davis, 1995). The phenotype of mutant CS3125 is slightly different from the phenotype of both *gl2-1* and *gl2::I/dSpm* mutants. The trichomes that are formed appear to be growing within the epidermal layer rather than protruding out of the surface. In the F_1 hybrid between CS3125 and *gl2::I/dSpm* this trichome phenotype can be observed next to leaf areas with normal trichomes due to excision of the *I/dSpm* element from *gl2::I/dSpm*. The phenotype of the hybrid between *gl2-1*

and *gl2::I/dSpm* is similar to *gl2::I/dSpm*, with slightly more and better developed leaf trichomes than *gl2-1*. The *gl2::I/dSpm* allele is thus part of an allelic series, with distinguishable mutant phenotypes.

The *GL2* locus has been mapped to the bottom of chromosome 1, at position 116.7 cM (Hauge et al., 1993). The *GL2* gene has already been isolated by T-DNA tagging (Rerie et al., 1994), and a genomic DNA sequence of the *GL2* gene is present in the GenBank database (acc. no. L32873). From this sequence data it is clear that the gene is covered by two HindIII fragments of 3598 and 1183 bp, plus a fragment of over 985 bp extending beyond the determined DNA sequence. The *gl2::I/dSpm* fragment of around 6 kb, which we found by DNA blot hybridization, corresponds well with insertion of an 2.2-kb *I/dSpm* element in the 3.6-kb HindIII fragment containing the 5' end of the *GL2* gene.

ANL2, a novel regulator of anthocyanin accumulation

The *anthocyaninless* (*anl*) phenotype is especially apparent under stress conditions that induce leaf anthocyanin accumulation. The mutant phenotype is highly variegated, with several hundred excision sectors per leaf (Figure 3). These vary in size from single cells to large parts of the leaf. Most of the anthocyanin mutants known for Arabidopsis, such as the *transparent testa* (*tt*) or *transparent testa glabra* (*ttg*) mutants, lack anthocyanins in general and have yellow seeds due to the lack of seed coat pigmentation (Shirley et al., 1995). In the *anl-m* mutant the seed coats are coloured normally and also anthocyanin production at the trichome base is unaffected. Excision sectors have sharp boundaries between anthocyanin pigmented and non-pigmented cells and the gene is therefore probably not an enzyme in the biochemical pathway (for which intermediate products or substrates can leak to nearby cells), but is more likely encoding a regulatory component of stress response controlling the expression of several genes in the anthocyanin pathway. There is one other *anl* mutant known, which also lacks anthocyanins in leaves and which has normally coloured seeds (M. Koornneef, unpublished results). This EMS induced *anl1* mutant is not allelic to any of the *tt* or *ttg* mutants, nor to *anl-m*. Consequently, the gene disrupted in the *anl-m* mutant is from now on referred to as the *ANL2* gene.

The *anl2::I/dSpm* mutant is genetically recessive. F_2's without TEn5 from a cross with wild type were used for DNA blot analysis to identify an *I/dSpm* element that cosegregated with the mutant phenotype (data not shown). The DNA flanking this *I/dSpm* has been cloned by IPCR and the corresponding genomic DNA and cDNA clones have been isolated. Genetic mapping of these sequences on the Columbia (Col) x Landsberg *erecta* (Ler) recombinant inbred lines (RILs; Lister and Dean, 1993) positions the locus at the top of chromosome 4 (Figure 4). The marker can be easily observed in senescing leaves, and is a useful phenotypic excision marker.

The cDNA corresponding to the *ANL2* locus has been characterized. It encodes a 803 amino acids long open reading frame, part of which shows 60-70% identity to the homeodomain box of the GL2 protein (Rerie et al., 1994). The *ANL2* gene appears to be a

new type of transcriptional regulator of the anthocyanin pathway, as previously identified regulators belong to the *myb* or *myc* class of transcription factors (Shirley et al., 1995). The insert is in an intron, such that excision nearly always results in restoration of the *ANL2* gene-expression. The high reversion frequency therefore illustrates the high excision frequency of the *En/Spm-I/dSpm* transposable element system.

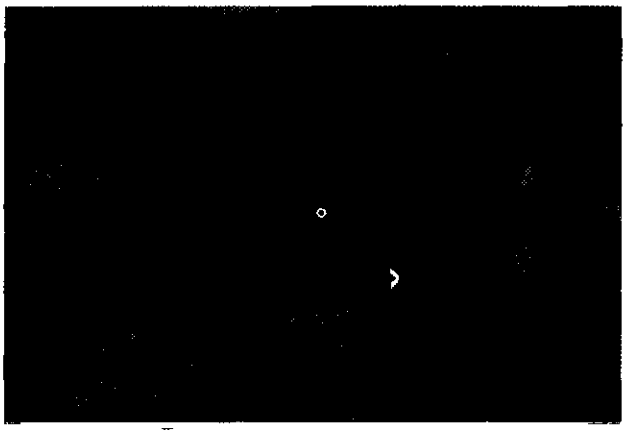

Figure 3: Variegated leaf anthocyanin pigmentation of the *I/dSpm* tagged *anthocyaninless2* (*anl2*) mutant. Senescing leaves have accumulated anthocyanins only in sectors of varying sizes due to excision of the *I/dSpm* insert from one of the *anl2::I/dSpm* alleles. Small sectors of only one cell (>), as well as large sectors covering a considerable leaf area can be seen (o).

Positioning transposed I/dSpm elements on the genetic map of Arabidopsis

In certain cases it is more desirable to use a transposon tagging system for targeted rather than random mutagenesis. With a known tendency for transposition to physically and therefore genetically linked sites (Dooner and Belachew, 1989; Dooner et al., 1994; Peterson, 1970; Tower et al., 1993), the highest insertion frequencies are to be expected using a transposon donor line containing an element linked to the target locus. The prerequisite for employing such a strategy is the availability of mapped *I/dSpm* elements, distributed over the Arabidopsis genome. A first set of mapped *I/dSpm* elements has already been presented (Aarts et al., 1995a). In an ongoing effort more elements have been mapped since then, as is shown in figure 4. Most *I/dSpm* elements have been mapped as RFLPs on the Col x Ler RILs (Lister and Dean, 1993). For some elements for which no RFLP was revealed between both parents, the position was determined by hybridization to Arabidopsis Yeast Artificial Chromosome (YAC) libraries at the John Innes Centre, Norwich, UK (Schmidt et al., 1995). The positions of these elements relative to other genetic markers has not been determined by recombination, but has been estimated after comparison of YAC mapping data to the latest genetic mapping data in the RILs population (Clare Lister, unpublished). The position of elements *I/dSpm 3, 41G, 75, 78, 104, 113* and *114*, genetically mapped to chromosomes 4 and 5, was confirmed by hybridization to the YAC library. Tagged alleles *ap1::I/dSpm* and *tad::I/dSpm* are discussed in Chapter 6.

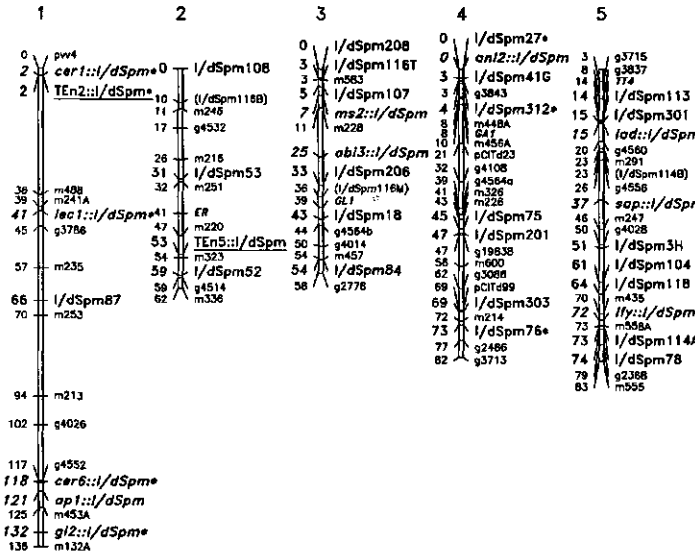

Figure 4: Genetic map positions of *I/dSpm* elements.
The positions of *I/dSpm* elements are given relative to RFLP markers (small case) mapped in a Col x Ler population of recombinant inbred lines (Lister and Dean, 1993). The T-DNA loci TEn2 and TEn5, containing the *En/Spm* transposase genes are underlined. *I/dSpm* insertions in known genes are given in italics. *I/dSpm* elements mapped by hybridization to YAC filters (*I/dSpm27*, *I/dSpm76*, *I/dSpm312*, *cer1::I/dSpm*) (Schmidt et al., 1995; J. Ecker c.s., unpublished) or by comparison of RFLP data on RILs with the combined classical marker-RFLP map (Hauge et al., 1993) (TEn2, *cer6::I/dSpm*, *gl2::I/dSpm*, *lec1::I/dSpm*) are annotated (*). The map positions of loci homologous to the *I/dSpm116* flanking DNA IPCR probe (I/dSpm116M and I/dSpm116B) are shown between brackets.

The lack of RFLPs between Ler and Col ecotypes is a serious problem in assigning a map position to *I/dSpm* elements. As this could be due to the small size of the IPCR derived probes, λ clones with around 15 kb of Ler genomic DNA flanking the insert were isolated for the *I/dSpm56*, *I/dSpm82*, and *cer1::I/dSpm89* inserts and used as probe. For the TEn2 T-DNA insert around 30 kb of flanking DNA was isolated and used as probe. All these regions seem to be very conserved between Col and Ler, as still no RFLP could be detected using a standard set of five restriction enzymes (see Material and methods). Only for *cer1::I/dSpm89* several RFLPs were found, but they mapped to different chromosomes and none of the polymorphic bands corresponded to a band of the λ clone.

Most of the IPCR probes detected single copy genomic DNA fragments and occassionally a weakly hybridizing homologous copy. However, the IPCR clones from *I/dSpm75* and *I/dSpm201*, detected multiple cosegregating bands. These elements both map to a region of chromosome 4, close to downy mildew resistance locus *RPP5* (Parker et al., 1993), which may be rich in duplicated sequences. The IPCR probes for *I/dSpm84* and *I/dSpm118* were Ler specific. The *I/dSpm84* probe detected only a weak hybridizing copy in Col and the *I/dSpm118* probe did not hybridize to Col DNA at all, not even under high stringency. Element *I/dSpm116* is another exceptional case. The IPCR probe detects two DNA copies in Ler, and one in Col, all three mapping at different loci (Figure 4).

Targeted tagging with W7 and W9 mutant lines

With a number of *I/dSpm* elements positioned on the genetic map, a pilot experiment was performed to test the possibility of targeted tagging. In this approach, transposon lines were crossed on a large scale with mutant plants. As most F_1 plants are heterozygous for the recessive mutant allele, their phenotype is wild-type. Only when an *I/dSpm* element has jumped into the wild-type allele, mutant F_1 plants are found. To facilitate large scale crossings, sterile *ms2* plants, derived from a plant containing the mapped elements *I/dSpm76, 78, 84* and *ms2::I/dSpm*, were chosen as the transposon donor parents (see Material and methods). They were crossed to multimarker line NW9 (mutant for *transparent testa glabra (ttg)* and *yellow inflorescence (yi)* alleles, both mapped to chromosome 5; Koornneef et al., 1983). Only one putative *yi* mutant (yi^p) was found among 6600 F_1 seeds screened for *ttg* and *yi*. This plant had narrow yellow-green thick leaves and was almost completely sterile with stunted growth and reduced apical dominance. In a progeny test, only 3 out of 16 plants carried the yi^p phenotype, which was not correlated with the presence of any of the *I/dSpm* elements. Selfing one of the mutant progeny plants showed the same low transmission of the mutant phenotype, with also the original *yi* phenotype segregating (*Yi* to yi^p to *yi* = 17 : 6 : 4). The yi^p mutant shows some similarity to the phenotype of plants containing an extra chromosome 3 (trisomic) (Koornneef and Van der Veen, 1983). Although this could not be confirmed in crosses with chromosome 3 marker lines due to the very poor fertility, it might explain the dominant nature of the yi^p mutant, the poor transmission to the next generation and the high frequency of reversion to normal looking plants.

In another targeted tagging experiment, different stable *ms2* lines containing TEn5 and between 2 to 6 unmapped *I/dSpm* elements, were crossed to NW9 or NW7 (*long hypocotyl 2 (hy2), glabra 1 (gl1)* and *transparent testa 5 (tt5)*, all chromosome 3 mutations; Koornneef et al., 1983). Linkage of the *I/dSpm* elements to any of the target genes was not determined. In total almost 20,000 F_1 plants were screened for *hy2* and *gl1*, but no mutant was found. Another 6600 F_1 plants were screened for *ttg* and *yi*, but again no such mutant was found.

Discussion

Random tagging

The major aim for the random tagging experiments using large bulked populations was to test if this would increase the mutation frequency as a first step towards *En/Spm-I/dSpm* mediated targeted tagging in Arabidopsis. Indeed more *I/dSpm* tagged mutants were found in the M_2 populations, although not at the high frequency as was theoretically expected. Ideally, an M_1 progeny of 4000 from one M_0 plant with at least 5 different *I/dSpm* element insertions, each with an estimated independent transposition frequency of 10%, is expected to harbour around 2000 new independent inserts (5 x 0.10 x 4000). In 1/8 of the total M_1 population (500 plants, giving one M_2 seedlot), 250 new inserts are expected on average. Assuming that most

of these will be late and thus hemizygous in single M_1 plants, mutants are expected to show up in the M_2 population. On the basis that each M_1 plant has contributed equally to each seedlot, there are on average 219 new inserts represented as homozygotes in a population of 1000 M_2 per seedlot (0.5 (chance of an M_1 plant having a new insert) x 7/16 (chance of having at least one plant with a homozygous insert when sowing 2 M_2 per M_1 plant) x 1000 M_2). With an estimated frequency of 2.4% of all inserts which reveal a mutant phenotype (as can be concluded from experiments with *Ac/Ds* (Bancroft et al., 1993; Long et al., 1993a)), there are on average 5.3 different mutants expected per seedlot, or 37 with seven seedlots tested. In reality only 2 mutants were found, which is around 20 times lower than expected. This discrepancy between reality and expectation is due to several factors. The main factor is probably that the independent transposition frequency of 10% was determined based on a small subset of all offspring from a plant (Aarts et al., 1995a; chapter 1). This estimate of the overall transposition frequency appeared adequate to measure transposition in progeny of up to a few hundred plants, but extrapolation of this estimate to nearly the whole offspring, as for the M_1 population, may well have led to an overestimatiation of the number of new and independent insertions. Another factor is that by sowing and harvesting in bulk there is a strong selection for large and healthy M_1 plants producing many seeds. As many mutations are likely to have a negative influence on the viability of a plant, there is also a selection against plants with insertions in genes, sometimes even in a hemizygous state. The composition of the M_2 seeds is therefore probably not equally distributed over the M_1 plants. A third factor is that the M_2 population was sown at a high density for screening, which may have caused that mutants with weak or subtle phenotypes have not been detected. A fourth factor is that data for the frequency of visible mutants found by *I/dSpm* tagging are lacking. Comparison with the *Ac/Ds* system has provided only a rough estimate of this frequency.

Taken all this into account it is not remarkable to find only two tagged mutants. Essential for future large scale random tagging experiments is to make sure that the mutated population is as diverse as possible. An equal sampling of the M_2 population over the M_1 plants and screening a large enough M_2 are ways to improve the mutation frequency. Alternatively, it may be advantageous to harvest M_2 seed per individual M_1 and test small (10-15) samples per line and not at the high density as was used in the experiments described here. This would allow the detection of weak and retarded mutants that are rapidly overgrown when sown in bulk. Note that in these experiments no positive selection was applied. When selecting for reduced seed dormancy, it has been possible to isolate tagged mutants at the *ABI3* and *LEC1* loci, from similar populations (M. Koornneef, H. Kubo and A.J.M. Peeters, unpublished results).

An important aspect of every transposon system is the opportunity to tag any gene, whatever location, size or function. The mutants found so far represent genes of various sizes, located on different chromosomes. Remarkable is that most of the tagged genes are expressed (although not all exclusively) in inflorescences: *ms2*, *ap1*, *lad*, *cer1*, *sap*, *lfy*, *wil*. This may be because the mutants have very obvious phenotypes and are easy to distinguish, but other easily detectable phenotypes, like chlorophyll mutants, have hardly been found.

Could it be that there is a preferential mutation of flower specific genes? It is known that transposable elements in maize prefer to jump to genetically and physically linked sites (Dooner and Belachew, 1989; Dooner et al., 1994; Peterson, 1970) and preliminary results have confirmed this tendency for the *En/Spm-I/dSpm* system in Arabidopsis (Aarts et al., 1995a). Linked sites may not only be favoured due to their close vicinity, but also due to a similar DNA conformation. Transposition requires the accessibility of transposase proteins to the transposon termini, which is more plausible when the DNA is in an open, instead of a closed conformation. Such open conformations exist during DNA replication or during DNA transcription, when proteins involved in DNA replication or transcription need access to the DNA molecule. Although speculative, this may be an explanation for the high frequency of tagging flower specific genes. Of course other genes can be tagged too, either by random tagging, like the *gl2* and *anl2* mutants, or when selected for, like the *abi3* and the *lec1* mutants. These inserts can be the source for spreading of *I/dSpm* insertions to regions with other DNA conformations.

Mapping of I/dSpm elements

Mapping of transposed *I/dSpm* elements is beneficial for targeted tagging, for which plants with elements linked to a target locus are chosen as transposon donors. In total 37 different *I/dSpm* elements have been placed on the genetic map of Arabidopsis. Plants containing one or a few of these elements are available for targeted tagging experiments. Especially valuable are the tagged and mapped mutants at the *ANL2*, *AP1*, *GL2*, *LFY* and *MS2* loci, which revert at relatively high frequencies resulting in selectable plants with a higher frequency of *I/dSpm* elements in the vicinity of the tagged gene.

Most IPCR probes, containing transposable element flanking DNA, detect single copy genomic DNA fragments. Only few, like *I/dSpm75, 116* and *201* detect several copies. *I/dSpm116* detects different loci in Ler and Col, each with a specific allele not present in the other ecotype. Sequence analysis has to show if the probe contains a low copy repetitive sequence, such as an endogenous (retro)transposon (Peleman et al., 1991; Tsay et al., 1993), which can explain the ecotype specific distribution of these loci. A similar phenomenon was observed when hybridizing the *I/dSpm118* IPCR probe. This probe is Ler specific and has no corresponding Col allele. Sequence analysis of the IPCR sequence did not reveal any significant DNA or protein homology.

Targeted tagging

The prerequisites for targeted tagging (mapped *I/dSpm* elements close to well defined marker genes) are available. Still, the targeted tagging experiments described here were not successful. Heterologous targeted tagging has been reported using *Ac-Ds* based systems in tomato, Arabidopsis and flax (James et al., 1995; Jones et al., 1994; Lawrence et al., 1995). From those experiments, it appeared that the tagging frequency decreased linearly by about 1 in 1000 for every cM distance between the target locus and transposon donor locus. With

the endogenous systems of maize it is often very difficult to determine a targeted tagging frequency, due to the abundance of transposable elements. For the *Spm* system, Nelson and Klein (1984) have described the targeted tagging of the *Bronze* (*Bz*) locus, using flanking tagged alleles at the *C* and *Wx* loci carrying an autonomous and a non-autonomous element at 5, respectively, 25 cM from *Bz*. Only when using the transposon donor as male parent in a test cross, they found one non-autonomously tagged *bz-m* kernel out of 4582 kernels. This would indicate a higher frequency in maize, but with only one tagged mutant found, the statistical significance of this estimate is low.

In the first of the two *En/Spm-I/dSpm* targeted tagging experiments described here, the *YI* target gene was within 10 cM of *I/dSpm78*. The estimated frequency of tagging this gene would be around 1 in 10,000. Testing of only 6600 progeny was therefore not sufficient to find a tagged mutant. In the second targeted tagging experiment, when *HY2*, *GL1*, *TTG* and *YI* were the targets, the distance between target and donor loci was unknown. Although more F_1 plants have been tested than in the first experiment, again no mutants were found, despite the relatively easy scorable mutant phenotypes of tagged genes. In all these targeted tagging experiments, the transposon donor was used as female in the cross. Nelson and Klein (1984) could not find any tagged *bz* alleles when using the transposon donor as female in the cross, although they screened over 122,000 kernels, compared to 4582 in the reciprocal cross. The lack of tagged *yi* mutants, as we have found, may therefore also be caused by a reduced transposition frequency in the female germ line.

In future targeted tagging experiments it will be wise to use plants hemizygous for the TEn5 *En/Spm*-transposase T-DNA locus and containing at least one well characterized homozygous *I/dSpm* element linked to the target locus. For the F_1 tagging approach, the transposon genotype should be represented by as many different flowers as possible, to increase the chance of including flowers from somatic sectors with an insertion in the target gene. As more tagged genes become available, there is an increasing chance of obtaining a tagged gene linked to the target gene. A linked tagged gene can be used in an *in vivo* excision assay, to enrich the target locus for transposed *I/dSpm* elements. Thus, revertant sectors or independent germinal revertants can be selected and chosen as a transposon donor in a cross to the target mutant. This approach has recently resulted in tagging and cloning of the *CER6* gene, for which an F_1 was screened, made by crossing *cer6* mutants to revertant sectors on plants containing the linked *ap1::I/dSpm* allele (A. Pereira, unpublished results).

For the production of a large F_1 for targeted tagging, it is advisable to use male sterility in one of the parents. When such is not available it may be easier to use somatic or germinal revertants as the M_1 population and their progeny as the M_2. A large population of reversions provides in theory a large pool of *I/dSpm* element inserts in the target region of the genome. For example, when assuming an independent *Ac-Ds* excision frequency of 10% and linked transposition for 40% of the transposed elements (Greenblatt, 1984; Dooner and Belachew, 1989; Osborne et al, 1991; Healy et al., 1993), a 1 in 1000 tagging frequency per cM (as was established in an *Ac/Ds* targeted tagging experiment (Jones et al., 1994)), was based on only 40 new local insertions. Comparatively, 40 insertions can be obtained from 200

independent *I/dSpm* element excision sectors, assuming that 20% of the excised elements transpose to linked sites. With a distance of 10 cM between target locus and mutant transposon donor locus on average 400 independent insertions are needed. When such populations are made, a local 'mutation machine' can be created that can be used to tag and isolate several genes in the region.

This can be extended to the whole genome. To get a suitable coverage of all chromosomes about 100,000 different inserts will be needed. To fit this in a manageable greenhouse population of around 5000 plants, 20 different inserts per plant are needed. A way to achieve this, is to accumulate about 40-50 transposons per plant by recurrent crosses for a few generations and selection for families homozygous for the TEn5 *En/Spm* transposase locus while carrying a large number of transposon inserts. The necessary diversity in transposon distribution can be obtained by starting with about 100 plants to produce 50 progeny each, after which this population of 5000 plants is propagated by single seed descent. When assuming a transposition frequency of 10% and 40 transposons to start with, about seven generations are needed to give on average 20 new inserts per plant, and 100,000 new inserts per population. Such a transposon mutagenized population would be a true 'mutation machine'.

Material and methods

Plant material

The parental plants providing seed for the large scale bulked tagging population were all T_3 plants containing the *En/Spm-I/dSpm* two element transposon tagging system (Aarts et al., 1995a), selected for a relatively high *I/dSpm* copy number. Plants H22-3 and H25-1 were homozygous for the TEn2 and TEn5 *En/Spm* transposase T-DNA loci (Aarts et al., 1995a), plants H26-1 and H32-2 homozygous for TEn5 alone, plant H27-1 homozygous for TEn2 alone, plant H33-1 hemizygous for TEn2 alone and plant H37-6 hemizygous for TEn5 alone. Of each selection, 4000 seeds were sown and grown at high density (4 plants/cm^2) to a plant population with transposing and transposed *I/dSpm* elements. Before harvesting, dry plants were divided in eight groups of around 500 plants and all seeds per plant were harvested per group in bulks labelled #1 to #8.

The transposon donors used for the first targeted tagging experiment were male sterile F_4 progeny descending from *ms2* F_2 plant 8-8 (Aarts et al., 1993). This F_2 plant contained mapped *I/dSpm* elements *76*, *78*, *84* and *ms2::I/dSpm*. For the second experiment male sterile (*ms2*) F_4 progeny descending from *ms2* F_2 plant 9-5 (Aarts et al., 1993) containing between two to six unmapped *I/dSpm* elements were used.

For complementation analysis of *gl2::I/dSpm*, the mutant was crossed to *gl2-1* and *gl3-1*, as well as to uncharacterized trichome mutants CS3125, CS3289 and CS3355 present in the Arabidopsis Biological Resource Center (Columbus, OH, USA) (Scholl and Davis, 1995).

The lines used for the isolation and mapping of *I/dSpm* element flanking DNAs are coming through different lineages from the original primary transformants (more specific information

is available on request). In comparison to figure 7 in chapter 1, some *I/dSpm* elements are not shown any more, as they were not transmitted to the next generation and are as such not any longer present in the population. The *I/dSpm* element IPCR DNA probes were first tested by DNA blot analysis for their ability to reveal polymorphisms between Ler and Col, standardly restricted with BglII, DraI, EcoRI, EcoRV or HindIII. Mapping was performed using the JoinMap program (Stam, 1993), with data from at least 50, but preferably all 98 RI-lines distributed through the Nottingham Arabidopsis Stock Centre. The map was drawn using the DrawMap program developed by Johan van Ooijen (CPRO-DLO).

DNA analysis

The techniques used for DNA analysis, and the genomic library used to isolate long fragments of *I/dSpm* element flanking DNA were all as has been described in chapter 4.

Acknowledgements

We want to thank the Arabidopsis Biological Resource Center (Columbus, OH) for providing the *gl* mutants, Renate Schmidt for mapping some *I/dSpm* elements by YAC hybridization and Gerrit Stunnenberg c.s. for taking care of the plants in the greenhouse.

Chapter 6

Expression of regulatory genes controlled by the *En/Spm-I/dSpm* transposon system in Arabidopsis

Mark G.M. Aarts, Bas te Lintel Hekkert, Maarten Koornneef[1], Willem J. Stiekema and Andy Pereira

[1]Department of Genetics, Wageningen Agricultural University, Dreijenlaan 2, 6703 HA Wageningen, The Netherlands.

Summary

The maize *En* or *Spm* transposable element system is known for its ability to induce transposase controlled gene expression of *I/dSpm* tagged mutant genes. A set of *I/dSpm* tagged mutants in Arabidopsis was searched for *En/Spm* controlled gene expression. An *En/Spm* suppressible and an *En/Spm* dependent *I/dSpm* mutant have been found and characterized. Both showed different phenotypes in the presence compared to the absence of *En/Spm* transposase protein TNPA. The interaction of the two tagged genes with the *En/Spm* transposase resembles that found for *En/Spm* regulated tagged mutants in maize. The possibility to induce *En/Spm* suppressible or dependent mutations in the heterologous species *Arabidopsis thaliana* offers new opportunities to study *En/Spm* regulated gene expression.

Introduction

The *En/Spm* transposable element family in maize frequently induces tagged mutants in which expression of the tagged gene is influenced by the presence or absence of an active so called controlling *En* or *Spm* element (Kim et al., 1987; Menssen et al., 1990; Schwarz-Sommer et al., 1987). With the increasing number of *I/dSpm* tagged mutants found in Arabidopsis (see Chapter 5), there is a good possibility to find such type of mutants in a heterologous species. For most of the maize mutants controlled by *En/Spm*, the expression of a tagged gene is suppressed in the presence of transposase proteins, but nearly normal in the absence of transposase proteins. A proposed explanation is that in the latter case, the insert can be spliced from the pre-mRNA, but in the presence of the transposase proteins the element's ends are bound and thus splicing is prevented. On the other hand, there is a tagged maize mutant in which the tagged allele (*a1-m2)* is only expressed in the presence of transposase proteins, while the plant has a mutant phenotype in the absence of transposase. In this mutant, the insertion is in the promoter upstream of the CAAT-box. Gene-activity is explained by the presence of transposase protein binding sites at the element's ends, replacing the function of upstream, but separated, transcription factor binding sites, by binding of transposase proteins and transcription factors (Schwarz-Sommer et al., 1987).

When the collection of *I/dSpm* tagged Arabidopsis mutants was screened, two were found that showed different F_2 segregation ratios in the presence compared to the absence of transposase genes, as is expected for mutations influenced by controlling elements. An *En/Spm* suppressible and an *En/Spm* dependent mutant have been characterized and the phenomenon of controlling elements, described here for the first time for a heterologous transposable element system, is discussed.

Results

En/Spm transposase suppressible expression of a mutant ap1::I/dSpm gene

A set of nine *I/dSpm* tagged mutants (*anl2, gl2, lfy* and *sap* (Chapter 5), *cer1* (Aarts et al., 1995b), *cer6* (A. Pereira, unpublished results), *ms2* (Aarts et al., 1993), *ap1* and *lad*), was examined for differences in phenotype other than variegation, in the presence or absence of *En/Spm* transposase. Two mutants with such properties were selected, which are a new mutant allele at the *APETALA1* (*AP1*) locus, and a mutant allele at the novel *Late Anther Dehiscence* (*LAD*) locus. The *AP1* locus has previously been mapped to position 99.5 of chromosome 1 (Hauge et al., 1993). The *AP1* gene has already been cloned and characterized as a MADS box type of transcriptional regulator gene (Mandel et al., 1992). Allelism of this *ap1* mutant was determined by an allelism test with the *ap1-1* mutant. To confirm the insertion of an *I/dSpm* element in the *AP1* gene, F_2 progeny from a cross with wild type was analysed by DNA blot analysis. After hybridization with an *I/dSpm* element specific probe, a homozygous *I/dSpm* element was found to cosegregate with the *ap1* phenotype. On hybridizing the same blot with the complete *AP1* cDNA, this cosegregating band was again

detected, while plants with wild-type phenotype contained a wild-type *AP1* fragment without the 2.2-kb *I/dSpm* insert. By comparative hybridization analysis, using a 9-kb genomic DNA clone comprising the *AP1* gene as probe to DNA blots from plants with and without the *ap1::I/dSpm* allele, the insert was localized on a 920-bp DraI fragment (Figure 1). This DraI fragment contains part of the *AP1* gene from within the first intron to 55 bp into the third intron. With a primer specific to the DraI fragment and the *I/dSpm* specific TIR primer (Aarts et al., 1995a), a genomic fragment was amplified from a homozygous *ap1::I/dSpm* plant by PCR. When sequenced, the *I/dSpm* element was found to be inserted in the first intron, 580 bp upstream from the start of the second exon (Mandel et al., 1992). The distance from the end of the first exon could not be determined due to lack of sequence information. The *I/dSpm* element is inserted in anti-parallel orientation when comparing the *AP1* mRNA transcription to the transcriptional direction of the *En/Spm* element from which the *I/dSpm* element was derived.

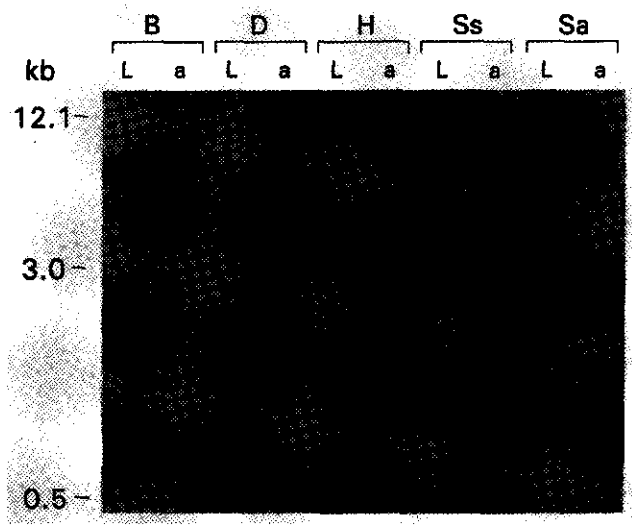

Figure 1: DNA blot analysis showing insertion of an *I/dSpm* element the *AP1* gene.
Blot containing DNA from a wild-type (L) plant and a stable apetala F_2 plant (a) from the cross *ap1::I/dSpm* x Ler, arranged in sets of two and digested with BglII (B), DraI (D), HincII (H), SstI (Ss) and SalI (Sa), was hybridized with a radiolabelled 920-bp DraI subfragment of a 9-kb genomic DNA clone comprising the *AP1* gene. This confirms insertion of an *I/dSpm* element in the *AP1* gene, especially in the 920-bp DraI fragment. The difference between the "L" fragment size and the "a" fragment size is often less than 2.2 kb (the size of the *I/dSpm* element), due to the presence of restriction sites in the *I/dSpm* element.

Reversion of the apetala to the wild-type phenotype is a useful marker to screen for excision of the *I/dSpm* element from the *ap1::I/dSpm* allele. In the presence of the TEn5 *En/Spm* transposase T-DNA, the *ap1::I/dSpm* allele reverts both germinally and somatically to wild type. Somatic reversion is characterized by sectors of wild-type flowers or with a phenotype intermediate between wild type and apetala. The intermediate phenotype may be caused by an excision in only one of the three developmental layers (L1, L2 or L3). Every *ap1::I/dSpm* plant showing somatic reversion, has in general between three to six wild type inflorescence

sectors of varying sizes.

The *ap1::I/dSpm* allele is not only of interest as a useful excision marker, but also for studying the influence of *En/Spm* transposase proteins on the expression of an *I/dSpm* tagged allele. Some of the mutant plants in the progeny of a self-fertilized *ap1::I/dSpm* plant (hemizygous for TEn5) showed a much weaker phenotype than its parent. Although clearly apetala at the bolting stage, this weak mutant phenotype gradually disappeared during the next 7 days of flowering. Just prior to bolting, the visible flower buds are open (as in the EMS induced *ap1-1* mutant), rather than closed (as in the wild type). Two days later, the youngest flowers develop petals, although narrower and often not as white as in wild-type plants. Another four days later, the open flowers are nearly normal, but flower buds are still more *ap1-1* like (Figure 2). The appearance of open buds is maintained throughout flowering. At the end of flowering, the only character reminiscent of the *ap1* mutation is the presence of bract-like sepals at the base of the mature fruit, as described for most *ap1* alleles (Bowman et al., 1993).

The weak apetala phenotype was correlated with the absence of the TEn5 locus. In the presence of TEn5, the expression of the *AP1* gene is suppressed and a strong apetala phenotype is displayed: sepals are bract-like, petals are missing and secondary flowers are frequently formed from axillary buds between sepals and stamen (Figure 2). As mentioned before, this phenotype is unstable, with wild type revertant sectors appearing due to excision of the *I/dSpm* element. Reconstitution of the suppressible phenotype was obtained by crossing a homozygous *ap1::I/dSpm* plant without transposase to a wild-type plant homozygous for TEn5. As a control the same *ap1::I/dSpm* plant was crossed to the Landsberg *erecta* (Ler) ecotype without transposons or transposase. In the F_2 with TEn5, both the strong and weak apetala phenotypes are segregating (Table 1). All apetala plants are homozygous for the *ap1::I/dSpm* allele. All and only the strong apetala plants contain the TEn5 locus. In the control cross to Ler, only the weak apetala phenotype segregated (Table 1). By DNA blot hybridization all and only apetala plants were found to be homozygous for the *ap1::I/dSpm* allele.

The TEn5 T-DNA locus contains two transposase genes which produce the TNPA and TNPD proteins needed for transposition. The TNPA protein has been shown before to be involved in *En/Spm* regulated expression of tagged genes (Kim et al., 1987; Menssen et al., 1990; Schwarz-Sommer et al., 1987). To see if the transposase suppressible effect could be induced by the expression of TNPA alone, the same *ap1::I/dSpm* plant as used before was crossed to a plant homozygous for the Tnp2 T-DNA construct (see Material and methods). This plant expresses the TNPA protein from the 2' promoter, and activity of this protein alone is not sufficient to induce *I/dSpm* element transposition. In the F_2, both the strong and the weak apetala phenotypes were segregating like in the F_2 with TEn5, as is expected when the suppressible effect relies on expression of TNPA (Table 1). All and only plants with an apetala phenotype are homozygous for the *ap1::I/dSpm* allele. Of these only the plants with a strong apetala phenotype contain the Tnp2 locus, confirming the influence of TNPA expression on the apetala phenotype.

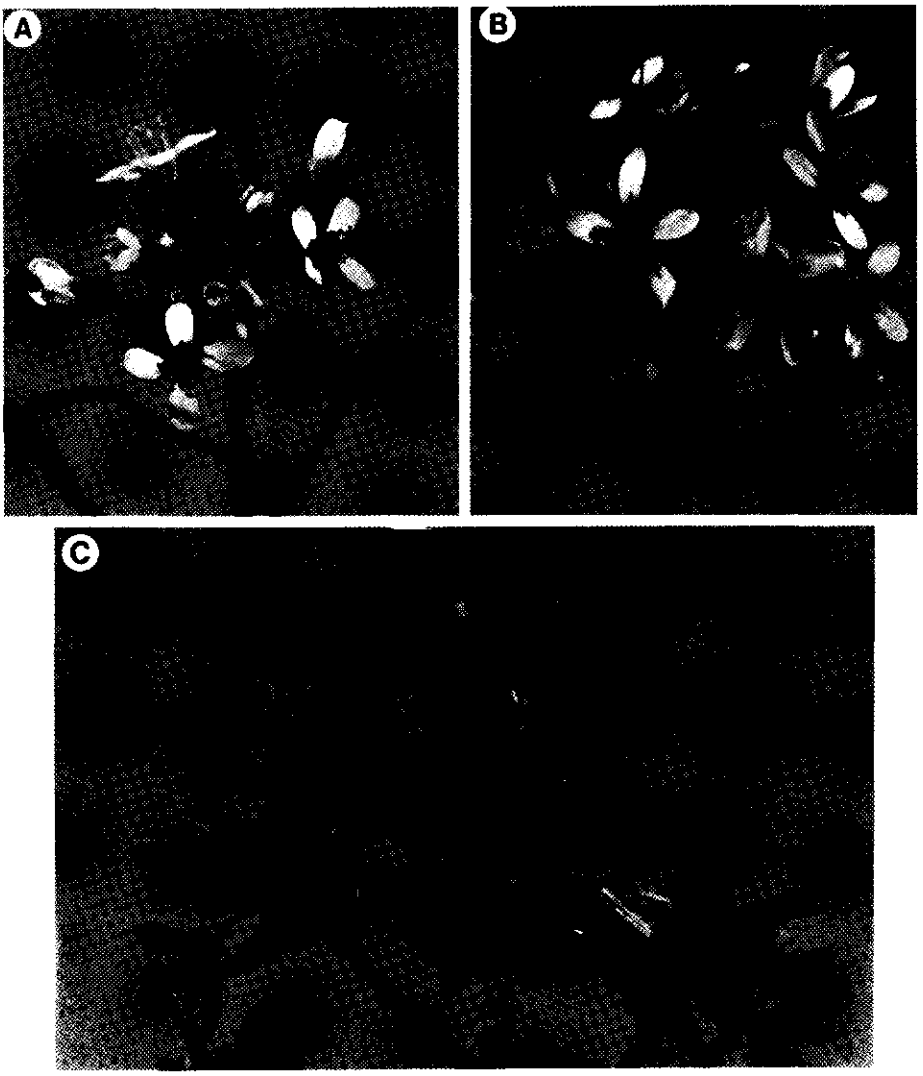

Figure 2: Phenotypes of weak and strong *ap1::I/dSpm* mutant inflorescences compared to wild type Landsberg *erecta*.

A. Wild-type inflorescence.

B. Weak *ap1::I/dSpm* mutant inflorescence in the absence of *En/Spm* transposase. Note the bract-like sepals, the open flower-type and the prematurely opened flower buds.

C. Strong *ap1::I/dSpm* mutant inflorescence in the presence of *En/Spm*-transposase. Sepals are converted to bracts, petals are absent, secondary flowers are formed in the axils of sepals and all flower buds are opened prematurely.

Table 1: Segregation ratios and χ^2 values[1] of several F_2 populations derived from *lad* and *ap1* mutants crossed to either wild type (Ler), or plants homozygous for TEn5 (TEn5) or homozygous for Tnp2 (Tnp2).

	mutant parent								
	ap1::I/dSpm[2]			*lad::I/dSpm*			*lad*		
	wt:*ap1*^{s}:*ap1*^{w}	$\chi^2_{(3:1)}$	$\chi^2_{(15:1)}$	wt:*lad*	$\chi^2_{(3:1)}$	$\chi^2_{(15:1)}$	wt:*lad*	$\chi^2_{(3:0:1)}$	$\chi^2_{(12:3:1)}$
crossed to:									
Ler	34:0:14	0.44	49**	20:4	0.89	4.4*	18:6	0	14.4**
TEn5	40:4:2	-	3.6	46:2	11.1**	0.36	25:9	0.039	24**
Tnp2	37:6:5	-	2.4	42:6	4.0*	3.2	36:12	0	28**

[1] χ^2 values indicate that the observed segregation ratio (w.t.:mutant) is significantly different from the tested ratio when $p<0.05$ (*) or when $p<0.01$ (**).
[2] *ap1*^{s}: strong apetala phenotype, *ap1*^{w}: weak apetala phenotype.

En/Spm transposase dependent expression of a mutant lad::I/dSpm gene

The *lad* mutant has a male sterile phenotype, characterized by the lack of silique-elongation after selfing, but not after pollinating with wild-type pollen. Pollen is formed, but due to a delay in the opening of mutant anthers, pollen arrives too late on the stigma surface to be functional for fertilization. Pollen that is manually removed from unopened mutant anthers germinates on the stigma surface, resulting in fertilization and nearly normal seed set. The male sterile phenotype is most prominent in the first 10-15 flowers of an inflorescence, with later flowers yielding increasingly more seeds. After self-pollination, *lad* plants produce only several hundreds of seeds, compared to several thousands for wild-type plants under similar conditions.

Around 45% of the initially tested progeny derived from plant H21-6 (Aarts et al., 1995a), displayed the male sterile phenotype. This is rather different from 100% male steriles expected when H21-6 was homozygous for a recessive mutant allele, or 25% male steriles when H21-6 was heterozygous for the mutant allele. Plant H21-6 did not show a male sterile phenotype nor did it produce notably less seed so that chimaerism of wild-type and mutant tissue is unlikely to exclusively account for this aberrant segregation ratio. To find a proper explanation, we first searched for a relation between the presence of an *I/dSpm* element and the mutation. From the H21-6 progeny, 12 fertile and 12 male sterile plants were analysed by DNA blot hybridization. This did not reveal a correlation of the mutant phenotype with the presence or absence of any of the detected *I/dSpm* elements. However, when the same set of plants was assayed for the presence of TEn5 and both results were combined, one element, *I/dSpm207*, turned out to be absent in male steriles containing the TEn5 locus, and present in male steriles without TEn5. When a *lad* plant homozygous for *I/dSpm207* and without TEn5 was crossed to Ler, the mutation cosegregated in the F_2 with the homozygous presence of *I/dSpm207* in a 20 to 4 ratio (w.t.:*lad*). This ratio is not significantly different from a Mendelian 3:1 ratio for a monogenic recessive mutation (Table 1). The presence of

the TEn5 locus thus appears to be an additional factor influencing the Mendelian segregation of male sterility. If this is true, the mutation should show a different F_2 segregation in the presence of the TEn5 locus. Indeed, when a TEn5 *En/Spm* transposase plant was crossed with the same *lad* plant as used previously (homozygous for *I/dSpm207*, no TEn5), the F_2 segregated only 2 male steriles among 48 plants, which is significantly different from a 3:1 ratio for w.t.:*lad* (Table 1). Out of 36 F_2 plants tested by DNA blot analysis, two male steriles were found to be homozygous for the *I/dSpm207* insert but without TEn5, while two plants homozygous for *I/dSpm207* and containing TEn5 did not show the male sterile phenotype. It means that for homozygous *I/dSpm207* plants, the mutant phenotype is only displayed in the absence of the TEn5 locus. As the mutant phenotype depends on two recessive alleles (*lad::I/dSpm* and absence of TEn5), a 15:1 segregation ratio is expected in the F_2. The observed segregation ratio is indeed not significantly different from a 15:1 ratio (Table 1).

To test the influence of TNPA on expression of the *I/dSpm207* tagged *LAD* gene, the same *lad* mutant plant as used before, was crossed to plants containing the Tnp2 T-DNA construct. Among 48 F_2 plants, 6 male steriles were found, which is again significantly different from a 3:1 segregation ratio but not from a 15:1 ratio (Table 1). When 24 of these F_2 plants were analysed by DNA blot hybridization, all male steriles were again homozygous for *I/dSpm207* and without Tnp2, while three plants homozygous for *I/dSpm207*, but containing the Tnp2, showed a fertile phenotype. As controls, the same crosses were performed with a plant containing a homozygous *lad* allele from which the *I/dSpm207* element was excised. Expression of TNPA has no influence on this allele, as the segregation ratios observed in all three F_2 s were not significantly different from a single locus 3:1 ratio (Table 1).

The restoration of fertility in homozygous *I/dSpm207* plants containing transposase (TEn5 or Tnp2) is not 100% compared to the wild type. When the average seed set was determined, homozygous *lad::I/dSpm207* plants containing transposase produced 27 ($\pm$ 7) seeds per silique, while wild-type plants or plants hemizygous for the *lad::I/dSpm207* allele produced 50 ($\pm$ 5) seeds per silique. There was no notable difference between plants containing TEn5 or Tnp2. For comparison, homozygous *lad::I/dSpm207* plants without transposase or homozygous *lad* plants without *I/dSpm207*, produced 6 ($\pm$ 2) seeds per silique.

The DNA flanking *I/dSpm207* was isolated by IPCR and with the resulting probe the *LAD* locus was genetically mapped in the Col x Ler RILs (Lister and Dean, 1993) to chromosome 5, 1 cM below *TT4* and 5 cM above RFLP marker g4560 (see Chapter 5). Preliminary sequence data indicate that the element is inserted in anti-parallel orientation in a long reading frame, part of which shows strong homology to the helix-loop-helix domain characteristic of a *myc*-like transcription factor.

Discussion

Among all *I/dSpm* tagged mutants found so far, the *ap1::I/dSpm* and the *lad::I/dSpm* mutants are remarkable, in that they show suppression of the mutant phenotype either in the presence

(*lad::I/dSpm207*) or absence (*ap1::I/dSpm*) of the *En/Spm* transposase TNPA protein. Such phenomena have not been described before for any of the heterologous transposon tagging systems, but they are well known for *En/Spm* controlling element mutations in maize. *En/Spm* suppressible or dependent alleles have been studied in detail for the *A1* (*a1-m1 5719* allele (Schwarz-Sommer et al., 1985a); *a1-m1 1112* allele (Tacke et al., 1986); *a1-m2 8004* allele (Schwarz-Sommer et al., 1987)), *A2* (*a2-m1* allele (Menssen et al., 1990)) and *Bz* loci (*bz-m13* allele (Kim et al., 1987; Schiefelbein et al., 1985)). All three loci are involved in anthocyanin pigmentation and their expression can be monitored by examining aleurone colour.

The *a1-m1*, *a2-m1* and *bz-m13* alleles carry an *I/dSpm* element inserted anti-parallel (comparing the original direction of *En/Spm* transcription in *I/dSpm* with that of the tagged genes) in the transcribed part of these genes (Kim et al., 1987; Menssen et al., 1990; Schwarz-Sommer et al., 1985a; Tacke et al., 1986). In the absence of any *En/Spm* element the aleurone in kernels of all mutants is nearly full-coloured. In the presence of *En/Spm* the phenotype of the aleurone is suppressed to colourless in *a1-m1* and *a2-m1* kernels and to bronze in *bz-m13* kernels. This *En/Spm* dependent colour change is caused by splicing of the *I/dSpm* element from the pre-mRNA transcript. The *A2* gene itself is without introns, but *I/dSpm* insertion has created new splice sites at the element's termini, leading to a nearly wild-type mRNA encoding an additional seven amino acids (Menssen et al., 1990). The *a1-m1* insert is 1 bp upstream of an exon-intron boundary, but the element can be spliced using a new exon-intron boundary 84 bp downstream of the insertion site (Tacke et al., 1986). For the *bz-m13* allele, alternative splicing uses a splice donor site from an intron upstream of the insert to a new splice acceptor site in the terminal inverted repeat at the downstream end of the insert (Kim et al., 1987). These three maize alleles demonstrate the Suppressor function of *En/Spm*, which suppresses the activity of a tagged gene by binding of the TNPA protein to (one of) the termini of the element, thus preventing splicing of the element from the pre-mRNA (Grant et al., 1990).

The effect of the *ap1::I/dSpm* insert resembles the effect of 'suppressible' alleles. Similar as was found for the maize alleles, the *I/dSpm* is inserted in opposite orientation (anti-parallel) to the *AP1* gene. In the *anl2::I/dSpm* mutant (see Chapter 5) the *I/dSpm* element is also inserted in an intron, but in parallel orientation. This mutant gene is not expressed in the presence, nor in the absence of *En/Spm* transposase, indicating that the orientation of the insertion is important in determining suppressibility. The *ap1::I/dSpm* cDNAs still have to be sequenced to determine if splicing uses the original splice donor/acceptor sites or if new sites are used. The wild-type phenotype is not only suppressed in the presence of the TEn5 *En/Spm*-transposase T-DNA, but also in combination with the Tnp2 T-DNA. This confirms the role of the TNPA protein as the suppressor factor. Like for the *a1-m1* alleles, the phenotype of *ap1::I/dSpm* mutants in the absence of *En/Spm* transposase can be distinguished from the wild type, so expression of the *AP1* gene is probably not complete. Remarkable is that in contrast to weak EMS or radiation induced alleles of the *AP1* gene, the mutant phenotype becomes less apparent during plant development. Higher up the inflorescence, the

flowers are nearly normal, whereas Bowman et al. (1993) report less normal flowers at the tip, compared to the base of an inflorescence, for weak alleles of the *AP1* gene. This points at an increased need for functional APETALA1 protein in flowers formed higher up the inflorescence. Apparently this need for functional APETALA1 protein can be fulfilled in the *ap1::I/dSpm* plants.

The effect found for *a1-m2 8004* maize kernels, in which the mutant allele carries an *I/dSpm* inserted in the promoter of the *A1* gene involved in anthocyanin pigmentation, is reversed compared to the situation for *a1-m1* kernels. In the absence of an *En/Spm* the aleurone is colourless, but in the presence of transposase proteins, presumably TNPA, the aleurone is coloured pale (Schwarz-Sommer et al., 1987). Binding of the TNPA protein to the element's termini is proposed to mediate *En/Spm* dependent restoration of *A1* expression, not by blocking the interaction with other proteins as in *a1-m1*, *a2-m1* and *bz-m13*, but by bringing the transposon termini together, in a way that the disrupted parts of the *A1* promoter are virtually reunited such that interaction with transcription factors results in mRNA transcription.

It can be concluded from the reconstitution experiments, in which *lad(::I/dSpm)* alleles are recombined with *En/Spm* transposase, that expression of the *LAD* gene from the *lad::I/dSpm207* allele depends on the expression of the TNPA transposase protein, as in the *a1-m2 8004* maize allele. In the absence of TNPA protein, the insertion of *I/dSpm207* results in a non-functional *LAD* gene and male sterility. In the presence of the TNPA protein, the mutagenic ability of *I/dSpm207* is abolished and the plant appears fertile. The *I/dSpm207* element is still able to excise in the presence of TEn5 (expressing both transposase proteins), which can lead to stable, non-depending *lad* alleles. The relatively large fraction of male steriles in the progeny of plant H21-6 (hemizygous for TEn5), can be explained when assuming that this plant was heterozygous for a *lad::I/dSpm207* insertion allele and a *lad* excision allele. 9/16 (56.2%) of the progeny will have at least one TEn5 and one *lad::I/dSpm207* allele, resulting in fertility, while the rest (43.8%) lacks either TEn5 or the *I/dSpm207* insert causing male sterility. This expected male sterile fraction is very close to the observed fraction (43.8 vs. 45%).

As in the maize *a1-m2* allele, the *I/dSpm* element is inserted anti-parallel in the *LAD* gene. Different from the situation in maize is that *I/dSpm207* does not seem to be inserted in the promoter of the *LAD* gene. There is an ATG startcodon 8 bp downstream of the insertion, but as there are two out of frame ATGs 7 bp and 35 bp upstream, and as the first upstream stop codon is 104 bp upstream of the insertion, *I/dSpm207* is inserted in the reading frame or in the 5' untranslated leader of the *LAD* gene. Insertion in the reading frame makes it difficult to envisage the positive influence of TNPA binding on the expression of the gene. Insertion in the leader allows the alternative explanation that the 5' part of an *I/dSpm* element has an outward directed promoter activity, but only in combination with the TNPA protein. This may also explain the observed *En/Spm* dependence of the maize *a1-m2* allele. To resolve these questions, mRNA transcription in wild-type plants, and in male sterile and fertile *lad::I/dSpm* plants has to be studied. It will provide evidence for transcription of the

open reading frame and allows the positioning of the insert relative to the transcription start. Studying the *ap1::I/dSpm* and *lad::I/dSpm* mutants has made clear that the Suppressor function of the *En/Spm-I/dSpm* system, as described in maize, can also be found in the heterologous host Arabidopsis. Understanding this phenomenon to its full extent will extend the application of the *En/Spm-I/dSpm* transposable element system in Arabidopsis beyond transposon mutagenesis, to a situation where *I/dSpm* element insertions can be used to study regulation of gene-expression.

Material and methods

Analysis of ap1::I/dSpm

The new *ap1* mutant was found in progeny of plant 9-5R4 (Aarts et al., 1993). Insertion of an *I/dSpm* element in the *AP1* gene was determined by DNA blot analysis using DNA from a stable *ap1::I/dSpm* mutant cut with BglII, HindIII, SstI, SphI or SalI. The blot was first hybridized with the full *AP1* cDNA as probe, then stripped and hybridized with the *I/dSpm* element specific I-LJ probe. Both probes detected one common band with all enzymes. To determine a restriction fragment containing the *I/dSpm* element insert suitable for cloning, DNA from Ler and a stable *ap1::I/dSpm* plant was cut with BglII, DraI, HincII, SstI or SalI and hybridized with a 9-kb *AP1* genomic DNA probe. The smallest fragment showing a band shift in the *ap1::I/dSpm* plant was a 0.9-kb DraI fragment, which was subcloned in pBluescript SK^+ from the 9-kb genomic DNA clone and sequenced. The position of the insert relative to the gene was determined using one genomic primer positioned at each end of this DraI fragment (TGA-ACT-AGT-GTT-GCT-TCT-TTT-GC (5') or TGG-AAC-CCG-ATC-GAT-TTT-CTA-GAT-AG (3')) and the TIR *I/dSpm* terminus specific primer (Aarts et al., 1995a), for PCR on DNA of a plant homozygous for the *ap1::I/dSpm* allele, and without any *En/Spm* transposase. One combination produced a fragment which was cloned and sequenced.

Suppression of the apetala phenotype in *ap1::I/dSpm* mutants was examined in three F_2 families made by crossing a stable homozygous *ap1::I/dSpm* plant to Ler; to a plant containing the TEn5 *En/Spm* transposase locus; and to a plant containing the Tnp2 *En/Spm* transposase locus. In the last two crosses, F_1 plants selected for the presence of the T-DNA loci on hygromycin medium (20 mg/l), were used to produce the segregating F_2 families. The phenotype of individual F_2 plants was observed for 22 days, starting 24 days after sowing. The TEn5 T-DNA locus is described in chapter 1 (Aarts et al., 1995a). The Tnp2 T-DNA locus contains a 2'-*tnpA* gene fusion as well as a chiB-*tnpD* gene fusion. The locus was designed to induce germinally transmitted anther specific *I/dSpm* transposition. However, no germinal excisions were detected in the F_2 of a Tnp2 plant crossed to plants containing an *I/dSpm* element inserted in a streptomycin resistance gene (Chapter 7). The 2'-*tnpA* gene cassette used in the Tnp2 construct was shown before to be functional (Grant et al., 1990) and therefore the chiB-*tnpD* gene is probably not active.

Analysis of lad

The *En/Spm* transposase dependent expression of the *lad::I/dSpm* allele was tested by crossing a stable homozygous *lad::I/dSpm* plant to Ler; to a TEn5 plant; and to a Tnp2 plant. At the same time, similar crosses were made with a homozygous *lad* plant (no *I/dSpm* insertion, but male sterile). Plants were sown on hygromycin containing medium (20 mg/l) to select for the presence of the *En/Spm* transposase T-DNA loci in the F_1. 48 F_2 plants from the crosses with TEn5 or Tnp2 and 24 plants from the cross to Ler were sown and their phenotypes notated. After flowering the average number of seeds per silique was determined in at least 50 consecutive siliques. First the weight of 100 seeds was determined to calculate the total number of seeds, which was divided by the number of siliques to get the average number of seeds per silique.

DNA analysis

The techniques used for DNA analysis, and the genomic library used to isolate long fragments of *I/dSpm* element flanking DNA were all as has been described previously (Aarts et al., 1995b).

Genetic mapping

The genetic mapping of the *LAD* locus was performed by RFLP analysis of the Col x Ler population of recombinant inbred lines (Lister and Dean, 1993), using the JoinMap mapping program (Stam, 1993).

Acknowledgements

We want to thank Martin Yanofsky for providing the *API* cDNA and genomic DNA clones, Hiroyoshi Kubo for the first segregation analyses of the tagged *ap1* mutant and Gerrit Stunnenberg c.s. for taking care of the plants in the greenhouse.

Chapter 7

Gene trapping with an *En/Spm-I/dSpm* transposon system

Mark G.M. Aarts, Krista Kooi and Andy Pereira

Summary

The analysis of transposed *I/dSpm* elements revealed many insertions in the transcribed part of genes, but without causing a mutant phenotype. These insertions cannot be discriminated from insertions outside genes, unless the *I/dSpm* element contains a reporter gene that is activated upon insertion in transcribed DNA regions. Such transposable elements, that are modified to detect gene expression, are known as gene trap elements. A set of *I/dSpm* gene trap elements has been designed and transformed to Arabidopsis. In this chapter, the transposition of these gene traps is reported and their use in detecting gene expression is demonstrated.

Introduction

A relatively new application of transposons is the use of modified elements as detectors for adjacent gene activity. This gene trapping technique, which is further specified as exon trapping, promoter trapping or enhancer trapping, has originally been developed for the P-element of *Drosophila melanogaster* (Bellen et al., 1989; Bier et al., 1989; Wilson et al., 1989). The two major goals of using such modified elements are: 1) the identification of insertions in coding regions by screening for the expression of translational or transcriptional fusions between a plant gene and a reporter gene; 2) the detection of insertions in the vicinity of transcriptional regulatory elements that control spatial or temporal reporter gene expression in response to developmental, hormonal or environmental stimuli. Enhancer trap elements are designed such that a reporter gene is cloned behind minimal promoter sequences, which induce transcription only when controlled by nearby enhancer sequences. In promoter trap elements the minimal promoter sequences are left out and reporter gene activity relies on insertion of the element downstream of a genomic promoter. Exon traps lack promoter sequences as well as the ATG startcodon of the reporter gene. Instead, they contain splice acceptor sites, which enable the formation of transcriptional fusions of the reporter gene to the trapped gene. It is expected for all these gene trap elements, that the expression of the reporter genes reflects the expression of the trapped gene, as both are regulated by the same enhancer or promoter sequences.

In Arabidopsis, promoterless *aph(3')II* gene T-DNA constructs were first introduced *in planta* by *Agrobacterium tumefaciens* transformation, for the detection of upstream gene regulatory sequences by selection for APH(3')II activity (Koncz et al., 1989). This was soon followed by transformation of *uidA* based gene trap T-DNAs (Goddijn et al., 1993; Kertbundit et al., 1991; Topping et al., 1991). The cell-autonomous *uidA* gene encodes for ß-glucuronidase or GUS, an enzyme that converts the added colourless substrate X-Gluc (5-bromo-4-chloro-3-indolyl β-D-glucuronide) into a blue product (Jefferson et al., 1986). The use of modified transposable elements for trapping purposes in Arabidopsis was first published by Fedoroff and Smith (1993), who used it to trap a ribosomal protein S16 gene homologue (Tsugeki et al., 1996). Recently an *Ac-Ds* based enhancer trap and exon trap system has been described, which gives high frequencies of gene detection as well as tagged mutants (Springer et al., 1995; Sundaresan et al., 1995). These two transposon systems also use the *uidA* gene as reporter gene.

With the gene trapping technique it is possible to obtain information on of the majority of transposon insertions that do not cause a visible phenotype. We have previously shown, that many plants which contain homozygous *I/dSpm* are without a mutant phenotype (Aarts et al., 1995a; Chapter 5). In this chapter is described that some of these elements are actually in a gene, or very close to a gene. For *Ac-Ds* transposon systems in Arabidopsis, it was estimated that only 2-4% of the *Ds* insertions give a visible mutation (Bancroft et al., 1993; Long et al., 1993a). In most cases, such transposons are inserted outside the transcribed part of a gene, or in an intron. Other inserts may knock out a gene function, but the mutant phenotype is not obvious under regular growing conditions, or the mutated gene is redundant

and no mutant phenotype can be observed. In these cases, gene trap elements are likely to reveal insert specific *uidA* expression. Another advantage of gene trap elements is the opportunity to screen plants with hemizygous insertions for reporter gene activity. This allows the rescue of plants with an insertion that causes lethality in a homozygous state.

To permit trapping studies with the *En/Spm-I/dSpm* system in Arabidopsis, *I/dSpm* elements have been modified to contain the *uidA* gene under control of minimal 35S promoter sequences or without any additional promoter sequences. Plants with mobilized *I/dSpm* gene trap elements have been studied and use of such elements for gene trapping is discussed.

Results

Insertion of I/dSpm elements in genes does not always lead to an obvious phenotype

Only few of the several hundreds of different *I/dSpm* element insertions that have been characterized from the two element *En/Spm-I/dSpm* transposon system in Arabidopsis cause an obvious phenotype. From the rest of the *I/dSpm* elements, some may have jumped into a gene, but have not caused a mutation or the function of the mutated gene is redundant under the screening condition. To test if this is indeed the case, the flanking DNAs from *I/dSpm* elements *18, 41G, 52, 53, 75, 76, 78* and *84* (see Chapter 5) were used as probe to hybridize to a cDNA library representing mRNAs from various plant tissues and developmental stages (Newman et al., 1994). Only the flanking DNAs from *I/dSpm41G* and *I/dSpm78* each detected one cDNA clone, which DNA sequences had no homology to sequences present in the DNA databases. The homozygous insertion of neither of these *I/dSpm* elements leads to a mutant phenotype.

For many of the *I/dSpm* flanking DNA fragments obtained by IPCR, the DNA sequences have been determined and compared to DNA or protein databases. From 18 tested sequences, the sequences flanking *I/dSpm* elements *62, 64, 75, 114* and *201* shared significant homology with a gene. *I/dSpm62* is inserted in the reading frame of a *myb* related gene deposited as *Atmyb5* (GenBank accession number U26935), sharing 95% amino acid (a.a) sequence identity. It further identifies two Arabidopsis ESTs. *I/dSpm64* appears to be inserted in a tyrosine aminotransferase gene, sharing 77% a.a. identity with a rat tyrosine aminotransferase protein. It also shares homology with one Arabidopsis EST. *I/dSpm75* appears to be inserted in an intron of a gene for which only two Arabidopsis EST sequences have been found. One of these, EST 134O4T7, shares homology with the N-gene from tobacco conferring TMV resistance. The *I/dSpm114* element is inserted in the open reading frame of a homologue of the Small-Auxin-Upregulated-RNA *AC1* gene (*SAUR-AC1* gene; Gil et al., 1994) and has homology to one Arabidopsis EST. The *I/dSpm201* element is inserted in a gene for which only an Arabidopsis and a *Brassica campestris* EST sequence has been found, with no homology to any functional gene in the databases. Including the inserts for which cDNAs have been found, only *I/dSpm41G, 75, 78, 114* and *201* were found to be transmitted to the next generation. The other elements were probably somatic inserts, not

transmitted through the germline. Under standard greenhouse growing conditions, none of the plants containing these elements displayed a mutant phenotype correlated with these inserts.

Modifying I/dSpm elements as gene traps

In analogy to *Ds* elements that have been modified as gene traps (Fedoroff and Smith, 1993; Sundaresan et al., 1995), *I/dSpm* elements have been converted into gene trap elements. A promoter trap element (SI2G, Figure 1) was constructed by replacing the central part of an *I/dSpm* element by a promoterless *uidA* gene, leaving only 119 bp of *I/dSpm* left junction (LJ) sequence in front of the *uidA* gene. This promoter trap element was altered into enhancer trap elements SI2G60 or SI2G90 (Figure 1), by cloning respectively the -58 or -90 bp minimal Cauliflower Mosaic Virus (CaMV) 35S promoter sequences in front of the *uidA* gene (see Material and methods). These altered *I/dSpm* elements were cloned between a full CaMV 35S promoter and the streptomycin phosphotransferase (SPT) gene such that excision of the *I/dSpm* element will restore the transciption of the SPT gene leading to streptomycin resistance (strep^r). The SPT and the *uidA* genes are transcribed in the same direction. The SI4iG enhancer trap element (Figure 1) was made in a similar way as the SI2G element, only 185 instead of 119 bp of *I/dSpm* LJ sequence were cloned in front of the *uidA* gene, and the whole *I/dSpm* element was cloned in reversed orientation between CaMV 35S promoter and SPT gene. The longer *I/dSpm* LJ sequence contains the weak *En/Spm* promoter, that may function as a minimal promoter. The SI4iG element was further modified by cloning the -58 or -90 minimal CaMV 35S sequences in front of the *uidA* gene resulting in the SI4iG60 and SI4iG90 elements (Figure 1). All SI-elements were cloned into binary T-DNA vectors, which can be selected for, after transformation, by growing plants on kanamycin containing medium.

In total 6 constructs containing the SI2G, SI2G60, SI2G90, SI4iG, SI4iG60 and SI4iG90 gene trap *I/dSpm* elements were transferred to Arabidopsis by *Agrobacterium tumefaciens* transformation (Table 1). The background *uidA* activity of the *I/dSpm* gene traps residing within the T-DNA insertion was determined histochemically. In four independent SI2G transformants and one SI4iG transformant the GUS staining was low and mainly in the vascular bundles of the leaf and stem. In two SI2G60 transformants, three SI4iG60 transformants and one SI4iG90 transformant, GUS staining was strong, especially in vascular bundles from roots and leaves. As there were no obvious differences in GUS staining between these transformants, the *uidA* gene activity is probably caused by enhancer sequences such as the full 35S promoter, present within the T-DNA construct. Unexpectedly, there was no marked difference in GUS expression between the SI2G60 and the SI4iG60 transformants, although the orientation of the enhancer detector in the latter transformants was reversed compared to that of the 35S-SPT gene.

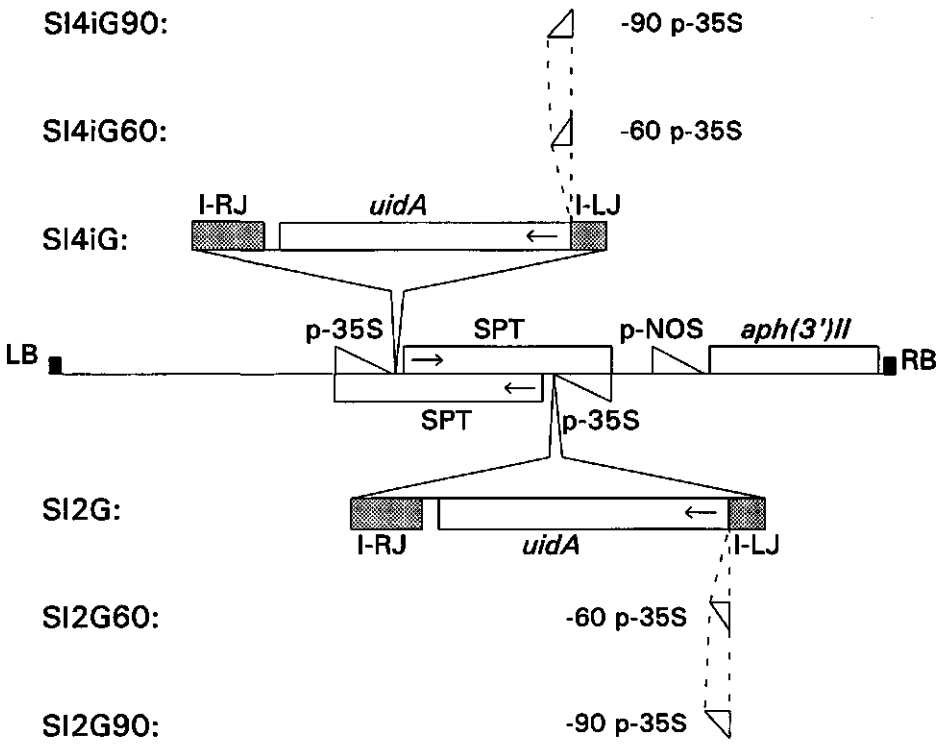

Figure 1: Schematic representation of six T-DNA constructs, containing promoter and enhancer trap *I/dSpm* elements.
The gene trap elements SI4iG, SI4iG60, SI4iG90, SI2G, SI2G60 and SI2G90 are inserted between the CaMV 35S promoter (p-35S) and the SPT gene. SI elements are flanked by *I/dSpm* left end right junction fragments (I-LJ and I-RJ) to permit transposition in the presence of *En/Spm* transposase. SI elements carry a *uidA* detector gene transcribed in either the same (SI2G-type of elements) or the reversed direction (SI4iG-type of elements) as the SPT gene (indicated with an arrow). In the SI4iG-types the I-LJ region is 185 bp long. In the SI2G-types, the I-LJ region is 120 bp long. The -60 or-90 minimal 35S promoters (-60 p-35S/-90 p-35S) in SI4iG60, SI4iG90, SI2G60 and SI2G90 are cloned 5' of the *uidA* gene. The T-DNA constructs are marked by a *aph(3')II* gene conferring kanamycin resistance to the plant. LB and RB are the left and right T-DNA borders.

Transposition of I/dSpm enhancer detector elements causes tissue specific uidA gene expression

The enhancer detector elements were induced to transpose after introducing the TEn5 *En/Spm* transposase T-DNA locus by crossing. F_1 hybrids containing both the transposon and

Table 1: Background GUS staining of gene trap T-DNA transformants.

transformant	seedling	leaf	stem	flower	fruit
SI2G-22	vascular bundles and shoot primordium, low	vascular bundles, medium	vascular bundles, medium	sepals, weak; stigma, medium	tip, medium
SI2G-27	vascular bundles, top of hypocotyl and shoot primordium, weak	id	id	sepals and stigma, weak	vascular bundles, strong; tip, weak
SI2G-29	vascular bundles and shoot primordium, low	id	no	id	carpels, strong
SI2G-31	vascular bundles, medium	vascular bundles, weak	no	id	no
SI2G60-15	root, strong; vascular bundles, weak	vascular bundles, strong	no	vascular bundles, stamen, stigma and pollen, strong	carpels, strong
SI2G90-7	vascular bundles and top hypocotyl, weak; root strong	vascular bundles, weak	no	sepals and stigma, medium; stamen, weak	id
SI2G90-9	root, strong	medium	no	sepals, stamen and stigma, strong; petals, weak	carpels, weak; pedicel, strong
SI4iG-1	hypocotyl, weak	major vascular bundle and trichome base, weak	no	no	pedicel, strong; tip, weak
SI4iG60-2	hypocotyl and vascular bundles, weak; root, strong	strong	no	anthers, weak; stamen filament and sepals, medium; stigma, strong	carpels, strong
SI4iG60-3	hypocotyl and root, strong; shoot primordium, medium; vascular bundles, weak	id	no	sepals, strong; stigma and anthers, medium; stamen filament, weak	pedicel, strong; tip, medium
SI4iG60-8	shoot primordium, strong; vascular bundles, weak	trichomes, strong	no	trichomes, strong; sepals, stigma and anthers, weak	pedicel, weak
SI4iG90-2	root, strong; hypcotyl and cotyledons, weak	medium	cutting edge, strong	stamen and pollen, strong; sepals and stigma, medium	pedicel and tip, medium

transposase T-DNA constructs were self-fertilized and the F_2 seedlings were assayed for *I/dSpm* excision on streptomycin containing medium (Table 2). In the F_2 offspring from plants containing the SI4iG-type of construct, some streptomycin resistant (strep^r) plants and plants with streptomycin resistant sectors (streptomycin variegated or strep^v) plants were found, indicating the excision of enhancer detector elements. The somatic and germinal excision frequencies varied from cross to cross, with most strep^r plants found in the crosses with plants containing the SI4iG90 construct. None of the F_2 offspring from crosses with three different SI2G transformants and one SI2G60 transformant showed any sign of streptomycin resistance. All four SI4iG-type parental lines that gave strep^v seedlings were analysed by DNA blot hybridization. Most plants contained single T-DNAs, except for plant SI4iG90-2, which contained at least three T-DNA copies presumably at one locus. DNA flanking the T-DNA inserts was isolated and used as probe for RFLP mapping. Only DNA flanking the T-DNA in plant SI4iG60-3 detected a polymorphism, which was mapped at 2 cM below RFLP marker g4715b and 16 cM above marker m247 on chromosome 5.

Table 2: Germinal excision assay for response to streptomcyin in the F_2 progeny of crosses between TEn5 plants and plants with gene trap *I/dSpm* elements.
The influence of streptomycin was scored as streptomycin resistance (strep^r), streptomycin variegated (strep^v) and streptoymycin sensitive (strep^s). The number of F_2 families (n) tested per transformant is indicated between brackets.

transformant (n F_2's)	strep^r (%)	strep^{var} (%)	strep^s (%)
SI2G-22 (3)	0 ± 0	0 ± 0	100 ± 0
SI2G-27 (3)	0 ± 0	0 ± 0	100 ± 0
SI2G-29 (2)	0 ± 0	0 ± 0	100 ± 0
SI2G60-16 (2)	0 ± 0	0 ± 0	100 ± 0
SI4iG-1 (5)	0 ± 0	91 ±4	9 ± 4
SI4iG60-2 (3)	0 ± 0	88 ± 6	12 ± 6
SI4iG60-3 (6)	0.3 ± 0.2	72 ±14	28 ± 14
SI4iG90-2 (6)	12 ± 10	55 ± 15	33 ± 11

To increase the number of lines with transposed enhancer detector elements, more strep^r and strep^v plants were selected and DNA from rosette leaves was analysed by blot hybridization using an *I/dSpm* specific DNA probe (Figure 2). Most of the strep^r plants contained a transposed enhancer detector *I/dSpm* element and only some of the strep^v plants. Within an F_2 progeny either one or several different transposed *I/dSpm* enhancer detector elements were found. Before DNA was isolated, cotyledons of some strep^r and strep^v plants were stained with X-gluc and the staining pattern was compared with that of the parental line without transposase. In all plants derived from the multi-copy T-DNA plant SI4iG90-2, the GUS staining pattern was very similar to that in the parental plant, which is probably due to the dominant expression of the *uidA* gene from the unexcised *I/dSpm* elements controlled by T-

DNA located enhancer sequences. Any additional staining caused by transposed enhancer detector elements is masked by this residual *uidA* expression and segregation of the T-DNA locus from the insertion locus is required to effectively use these plants for enhancer detection.

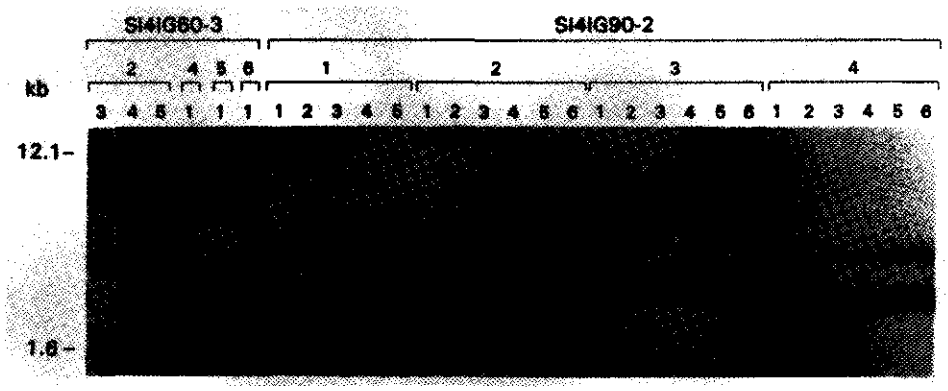

Figure 2: DNA blot hybridization showing transposition of SI4iG60 and SI4iG90 *I/dSpm* elements in F_2 plants (SI-plant x TEn5 plant) selected for streptomycin resistance.
DNA was digested with HindIII and the blot was hybridized with the *I/dSpm* specific SalI I-LJ fragment (Aarts et al., 1995a). Most of the plants with the SI4iG90-2 T-DNA still contain unexcised *I/dSpm* elements, detected as a 2.1-kb fragment, while this is not seen in most SI4iG60-3 F_2 plants. Among eight F_2 families, seven new independent *I/dSpm* element insertions can be observed.

GUS staining in strep^v SI4iG60-3 F_2 seedlings was often only observed in the streptomycin sensitive (strep^s) parts and not in the strep^r excision sectors (data not shown), indicating that the *uidA* gene expression in these sectors was no longer determined by the T-DNA located enhancer sequences. Whether the excised elements had inserted in a part of the genome that is not influenced by enhancers of cotyledon expression, or whether they had not inserted at all, could not be determined. In one cotyledon, from a completely strep^r plant in the same offspring, only one little blue spot was seen at the position of the terminal hydathode of the major leaf vein. This suggests an insertion close to a hydathode specific transcriptional enhancer sequence, but it cannot be ruled out that staining was caused by a clonal sector carrying an *I/dSpm* element reinsertion in a region enhancing cotyledon specific expression. Unfortunately the plant was lost at the cotyledon stage, and the tissue specific staining pattern could not be confirmed in other leaves.

Two of the six SI4iG60-3 F_2 plants shown in Figure 2 have independent *I/dSpm* reinsertions. F_2 plant 2-3 showed staining at the base of the siliques, very weak in petals and very strong in the stamen filaments of mature flowers, which is clearly different from that of the parental plant (Figure 3A and B). This plant, nor its offspring, showed a mutant phenotype. In the mature F_2 plant 5-3 no *uidA* expression was found in the leaves, stem, flowers or roots tested, and therefore the enhancer detector has not transposed to a region controlled by a transcriptional enhancer.

Discussion

On examining the DNA sequences flanking 21 different *I/dSpm* insertions, it was found that at least one third appeared to be inserted in genes. None of the five heritable insertions caused a mutant phenotype. It emphasizes the need to modify *I/dSpm* elements to contain enhancer, promoter or exon traps, enabling the detection and study of gene-expression of genes tagged but not mutated by an *I/dSpm* element. For these purposes, an *I/dSpm* element was modified into a promoter trap element, such as in the SI2G element, which does not have any additional minimal promoter upstream of the *uidA* gene, and into an enhancer trap element, as in the SI2G60, SI2G90, SI4iG, SI4iG60 and SI4iG90 elements. In the enhancer traps either the endogenous weak *En/Spm* promoter or two short characterized sequences from the CaMV 35S promoter were chosen as minimal promoters. The -90 CaMV 35S promoter domain (in the SI#90 elements) alone was expected to confer some promoter activity itself, as it contains an ASF-1 transcription factor binding site (Benfey and Chua, 1990). In tobacco, the -90 domain gave detectable *uidA* expression in roots and meristems. The -58 domain of the CaMV 35S promoter (in SI#60 elements) was unlikely to confer detectable *uidA* expression as neither the -72 nor the -46 domain of the CaMV 35S promoter were able to give detectable *uidA* expression in tobacco (Benfey and Chua, 1990). The strong GUS staining that was observed in transformants containing only the -60 and -90 minimal promoter CaMV 35S enhancer detector element T-DNA constructs, thus can not be explained by activity of the minimal 35S promoters alone, but must be caused by T-DNA located enhancer sequences. Such enhancer sequences are known to be present in the full 35S promoter (Raina et al., 1993), which drives the expression of the nearby SPT marker gene. Unfortunately *uidA* activity could not be abolished by reversion of the *I/dSpm* element relative to the 35S promoter in the SI4iG-type of constructs. It however demonstrates the nature of the 35S enhancer, which acts irrespective of orientation. The consequence was that all the constructs containing a 35S minimal promoter inevitably showed high background GUS staining when an unexcised copy of the *I/dSpm* element was present in the T-DNA.

When the gene trap elements were tested for transposition, the SI2G-types of elements were never found to excise, probably because three of the nine *I/dSpm* Left Border TNPA binding motifs have been deleted. All SI4iG-type of *I/dSpm* elements, which contain all TNPA binding motifs, can transpose, although excision is at a lower level than what has been found for the wild-type 2.2-kb *I/dSpm* element (Aarts et al., 1995a). The increased total length of the gene trap elements and the reduced left subterminal region, may have attributed to this lower excision frequency. Plants with transposed copies of the enhancer detector were found for all three SI4iG-types of constructs. When tested for *uidA* expression, none of the SI4iG enhancer trap transposants showed tissue specific expression of the *uidA* reporter gene. In these elements, the *uidA* gene is cloned downstream of the weak *En/Spm*-transposase promoter, which was shown to be insensitive to the normally strong enhancer of the CaMV 35S promoter and probably also to various other enhancers (Raina et al., 1993). Expression of the *uidA* gene can only be expected in the rare case of an occasional transcriptional fusion of the reporter gene to an *Arabidopsis* gene.

Plants SI4iG60-2, 3 and 8 were the best transformants for testing enhancer detection, as each plant contained a single copy T-DNA insert. As a consequence, no background staining was observed in plants with excised enhancer-traps, contrary to the strong background found in transposants derived from the multicopy T-DNA SI4iG90-2 plant. Only a relatively small sample of plants with transposed enhancer detector elements was further examined for enhancer detection. This resulted in two strep^r transposants with new GUS staining patterns.

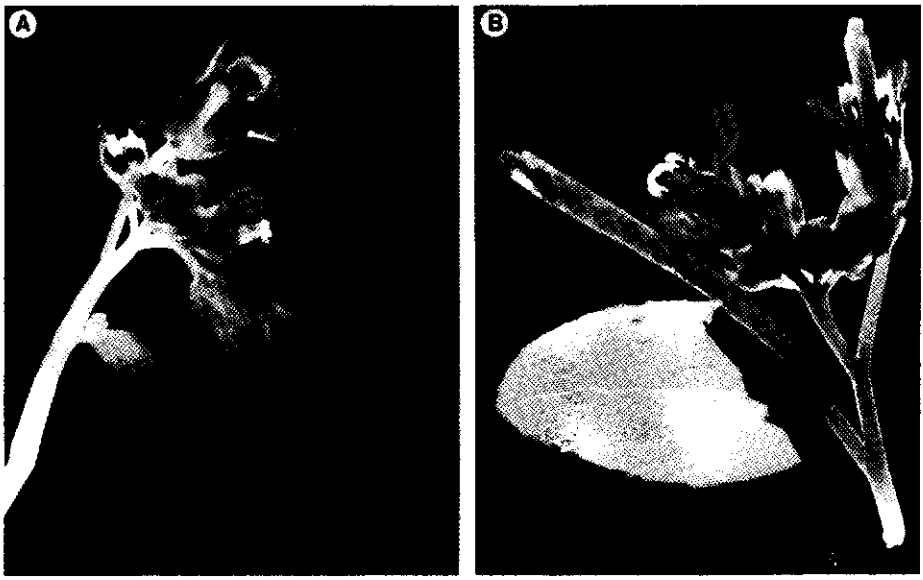

Figure 3: GUS staining in enhancer detector SI4iG60 plants.
A. GUS staining pattern in parental plant SI4iG60-3.
B. GUS staining pattern in streptomycin resistant SI4iG60-3 x TEn5 F_2 plant 2-3 containing a transposed copy of the *I/dSpm* element and no *I/dSpm* element residing in the T-DNA. New *uidA* expression can be seen at the base of the silique and in the stamen filament.

The experiments described and discussed in this chapter indicate that the *En/Spm-I/dSpm* based enhancer trap elements are effective as detectors of enhancer activity. However, for efficient use of these elements some improvements will be required. First of all it would be wise to select against the TEn5 T-DNA (by DNA blot analysis) to obtain plants with stable inserts. This will allow an accurate determination of the induced GUS staining pattern without the problem that the pattern may change due to transposition of the element. It is also preferred to select against unexcised *I/dSpm* elements to be sure that no background staining will be found. This involves a PCR assay on strep^r F_2 seedlings, using one primer flanking the *I/dSpm* element cloning site (e.g. in the CaMV 35S promoter) and one of the *I/dSpm* Right Junction (I-RJ) specific primers as has been described before (Aarts et al., 1995a). As a control, primers flanking the *I/dSpm* element cloning site can be used. At least one combination should give a PCR fragment. Full strep^r seedlings that do not give a PCR product with CaMV 35S and I-RJ primers will not have any *I/dSpm* elements at the T-DNA locus any more.

Recently, an *Ac-Ds* based entrapment system has been published which allows the generation of unlinked transposed *Ds* elements by selection against both the transposase and the transposon donor T-DNAs (Sundaresan et al., 1995). In this system, the *Ds* element contains a kanamycin resistance gene and both the *Ac*-transposase and the *Ds* donor T-DNAs contain the negative selection gene *iaaH,* coding for indole acetic acid hydrolase, which renders a plant sensitive to naphthalene acetamide (NAM). Now that it has been shown that the *En/Spm-I/dSpm* system can be used for enhancer trapping, a similar strategy can be employed to design new *En/Spm* transposase and *I/dSpm* gene trap T-DNA constructs. A disadvantage of the *Ac-Ds* system is that only insertions unlinked to either T-DNA can be found. As an alternative, non-lethal selectable marker genes like the *Lc* gene (Lloyd et al., 1992) can be used instead of the *iaaH* gene, which will allow the recovery of plants still containing T-DNAs. It will be advantageous to have an improved *En/Spm-I/dSpm* based entrapment system in addition to the *Ac-Ds* based system, as the two transposon systems have a different transposition behaviour in terms of the state of the insertion site and the range of insertion sites. Together, they are likely to give a broader range of insertion types than each of the transposable element systems alone.

Material and methods

DNA analysis, cDNA library screening and database searches

The isolation of genomic DNA, *I/dSpm*-IPCR, blotting, hybridization, washing of blots etc. has been described previously (Aarts et al., 1995a). The cDNA library that was screened with different *I/dSpm* flanking DNA probes represented mRNAs isolated from different parts of Arabidopsis ecotype Columbia plants (Newman et al., 1994). DNA sequences were determined as described previously (Aarts et al., 1995a). DNA sequences flanking *I/dSpm* elements *4, 18, 52, 56, 62, 64, 67, 70, 71, 72, 75, 76, 78, 84, 113, 114, 117, 118* and *201* have been checked against the database of Expressed Sequence Tags (dbEST) and the non-redundant database of protein sequences to search for homologues, using the BLAST program (Altschul et al., 1990). The accession numbers for DNA or protein database entries reported to have significant homology to *I/dSpm* element flanking sequences are as follows. For *I/dSpm62,* Arabidopsis ESTs 138E3 (T46204) and 180F18 (H36879) and cDNA *Atmyb5* (U26935) were found. For *I/dSpm64,* Arabidopsis EST 149B19 (T76370) and the rat tyrosine aminotransferase (P04694) were found. For *I/dSpm75,* Arabidopsis ESTs 134O4 (T46064) and G1C3 (N96078) were found. For *I/dSpm114,* Arabidopsis EST 126C19 (T44528) and *SAUR-AC1* (S70188) were found and for *I/dSpm201,* Arabidopsis EST ATTS1353 (Z26081) and *Brassica campestris* EST BNAESTAH (L35806) were found.

Construction of gene trap elements

All gene trap elements were cloned in binary vector plasmid pPCV701 (Walden et al., 1990).

The plasmids pSI2G, pSI2G60 and pSI2G90, containing respectively the promoter trap SI2G and enhancer traps SI2G60 and SI2G90 (SI2G type of plasmids), were made in one ligation step combining multiple components in equimolar amounts. The different components were isolated as follows. The cloning sites in the vector were created by releasing a 870-bp EcoRI-HindIII fragment. The CaMV 35S promoter was isolated as a 0.43-kb HindIII-BamHI fragment from plasmid pDH51 (Pietrzak et al., 1986)). The *I/dSpm* element Left Junction fragment (I-LJ) was isolated as a BamHI-PstI (pSI2G) or BamHI-EcoRV fragment (pSI2G60, pSI2G90) from a 0.12-kb AvaII fragment subclone of the *En-1* element (Pereira et al., 1986) in the SmaI site of pBluescript SK^+ (SK^+). The promoter-less *uidA* gene was first cloned as a 1.86-kb blunt-ended NcoI-BamHI fragment (Jefferson et al., 1986; V. Walbot, unpublished) in the EcoRV site of SK^+, from where it was isolated as a PstI-SalI fragment. The CaMV 35S terminator was isolated as a 0.22-kb SalI-EcoRI fragment from plasmid pDH51. The 0.66-kb *I/dSpm* Right Junction fragment (I-RJ) was isolated as a EcoRI-ClaI fragment from a NlaIII subclone of the *En-1* element in the SmaI site of SK^+. The SPT gene was isolated as a 2.24-kb ClaI-EcoRI fragment from plasmid pJJ2111 derived from pJJ1571 (Jones et al., 1987). The CaMV 35S minimal promoter sequences were isolated from plasmid pDH51 as EcoRV-PstI (from positions -90 to +8 of the 35S promoter) or FokI (blunt)-PstI (-58 to +8) fragments for respectively pSI2G90 and pSI2G60.

For plasmids pSI4iG, pSI4iG60 and pSI4iG90 (SI4iG type of plasmids), essentially the same components have been used as in the SI2G type of plasmids, with the exception that the whole *I/dSpm* element is cloned in reversed orientation (Figure 1). This consequently chanced the restriction sites used for cloning. First the SPT gene was cloned from a 35S-SPT SK^+ vector as an EcoRI-SmaI fragment in the EcoRI-HindIII (blunt) sites of pPCV701, creating pPCV701-SPT. For cloning of the *I/dSpm* element, the 35S promoter was removed as a EcoRI-ClaI fragment, opening the vector. The ligation components were as follows. The full CaMV 35S promoter was supplied as an EcoRI-HindIII fragment from 35S-SPT. An I-RJ/CaMV 35S terminator-fusion was supplied as a HindIII-SalI fragment isolated from pSI2G. The *uidA* gene was supplied as a SalI-PstI (blunt) fragment (for pSI4iG), or as a SalI-PstI fragment (for pSI4iG60 and 90). Minimal 35S promoter fragments were supplied as PstI-FokI (pSI4iG60) or PstI-EcoRV fragments (pSI4iG90). The I-LJ was supplied as a 0.18-kb MaeIII (blunt)-ClaI fragment (3'-5') isolated from a DdeI subclone of *En-1* in the SmaI site of SK^+. These four (pSI4iG) or five fragments (pSI4iG60 and 90) were cloned in the EcoRI-ClaI opened vector pPCV701-SPT.

Plant transformation and plant growth

Arabidopsis was transformed according to two protocols (Clarke et al., 1992; Márton and Browse, 1991), both based on the method described by Valvekens et al. (1988). The major adaptation in the protocol described by Clarke et al. (1992) is the addition of silver thiosulfate to the media. Márton and Browse (1991) reevaluated the protocol by Valvekens et al. (1988) and introduced cold treatment, and an adjusted hormone balance leading to regeneration through adventitious somatic embryogenesis. When this protocol was used, an

about 10 times higher transformation frequency was obtained compared to the protocol by Clarke et al. (1992).

Gene trap transformants were selected on MS medium containing 1% sucrose (MS10) and 30 to 50 mg/l kanamycin. F_1 hybrids between progeny of gene trap transformants and plants containing the TEn5 *En/Spm* transposase locus were selected on MS10 medium containing 50 mg/l kanamycin and 20 mg/l hygromycin. Excision of the *I/dSpm* gene trapping elements was tested by sowing at least 50 F_2 seeds on MS medium with 1% glucose and 200 mg/l streptomycin. Green and green/white variegated seedlings were transferred to the greenhouse soon after selection, and kept at high (near 100%) relative humidity until the first leaves were visible.

Isolation of T-DNA flanking DNA and genetic mapping

The isolation of T-DNA flanking DNA was by plasmid rescue, as described previously (Aarts et al., 1995a). T-DNA copy number was determined by DNA blot analysis after digestion with HindIII and hybridization with a T-DNA right border specific probe. Genetic mapping of T-DNA flanking probes was performed on the Col x Ler set of recombinant inbred lines (Lister and Dean, 1993) using the Joinmap program (Stam, 1993) as has been described previously (see Chapter 5).

GUS staining procedure

Plant parts assayed for in situ *uidA* activity were cut from the plant with a scalpel and transferred to a 0.1 M sodium phosphate buffer (pH 7.2) containing 1mM EDTA, 0.1% Triton, 2 mM $K_3[Fe(CN)_6]$ and 1 mM X-Gluc. The samples were placed at 37°C for at least 1 hour. The tissue was cleared from chlorophyll by washing with 96% ethanol (1 hour) and 70% ethanol (several hours). GUS staining was observed by placing the sample on a glass slide and examining it under a binocular microscope.

Acknowledgements

We thank Maarten Koornneef for critically reading the manuscript, the Nottingham Arabidopsis Stock Centre for distributing the RI-lines, and Gerrit Stunnenberg for taking care of most of the plants in the greenhouse. This work was supported by grant BIOT-CT90-0207 from the European Community under the BRIDGE T-project on 'Development and use of *Arabidopsis thaliana* as a tool for isolating genes of agronomic importance.'

General conclusions and summary

At the start of the research described in this thesis, the main aim was to develop, study and apply an efficient *En/Spm-I/dSpm* based transposon tagging system in *Arabidopsis thaliana* to generate tagged mutants and to provide insights in the possibilities for future applications of such a transposon tagging system in studying plant gene functions. The first step was the introduction of an active *En/Spm-I/dSpm* system into Arabidopsis. Initially a very simple T-DNA construct was transformed, containing a nearly full *En-1* element, without left and right border sequences, and with its promoter replaced by the stronger CaMV 35S promoter. As the same construct harboured a non-autonomous *I/dSpm* element, only one T-DNA transformation was needed. Transformation of this 'in cis two-element *En/Spm-I/dSpm* system' yielded one transformant with two T-DNA insertion loci, TEn2 and TEn5, each having one, respectively, five 35S-*En/Spm* transposase gene copies (Chapter 1). The transposition activity of the *I/dSpm* elements turned out to be surprisingly high. Instead of a germinal excision frequency, which was generally used to express the activity of heterologous transposable element systems, the term independent transposition frequency or itf was coined, as a measure accounting for the entire transposition process. Although not always easy to compare for different transposable element systems, an average itf of over 10%, as was found for this *En/Spm-I/dSpm* system (Chapter 1), has not been reported for any of the *Ac-Ds* based heterologous transposon tagging systems developed for Arabidopsis (Bancroft et al., 1992; Bhatt et al., 1996; Fedoroff and Smith, 1993; Honma et al., 1993; Long et al., 1993b; Swinburne et al., 1992). Obtaining a high transposition frequency in *Ac-Ds* systems is hampered by the fact that the transposase acts as an inhibitor of *Ac* or *Ds* transposition when its expression exceeds a certain level. Apparently such autoregulatory mechanism is not present in the 'in cis two-element *En/Spm-I/dSpm* system'.

The need for a more sophisticated system diminished with the availability of this simple, but efficient *En/Spm-I/dSpm* transposon system and it was therefore studied in more detail to determine: 1) the ability to transpose continuously even after many plant generations; 2) the distribution of elements after transposition; 3) the ability to transpose to transcribed regions and 4) the ability to cause mutations. To start with the first issue, transposition has been studied in up to 12 generations, starting from the primary transformant. In all these generations, there was no apparent reduction in itf, demonstrating a continuous transposition of *I/dSpm* elements, irrespective of the generation number. The second issue, the distribution of elements after transposition, is another important aspect of a transposable element system. For maize transposons it is reported that the insertion site is preferentially physically and often genetically linked to the excision site (Dooner and Belachew, 1989; Peterson, 1970). This is not so remarkable considering that the proteins that perform the transposition steps will have a higher chance of encountering a nearby site on the genome instead of a distant DNA sequence. Like in maize (Peterson, 1970), the *I/dSpm* elements show a preference for insertion in genetically linked sites (Chapter 1), but the preference is not very explicit. Based on the observations of transpositions from tagged genes to sites within a few cM or only

several kb away, and on the analogy to the *En/Spm-I/dSpm* elements in maize, the overall estimate is that about 30% of the elements transpose to sites genetically linked to the excision site. The mapped elements (Chapter 5) show a fairly even distribution over the genome, although there seems to be some clustering of elements (from different origin) to certain genomic regions, like the top of chromosome 4, the bottom of chromosome 1 and the lower half of chromosome 5.

In accordance with the idea that DNA must be in an open confirmation to allow the access of transposase proteins before transposition (Zhang and Spradling, 1993), there are many indications that *I/dSpm* elements insert in regions of the Arabidopsis genome containing genes: a) *I/dSpm* flanking DNAs rarely contain repetitive sequences, but are mostly single copy sequences, as are genes (Chapters 1 and 5); b) about half of the *I/dSpm* elements are inserted in relatively conserved genomic regions, with no RFLPs for five restriction enzymes (Chapters 1 and 5); c) at least one third of the examined *I/dSpm* elements is inserted in close vicinity of transcriptionally active genomic sequences (Chapter 7). Insertion in unique, conserved and often transcribed DNA may not seem surprising for a plant species with little repetitive DNA and a small genome with a high gene density (Meyerowitz, 1989). However, a high frequency of insertion into genic regions of the genome offers the best chances for gene tagging.

The most important aspect of a transposable element system is the possibility to generate tagged mutants. The *En/Spm-I/dSpm* system is mutagenic, with as much as 12 tagged mutants found so far. Most of these were obtained after screening for random mutant phenotypes. When screening for specific mutants, such as reduced seed dormancy, tagged alleles of the *ABI3* and *LEC1* genes were found (M. Koornneef et al., unpublished results). Mutants at the *CER6* locus were obtained by targeted tagging, using the nearby *ap1::I/dSpm* allele as the *I/dSpm* element donor (A. Pereira, unpublished results). These selected and targeted transposon tagging experiments are illustrative examples of the feasibility to efficiently isolate tagged mutants of a special phenotypic or genotypic class.

Ideally, a population saturated with different *I/dSpm* insertions can be made, allowing the isolation of mutants for virtually every gene. Such a 'mutation machine' can be further used for PCR based targeted gene inactivation. This novel technique, which was originally developed for *Drosophila melanogaster* (see O'Hare, 1990), exploits the abundance of transposons for the identification of insertions in genes with known DNA sequence, but no known mutant phenotype. In general, DNAs from multidimensional pools of individuals from a large population are used for a PCR using two primers. One primer is specific for the transposon terminus (directed outwards), the other is specific for the target gene. A fragment can only be amplified when a transposon insert is close enough to the target primer. This technique has been shown to work in *Drosophila* (Ballinger and Benzer, 1989; Kaiser and Goodwin, 1990), *Caenorhabditis elegans* (Zwaal et al., 1993), *Petunia hybrida* (Koes et al., 1995) and maize (Das and Martienssen, 1995), using 'mutation machine' transposable element systems.

Two *I/dSpm* tagged genes that have been studied in great detail are the *MS2* gene (Chapters 2 and 3) and the *CER1* gene (Chapter 4). Mutants for both genes display a

conditional male sterile phenotype, which is the only mutant phenotype for the *ms2* mutant, but for the *cer1* mutant it is a pleiotropic effect of a deficiency in epicuticular wax biosynthesis. The *ms2* mutants are occasionally able to self-fertilize, especially in high relative humidity and late in plant development, but seed set rarely reaches more than a few percentages of wild-type seed set. This in contrast to the *cer1* mutants, which are male sterile in low relative humidity ($\leq$50% RH) and fertile in high relative humidity ($\geq$95% RH). Fertility cannot be completely restored by environment in the *ms2* mutants due to the drastic effect of the mutation on pollen development. The *MS2* gene is expressed in the tapetum around the time of microspore release from the microspore mother cells. The gene is needed for the development of a proper exine layer, protecting the microspore from harmful environmental influences. Consequently the few *ms2* microspores that are produced have very feeble pollen walls, which leaves only very few pollen grains intact for fertilization.

The *CER1* gene acts much later in pollen development. Phenotypically *cer1* and wild-type pollen cannot be distinguished, apart from a difference in germination ability (Chapter 4). As *cer1* pollen germination is like wild type when applying *CER1* pollen or by pollinating in high relative humidity, there appears to be a substance missing from the pollen coat that is required for the necessary rehydration of a pollen grain. Although essential under low relative humidity conditions, this is only a minor defect, which can be easily overcome.

Besides the similarity in mutant phenotypes, the *MS2* gene and the *CER1* gene share the characteristic of encoding proteins with homology to enzymes in the fatty acid biosynthesis pathway. The MS2 protein has most resemblance with the wax fatty acid reductase protein from the desert shrub jojoba, which is involved in the conversion of wax fatty acids to wax alcohols (Chapter 3). The CER1 protein shares structural features of fatty acid desaturases, and it has a proposed function as a decarbonylase, converting long carbon chain aldehydes to alkanes (Chapter 4). There are more examples of a correlation between male fertility and wax biosynthesis, as also other *cer* mutants such as *cer3, 6, 8* and 10 are known to be disturbed in male fertility. It demonstrates the general importance of fatty acid biosynthesis for male gametogenesis.

The last part of this thesis has been devoted to further applications of the *En/Spm-I/dSpm* tagging system in Arabidopsis for the analysis of plant gene functions. The first description of transposable elements as controlling elements (McClintock, 1948), was based on the effect transposons had on the expression of maize genes. Especially *(d)Spm* insertions were known to cause *Spm* dependent or suppressible gene expression (Fedoroff, 1989). This effect is now also described for the *En/Spm-I/dSpm* system in Arabidopsis (Chapter 6). As in maize, an *En/Spm* suppressible allele contains an anti-parallel *I/dSpm* element insertion, which can be spliced from the mRNA. This knowledge can be further used to design an artificial gene expression system, in which an introduced gene containing an anti-parallel *I/dSpm* element, can be negatively controlled in the presence of an *En/Spm* transposase source. The reverse effect, *En/Spm* dependent gene expression, is now also described for Arabidopsis (Chapter 6), but the mechanism for dependence is not yet understood. The availability of the *En/Spm* dependent *lad::I/dSpm* mutant will be useful for further research.

A more general way to study gene expression is the use of gene traps as detectors of gene activity. The pilot experiments described in chapter 7, have shown the possibility to adapt *I/dSpm* elements as gene traps. Especially their advantage in detecting genes without the need for mutation, and the possibility of studying the activity of genes which are lethal as homozygous mutants, are important additional properties of gene trap systems over "traditional" transposon tagging. In combination with its efficient transposition behaviour, an *En/Spm-I/dSpm* based gene trap tagging system seems an attractive alternative for the existing *Ac-Ds* or T-DNA based gene trap systems.

Summarizing, an *En/Spm-I/dSpm* transposon tagging system has been well developed for Arabidopsis and many of its basic characteristics are studied and described. *I/dSpm* tagged mutants can be found with reasonable frequencies, either by random, selected or targeted tagging strategies. The cloning and characterization of two genes affecting male fertility has been described. Further ways to improve tagging frequencies, based on phenotypic or on genotypic selection have been discussed. In addition, the system can be exploited to study plant gene expression and gene function either by *En/Spm* controlled activity of *I/dSpm* tagged genes, or by using *I/dSpm* gene detector elements.

References

Aarts, M.G.M., Corzaan, P., Stiekema, W.J. and Pereira, A. (1995a). A two-element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **247**, 555-564.

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

Aarts, M.G.M., Keijzer, C.J., Stiekema, W.J. and Pereira, A. (1995b). Molecular characterization of the *CER1* gene of Arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* **7**, 2115-2127.

Albertsen, M.C. and Philips, R.L. (1981). Developmental cytology of 13 genetic male sterile loci in maize. *Can. J. Genet. Cytol.* **23**, 195-208.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H.M. and Somerville, C.R. (1992). Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis. *Science* **258**, 1353-1355.

Arthington, B.A., Bennet, L.G., Skatrud, P.L., Guynn, C.J., Barbuch, R.J., Ulbright, C.E. and Bard, M. (1991). Cloning, disruption and sequence of the gene encoding yeast C-5 sterol desaturase. *Gene* **102**, 39-44.

Baker, B., Schell, J., Lörz, H. and Fedoroff, N. (1986) Transposition of the maize controlling element "Activator" in tobacco. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4844-4848.

Baker, B., Coupland, G., Fedoroff, N., Starlinger, P. and Schell, J. (1987). Phenotypic assay for excision of the maize controlling element *Ac* in tobacco. *EMBO J.* **6**, 1547-1554.

Baker, E. A. (1974). The influence of environment on leaf wax development in *Brassica oleracea* var. *gemmifera*. *New Phytol.* **73**, 955-966.

Balcells, L., Swinburne, J. and Coupland, C. (1991). Transposons as tools for the isolation of genes. *Trends Biotechnol.* **9**, 31-37.

Ballinger, D.G. and Benzer, S. (1989). Targeted gene mutations in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **86**, 9402-9406.

Bancroft, I., Bhatt, A.M., Sjodin, C., Scofield, S., Jones, J.D.G. and Dean, C. (1992). Development of an efficient two-element transposon tagging system in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **233**, 449-461.

Bancroft, I., Jones, J.D.G. and Dean, C. (1993). Heterologous transposon tagging of the *DRL1* locus in Arabidopsis. *Plant Cell* **5**, 631-638.

Bedinger, P. (1992). The remarkable biology of pollen. *Plant Cell* **4**, 879-887.

Beevers, H. (1979). Microbodies in higher plants. *Ann. Rev. Plant Physiol.* **30**, 159-193.

Bellen, H.J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K. and Gehring, W.J. (1989). P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Genes Dev.* **3**, 1288-1300.

Belzile, F. and Yoder, F.I. (1992). Pattern of somatic transposition in a high copy *Ac* tomato line. *Plant J.* **2**, 173-179.

Benfey, P.N. and Chua, N.-H. (1990). The Cauliflower Mosaic Virus 35S promoter: combinatorial regulation of transcription in plants. *Science* **250**, 959-966.

Bhatt, A.M., Page, T., Lawson, E.J.R., Lister, C. and Dean, C. (1996). Use of *Ac* as an insertional mutagen in *Arabidopsis*. *Plant J.* **9**, 935-945.

Bianchi, A., Bianchi, G., Avato, P. and Salamini, F. (1985). Biosynthetic pathways of epicuticular wax of maize as assessed by mutation, light, plant age and inhibitor studies. *Maydica* **30**, 179-198.

Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Ellsworth, G., Jan, L.Y. and Jan, Y.N. (1989). Searching for pattern and mutation in the *Drosophila* genome with a P-*lacZ* vector. *Genes Dev.* **3**, 1273-1287.

Bingham, P.M., Levis, R. and Rubin, G.M. (1981). Cloning of DNA sequences from the *white* locus of D. melanogaster by a novel and general method. *Cell*, **25**, 693-704.

Black, T.A. and Wolk, C.P. (1994). Analysis of a Het- mutation in *Anabaena* sp. strain PCC 7120 implicates a secondary metabolite in the regulation of heterocyst spacing. *J. Bacteriol.* **176**, 2282-2292.

Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M. and Smyth, D.R. (1993). Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* **119**, 721-743.

Bradley, D., Carpenter, R., Sommer, H., Hartley, N. and Coen, E. (1993). Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of Antirrhinum. *Cell* **72**, 85-95.

Brown, J.W.S. (1986). A catalogue of splice junction and putative branch point sequences from plant introns. *Nucl. Acids Res.* **14**, 9549-9559.

Cardon, G.H., Frey, M., Saedler, H. and Gierl, A. (1991). Transposition of *En/Spm* in transgenic tobacco. *Maydica* **36**, 305-308.

Cardon, G.H., Frey, M., Saedler, H. and Gierl, A. (1993a). Mobility of the maize transposable element *En/Spm* in *Arabidopsis thaliana*. *Plant J.* **3**, 773-784.

Cardon, G.H., Frey, M., Saedler, H. and Gierl, A. (1993b). Definition and characterization of an artificial *En/Spm*-based transposon tagging system in transgenic tobacco. *Plant. Mol. Biol.* **23**, 157-178.

Chaudhury, A.M. (1993). Nuclear genes controlling male fertility. *Plant Cell* **5**, 1277-1283.

Chaudhury, A.M., Lavithis, M., Taylor, P.E., Craig, S., Singh, M.B., Signer, E.R., Knox, R.B. and Dennis, E.S. (1994). Genetic control of male fertility in *Arabidopsis thaliana*: structural analysis of premeiotic developmental mutants. *Sex. Plant Reprod.* **7**, 17-28.

Cheesbrough, T.M. and Kolattukudy, P.E. (1984). Alkane biosynthesis by decarbonylation of aldehydes catalyzed by a particulate preparation from *Pisum sativum*. *Proc. Natl. Acad. Sci. USA* **81**, 6613-6617.

Chirala, S.S., Kasturi, R., Pazirandeh, M., Stolow, D.T., Huang, W.-Y. and Wakil, S.J. (1989). A novel cDNA extension procedure. *J. Biol. Chem.* **264**, 3750-3757.

Chuck, G., Robbins, T., Nijjar, C., Ralston, E., Courtney-Gutterson, N. and Dooner, H.K. (1993). Tagging and cloning of a petunia flower colour gene with the maize transposable element *Activator*. *Plant Cell* **5**, 371-378.

Clarke, M.C., Wei, W. and Lindsey, K. (1992). High-frequency transformation of *Arabidopsis thaliana* by *Agrobacterium tumefaciens*. *Pl. Mol. Biol. Rep.* **10**, 1992.

Coen, E.S., Robbins, T.P., Almeida, J., Hudson, A. and Carpenter, R. (1989). Consequences and mechanism of transposition in *Antirrhinum majus*. In *Mobile DNA*. D.E. Berg and M.M. Howe (eds). Am. Soc. Microbiol., Washington DC. pp. 413-436.

Cone, K. (1994). Transposon tagging with *Spm*. In *The maize handbook*. M. Freeling and V. Walbot (eds). Springer-Verlag, New York. pp 240-242.

Cox, K. and Goldberg, R.B. (1988). Analysis of plant gene expression. In *Plant Molecular Biology - A practical approach*. C. Shaw (ed). IRL Press, Oxford. pp. 1-34.

Crute, I., Beynon, J., Dangl, J., Holub, E., Mauch-Mani, B., Slusarenko, A., Staskawicz, B. and Ausubel, F. (1994). Microbial pathogenesis of *Arabidopsis*. In *Arabidopsis*. E.M. Meyerowitz and C.R. Somerville (eds). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA. pp. 705-747.

Das, L. and Martienssen, R. (1995). Site-selected transposon mutagenesis at the *hcf106* locus in maize. *Plant Cell* **7**, 287-294.

Dash, S. and Peterson, P.A. (1994). Frequent loss of the *En* transposable element after excision and its relation to chromosome replication in maize (*Zea mays* L.). *Genetics* **136**, 653-671.

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

Dellaporta, S.L., Woods, J. and Hicks, J.B. (1983). A plant DNA minipreparation version II. *Plant Mol Biol Rep* **4**, 19-21.

Dennis, M. and Kolattukudy, P.E. (1992). A cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO. *Proc. Natl. Acad. Sci. USA* **89**, 5306-5310.

Detrez, C., Sangwan, R.S., Sangwan-Norreel, B.S. (1989). Phenotypic and karyotypic status of *Beta vulgaris* plants regenerated from direct organogenesis in petiole culture. *Theor. Appl. Genet.* **77**, 462-468.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehenxive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.

Domon, C. and Steinmetz, A. (1994). Exon shuffling in anther-specific genes from sunflower. *Mol. Gen. Genet.* **244**, 312-317.

Dooner, H.K. and Belachew, A. (1989). Transposition pattern of the maize element *Ac* from the *bz-m2(Ac)* allele. *Genetics* **122**, 447-457.

Dooner, H.K., Belachew, A., Burgess, D., Harding, S., Ralston, M. and Ralston, E. (1994). Distribution of unlinked receptor sites for transposed *Ac* elements from the *bz-m2(Ac)* allele in maize. *Genetics* **136**, 261-279.

Eggink, G., Engel, H., Vriend, G., Terpstra, P. and Witholt, B. (1990). Rubredoxin reductase of *Pseudomonas oleovorans*: Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *J. Mol. Biol.* **212**, 135-142.

Eigenbrode, S.D. and Espelie, K.E. (1995). Effects of plant epicuticular lipids on insect herbivores. *Annu. Rev. Entomol.* **40**, 171-194.

Erdtman, G. (1960). The acetolysis method. *Svensk Botanisk Tidskrift* **54**, 561-564.

Evrard, J.-L., Jako, C., Saint-Guily, A., Weil, J.-H. and Kuntz, M. (1991). Anther-specific, developmentally regulated expression of genes encoding a new class of proline-rich proteins in sunflower. *Plant Mol. Biol.* **16**, 271-281.

Fedoroff, N., Furtek, D. and Nelson, O., Jr. (1984). Cloning of the *bronze* locus in maize by a simple and generalizable procedure using the transposable controlling element *Ac*. *Proc. Natl. Acad. Sci. USA* **81**, 3825-3829.

Fedoroff, N.V. (1989). About maize transposable elements and development. *Cell* **56**, 181-191.

Fedoroff, N.V. and Smith, D.L. (1993). A versatile system for detecting transposition in *Arabidopsis*. *Plant J.* **3**, 273-289.

Fedoroff, N., Wessler, S. and Shure, M. (1983). Isolation of the transposable maize controlling elements *Ac* and *Ds*. *Cell* **35**, 235-242.

Feldmann, K.A. (1991). T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *Plant J.* **1**, 71-82.

Feldmann, K.A., Marks, M.D., Christianson, M.L., and Quatrano, R.S. (1989). A dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis. *Science* **243**, 1351-1354.

Frey, M., Reinecke, J., Grant, S., Saedler, H. and Gierl, A. (1990). Excision of the *En/Spm* transposable element of *Zea mays* requires two element-encoded proteins. *EMBO J.* **9**, 4037-4044.

Frey, M., Tavantzis, S.M. and Saedler, H. (1989). The maize *En-1/Spm* element transposes in potato. *Mol. Gen. Genet.* **217**, 172-177.

Gerats, A.G.M., Huits, H., Vrijlandt, E., Maraña, C., Souer, E. and Beld., M. (1990). Molecular characterization of a nonautonomous transposable element (*dTph1*) of petunia. *Plant Cell* **2**, 1121-1128.

Gierl, A., Schwarz-Sommer, Zs. and Saedler, H. (1985). Molecular interactions between the components of the En-I transposable element system of *Zea mays*. *EMBO J.* **4**, 579-583.

Gil, P., Liu, Y., Orbovi, V., Verkamp, E., Poff, K.L. and Green, P.J. (1994). Characterization of the auxin-inducible *SAUR-AC1* gene for use as a molecular genetic tool in *Arabidopsis*. *Plant Physiol.* **144**, 777-784.

Goddijn, O.J.M., Lindsey, K., van der Lee, F.M., Klap, J.C. and Sijmons, P.C. (1993). Differential gene expression in nematode-induced feeding structures of transgenic plants harbouring promoter-*gusA* fusion constructs. *Plant J.* **4**, 863-873.

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

Grandbastien, M.-A. (1992). Retroelements in higher plants. *Trends Genet.* **8**, 103-108.

Grant, S.R., Gierl, A. and Saedler, H. (1990). En/Spm encoded tnpA protein requires a specific target sequence for suppression. *EMBO J.* **9**, 2029-2035.

Gray, M.W (1990). The mitochondrial genome and its expression. In *Plant physiology, biochemistry and molecular biology*. D.T. Dennis and D.H. Turpin (eds). Longman, Burnt Mill. pp 147-159.

Greenblatt, I.M. (1984). A chromosome replication pattern deduced from pericarp phenotypes resulting from movements of the transposable element *Modulator* in maize. *Genetics* **108**, 471-485.

Greenblatt, I.M. and Brink, R.A. (1961). Twin mutations in medium variegated pericarp maize. *Genetics* **47**, 489-501.

Grevelding, C., Becker, D., Kunze, R., Von Menges, A., Fantes, V., Schell, J. and Masterson, R. (1992). High rates of *Ac/Ds* germinal transposition in *Arabidopsis* suitable for gene isolation by insertional mutagenesis. *Proc. Natl. Acad. Sci. USA* **89**, 6085-6089.

Hall, D.M. and Jones, R.L. (1961). Physiological significance of surface wax on leaves. *Nature* **191**, 95-96.

Hannoufa, A., McNevin, J. and Lemieux, B. (1993). Epicuticular waxes of *eceriferum* mutants of *Arabidopsis thaliana*. *Phytochemistry* **33**, 851-855.

Hanson, M.R. (1991). Plant mitochondrial mutations and male sterility. *Annu. Rev.Genet.* **25**, 461-486.

Haque, M.M., Mackill, D.J. and Ingram, K.T. (1992). Inheritance of leaf epicuticular wax content in rice. *Crop Sci.* **32**, 865-868.

Haring, M.A., Rommens, C.M.T., Nijkamp, H.J.J. and Hille, J. (1991). The use of transgenic plants to understand transposition mechanisms and to develop transposon tagging strategies. *Plant Mol. Biol.* **16**, 449-461.

Hauge, B.M., Hanley, S.M., Cartinhour, S., Cherry, J.M., Gooodman, H.M., Koornneef, M., Stam, P., Chang, C., Kempin, S., Medrano, L. and Meyerowitz, E.M. (1993). An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. *Plant J* **3**, 745-754.

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

Healy, J., Corr, C., DeYoung, J. and Baker, B. (1993). Linked and unlinked transposition of a genetically marked *Dissociation* element in transgenic tomato. *Genetics* **134**, 571-584.

Henzel, L.L., Nelson, M.A., Richmond, T.A. and Bleecker, A.B. (1994). The fate of inflorescence meristems is controlled by developing fruits in *Arabidopsis*. *Plant Physiol.* **106**, 863-876.

Hird, D.L., Worrall, D., Hodge, R., Smartt, S., Paul, W. and Scott, R. (1993). The anther-specific protein encoded by the *Brassica napus* and *Arabidopsis thaliana* A6 gene displays similarity to ß-1,3-glucanases. *Plant J.* **4**, 1023-1033.

Hodge, R., Paul, W., Draper, J. and Scott, R. (1992). Cold-plaque screening: a simple technique for the isolation of low abundance, diffentially expressed transcripts from conventional cDNA libraries. *Plant J.* **2**, 257-260.

Hoffman, B.M. and Petering, D.H. (1970). Coboglobins: Oxygen-carrying cobalt-reconstituted haemoglobin and myoglobin. *Proc. Natl. Acad. Sci. USA* **67**, 637-643.

Honma, M.A., Baker, B.J. and Waddell, C.S. (1993). High-frequency germinal transposition of Ds^{ALS} in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **90**, 6242-6246.

Hülskamp, M., Miséra, S. and Jürgens, G. (1994). Genetic dissection of trichome cell development in Arabidopsis. *Cell* **76**, 555-566.

Iba, K., Gibson, S., Nishiuchi, T., Fuse, T., Nishimura, M., Arondel, V., Hugly, S. and Somerville, C. (1993). A gene encoding a chloroplast -3 fatty acid desaturase complements alterations in fatty acid desaturation and chloroplast copy number of the *fad7* mutant of *Arabidopsis thaliana*. *J. Biol. Chem.* **268**, 24099-24105.

Igeno, M.I., Simon, R. and Coupland, C. (1996). Increased expression of the *CONSTANS* gene causes early flowering. Abstract in *7th International Conference on Arabidopsis Research*, Norwich, UK.

Jackson, D.P. (1991). In situ hybridization in plants. In *Molecular Plant Pathology*. D.J. Bowles, S.J. Gurr and M. McPhereson (eds). IRL Press, Oxford. pp. 163-174.

James, D.W. Jr., Lim, E., Keller, J., Plooy, I., Ralston, E. and Dooner, H.K. (1995). Directed tagging of the Arabidopsis *FATTY ACID ELONGATION1* (*FAE1*) gene with the maize transposon *Activator*. *Plant Cell* **7**, 309-319.

Jefferson, R.A., Burgess, S.M. and Hirsh, D. (1986). β-Glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc. Natl. Acac. Sci. USA* **83**, 8447-8451.

Jenks, M.A., Tuttle, H.A., Eigenbrode, S.D. and Feldmann, K.A. (1995). Leaf epicuticular waxes of the *eceriferum* mutants in *Arabidopsis*. *Plant Physiol.* **108**, 369-377.

Jones, D.A., Thomas, C.M., Hammond-Kosack, K.E., Balint-Kurti, P.J. and Jones, J.D.G. (1994). Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science* **266**, 789-793.

Jones, J.D.G., Svab, Z., Harper, E.C., Hurwitz, C.D. and Maliga, P. (1987). A dominant nuclear streptomycin resistance marker for plant cell transformation. *Mol. Gen. Genet.* **210**, 86-91.

Kaiser, K. and Goodwin, S.F. "Site-selected" transposon mutagenesis of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **87**, 1686-1690.

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

Kaulen, H., Schell, J. and Kreuzaler, F. (1986). Light-induced expression of the chimeric chalcone synthase-NPTII gene in tobacco cells. *EMBO J.* **5**, 1-8.

Kertbundit, S., De Greve, H., DeBoeck, F., van Montagu, M. and Hernalsteens, J.P. (1991). *In vivo* random ß-glucuronidase gene fusions in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **88**, 5212-5216.

Kim, H.-Y., Schiefelbein, J.W., Raboy, V., Furtek, D.B. and Nelson, O.E., Jr. (1987). RNA splicing permits expression of a maize gene with a defective Suppressor-mutator transposable element insertion in an exon. *Proc. Natl. Acad. Sci. USA* **84**, 5863-5867.

Knapp, S., Larondell, Y., Rossberg, M., Furtek, D. and Theres, K. (1994). Transgenic tomato lines containing *Ds* elements at defined genomic positions as tools for targeted transposon tagging. *Mol. Gen. Genet.* **243**, 666-673.

Koes, R., Souer, E., van Houwelingen, A., Mur, L., Spelt, C., Quattrocchio, F., Wing, J., Oppedijk, B., Ahmed, S., Maes, T., Gerats, T., Hoogeveen, P., Meesters, M., Kloos, D. and Mol, J.N.M. (1995). Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. *Proc. Natl. Acad. Sci. USA* **92**, 8149-8153.

Kok, M., Oldenhuis, R., van der Linden, M.P.G., Raatjes, P., Kingma, J., van Lelyveld, P.H. and Witholt,

B. (1989). The *Pseudomonas oleovorans* alkane hydroxylase gene. *J. Biol. Chem.* **264**, 5435-5441.

Kolattukudy, P.E. (1975). Biochemistry of cutin, suberin and waxes, the lipid barriers on plants. In *Recent advances in the chemistry and biochemistry of plant lipids*. T. Galliard and E.I. Mercer (eds). Academic Press, New York. pp 203-246.

Kolattukudy, P.E. (ed) (1976). *Chemistry and biochemistry of natural waxes*. Elsevier, Amsterdam.

Kolattukudy, P.E. (1980). Cutin, suberin and waxes. In *The biochemistry of plants, vol 4: Lipids: structure and function*. P.K. Stumpf (ed). Academic Press, New York. pp 571-645.

Koltunow, A.M., Truettner, J., Cox, K.H., Wallroth, M. and Goldberg, R.B. (1990). Different temporal and spatial gene expression patterns occur during anther development. *Plant Cell* **2**, 1201-1224.

Koncz, C. and Schell, J. (1986). The promoter of T_L-DNA gene *5* controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* **204**, 383-396.

Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Körber, H., Redei, G.P. and Schell, J. (1989). High-frequency T-DNA-mediated gene tagging in plants. *Proc. Natl. Acad. Sci. USA* **86**, 8467-8471.

Koncz, C., Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Reiss, B., Redei, G.P. and Schell, J. (1990). Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana*. *EMBO J.* **9**, 1337-1346.

Koornneef, M. (1981). The complex syndrome of *ttg* mutants. *Arab. Inf. Serv.* **18**, 45-51.

Koornneef, M. and van der Veen, J.H. (1983). Trisomics in *Arabidopsis thaliana* and the location of linkage groups. *Genetica* **61**, 41-46.

Koornneef, M., Hanhart, C.J. and Thiel, F. (1989). A genetic and phenotypic description of *eceriferum* (*cer*) mutants in *Arabidopsis thaliana*. *J. Hered.* **80**, 118-122.

Koornneef, M., van Eden, J., Hanhart, C.J., Stam, P., Braaksma, F.J. and Feenstra, W.J. (1983). Linkage map of *Arabidopsis thaliana*. *J. Hered.* **74**, 265-272.

Kunkel, B.N. (1996). A useful weed put to work: genetic analysis of disease resistance in *Arabidopsis thaliana*. *Trends Genet.* **12**, 63-69.

Lawrence, G.J., Finnegan, E.J., Ayliff, M.A. and Ellis, J.G. (1995). The *L6* gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. *Plant Cell* 7, 1195-1206.

Lawson, E., Scofield, S., Sjodin, C., Jones, J.D.G. and Dean, C. (1994). Modification of the 5' untranslated leader region of the maize *Activator* element leads to increased activity in *Arabidopsis*. *Mol.Gen.Genet.* **245**, 608-615.

Lemieux, B., Koornneef, M. and Feldmann, K.A. (1994). Epicuticular wax and *eceriferum* mutants. In *Arabidopsis*. E.M. Meyerowitz and C.R. Somerville (eds). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA. pp 1031-1047.

Lister, C. and Dean, C. (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J.* **4**, 745-750.

Lloyd, A.M., Walbot, V. and Davis, R.W. (1992). *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regualtors *R* and *C1*. *Science* **258**, 1773-1775.

Long, D., Martin, M., Sundberg, E., Swinburne, J., Puangsomlee, P. and Coupland. G. (1993a). The maize transposable element system *Ac/Ds* as a mutagen in *Arabidopsis*: Identification of an *albino* mutation induced by *Ds* insertion. *Proc. Natl. Acad. Sci. USA* **90**, 10370-10374.

Long, D., Swinburne, J., Martin, M., Wilson, K., Sundberg, E., Lee, K. and Coupland, G. (1993b). Analysis of the frequency of inheritance of transposed *Ds* elements in *Arabidopsis* after activation by a CaMV 35S promoter fusion to the *Ac* transposase gene. *Mol. Gen. Genet.* **241**, 627-636.

Lucas, H., Feuerbach, F., Kunert, K., Grandbastien, M.-A. and Caboche, M. (1995). RNA-mediated transpostion of the tobacco retrotransposon Tnt1 in *Arabidopsis thaliana*. *EMBO J.* **14**, 2364-2373.

Mandel, M.A. and Yanofsky, M.F. (1995). A gene triggering flower formation in *Arabidopsis*. *Nature* **377**, 522-524.

Mandel, M.A., Gustafson-Brown, C., Savidge, B. and Yanofsky, M.F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**, 273-277.

Mariani, C., De Beuckeleer, M., Truettner, J., Leemans, J. and Goldberg, R.B. (1989). Induction of male sterility in plants by a chimaeric ribonuclease gene. *Nature* **347**, 737-741.

Martin, C., Carpenter, R., Sommer, H., Saedler, H. and Coen, E.S. (1985). Molecular analysis of instability in flower pigmentation of *Antirrhinum majus*, following isolation of the pallida locus by transposon tagging. *EMBO J.* **4**, 1625-1630.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D. and Tanksley, S.D. (1993). Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432-1436.

Márton, L. and Browse, J. (1991). Facile transformation of *Arabidopsis*. *Pl. Cell Rep.* **10**, 235-239.

Masson, P. and Fedoroff, N. (1989). Mobility of the maize *Suppressor-mutator* element in transgenic tobacco cells. *Proc. Natl. Acad. Sci. USA* **86**, 2219-2223.

Masson, P., Strem, M. and Fedoroff, N. (1991). The *tnpA* and *tnpD* gene products of the *Spm* element are required for transposition in tobacco. *Plant Cell* **3**, 73-85.

Masson, P., Surosky, R., Kingsbury, J. and Fedoroff, N.V. (1987). Genetic and molecular analysis of the *Spm-dependent a-m2* alleles of the maize *a* locus. *Genetics* **177**, 117-137.

Masterson, R.V., Furtek, D.B., Grevelding, C. and Schell, J. (1989). A maize *Ds* transposable element containing a dihydrofolate reductase gene transposes in *Nicotiana tabacum* and *Arabidopsis thaliana*. *Mol. Gen. Genet.* **219**, 461-466.

McClintock, B. (1948). Mutable loci in maize. *Carnegie Inst. Washington Yearbook* **47**, 155-169.

McClintock, B. (1951). Mutable loci in maize. *Carnegie Inst. Washington Yearbook* **50**, 174-181.

McClintock, B. (1954). Mutations in maize and chromosomal aberrations in Neurospora. *Carnegie Inst. Washington Yearbook* **53**, 254-260.

McNevin, J.P., Woodward, W., Hannoufa, A., Feldmann, K.A. and Lemieux, B. (1993). Isolation and characterization of *eceriferum* (*cer*) mutants induced by T-DNA insertions in *Arabidopsis thaliana*. *Genome* **36**, 610-618.

Menssen, A., Höhmann, S., Martin, W., Schnable, P.S., Peterson, P.A., Saedler, H. and Gierl, A. (1990). The En/Spm transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a dSpm element in the *A2* gene. *EMBO J.* **9**, 3051-3057.

Metz, J.G., Pollard, M.R. and Lassner, M.W. (1994). *Fatty acyl reductases*. US Patent 5,370,996, published Dec. 6, 1994.

Meyerowitz, E.M. (1989). Arabidopsis, a useful weed. *Cell* **56**, 263-269.

Meyerowitz, E.M. (1992). Introduction to the *Arabidopsis* genome. In *Methods in Arabidopsis Research*. C.Koncz, N.-H. Chua and J. Schell (eds). World Scientific, Singapore. pp 100-118.

Meyerowitz, E.M. and Somerville, C.R. (eds) (1994). *Arabidopsis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Mulligan, B., Wilson, Z., Dawson, J., Kalantidis, K., Vizir, I. Briarty, 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.

Murashige, T. and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473-497.

Nelson, O.E. and Klein, A.S. (1984). Characterization of an *Spm*-controlled bronze-mutable allele in maize. *Genetics* **106**, 769-779.

Newman, T., de Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E. and Somerville, C. (1994). Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol.* **106**, 1241-1255.

Ochman, H., Gerber, A.S. and Hartl, D.L. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-623.

Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E. and Browse, J. (1994). Arabidopsis *FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. *Plant Cell* **6**, 147-158.

O'Reilly, C., Shepherd, N.S., Pereira, A., Schwarz-Sommer, Zs., Bertram, I., Robertson, D.S., Peterson, P.A. and Saedler, H. (1985). Molecular cloning of the a1 locus of *Zea mays* using the transposable elements En and Mu1. *EMBO J.* **4**, 877-882.

Osborne, B.I., Corr, C.A., Prince, J.P., Hehl, R., Tanksley, S.D., McCormick, S. and Baker, B. (1991). *Ac* transposition from a T-DNA can generate linked and unlinked clusters of insertions in the tomato genome. *Genetics* **129**, 833-844.

O'Toole, J.C. and Cruz, R.T. (1983). Genotypic variation in epicuticular wax of rice. *Crop Sci.* **23**, 392-394.

Paul, W., Hodge, R., Smartt, S., Draper, J. and Scott, R. (1992). The isolation and characterisation of the tapetum-specific *Arabidopsis thaliana* A9 gene. *Plant Mol. Biol.* **19**, 611-622.

Parker, J.E., Szabò, V., Staskawicz, B.J., Lister, C., Dean, C., Daniels, M.J. and Jones, J.D.G. (1993). Phenotypic characterization and molecular mapping of the *Arabidopsis thaliana* locus *RPP5*, determining disease resistance to *Peronospora parasitica*. *Plant J.* **4**, 821-831.

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. *Sex. Plant Reprod.* **9**, 1-16.

Peleman, J., Cottyn, B., Van Camp, W., Van Montagu, M. and Inzé, D. (1991). Transient occurrence of extrachromosomal DNA of an *Arabidopsis thaliana* transposon-like element, *Tat1*. *Proc. Natl. Acad. Sci. USA* **88**, 3618-3622.

Pereira, A. and Saedler, H. (1989). Transpositional behavior of the maize *En/Spm* element in transgenic tobacco. *EMBO J.* **8**, 1315-1321.

Pereira, A., Cuypers, H., Gierl, A., Schwarz-Sommer, Zs. and Saedler, H. (1986). Molecular analysis of the *En/Spm* transposable element system of *Zea mays*. *EMBO J.* **5**, 835-841.

Pereira, A., Schwarz-Sommer, Z., Gierl, A., Bertram, I., Peterson, P.A. and Saedler, H. (1985). Genetic and molecular analysis of the *Enhancer* (*En*) transposable element system of *Zea mays*. *EMBO J.* **4**, 17-23.

Peterson, P.A. (1953). A mutable pale green locus in maize. *Genetics* **38**, 682-683.

Peterson, P.A. (1960). The pale green mutable system in maize. *Genetics* **45**, 115-133.

Peterson, P.A. (1970). The *En* mutable system in maize. III. Transposition associated with mutational events. *Theor. Appl. Genet.* **40**, 367-377.

Pietrzak, M., Shillito, R.D., Hohn, T. and Potrykus, I. (1986). Expression in plants of two bacterial antibiotic resistance genes after protoplast transformation with a new plant expression vector. *Nucl. Ac. Res.* **14**, 5857-5868.

Podila, G.K., Rogers, L.M. and Kolattukudy, P.E. (1993). Chemical signals from avocado surface wax trigger germination and appressorium formation in *Colletotrichum gloeosporioides*. *Plant Physiol.* **103**, 267-272.

Pospiech, A., Cluzel, B., Bietenhader, J. and Schupp, T. (1995). A new *Myxococcus xanthus* gene cluster for the biosynthesis of the antibiotic saframycin Mx1 encoding a peptide synthetase. *Microbiology* **141**, 1793-1803.

Post-Beittenmiller, D. (1996). Biochemistry and molecular biology of wax production in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 405-430.

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

Pyee, J., Hongshi, Y. and Kolattukudy, P.E. (1994). Identification of a lipid transfer protein as the major protein in the surface wax of broccoli (*Brassica oleracea*) leaves. *Arch. Biochem. Biophys.* **311**, 460-468.

Raina, R., Cook, D. and Fedoroff, N. (1993). Maize *Spm* transposable element has an enhancer-insensitive promoter. *Proc. Natl. Acad. Sci. USA* **90**, 6355-6359.

Reddy, A.S., Nuccio, M.L., Gross, L.M. and Thomas, T.L. (1993). Isolation of a Δ6-desaturase gene from the cyanobacterium *Synechocystis* sp. strain PCC6803 by gain-of-function expression in *Anabaena* sp. strain PCC7120. *Plant Mol. Biol.* **27**, 293-300.

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

Rerie, W., Feldmann, K. and Marks, M. (1994). The *GLABRA2* gene encodes a homeo domain protein required for normal trichome development in *Arabidopsis*. *Genes Dev.* **8**, 1388-1399.

Rick, C.M. (1948). Genetics and development of nine male-sterile tomato mutants. *Hilgardia* **18**, 599-633.

Saedler, H. and Nevers, P. (1985). Transposition in plants: a molecular model. *EMBO J.* **4**, 585-590.

Sakamoto, T., Wada, H., Nishida, I., Ohmori, M. and Murata, N. (1994). Identification of conserved domains in the Δ12 desaturases of cyanobacteria. *Plant Mol. Biol.* **24**, 643-650.

Sambrook, J., Fritsch, E.F. and Maniatis, R. (1989). *Molecular cloning, a laboratory manual* (2nd edn), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Schiefelbein, J.W., Raboy, V., Fedoroff, N.V. and Nelson, O.E., Jr. (1985). Deletions within a defective Suppressor-mutator element in maize affect the frequency and developmental timing of its excision from the *bronze* locus. *Proc. Natl. Acad. Sci. USA* **82**, 4783-4787.

Schläppi, M., Raina, R. and Fedoroff, N. (1994). Epigenetic regulation of the maize *Spm* transposable element: novel activation of a methylated promoter by TnpA. *Cell* **77**, 427-437.

Schmidt, R. and Willmitzer, L. (1989). The maize autonomous element *Activator* (*Ac*) shows a minimal germinal excision frequency of 0.2%-0.5% in transgenic *Arabidopsis thaliana* plants. *Mol. Gen. Genet.* **220**, 17-24.

Schmidt, R., West, J., Love, K., Lenehan, Z., Lister, C., Thompson, H., Bouchez, D. and Dean, C. (1995). Physical map and organization of *Arabidopsis thaliana* chromosome 4. *Science* **270**, 480-483.

Scholl, R. and Davis, K. (1995). *Seed and DNA stock list*. Arabidopsis Biological Resource Center, Columbus OH, USA.

Schwarz-Sommer, Zs., Gierl., A., Berndtgen, R. and Saedler, H. (1985a). Sequence comparison of 'states' of *a1-m1* suggests a model of Spm (En) action. *EMBO J.* **4**, 2439-2443.

Schwarz-Sommer, Zs., Gierl, A., Cuypers, H., Peterson, P.A. and Saedler, H. (1985b). Plant transposable elements generate the DNA sequence diversity needed in evolution. *EMBO J.* **4**, 591-597.

Schwarz-Sommer, Zs., Shepherd, N., Tacke, E., Gierl, A., Rohde, W., Leclercq, L., Mattes, M., Berndtgen, R., Peterson, P.A. and Saedler, H. (1987). Influence of transposable elements on the structure and function of the *A1* gene of *Zea mays*. *EMBO J.* **6**, 287-294.

Scofield, S.R., English, J.J. and Jones, J.D.G. (1993). High level expression of the *Activator* transposase gene inhibits the excision of *Dissociation* in tobacco cotyledons. *Cell* **75**, 507-517.

Scott, R., Hodge, R., Paul, W. and Draper, J. (1991). The molecular biology of anther differentiation. *Plant Science* **80**, 167-191.

Scott, R.J. (1994). Pollen exine - the sporopollenin enigma and the physics of pattern. In *Molecular and cellular aspects of plant reproduction*. R.J. Scott and A.D. Stead (eds). Cambridge University Press, Cambridge, UK. pp 49-81.

Seurinck, J., Truettner, J. and Goldberg, R.B. (1990). Nucleotide sequence of an anther-specific gene. *Nucl. Acids Res.* **18**, 3403.

Shanklin, J., White, E. and Fox, B.G. (1994). Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochemistry* **33**, 12787-12794.

Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M. and Goodman, H.M. (1995). Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. *Plant J.* **8**, 659-671.

Shockey, J.M., Rajasekharan, R. and Kemp, J.D. (1995). Photoaffinity labeling of developing jojoba seed microsomal membranes with a photoreactive analog of acyl-coenzyme A (acyl-CoA). *Plant Physiol.* **107**, 155-160.

Shure, M., Wessler, S. and Fedoroff, N. (1983). Molecular identification and isolation of the *Waxy* locus in maize. *Cell* **35**, 225-233.

Song, W.-Y., Wang, G.-L., Chen, L.-L., Kim, H.-S., Pi, L.-Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.-X., Zhu, L.-H., Fauquet, C. and Ronald, P. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* **270**, 1804-1806.

Southworth, D. (1990). Exine biochemistry. In *Microspores. Evolution and ontogeny*. S. Blackmore and R.B. Knox (eds). Academic Press, Londen. pp 193-212.

Spencer, D.F., Schnare, M.N., Coulthart, M.B. and Gray, M.W. (1992). Sequence and organization of a 7.2 kb region of wheat mitochondrial DNA containing the large subunit (26S) rRNA gene. *Plant Mol. Biol.* **20**, 347-352.

Springer, P.S., McCombie, W.R., Sundaresan, V. and Martienssen, R.A. (1995). Gene-trap tagging of *PROLIFERA*, an essential *MCM2-3-5*-like gene in *Arabidopsis*. *Science* **268**, 877-880.

Spurr, A.R. (1969). A low viscosity epoxy resin embedding medium for electron microscopy. *J. Ultrastruct. Res.* **71**, 173-184.

Städler, E. (1986). Oviposition and feeding stimuli in leaf surface waxes. In *Insects and the plant surface*. B. Juniper and R. Southwood (eds). Edward Arnold, London. pp 105-121.

Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.* **3**, 739-744.

Staskawicz, B.J., Ausubel, F.M., Baker, B.J., Ellis, J.G. and Jones, J.D.G. (1995). Molecular genetics of plant disease resistance. *Science* **268**, 661-667.

Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A. and De Vries, S.C. (1991). Cell-specific expression of the carrot EP2 lipid transfer protein gene. *Plant Cell* **3**, 907-921.

Strittmatter, P., Spatz, L., Corcoran, D., Rogers, M.J., Setlow, B. and Redline, R. (1974). Purification and properties of rat liver microsomal stearyl Coenzyme A desaturase. *Proc. Natl. Acad. Sci. USA* **71**, 4565-4569.

Stukey, J.E., McDonough, V.M. and Martin, C.E. (1990). The *OLE1* gene of *Saccharomyces cerevisiae* encodes the Δ9 fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase gene. *J. Biol. Chem.* **265**, 20144-20149.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D.G., Dean, C., Ma, H. and Martienssen, R. (1995). Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* **9**, 1797-1810.

Suzuki, M., Hayakawa, T., Shaw, J.P., Rekik, M. and Harayama, S. (1991). Primary structure of xylene

monooxygenase: similarities to and differences from the alkane hydroxylation system. *J. Bacteriol.* **173**, 1690-1695.

Swinburne, J., Balcells, L., Scofield, S.R., Jones, J.D.G. and Coupland, G. (1992). Elevated levels of *Activator* transposase mRNA are associated with high frequencies of *Dissociation* excision in Arabidopsis. *Plant Cell* **4**, 583-595.

Tacke, E., Schwarz-Sommer, Zs., Peterson, P.A. and Saedler, H. (1986). Molecular analysis of states of the *A1* locus of *Zea mays*. *Maydica* **31**, 83-91.

Thiede, M.A., Ozols, J. and Strittmatter, P. (1986). Construction and sequence of cDNA for rat liver stearyl Coenzyme A desaturase. *J. Biol. Chem.* **261**, 13230-13235.

Thoma, S., Kaneko, Y. and Somerville, C. (1993). A non-specific lipid transfer protein from *Arabidopsis* is a cell wall protein. *Plant J.* **3**, 427-436.

Thoma, S., Hecht, U., Kippers, A., Botella, J., De Vries, S. and Somerville, C. (1994). Tissue-specific expression of a gene encoding a cell wall-localized lipid transfer protein from *Arabidopsis*. *Plant Physiol.* **105**, 35-45.

Thompson, K.F. (1963). Resistance to the cabbage aphid (*Brevicoryne brassicae*) in Brassica plants. *Nature* **198**, 209

Töpfer, R., Martini, N. and Schell, J. (1995). Modification of plant lipid synthesis. *Science* **268**, 681-686.

Topping, J.F. and Lindsey, K. (1995). Insertional mutagenesis and promoter trapping in plants for the isolation of genes and the study of development. *Trans. Res.* **4**, 291-305.

Topping, J.F., Wei, W. and Lindsey, K. (1991). Functional tagging of regulatory elements in the plant genome. *Development* **112**, 1009-1019.

Tower, J., Karpen, G.H., Craig, N. and Spradling, A.C. (1993). Preferential transposition of Drosophila *P*-elements to nearby chromosomal sites. *Genetics* **133**, 347-359.

Tsay, Y.-F., Frank, M.J., Page, T., Dean, C. and Crawford, N.M. (1993). Identification of a mobile endogenous transposon in *Arabidopsis thaliana*. *Science* **260**, 342-344.

Tsuchiya, T., Toriyama, K., Ejiri, S. and Hinata, K. (1994). Molecular characterization of rice genes specifically expressed in the anther tapetum. *Plant Mol. Biol.* **26**, 1737-1746.

Tsugeki, R., Kochieva, E.Z. and Fedoroff, N.V. (1995). A transposon insertion in the *Arabidopsis SSR16* gene causes an embryo-defective lethal mutation. *Plant J.* **10**, 479-489.

Uknes, S., Dincher, S., Friedrich, L., Negrotto, D., Williams, S., Thompson-Taylor, H., Potter, S., Ward, E. and Ryals, J. (1993). Regulation of pathogenesis-related protein-1a gene expression in tobacco. *Plant Cell* **5**, 159-169.

Valvekens, D., Van Montagu, M. and Van Lijsebettens, M. (1988). *Agrobacterium tumefaciens* mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. *Proc. Natl. Acad. Sci. USA* **85**, 5536-5540.

Van den Elzen, P., Townsend, J., Lee, K. and Bedbrook, J. (1985). A chimaeric hygromycin resistance gene as a selectable marker in plant cells. *Plant Mol. Biol.* **5**, 299-302.

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.

Van Sluys, M.A., Tempé, J. and Fedoroff, N. (1987). Studies on the introduction and mobility of the maize *Activator* element in *Arabidopsis thaliana* and *Daucus carota*. *EMBO J.* **6**, 3881-3889.

Van Tunen, A.J., van der Meer, I.M. and Mol, J.N.M. (1994). Flavonoids and genetic modification of male fertility. In *Genetic control of self-incompatibility and reproductive development in flowering plants*. E.G. Williams, A.E. Clarke and R.B. Knox (eds). Kluwer Academic Publishers, Dordrecht, The Netherlands. pp 423-442.

Von Wettstein-Knowles, P. (1979). Genetics and biosynthesis of plant epicuticular waxes. In *Advances in the biochemistry and physiology of plant lipids*. L.-Å. Appelqvist and C. Liljenberg (eds). Elsevier/North-Holland Biomedical Press, Amsterdam. pp 1-26.

Von Wettstein-Knowles, P. (1994). Biosynthesis and genetics of waxes. In *Waxes: chemistry, molecular biology and functions*. R.J. Hamilton (ed). The Oily Press, Dundee, UK. pp 91-129.

Wada, H., Gombos, Z. and Murata, N. (1990). Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation. *Nature* **347**, 200-203.

Walbot, V. (1992). Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 49-82.

Walden, R., Koncz, C. and Schell, J. (1990). The use of gene vectors in plant molecular biology. *Meth. Mol. Cell. Biology* **1**, 175-194.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F. and Meyerowitz, E.M. (1992). *LEAFY* controls floral meristem identity in Arabidopsis. *Cell* **69**, 843-859.

Weigel, D. and Meyerowitz, E.M. (1993). Activation of floral homeotic genes in *Arabidopsis*. *Science* **261**, 1723-1726.

Weigel, D. and Nilsson, O. (1995). A developmental switch sufficient for flower initiation in diverse plants. *Nature* **377**, 495-500.

Wierenga, R.K., Terpstra, P. and Hol, W.G.J. (1986). Prediction of the occurrence of the ADP-binding ßαß-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* **187**, 101-107.

Wilson, C., Pearson, R.K., Bellen, H.J., O'Kane, C.J., Grossniklaus, U. and Gehring, W.J. (1989). P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes Dev.* **3**, 1301-1313.

Wilson, K., Long, D., Swinburne, J. and Coupland, G. (1996). A *Dissociation* insertion causes a semidominant mutation that increases expression of *TINY*, an Arabidopsis gene related to *APETALA2*. *Plant Cell* **8**, 659-571.

Wing, D., Koncz, C. and Schell, J. (1989). Conserved function in *Nicotiana tabacum* of a single *Drosophila hsp70* promoter heat shock element when fused to a minimal T-DNA promoter. *Mol. Gen. Genet.* **219**, 9-16.

Woodhead, S. and Padgham, D.E. (1988). The effect of plant surface characteristics on resistance of rice to the brown planthopper, *Nilaparvata lugens*. *Entomol. Exp. Appl.* **47**, 15-22.

Worrall, D., Hird, D.L., Hodge, R., Paul, W., Draper, J. and Scott, R. (1992). Premature dissolution of the microsporocyte callose wall causes male sterility in transgenic tobacco. *Plant Cell* **4**, 759-771.

Xu, H., Knox, R.B., Taylor, P.E. and Singh, M.B. (1995). *Bcp1*, a gene required for male fertility in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **92**, 2106-2110.

Zhang, P. and Spradling, A.C. (1993). Efficient and dispersed local *P* element transposition from Drosophila females. *Genetics* **133**, 361-373.

Zwaal, R.R., Broeks, A., van Meurs, J., Groenen, J.T.M. and Plasterk, R.H.A. (1993). Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc. Natl. Acad. Sci. USA* **90**, 7431-7435.

Samenvatting

Het doel van het onderzoek beschreven in dit proefschrift is het ontwikkelen, bestuderen en toepassen van een efficiënt gen-isolatie systeem in zandraket (*Arabidopsis thaliana*). Hierbij wordt gebruik gemaakt van de mutagene eigenschappen van *En/Spm-I/dSpm* transposons, zoals die van nature aanwezig zijn in maïs. Transposons zijn mobiele stukken DNA die zich door middel van excisie en insertie door het celkern-DNA van een organisme, bijvoorbeeld een plant, kunnen verplaatsen. Een insertie in een gen kan leiden tot de uitschakeling van dat gen, hetgeen het uiterlijk van de plant kan veranderen. Een gen dat gemarkeerd is ("tagged") door een transposon insertie, kan met behulp van standaard moleculair biologische technieken geïsoleerd worden voor verdere studie. Bij de ontwikkeling van een efficiënt *En/Spm-I/dSpm* transposon tagging systeem in Arabidopsis zijn een aantal belangrijke parameters bestudeerd, namelijk: 1) de mogelijkheid tot continue transpositie, onafhankelijk van het aantal generaties verstreken sinds transformatie; 2) de verdeling van *I/dSpm* elementen over het genoom na transpositie; 3) de mogelijkheid tot transpositie naar genomische regio's met transcriptionele activiteit; 4) de mogelijkheid tot het veroorzaken van mutaties.

Allereerst wordt in hoofdstuk 1 beschreven dat transgene Arabidopsis planten gemaakt zijn door introductie van een T-DNA fragment met daarop *En/Spm* transposase genen en een *I/dSpm* transposon. In de aanwezigheid de *En/Spm* transposase genen, verplaatst het *I/dSpm* transposon zich door het Arabidopsis-genoom. De transpositie-activiteit is gedefinieerd met behulp van de "onafhankelijk transpositie frequentie" (otf), waarmee het aandeel nieuwe, unieke *I/dSpm* inserties in het totale aantal *I/dSpm* inserties in de planten van een gegeven populatie wordt uitgedrukt. De otf van *I/dSpm* elementen is ongeveer 10% en deze blijft constant over meerdere generaties. Een otf van 10% houdt in dat in een populatie van 10 planten met elk één transposon, 1 plant een nieuwe transposon insertie bevat.

De verdeling van transposons over het genoom is bestudeerd door de genetische positie van verplaatste *I/dSpm* elementen ten opzichte van bekende RFLP markers te bepalen (hoofdstukken 1 en 5). De gekarteerde *I/dSpm* elementen zijn nagenoeg willekeurig verdeeld over het genoom. Toch lijkt er na transpositie een lichte voorkeur te zijn voor insertie op een genetisch gekoppelde nieuwe positie (hoofdstuk 1).

De studie van het *En/Spm-I/dSpm* systeem heeft meerdere aanwijzingen opgeleverd dat *I/dSpm* elementen in of dichtbij genen inserteren: *I/dSpm* flankerend genomisch DNA is vaak uniek; het is qua DNA volgorde geconserveerd tussen verschillende accessies (ecotypes) van Arabidopsis (hoofdstukken 1 en 5); en tenminste één derde van de *I/dSpm* flankerende genomische DNAs bevat gedeelten waarvan mRNA afgeschreven wordt (hoofdstuk 7).

Gedurende het onderzoek zijn tenminste 12 getagde mutanten gevonden. Hiervan zijn de *ms2*, *cer1*, *ap1* en *lad* mutanten in detail onderzocht, zoals beschreven in de hoofdstukken 2, 3, 4 en 6. De mannelijk steriele *ms2* mutant is ontstaan door insertie van een *I/dSpm* element in het *MALE STERILITY2* (*MS2*) gen betrokken bij de vorming van stuifmeelkorrels (hoofdstuk 2). Het *MS2* gen komt tot expressie in het tapetum op het moment dat losse microsporen gevormd worden (hoofdstuk 3). Dit heeft tot gevolg dat er om de *ms2*

microsporen geen beschermende exine laag wordt aangelegd, zodat uiteindelijk slechts weinig en zeer kwetsbare pollenkorrels gevormd worden. Het MS2 eiwit, afkomstig van het *MS2* gen, komt qua aminozuurvolgorde overeen met een was-vetzuurreductase uit jojoba. Dit wijst erop dat vetzuurreductie waarschijnlijk een belangrijke rol speelt bij de aanleg van pollenexine.

Het *CER1* gen (hoofdstuk 4) is betrokken bij de epicuticulaire waslaagvorming. Mutante *cer1* planten hebben een glanzend heldergroen uiterlijk door een afwijkende waslaagsamenstelling, terwijl planten normaal mat grijsgroen gekleurd zijn. Ook *cer1* planten zijn mannelijk steriel. Steriliteit wordt hier veroorzaakt doordat *cer1* stuifmeelkorrels na bestuiving geen kiembuis vormen. Het CER1 eiwit bevat enkele karakteristieke aminozuurvolgordes die ook gevonden worden in vetzuurdesaturases. Mede op basis van de biochemische analyses van *cer1* waslaagbestanddelen is geconcludeerd dat het CER1 eiwit betrokken is bij de decarbonylatie van langketenige aldehyden naar alkanen, die nodig zijn in de waslaag en in pollenkit.

Verdere toepassingen van het *En/Spm-I/dSpm* transposon systeem zijn bestudeerd, zoals het effect van *En/Spm* expressie op de expressie van genen gemuteerd door *I/dSpm* insertie (hoofdstuk 6). Dergelijke inserties zijn gevonden in het *AP1* en het *LAD* gen. Het *AP1* gen zorgt normaliter voor de vorming van bloemblaadjes. Het *LAD* gen is nodig voor de tijdige opening van helmhokjes. De *ap1::I/dSpm* mutant vertoont een sterk mutant fenotype (geen bloemblaadjes) in aanwezigheid van het transposase T-DNA, en een zwak mutant fenotype in afwezigheid van dit T-DNA. Dit in tegenstelling tot *lad::I/dSpm* planten, die in aanwezigheid van het transposase T-DNA nauwelijks afwijken van het wild-type, maar in afwezigheid van het transposase T-DNA gedeeltelijk mannelijk steriel zijn. Dit effect van *En/Spm* transposase eiwitten lijkt voornamelijk bepaald te worden door de oriëntatie en de positie van de *I/dSpm* insertie in het gemuteerde gen.

Tenslotte wordt in hoofdstuk 7 beschreven hoe Arabidopsis is getransformeerd met een serie van verschillende *I/dSpm* "gene trap elements", die vervolgens getest zijn voor transpositie. De *I/dSpm* elementen zijn uitgerust met een promoterloos β-glucuronidasegen (GUS gen), dat niet tot expressie komt, tenzij in aanwezigheid van naburige transcriptieregulatie sequenties. Omdat deze sequenties over het algemeen dichtbij genen liggen, wordt een dergelijk GUS-transposon gebruikt voor het opsporen van genactiviteit. In een aantal planten met een gesprongen *I/dSpm* "gene trap element" is weefselspecifieke GUS-expressie gevonden, en daarmee is aangetoond, dat deze gemodificeerde *I/dSpm* transposons inderdaad gebruikt kunnen worden voor het opsporen van genactiviteit. Dit kan zonder dat een gen daarvoor gemuteerd hoeft te worden.

Concluderend, er is een efficiënt *En/Spm-I/dSpm* transposon tagging systeem ontwikkeld en bestudeerd, dat verdere perspectieven biedt voor genisolatie en genfunctieanalyse in Arabidopsis. Het transposon systeem is succesvol gebruikt voor de isolatie van drie, tot dan toe onbekende, genen, die betrokken zijn bij mannelijke fertiliteit.

Nawoord

Dit was het dan, mijn proefschrift. Althans bijna. Wat rest is het nawoord, waarin ik de mensen met name wil noemen die de afgelopen jaren bijgedragen hebben tot het tot stand komen van dit proefschrift.

Allereerst mijn co-promoter, Andy Pereira. Andy, ik herinner me nog ons eerste gesprek ter introductie van mijn afstudeervak plantenveredeling, waarbij je me uitlegde wat transposons zijn. Het hele idee van mobiele stukjes DNA die simpelweg genen kunnen uitschakelen leek complex en druiste volkomen in tegen het statische beeld dat ik had van genetica. Gelukkig zag ik de mogelijkheden die transposon tagging biedt voor de moleculaire plantenveredeling en ging ik de uitdaging aan om een fenotypische excisie assay te maken voor transpositie van het *Ac* element in aardappel. Mijn belangstelling voor transposons was zodanig gewekt dat ik maar wat graag ja zei op je vraag of ik zin had om na mijn afstuderen nog drie jaar te komen werken aan een EG-project voor het ontwikkelen van een transposon tagging systeem in Arabidopsis. Misschien zat er zelfs een promotie in! Inmiddels zijn dit alweer ruim vijf prettige jaren, waarin je me de fijne kneepjes van het wetenschappelijk onderzoek bijgebracht hebt en met die promotie is het ook nog zover gekomen. Andy, bedankt voor je support, het is mede dankzij jouw inzet en enthousiasme dat ik zover gekomen ben.

Dan mijn promoter, Maarten Koornneef. Het nadeel van een driejarig onderzoekscontract was dat er niet direct de voorzieningen waren om het project af te sluiten met een promotie. Een promotor bijvoorbeeld heeft lange tijd ontbroken. Gelukkig werd jij in de tussentijd hoogleraar, en ik ben blij dat je bereid was mijn promotor te worden. Hoewel het merendeel van het onderzoek al was verricht, heb jij een belangrijke invloed gehad op de inhoud en opbouw van dit proefschrift. Dankzij jouw kennis van de Arabidopsis wereld heb ik het onderzoek in een breder kader kunnen plaatsen dan alleen transposon tagging. Wat begon als promotiebegeleiding is inmiddels uitgegroeid tot een goede wetenschappelijke samenwerking, en ik hoop dat nog lange tijd zo door mag gaan.

Al mijn tijd bij het ITAL, het CPO en vervolgens het CPRO-DLO, heb ik gewerkt binnen de afdeling Moleculaire Biologie van Willem Stiekema. Willem, door jou als afdelingshoofd heb ik me altijd gesteund geweten. Ondanks dat je transposon tagging in het begin waarschijnlijk meer genetica dan moleculaire biologie vond en het daardoor wat verder van je bed stond dan je lief was. In dit geval was onbekend niet onbemind. De vrijheid van onderzoek die ik gekregen heb, en je geduld dat nodig was voor de voltooiing van dit proefschrift heb ik erg gewaardeerd. En niet te vergeten dat jouw bemoeiienis met het aanvragen van patenten er toe geleid heeft dat ik me nu "uitvinder" kan noemen.

In de loop der jaren heb ik een flink aantal collega's gekend. Sommige kwamen en gingen, maar één is altijd gebleven. Bas, van het begin af aan hebben we samengewerkt en al lange tijd ben je mijn kamergenoot. Zonder jouw practische hulp en inzet voor het reilen en zeilen in het lab was ik waarschijnlijk vele uren extra kwijt geweest aan het voorbereiden en uitvoeren van mijn experimenten. Ik hoop dat je ook het een en ander van mij opgestoken

hebt, al geloof ik dat je aversie tegen planten wel helemaal nooit zal verdwijnen!

Van alle andere collega's die het CPRO tot een wetenschappelijk inspirerende werkomgeving maakten, zijn er een paar die ik met name wil noemen. Wim, Hans, Dion, Jeanne, Nelleke, Ruud, met z'n allen hebben we er binnen en buiten het lab altijd een gezellige boel van kunnen maken. Also, Christiane, I should not forget to mention you. You were the first with whom I really had to cooperate. Sharing the monopoly on Arabidopsis transposon tagging was an educative experience for me. I am glad it turned out all right in the end, and I am still sorry that you were not able to continue your stay much longer.

Drie studenten kruisten mijn pad en leverden een aanzienlijke bijdrage aan het *En-I* tagging onderzoek. Paul Corzaan, hoewel onze tijd op het ITAL maar een maand overlapte, was het zonder jouw eerste Arabidopsis-transformatie allemaal een stuk langzamer gegaan. Anne Mulder, jij was mijn eerste echte student. Behalve met de nodige vrolijke noten in het lab ben ik nog steeds blij met de hoeveelheid werk die je hebt kunnen doen tijdens je stage. Tenslotte Krista Kooi, zoals je kunt zien heeft jouw inzet grotendeels geleid tot de data waarmee ik hoofdstuk 7 heb kunnen schrijven. Allemaal bedankt.

Ik ben ook de nodige dank verschuldigd aan Gerrit Stunnenberg, Gerard Scholten en Henk Oostingh. Hoewel het even duurde voordat jullie Arabidopsis onder de knie hadden (wat is het nou ook voor gewas!), zou ik niet zonder jullie verzorging van de planten gekund hebben. Verder zijn de meeste foto's in dit proefschrift niet door mijzelf genomen, maar dankzij het vakkundig inzicht van Peter Stadt, Joop van Os, Frank Klinge en Dick Vermeer, van de fotodienst van DLO tot stand gekomen.

Some of the chapters bear the names of others then CPRO-DLO authors. Chapter 3 would have been very different, and not as far as interesting, without the contributions of Rachel Hodge and Rod Scott from the University of Leicester and Kriton Kalantidis, Zoe Wilson and Bernard Mulligan from the University of Nottingham. Many thanks for sharing your results, that will make publication of this chapter possible. Special thanks to Zoe and Bernard, for introducing male sterility in Arabidopsis to me during my stay in Nottingham as part of my undergraduate degree. Hoofdstuk 4 zou waarschijnlijk in een heel wat minder aanzienlijk tijdschrift gepubliceerd zijn zonder de enthousiaste inzet van Koos Keijzer bij de karakterisatie van de *cer1* waslaagmutant. Zo'n soepele samenwerking hoop ik nog eens mee te maken. Mijn hartelijke dank gaat ook uit naar Ton Peeters en Hiroyoshi Kubo, voor het gebruik van ongepubliceerde gegevens voor het schrijven van hoofdstuk 5. Also my special thanks to Marina Byzova, for communicating her results on the isolation of the *SAP1* gene.

Tenslotte het thuisfront. Over het gezinsleven van promovendi is al het nodige geschreven, en ik ben bang dat ik geen verandering kan brengen in het beeld van schrijnende verwaarlozing dat de niet-promoverende helft van dit tweeverdienersstel meestal ten deel valt. Promotieonderzoek is geen negen tot vijf baan, dat is duidelijk. Desalniettemin, Coco, heb je me enorm gesteund om dit onderzoek tot een goed einde te brengen. Zelfs als ik voor de zoveelste keer te laat thuis kwam of als mijn inbreng in het avondeten niet verder reikte dan alweer macaroni met ham en kaas. En als ik het al een keer niet meer zag zitten, wist jij me toch altijd wel weer op te peppen. Jouw inbreng in mijn proefschrift is daarom waarschijnlijk het grootst en het minst zichtbaar. Ik zal het niet vergeten. Bedankt!

Curriculum vitae

Martinus Gerardus Maria Aarts werd op 20 juni 1968 in Boxmeer geboren. Na het behalen van het VWO diploma aan het Elzendaalcollege in Boxmeer, begon hij in september 1986 met de studie Plantenveredeling aan de Landbouwuniversiteit te Wageningen. In september 1987 slaagde hij met lof voor het propaedeuse examen, en in maart 1991 werd de studie afgesloten met het behalen van het ingenieursdiploma, eveneens met lof. Tijdens de doctoraalfase volgde hij de afstudeervakken Plantenveredeling (bij de toenmalige stichting Ital) en Erfelijkheidsleer. Het laatste vak werd gecombineerd met een praktijktijd Erfelijkheidsleer bij het 'Department of Botany, University of Nottingham' in het Verenigd Koninkrijk. Tussen april 1991 en april 1994 werkte hij als wetenschappelijk onderzoeker bij de afdeling Moleculaire Biologie van de Stichting ITAL, in 1992 opgegaan in het Centrum voor Plantenveredelings- en Reproductieonderzoek van de Dienst Landbouwkundig Onderzoek (CPRO-DLO). Tijdens deze periode werd een door de Europese Gemeenschap gefinancierd BRIDGE onderzoeksproject uitgevoerd, waarvan de resultaten in dit proefschrift beschreven zijn. Vanaf oktober 1994 is hij in dienst bij de Nederlands Organisatie voor Wetenschappelijk Onderzoek (NWO), waarvoor hij werkt aan de isolatie van resistentiegenen tegen valse meeldauw uit Arabidopsis, als wetenschappelijk onderzoeker gedetacheerd bij het CPRO-DLO.

Appendix A

Protocol for the analysis of a putatively *I/dSpm* tagged mutant and the cloning of a tagged gene.

1. Harvest leaf material from a mutant for DNA isolation. Cross the mutant with wild type Landsberg *erecta* (cross 1) and with the *En/Spm* transposase TEn5 line if the mutant was not known to express transposase (cross 2).
2. Determine the heritability of the phenotype in the selfed offspring of the mutant and determine the presence of a transposase locus on hygromycin containing medium. If the mutant contains a transposase locus, screen 500-1000 progeny for wild type looking revertants to establish the stability of the mutant phenotype.
3. If the mutant does not contain a transposase locus, screen siblings to find a family expressing transposase. If found, screen the progeny as in 2. Alternatively, screen the F_2 from cross 2 for families with transposase. When found, screen progeny as in 2.
4. Identify a transposon insertion cosegregating with the mutation. Preferably use a population without transposase (e.g. the F_2 from cross 1), segregating 3:1 for wild type to mutant. Alternatively, use a population with transposase (e.g. the F_2 from cross 2), or revertant and mutant progeny of the original mutant when the mutant contained the transposase locus. Perform a DNA blot analysis on about 20-50 plants. When equal amounts of DNA are loaded per lane it is possible to distinguish between homozygous and hemizygous inserts. Use other populations if no cosegregating transposon can be found.
5. Isolate genomic DNA flanking the cosegregating transposable element by IPCR (Appendix B), using preferably DNA template from a plant lacking the transposase locus and carrying five or less copies of transposon inserts. If no such plant can be found, make a backcross of the mutant with wild type (F_1 cross 1 x wild type) to generate a new population with reduced transposon copy number.
6. Confirm the cloning of genomic DNA flanking the transposon by:
   a. hybridizing the IPCR probe to a DNA blot of mutant and revertant plants. Mutant plants must show homozygous inserts, revertant plants must show hemizygous inserts.
   b. analysis of the insertion site. Determine the sequence of the IPCR fragments and design PCR primers flanking the insertion site. PCR amplify the insertion site sequences from the wild-type allele and from mutant and revertant excision alleles. Clone the PCR products and determine their DNA sequence. All revertants should have at least one allele with (near) wild type DNA sequence. All mutants should have only alleles featuring an insert, frame shifts, aberrant terminations or amino acid exchanges.
7. Isolate genomic and cDNA clones from the appropriate libraries using the IPCR products as probes, and determine their DNA sequence. There have to be parts of the genomic and cDNA clones with (nearly) 100% identity to the *I/dSpm* IPCR fragment to be derived from the tagged gene.

Appendix B

Protocol for the isolation of *I/dSpm* flanking Arabidopsis DNA by Inverse PCR

1x DNA extraction buffer: 0.3 M NaCl, 50 mM Tris-HCl (pH 7.5), 20 mM EDTA, 2% sarkosyl, 0.5% SDS, 5M Urea, 5% phenol (equilibrated) (Shure et al., 1983). The first five ingredients are mixed as a 2X stock solution before urea and phenol are added and the total volume is adjusted with water.

A. *Isolation of DNA from a single plant* (after Shure et al. (1983), adapted for Arabidopsis by Robert Whittier, Mitsui Plant Biotechnology Research Institute, Tsukuba, Japan)

1. Harvest 100-150 mg of leaves (two to three young leaves) or inflorescence tips (flowers plus buds) per plant in a 1.5 ml micro tube, freeze in liquid N_2.
2. Cool steel pestle for 3 seconds in liquid N_2. Grind material in micro tube to a fine powder. Add 200 μl of DNA extraction buffer (room temperature) and grind once more. Add another 400 μl of extraction buffer and mix by vortexing. Leave samples at room temperature (RT) until 18 or 24 samples are prepared to fit a micro tube centrifuge.
3. Extract the samples by adding (600 μl) phenol/chloroform. Take 500 μl of supernatant and precipitate DNA by adding 350 μl i-propanol. Keep tubes at RT for 5 min, if DNA precipitate can be seen, then spin for 5 min in micro tube centrifuge at full speed. Otherwise, keep sample at -20°C for 20 min, before centrifugation. Wash DNA pellet with 70% ethanol and dry briefly.
4. Dissolve DNA pellets in 100 μl TE (Sambrook et al., 1989) containing 10 mg/ml RNase. DNA samples may be stored at 4°C (few months) or at -20°C. Determine DNA concentration by comparing on agarose gel with standard amounts of λ DNA.

B. *Preparation of DNA template for IPCR*

1. Digest 200-300 ng DNA with 10 units HinfI in 100 μl 1xHinfI buffer containing 1mM spermidine ($\geq$3 hours at 37°C).
2. Add 1 μl 2.5mM dNTPs and 1 unit of DNA polymerase I Klenow fragment. Incubate for 5 min at RT. Phenol/chloroform extract and precipitate DNA with NaAc/i-propanol for 20 min at -20°C (Sambrook et al., 1989).
3. Centrifuge samples at maximum speed for 20 min, wash DNA pellet in 70% ethanol, and dry. Resuspend in 149 μl 1x ligation buffer (Sambrook et al., 1989). Add 2.5 units (1μl) of T4 DNA ligase and self-ligate DNA fragments overnight (o/n) at 14°C.
4. Inactivate the ligase by heating the samples at 70°C for 5 min.
5. Add 25 μl SalI-buffer (0.25 M Tris-HCl, pH7.5; 10 mM $MgCl_2$; 0.6 M NaCl) and 10 units of SalI to 75 μl of ligated DNA. Add 25 μl XbaI-buffer (50 mM Tris-HCl, pH8.0; 10 mM $MgCl_2$; 0.2 M NaCl) and 10 units of XbaI to the remaining 75 μl ligated DNA and incubate both digests at 37°C ($\geq$3 hours).

6. NaAc/i-propanol precipitate both SalI and XbaI digested DNA samples (see step 3), wash the pellet with 70% ethanol and dry. Resuspend DNAs in 30 μl H_2O.

C. *Inverse PCR*

1. Transfer DNA template into a PCR tube and add 4 μl 10x PCR buffer, 2 μl primer LJ1(see below), 2 μl primer RJ1 (both at 120 ng/μl) and 2 μl dNTPs (2.5 mM each). Prepare 10 μl of 1x PCR buffer with 2.5 units of Taq DNA polymerase. More samples can be processed in batch.
2. PCR reaction: -3 min at 92°C
   -add 10 μl Taq DNA polymerase solution.
   -set 25 cycles of PCR: 30 sec 95°C, 30 sec 55°C, 3 min 72°C.
   -elongation for 5 min at 72°C.
3. Transfer 2 μl aliquot to a new PCR tube. Add 38 μl 1x PCR buffer, containing 2 μl primer LJ2, 2 μl primer RJ2 (both at 120 ng/μl), and 2 μl dNTPs (2.5 mM each).
4. Second PCR for 25 cycles using the same conditions as described in step 2.

D. *Cloning of IPCR fragments*

1. Dilute 45 μl IPCR product to 100 μl by adding 55 μl 1x Klenow buffer containing 1 unit DNA polymerase I Klenow fragment and dNTPs (Sambrook et al., 1989). Incubate for 5 min, phenol/chloroform extract and NaAc/i-propanol precipitate. Centrifuge sample (20 min), wash DNA pellet with 70% ethanol, resuspend in 25 μl TE and size separate on a 1.2% TBE-agarose gel. Cut out the DNA bands from the gel, elute and clone (Sambrook et al., 1989) in an appropriate vector (e.g. pBluescript SK^+).
2. To obtain probes with very little *I/dSpm* element sequence, use the cloned IPCR fragments for another PCR with one primer annealing to the terminal inverted repeats (TIR) of the *I/dSpm* element. As an alternative to D1, the gel-eluted DNA from the IPCR, or 2 μl of the whole IPCR product can be used for this PCR.
3. Use 25 ng of linearized plasmid in a 50 μl PCR (see C), containing 2 μl of TIR primer (at 105 ng/μl), but with annealing at 50°C instead of 55°C.
4. Clone PCR fragment as described in step 1. Alternatively, dilute 10 μl of PCR product in 90 μl 1xBglII reaction mix and digest with BglII. Gel elute DNA fragments, and clone in BamHI digested and phosphatase treated (Sambrook et al., 1989) pBluescript SK^+.

Primer sequences:		5' 3'
1st PCR:	LJ1:	GAA TTT AGG GAT CCA TTC ATA AGA GTG T
	RJ1:	TTG TGT CGA CAT GGA GGC TTC CCA TCC GGG GA
2nd PCR:	LJ2:	ATT AAA AGC CTC GAG TTC ATC GGG A
	RJ2:	AGG TAG TCG ACT GAT GTG CGC GC
TIR primer:		GAC ACT CCT TAG ATC TTT TCT TGT AGT G