

# T-DNA tagging in *Medicago truncatula*

Marije Scholte

Promotor: prof.dr. T. Bisseling, hoogleraar in de moleculaire biologie

Promotiecommissie:

Dr. P. Ratet (Centre National de la Recherche Scientifique, Gif sur Yvette Cedex, France)

Prof.dr. H. Spaink (Universiteit Leiden)

Prof.dr. S.C. de Vries (Wageningen Universiteit)

Dr. A.B. Pereira (Plant Research International)

Dr. A.D.L. Akkermans (Wageningen Universiteit)

Marije Scholte

T-DNA tagging in *Medicago truncatula*

Proefschrift  
ter verkrijging van de graad van doctor  
op gezag van de rector magnificus  
van Wageningen Universiteit,  
prof.dr. ir. L. Speelman,  
in het openbaar te verdedigen  
op maandag 18 november 2002  
des namiddags te vier uur in de Aula.

Scholte, Marije

T-DNA tagging in *Medicago truncatula*

Thesis Wageningen- with ref.- with summary in Dutch.

ISBN 90-5808-740-9

Subject headings: *Medicago truncatula*, nodulation, T-DNA tagging.

# Contents

<b>Outline</b>	<b>7</b>
<b>1 General introduction</b>	<b>9</b>
1.1 The symbiotic partners involved in nitrogen fixation	9
1.2 Nodule development	10
1.3 Nod factor perception and signal transduction	13
1.4 Physiological control of nodulation and the role of plant growth regulators	15
1.5 Nodulins	16
1.6 Nodulation mutants	18
1.7 T-DNA tagging	24
1.8 Aim of the thesis	26
<b>2 T-DNA tagging in the model legume <i>Medicago truncatula</i> allows efficient gene discovery</b>	<b>27</b>
2.1 Introduction	27
2.2 Materials and methods	28
2.3 Results	30
2.4 Discussion	35
<b>3 Generation and screening of a tagged <i>M. truncatula</i> population</b>	<b>39</b>
3.1 Introduction	39
3.2 Results	39
3.3 Discussion	47
3.4 Material and methods	49
<b>4 Isolation and sequence analysis of T-DNA flanking regions in the tagged plants</b>	<b>53</b>
4.1 Introduction	53
4.2 Results	54
4.3 Discussion	61
4.4 Materials and methods	63
<b>5 T-DNA insertion in line B9-1 disrupted the <i>MtN3</i>-like gene which encodes a putative seven transmembrane protein</b>	<b>65</b>
5.1 Introduction	65
5.2 Results	65
5.3 Discussion	73
5.4 Materials and Methods	74
<b>6 Disruption of the <i>NARF</i>-like gene in line A7-1 causes a dwarf phenotype</b>	<b>77</b>
6.1 Introduction	77
6.2 Results	77
6.3 Discussion	83
6.4 Materials and methods	84
<b>7 Generation of an improved T-DNA tagging vector</b>	<b>87</b>
7.1 Introduction	87
7.2 Results	89
7.3 Discussion	92
7.4 Materials and methods	93
<b>8 Concluding remarks</b>	<b>95</b>
<b>9 Bibliography</b>	<b>97</b>
<b>Samenvatting</b>	<b>107</b>
<b>Curriculum vitae</b>	<b>109</b>



# Outline

Symbiotic interaction between rhizobia and leguminous plants leads to the formation of N<sub>2</sub>-fixing root nodules. This interaction shows a high degree of host specificity based on the exchange of chemical signals between the symbiotic partners. Although much is known about the bacterial genes required for the establishment of this interaction, less is known about the plant genes involved.

Symbiotic plant mutants serve as a tool to identify genes involved in nodulation and to study their function. However, the isolation of the affected gene(s) found in natural populations or induced by 'classical' methods like EMS treatment or irradiation is difficult, and so far only few of such genetically identified genes resulting in a mutant phenotype have been isolated. Problems to isolate a mutated gene can be overcome by the use of T-DNA as a mutagen, because the known nucleotide sequence of the T-DNA can serve as a starting point for the identification of the flanking sequences. T-DNA tagging not only facilitates the cloning of new genes, but at the same time provides insight in the function of the identified genes. In addition, when the T-DNA tag contains a promoter-less reporter gene transcriptional and translational gene fusions may provide new marker genes for the different stages of nodule development. The T-DNA inserts randomly in the euchromatic region of the plant genome and the chance to tag a gene of interest will increase with the size of the mutagenized population. This implies that an efficient transformation of the plant should be possible. A simple and efficient protocol for the transformation by co-cultivation with *Agrobacterium tumefaciens* and the subsequent regeneration of *Medicago truncatula* line R108-1 has been developed (Trinh *et al.*, 1998; Kamaté *et al.*, 2000). Besides the ease of transformation other features such as a small diploid genome, the ability to establish effective nitrogen-fixing symbiosis with a wide range of *Sinorhizobium* strains, high synteny with *Medicago sativa* and pea and the existence of numerous symbiotic mutants make *M. truncatula* a good model plant to study the establishment of symbiosis of the indeterminate nodule type.

The aim of this thesis work was to assess the possibilities of T-DNA tagging as a tool to discover *M. truncatula* genes involved in symbiosis and to study their function. The practical work was carried out in the laboratory of Dr. E. Kondorosi at the Institut des Sciences Vegetales (Centre National de la Recherche Scientifique) in Gif-sur-Yvette (France) within a TMR network project (project ERB4061PL970121).

The introducing **chapter 1** gives an overview of morphological and physiological aspects of the establishment of the symbiosis between rhizobia and leguminous host plants forming indeterminate root nodules. In addition, several T-DNA tagging strategies are described. In **chapter 2** the results obtained using different T-DNA vectors are compared to explore their use for gene tagging in *M. truncatula*. **Chapter 3** describes the production and screening of a tagged *M. truncatula* population using one of the vectors tested in chapter 2 and in **chapter 4** the isolation and characterization of the T-DNA flanking regions of a selection of these tagged lines is described. In **chapter 5** and **chapter 6** two tagged plant lines are discussed in more detail, and in **chapter 7** some considerations about the improvement of the tagging vector are presented. Finally, in **chapter 8**, the possibilities of T-DNA tagging in *M. truncatula* are discussed.

You will find a full-colour pdf version of this thesis at <http://www.bib.wau.nl/wda/>.





# 1 General introduction

Nitrogen is a key element in many of the compounds present in (plant) cells. The atmosphere contains vast quantities of molecular nitrogen ( $N_2$ ), about 78 % of the atmosphere by volume.  $N_2$  however, can not be utilized by animals, plants and most prokaryotes. One way to convert atmospheric  $N_2$  to utilizable combined N is biological nitrogen fixation by reduction of nitrogen gas into ammonia. This can be performed by a few prokaryotes, either alone by free-living bacteria or by bacteria in symbiosis with plants. Among the symbiotic associations, the *Rhizobium*-Leguminosae interactions are the most important with respect to biological nitrogen fixation. During this symbiosis the bacteria are hosted in a special plant organ, the nodule. The nodules provide the optimal micro-aerobic conditions for the nitrogen fixation process. In return for the fixed nitrogen the plant host supplies the bacteria with the photosynthetic products, which the bacteria use as carbon source. In the following, several aspects involved in establishing the symbiosis between rhizobia and their host plants will be discussed.

## 1.1 The symbiotic partners involved in nitrogen fixation

Soil bacteria that are able to establish a symbiosis with legumes belong to the genera *Rhizobium*, *Bradyrhizobium*, *Allorhizobium*, *Azorhizobium*, *Sinorhizobium* and *Mesorhizobium* (Dénarié *et al.*, 1992; Van Berkum and Eardly, 1998; De Lajudie, 1998). The genera *Rhizobium* and *Sinorhizobium* in general nodulate roots of temperate legumes like pea, alfalfa and vetch. *Bradyrhizobium*, *Mesorhizobium*, *Allorhizobium* and *Azorhizobium* rather associate with tropical and subtropical legumes like soybean and bean (see Table 1.1).

The Leguminosae family contains between 1600 to 1900 species, distributed among three subfamilies: the Papilionoideae, the Mimosoideae and the Caesalpinoideae. In contrast to the Caesalpinoideae, which have only few nodulating species, the majority of the members from the Papilionoideae and the Mimosoideae can establish  $N_2$ -fixing symbiosis with rhizobia. The only non-legume known to establish a symbiosis with rhizobia is *Parasponia andersonii*, from the Ulmaceae family. Phylogenetic analysis of DNA sequences of the chloroplast gene *RBCL* and the ribosomal 18S rRNA gene indicates that both the Leguminosae and the Ulmaceae belong to the Rosid I clade and therefore are rather close relatives (Soltis *et al.*, 1995; Doyle, 1998).

Table 1.1 Some examples of symbiotic host range (after Brewin, 1998;Dénarié *et al.*, 1992; De Lajudie *et al.*, 1998).

<i>Rhizobia</i>	Host plants
<i>Azorhizobium</i>	
<i>Azorhizobium caulidans</i>	<i>Sesbania</i>
<i>Bradyrhizobium</i>	
<i>Bradyrhizobium japonicum</i>	<i>Glycine</i> , <i>Vigna</i> , <i>Macroptilium</i>
<i>Bradyrhizobium</i> ssp.	<i>Vigna</i> , <i>Parasponia</i> (non legume)
<i>Mesorhizobium</i>	
<i>Mesorhizobium loti</i>	<i>Lotus</i> , <i>Lupinus</i> , <i>Anthyllis</i> , <i>Dorycnium</i>
<i>Rhizobium</i>	
<i>Rhizobium leguminosarum</i>	
biovar <i>phaseoli</i>	<i>Phaseolus</i>
biovar <i>trifolii</i>	<i>Trifolium</i>
biovar <i>viciae</i>	<i>Pisum</i> , <i>Vicia</i> , <i>Lathyrus</i> , <i>Lens</i>
<i>Rhizobium lupini</i>	<i>Lupinus</i> , <i>Ornithopus</i> , <i>Lotus</i>
<i>Rhizobium tropici</i>	<i>Phaseolus</i> , <i>Leucaena</i>
<i>Rhizobium</i> sp.	
BR816	<i>Leucaena</i> , <i>Phaseolus</i> , <i>Parasponia</i> (non legume)
NGR234, MPIK1030	Broad host range and also the non-legume <i>Parasponia</i>
<i>Rhizobium galegae</i>	<i>Galega</i>
<i>Sinorhizobium</i>	
<i>Sinorhizobium fredii</i>	<i>Glycine</i> , <i>Vigna</i>
<i>Sinorhizobium meliloti</i>	<i>Medicago</i> , <i>Melilotus</i> , <i>Trigonella</i>
<i>Allorhizobium</i>	
<i>Allorhizobium undicola</i>	<i>Neptunia</i>

## 1.2 Nodule development

Two types of legume root nodules exist: nodules of the determinate type and nodules of the indeterminate type (Fig. 1.1). The type of root nodule that will be formed depends on the host plant and not on the rhizobial strain (Hirsch, 1992). Differences between the two nodule types are the site of first internal cell divisions, nodule growth, and the form of the mature nodule (Hirsch, 1992). Determinate nodules are formed on tropical and subtropical legumes such as *L. japonicus*, soybean and bean and indeterminate nodules are formed on temperate legumes such as *M. truncatula*, pea, alfalfa and vetch. A comparison between the two nodule types is given in Table 1.2.

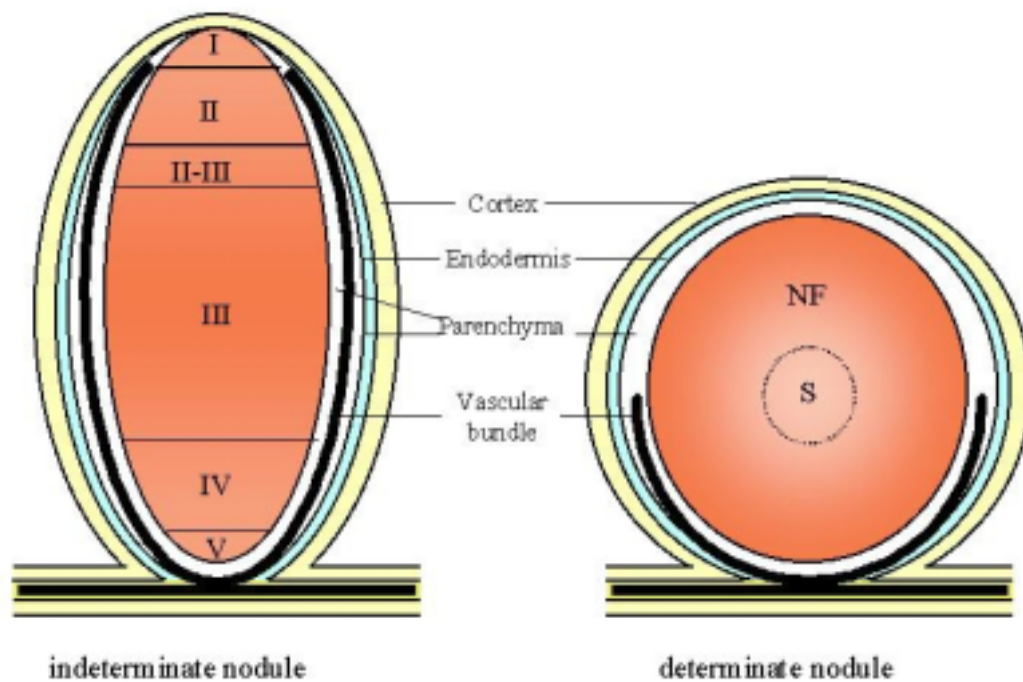


Figure 1.1 Structure of indeterminate and determinate nodules. I: meristem, II: prefixation zone, II-III: interzone II-III, III: nitrogen fixation zone, IV: senescent zone, V: saprophytic zone, S: senescent zone, NF: nitrogen fixation zone (adapted from Pawslowski and Bisseling, 1996).

Table 1.2 Comparison of several characteristics of indeterminate and determinate nodule types (adapted from Hirsch, 1992).

	Indeterminate nodule	Determinate nodule
Site of initial cell divisions	Inner cortex	Outer cortex
Meristem	Persistent meristem	No persistent meristem
Nodule form	Elongated	Round shaped
Geographic region of plant origin	Temperate regions	Subtropical and tropical
Example of symbiotic association	<i>Medicago-Sinorhizobium</i>	<i>Glycine-Bradyrhizobium</i>

This thesis work has been carried out with the legume *M. truncatula*, which forms indeterminate nodules and therefore only the development of this nodule type will be described.

### 1.2.1 Pre-infection stage

The interaction between the two symbiotic partners starts when the rhizobia are attracted by chemotaxis by flavonoids excreted in the rhizosphere from the host plant (Fig. 1.2 a). Although this chemotaxis is not indispensable for the establishment of the symbiosis (Kijne, 1992), it was found that root flavonoid production in alfalfa increased when the plants were grown under nitrogen limitation (Coronado *et al.*, 1995).

Once the bacteria are attracted by the root exudates, the rhizobia become attached to the surface of the root hair tips (Fig. 1.2 a). A variety of bacterial cell-surface components like exopolysaccharides, lipopolysaccharides, capsular polysaccharides, rhicadsin, cellulose fibrils and bacterial lectins seem to contribute to attachment and bacterial aggregation at the root surface, although for none of these components it has been demonstrated that they are essential (reviewed in Kannenberg and Brewin, 1994). Also the plant symbiotic partner seems to have a role in the attachment and Hirsch (1999) proposed that lectins could act as a 'glue' attaching the rhizobia to a site on a susceptible root hair.

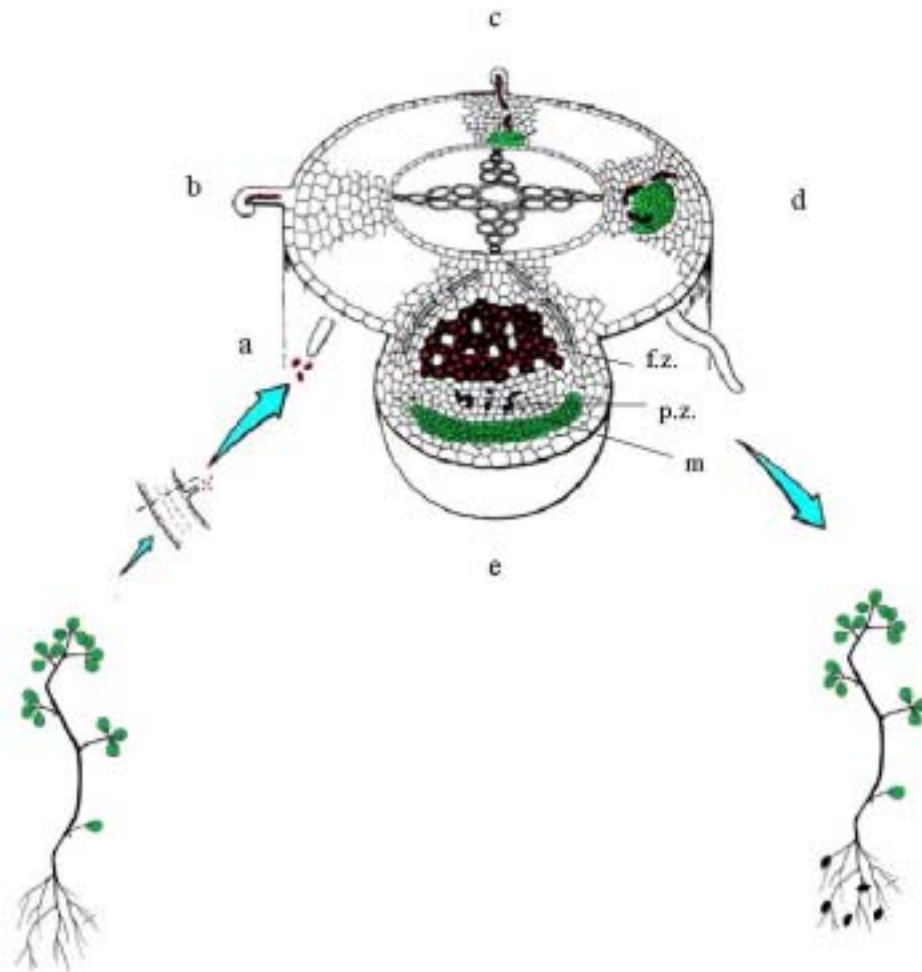


Figure 1.2 Different stages of indeterminate nodule development. **a:** attraction and attachment of the *Rhizobium* to the root hair tip, **b:** curling of root hairs and subsequent entrapment of bacteria followed by infection thread formation, **c:** penetration of the infection threads and simultaneous induction of cell divisions in the inner cortex, **d:** development of a nodule primordium and release of the bacteria into the plant cells, **e:** a major, nitrogen fixing nodule with a meristemic zone (m), a prefixation zone (p.z.) and nitrogen fixation zone (f.z.) (after Hirsch *et al.*, 1992).

For most legumes, such as *V. sativa* and *M. truncatula*, the tip of emerging root hairs is the primary target for infection by rhizobia, probably because the thinner and less cross-linked cell wall allows penetration by rhizobia more easily (Hirsch, 1992). The first visible morphological response of the host plant upon contact with rhizobia is root hair deformation (Had phenotype, for hair deformation). This deformation starts by swelling of the root hair tip (Has phenotype, for hair swelling) followed by reinitiation of polar growth from the swelling (Hab phenotype, for hair branching). Rhizobia attached to the root hair tip become entrapped by curling of the root hair (Hac phenotype, for hair curling). The Had phenotype can be induced by Nod factors alone, but root hair curling only appears in the presence of rhizobia showing that other factors, such as localized Nod factor production, are needed for a successful start of the symbiotic association (Heidstra *et al.*, 1994; Catoira *et al.*, 2000).

### 1.2.2 Infection and nodule formation

Within the pocket formed by the curled hair the infection starts with a local modification of the cell wall and invagination of the root hair plasma membrane. By localized deposition of plant membrane and cell wall-like material a tunnel (the infection thread) is formed within the root hair (Fig. 1.2 b and c). Most likely via tip growth, the infection thread grows towards the inner root cortex where simultaneously the nodule primordium will be formed. Formation of the infection thread requires local cell wall degradation that likely involves both bacterial and plant enzymes (Van Spronsen *et al.*, 1994). Within the infection thread the bacteria multiply (Brewin, 1998). By growth of the infection thread the bacteria penetrate the cortical cell layers, however the membrane of the infection thread prevents direct contact between the bacteria and the cytoplasm of the plant cells.

The infection thread is guided by the so called pre-infection threads (Van Brussel *et al.*, 1992). Cells in both inner and outer cortex enter the cell cycle but only those in the inner cortex complete the cell cycle (Yang *et al.*, 1994). In the outer cortical cells the cytoplasm collects to form conical structures aligned with the radial division walls of the inner cortex cells. These are the cytoplasmic bridges that constitute the pre-infection threads (Van Brussel *et al.*, 1992).

When the infection threads reach the nodule primordium, they ramify intracellularly before the bacteria are released by a process that resembles endocytosis. In *Medicago*, pea or vetch for example, each bacterium released from the infection thread is surrounded by a plant derived membrane, the peribacteroid membrane (PBM). The PBM enveloped bacteria, called symbiosomes, continue to divide within the host cells before they differentiate into bacteroids and start to fix nitrogen (Roth and Stacey, 1989; Newcomb, 1981).

### 1.2.3 Architecture of a mature indeterminate nodule

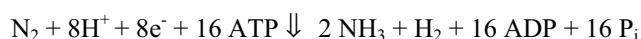
Upon infection with bacteria the nodule primordium will develop in a nodule. At the apex a persistent meristem is formed which constantly adds new, uninfected cells to the nodule. The indeterminate nodules are characterized by specific peripheral tissues and a central zone that contain successive stages of development; the youngest cells close to the meristem (apex) and the oldest near the basis. Due to the presence of cells at different stages of development the central tissue can be divided in several zones (Vasse *et al.*, 1990; Timmers *et al.*, 2000; Figs. 1.1 and 1.2 e):

Zone I contains the apical meristem which divides and ensures continued nodule growth over several weeks. Zone II is the prefixation zone where the cells become infected. In this zone, cell division is stopped, but most cells become polyploid and enlarge. The interzone II-III is characterized by the accumulation of starch in amyloplasts. In this zone the bacterial nitrogen-fixation starts (Yang *et al.*, 1991). Zone III is the nitrogen fixation zone. This region can easily be recognized by its pink color, due to the presence of leghemoglobin (see below). In older nodules, a senescent zone (zone IV) will be formed where both symbiotic partners degenerate. In old alfalfa nodules, after six weeks of infection, an additional zone was recently described where senesced plant cells become reinvaded by bacterial release from the remaining infection threads. In this zone no nitrogen fixation takes place and the bacteria become saprophytic (Timmers *et al.*, 2000).

The central tissue of the nodule is surrounded by several cell layers forming different tissues. The nodule parenchyma is located adjacent to the central tissue. The nodule vascular bundles are embedded in this tissue and link the nodules to the root vascular tissue. The parenchyma is surrounded by the nodule endodermis and the outer cortex, which form a continuum with the root endodermis and cortex respectively (Fig. 1.1).

### 1.2.4 Nodule functioning

Photosynthetic products are transported via the phloem to the nodules. In the nodules, these compounds provide the energy source for the bacteria as well as the C-backbone for several N-containing compounds produced during nitrogen fixation and assimilation. Symbiotic nitrogen fixation takes place in the differentiated rhizobia, the bacteroids. The bacterial nitrogenase complex catalyzes the following reaction:



The ammonium is transported through the peribacteroid membrane to the plant. Once the ammonium is in the plant cell cytoplasm, it is assimilated by the glutamine synthetase (GS)/glutamate synthase (GOGAT) cycle to form glutamine. Then the glutamine is converted to the N-containing amides (in temperate legumes) or ureides (in tropical legumes). In these forms the fixed nitrogen is transported via the xylem to the other organs of the plant.

The nitrogenase complex is inactivated by O<sub>2</sub>. However some O<sub>2</sub> is essential for bacteroid respiration. The high affinity oxygen carrier leghemoglobin ensures that O<sub>2</sub> is transported to the bacteroids in the microaerobic nodule environment at accurately controlled rates (Appleby, 1984). In addition, the peripheral nodule cell layers, particularly the parenchyma containing small and densely packed cells, may function as a diffusion barrier for gasses, regulating the supply of oxygen to the nitrogen-fixing tissues (Brewin, 1998).

### 1.2.5 Recognition between the symbiotic partners

In general, rhizobia have the ability to interact with only a small number of plants (Mylona *et al.*, 1995). For example, *Rhizobium leguminosarum* bv *viciae* nodulates pea and vetch, whereas *Bradyrhizobium japonicum* nodulates soybean. Which factors determine this host specificity?

Roots secrete special flavonoids or isoflavonoids that, in concert with the constitutively produced bacterial NodD protein, induce transcription of the other bacterial *nod* genes whose products are involved in synthesis or secretion of Nod factors (Lerouge *et al.*, 1990; Schultze *et al.*, 1992). In general, a NodD protein from a certain *Rhizobium* species is most active in the presence of flavonoids from its host plant. Therefore, the

interaction between these flavonoids and NodD may constitute a first level of host range control (Spaink *et al.*, 1987).

Secondly, the host range of a certain *Rhizobium* strain is determined by the structure of the secreted Nod factors. The basic Nod factor consists of 3 to 6  $\eta$ -1,4-linked N-acetylglucosamine (GlcNAc) residues that carry an N-acyl chain at the non-reducing residue (Fig. 1.3 A). The proteins NodA, NodB, NodC catalyze the synthesis of the backbone of the Nod factors. The N-acetylglucosaminyl-transferase NodC controls the length of the backbone and catalyses the synthesis of the chitin oligomer by coupling UDP-GlcNAc molecules, the deacetylase NodB removes the N-acetyl group that is present at the non-reducing end of the GlcNAc-residue and NodA acylates the deacetylated backbone. Two other common Nod proteins are NodI and NodJ that are involved in the secretion of the Nod factors. In addition to the common Nod factors, each *Rhizobium* species possesses specific nodulation genes that determine further modifications at the reducing and non-reducing end of the basic Nod factor (Table 1.3). For example, *S. meliloti* produces four different Nod factors (Fig. 1.3 B). The presence of the sulfate group is decisive for activity of the Nod factors on its host-plant *Medicago* and strongly reduces the activity on the non-host plant *Vicia*. On the contrary non-sulfated Nod factors produced by *nodH* mutants of *S. meliloti* are active on *Vicia* but not on *M. sativa* (Roche *et al.*, 1991).

Table 1.3 Role of Nod proteins in Nod factor biosynthesis (adapted from Schultze and Kondorosi, 1996).

Nod protein	Function
NodA,B,C	Synthesis of Nod 'backbone'
NodE,F	Synthesis of unsaturated fatty acids and determinant of acyl chain length
NodS	N-methyl transferase
NodL	O-acetylation
NodU	O-carbomoylation
NodP,Q,H	O-sulfation
NodX	O-acetylation
NodZ	Fucosylation

Besides Nod factors, the rhizobia excrete other molecules that may be involved in the determination of the host specificity. For example, upon introduction into *A. caulinodans* OR571 of the *nodO* gene from *Rhizobium leguminosarum* bv. *viciae*, encoding a secreted protein that can form  $\text{Ca}^{2+}$  transporting ion-channels *in vitro*, this strain could form nitrogen-fixing nodules on *Leucaena leucocephala*, whereas wild-type *A. caulinodans* could only trigger the formation of few Fix<sup>+</sup> nodules on this non-host plant (Vlassak *et al.*, 1998).

Also lectins may be involved in the recognition of the symbiotic partners. Transgenic white clover roots carrying the pea lectin gene *Psl* were nodulated by the heterologous strain *R. leguminosarum* bv. *viciae*, whereas no nodules formed on clover plants that lacked the pea gene (Diaz *et al.*, 1989). Van Rhijn *et al.* (1998) showed that nodule-like outgrowths developed on transgenic *L. corniculatus* expressing the soybean lectin *Le1* when inoculated with the heterologous strain *B. japonicum* which nodulates soybean and not *Lotus* spp.

Besides symbionts also pathogens like *Agrobacterium* are attracted by the root exudates and the plant raises defense reactions against invading microorganisms, including rhizobia. Surface polysaccharides (exopolysaccharides, lipopolysaccharides and capsular polysaccharides) of rhizobia seem to have a role in the suppression of these defense reactions and thus may allow the plant to differentiate between pathogens and symbionts (Schultze and Kondorosi, 1998).

## 1.3 Nod factor perception and signal transduction

Nod factors are extremely powerful mitogens, acting at nanomolar concentrations. In order to trigger nodule organogenesis the plant must perceive them and transduce the perceived signal to the tissues that will be involved in the establishment of the symbiosis. This section gives a brief overview of the current knowledge of Nod factor perception and signal transduction.

The high specificity and the low concentration at which Nod factors are active suggest the existence of high affinity receptors at the host plant root. In several studies it was found that different plant responses have different requirements with regard to the Nod factor structure and the existence of a perception mechanism using two receptors was proposed: one having stringent requirements for the non-reducing end substitutions of Nod factors and another having less stringent requirements for these substitutions (Ardourel *et al.*, 1994; Geurts *et al.*, 1997; Felle *et al.*, 1996). This is illustrated clearly by the study of the alkalization of the cytosol of alfalfa root hairs upon addition of Nod factors (Felle *et al.*, 1996). A first dose of Nod factor inhibited further alkalization of the root hair cytoplasm when a second dose of the same molecule was added. However, addition of a sulfated Nod factor followed by the addition of a non-sulfated Nod factor, or vice versa, resulted in an increased alkalization of the cytoplasm, suggesting that sulfated and non-sulfated Nod factors are perceived by different receptors (Felle *et al.*, 1996).

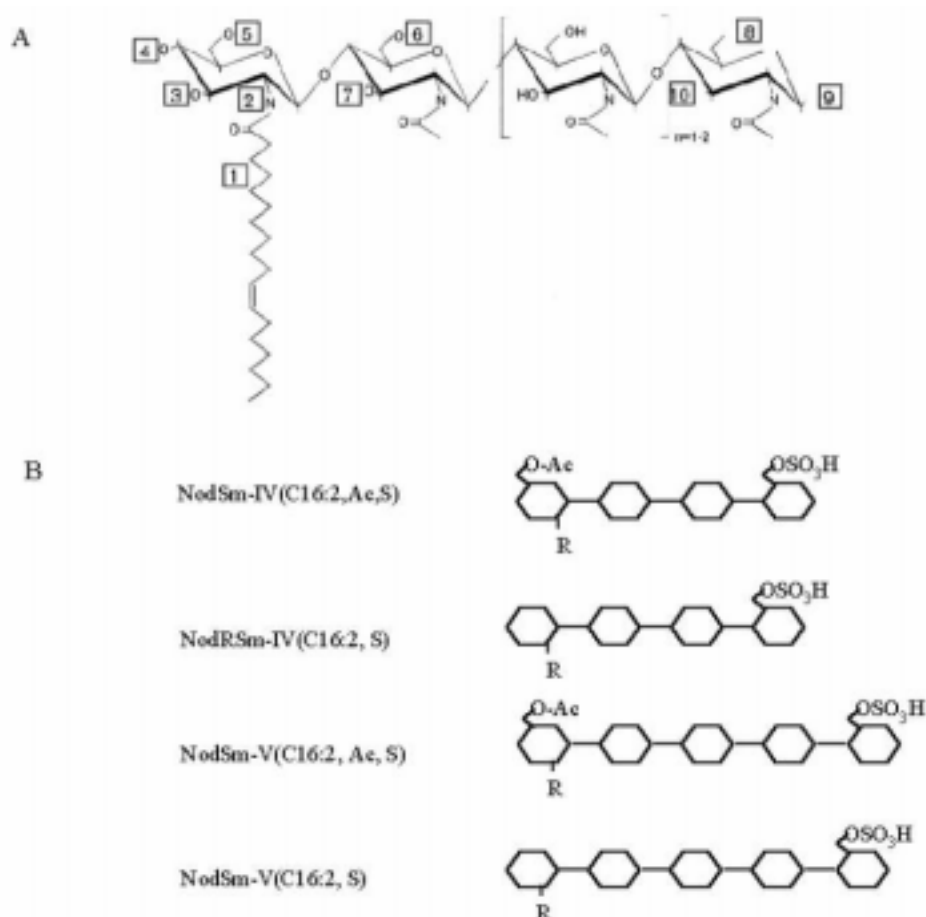


Figure 1.3 **A**: Generic Nod factor structure with potential sites for modifications. Known Modifications by various rhizobial isolates include (1) unsaturations in the fatty acyl chain and variation in chain length. Substitution of the following groups at the positions indicated: (2) methyl; (3) carbamoyl; (4) carbamoyl; (5) carbamoyl and acetyl; (6) acetyl; (7) fucosyl; (8) sulphuryl, acetyl, fucosyl and other fucosyl derivatives; (9) mannosyl and glycerol; (10) arabinosyl. Figure taken from Downie and Walker (1999). **B**: Schematic representation of the four major Nod factors produced by *S. meliloti* (from Schultze and Kondorosi, 1996). The nomenclature of the Nod factors is based on the name of the bacteria (Sm for *S. meliloti*), the number of  $\eta$ -1,4-linked N-acetyl glucosamines (IV or V), and the substitutions on the two terminal glucosamine-residues (C16:2 = length and saturation of the acyl chain; Ac: O-acetyl group at non-reducing end; S: sulphate group at the reducing end). R=N-acyl chain and  $\square$  =N-acetylglucosamine.

It is not clear whether Nod factors are perceived at the cell surface or whether they are internalized before perception. Philip-Hollingsworth *et al.* (1997) showed that Nod factors are taken up by the plant cell, but Heidstra *et al.* (1994) demonstrated that Nod factors were localized at the cell membrane and a more recent study from Goedhart *et al.* (2000) showed that fluorescently labeled Nod factors accumulated in the cell wall. The authors hypothesized that the Nod factors were concentrated and immobilized to reach a certain threshold before further perception would take place. Concentrating the Nod factors at the root surface was also thought to be important according to Van Rhijn *et al.* (1998), who proposed that plant lectins would play a role in the attachment and aggregation of rhizobia at the infection site, thus increasing the local Nod factor concentration. In a microtargetting experiment the fatty acid chain of the *R. leguminisarum* bv. *viciae* Nod factor was not necessary to elicit cell divisions in *Vicia* roots (Schlaman *et al.*, 1997) indicating that the acyl chain may be required for the uptake of the Nod factors while the other part of the Nod factor might act intracellularly. However, Nod factor uptake is at present not convincingly suggested by localization studies.

So far, 'the' Nod factor receptor has not yet been found, but several candidates exist. The Nod factor binding site 1 (NFBS1, Bono *et al.*, 1995) was found in a high density fraction of *M. truncatula* roots containing, among others, cell wall material. It has a high affinity for Nod factors ( $K_d=86$  nM), but a poor selectivity for the recognition of both the reducing and non-reducing end. Nod factor binding site 2 (NFBS2, Niebel *et al.*, 1997; Gressent *et al.*, 1999), which was found in a plasma-membrane enriched fraction of a *Medicago varia* cell suspension culture has a higher affinity for Nod factors than NFBS1 ( $K_d=4$  nM) and recognizes the O-acetyl group and the hydrophobicity of the acyl chain, but does not discriminate between sulfated and non-sulfated factors.

In addition, two apyrases with Nod factor binding activity have been found: The first is a lectin nucleotide phosphohydrolase (LNP) localized on the root hair surface of *Dolichos biflorus* (Etzler *et al.*, 1999; Kalsi and Etzler, 2000). Chitin could bind to Db-LNP, but the best competition for this binding was achieved using purified Nod factors from *B. japonicum* and *Rhizobium* sp. NGR234. GS52, an orthologue of Db-LNP, was located in the plasmamembrane of soybean roots (Day *et al.*, 2000). Antisera raised against either protein inhibited the ability of rhizobia to nodulate their hosts, *D. biflorus* and soybean, indicating the involvement of this lectin in the nodulation process.

The earliest observed response of alfalfa root hairs to the addition of Nod factors, measured with ion-selective microelectrodes, is an influx of  $\text{Ca}^{2+}$  in the cytoplasm (Felle *et al.*, 1998). The  $\text{Ca}^{2+}$  influx is rapidly followed by a  $\text{Cl}^-$  efflux, and then efflux of  $\text{K}^+$  to balance the charge. The  $\text{Cl}^-/\text{K}^+$  movements and the associated movement of  $\text{H}^+$  ions seem to account for a membrane depolarization and an alkalization both inside as well as outside the root hair (Felle *et al.*, 1998). This depolarization was elicited at concentrations in the range required for root hair deformation, and depolarization was not observed in cells of tomato, a non-host species (Ehrhardt *et al.*, 1992). Moreover, it was found that the ability of Nod factors to elicit plasma membrane depolarization correlated well with their bio-activity (Felle *et al.*, 1995).

Ten minutes after Nod factor application, alfalfa and pea root hairs responded with a regular  $\text{Ca}^{2+}$  spiking in the cytoplasm around the nucleus (Ehrhardt *et al.*, 1996; Walker *et al.*, 2000). The structural requirements for the calcium spiking response paralleled the requirements for other nodulation-specific plant responses and the non-nodulating alfalfa mutant MN-NN1008 did not spike. An overview of the temporal order of plant responses to Nod factors is shown in Figure 1.4. It was proposed that  $\text{Ca}^{2+}$  may play a role as second messenger during the establishment of the symbiotic interaction (Felle *et al.*, 2000).

The conclusion that heterotrimeric G-proteins are involved in Nod factor signaling is based on studies in which mastoporan mimicked Nod factor activity (Pingret *et al.*, 1998; Den Hartog *et al.*, 2001). However, it should be kept in mind that so far no gene for the  $\alpha$ -subunit of the heterotrimeric G proteins is found in plants (e.g. in the completely sequenced *Arabidopsis* genome; AGI, 2000).

## 1.4 Physiological control of nodulation and the role of plant growth regulators

The formation of nodules is a complex and costly process for the plant (Caetano-Anolles and Gresshoff, 1991). For example, an adequate supply of energy by photosynthesis is required for efficient nodulation. Moreover, plants will only invest in the establishment of symbiosis when combined nitrogen in the soil, such as nitrate and ammonium, is limiting (Schultze and Kondorosi, 1998). Once a critical number of nodules has been formed, the plant suppresses the formation of new nodules by a process called autoregulation (Caetano-Anolles and Gresshoff, 1991). This autoregulation seems to be independent of the presence of *Rhizobium* or the nutritional status of the plant (Kosslak and Bohlool, 1984; Caetano-Anolles and Bauer, 1988; Caetano-Anolles, 1990; Takats, 1990). For example, spontaneous formed nodules (NAR) in alfalfa (Caetano-Anolles, 1990) or empty nodules induced by application of invasion-defective *S. meliloti* (Takats, 1990) do suppress the formation of new nodules. Since these nodules are free of bacteria, they cannot influence the nutritional status of the plant.

Under normal conditions, only a few cells in the emerging root hair zone behind the root apical meristem opposite the protoxylem poles gain a transient competence for re-entering the cell cycle at the start of the nodule developmental program (Foucher and Kondorosi, 2000). Libbenga *et al.* (1973) suggested that in pea longitudinal gradients of auxin and cytokinin as well as a transversal gradient of a stele factor, later identified as uridine (Smit *et al.*, 1995) determine the competence of root cortical cells to nodule formation.

Contrary to uridine, in pea ethylene seems to have a negative effect on the formation of nodules opposite to phloem poles as shown by Heidstra *et al.* (1997a) by the use of ethylene antagonists. Moreover, *in situ* hybridization showed that the enzyme 1-aminocyclopropane-1-carboxylate (ACC) oxydase, which catalyzes the last step in ethylene synthesis, is localized opposite to the phloem poles, the position where normally no nodules will form (Heidstra *et al.*, 1997a). Besides controlling the position of the nodules, ethylene also seems to be involved in the nodule number control. Ethylene biosynthesis or perception inhibitors increase nodulation formation (reviewed in Schultze *et al.*, 1994). The study of nodulation mutants confirms the role of ethylene in nodule number control. For example, the ethylene hypersensitive Nod<sup>-</sup> pea mutant *sym5* can be rescued by ethylene perception inhibitors (Fearn and LaRue, 1991) and the ethylene insensitive *M. truncatula* mutant *sickle* shows an increase in persistent infections and number of nodules (Penmetsa and Cook, 1997).

Activation and completion of the cell cycle in plants is under the control of phytohormones, particularly auxin and cytokinin (Zhang *et al.*, 1996). Therefore, these hormones are most likely also involved in nodule initiation. Incubation of alfalfa, sweet clover and pea roots with auxin transport inhibitors results in formation of empty nodule-like structures and expression of early nodulin genes such as *ENOD12*, *ENOD8*, and *ENOD2*, although the morphology of these pseudonodules was different from Nod factor-induced nodules and did not require nitrogen starvation (reviewed in Schultze and Kondorosi, 1998). In addition, application of auxin



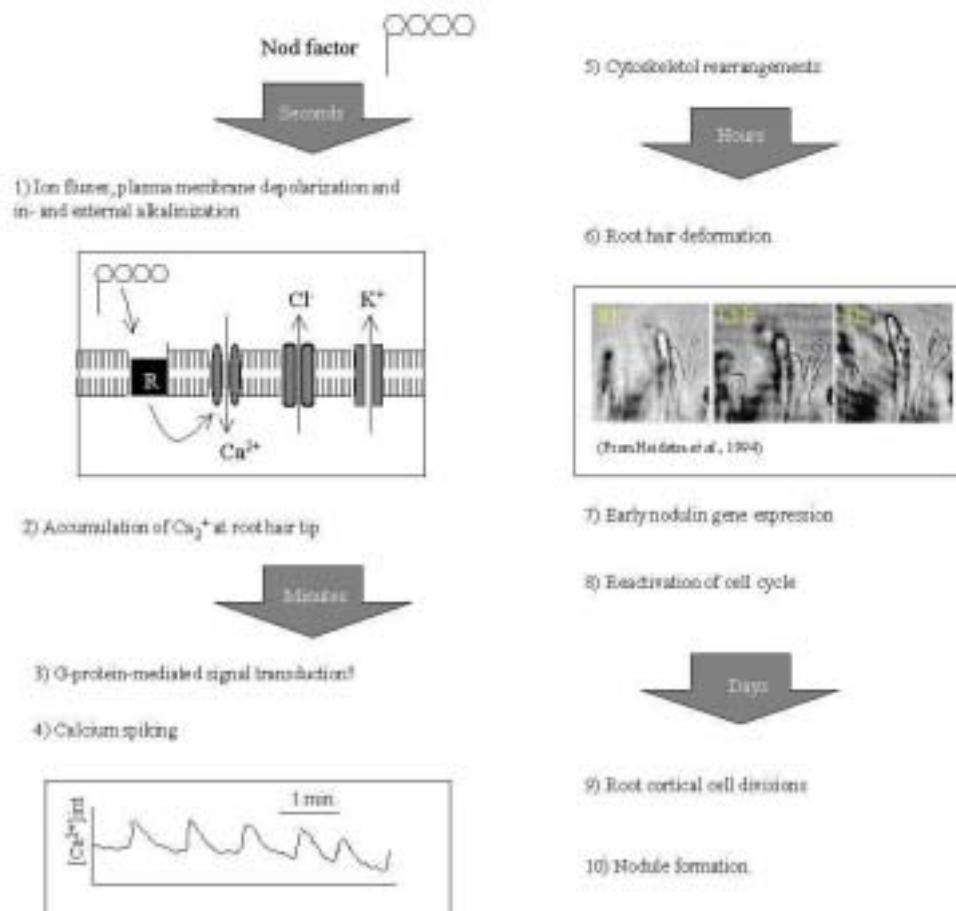


Figure 1.4 Temporal order of plant responses to Nod factors for a 'typical' legume. Figure based on Downie and Walker (1999) and Felle et al. (1998).

transport inhibitors or Nod factors to the root zone of white clover susceptible to rhizobial infection, led to an increased expression of an auxin-sensitive reporter gene fusion above and a decrease below the point of application (Mathesius *et al.*, 1998b). Also in *Vicia sativa* roots polar auxin transport was inhibited upon application of Nod factors and auxin transport inhibitors, respectively (Boot *et al.*, 1999). Flavonoids can act as auxin-transport inhibitors (Jacobs and Rubery, 1988) and it was shown by fluorescence microscopy that flavonoids accumulated in root cortical cells during nodule initiation at the site of rhizobial inoculation (Mathesius *et al.*, 1998a). Addition of certain flavonoids to white clover roots indeed resulted in similar expression patterns of the auxin-sensitive reporter gene as found for Nod factors or auxin transport-inhibitor (Mathesius *et al.*, 1998b). This led to the conclusion that the Nod factor induced cell divisions might be mediated by a perturbation of the auxin flow (Mathesius *et al.*, 1998a).

Nod factor synthesis defective *S. meliloti* mutants could partly be rescued by the introduction of an *Agrobacterium* gene encoding trans-zeatin which is a cytokinin (Cooper and Long, 1994). In alfalfa cytokinin can mimic several rhizobia-induced responses, such as inner cortical cell-divisions, amyloplast deposition and expression of nodulins like *ENOD12*, *ENOD40* and *ENOD2* (Bauer *et al.*, 1996; Fang and Hirsch, 1998).

Thus, gradients of cytokinins, auxins, ethylene and uridine probably determine a specific spatial position where cortical cells might be able to respond to bacterial Nod factors (Schultze and Kondorosi, 1998; factors involved in the positioning of nodules are summarized in Fig. 1.5)

## 1.5 Nodulins

During nodule development consecutive sets of genes become activated. Those, which showed no detectable expression in the roots but were induced during nodule development were thought to be nodule specific, and called nodulins (Van Kammen, 1984). However, development of more sensitive techniques as well as the detailed analysis of the expression patterns in other organs revealed that expression of many genes was not restricted to the nodulation process. For example, the early nodulin *ENOD40* is expressed in nodules but also in stems, leaves, and flowers of nodulating species (Crespi *et al.*, 1994). Moreover, *ENOD40*-like transcripts were



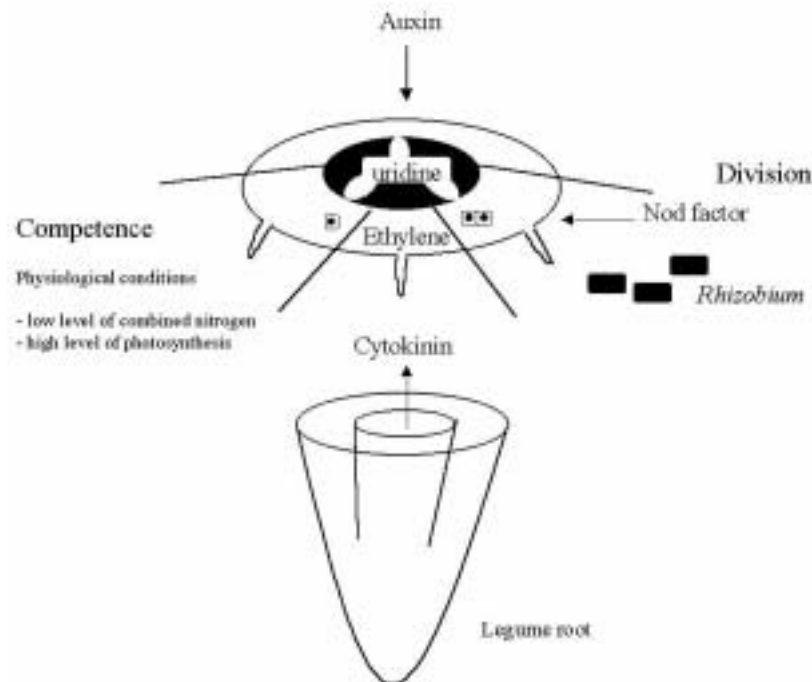


Figure 1.5 Legume root with cross section showing environmental and plant physiological factors required for nodulation and determination of spatial nodule position. Explanation in text (After Schultze and Kondorosi, 1998).

detected also in non-nodulating plants, for example in tobacco (Van de Sande *et al.*, 1996). This suggests that many nodulins were recruited for nodule organogenesis and functioning from genes with a similar function elsewhere in the plant (Muñoz *et al.*, 1996). The nodulin genes were classified according to the expression kinetics during nodule development (Govers *et al.*, 1985). Early nodulins (*ENODS*) are expressed before the beginning of  $N_2$  fixation. They are thought to participate in the early interactions between the bacteria and the plant, as well as being involved in nodule organogenesis. Late nodulins are expressed just before or during  $N_2$  fixation and are thought to be involved in the functioning and maintenance of the mature nodules (Govers *et al.*, 1985).

### 1.5.1 The function of nodulins

So far, the function of the early nodulin genes have been proposed mainly on the basis of their spatial and temporal expression patterns during symbiosis and their homology to genes with a known function. Although the function of most nodulins is not known, their expression patterns serve as molecular marker for different stages of nodule development. For example, the early nodulin *ENOD12* (Scheres *et al.*, 1990; Pichon *et al.*, 1992; Allison *et al.*, 1993) is expressed in root hairs, epidermal cells, and in cells where the infection thread is present or in cells preparing for the passage of these structures, and later in the nodule primordium and invasion zone. *ENOD12* encodes a putative prolin-rich cell wall protein (Kieliszewski and Lamport, 1994) and is thought to have a role in the infection process.

Another early nodulin being part of the cell wall is *ENOD2* (refs in Hirsch, 1997). This nodulin is expressed specifically in the nodule parenchyma and since the nodule parenchyma was shown to regulate the concentration of the free  $O_2$  in the nodule, it was hypothesized that *ENOD2* might contribute to the morphology of parenchyma cells creating the oxygen diffusion barrier (Van de Wiel *et al.*, 1990). However, Wycoff *et al.* (1998) found that changes in *ENOD2* gene expression or protein synthesis do not appear to affect the regulation of  $O_2$  diffusion.

*ENOD40* is expressed within 1-3 hours after addition of Nod factors in pericycle cells opposite protoxylem poles and in the nodule primordium (Kouchi and Hata, 1993; Yang *et al.*, 1993; Crespi *et al.*, 1994; Compaan *et al.*, 2001). Sequence comparison of *ENOD40* genes of different species showed that all *ENOD40* genes encode a short peptide and contain a conserved 3' non-coding region (Van de Sande *et al.*, 1996; Crespi *et al.*, 1994). It was proposed that the *ENOD40* gene product functions as a small peptide signaling molecule or as an RNA molecule involved in triggering a localized hormone imbalance followed by cell divisions in the root cortex (Van de Sande *et al.*, 1996; Crespi *et al.*, 1994; Fang and Hirsch, 1998).

Stable and transient overexpression of *ENOD40* in *M. truncatula* induced cell divisions in the inner cortex (Charon *et al.*, 1997). Moreover, overexpression of *ENOD40* accelerated nodulation due to extensive proliferation of cortical cells. In addition, *M. truncatula* lines co-suppressing *ENOD40* formed few nodules with

arrested meristems (Charon *et al.*, 1999). In addition to nodulation, *ENOD40* also seems to play a role in the symbiotic associations of legume roots with arbuscular mycorrhizal fungi (AM). For example, mycorrhizal colonization with the AM fungus *Glomus mosseae* was significantly stimulated by overexpression of *ENOD40* in transgenic *M. truncatula* plants (Stahelin *et al.*, 2001).

Late nodulins participate in various aspects of nodule functioning, such as nitrogen assimilation, nitrogen or carbon metabolism and oxygen transport (Muñoz *et al.*, 1996). Although the leghemoglobins also were found before the onset of nitrogen fixation (Heidstra *et al.*, 1997b; Györgyey *et al.*, 2000), they are the most abundant proteins in mature nodules. These proteins act as oxygen carriers providing the bacteroids with sufficient oxygen to generate energy for the functioning of the nitrogenase enzyme at microaerobic conditions. Examples of late nodulins involved in the carbon and nitrogen metabolism are sucrose synthase, uricase and glutamine synthase (Muñoz *et al.*, 1996).

### 1.5.2 The search for nodulins

Most nodulins were isolated by differential display techniques, comparing either proteomes or transcriptomes of root nodules to roots. Using nodule-specific antisera, nodulins were identified by comparing either proteins (Bisseling *et al.*, 1983) or *in vitro* translation products (Legocki and Verma, 1980) from nodules and roots of pea and soybean, respectively. Differences in gene expression between nodules and roots were revealed by two-dimensional gel electrophoresis of <sup>35</sup>S methionine-labeled *in vitro* translation products from nodules and roots (Govers *et al.*, 1985). The differentially expressed proteins were isolated and their microsequences served as base for the synthesis of degenerated primers used to obtain cDNA encoding the isolated protein. In this way for example *GmNOD53b*, encoding a symbiosome membrane protein of soybean was isolated (Winzer *et al.*, 1999). A disadvantage of comparing proteomes for the isolation of nodulins is, that only relatively abundant proteins can be identified.

Less abundant transcripts can be isolated by differential display and differential hybridizations of cDNAs. With the differential display technique differentially expressed mRNAs are detected in different RNA samples (e.g. from nodules and roots) by comparing the patterns of partial cDNAs generated after reverse transcription and PCR amplification with arbitrary primers. By using this technique several nodulin genes were identified in *Sesbania rostrata*, *Vicia sativa*, *L. japonicus*, and *M. truncatula* (Goormachtig *et al.*, 1995; Heidstra *et al.*, 1997b; Szczyglowski *et al.*, 1997; De Carvalho Niebel *et al.*, 1998).

The most frequently used method is the differential hybridization technique (Franssen *et al.*, 1987). In this case, radiolabeled cDNAs are synthesized from mRNAs isolated from nodules and roots and used as hybridization probe on nodule cDNA libraries displayed on filters. By using subtractive probes the sensitivity of the differential screening could be improved, although one may lose the nodule-stimulated transcripts (Kouchi and Hata, 1993; Gamas *et al.*, 1996).

In addition, the growing number of sequenced ESTs (e.g. Györgyey *et al.*, 2000) in combination with recent developments of microarray techniques allows reproducible differential hybridization of many genes. Moreover, by comparing the expression pattern of genes with unknown functions during different growth conditions to the expression pattern of genes with known functions, the unknown functions might be predicted (Hughes *et al.*, 2000).

## 1.6 Nodulation mutants

Four basic categories of symbiotic mutant phenotypes have been identified: (i) no nodulation or nodulation that is blocked at a very early stage, (Nod<sup>-</sup>); (ii) few nodules, Nod<sup>+/-</sup>; (iii) ineffective nodulation, (Nod<sup>+</sup>Fix<sup>-</sup>); and (iv) hypernodulation/nitrate tolerance (Nod<sup>++</sup>, Fix<sup>+</sup>). Further refinement can be added by characterizing more accurately at which step of nodule formation the interaction is blocked, like root hair swelling (has), root hair deformation (had), root hair curling (hac), nodule initiation (noi), nodule persistence (nop), nodule functioning (nof) etc....

The first symbiotic mutant phenotypes were found in pea and were due to natural variations. For example, the pea ecotype Afghanistan, containing the *sym2* locus, can be nodulated by Middle East strains of *R. leguminosarum* bv. *viciae* such as TOM, but not by the European strains (Firmin *et al.*, 1993; Geurts *et al.*, 1997). Later, mutations were induced chemically by the use of EMS or physically by the use of v-rays, resulting in a large collection of pea lines with mutant nodulation phenotypes (Table 1.4). However, molecular analysis in pea is difficult, due to inefficient transformation, regeneration and to its large genome size. Recently two legume plants, *L. japonicus* and *M. truncatula* were accepted by the scientific community as model legumes for determinate and indeterminate nodules, respectively (Schauser *et al.*, 1998; Cook *et al.*, 1999). Both legume species are diploid and autogamous, have small genome sizes and can be efficiently transformed and regenerated. During the last couple of years also for *L. japonicus* and *M. truncatula* numerous lines with natural or induced mutant phenotypes became available (Table 1.5 and 1.6).

Complementation analysis of mutant lines (found in natural populations or induced by EMS or irradiation) allowed the identification of a number of genes involved in nodulation, and most loci were placed on genetic linkage maps. In principle, this allows map-based cloning of the mutated genes. However, map-based cloning is a very time consuming process, and by this method so far only the receptor-like kinases responsible for the Nod<sup>-</sup> phenotypes in the *M. sativa* mutant line MN-1008 (Endre *et al.*, 2002) and in the *L. japonicus* mutant lines 282-287, cac41.5 and EMS61 (Stracke *et al.*, 2002) were isolated.

Problems to isolate a mutated gene can be overcome by the use of T-DNA or transposons as a mutagen (see below). The isolation of tagged genes is relatively easy, because the known nucleotide sequence of the tag can serve as a starting point for the identification of flanking sequences. In legumes, T-DNA insertional mutagenesis has been initiated in *L. japonicus* (Martirani *et al.*, 1999) and *M. truncatula* (this work). Transposon tagging programs have been started in *M. truncatula* (d'Erfurth, unpub.) and *L. japonicus* (Schauser *et al.*, 1999). The first successful identification of a mutated plant nodulation gene was obtained in *L. japonicus* using the maize transposon *Ac* (Schauser *et al.*, 1999). The *nin* (for 'nodule inception') mutant does not form nodules and the tagged gene encodes a putative transcription factor. When the *Ac* transposon excised from the affected gene, the mutant plants reverted to the wild-type phenotype. This showed that the mutant phenotype indeed had been caused by insertion of the transposon. Based upon the sequence of *LjNIN*, the *PsNIN* gene was isolated and it was found that pea *sym35* mutants, showing the same phenotypic characteristics as the *LjNIN* mutant, were mutated in the pea homologue of *LjNIN* (Madsen *et al.*, 2001, EuroConference Golm). An updated overview of the *P. sativum*, *M. truncatula* and *L. japonicus* mutant lines is given in Tables 1.4, 1.5 and 1.6.

Symbiotic mutants serve as tool to identify genes involved in the nodulation process. The phenotypic and genetic study of these mutants allows to dissect the steps involved in nodule development and to determine the order in which they take place.

Besides symbiotic interactions with rhizobia, leguminous plants also have the ability to interact symbiotically with the arbuscular mycorrhizal fungi belonging to the order of *Glomales*. Molecular and genetic studies show that rhizobial and mycorrhizal infection involves common mechanisms, and a large proportion of nodulation-resistant mutants were also resistant to AM fungi (reviewed in Albrecht *et al.*, 1999). Examples of the use of nodulation mutants in combination with the study of phenotypic and genetic characteristics to dissect events during nodulation and mycorrhization are given by the work of Catoira *et al.* (2000) and Wais *et al.* (2000) in *M. truncatula* and Walker *et al.* (2000) in *P. sativum*. The behavior of different nodulation mutant lines upon addition of Nod factors or mycorrhizal fungi with respect to responses like calcium spiking and root hair deformation were studied. For example none of the *M. truncatula* plants mutated in *dmi1*, *dmi2* or *dmi3* and none of the pea plants mutated in *sym10*, *sym8*, *sym19* or *sym30* showed root hair deformation. However, in contrast to the other mutants, plants mutated in *dmi3* or *sym30* did show calcium spiking (Fig. 1.6). This led to the hypothesis that DMI1, DMI2, SYM10, SYM8 and SYM19 are involved in steps preceding calcium spiking and that DMI3 and SYM30 must act downstream of calcium spiking. By studying other pea and *M. truncatula* mutants the authors have proposed models for the potential order of functioning of *M. truncatula* and pea nodulation genes in nodulation and mycorrhizal signaling (Fig. 1.6). In addition, these experiments showed that *M. truncatula* indeed can serve as a model plant for the dissection of events during nodulation in pea, since in *M. truncatula* mutant lines are found that showed phenotypes comparable to those obtained in mutant lines of pea.

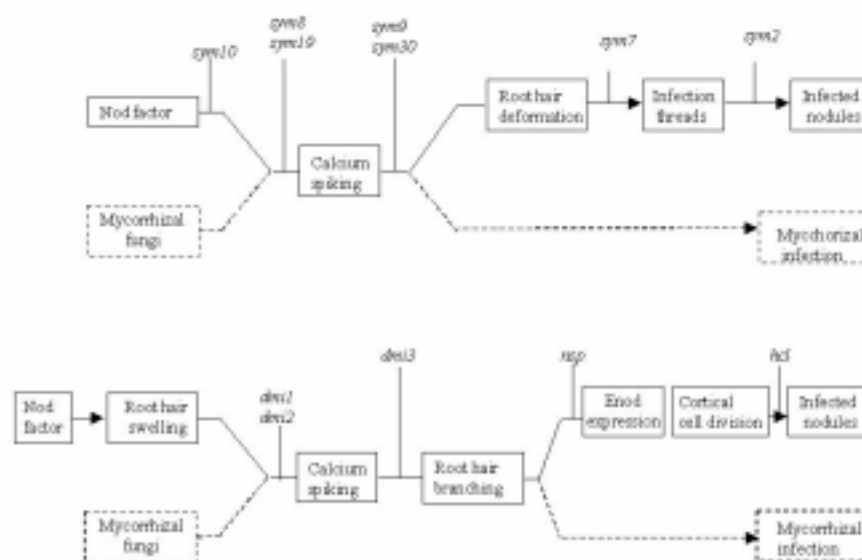


Figure 1.6 Models based upon the study of *Pisum* (a) and *Medicago* (b) nodulation and mycorrhizal mutants showing the potential order of involvement of the corresponding *Pisum* and *Medicago* wild-type loci in nodulation and mycorrhizal signalling (after Walker *et al.*, 2000 and Wais *et al.*, 2000).

Table 1.4 Overview of some *P. sativum* lines with mutant nodulation phenotypes.

Line	Gene (a)	Origin (b)	Phenotype	Remarks	Reference
Nod3 <sup>6</sup>		EMS	Nod <sup>+/+</sup> Fix <sup>+</sup>	Nitrate tolerant. Phenotype root controlled small shoots and compact roots.	Postma <i>et al.</i> , 1988
FN1 (mutated Nod3 line)		EMS	Nod <sup>+/+</sup> Fix <sup>-</sup>		Postma <i>et al.</i> , 1990
K5 <sup>6</sup>		EMS	Nod <sup>+/+</sup> Hac <sup>+</sup> Inf <sup>+</sup>	Phenotype root controlled.	Postma <i>et al.</i> , 1988
K24 <sup>6</sup>		EMS	Nod <sup>+/+</sup> Hac <sup>+</sup> Inf <sup>-</sup>	Phenotype controlled by shoot and root.	Postma <i>et al.</i> , 1988
cv. Iran	<i>SYM1</i>	-		Nod <sup>-</sup> when inoculated with European <i>R. leguminosarum</i> bv. <i>viciae</i> strains at 18°C	Kozik <i>et al.</i> , 1995
				Nod <sup>+</sup> at 26°C.	
cv. Afghanistan	<i>SYM2</i>	-		Nod <sup>-</sup> when inoculated with European <i>R. leguminosarum</i> bv. <i>viciae</i> strains at 18°C	Kozik <i>et al.</i> , 1995
				Nod <sup>+</sup> at 26°C.	
	<i>SYM3</i>	-	Nod <sup>+</sup> Fix <sup>-</sup>		Ref. in Caetano-Anolles and Gresshoff, 1991
	<i>SYM4</i>	-		Nodulation resistance in cv. Iran and Afghanistan.	Ref. in Caetano-Anolles and Gresshoff, 1991
E2 <sup>1</sup>	<i>SYM5</i>	EMS		Nod <sup>+</sup> at 16°C, Nod <sup>+/+</sup> at 20°C. Ethylene inhibitors restore phenotype. Phenotype root controlled.	Weeden <i>et al.</i> , 1990; Fearn and LaRue, 1991; Markwei and LaRue, 1997
	<i>SYM6</i>	I	Nod <sup>+</sup> Fix <sup>-</sup>	Ineffective nodulation in cv. Afghanistan.	Ref. in Caetano-Anolles and Gresshoff, 1991
E69 <sup>1</sup>	<i>SYM7</i>	EMS	Nod <sup>-</sup> Had <sup>+</sup> Hac <sup>+</sup> Inf <sup>-</sup>		Walker <i>et al.</i> , 2000; Weeden <i>et al.</i> , 1990
R25 <sup>1</sup> , P2 <sup>2</sup>	<i>SYM8</i>	GAM EMS	Nod <sup>-</sup> Had <sup>+</sup> Hac <sup>+</sup> Inf <sup>-</sup> Myc <sup>-</sup>	No NF induced calcium spiking. <i>SYM8</i> required for <i>PsENOD5</i> and <i>PsENOD12A</i> induction during interaction with endomycorrhiza and rhizobia. Phenotype root controlled.	Walker <i>et al.</i> , 2000; Weeden <i>et al.</i> , 1990; Albrecht <i>et al.</i> , 1998
P54 <sup>2</sup>	<i>SYM9</i>	EMS, GAM	Nod <sup>-</sup> Has <sup>+</sup> Had <sup>+</sup> Hac <sup>+</sup> Inf <sup>-</sup> Myc <sup>-</sup>	Phenotype root controlled.	Walker <i>et al.</i> , 2000; Weeden <i>et al.</i> , 1990
R72 <sup>1</sup>					
P5/P56 <sup>2</sup>	<i>SYM10</i>	EMS	Nod <sup>-</sup> Had <sup>+</sup> Hac <sup>+</sup> Inf <sup>-</sup> Myc <sup>+</sup>	No NF induced calcium spiking.	Walker <i>et al.</i> , 2000; Weeden <i>et al.</i> , 1990;
N15 <sup>1</sup>		N		Phenotype root controlled.	Duc <i>et al.</i> , 1989; Duc and Messenger, 1989
N24 <sup>1</sup>	<i>SYM11</i>	N	Nod <sup>-</sup>		Weeden <i>et al.</i> , 1990
	<i>SYM12</i>	I	Nod <sup>-</sup>		Ref. in Caetano-Anolles and Gresshoff, 1991
E135F <sup>1</sup> P58 <sup>2</sup>	<i>SYM13</i>	EMS	Nod <sup>+</sup> Fix <sup>-</sup>	Affected in symbiosome development. Early senescence. Phenotype root controlled.	Borisov <i>et al.</i> , 1997; Sagan <i>et al.</i> , 1993;
E135N <sup>1</sup>	<i>SYM14</i>	EMS	Nod <sup>-</sup>		Weeden <i>et al.</i> , 1990
E151 <sup>1</sup>	<i>SYM15</i>	EMS	Nod <sup>+/+</sup>	Short lateral roots.	Weeden <i>et al.</i> , 1990
R50 <sup>1</sup>	<i>SYM16</i>	GAM	Nod <sup>+/+</sup>	Mutant phenotype restored by ethylene inhibitors.	Weeden <i>et al.</i> , 1990
				Phenotype root controlled. Short lateral roots.	Guinel and Sloetjes, 2000
R82 <sup>1</sup>	<i>SYM17</i>	GAM	Nod <sup>+/+</sup>	Stubby roots and shoots.	Weeden <i>et al.</i> , 1990
E54 <sup>1</sup>	<i>SYM18</i>	EMS	Nod <sup>+/+</sup>	Strain specific phenotype.	Weeden <i>et al.</i> , 1990
P4/P6/P55 <sup>2</sup>	<i>SYM19</i>	EMS	Nod <sup>-</sup> Has <sup>+</sup> Had <sup>+</sup> Hac <sup>+</sup> Inf <sup>-</sup> Myc <sup>-</sup>	No NF induced <i>ENOD12A</i> induction. Phenotype root controlled. P4 and P5 mutated in <i>NORK</i> gene.	Walker <i>et al.</i> , 2000; Sagan <i>et al.</i> , 1993
NEU5/NMU1 <sup>1</sup>					Schneider <i>et al.</i> , 1999; Weeden <i>et al.</i> , 1990;
					Endre <i>et al.</i> , 2002.
E132 <sup>1</sup>	<i>SYM21</i>	EMS	Nod <sup>+/+</sup> Fix <sup>+</sup>	Phenotype is shoot controlled. Short, highly branched roots. Less nodule meristems. Ethylene inhibitors can partly restore nodule number, but not root phenotype.	Markwei and LaRue, 1997

Line	Gene (a)	Origin (b)	Phenotype	Remarks	Reference
P59 <sup>2</sup>	<i>SYM23</i>	EMS	Nod <sup>+/+</sup> Fix <sup>-</sup>	Phenotype root controlled.	Sagan <i>et al.</i> , 1994
P60 <sup>2</sup>	<i>SYM24</i>	EMS	Nod <sup>+/+</sup> Fix <sup>-</sup>	Phenotype root controlled.	Sagan <i>et al.</i> , 1994
P61 <sup>2</sup>	<i>SYM25</i>	EMS	Nod <sup>+</sup> Fix <sup>-</sup>	Phenotype root controlled.	Sagan <i>et al.</i> , 1994
P62 <sup>2</sup>	<i>SYM26</i>	EMS	Nod <sup>+/+</sup> Fix <sup>-</sup>	Phenotype root controlled.	Sagan <i>et al.</i> , 1994
P12 <sup>2</sup>	<i>SYM27</i>	EMS	Nod <sup>+</sup> Fix <sup>-</sup>	Phenotype root controlled.	Sagan <i>et al.</i> , 1994
P77 <sup>2</sup>	<i>SYM28</i>	EMS	Nod <sup>++</sup> Fix <sup>+</sup>	Nitrate-tolerant mutant. Phenotype shoot controlled.	Sagan and Duc, 1996
P87-P94 <sup>2</sup>	<i>SYM29</i>	EMS	Nod <sup>++</sup> Fix <sup>+</sup>	Nitrate-tolerant mutant Phenotype shoot controlled.	Sagan and Duc, 1996
P53 <sup>2</sup>	<i>SYM30</i>	EMS	Nod <sup>+</sup> Has <sup>+</sup> Had <sup>+</sup> Hac <sup>-</sup> Inf <sup>-</sup>	Phenotype root controlled.	Walker <i>et al.</i> , 2000; Sagan <i>et al.</i> , 1994
P54 <sup>2</sup>			Myc <sup>-</sup>		Duc and Messenger, 1989
RisFixA <sup>3</sup>		EMS	Nod <sup>+</sup> Fix <sup>-</sup>	Early nodule senescence.	Morzhina <i>et al.</i> , 2000
Sprint-2Fix <sup>-5</sup>	<i>SYM31</i>	I	Nod <sup>+</sup> Fix <sup>-</sup>	Affected in symbiosome development.	Borisov <i>et al.</i> , 1997
RisFixL <sup>3</sup>	<i>SYM32</i>	EMS	Nod <sup>+</sup> Fix <sup>-</sup>	Bacteroid differentiation blocked.	Morzhina <i>et al.</i> , 2000
FixRisO <sup>3</sup>					
RisFixM <sup>3</sup> , RisFixT <sup>3</sup>	<i>SYM26</i>	EMS	Nod <sup>++</sup> Fix <sup>-</sup>	Premature degradation of symbiotic structures.	Morzhina <i>et al.</i> , 2000
RisFixN <sup>3</sup>		EMS	Nod <sup>++</sup> Fix <sup>-</sup>	Premature degradation of symbiotic structures.	Morzhina <i>et al.</i> , 2000
RisFixQ <sup>3</sup>	<i>SYM27</i>	EMS	Nod <sup>++</sup> Fix <sup>-</sup>	Premature degradation of symbiotic structures.	Morzhina <i>et al.</i> , 2000
RisFixK <sup>3</sup>		EMS	Nod <sup>++</sup> Fix <sup>-</sup>	Premature degradation of symbiotic structures.	Morzhina <i>et al.</i> , 2000
RisFixV <sup>3</sup>		EMS	Nod <sup>+</sup> Fix <sup>-</sup>	Premature degradation of symbiotic structures.	Morzhina <i>et al.</i> , 2000
SGEFix—1 <sup>4</sup>	<i>SYM40</i>	EMS	Nod <sup>++</sup> Fix <sup>-</sup> Nop <sup>-</sup>	Premature degradation of invaded cells. Phenotype root controlled.	Tsyganov <i>et al.</i> , 1998
P57 <sup>2</sup>		EMS	Nod <sup>+/+</sup> Fix <sup>-</sup> Has <sup>+</sup> Hac <sup>+</sup> Inf <sup>-</sup> Myc <sup>+</sup>	Phenotype root controlled.	Sagan <i>et al.</i> , 1994; Duc and Messenger, 1989
DK24 <sup>3</sup>		I	Nod <sup>+/+</sup> Fix <sup>-</sup> Myc <sup>-</sup>		Engvild <i>et al.</i> , 1987
P63 <sup>2</sup>		EMS	Nod <sup>+</sup> Fix <sup>-</sup>	Phenotype root controlled.	Sagan <i>et al.</i> , 1994
E107 <sup>1</sup>	<i>BRZ</i>	EMS	Nod <sup>-</sup>	Nodule phenotype partly restored by ethylene inhibitors.	Guinel and LaRue, 1992;
					Weeden <i>et al.</i> , 1990
SGEFix—2 <sup>4</sup>	<i>SYM33</i>	EMS	Nod <sup>+/+</sup> Fix <sup>-</sup>	Premature degradation of invaded cells. Phenotype root controlled.	Tsyganov <i>et al.</i> , 1998

a = Except *SYM19*, the other *SYM* genes were not yet cloned. In uppercase numbers the cultivar is indicated: 1 = Sparkle, 2 = Frisson, 3 = Finale, 4 = SGE, 5 = Sprint, 6 = Rondo. b: dash = natural occurrence, EMS = ethylmethane sulfonate induced, N = nitrosoethylurea or nitrosomethylurea induced, GAM = v-ray induced., I = induced mutation.

Table 1.5 Overview of some *Medicago truncatula* cv. Jemalong lines with mutant nodulation phenotypes.

Line	Gene (a)	Origin (b)	Mutant phenotype	remarks	reference
B129, C71 (domi), Y6	<i>DMI1</i>	EMS	Has <sup>+</sup> Hac <sup>+</sup> Hab <sup>+</sup> Inf <sup>+</sup> Nod <sup>+</sup> Myc <sup>-</sup>	No <i>MtENOD11</i> and <i>MtENOD40</i> expression, reduced <i>MtRIP</i> expression, no cell divisions, no calcium spiking.	Penmetsa and Cook, 1997
TR25, TR26, P1	<i>DMI2</i>	GAM	Has <sup>+</sup> Hac <sup>+</sup> Hab <sup>+</sup> Inf <sup>+</sup> Nod <sup>+</sup> Myc <sup>-</sup>	No <i>MtENOD11</i> and <i>MtENOD40</i> expression, reduced <i>MtRIP</i> expression no cell divisions, no calcium spiking. Mutated in <i>NORK</i> gene.	Catoira <i>et al.</i> , 2000; Wais <i>et al.</i> , 2000
TRV25	<i>DMI3</i>	EMS	Has <sup>+</sup> Hac <sup>+</sup> Hab <sup>+</sup> Inf <sup>+</sup> Nod <sup>+</sup> Myc <sup>-</sup>	No <i>MtENOD11</i> and <i>MtENOD40</i> expression, reduced <i>MtRIP</i> expression no cell divisions, normal calcium spiking.	Catoira <i>et al.</i> , 2000; Sagan <i>et al.</i> , 1995, Wais <i>et al.</i> , 2000, Endre <i>et al.</i> , 2002.
B85, C54	<i>NSP</i>	GAM	Has <sup>+</sup> Hac <sup>+</sup> Hab <sup>+</sup> Inf <sup>+</sup> Nod <sup>+</sup> Myc <sup>+</sup>	No <i>MtENOD40</i> expression, reduced <i>MtENOD11</i> and <i>MtRIP</i> expression.	Catoira <i>et al.</i> , 2000
B56	<i>HCL</i>	EMS	Has <sup>+</sup> Hac <sup>+</sup> Hab <sup>+</sup> Inf <sup>+</sup> Nod <sup>+</sup> Myc <sup>+</sup>	Normal <i>MtENOD40</i> , <i>MtENOD11</i> and <i>MtRIP</i> expression, normal calcium spiking.	Wais <i>et al.</i> , 2000
TE7	<i>MtSYM1</i>	EMS	Nod <sup>+</sup> Fix <sup>+</sup> and Fix <sup>-</sup>		Benaben <i>et al.</i> , 1995
TR3, TR9, TR13, TR36, TR62, TR69, TR74, TRV15		GAM	Nod <sup>+</sup> Fix <sup>+</sup> Myc <sup>+</sup>		Sagan <i>et al.</i> , 1995
TR36		GAM	Nod <sup>+</sup> Myc <sup>+</sup>		Sagan <i>et al.</i> , 1995
TR34, TR79, TR89, RV9		GAM	Nod <sup>+/+</sup> Myc <sup>+</sup>		Sagan <i>et al.</i> , 1995
TR122, TRV3, TRV8		GAM	Nod <sup>++</sup> Myc	Nitrate tolerant.	Sagan <i>et al.</i> , 1995
Sickle	<i>SKL1</i>	EMS	Nod <sup>++</sup> Fix <sup>+</sup>	Ethylene-insensitive.	Penmetsa and Cook, 1997
Jemalong6	<i>MtSYM6</i>	-	Has <sup>+</sup> Hac <sup>+</sup> Hab <sup>+</sup> Inf <sup>+</sup>	Fix <sup>-</sup> when invaded by <i>S. meliloti</i> A145.	Tirichine <i>et al.</i> , 2000
			Nod <sup>+</sup> Fix <sup>+</sup> Nop <sup>+</sup>	Phenotype root controlled.	

a = Except *DMI2*, the other mutated genes were not yet cloned. b: dash = natural occurrence, EMS = ethylmethane sulfonate induced, GAM = v-ray induced.

Table 1.6 Overview of some *Lotus japonicus* lines with a mutant nodulation phenotype.

Line	gene (a)	Source (b)	phenotype	remarks	reference
alb1	<i>ALB1</i>	EMS	Nod <sup>+</sup> Fix <sup>+</sup> and Fix <sup>-</sup>	Fix <sup>+</sup> nodule formation delayed.	Imaizumi-Anraku <i>et al.</i> , 1997
fen1	<i>FEN1</i>	EMS	Nod <sup>+</sup> Fix <sup>+/-</sup>	Nodules fix seven times less N <sub>2</sub> than wt. Pre-mature senescence.	Imaizumi-Anraku <i>et al.</i> , 1997
282-118, 282-665	<i>LjSYM1</i>	'T-DNA'	Nod <sup>+</sup> Myc <sup>+</sup>		Schauser <i>et al.</i> , 1998
282-287, 282-288	<i>LjSYM2</i>	'T-DNA'	Nod <sup>+</sup> Myc <sup>-</sup>	282-287 mutated in <i>SYMRK</i> gene.	Schauser <i>et al.</i> , 1998, Stracke <i>et al.</i> , 2002.
5371-22, 2557-1	<i>LjSYM3</i>	'T-DNA'	Nod <sup>+</sup> Myc <sup>-</sup>		Schauser <i>et al.</i> , 1998
282-227; EMS1749	<i>LjSYM4</i>	'T-DNA'	Has <sup>+</sup> Hac <sup>-</sup> Inf <sup>+</sup> Nod <sup>-</sup>		Schauser <i>et al.</i> , 1998
		EMS	Myc <sup>+/-</sup>		Bonfante <i>et al.</i> , 2000
282-894	<i>LjSYM5</i>	'T-DNA'	Nod <sup>+</sup> Myc <sup>+</sup>		Schauser <i>et al.</i> , 1998
10512.9, 1962-124	<i>LjSYM6</i>	'T-DNA'	Fix <sup>-</sup>		Schauser <i>et al.</i> , 1998
312-133	<i>LjSYM7</i>	'T-DNA'	Fix <sup>+</sup> Myc <sup>+</sup>		Schauser <i>et al.</i> , 1998
5361-33	<i>LjSYM8</i>	'T-DNA'	Fix <sup>-</sup>	T-DNA insertion linked to mutant phenotype.	Schauser <i>et al.</i> , 1998
282-XI	<i>LjSYM9</i>	'T-DNA'	Fix <sup>-</sup>		Schauser <i>et al.</i> , 1998
2572-77.1	<i>LjSYM10</i>	'T-DNA'	Fix <sup>-</sup>		Schauser <i>et al.</i> , 1998
KU3-13	<i>LjSYM11</i>	'T-DNA'	Nod <sup>+</sup> Fix <sup>-</sup>		Schauser <i>et al.</i> , 1998
282-483	<i>LjSYM12</i>	'T-DNA'	Nod <sup>+</sup> Fix <sup>-</sup>		Schauser <i>et al.</i> , 1998
282-936	<i>LjSYM13</i>	'T-DNA'	Nod <sup>+</sup> Fix <sup>-</sup> Myc <sup>+</sup>	T-DNA insertion linked to mutant phenotype.	Schauser <i>et al.</i> , 1998
PASac123	<i>LjSYM14</i>	'T-DNA'	Nod <sup>+</sup> Fix <sup>-</sup> Myc <sup>+</sup>		Schauser <i>et al.</i> , 1998
282-1078	<i>LjSYM15</i>	'T-DNA'	Nod <sup>+/-</sup> Fix <sup>+</sup>		Schauser <i>et al.</i> , 1998
1962-125	<i>HAR1</i>	'T-DNA'	Nod <sup>++</sup> Fix <sup>+</sup> Myc <sup>+</sup>	Nitrate tolerant, inhibition of root and shoot growth. Cytokinin in presence of AVG stimulates root growth.	Schauser <i>et al.</i> , 1998; Szczygłowski <i>et al.</i> , 1998
LjEMS102	( <i>LjSYM16</i> )	EMS		presence of AVG stimulates root growth.	Wopereis <i>et al.</i> , 2000
LjEMS34,61	<i>LjSYM21</i>	EMS	Nod <sup>+</sup> Myc <sup>-</sup>	Mutated in <i>SYMRK</i> gene	Szczygłowski <i>et al.</i> , 1998, Stracke <i>et al.</i> , 2002
LjEMS46, 40	<i>LjSYM22</i>	EMS	Nod <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS70, 167	<i>LjSYM23</i>	EMS	Nod <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS76	<i>LjSYM24</i>	EMS	Nod <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS223	<i>LjSYM25</i>	EMS	Nod <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS247	<i>LjSYM26</i>	EMS	Nod <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS81	<i>LjSYM27</i>	EMS	Nod <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS236	<i>LjSYM28</i>	EMS	Nod <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS126	<i>LjSYM30</i>	EMS	Nod <sup>+</sup> Fix <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS75	<i>LjSYM31</i>	EMS	Nod <sup>+</sup> Fix <sup>-</sup> and Fix <sup>+</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS208	<i>LjSYM32</i>	EMS	Nod <sup>+</sup> Fix <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS79	<i>LjSYM33</i>	EMS	Nod <sup>+/-</sup> Fix <sup>+/-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS40	<i>HAR1, LjSYM22</i>	EMS	Nod <sup>-</sup>	Doublemutant, see also <i>LjSYM16</i> .	Szczygłowski <i>et al.</i> , 1998; Wopereis <i>et al.</i> , 2000
LjEMS45, 88, 217		EMS	Nod <sup>+</sup> Fix <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
LjEMS237		EMS	Nod <sup>-</sup>		Szczygłowski <i>et al.</i> , 1998
Nin	<i>NIN</i>	Ac	Has <sup>+</sup> Had <sup>+</sup> Hac <sup>+</sup> Inf <sup>+</sup>	First mutation coupled to an identified gene, Ac excision restores <i>NIN</i> function.	Schauser <i>et al.</i> , 1999

a = Except *NIN* and *LjSYM2*, the other mutated genes were not yet cloned. b: EMS = ethylmethane sulfonate induced, 'T-DNA' = found in T-DNA tagged plant line, Ac = induced by transposon tagging using the *Ac* transposon.

## 1.7 T-DNA tagging

As mentioned before, T-DNA tagging may represent a powerful tool for the identification and functional analysis of genes. In this section, tagging methods and the advantages of T-DNA tagging compared to other methods will be discussed.

### 1.7.1 No T-DNA tagging without *Agrobacterium*

The plant pathogenic bacterium *Agrobacterium tumefaciens* has the capacity to transfer part of its DNA (the T-DNA) into the nuclear genome of plant cells. The T-DNA transfer to the plant nucleus depends on the expression of the bacterial *vir* genes that delimit the DNA region transferred to the nucleus by recognizing specific sequences called T-DNA right and left borders (RB and LB), respectively. Gene fusion studies indicate that the insertion preferably takes place in the transcribed regions or in their vicinity (Koncz *et al.*, 1989; Mathur *et al.*, 1998; Lindsey *et al.*, 1993). For example, although tobacco, *Arabidopsis* and potato have different genome sizes and different amounts of repetitive DNA, upon transformation with a promoter-less *GUS* gene, about 20-30 % of the transformants of each species exhibited *GUS* activation (Lindsey *et al.*, 1993). How the integration of the T-DNA takes place is not known, but it is believed to arise from illegitimate recombinations between the T-DNA and the genome in which the plant DNA repair machinery would be involved (Tinland, 1996). In most cases only one T-DNA is integrated, but also integrations of more than one copy (either at one locus or at different loci) and even more complex integration patterns were found (Tinland, 1996; Nacry *et al.*, 1998; Laufs *et al.*, 1999).

In nature, the T-DNA encodes proteins involved in the biosynthesis of plant growth factors and bacterial nutrients (the opines). But since these functions are not necessary for the transfer, they can be replaced by any DNA to be introduced into the genome (Tinland *et al.*, 1996). For example resistance markers allow selection for transformed plants or tissues, and reporter genes can be used to monitor the expression patterns of genes.

Besides being used as a carrier of DNA into the plant genome, the T-DNA can also be used as a mutagen or as a vehicle for transposable elements, which in their turn may act as mutagens (reviewed in Springer, 2000). Different approaches for insertional mutagenesis were tried out in the laboratory; however, the thesis work was concentrated on T-DNA tagging in *M. truncatula*.

#### T-DNA tagging strategies

Several different types of gene identification systems using T-DNA insertion have been developed. The T-DNAs used carried activation tags, enhancer traps or promoter traps (Fig. 1.7). Since all these methods aim at 'tagging' a gene by the use of T-DNA, they will commonly be referred to as T-DNA tags.

Activation tags and enhancer traps both use the property of enhancers to activate gene expression. For activation tagging (Fig. 1.7 B), T-DNAs containing multiple enhancers orientated in tandem (from for example the 35S RNA promoter of cauliflower mosaic virus), cloned to one of the T-DNA borders, is used. After insertion in the genome, the enhancers may activate the expression of plant genes located in their vicinity. Because this will result in a dominant phenotype, selection for a specific phenotype can be carried out already in primary transformed, hemizygous individuals.

In enhancer traps (Fig. 1.7 C) a reporter gene is fused to a minimal promoter containing a TATA box and transcription start site that is unable to drive reporter gene expression alone. When this T-DNA inserts in the vicinity or within a plant gene, the expression of the reporter gene can be activated by the enhancer elements of this plant gene. However, by using enhancer traps, mutant phenotypes will only be visible in transformants homozygous for the T-DNA insertion within a non-redundant plant gene. Enhancers can activate gene expression at considerable distances. On the one hand this may increase the number of genes that can be identified, but on the other hand this may complicate the isolation of the tagged gene.

Promoter traps (Figs. 1.7 D and E) contain a promoter less reporter gene and insertion of the T-DNA may lead to transcriptional and translational gene fusions between the reporter gene and the tagged plant gene. To obtain translational gene fusions with promoter traps, the reporter gene should insert in the correct orientation and correct reading frame within an exon of the tagged gene. In addition, promoter trap constructs containing one or more splice acceptor sequences preceding the reporter gene (Fig. 1.7 E), allow transcriptional and translational gene fusions if insertion occurs in an intron. The presence of the reporter gene allows screening for tagged genes in the hemizygous, primary transformed mutants.

T-DNA mediated gene tagging methods have been used successfully in several species, such as *Arabidopsis* (e.g. Bouchez *et al.*, 1993; Weigel *et al.*, 2000), tobacco (e.g. Foster *et al.*, 1999), rice (Chin *et al.*, 1999), *L. japonicus* (Schauser *et al.*, 1999; Webb *et al.*, 2000), and tomato (e.g. Takken *et al.*, 1998; Meissner *et al.*, 2000).



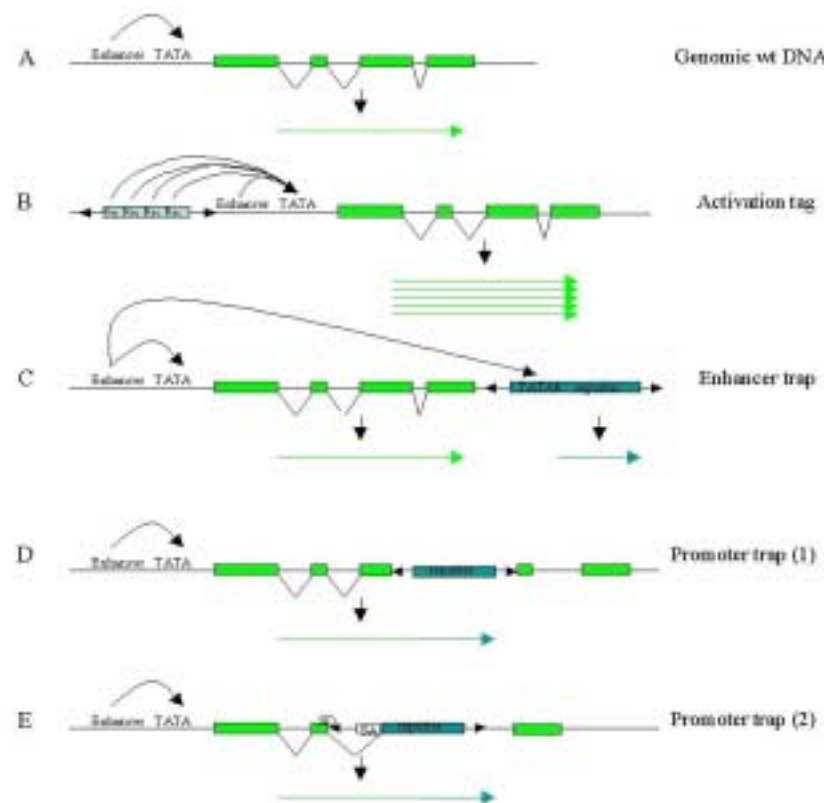


Figure 1.7 Strategies of T-DNA tagging. **A:** Example of the organization and expression of a wild-type chromosomal gene with exons (boxes) and introns (lines). **B:** An activation tag. The gene expression is activated by the four enhancers. **C:** An enhancer trap. The minimal promoter of the reporter gene (TATA) is activated by a chromosomal enhancer element, resulting in expression of the reporter gene. **D:** A promoter trap (1). The promoter-less reporter gene can be expressed when inserted in frame in an exon. **E:** A promoter trap (2). The promoter-less reporter gene contains splice acceptor (SA) sequences. Expression of the reporter gene occurs upon its insertion into an intron. Splicing from the chromosomal splice donor site (SD) to the SA sequence results in a transcriptional fusion. Arrows in each panel represent the transcripts that are produced as a consequence of the T-DNA insertion.

### 1.7.2 Advantages of the use of T-DNA tags for gene identification in comparison to other methods

Except for activation tagging, genes tagged by enhancer traps, or promoter traps, can be identified based on reporter gene expression and therefore mutant phenotypes are not required. This allows the identification of genes that are difficult to identify mutants obtained by for example classical mutagenesis methods (EMS, v-ray). For example classical mutagenesis of functional redundant genes does not lead to a mutant phenotype, since the loss of function would be compensated by another gene. Also mutations leading to subtle phenotypes could be overlooked in screens of classical mutants. However, the expression of the reporter gene used for gene tagging makes also the identification of these genes possible. Another class of genes that cannot easily be identified using classical mutants, are genes that are essential for the viability of the plant. Although plants homozygous for these mutations would not survive, the expression of a reporter gene makes these essential genes visible in heterozygous, viable plants. Furthermore, even if a T-DNA insertion does not cause a mutant phenotype, the expression of a reporter gene by translational or transcriptional gene fusion can serve as a marker to identify particular cells or tissues during, for example, nodule development.

Transgenic lines obtained after insertion of promoter or gene traps do not always show expression of the reporter gene since the reporter gene may not be inserted in the right orientation or reading frame. However, also the genes tagged in such lines may be identified if the plants homozygous for the mutation show a mutant phenotype. This implies that even those genes that are expressed at a too low level to be identified by techniques like differential display or differential hybridization techniques can be identified by using a promoter or gene tag. The tag inserts randomly in the genome, giving all genes equal chances to be identified and even if a gene fusion between a plant gene and the reporter gene shows an expression level that is too low to be detected, the mutation may cause a mutant phenotype in plants homozygous for the gene tag, thus making gene identification possible.

Probably the most important advantage of the tagging methods described here in comparison to other techniques to identify genes, concerns the isolation of the tagged gene. Once a gene has been identified (either by the expression of the reporter gene or by a mutant phenotype linked to the insertion of the T-DNA) the known DNA sequence of the tag can serve as a starting point to isolate the genes responsible for the reporter gene expression, the mutant phenotype or both.

## **1.8 Aim of the thesis**

The aim of this thesis work was to create and analyze a T-DNA tagged *M. truncatula* population in order to assess the possibilities of T-DNA tagging for the isolation of plant genes involved in nodulation from this model plant.

## 2 T-DNA tagging in the model legume *Medicago truncatula* allows efficient gene discovery

*This chapter is a reproduction of an article in Mol. Breeding 10:203-215 (2002)*

**Marije Scholte, Isabelle d'Erfurth, Sonia Ripa, Samuel Mondy, Viviane Jean, Patricia Durand, Colette Breda, Hanh Trinh, Ignacio-Rodriguez-Llorente, Eva Kondorosi, Michael Schultze, Adam Kondorosi, Pascal Ratet**

The annual legume *Medicago truncatula* has been proposed as a model plant to study various aspects of legume biology including rhizobial and mycorrhizal symbiosis because it is well suited for the genetic analysis of these processes (Barker *et al.*, 1990). To facilitate the characterization of *M. truncatula* genes participating in various developmental processes we have initiated an insertion mutagenesis program in this plant using three different T-DNAs as tags. To investigate which type of vector is the most suitable for mutagenesis we compared the behavior of these T-DNAs. One T-DNA vector was a derivative of pBin19 and plant selection was based on kanamycin resistance. The two other vectors (Bouchez *et al.*, 1993; Tissier *et al.*, 1999) carried T-DNA conferring Basta resistance in the transgenic plants. For each T-DNA type, we determined the copy number in the transgenic lines, the structure of the T-DNA loci and the sequences of the integration sites. The T-DNA derived from pBin19 generated complex T-DNA insertion patterns. The two others generally gave single copy T-DNA inserts that could result in gene fusions for the pGKB5 T-DNA. Analysis of the T-DNA borders revealed that several *M. truncatula* genes were tagged in these transgenic lines and *in vivo gus* fusions were also obtained. These results demonstrate that T-DNA tagging can efficiently be used in *M. truncatula* for gene discovery.

**Abbreviations:** EST: expressed sequence tags, *gus*: *uidA* gene coding for the  $\eta$ -glucuronidase, T-DNA: transferred DNA.

### 2.1 Introduction

Legume plants provide an important source of protein for human and animal nutrition as well as combined nitrogenous compounds for the biosphere (Tissier *et al.*, 1999). Legume growth can partly be independent of the reduced nitrogen present in the soil due to nitrogen-fixing symbiosis with bacteria of the genus *Rhizobium*, *Sinorhizobium* and *Azorhizobium* leading to the formation of a new organ, the root nodule. In root nodules, bacteria fix atmospheric nitrogen and export it to the plant, which in turn provides carbohydrates necessary for this energy-consuming process (Schultze and Kondorosi, 1998). Recognition between the two partners takes place at several steps during the establishment of the symbiosis, controlled by molecular signal exchanges (Schultze and Kondorosi, 1998). In *Rhizobium*, genetic studies have allowed the identification of most of the genes involved in the symbiotic interaction. In contrast, genetic studies on nodule development are rather difficult in a species such as *Medicago sativa* because of its tetraploid and allogamous nature. Recently, two leguminous plants (*Medicago truncatula* and *Lotus japonicus*) were chosen as model plants by the scientific community (Cook *et al.*, 1997) to study nitrogen-fixing symbiosis as well as mycorrhiza. Symbiotic mutants have already been obtained in both plants (Catoira *et al.*, 2001; Penmetsa and Cook, 1997; Schauser *et al.*, 1999), however, so far only one mutant gene has been identified (Schauser *et al.*, 1999).

*M. truncatula* is an annual relative of alfalfa (*M. sativa*) with mediterranean origin and represented by numerous ecotypes (Bonnin *et al.*, 1996). Considerable genetic variability of these ecotypes was observed. For example, various symbiotic combinations with different *Sinorhizobium meliloti* strains were described (Tirichine *et al.*, 2000). The genome of this microsymbiont has been sequenced (Galibert *et al.*, 2001) and numerous symbiotic mutants of *S. meliloti* are available. The small, diploid genome of *M. truncatula* (Blondon *et al.*, 1994) and the ease of transformation (Trinh *et al.*, 1998) make this plant a suitable model for studies on various aspects of legume molecular genetics and genomics (Cook *et al.*, 1999).

Insertion mutagenesis is a powerful tool for gene discovery, allowing rapid characterization of mutated genes. In plants, either endogenous or heterologous transposable elements or the *Agrobacterium* T-DNA are applied for gene tagging. In maize and snapdragon, efficient transposon mutagenesis has been developed using endogenous transposable elements (*Ac/Ds*; *En/Spm*; *Mu*; *Tam*; (Sundaresan, 1996; Walbot, 1999)). In the model plant *Arabidopsis* and in tomato, the maize *En/Spm* and *Ac/Ds* transposable elements were used successfully to construct mutant collections (Meissner *et al.*, 2000; Parinov *et al.*, 1999; Tissier *et al.*, 1999; Wisman *et al.*, 1998). In *L. japonicus* the first tagged symbiotic mutant was described by using *Ac/Ds* as mutagen (Schauser *et al.*, 1999). This is the only legume mutant obtained by insertion mutagenesis described today.

The capacity of *agrobacteria* to transfer the T-DNA into the nuclear genome of plant cells (Tinland, 1996) has been used not only to introduce genes into plants but also to inactivate plant genes. Transfer of the T-DNA to the plant nucleus requires the expression of the bacterial *vir* genes that delimit the extent of the DNA sequence transferred to the nucleus by recognizing specific sequences, the right and left borders of the T-DNA (RB and LB). Integration of T-DNA in the plant genome probably depends on the DNA repair machinery of the plant. In general, one copy of the T-DNA is inserted randomly in the plant genome, and gene fusion studies indicated that these insertions occur preferably in transcribed regions or in their vicinity (Mathur *et al.*, 1998). Thus, when large-scale transformation is available, T-DNA mutagenesis is a convenient technique for gene discovery. By using the *Agrobacterium* infiltration/transformation method (Bechtold *et al.*, 1998), T-DNA-insertion mutant collections for *Arabidopsis* have been established (Springer, 2000).

Previously we have set up a very efficient transformation/regeneration protocol for *M. truncatula* (Trinh *et al.*, 1998) that allows us to initiate an insertion mutagenesis program in this plant. We report here the tagging of *M. truncatula* by using three T-DNA vectors. The T-DNA copy number, the structure of the tagged loci and their sequences were determined. This study revealed that a pBin19 derived T-DNA was inserted in several copies, but the use of the other T-DNA constructs resulted mostly in single copy T-DNA inserts which are more suitable for tagging experiments. In addition, several genes were shown to be tagged or found in the vicinity of the T-DNA, indicating that T-DNA tagging is efficient in *M. truncatula*. Finally, *in vivo* gene fusions with the *gus* reporter gene were also obtained. Thus, this study opens the way for large scale insertion mutagenesis in this model legume.

## 2.2 Materials and methods

### 2.2.1 Plant material and plant growth conditions

*M. truncatula* (Gaertn.) line R108-1 (c3) is described in Trinh *et al.* (Trinh *et al.*, 1998) and called R108 throughout the text. Plants were grown in the greenhouse in vermiculite and watered alternately by deionized water and a nutritive solution (Soluplant, N/P/K: 18/6/26, Duclos International, Lunel Viel, France) at 60% relative humidity under a 16-h photoperiod, and 25 °C and 15 °C for day and night, respectively.

For seed sterilization, seeds were scarified with sand paper, sterilized for 30 min in a Bayrochlor solution (7g/l, Bayrochlor; BAYROL GMBH, Germany), rinsed 4 times in sterile water and germinated in the dark at room temperature on wet Whatmann 3MM paper, in petri dishes.

Plant nodulation was performed in aeroponic conditions as described previously (Coronado *et al.*, 1995) or in perlite (see below) in growth chambers under the same growing conditions as described above for the greenhouse. Plants were inoculated with *S. meliloti* Rm41 (Kondorosi *et al.*, 1984) and scored for nodule numbers two weeks after inoculation. For experiments in perlite, approximately 40 sterilized seeds were germinated in 30 X 50 cm plastic trays and watered with sterile water until the seedlings appeared. Then they were watered with nutritive solution until the first unifoliate leaf was expanded and with nitrogen-limiting solution (Coronado *et al.*, 1995) for the rest of the experiment. Inoculation was done when the first trifoliate leaf was fully developed.

### 2.2.2 Plant transformation and selection

Plant transformation was done according to Trinh *et al.* (Trinh *et al.*, 1998). *In vitro* Basta selection was done by incorporating 3 mg/l glufosinate-ammonium (Hoechst Schering AgrEvo GmbH, Frankfurt/Main) in the SHMab medium. In some experiments, plants were further selected in the greenhouse by spraying plants twice with the glufosinate-ammonium solution at 120 mg/l.

### 2.2.3 *Agrobacterium* strain and T-DNA vectors

*Agrobacterium tumefaciens* EHA105 strain (Hood *et al.*, 1993) was used in all experiments. Plasmids were introduced in this strain by triparental mating as previously described (Ratet *et al.*, 1988). Three types of vectors were used for these experiments. The first series of vectors was constructed using plasmid pPR97 that is itself derived from the pBin19 vector (Bevan, 1984; Szabados *et al.*, 1995). For the second transformation experiment we used plasmids pSLJ8313 and pSLJ8337 (Tissier *et al.*, 1999). In the third experiment transgenic plants were generated using the pGKB5 vector (Bouchez *et al.*, 1993) which can generate transcriptional as well as translational *gus* gene fusions upon integration in a gene.

### 2.2.4 $\eta$ -glucuronidase (GUS) assays

Histological detection of Gus activity on all plant pieces or 100- $\mu$ m tissue sections of different tissues was done according to Trinh *et al.* (Trinh *et al.*, 1998).

### 2.2.5 Isolation of genomic DNA and Southern blot analysis

Plant genomic DNA was isolated as described by Dellaporta *et al.* (Dellaporta *et al.*, 1983). Southern blot analysis was performed as described by Church and Gilbert (Church and Gilbert, 1984). The *gus* probe consisted of a 702-bp PCR fragment (nucleotides 240 to 942 downstream of the *gus* translation initiation site; (Jefferson *et al.*, 1986)) amplified using the oligonucleotides 5'GGCCAGCGTATCGTGCTGCG3' and 5'GGTCGTGCACCATCAGCACG3'. The *nptII* probe was amplified with oligonucleotides 5'GAGGCTATTCGGCTATGACTG3' and 5'ATCGGGAGCGCGCATACCGTA3' resulting in a 699-bp PCR fragment located between nucleotides 51-750 downstream of the *nptII* translation initiation site (Beck *et al.*, 1982). The *bar* probe consisted of a 372-bp fragment (nucleotide 279 to 651 downstream of the *bar* translation initiation site) amplified from the *bar* gene using oligonucleotides 5'GCAGGAACCGCAGGAGTGG3' and 5'CCAGAAACCCACGTCATG3' (Thompson *et al.*, 1987). The *cytP450* probe consisted of a 614 bp PCR fragment amplified from the *cytP450* gene (nucleotide 252 to 866 downstream of the translation initiation site; (Omer *et al.*, 1990)) using oligonucleotides 5'TTCGAGGCCGTCCGGGAGAG3' and 5'GGCGAGGTAGCGGAGCAGTT3'.

### 2.2.6 Characterization of T-DNA borders

T-DNA borders were isolated using a modification of the PCR walk protocol described by Siebert *et al.* (Siebert *et al.*, 1995) and Devic *et al.* (Devic *et al.*, 1997).

The oligonucleotide adaptor consisted of oligonucleotide 5'CTAATACGACTCACTATAGGGCTCGAGCGGCCCGGGGAGGT3' annealed to the octamer oligonucleotide 5'ACCTCCCC3', 5'phosphorylated and 3' aminated. Oligonucleotides 5'GGATCCTAATACGACTCACTATAGGGC3' (Ap1) and 5'CTATAGGGCTCGAGCGGC3' (Ap2) were used in the first and second PCR reactions respectively, in combination with T-DNA specific oligonucleotides.

For the pBin19 derived vector (Bevan, 1984; Szabados *et al.*, 1995) the first and second oligonucleotides used for amplification of the right border are respectively 5'AAGCGGCCGAGAACTGCGTGCAA3' and 5'CTAATTGGATACCGAGGGGA3'. They correspond respectively to nucleotide positions 8898-8922 and 8997-9015 of the pBin19 sequence (Accession number U09365). For the amplification of the left border we used oligonucleotides 5'CTGGACTGGCATGAACTTCGGTGAAA3' and 5'AGCAGGGAGGCAAACAATGA3' corresponding respectively to nucleotides 2056-2081 1989-2108 of the *gus* gene sequence (Accession number A00196; (Jefferson *et al.*, 1986)).

For the pSLJ8313 and pSLJ8337 vectors the positions of the oligonucleotides are given relative to their position in the pSLJ8313 sequence (Ratet and Rippa, 1999). The first and second oligonucleotides used for the amplification of the right border are respectively 5'CGAACTTAACCCGTTCCGCTTTGTAGA3' (nucleotide position 1095-1121) and 5'CCGCCGACAGAGGTGTGA3' (nucleotide position 827-844) and for the left border 5'CTGGACTGGCATGAACTTCGGTGAAA3' (nucleotide positions 21892-21919) and 5'AGCAGGGAGGCAAACAATGA3' (nucleotide positions 21924-21944).

For the pGKB5 vector ((Bouchez *et al.*, 1993); Accession number AX063413) the first and second oligonucleotides used for amplification of the right border from the *gus* coding sequence are respectively 5'CCAGACTGAATGCCACAGGCCGTC3' and 5'TCACGGGTTGGGGTTTCTACAGGAC3'. For the left border we used oligonucleotides 5'CAACCCTCAACTGGAAACGGGCCGGA3' and 5'CGTGTGCCAGGTGCCACGGAATAGT3'.

10 $\mu$ g genomic DNA was digested over night by restriction enzymes *DraI* or *SspI* or *HincII* or *EcoRV* plus *PvuII* or *ScaI* plus *HpaI* in a final volume of 100 $\mu$ l. The five digestions were phenol/chloroform extracted

and ligated overnight at 16°C in a 20 µl final volume in the presence of the adaptor at a final concentration of 5 µl and 10 units of T4 ligase. The ligation was diluted 25 times in water and 2 µl were used for the first PCR reaction (94°C: 45s; 72°C: 8mn, 7 cycles; 94°C: 45s; 67 °C: 8mn, 25 cycles). This first PCR reaction was diluted 100 times in water and 1 µl was used for the second PCR reaction. In order to amplify fragments of large size (more than 1 kb), we used a mixture of the two enzymes EurobioTaq (Eurobio, France) and Elongase (Gibco BRL). For the first PCR reaction the reaction mix consisted of 0.3u Elongase with 0.1u Eurobiotaq and for the second PCR reaction it consisted of 1u Elongase with 0.25u Eurobiotaq. When the reaction yielded a single fragment of a size compatible with a T-DNA border amplification, this fragment was recloned in the pGEM-T vector (Promega) and sequenced.

### 2.2.7 Sequence analysis

The sequences of the tagged loci were compared to sequences of the data base using the blast program <http://www.blast.genome.ad.jp/SIT/Blast.html> or the Tigr *Medicago* gene index <http://www.tigr.org/tdb/mtgi>.

## 2.3 Results

### 2.3.1 pBin19 derivatives generate multiple T-DNA inserts in *M. truncatula*

Plants containing pBin19 derived T-DNA belong to a population of 100 transgenic lines originally generated for studying the promoter of the polygalacturonase *MsPG3* gene (Muñoz *et al.*, 1998). The vector pPR97 (Szabados *et al.*, 1995) used allows the analysis of promoter-*gus* fusions in transgenic plants. In addition to this promoter study (I.R. Llorente and P. Ratet, unpublished), the progeny of this population was analyzed for the structure of the T-DNA inserts. For this, genomic DNA was prepared from individual T1 plants and analyzed by Southern blot experiment. This analysis indicated that the majority of these lines contained multiple T-DNA borders and thus multiple T-DNA inserts (Fig. 2.1). These borders did not segregate independently but rather as clusters (Fig. 2.1), suggesting that the corresponding T-DNAs were probably inserted in the plant genome as tandem or inverted copies (concatemers). Thus, transformation of *M. truncatula* with the pPR97 T-DNA resulted in complex T-DNA integration patterns.

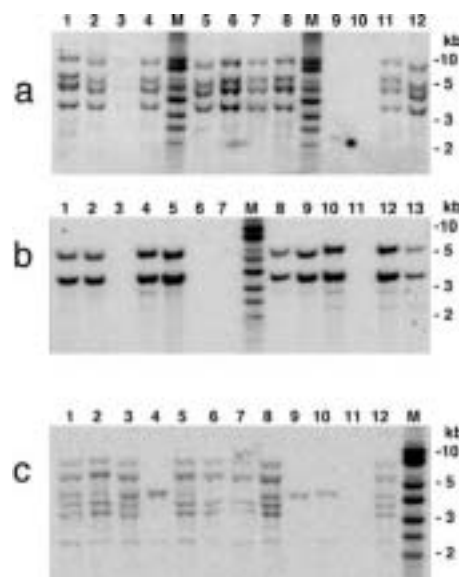


Figure 2.1 Southern blot analysis of the progeny of three transgenic plants containing pBin19 type T-DNA. Plant DNA was digested with restriction enzyme *EcoR*I and *Nco*I. A fragment corresponding to the *gus* gene was used as a probe. This probe revealed the left border (LB) of the T-DNA. (a) progeny of plant PG5; (b) progeny of plant PG107; (c) progeny of plant PG110. M: molecular weight marker.

In one plant (line N126) isolated from this transgenic population, the progeny segregated plants devoid of anthocyanin in a ratio (26 out of 126) corresponding to a monogenic recessive mutation. Southern blot experiment on a small plant sample indicated that in this plant a single T-DNA locus was genetically linked to the mutation. However, this T-DNA locus contained multiple T-DNA copies (data not shown). PCR experiments

revealed that in the population with wild-type phenotype, the *nptII* and *gus* gene sequences could be detected in only two third of the plants, however these sequences were present in all the plants with the mutant phenotype, indicating a tight genetic linkage between the mutation and the T-DNA locus. Genes involved in anthocyanin production have been characterized and their sequences are conserved in plants. We thus hypothesized that if the T-DNA insertion altered one of these genes in this *M. truncatula* line, the isolation of T-DNA borders might help rapidly characterize the mutated gene. A PCR walk experiment using genomic DNA from this line allowed us to isolate four different LB- and two RB-sequences confirming the complexity of this T-DNA locus. The sequences of these borders revealed T-DNA/T-DNA junctions as well as unmodified T-DNA-vector junctions. Only two LB were linked to sequences unrelated to the original vector. Data base search indicated that they probably represent sequences of a *M. truncatula* retrotransposon from the Athila family. From these results we deduced that in this line the transgenic locus contains four T-DNA copies as well as the vector sequence inserted in a single locus corresponding to a retrotransposable element. Thus, the mutation present in this line did not result from the insertion into a gene of the anthocyanin pathway, even if the mutation seems to be genetically linked to the T-DNA locus. This analysis also revealed that, in at least one *Medicago* plant, the complete vector sequence was integrated in the plant genome.

### 2.3.2 The pSLJ8313 and pSLJ8337 T-DNAs insert as single copy T-DNAs in *M. truncatula* genes

The T-DNAs carried on plasmids pSLJ8313 and pSLJ8337 contain defective transposable elements derived from the maize *En/Spm* transposable element (Ratet and Rippa, 1999; Tissier *et al.*, 1999). A population of 82 transgenic plants containing these T-DNAs was originally generated to evaluate the transposition activity of this element in *M. truncatula*. This transposition behavior in *M. truncatula* will be described elsewhere.

Southern blot analysis of plants from this population using a RB probe (*gus*) or a LB probe (*cytP450*; Fig 2.2) indicated that the majority of them contains a single T-DNA insert and that the entire T-DNA was integrated in the plant genome despite its large size (22 kb). In at least two plants containing multiple T-DNA inserts, the restriction pattern using restriction enzyme *Clal* indicated that methylation had occurred at these loci (data not shown).

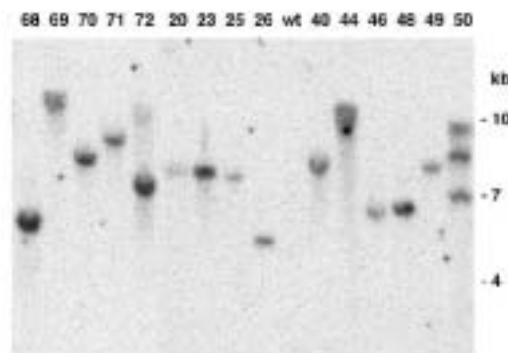


Figure 2.2 Southern blot analysis of plants transformed with the pSLJ8313 T-DNA. Genomic DNA was digested with restriction enzyme *Bam*HI and the Southern blot was hybridized with a *SuaC* probe (Tirichine *et al.*, 2000), corresponding to the left T-DNA border region. The molecular weight scale is given on the right of the figure.

To further characterize these T-DNA insertion sites, we isolated sequences flanking the T-DNA. Flanking sequences for both RB and LB were isolated for eleven lines, and only for the RB from one line. The size of the border fragments that we could amplify from these transgenic plants varied from 302 bp to 2785 bp (Table 2.1). For 8 lines the wild-type locus could also be amplified by using flanking sequence information. By comparing these wild-type sequences with the sequence of the tagged loci, small insertions or deletions were detected at each locus (Table 2.1). Deletions were detected in all the lines analyzed and varied from 5 to 404 bp. Insertions corresponded to filler sequences (Tinland, 1996) of unknown origin (392 bp for line Sy4, 22 bp for line Sy6 and 3 bp for line SP11) or originated from an internal T-DNA sequence (38 bp for line Sy8).

The positions of the break points between the T-DNA and the genomic sequences (Table 2.1) indicate that deletions from 8 bp up to 201 bp occurred upon integration of the T-DNA at the LB borders, in agreement with the T-DNA integration model described by Tinland (Tinland, 1996). However deletions of the RB as large as 161 bp were also found (line SP41), indicating that in some cases this border can also be degraded upon insertion in the *M. truncatula* genome.

Table 2.1 T-DNA border analysis of transgenic plants.

lines	Breakpoint (a)		cloned border (b)		wt locus (b)	del. (b)	ins. (b)
	LB	RB	LB	RB			
Plants transformed with pSLJ8313 and pSLJ8337 T-DNA.							
Sy4	39	3	1048	2785	3529	96	392
Sy5	205	33	1068	2072	ni		
Sy6	61	22	1265	600	1896	52	22
Sy8	94	36	338	302	818	206	38
Sy11	77	85	612	947	ni		
Sy12	54	140	423	2500	3083	162	0
Sy13	60	45	1178	744	2089	163	0
SP2	7	2	436	1713	2148	5	0
SP11	24	26	1503	338	2242	404	3
SP16	37	19	1284	754	2214	137	0
SP17	145	20	531	879	ni		
SP41	Ni	164	ni	2105	ni		
Plants transformed with GKB5 T-DNA.							
A7-1	77	9	916	1956	3095	223	0
B7-2	26	nd	575	Ni	ni		
B9-1	1	2	292	2256	2667	119	0
L6-1	202	2	123	429	ni		
L8-1	6	nd	866	Ni	ni		
M9-8	24	0	451	641	1131	39	0
M17-1	68	nd	1073	Ni	ni		
N4-1	33	0	609	515	1467	343	0
N27-2	68	20	399	202	ni		

(a): the break point represents the end of the T-DNA sequence incorporated in the plant genome. For the right border (RB) and the left border (LB), these values represent the number of bp of the T-DNA sequence lost upon integration in the plant genome. (b): The size of the cloned LB, RB, PCR-amplified wild-type (wt) loci and of the deletions (del.) or the insertions (ins.) at the transgenic locus are given in bp. nd: not determined; ni: not isolated

To estimate the proportion of T-DNA insertions in genes into intergenic sequences or in repetitive sequences, we compared the sequences of the tagged loci with those of the data base (Table 2.2). This analysis indicated that in the 12 lines studied, the T-DNA was inserted 5 times in genes or in their vicinity. In line Sp16 the T-DNA was inserted in the coding sequence of a gene coding for a protein homologous to the putative protein ATF17J16 of *A. thaliana*. In line Sy12 the T-DNA was inserted in a gene corresponding to the *M. truncatula* EST 422192. In line Sp41, the T-DNA was inserted in a gene corresponding to *M. truncatula* EST AW586897 and in line Sy5, 1,5 kb downstream of the coding sequence of a gene (putative anthranilate N-benzoyltransferase) corresponding to the *M. truncatula* expressed sequences TC39916. In line Sy4 the T-DNA was inserted 650 bp downstream of the coding sequence of a gene showing significant homology to the plant RAF kinase CTR1 involved in ethylene perception (Chang *et al.*, 1999). From line Sy11 the LB corresponds to the sequence of a *M. truncatula* transposable element homologous to the transposable elements *Tam1* of *Antirrhinum* and *En/Spm* of maize. The other sequences do not correspond to known sequences. These results thus indicate that in *M. truncatula* the T-DNA can frequently insert in genes or in their vicinity.

### 2.3.3 pGKB5 T-DNA can generate gene fusions with the *gus* reporter gene in *M. truncatula*

In addition to their use as gene tag, T-DNAs can be used to generate *in vivo* gene fusions to reporter genes as for example with the *gus* gene (Springer, 2000). The lines obtained in this way can reveal new patterns of gene expression and can be valuable tools for various developmental studies, in addition to providing direct information about the pattern of expression of the tagged gene.

We have constructed a population of 187 transgenic plants using the pGKB5 vector that allows to make gene fusions with the *gus* reporter gene and carries the kanamycin (*nptII* gene) and Basta (*bar* gene) selectable markers (Bouchez *et al.*, 1993). Because the Basta selection is very efficient in *M. truncatula*, the kanamycin resistance marker was not used in this experiment. Plants of this population as well as their progenies were analyzed for their T-DNA content and for the expression of the transgene during symbiosis with *S. meliloti*.

Southern blot analysis of randomly chosen plants indicated that the T-DNA of this vector preferentially inserted as single copy (9 out of 10) in the plant genome (Fig. 2.3), as found in the previous experiment with the pSLJ vectors. In addition, analysis of the progeny of randomly chosen T0 Basta resistant plants revealed segregation of the resistance marker which was compatible with the presence of a single T-DNA locus in the majority (21/26) of these lines (data not shown). Thus, transgenic plants generated using the Basta selection yielded plants containing mostly single T-DNA inserts.



Table 2.2 Identification of the *M. truncatula* tagged loci.

Lines	Acc. N°	MtEST(a)	<i>A. thaliana</i> (b)	function (c)	disrupted(d)
Plants transformed with pSLJ8313 and pSLJ8337 T-DNA.					
Sy4	AJ310823		T04683	RAF kinase	no
Sy5(RB)	AJ310831	TC39916	T45961	Anthranilate N-benzoyltransferase	no
Sy5(LB)	AJ310832				
Sy6	AJ310824				
Sy8	AJ310825				
Sy11(RB)	AJ310829				
Sy11(LB)	AJ310830			transposon ( <i>Tam 1</i> )	yes
Sy12	AJ310822	422192	A96735	Hypothetical	yes
Sy13	AJ310833				
SP2	AJ310821				
SP11	AJ311710				
SP16	AJ310820		ATF17J16	hypothetical	yes
SP17(RB)	AJ310826				
SP17(LB)	AJ310827				
SP41	AJ310828	AW586897	AC003105	hypothetical	yes
Plants transformed with GKB5 T-DNA.					
A7-1	AJ311927	TC11443	B71431	hypothetical	yes
B7-2	AJ311928				
B9-1	AJ307887	TC12312	AB025617	hypothetical	yes
L6-1(RB)	AJ311929				
L6-1(LB)	AJ311930				
L8-1	AJ311931				
M9-8	AJ311932			retrotransposon (copia)	no
M17-1	AJ311933				
N4-1	AJ311934	AW683943	A96753	threonine synthase	no
N27-2(RB)	AJ311935				
N27-2(LB)	AJ311936				

(a): *M. truncatula* TC and EST expressed sequences are from the TIGR Medicago gene index (<http://www.tigr.org/tdb/mtgi>).

(b): The accession number for the *A. thaliana* gene with the highest homology is given in the table.

(c): Putative functions are deduced from sequence comparisons. "Hypothetical" means that the function of the gene cannot be deduced from sequence comparison.

(d): yes or no indicates whether the transcription unit identified at this locus was disrupted by the insertion of the T-DNA.

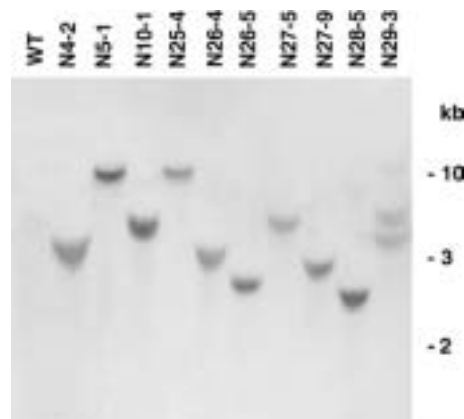


Figure 2.3 Southern blot analysis of independent transgenic plants transformed with the pGKB5 T-DNA. Genomic DNA was digested with restriction enzyme *Eco*R1 and *Bam*H1 and the Southern blot was hybridized with a *gus* probe, corresponding to the right T-DNA border region. The molecular weight scale is given on the right of the figure.

All T0 plants were tested for Gus activity in roots or nodules induced by *S. meliloti*. Among them, 19 plants showed Gus activity in nodules and were studied in more detail. Gus activity could also be observed in other parts of the plants (Table 2.3), indicating that the obtained gene fusions were not specific for the symbiotic organ. Different patterns of expression were observed in these plants (Table 2.3; Fig. 2.4), indicating that fusions to different genes were obtained. Interestingly, in nodules, we obtained fusions expressed in the meristematic region, in the apical part, in zone II and around the nitrogen fixation zone, in addition to fusions expressed in the vascular tissue. Several representatives of these fusions are shown in Fig. 2.4. Thus, gene fusions expressed in different regions of this organ (Timmers *et al.*, 2000; Vasse *et al.*, 1990) could be obtained even in this small scale experiment.

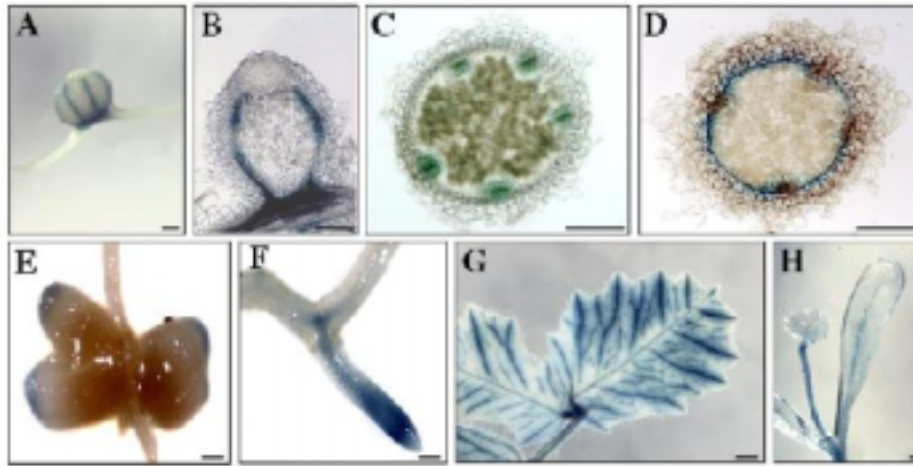


Figure 2.4 Gus expression in *M. truncatula* transgenic plants transformed with the pGKB5 T-DNA. A: mature nodule of line M3-3 showing Gus activity around the vascular tissue and at the nodule basis. B: longitudinal section of a mature nodule of line B9-1. The Gus staining is localized around vascular tissue. C and D: transversal section of nodules from lines N21-1 (C) and F8-1 (D). The Gus staining is localized around the nodule vascular tissue and between the inner and outer cortex respectively. E: mature nodules of line N4-1. Gus staining is visible at the nodule apex. F and G: Gus staining in the root elongation zone and in the leaf vascular tissue of line A7-1. Note that in these leaves only the side veins are stained. H: Gus staining in the vascular tissue of the cotyledons and the petiole of the first leaf (unifoliolate) of line L6-1. Bars = 200µm.

Table 2.3 Pattern of GUS expression of *in vivo* made *gus* gene fusions.

line	Cotyledon	unifoliolate leaf	trifoliolate leaf	stem	root	Nodule
A7-1	+	+	+	+	++	++
B9-1	+	+	+	+	+	++
F2-6*	nt	nt	nt	nt	+	++
F8-1	-	-	-	-	+	++
L6-1	+	-	-	+	+	+
L8-1	+	-	-	-	+	+
M3-3	-	-	-	-	+	++
M5-1	-	-	-	-	+	+
M9-8	-	-	-	-	+/-**	+/-**
M12-2	+	+	-	+	+	+
M17-1	-	-	-	-	++	+
N4-1	-	-	-	-	+	++
N14-7	-	-	-	-	++	+
N21-1	nt	Nt	-	nt	+	+
N27-2	+	+	+	-	+	+
N27-5	+	-	-	-	+	+
N27-8	-	-	-	-	+	++
N27-9	-	-	-	-	+	+
N36-1	-	-	-	-	+	+

For each line, at least 5 plants were tested (nb. all these plants were initially isolated as expressing Gus in roots or nodules). The intensity of Gus staining is indicated by +, - indicates no detectable signal. \* not yet confirmed in T1. \*\* staining pattern variable. nt: not tested.

The T-DNA borders of 8 plants expressing the fusion in the nodules and of one plant without Gus activity (line B7-2) were recloned in order to characterize the tagged loci as well as to gain information on the behavior of this T-DNA (Table 2.1). The RB sequence was obtained for 6 lines and the LB sequences for all the lines. For 4 lines we could isolate the corresponding wild-type locus. The sequence of the non-mutated loci indicated that the integration of the T-DNA in the transgenic plants induced deletions of 223 bp in line A7-1, 119 bp in line B9-1, 39 bp in line M9-8 and 343 bp in line N4-1 (Table 2.1).

The sequences of the border regions also indicated that deletions of the T-DNA left border ranging from 1 (B9-1 line) to 202 bp (L6-1 line) and of the right border ranging from 0 (M9-8 and N4-1) to 21 bp (N27-2 line) occurred upon T-DNA integration into the plant genome. Surprisingly, despite that Gus expression was detected in each of these 8 lines, gene fusion was found only for line B9-1 (see below).

In line N4-1 expressing Gus activity in the meristematic region of the nodules, the T-DNA was inserted 189 bp downstream of the stop codon of a threonine synthase homologue gene. The sequence of the *M. truncatula* EST (AW683943; Table 2.2) corresponding to this locus indicates that the insertion had occurred 10 bp after the transcribed region of this gene. In addition, the *gus* coding sequence is in opposite orientation of this gene, indicating that the pattern of expression of the reporter gene probably does not reflect the pattern of expression of this gene. Sequencing of a larger region will indicate if the presence of another gene upstream of the reporter gene could explain the pattern of GUS expression observed in this line.

In line N27-2 showing Gus expression in leaves, roots, nodules and cotyledons (Table 2.3), the sequence flanking the right T-DNA borders corresponds to the CaMV35S promoter region indicating that two T-DNA copies are physically linked at the same locus in this transgenic line. In fact, the *gus* gene is probably expressed in this line from the CaMV35S promoter region of the other T-DNA copy. In the other lines tested, we could not detect any gene in the vicinity (150 bp to 1 kb, Table 2.1) of the T-DNA.

In line A7-1 expressing the *gus* reporter gene in the vascular tissues of leaves, roots, stem, flower and nodules (Table 2.4, Fig. 2.4) a translation initiation site was fused in frame with the *gus* gene 20 codons before the *gus* AUG. This AUG is placed in a context (AAAAUGGUGA) compatible with an efficient translation initiation in plants. However, no open reading frame could be detected downstream of this sequence in the wild-type locus. Inspection of this sequence for coding capacity did not reveal any putative gene in this region in the same direction as the *gus* gene, moreover the T-DNA was inserted 3' and in opposite direction of a gene corresponding to the *M. truncatula* expressed sequence TC11443 (Table 2.2). Comparison of the sequence from the wild-type sequence with the border sequence revealed that 223 bp were deleted upon T-DNA insertion in this line resulting in the deletion of the last 49 amino acids from the putative protein encoded by this locus. Thus, despite that a gene was tagged in this line, the observed Gus expression pattern did not result from a gene fusion.

In line B9-1 the *gus* gene was fused in frame with the coding sequence of a *M. truncatula* expressed sequence (TC12312; Table 2.2) homologous to the *M. truncatula* nodulin MtN3 described by Gamas *et al.* (Gamas *et al.*, 1996). The sequence of the wild-type locus indicated that a deletion of 119 bp had occurred upon insertion of the T-DNA, resulting in a deletion of 39 amino acids from the MtN3-like protein. In this line the fusion protein was expressed around the vascular tissues of the nodules (Fig. 2.4) but not in the root. Expression of the fusion was also detected in the leaves and the flowers, indicating that its expression is not specific for symbiosis.

## 2.4 Discussion

In this work we have compared the insertion behavior of three different T-DNAs into the *M. truncatula* genome. These DNA sequences differed in their length, in the origin of their T-DNA borders, in the marker genes used for selection and in the vectors carrying them. We studied those characteristics which are critical for carrying out efficient mutagenesis. These are the copy number of the insert, the efficiency of the antibiotic resistance markers and the insertion in transcribed regions.

Surprisingly, marked differences concerning the T-DNA copy number were observed in the transgenic plants. The two constructs that were selected using Basta resistance yielded mostly single copy T-DNA inserts in the transgenic plants. In contrast, the T-DNA from the pBin19 derived vector that was selected in the plant using kanamycin, gave mostly multiple inserts in the plants. In addition, the insertion of this element seemed to have occurred at a restricted number of loci in each plant, because these copies did not segregate independently in the progeny of the transgenic plants. This indicates that these T-DNA copies were inserted in the plant genome as tandems or inverted repeats (concatamers). The most remarkable example was the line N126, where four copies of the T-DNA as well as the vector were inserted in a single locus. Such vector sequence integration has already been described for other plants and T-DNA constructs (De Buck *et al.*, 2000).

Since this multiple insertion phenomenon could have occurred only in this particular experiment, we looked back at results obtained with other transgenic plants previously obtained in the laboratory. Derivatives of the pBin19 vector were also used in a previous work (Trinh *et al.*, 1998) and we interpreted the Southern blot experiments as indications of multiple linked T-DNA inserts in independently obtained *M. truncatula* transgenic plants. Similar results were obtained with other pBin19 derivatives in the laboratory (I. d'Erfurth, unpublished). In a similar T-DNA tagging experiment in *L. japonicus* where a vector derived from pBin19 was also used, Martirani *et al.* (Martirani *et al.*, 1999) detected multiple inserts in more than 57% of the regenerated plants. Thus, all together these results indicate that vectors derived from pBin19 generated preferentially multiple inserts in *M. truncatula* and *L. japonicus*. Similarly, in tobacco Fobert *et al.* (1991) detected differences in the T-DNA copy number between pBin19 derivatives and another vector. These differences may arise from the combination of border sequences originating from the Ti plasmid pTiT37 (Bevan, 1984) with the *vir* genes originating from Ti plasmid pTiBo542 (Hood *et al.*, 1993). An alternative explanation is that we used the kanamycin resistance marker (*nptII* gene) during transformation. It is possible that the expression of only one copy of the *nptII* gene is

not enough for conferring resistance to this antibiotic in the first steps of the regeneration, resulting in the systematic selection for multiple insertions in *M. truncatula* in these experiments. A further possibility is that T-DNA transfer is very efficient with this *Agrobacterium* strain, resulting in a high copy number of the T-DNA molecules in the cells, which will preferentially integrate in a restricted number of competent loci (Tinland, 1996). Consequently, this type of vector is not suitable for T-DNA tagging and for *in vivo* gene fusions in *M. truncatula*. In contrast, the two other T-DNAs that we used in this work and whose selection was based on Basta resistance are suitable for the gene tagging strategy, because they preferentially insert in the plant genome as single copies.

Analysis of the border sequence break points (ends of the T-DNA sequence in the plant genome) revealed that during the integration process, both ends (RB and LB) were submitted to moderate loss of DNA sequence (up to 164 bp and 205 bp respectively for the pSLJ vectors and up to 20 bp and 202 bp respectively for the pGKB5 T-DNA). In fact larger deletions, resulting from partial T-DNA integration, may not have been possible to detect because of the method used for recloning or because of the position of the selectable marker gene or reporter gene in the vectors. However, the position of the oligonucleotides used for the PCR walk experiments allowed the detection of deletions up to 600 bp at the RB and 530 bp at the LB of the T-DNA carried by the pSLJ vectors. In the case of the pGKB5 vector used for gene fusion experiments, only small deletions were detected at the RB in the lines expressing GUS activity, because the *gus* coding sequence starts 40 bp after the right border. Previous experiments indicated that the sequence adjacent to the right border is generally not degraded upon T-DNA integration in plants (Tinland, 1996), however partial degradation of both T-DNA ends was also described by Greveling *et al.* (1993) in *Arabidopsis*. This degradation of the RB indicates that protection of the T-DNA by the *Agrobacterium* virD2 protein during the transfer to the plant genome might be less efficient in *M. truncatula* than in previously tested plants. Sequencing of more T-DNA borders may indicate whether this is a general feature of T-DNA integration in *M. truncatula*.

Comparison of the T-DNA border sequences with the sequence of the original genomic locus indicated that small rearrangements corresponding to deletions and insertions up to 400 bp occurred at these transgenic loci. Such rearrangements were previously described at the T-DNA integration loci (Tinland, 1996). Thus, the modifications of both T-DNA ends and the rearrangements of the tagged loci indicate that integration of the T-DNA in *M. truncatula* uses the same mechanism of illegitimate recombination that was previously described for its integration into the genome of other plants.

Sequence analysis of the 21 tagged loci obtained in this work revealed that in eight of them the T-DNA inserted either in the transcribed region of a gene (5 lines) or in its vicinity (3 other lines). Interestingly, sequences corresponding to transposable elements were also detected at the integration site of three additional lines. This indicates that in *M. truncatula* the T-DNA can also insert in repeated elements.

The frequency of tagging (disruption of a transcribed region) was 20% in our experiment. In a similar experiment with *Arabidopsis*, Mathur *et al.* (Mathur *et al.*, 1998) found 5 transgenic plants among 21 plants analyzed where the T-DNA was inserted in the coding region of a gene. In the other 16 plants the T-DNAs inserted in the promoter region of genes or in their vicinity or in unknown regions. Thus, in these small scale experiments the frequency of gene tagging in the two plants were similar, indicating that for gene discovery T-DNA tagging in *M. truncatula* can be as powerful as in *Arabidopsis*. Another interesting result of this small scale experiment is that the gene coding for the RAF kinase in line Sy4 does not have orthologues in *A. thaliana* and could represent a legume-specific gene (legumin).

Using the pGKB5 vector we have isolated a fusion between a gene homologous to the nodulin MtN3 and the *gus* reporter gene. In this line the GUS activity was detected in the vascular tissue of the symbiotic nodule but not in the root. This fusion is thus a valuable marker to study vascular tissue differentiation during nodule organogenesis in *M. truncatula*. It also indicates that by using this vector, it is possible to construct *in vivo* gene fusions by T-DNA insertion in the *M. truncatula* genome.

In this work we isolated plants expressing GUS fusions that do not seem to correspond to fusions to *M. truncatula* genes. Considering the sensitivity of the Gus system, they could be explained by the fact that we retained plants expressing fusions at low level that could result from the activation of cryptic promoters, as a consequence of the disruption of the original tagged locus upon insertion of the T-DNA. The expression of these cryptic promoters might be increased by enhancer sequences (CaMV35S promoter) present in the T-DNA. Such cryptic promoters were described in tobacco by Foster *et al.* (Foster *et al.*, 1999).

In conclusion the high frequency of gene tagging obtained in our work suggests that the T-DNA approach is very efficient in *M. truncatula* for gene discovery and systematic sequencing of the T-DNA borders will be feasible and valuable in this plant. Thus, our experiments open the way for larger gene tagging experiments in *M. truncatula*.

**Acknowledgements**

We thank Dr. David Bouchez for providing the pGKB5vector and N. Mansion for photographic work. This work was supported by grants from the Centre National de la Recherche Scientifique (Genome program, 1998-2000) and the Institut National de la Recherche Agronomique (Genomes and function program, 1998-2000). M. Scholte was supported by the EEC TMR program n° ERB4061PL970121



## 3 Generation and screening of a tagged *M. truncatula* population

### 3.1 Introduction

Although a number of plant genes induced during various steps of nodule formation and nodule development have been described (for a review see Miklashevichs, 2001), their exact functions remain mostly unknown and it is unclear whether they are essential for the nodulation process.

Genes involved in nodule development can be identified by genetic approaches such as analysis of mutant phenotypes, isolation of the mutated gene and complementation of the mutation. The classical and simplest methods for the creation of plant mutants are based on chemical or v-ray mutagenesis. Although in this way numerous symbiotic mutants have been found, the isolation of the mutated gene(s) is difficult, since it requires map-based cloning and so far only a few legume genes have been identified in this way.

This disadvantage can be overcome by insertional mutagenesis by introducing T-DNA with a known sequence in the plant genome. Besides causing a gene disruption, the T-DNA of *A. tumefaciens* provides a tag for the isolation of the mutated locus. In addition, if the T-DNA contains a reporter gene that is expressed upon insertion in a transcribed region of the plant genome, the resulting gene fusion may serve as a convenient marker for the analysis of gene expression.

In this work we used T-DNA tagging to gain more insight in the nodulation process and to identify novel genes involved in nodulation. To investigate the possibility of T-DNA tagging in the model legume plant *M. truncatula*, we aimed at the creation and analysis of a small tagged population using a binary vector with a promoter-less reporter gene. In this chapter the choice of the reporter gene, the production of the T-DNA tagged *M. truncatula* population and phenotypical and molecular analysis of the transgenic lines will be described.

### 3.2 Results

#### 3.2.1 Choice of the reporter gene

The presence a promoter-less reporter gene close to the border of the T-DNA may result in gene fusion upon insertion into the plant genome and this will allow direct screening for plants tagged in a gene. Since the T-DNA inserts preferentially randomly into the transcribed genome regions or in their vicinity, the chance to find insertions in genes of interest increases when the tagged population is larger. This implies that a reporter gene whose expression can be detected easily would be convenient to screen large numbers of tagged plants. The most widely used reporters in plants are the  $\beta$ -glucuronidase (GUS), luciferase and green fluorescent protein (GFP) (Jefferson *et al.*, 1986; Schneider *et al.*, 1990; Leffell *et al.*, 1997). In the case of GFP no substrate is required to monitor reporter activity. After excitation with UV or blue light (optimally at 396 nm or 475 nm) GFP emits green light with a spectrum having a maximum at 508 nm. This non-destructive detection of the reporter allows *in planta* screening for GFP expression. Luciferase allows also *in vivo* detection but requires an exogenously applied substrate and  $\beta$ -glucuronidase activity can not be detected *in vivo*. Therefore, GFP seemed to be the most convenient reporter gene. However at the beginning of this thesis, GFP had not yet been used in *M. truncatula*, and therefore the usefulness of this reporter gene first had to be tested. Two GFP constructs (*mGFP5-ER* (Haseloff *et al.*, unpublished) and *S65T-intron* (Pang *et al.*, 1996)) were tested. In the *mGFP5-ER* construct the cryptic intron splicing site present in wild-type GFP was eliminated and a signal peptide was added by which it became targeted to the lumen of the endoplasmatic reticulum (Haseloff *et al.*, 1997). Moreover, amino acid substitutions at V163A and S175G improved the GFP-fluorescence properties in comparison to wild-type GFP. The advantage of the *S65T-intron* construct was that the S65T substitution yielded a 100- to 120-fold brighter fluorescence than the wild-type GFP and the presence of the second intron of the potato *ST-LS1* gene avoided

expression of *GFP* in *Agrobacterium* (Pang *et al.*, 1996).

The potential use of these constructs expressed from the *CAMV 35S* promoter in the binary vector pCP60 was tested in hairy roots of *M. truncatula* plants obtained with *A. rhizogenes* mediated transformation. In the case of pCP60/35S::*S65T-intron* only 20 out of 93 *A. rhizogenes* infected plants had developed hairy roots, while in the case of pCP60/35S::*mGFP5-ER* only 3 out of 89 did so. The only plant developing fluorescent hairy roots was transformed with the 35S::*S65T-intron* fusion (Fig. 3.1 A) and no fluorescence was observed in the other hairy roots (result not shown).

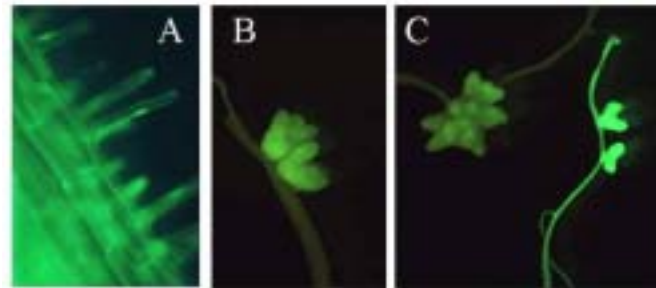


Figure 3.1 GFP in *M. truncatula*. **A:** Hairy root transformed with the *CAMV 35S* promoter::*S65T-intron*. Untransformed hairy roots were comparable to the root shown in B. **B:** Autofluorescence in nodules of a *M. truncatula* plant carrying the *MsENOD12A* promoter::*mGFP5-ER* fusion. **C:** wild-type root and nodule (left) and root and nodule carrying the *CAMV 35S* promoter *mGFP5-ER* fusion (right).

In addition, the expression of *mGFP-ER* from the *CAMV 35S* promoter in the binary vector pBIN was studied in stably transformed transgenic *M. truncatula* plants obtained with *A. tumefaciens* mediated transformation. Nine out of eleven independent primary transformants originating from different leaf explants, showed fluorescence in the roots. On the contrary, none of ten independent primary transformants carrying *MsENOD12A* promoter::*GUS* fusion, that were used as controls, showed fluorescence above autofluorescence level (result not shown). Thus in both hairy roots of *M. truncatula* and in stably transformed *M. truncatula* plants detectable *GFP* expression was possible.

In order to determine whether *GFP* can be used as a marker to study gene expression in a tagging experiment, both *GFP* variants were cloned behind the minimal promoter of the early nodulin *MsENOD12A* in the binary vector pBP12A and transgenic *M. truncatula* plants were obtained. Nineteen independent plants carrying the *MsENOD12A* promoter::*S65T-intron* fusion and ten independent plants carrying the *MsENOD12A* promoter::*mGFP5-ER* fusion were obtained.

The *GFP* expression in nodules was studied in progeny from primary transformants containing the *CAMV 35S* promoter::*mGFP5-ER*, the *MsENOD12A* promoter::*mGFP5-ER*, the *MsENOD12A* promoter::*S65T-intron* or the *MsENOD12A* promoter::*GUS* construct. For each construct and wild-type control plants 30 seedlings were inoculated with *S. meliloti* and the *GFP* expression was studied in mature, nitrogen fixing-nodules. The nodules from plants transformed with one of the two different *MsENOD12A* promoter::*GFP* constructs all showed a diffuse fluorescence (shown for *MsENOD12A* promoter::*mGFP5-ER* in Fig. 3.1 B) that did not differ from the autofluorescence detected in the wild-type plants or the plants transformed with the *MsENOD12A* promoter::*GUS* fusion (a wild-type nodule is shown in Fig. 3.1C). In contrast, the plants containing the *CAMV 35S* promoter::*mGFP5-ER* construct produced brightly fluorescing nodules and roots (Fig. 3.1 C). Although in the nodules the *GFP* could be detected when expressed from a strong promoter, due to the high level of autofluorescence in this organ, *GFP* seemed not to be a convenient reporter gene for the insertional mutagenesis of *M. truncatula* aiming at the discovery of nodulin genes (see discussion). Therefore, we decided to use the more labor-intensive but clearly detectable *GUS* reporter gene that was used successfully in *M. truncatula* in other studies.

### 3.2.2 Transformation and regeneration

In *Arabidopsis* the binary vector pGKB5 (Fig. 3.3 A) was used successfully to produce a large tagged population (Bouché *et al.*, 1993). The T-DNA of this vector contains a promoter-less *GUS* gene linked to its right border and confers Basta and kanamycin resistance to transformed plants. The absence of stop codons in the three reading frames between the right T-DNA border and the start codon of the *GUS* gene makes transcriptional and translational gene fusions possible upon insertion in the genome. Figure 3.2 gives an overview of the procedure that was followed to produce a small T-DNA tagged population with this vector, and the subsequent analysis of the obtained plants.



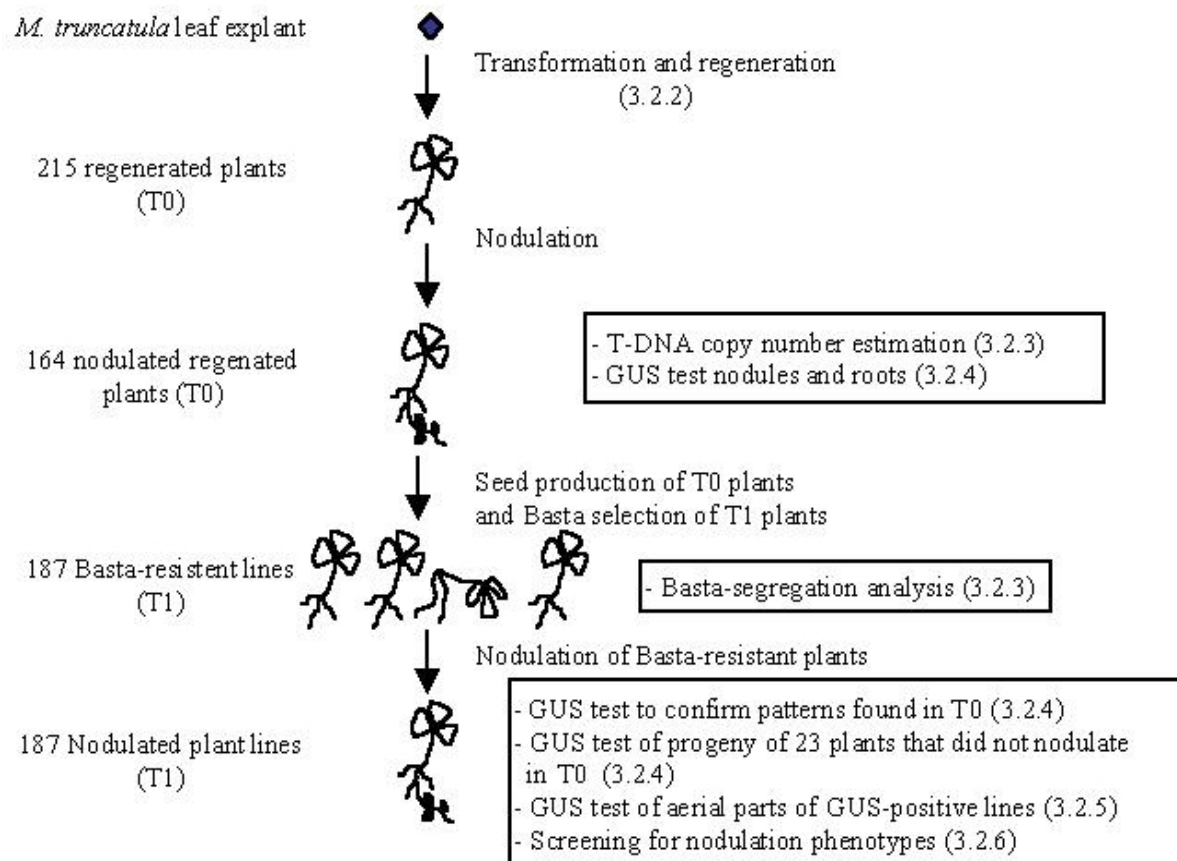


Figure 3.2 Overview of the procedure that was followed for the production and analysis of the T-DNA tagged population. The numbers between brackets refer to the paragraphs where the corresponding results are described.

In 14 independent transformation experiments (A-N) with pGKB5 (Bouchez *et al.*, 1993) a total of 236 plants was regenerated following the transformation and regeneration protocol developed by Trinh *et al.* (1998). Each leaf explants used for transformation was numbered and once the somatic embryos had formed, they were immediately separated and numbered. To each regenerated plant obtained a code was given that indicated the transformation experiment, the leaf explant number and the embryo number. For example plant M17-1 was obtained in transformation experiment M from leaf explant number 17 as the first embryo. For transformation rounds A and B however, the somatic embryos were not separated from each other, thus creating the possibility that two regenerated plants originated from the same transformed cell and were genetically identical. In order to minimize the chance to obtain siblings, during all other transformation rounds only plants originating from somatic embryos that had developed on different parts of the leaf explant were kept. In total 237 primary transformants were obtained.

The primary transformants produced between 5-100 seeds per plant, except for primary transformant A9-4 that died before flowering. In this thesis the entire progeny of a primary transformant (homozygous T-DNA insertion mutants, heterozygous T-DNA insertion mutants and wild types) will be called a line. From each regenerated plant 5-60 seeds were sown in Perlite. In order to eliminate Basta-sensitive plants (thus not transformed or carrying a truncated T-DNA that missed the *BAR* gene), the seedlings were screened for Basta resistance. In total 187 plant lines contained Basta resistant plants. Progeny of the other 49 primary transformants did not survive the Basta treatment and thus appeared to originate from escapes of the Basta selection during the regeneration procedure or, from a primary transformant that did not produce transformed seeds. There seemed to be a positive correlation between the presence of Basta in the SH9 medium and the number of transformants: 133 (=86%) of the 154 plants originating from leaf explants from transformation rounds F, K, L, M and N that were cultured at least two weeks in the light on SH9 medium containing Basta were Basta-resistant, whereas only 54 (=66%) of the 82 plants originating from the other transformation rounds, where the selection was stopped when the somatic embryo's were transferred to SH9, were Basta-resistant.

### 3.2.3 Estimation of the T-DNA copy number

In order to estimate the T-DNA copy number of tagged plants, Southern analysis of 45 primary transformants using the *GUS* probe was carried out. The *GUS* gene has two *EcoRV* restriction sites and there is one *EcoR*I site just downstream from the *GUS* gene. Therefore, when genomic DNA from the transgenic plants was digested with *EcoR*I and *EcoRV* the *GUS* fragment used as a probe revealed 2 internal fragments of the T-DNA (230 and 1400 bp) and fragments corresponding to the flanking regions of the T-DNA right border (Fig. 3.3 B). The number of these fragments corresponded with the number of insertions. When the DNA was digested with *EcoR*I and *Bam*H1, none of them cutting inside the *GUS* gene, the *GUS* probe revealed DNA fragments of at least 2300 bp, corresponding to the right border flanking region (Fig. 3.3 B). The Southern analysis indicated that 41 of the 45 tested plants contained a single T-DNA insert. Table 3.1 shows a summary of these results. In addition, the Southern analysis showed, that the strategy to maximize the number of independent transformants used for transformation rounds C-N (§ 3.2.2) was successful, since none of the plants originating from the same leaf explant showed identical hybridization patterns.

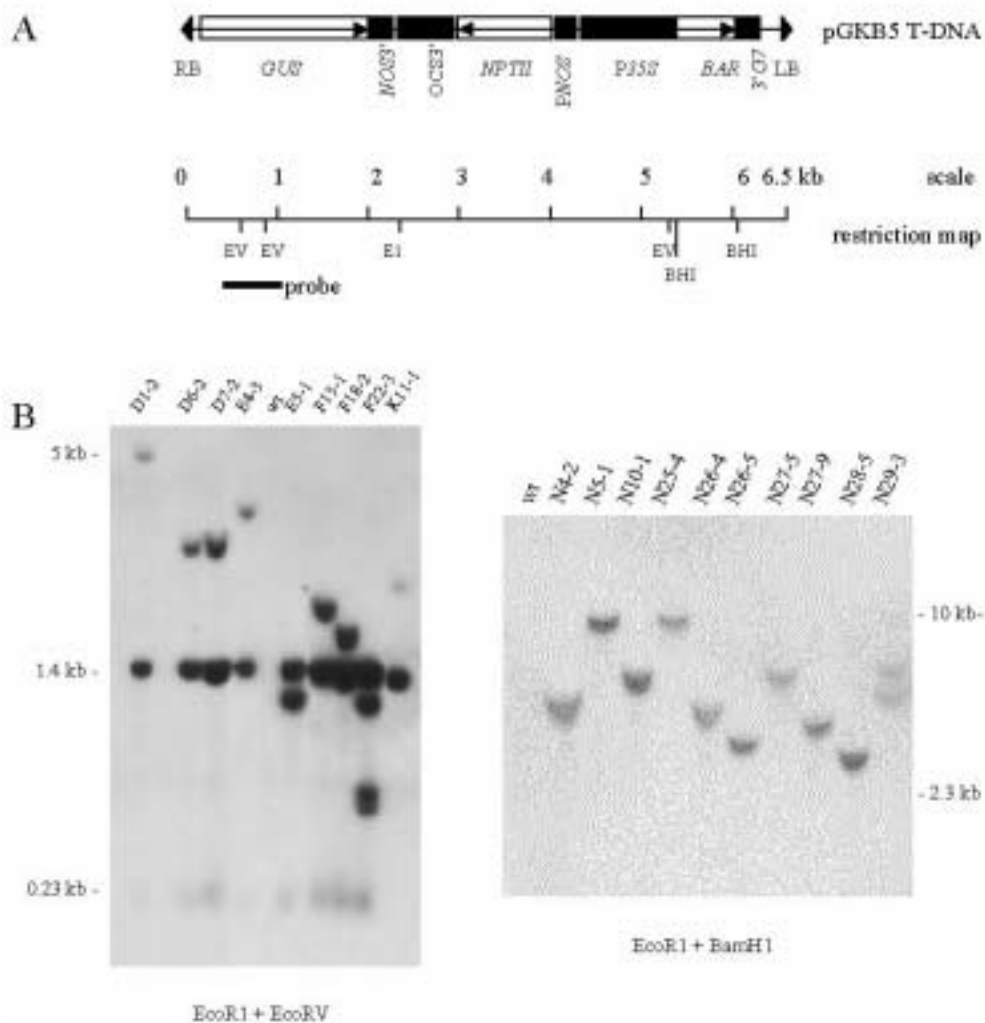


Figure 3.3 Southern blot analysis of *M. truncatula* plants carrying the GKB5 T-DNA. **A**: Functional map of the pGKB5 T-DNA. Arrows indicate coding sequences, and black boxes promoter/terminator regions. *GUS*: coding region of the β-glucuronidase gene from *E. coli*; *NOS* 3': 3' region of the nopaline synthase gene from pTiC58; *OCS* 3': 3' region of the octopine synthase gene from pTiAch5; *NPTII* neomycin phosphotransferase II; *PNOS*: promoter region of the nopaline synthase gene; *P35S*: promoter of the 35S gene of the Cauliflower Mosaic Virus; *BAR*: coding sequence of the Basta resistance gene from *Streptomyces hygroscopicus*; 3'*G7*: 3' region of the gene 7 from the T-DNA of pTi15955. Below the functional pGKB5 T-DNA map the restriction map of the GKB5 T-DNA showing the restriction sites of the enzymes used for digestions of genomic DNA for Southern hybridizations. The black bar indicates the *GUS* probe. **B**: Southern blot hybridizations with *GUS* probe of genomic DNA digested with *EcoR*I and *EcoRV* (left) and *EcoR*I and *Bam*H1 (right).

Table 3.1 Southern analysis of transgenic plants carrying the pBGKB5-TDNA.

Line	Digestion	Size of the T-DNA border in bp	line	Digestion	size of the T-DNA border in bp
A7-1	EcoR1/BamH1	3500	F22-3	EcoR1/EcoRV	700+750+1200
B1-1	EcoR1/EcoRV	900	I2-2	EcoR1/BamH1	2500+4500
B1-2	EcoR1/EcoRV	900	K3-2	EcoR1/BamH1	4000
	EcoR1/BamH1	4000	K11-1	EcoR1/EcoRV	9000
B4-5	EcoR1/EcoRV	1000	L6-1	EcoR1/BamH1	2000
B6-1	EcoR1/EcoRV	1000	L8-1	EcoR1/EcoRV	7000
B7-2	EcoR1/EcoRV	5000	M9-8	EcoR1/EcoRV	2800
B9-1	EcoR1/EcoRV	2700		EcoR1/BamH1	3000
	EcoR1/BamH1	4000	M17-1	EcoR1/EcoRV	800
B9-2	EcoR1/EcoRV	1000	N4-1	EcoR1/EcoRV	1500
	EcoR1/BamH1	2500	N4-2	EcoR1/BamH1	4500
B11-1	EcoR1/EcoRV	8000	N5-1	EcoR1/BamH1	10000
B13-1	EcoR1/EcoRV	2700	N10-1	EcoR1/BamH1	6000
B14-1	EcoR1/EcoRV	2500	N12-1	EcoR1/BamH1	3000
B14-3	EcoR1/EcoRV	2500	N20-1	EcoR1/BamH1	9000
C5-4	EcoR1/EcoRV	8000	N21-2	EcoR1/BamH1	4500
D1-2	EcoR1/EcoRV	5000	N25-4	EcoR1/BamH1	9000
D6-2	EcoR1/EcoRV	3000	N26-4	EcoR1/BamH1	4000
D7-2	EcoR1/EcoRV	3000	N26-5	EcoR1/BamH1	3000
E4-3	EcoR1/EcoRV	3500	N27-2	EcoR1/BamH1	2500
	EcoR1/BamH1	4500	N27-5	EcoR1/BamH1	6500
E5-1	EcoR1/EcoRV	1200	N27-9	EcoR1/BamH1	3500
F13-1	EcoR1/EcoRV	2000	N28-5	EcoR1/BamH1	2500
F16-2	EcoR1/BamH1	4000	N29-3	EcoR1/BamH1	4500+6500
F18-2	EcoR1/EcoRV	1800	N31-6	EcoR1/BamH1	3500+6000+10000

For each plant the size of the T-DNA RB, revealed by the *GUS* probe, is indicated in bp. Bands corresponding to internal fragments are not mentioned

Table 3.2 Segregation analysis of the Basta resistance marker from pGKB5 primary transformants.

Lines tested	Total plants tested	Basta-s (a)	Basta-r (b)	$\theta^2$ (c)	P value (d)	Lines tested	total plants tested	Basta-s (a)	Basta-r (b)	$\theta^2$ (c)	P value (d)
A9-3	37	11	26	0,44	0.50-0.75	N9-3	50	10	40	0,67	0.25-0.50
A10-5	33	11	22	1,22	0.25-0.50	N10-1	39	16	23	5,34	0.01-0.025*
B9-2	50	8	42	2,16	0.10-0.25	N12-1	26	4	22	1,28	0.25-0.50
D6-2	23	5	18	0,13	0.50-0.75	N19-4	33	13	20	3,65	0.05-0.10
F9-2	41	14	27	1,83	0.10-0.25	N19-7	46	12	34	0,03	0.75-0.90
F11-1	23	11	12	6,39	0.01-0.025*	N21-1	36	10	26	0,15	0.50-0.75
F18-3	35	7	28	0,47	0.25-0.50	N21-2	32	8	24	0	0.999
I2-2	37	9	28	0,01	0.90-0.95	N23-5	44	14	30	1,09	0.25-0.50
I8-9	31	9	22	0,27	0.50-0.75	N25-1	37	6	31	1,52	0.10-0.25
I13-1	48	14	34	0,44	0.50-0.75	N25-4	41	7	34	1,37	0.10-0.25
K3-2	59	14	45	0,05	0.75-0.90	N26-4	56	16	40	0,38	0.50-0.75
K7-2	46	28	18	31,57	<0.001*	N26-5	40	6	34	2,13	0.10-0.25
L2-3	41	12	29	0,40	0.50-0.75	N26-9	45	8	37	1,25	0.25-0.50
L2-4	41	13	28	0,98	0.25-0.50	N27-2	23	4	19	0,71	0.25-0.50
L8-1	42	14	28	1,56	0.10-0.25	N27-5	60	10	50	2,22	0.10-0.25
L8-8	48	22	26	11,11	<0.001*	N27-7	42	10	32	0,03	0.75-0.90
M3-1	29	13	16	6,08	0.01-0.025*	N27-8	51	12	39	0,06	0.75-0.90
M4-2	43	9	43	0,38	0.50-0.75	N27-9	46	12	34	0,03	0.75-0.90
M5-1	53	14	39	0,06	0.75-0.90	N27-11	24	3	21	2	0.10-0.25
M8-1	50	10	40	0,67	0.25-0.50	N28-3	29	9	20	0,56	0.25-0.50
M9-1	49	13	36	0,06	0.75-0.90	N28-5	56	12	44	0,38	0.50-0.75
M9-2	19	15	34	17,29	<0.001*	N29-3	37	8	29	0,23	0.50-0.75
M9-8	29	7	22	0,01	0.90-0.95	N31-6	25	5	20	0,33	0.50-0.75
M9-9	35	11	24	0,77	0.25-0.50	N33-2	61	15	46	0,01	0.90-0.95
M13-5	37	19	18	13,70	<0.001*	N34-3	45	13	32	0,36	0.50-0.75
N4-2	45	11	34	0,01	0.90-0.95	N35-2	29	13	16	6,08	0.01-0.025*
N4-4	40	6	34	2,13	0.10-0.25	N37-1	41	8	33	0,66	0.25-0.50
N5-1	32	4	28	2,67	0.10-0.25						

If more than 20 T1 seeds were available, the ratio of Basta-resistant: Basta-sensitive plants was scored.

(a): Basta-s: Basta sensitive plants; (b): Basta-r: Basta resistant plants; (c):  $\theta^2$  value calculated for the expected segregation ratio 3:1

(d): Tabulated  $\theta$  values for the calculated  $\theta^2$  values. P = 0.05 was chosen as a critical limit, such that the predicted segregation was not rejected for P values >0.05. \* indicates a significant segregation of the resistance marker in a 1:1 ratio.

Twenty out of the 45 primary transformants that were used for the Southern analysis and 35 other primary transformants produced enough seeds to test at least twenty T1 progeny plants for the segregation of the Basta-resistance marker. Chi square analysis indicated that 47 (=86%) of these 55 lines showed the Mendelian 1:3 segregation for Basta resistance (Table 3.2). This indicated that the vast majority of the lines contained a single T-DNA insertion, which is according to the Southern analysis. However, lines I2-2, N29-3 and N31-6 showed several bands by Southern analysis, but a 1:3 segregation which may indicate that these plants contain closely linked T-DNA insertions. In addition, Southern analysis of line N10-1 indicated one T-DNA insertion, while (as was the case in lines F11-1, K7-2, L8-8, M3-1, M9-2, M13-5 and N35-2) the Basta-resistance marker segregated in a 1:1 ratio. One possible explanation is that in this line the T-DNA insertion caused male or female sterility, thus resulting in a 1:1 segregation of the Basta-resistance marker upon selfing of T0 plants.

### 3.2.4 GUS staining patterns of the regenerated plants

The presence of GUS activity resulting from insertion of the promoter-less *GUS* gene downstream of a promoter or *in frame* with the coding part of a gene, was tested in all the 236 primary transformants (Basta-resistant and Basta-sensitive). Each transformant was grown individually in a pot and watered with low-nitrogen nutritive solution. After two weeks of nitrogen starvation and adaptation to greenhouse conditions, the plantlets were inoculated with *S. meliloti* and 4 weeks post-inoculation 199 plants had formed nodules that could not be distinguished from wild-type nodules. The fact that the other 38 plants did not nodulate could be due to dominant mutations but probably was caused by their bad physiological condition (see discussion). From each nodulated plant part of the roots and nodules were harvested and incubated over night in a buffer containing 2 mM X-Gluc. 18 primary transformants showed GUS staining in the nodules and/or the roots. After the GUS assay the plants were transferred to vermiculite and watered with commercial nutritive solution and allowed to set seed.

To confirm that the GUS staining found in the primary transformants was maintained in the T1 generation, T1 plants were first selected with Basta and then nodulated (except for T0 plant A9-4 that died before flowering and F2-6, that produced less than 5 (Basta-resistant) seedlings). From each Basta-resistant line 5-10 T1 plants were tested for GUS staining. The GUS staining patterns found in the primary transformants were maintained in all T1 progenies, except in plant line M9-8 (see further).

The progeny of the T0 plants that did not form nodules has been tested simultaneously and now all the plants formed normal, nitrogen-fixing nodules. GUS staining of these nodules resulted in the discovery of one additional *GUS* expressing lines (line N36-1), making the total number of plant lines expressing *GUS* 19 out of 187 Basta-resistant lines.

The GUS staining (Fig. 3.4) varied from weak and diffuse (e.g. line A9-4, data not shown) to very strong and localized (e.g. line M3-3). None of the plants showed GUS staining exclusively in the nodules. However, in lines B9-1, F2-6, F8-1, M3-3, N4-1, and N27-8 the GUS staining was more intense in the nodules than in the roots. On the contrary, lines M12-2 and M17-1 showed a stronger GUS staining in the roots than in the nodules. In the majority of the tested lines the GUS staining was localized around the vascular tissue (in 88% of the GUS-positive tested roots and in 52% of the GUS-positive tested nodules).

In this small T-DNA tagged population, several plant lines displayed a unique GUS expression pattern in the nodules and the lines were classified according to this GUS staining pattern. Based on this, the plant lines are ordered into two groups: The first group exhibited GUS staining in the nodule vascular tissue, while the second group showed GUS staining elsewhere in the nodules. Plant lines A7-1, B9-1, L6-1, M3-3, M5-1, M17-1, N21-1, N27-2, N27-8, and N36-1 belonged to the first group and lines F2-6, F8-1, L8-1, M12-2, N4-1, N14-7, N27-5 and N27-9 to the second group.

Line M9-8 was, however, an exceptional case, since the GUS staining pattern varied in the different generations. In the primary transformant of line M9-8 the GUS staining was observed around the root vascular tissue (Fig. 3.4 A1) and all over the nodules (results not shown). In its progeny the GUS staining patterns varied among the tested plants. For example, one plant showed a patchy pattern in the root (Fig. 3.4 A2) and no staining in the nodules while another one showed a very intense, but not homogenous GUS staining in the nodules (Fig. 3.4 A3) whereas the root was not GUS stained. Because of this variability, the GUS staining pattern of line M9-8 will not be discussed further. Below representatives belonging to one of the two GUS staining pattern groups will be discussed.

#### Plant lines with GUS staining in the nodule vascular tissue

In the nodules of lines A7-1 (Fig. 3.4 B1), L6-1 (Fig. 3.4 D1), M17-1 (result not shown) and N27-8 (Fig. 3.4 G), the GUS staining was restricted to the vascular tissue. In lines M3-3 (Fig. 3.4 E) and B9-1 (Fig. 3.4 C) GUS staining was also found at the nodule basis, and in line N27-2 also in the nodule parenchyma (Fig. 3.4 I1). In the nodules of lines N21-1 and N36-1 the nodule apex (probably zone I, II and zone II-III) was GUS-positive (Figs. H1 and J1). Although the nodule expression pattern was similar, these lines differed in their root GUS staining pattern: In addition to root vascular tissue, line N36-1 showed GUS staining in the elongation zone of main and lateral roots (Fig. 3.4 J2), whereas line N21-1 did not (Fig. 3.4 H2). Line L6-1 showed clear GUS staining in the root vascular tissue (Fig. 3.4 D2) and line N27-2 showed GUS staining in the vascular tissue and the elongation

zone (Fig. 3.4 I2). In lines A7-1 and M17-1, the GUS staining was predominant at the tips of the (emerging) secondary roots (Figs. 3.4 B2 and F). The other lines from this group showed very weak GUS staining in the root vascular tissue (data not shown). Lines A7-1 and B9-1 will be discussed in more detail in chapters 5 and 6, respectively.

#### Plant lines with GUS staining in other tissues than nodule vascular bundles

From this group, line L8-1 showed the most apical GUS staining pattern: only a few cell layers of the nodule cortex or the nodule meristem were GUS-positive (Figs. 3.4 K1 and K2). Line N4-1 showed strong GUS staining in the nodule region corresponding to the meristematic region (Fig. 3.4 L1) and lines F2-6 and M12-2 in the nodule region corresponding to zone I and the prefixation zone II (Figs. 3.4 M1 and N1). Lines N14-7 and N27-9 both showed GUS staining in the peripheral nodule tissues surrounding the central tissue (Figs. 3.4 O1 and P1). Transversal and longitudinal sections of nodules from line F8-1 revealed that the GUS staining was localized at a very narrow zone adjacent to the infected cells (Figs. 3.4 Q1 and Q2). In addition, the beginning of the prefixation zone II showed GUS staining (Fig. 3.4 Q2). Line N27-5 showed GUS staining in the epidermis and parenchyma around the fixation zone in the nodules (Figs. 3.4 R1 and R2). In lines F2-6 and N4-1 the root primordia were GUS stained (shown for line F2-6 in Fig. 3.4 M2). Line M12-2 showed GUS staining in the elongation zone of (side) roots (Fig. 3.4 N2) and line N14-7 showed strong GUS staining in the root vascular bundle (Fig. 3.4 O2).

### 3.2.5 Organ specificity of the GUS staining

In order to test whether the expression of the *GUS* fusions was root/nodule specific, GUS staining was performed in seeds, unifoliate leaves, cotyledons, leaves, stems, flowers and seed pods from 18 independent GUS-positive lines. All lines, including the wild type, showed a weak and overall GUS staining of the seeds and seedpods (result not shown), suggesting a background activity of *GUS* in these organs. However, lines A7-1 and B9-1 showed a GUS staining pattern in the vascular tissues of the seedpods that differed clearly from the background (Figs. 3.5 A6 and B4). Nine lines (line F8-1, M3-3, M5-1, M9-8, M17-1, N21-1, N27-8, N27-9, and N36-1) did not show any GUS staining in the tested aerial parts, whereas the other lines (line A7-1, B9-1, L6-1, L8-1, M12-2, N4-1, N14-7, N27-2, and N27-5) did (Table 3.3 and for some examples see Fig. 3.5).

Table 3.3 GUS staining in aerial parts of tagged *M. truncatula* plants showing GUS staining in the root system.

Line	cotyledon	unifoliate	leaf	stem	flower
A7-1	+	+1,2	+	+	+
B9-1	+	+1,2	+	+	+
F8-1	-	-	-	-	nt
L6-1	+	+2	-	+	-
L8-1	+	+	-	-	-
M3-3	-	-	-	-	nt
M5-1	-	-	-	-	nt
M9-8	-	-	-	-	nt
M12-2	+	+2	-	+	+
M17-1	-	-	-	-	-
N4-1	-	+2	-	-	-
N14-7	-	-	-	-	+
N21-1	nt	nt	-	nt	nt
N27-2	+	+1,2	-	-	-
N27-5	+	+1,2	-	-	nt
N27-8	-	-	-	-	-
N27-9	-	-	-	-	-
N36-1	-	-	-	-	-

1 = GUS staining in the cotyledon and/or unifoliate leaf.

2 = GUS staining in the petiole.

### 3.2.6 Screening for mutant nodule phenotypes

In the GUS-negative plants, the T-DNA insertion might have caused a mutant nodulation phenotype as the result of disruption of an indispensable nodulation gene. Since the primary transformants are hemizygous for the T-DNA, in 25% of their progeny both alleles should be tagged. Therefore, after the Basta-selection (§3.2.2) progeny of all the 187 Basta-resistant lines, with or without *GUS* expression in the roots and/or nodules were inoculated with *S. meliloti* and, after ca. 4 weeks tested for mutant nodulation phenotypes by a visual screen. For 55 lines between 12 and 50 Basta-resistant plants (Table 3.2) were tested and for the other lines less than 12

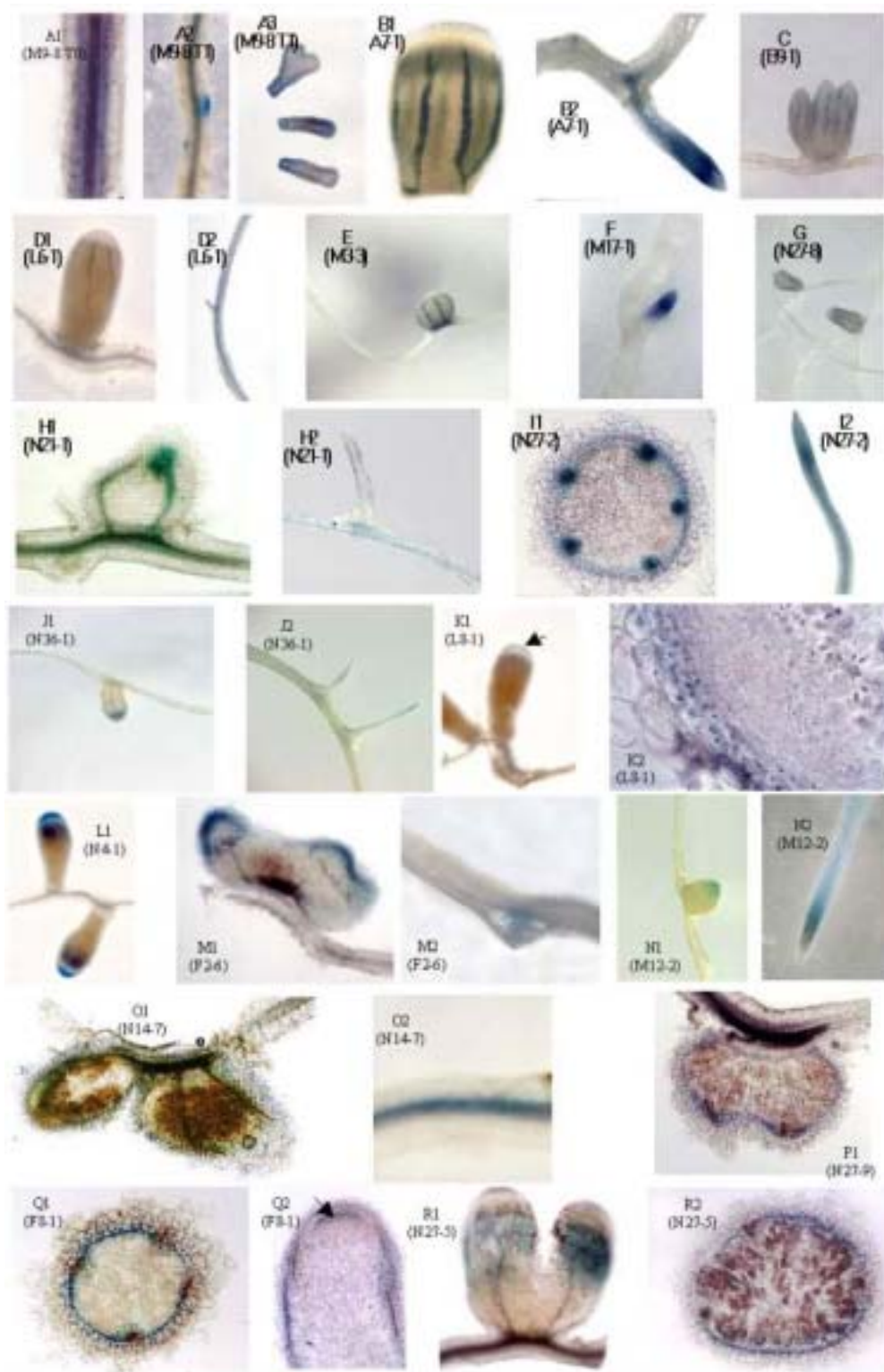


Figure 3.4 GUS staining in the nodules and roots of *M. truncatula* plants transformed with the pGKB5 T-DNA. **A1**: root of primary M9-8 transformant. **A2** and **A3**: root and nodules of T1 plant of line M9-8. **B1** and **B2**: nodule and side root of line A7-1. **C**: B9-1 nodule. **D1** and **D2**: nodule and root of line L6-1. **E**: nodule of line M3-3. **F**: side root of line M17-1. **G**: nodule of line N27-8. **H1** and **H2**: longitudinal nodule section and root of line N21-1. **I1** and **I2**: transversal nodule section and root of line N27-2. **J1** and **J2**: nodule and root of line N36-1. **K1** and **K2**: nodule and longitudinal section of a nodule tip of line L8-1 (arrow points at GUS staining). **L1**: nodules of line N4-1. **M1** and **M2**: longitudinal nodule section and root of line F2-6. **N1** and **N2**: nodule and root tip of line M12-2. **O1** and **O2**: longitudinal section of nodules and entire root of line N14-7 (arrow points at GUS staining). **P1**: longitudinal section of nodules of line N27-9. **Q1** and **Q2**: transversal and longitudinal nodule section of line F8-1. **R1** and **R2**: longitudinal and transversal nodule section of line N27-5.



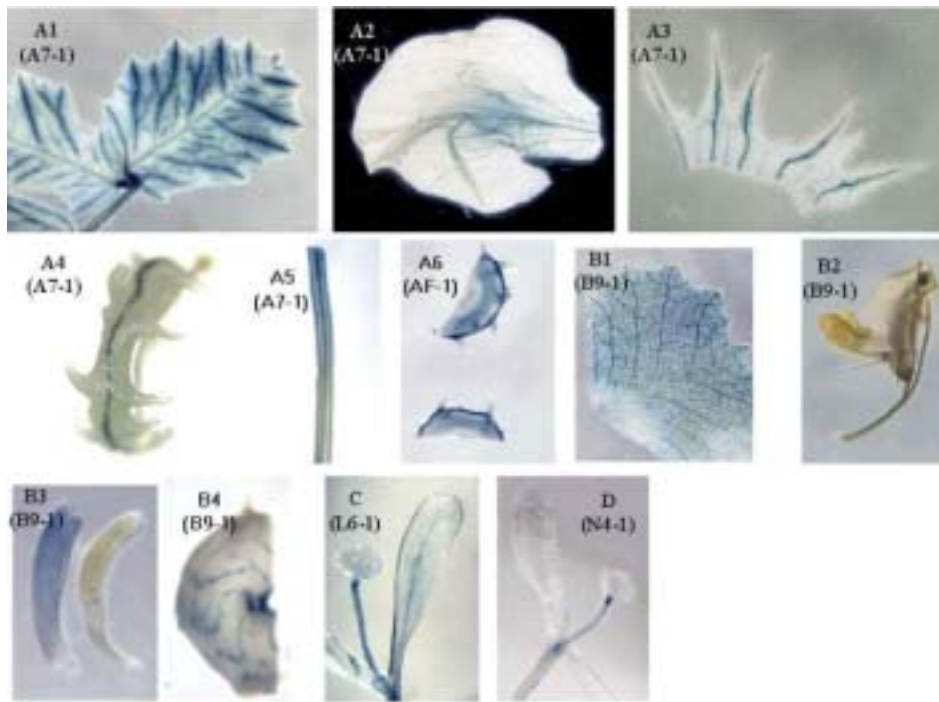


Figure 3.5 GUS staining in aerial parts of transgenic *M. truncatula* plants. **A1, A2, A3, A4, A5, A6:** Leaf, petal, sepal, seed pod, stem, and seed pod turns of line A7-1. **B1, B2, B3, and B4:** Leaf, flower, stamen filaments and ovary, and seed pod turn of line B9-1. **C and D:** unifoliar stage plantlets of lines L6-1 and N4-1.

Basta-resistant plants were tested. However, all progeny plants from each tested line formed nodules with a similar overall morphology as wild-type nodules, indicating that in none of the lines a gene has been knocked-out that is essential for a step in nodulation.

However, 1 line showed a dwarf phenotype (line A7-1). This line will be discussed in more detail in chapter 6.

### 3.3 Discussion

#### 3.3.1 GUS is the reporter of choice

The GFP protein is a very attractive reporter for the creation of a T-DNA tagged mutant collection. The reason for this is that GFP can be detected *in vivo* without the need of addition of a substrate which allows the quick screening of large numbers of plants.

Nodules of wild-type *M. truncatula* plants appeared to have a high level of autofluorescence. Therefore, it was determined whether the GFP could be detected in *M. truncatula* when expressed at a high level or at a low level but in a localized manner. Indeed, hairy roots that constitutively expressed the GFP variant *S65T-intron* (Pang *et al.*, 1996) and roots and nodules of stably transformed plants constitutively expressing mGFP5-ER (Haseloff *et al.*, unpublished) showed bright fluorescence that could be distinguished from autofluorescence. This indicates that high amounts of GFP are not toxic for *M. truncatula* and that GFP can be detected in roots and nodules (despite the high level of autofluorescence observed in the nodules).

However, nodules of wild-type plants and plants carrying *MsENOD12A* promoter::*GFP* fusions could not be distinguished from each other. In contrast, *M. truncatula* plants carrying an *MsENOD12A* promoter::*GUS* fusion did show GUS staining at the distal end of the nodule as previously described by Bauer *et al.* (1997). The minimal promoter region used for the *MsENOD12A* promoter::*GFP* fusions contained the 300 bp region upstream of the transcription start of *MsENOD12A*, whereas the *MsENOD12A* promoter::*GUS* fusion contained 3.7 kb of the *MsENOD12A* promoter region. It was found that the 200 bp region present immediately upstream of the transcription initiation site of the pea *ENOD12A* was sufficient for nodule specific expression in transgenic *Vicia hirsuta* nodules (Vijn *et al.*, 1995). Therefore it is not probable that the lack of visible GFP expression in nodules is caused by the use of a smaller promoter region and it is more likely that the expression of the *MsENOD12A* promoter::*GFP* fusion has been masked by the autofluorescence of the nodules.

This autofluorescence might mask low level expressed *GFP*-gene fusions, and it was decided that instead of *GFP*, *GUS* would be used as reporter gene in the tagging project of *M. truncatula*. For this the binary vector pGKB5 was used.

### 3.3.2 Transformation and regeneration of *M. truncatula* with pGKB5 T-DNA

The presence of chlorophyll seemed to be important for effective selection with Basta. The target of its active compound phosphinothricin (PTT) is the enzyme glutamine synthase that converts ammonium to glutamine. It was hypothesized that the lack of glutamine could result in the development of chlorosis, followed by plant death (Perez-Garcia *et al.*, 1998). The absence of chlorophyll in the dark-grown regenerating *M. truncatula* cells may explain the relatively high number of escapes to Basta in comparison to the embryo's that were cultured in the light (and thus photosynthesized). Therefore, for future tagging projects it may be recommended to maintain the regenerating tissues for at least two weeks on Basta-containing medium upon transfer to the light. In addition, in order to avoid time and space investments for lines that escaped the Basta selection during the regeneration, it may be recommended to spray the primary transformants with Basta after their transfer to the green house.

### 3.3.3 GUS staining of the T-DNA tagged lines

Nineteen (=10%) of 187 transformed plant lines showed GUS staining. Only nodulated roots were systematically screened for GUS activity and it is probable that more GUS positive lines will be found when also other organs or the effect of for example chemicals, hormones, mycorrhizal colonization or temperature on the GUS activity would be tested. In other T-DNA tagging experiments different frequencies of gene fusions to a promoter-less reporter or antibiotic gene were found: for example 35% - 50% in tobacco micro calli (Herman *et al.*, 1990), 30% in transformed and regenerated *Arabidopsis* (Koncz *et al.*, 1989), 1.6% - 2.1 % in transformed and regenerated rice (Jeon *et al.*, 2000) and 75% in hairy roots of *Lotus japonicus* (Martirani *et al.*, 1999). Mollier *et al.* (1995) who used the pGKB5 for *in planta* transformation of *Arabidopsis* found that 2.4% of the plants showed GUS staining in the roots. That is 5 times less than the frequency that we found by using the same vector in transformed and regenerated *M. truncatula* (12%). The *Arabidopsis* genome contains about 1 gene in every 5 kb (The Arabidopsis Genome Initiative, 2000), while it is estimated that the gene-rich regions of *M. truncatula* contain approximately 1 gene in every 10 kb (D. Cook, personal communication). Therefore it is surprising that the percentage of GUS-positive lines is higher in *Medicago* than in *Arabidopsis*. The reason for this is unclear.

Lines showing different GUS staining patterns in the nodules were already obtained in this small tagged population. This indicates, that in *M. truncatula* a T-DNA with a promoter-less *GUS* gene can lead to gene fusions and is thus a valuable tool to search for genes involved in nodulation or for the creation of new markers to study symbiosis.

More than half of the GUS-positive lines showed GUS staining in the vascular tissue of the nodules and/or roots. However, the data presented in this chapter do not make it possible to conclude whether this was due to an artifact or whether indeed many different genes are expressed in the vascular bundle (see chapter 4). In addition, it is very well possible that the promoters of the genes responsible for the observed GUS staining do not have a nodule or root-specific function, since several lines also showed GUS staining in the aerial parts of the plant.

### 3.3.4 Search for phenotypes

In *Arabidopsis* about 4% of T-DNA tagged lines showed a mutant phenotype, and in about 25% of these mutants the mutation was found to be linked to the presence of T-DNA (D. Bouchez, personal communication). Thus only 1% of the T-DNA tagged *Arabidopsis* collection showed a mutant phenotype caused by the T-DNA insertion. If the percentage of mutants found in T-DNA tagged *M. truncatula* would be comparable and if one considers that only a small fraction of the *M. truncatula* genome is specifically required for nodule development, it is not surprising that none of the 187 tested T-DNA tagged *M. truncatula* lines showed a mutant nodulation phenotype.

The primary transformants produced relatively few seeds, probably due to the fact that they were exposed twice to considerable stress: first the transfer from the *in vitro* conditions to the greenhouse and second the nitrogen starvation required for the nodulation experiment. Due to the low seed production, for only 54 primary transformants more than twenty progeny plants could be tested for a nodulation phenotype. Therefore, nodulation phenotypes may have been overlooked due to the small numbers of progeny plants that were tested for the majority of the lines. For future tagging projects it may be recommended to grow the regenerated plants under optimal growth conditions in order to obtain a maximum seed yield and screen the plant lines for (nodule) phenotypes and GUS staining in the T1. The lines that did not produce enough T1 seeds are now being propagated in order to obtain more seeds to make screening for nodulation phenotypes possible.



## 3.4 Material and methods

### 3.4.1 Plasmids

The plant binary vector pBin/*mGFP5-ER* (Haseloff *et al.*, unpublished) was kindly provided by Jim Haseloff. In this vector the *mGFP5-ER* is under the control of the *CAMV 35S* promoter. The T-DNA contains a neomycin phosphotransferase II (*NPTII*) gene that allows selection of transformed plant cells with kanamycin. The vector also confers kanamycin resistance to bacteria.

The vector pCP60/35S::*mGFP5-ER* was constructed by cloning the BamH1-SstI fragment from pBIN/*mGFP5-ER* containing *mGFP5-ER* behind the *CAMV 35S* promoter of pCP60 (P. Ratet, unpublished). The vector pCP60/35S::*S65T-intron* was constructed by excising the *S65T-intron* as a BamH1-EcoRI fragment from the monocot expression vector pMON30049 (Pang *et al.*, 1996) and cloning it under the control of the *CAMV 35S* promoter of pCP60. The T-DNA of pCP60 contains a neomycin phosphotransferase II (*NPTII*) gene that allows selection of transformed plant cells with kanamycin. The vector confers kanamycin resistance to bacteria.

In pB12A/*ENOD12A* promoter::*S65T-intron* the *S65T-intron* was cloned as a BamH1-EcoRI fragment in pBluescript II KS (+) phagemid (Stratagene) and then cloned as a Xba1-XhoI fragment under the control of the minimal *MsENOD12A* promoter (containing a 300 bp fragment just upstream of the *MsENOD12A* start codon) in the plant binary vector pBP12A (A. Cebolla, unpublished). pBP12A is a derivative of the binary vector pGPTV-BAR (Becker *et al.*, 1992) and confers Basta-resistance to plants and kanamycin-resistance to bacteria.

pB12A/*ENOD12A* promoter::*mGFP5-ER* was constructed by cloning a Xba1-SstI fragment containing *mGFP5-ER* behind a minimal *MsENOD12A* promoter in the plant binary vector pBP12A (A. Cebolla, unpublished).

The plant binary vector pBGK5 (Bouchez *et al.*, 1993; Fig. 3.3A) was kindly provided by David Bouchez. The T-DNA of this vector contains a promoter-less *GUS* gene linked to its right border and confers Basta and kanamycin resistance to transformed plants. Bacteria containing pBGK5 are kanamycin resistant.

### 3.4.2 Bacterial strains

*Escherichia coli* strain DH5 $\alpha$  (Promega) was used for the amplification of the different constructs. *E. coli* was transformed by the use of the calcium chloride method as described by Sambrook *et al.* (1989) and grown at 37°C in liquid or on solid LB medium (Ausubel *et al.*, 1989) containing the appropriate antibiotics.

*Agrobacterium tumefaciens* strain EHA105 (Hood *et al.*, 1993) was used to transform *M. truncatula*. *A. tumefaciens* was transformed with the binary constructs by electroporation as described by Mozo and Hooykaas (1991) and cultured at 30°C in liquid or on solid YEB medium (Gartland and Davey, 1995) containing the appropriate antibiotics.

*Agrobacterium rhizogenes* strain Arqua1 (Quandt *et al.*, 1993) was used to induce hairy roots of *M. truncatula*. The binary constructs pCP60/35S::*S65T-intron* and pCP60/35S::*mGFP5-ER* were transferred from *E. coli* into *A. rhizogenes* through triparental mating in the presence of the helper *E. coli* strain RK2013 (Ditta *et al.*, 1980). For this *A. rhizogenes* was grown during 2 days at 30°C on solid YEB plates containing rifampicin (100 mg/l) and the three *E. coli* strains (containing the binary vectors and the helper plasmid pRK2013) were grown overnight at 37 °C on solid LB plates containing kanamycin (25 mg/l). For each of the triparental matings an aliquot of the *A. rhizogenes* and the helper strain was removed from the solid plates and mixed on a solid YEB plate together with one of the two *E. coli* strains containing the binary vectors pCP60/35S::*S65T-intron* and pCP60/35S::*mGFP5-ER*. After over night growth at 30°C transformed *A. rhizogenes* was selected on solid YEB plates containing rifampicin (100 mg/l) and kanamycin (100 mg/l).

*Sinorhizobium meliloti* strain Rm41 (Kondorosi *et al.*, 1973), forming Fix<sup>+</sup> nodules on *M. truncatula*, was used for nodulation experiments. This strain was cultured at 30°C in liquid or on solid TA medium (10g/l bacto tryptone, 1 g/l bacto yeast extract, 5 g/l NaCl, 1mM CaCl<sub>2</sub>, 1 mM MgSO<sub>4</sub>).

### 3.4.3 Plant material and plant growth conditions

*Medicago truncatula* (Gaertn.) line R108-1 (c3), which had been selected for regeneration and transformation competence (Hoffmann *et al.*, 1997) was used for all the described experiments. *M. truncatula* seeds were scarified with sandpaper and surface sterilized with Barychlor (3.5 g/ 400ml) (Bayrol GMBH) during gentle agitation for 20'. After three washes with sterile water the seeds were transferred to petri-dishes containing humidified sterile Whatman paper. In order to synchronize the germination the seeds were incubated for 2-3 days at 4°C. Then the seeds were kept over night at RT in the dark and the seedlings were planted in Perlite (Puteau, France) or transferred to solid modified MS medium (0.5 X MS 'salt and vitamins' (Sigma); 10 g/l saccharose; 9

g/L kalys 575 agar; Mayoly Spindler, France).

In the greenhouse *Medicago truncatula* was grown at 50-55% relative humidity under a 16h photoperiod at 25 °C and 15 °C during day and night respectively. During *in vitro* culturing *M. truncatula* was placed in a plant culture room (24 °C, 12 h photoperiod).

### 3.4.4 Production of hairy roots of *M. truncatula*

Ten seedlings were placed on 12 cm square plates containing solid modified MS medium which were then incubated vertically in a 25°C growth cabinet in the dark. After 24 hours the plantlets were transferred to a growth chamber with a 16 h photoperiod at 25°C. After three days the first leaves appeared and the seedlings were infected with *A. rhizogenes* containing the binary vector of choice by scarifying the hypocotyl (approximately 1/3 distance from the hook of the primary root to the cotyledons) gently with a needle dipped in plate grown *A. rhizogenes* culture. Plates containing infected plantlets were returned to the growth chamber in the conditions previously described.

Between 2-3 weeks after this procedure, the first hairy roots were obtained. Once these roots were at least 2 cm long, the main root was excised and the resulting composite plants were transferred to fresh Gibson. The composite plants were grown in the conditions described above keeping the roots in the dark. One week later, hairy roots with a length of 4 cm were checked for GFP expression by epifluorescence microscopy

### 3.4.5 Stable transformation and regeneration of *M. truncatula* plants

The transformation and *in vitro* regeneration of transformed *M. truncatula* plants was performed according to Trinh *et al.* (1998).

For each transformation ca. 30 young leaves from 6 to 7 week-old plants were rinsed in about 50 ml sterilized tap water containing 2 to 3 droplets of Teepol for 15 minutes and then surface sterilized with commercial bleach diluted with sterilized tap water (1:1) to a final sodium hypochlorite concentration of 6% for 15 minutes, then washed three times with sterile distilled water and cultivated in the dark at 25 °C on plastic-foil closed petri-dishes containing SH3a medium over night.

100 ml of YEP medium containing kanamycin (100 mg/l) was inoculated with 1 ml of an overnight *A. tumefaciens* starter culture and grown overnight until a culture OD600 of ca. 0.3 was reached. Then the cells were spun down and resuspended in SH3a medium.

To optimize the access of the *A. tumefaciens* to the plant cells, the overnight cultured *M. truncatula* leaves were placed in a petri-dish lid containing a part of the *A. tumefaciens* suspension and the leaf borders were cut off with a scalpel. The cut leaves were put together with 100 ml of the cell suspension in a vacuum flask connected to a water pump and kept under vacuum (760 mm Hg) for 10 minutes. After the leaves had been shaken gently in the bacterial suspension at room temperature for 1 hour, they were removed from the suspension and placed per 5 - 6 on petri-dishes containing solid SH3a supplemented with acetosyringone (0.035 µM) to switch on the *vir*-genes of the *A. tumefaciens*. The petri-dishes were closed with plastic foil and put in the dark at 24 °C for 3 days.

After 3 days co-cultivation of the leaf explants with *A. tumefaciens*, they were placed on petri-dishes containing solid SH3a medium, augmentin (600 mg/l) to suppress bacterial growth, and Basta (6 mg/l, Rhône-Poulenc) or kanamycin (40 mg/l) for selection of transgenic tissue. The plates were incubated in the dark and the explants were transferred to fresh SH3a plates containing the appropriate selective agent in every two weeks. After ca. 40 days somatic pro-embryo's had formed and the calli were transferred to solid SH9 and placed in a plant culture room (24 °C, 12 h photoperiod). Except for the first two transformation experiments with pBGK5, for the further transformations the SH9 was supplemented with the appropriate selective agents (Basta or kanamycin). After ca. 2 weeks the somatic pro-embryo's became green and the green embryo's were transferred to SH9 without selective agent.

When the plantlets had both root as well as shoot, they were planted in sand soaked with sterilized tap water and transferred to the green house. After ca. one week being covered with a plastic transparent lid to keep humidity high, the plantlets were used for the various experiments.

### 3.4.6 GFP detection

Hairy roots were examined using a Polyvar microscope with two types of filters giving an excitation spectrum between 450 and 495 nm (B1) or between 475 and 495 nm (B4) and a stop filter at 520 nm (B1) or between 520 to 560 nm (B4). For the examination of roots and nodules a fluorescence stereomicroscope (Leica MZFLIII) with the GFP3 filter with an excitation spectrum of 470/40 nm and a stop filter between 525 to 550 nm was used.

### 3.4.7 GUS enzymatic assays

For histological detection of the GUS activity transformed *M. truncatula* tissues or 100- $\mu$ m tissue sections were stained overnight at 37 °C in 2mM X-gluc according to Trinh *et al.* (1998). Green tissues were cleared before observation by incubation in 50% ethanol.

### 3.4.8 Basta selection and segregation analysis of Basta resistance

For Basta selection of plants, germinated seedling were planted in Perlite (Puteau, France) and sprayed with 120 mg/l Basta (Rhône-Poulenc) when the first trifoliolate leaf had appeared. One week after the first treatment a second Basta treatment was carried out. Generally, two weeks after the first treatment the Basta sensitive plants were dead.

If more than 20 T1 seeds were available, the ratio of Basta-resistant: Basta-sensitive plants was scored. A  $\chi^2$  analysis was performed on the segregation ratio 3:1.  $P = 0.05$  was chosen as a critical limit, such that the predicted segregation was not rejected for  $P$  values  $>0.05$ .

### 3.4.9 Nodulation assays

For nodulation of regenerated plants, the plants were potted individually in sand in plastic disposable pots. The pots were placed in a tray in the green house (24 pots/tray) After one week watering with tap water and two weeks watering with low nitrogen solution (solution I: 0.25 mM KNO<sub>3</sub>, 0.25 mM KH<sub>2</sub>PO<sub>4</sub>, 0.62 mM CaCl<sub>2</sub>, 0.12 mM MgSO<sub>4</sub>, 0.31 mM K<sub>2</sub>SO<sub>4</sub>, 9.25  $\sigma$ M KCl, 12.2.10<sup>-2</sup>  $\sigma$ M H<sub>3</sub>BO<sub>4</sub>, 1.9  $\sigma$ M MnSO<sub>4</sub>, 0.3  $\sigma$ M ZnSO<sub>4</sub>, 0.137  $\sigma$ M CuSO<sub>4</sub>, 0.0225  $\sigma$ M (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>, 27.5  $\sigma$ M C<sub>10</sub>H<sub>12</sub>FeN<sub>2</sub>NaO<sub>8</sub>(CH<sub>2</sub>N(CN<sub>2</sub>COOO)<sub>2</sub>)<sub>2</sub>FeNa) the plants were inoculated with an overnight *S. meliloti* culture (50 ml/tray, resuspended in solution I).

For nodulation of T1 plants, seedlings were selected with Basta and watered with solution I. When the plantlets had about 4 trifoliolate leaves, they were inoculated with an overnight *S. meliloti* culture (50 ml/bac, resuspended in solution I). Nodulation was scored 3-4 weeks after inoculation.

### 3.4.10 Molecular cloning techniques and Southern hybridization

Standard DNA manipulations were according to Sambrook *et al.* (1989). Genomic DNA was extracted from leaves according to Dellaporta *et al.* (1983). After double digestion of 10  $\mu$ g of genomic DNA with EcoR1/EcoRV or EcoR1/BamH1, the digestion mixtures were run on a 0.7 % agarose gel and the DNA was transferred to a nylon membrane (Biotrans Nylon Membrane, ICN) according to the manual. The recognition sites of the used restriction enzymes are indicated in Figure 3.3. For the detection of the T-DNA a 702 bp PCR fragment (nucleotides 240 to 942 downstream of the *GUS* translation initiation site; Jefferson *et al.*, 1986) amplified using the oligonucleotides 5' GCCCAGCGTATCGTGCTGCG3' and 5'ATCGGGAGCGGCGATACCGTA3' was used. The probe was labeled with P<sup>32</sup> by random priming using the Megaprime DNA Labeling System (Amersham). The hybridization was carried out over night at 65°C in Church and Gilbert buffer (Church and Gilbert, 1984) and the membranes were washed according to Sambrook *et al.* (1989).



## 4 Isolation and sequence analysis of T-DNA flanking regions in the tagged plants

### 4.1 Introduction

As described in chapter 3, by using pGKB5 (Bouchez *et al.*, 1993) a small T-DNA tagged *M. truncatula* population was created. The presence of T-DNA was confirmed in 187 lines and 19 of them showed GUS staining in the nodules and/or the roots. This population was made to test whether large scale T-DNA insertion mutagenesis in *M. truncatula* is feasible, with the final goal to unravel mechanisms controlling the nodulation process.

One of the advantages of the use of T-DNA as a mutagen is, that it can be used as a starting point to isolate the mutated locus. For this, the flanking regions of the known T-DNA DNA sequence can be amplified either in phages or bacteria or by PCR based techniques. For example a genomic  $\phi$  phage library has been used for the isolation of T-DNA flanking regions of tobacco (Fobert *et al.*, 1994) and *Craterostigma plantagineum* (Furini *et al.*, 1997) and plasmid rescue has been used to isolate T-DNA flanking regions of *A. thaliana* (Weigel *et al.*, 2000).

Thermal Asymmetric Interlaced (TAIL-) PCR has been applied for the isolation of genomic sequences flanking T-DNA insertions in *A. thaliana* and *L. japonicus* (Smith *et al.*, 1996; Liu *et al.*, 1995; Campisi *et al.*, 1999; Schauser *et al.*, 1998), inverse PCR (IPCR) for the amplification of T-DNA flanking regions in *A. thaliana* (Lindsey *et al.*, 1993; Mathur *et al.*, 1998) and *L. japonicus* (Schauser *et al.* 1999) and the amplification of insertion mutagenised sites (AIMS) method for gene isolation in T-DNA tagged *A. thaliana* (Frey *et al.*, 1998).

The PCR walking technique has been used in *A. thaliana* by Devic *et al.* (1997). PCR walking was also applied successfully in our laboratory and therefore this method was chosen for the isolation of the T-DNA border region of our T-DNA tagged *M. truncatula* lines. For this method, genomic DNA samples of the tagged plant are digested with different blunt-end generating restriction enzymes and subsequently the generated fragments are ligated to oligo-cassettes (so called adaptors) which also have one blunt end (Fig. 4.1). Adaptor ligated DNAs constitute libraries and each library is then used for a PCR using a T-DNA-specific primer and an adaptor-specific primer. This PCR is followed by a second reaction using nested primers (Fig. 4.1).

The adaptor consists of an upper strand containing the annealing sites for both nested primers AP1 and AP2, and a shorter lower strand that lacks the AP1 site. This shorter strand contains an amine group at its 3'-end, which blocks the generation of the AP1 annealing site by the DNA polymerase during the first extension step.

Thus, the adaptor-specific primer sites are only generated when a strand complementary to the upper strand of the adaptor is synthesized by extension from a T-DNA specific primer. If, despite the presence of the amine group, PCR products are generated which contain double-stranded adaptor sequences at both ends (due to non-specific DNA synthesis), the ends of the individual DNA strands will form 'panhandle' structures following every denaturation step, due to the presence of inverted terminal repeats. These products are more stable than the template-primer hybrid and therefore will suppress exponential amplification. However, when the T-DNA specific primer extends a DNA strand through the adaptor sequence, the extension product will contain the adaptor sequence only at one end and thus cannot form the 'panhandle' structure. PCR amplification can then proceed normally.

This chapter will describe the isolation of T-DNA flanking regions of nine plants, one without and eight with *GUS* expression in nodules and/or roots as well as the results of nucleotide sequencing of the T-DNA flanking regions.

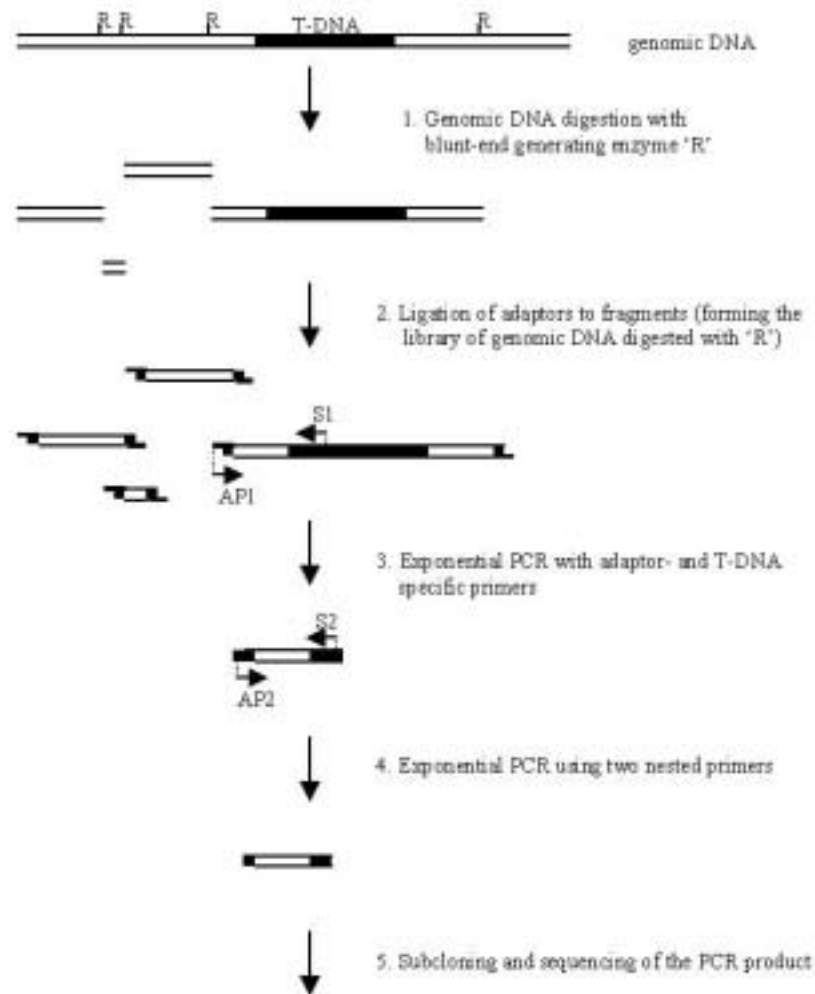


Figure 4.1 The scheme for PCR walking on T-DNA tagged genomic DNA. R = restriction site, AP1 = adaptor-specific primer 1, AP2 = nested adaptor-specific primer. S1 = T-DNA-specific primer 1. In the experiment described in this chapter S1 is the primer RBGUS1 for the amplification of T-DNA RB flanking sequence and GKB5A or LBBAR1 for the amplification of the T-DNA LB flanking sequence. S2 = nested T-DNA-specific primer. In the described experiment S2 is RBGUS2 for the amplification of the T-DNA RB and GKB5B or LBBAR2 for amplification of the T-DNA LB flanking sequence. Explanations in text.

## 4.2 Results

### 4.2.1 Amplification of T-DNA flanking regions

The T-DNA flanking regions of nine T-DNA tagged plant lines were isolated. For this, genomic DNA of each plant line was digested with EcoRV/PvuII, HpaI/ScaI, DraI, HincII or SspI. However, in some cases PCR walking on the genomic DNA libraries made from DNA digested with these enzymes did not permit the isolation of both T-DNA flanking regions (see below). Therefore, additional enzymes AatII, Eco47III, SmaI, NaeI and BbrPI were tested for lines A7-1, L8-1 and M9-8.

All the enzymes used for the digestion of genomic DNA generate blunt ends and do not cut between the left pBGKB5 T-DNA border and the sequence of GKB5A and the right T-DNA border and the sequence of RBGUS1. The only exception is the recognition site of DraI that is present between the annealing sites of GKB5A and the left T-DNA border but not between the T-DNA borders and the other annealing sites. The positions of the annealing sites of RBGUS1, RBGUS2, LBBAR1, LBBAR2, GKB5A and GKB5B are shown in Figure 4.2.



Figure 4.2 5' and 3' ends of the pGKB5 T-DNA. The PCR walk primers RBGUS1, RBGUS2, LBBAR1, LBBAR2, GKB5A and GKB5B are shown as arrows under the sequence. LB and RB are the left and right T-DNA border sequences respectively. For each transgenic plant line analyzed the last T-DNA nucleotide that was conserved during transfer is marked in red. The coding sequences of *GUS* and *BAR* are shown in blue and green, respectively.

In order to check whether the digestions were complete, an aliquot of the digestion mixtures was analyzed by gel electrophoresis. This gel analysis revealed that the digestion with the enzymes AatII, Eco47III, SmaI, NaeI and BbrPI generated mostly fragments of high molecular weight. To increase the concentration of fragments that were small enough to be amplified by PCR, the restriction products of AatII, Eco47III, SmaI and those of the restriction enzymes NaeI and BbrPI were pooled before ligation to the adaptor sequences. Also the fragments obtained by the other digestions were ligated to the adaptor sequences and in this way for each of the lines B9-1, B7-2, L6-1, M17-1, N4-1 and N27-2 five different libraries were constructed, and for the lines A7-1, L8-1 and M9-8 seven libraries were constructed.

In order to amplify the T-DNA border flanking sequences, for each flanking region two PCRs were carried out. The first PCR reaction used the library DNA as a template and one primer that annealed to the T-DNA and another primer that annealed to the adaptor sequence (such as primers S1 and AP1 in Fig. 4.1). The second PCR was carried out on the PCR products obtained with the first PCR reaction. For this second PCR the primers were nested in comparison to the first primer pair (such as primers S2 and AP2 in Fig. 4.1). The T-DNA specific primers that were used to amplify the T-DNA left border flanking regions are LBBAR1, LBBAR2, GKB5A and GKB5B. For the amplification of right T-DNA border flanking regions the T-DNA specific primers RBGUS1 and RBGUS2 were used.

After the first PCR only very faint bands could be observed (result not shown). In most cases clear bands were obtained after the second PCR (a representative gel is shown in Fig. 4.3). In cases where the obtained band was faint, the second PCR was repeated on an aliquot of the PCR product obtained after the second PCR in order to increase the amount of amplification product. Table 4.1 gives an overview of the different PCRs that were carried out. Left border fragments were obtained for transgenic lines, B7-2, L6-1, L8-1, M9-8, M17-1, N4-1 and N27-2.

Table 4.1 Overview of primer pairs used in PCR walk experiments for the amplification of the T-DNA flanking regions in different plant lines.

Library	Primers	Plant line								
		A7-1	B7-2	B9-1	L6-1	L8-1	M9-8	M17-1	N4-1	N27-2
EcoRV- PvuI	BAR/AP	-	-	-	-	-	-	-	-	-
	GKB/AP	-	nt	-	-	-	-	nt	nt	-
	GUS/AP	+	-	+	+	-	-	-	-	+
HpaI-ScaI	BAR/AP	-	-	-	-	+	-	+	+	-
	GKB/AP	-	nt	-	+	-	+	nt	nt	+
	GUS/AP	-	-	+	-	-	+	+	-	-
DraI	BAR/AP	-	+	-	-	-	-	+	+	-
	GKB/AP	nt	nt	nt	nt	nt	nt	nt	nt	nt
	GUS/AP	+	-	+	+	-	-	-	+	+
HincII	BAR/AP	-	-	-	-	+	-	+	+	-
	GKB/AP	-	nt	-	+	-	-	nt	nt	+
	GUS/AP	-	-	+	-	-	+	-	-	+
SspI	BAR/AP	-	-	-	-	-	-	-	-	-
	GKB/AP	-	nt	-	-	-	-	nt	nt	-
	GUS/AP	+	-	+	+	-	+	-	-	+
AattI- Eco47III- SmaI	BAR/AP	Nt	nt	nt	nt	nt	nt	nt	nt	nt
	GKB/AP	-	nt	nt	nt	-	+	nt	nt	nt
	GUS/AP	nt	nt	nt	nt	nt	-	nt	nt	nt
NaeI- BbrPI	BAR/AP	Nt	nt	nt	nt	nt	nt	nt	nt	nt
	GKB/AP	-	nt	nt	nt	-	-	nt	nt	nt
	GUS/AP	nt	nt	nt	nt	nt	nt	nt	nt	nt

BAR/AP = primer pairs LBBAR1/AP1 and LBBAR2/AP2, GKB/AP = primer pairs GKBA/AP1 and GKBB/AP2, GUS/AP = primer pairs GUSRB1/AP1 and GUSRB2/AP2. + = the primer pair did amplify a fragment, - the primer pair did not amplify a fragment, nt = not tested.

The sizes of the amplified borders were estimated by gel electrophoresis (subtracting the part belonging to the T-DNA) and ranged from 350 bp (line N27-2, HpaI-ScaI library) to 2000 bp (line M17-1, HpaI-ScaI library). Right T-DNA border flanking fragments were obtained for lines A7-1, B9-1, L6-1, M9-8, N4-1 and N27-2, the fragment size ranging from <200 bp (line M9-8, SspI library) to 3000 bp (line L6-1, EcoRV-PvuII library). An overview of the estimated sizes of the flanking regions is given in Table 4.2.

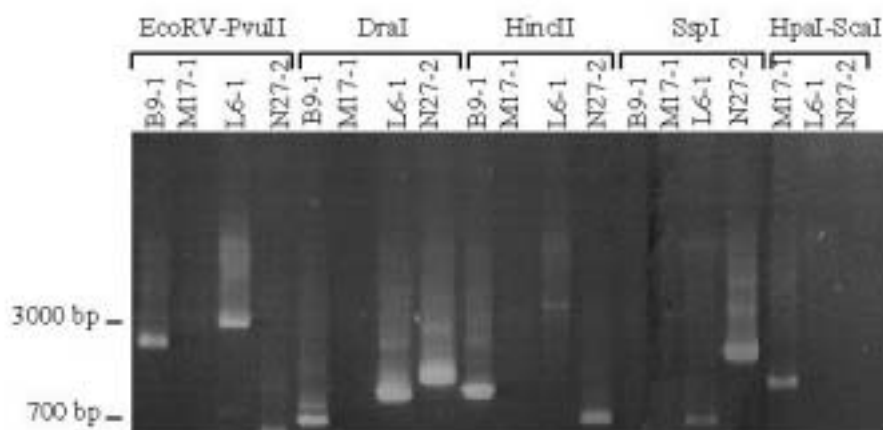


Figure 4.3 PCR amplification of the RB flanking regions with primers AP2 and RBGUS2 on EcoRV-PvuII, DraI, HincII, SspI and HpaI-ScaI libraries from transgenic lines B9-1, M17-1, L6-1 and N27-2.



#### 4.2.2 Sequence analysis of the subcloned T-DNA flanking regions

For each plant line, except M17-1, at least one amplified T-DNA flanking region was subcloned in the vector pGEMTeasy (Table 4.2, subcloned fragments are underlined). Despite several trials, the 1500 bp RB flanking region fragment of line M17-1 could not be subcloned. Sequencing of the cloned fragments was performed with primers used for the amplifications. In those cases where the obtained sequence did not cover the entire subcloned fragment, new primers were designed to obtain the complete sequences. All the fragments (with the exception of the RB fragments of the lines L6-1 and N27-2) were sequenced entirely. The isolation of the LB flanking regions of lines A7-1 and B9-1, which could not be amplified by the PCR walks using the primer pairs described above, was achieved with the help of their right border flanking sequences (see chapters 5 and 6).

Table 4.2 Size of the amplified T-DNA border flanking regions in plant lines tagged with pGKB5 T-DNA.

Plant Line	Library						RB				
	EcoRV-PvuI	HpaI-ScaI	DraI	Hinc II	SspI	AattI-mix	EcoRV-PvuI	HpaI-ScaI	DraI	Hinc II	SspI
A7-1	-	-	-	-	-	-	<u>830</u>	-	<u>630</u>	-	<u>2000</u>
	-	916	397 648	-	87	-	829 1313 1344	566	651 1301	1576 1605 1771	1986
B7-2	<u>600</u>	-	<u>500</u>	-	-	nt	-	-	-	-	-
	575	-	496	-	118	nt	-	-	-	-	-
B9-1	-	-	-	-	-	-	<u>2300</u>	<u>1300</u>	<u>700</u>	<u>1300</u>	320
	-	-	159	-	81	-	-	1147 1799	723 1175 1886 1985 2103	1147	297 456 1289 1861
L6-1	-	<u>800</u>	-	600	-	-	<u>3000</u>	-	<u>1000</u>	-	<u>700</u>
L8-1	-	-	-	-	-	-	-	-	-	-	-
	-	<u>1000</u>	-	<u>800</u>	-	-	-	-	-	-	-
M9-8	-	866	419 785	755	-	-	-	-	-	-	-
	-	1600	-	-	-	<u>450</u>	-	<u>600</u>	-	<u>600</u>	<200 500
M17-1	-	-	-	-	158	451	-	641	-	641	27 510
	-	2000	<u>1000</u>	800 2000	<u>600</u>	-	-	1500	-	-	-
N4-1	-	-	1073	-	550	-	-	-	-	-	-
	-	600	<u>600</u>	100	-	-	-	-	<u>500</u>	-	-
N27-2	-	-	609	81	293	-	-	-	515	-	-
	-	<u>350</u>	-	<u>380</u>	-	-	600	-	<u>1500</u>	800	1500
	-	399	-	350	-	-	668 1293 1463	135	1498 1370	341 873 900 1064	1486

Grey background: size of fragments estimated by gel electrophoresis, white background: expected fragment sizes on the bases of DNA sequences. AattI-mix = library made with restriction enzymes AattI, Eco47III and SmaI. Subcloned fragments are underlined.

Sequence analysis indicated that in each case half of the recognition site of the enzyme used to create the genomic DNA library was present at the junction of the adaptor sequence and the plant genomic DNA. In addition, when for the same T-DNA border flanking region fragments of different size were sequenced, the restriction sites used for the production of the smaller fragments were present in the sequence of the larger fragment(s) (Table 4.2).

In contrast, not all the predicted restriction sites led to amplification of PCR fragments. For example, although in line L8-1 the LB T-DNA flanking sequence predicts the presence of two DraI sites, no fragment was amplified on the L8-1 DraI library (Table 4.2).

### 4.2.3 Analysis of the T-DNA integration site

Southern analysis indicated that the T-DNA tagged plants used for the amplification of the flanking regions contained single T-DNA insertions (chapters 2 and 3). T-DNA borders are conserved to different extents during T-DNA integration in the plant genome (Tinland, 1996). In some cases the T-DNA integration may cause deletions of genomic DNA or chromosomal breaks (Laufs *et al.*, 1999; Nacry *et al.*, 1998). To examine whether the T-DNA borders were conserved in the *M. truncatula* tagged lines, six RB junctions and all LB junctions were sequenced (Table 4.3). The sequencing results indicated deletions ranging from 1 (line B9-1) to 202 bp (line L6-1) in the case of the T-DNA left border and deletions of the right border ranging from 0 (lines M9-8 and N4-1) to 20 bp (line N27-2) (Table 4.3 and Fig. 4.2).

Table 4.3 T-DNA border analysis of transgenic plants.

Plant line	border deletions (a)		cloned border (b)		wt locus (b)	del. (b)
	LB	RB	LB	RB		
A7-1	77	9	916	1956	3095	223
B7-2	26	nd	575	ni	ni	
B9-1	1	2	292	2256	2667	119
L6-1	202	2	123	429	ni	
L8-1	6	nd	866	ni	ni	
M9-8	24	0	451	641	1131	39
M17-1	68	nd	1073	ni	ni	
N4-1	33	0	609	515	1467	343
N27-2	68	20	399	202	ni	

a: The border deletions represent the number of bp of the T-DNA sequence lost upon integration in the plant genome. b: the size of the cloned LB, RB, PCR-amplified wild-type (wt) loci and of the deletions (del.) or the insertions (ins.) at the transgenic locus are given in bp. nd: not determined; ni: not isolated

To examine whether the T-DNA insertion had caused deletions of the genomic DNA, for four lines the corresponding wild-type locus was obtained by PCR using wild-type genomic DNA and primers based on the left and right T-DNA border flanking sequences. The sequences of the wild-type loci in these four lines indicated, that the integration of the T-DNA induced deletions of 223 (line A7-1), 119 (line B9-1), 39 (line M9-8), and 343 base pairs (line N4-1) in the transgenic plants (Table 4.3).

### 4.2.4 Search for coding sequences in the vicinity of the T-DNA

In order to determine whether the GUS activity was caused by transcriptional or translational gene fusions, the T-DNA flanking sequences were screened for open reading frames and compared to sequences in the data bases.

Despite *GUS* expression in eight lines, a gene fusion was found only for one line (line B9-1). In this plant the *GUS* gene was fused in frame with the coding sequence of a *M. truncatula* expressed sequence (TC12312; Table 4.4; Fig. 4.4 A) homologous to the *M. truncatula* nodulin gene *MtN3* (Gamas *et al.*, 1996). Comparison of the tagged locus to the wild-type locus indicated that the insertion of the T-DNA resulted in the deletion of 119 bp encoding the 39 C-terminal amino acids of the tagged protein (Table 4.3).

In line A7-1, no open reading frame could be predicted upstream of the reporter gene, but the T-DNA was inserted in opposite direction downstream of a gene corresponding to the *M. truncatula* expressed sequence TC11443 (Table 4.4; Fig. 4.4 B). Comparison of the wild-type locus sequence to the one of the tagged locus revealed that 223 bp were deleted upon T-DNA insertion in this line, resulting in the deletion of the 49 amino acids at the C-terminus from the protein encoded by this locus (Table 4.3).

In line N4-1, the T-DNA was inserted 189 bp downstream of the stop codon of a threonine synthase like gene. The sequence of the *M. truncatula* EST (AW683943; Table 4.4) corresponding to this locus indicates that the insertion had occurred 10 bp downstream of the polyA addition signal. In addition, in this line the *GUS* coding sequence is in opposite orientation to this gene, and it is therefore unlikely that the expression of the *GUS* gene would be driven by the promoter of the threonine synthase like gene (Fig. 4.4 C).

In line M9-8 the T-DNA was inserted 39 bp upstream of a gag-pol polyprotein fragment of a retrotransposon from the copia like family (most significant alignment with sequence AFO53008.1 from *Glycine max*; but in the opposite direction of the putative direction of transcription of this element (Fig. 4.4 D).

In line N27-2, the T-DNA was linked in opposite direction to another, truncated T-DNA copy, containing ca. 800 bp of the *CAMV 35S* promoter *BAR* gene fusion (Fig. 4.4 E). It is likely that expression of the *GUS* gene was controlled by the *CAMV 35S* promoter region present on the second T-DNA copy.

In the other tested lines no gene was detected in the vicinity of the T-DNA (Figs. 4.4F, G, H and I).

Table 4.4 Identification of the *M. truncatula* tagged loci.

Lines	Acc. N°	MtEST(a)	<i>A. thaliana</i> (b)	put. function (c)	Disrupted (d)
A7-1(RB)	AJ311927	TC11443	B71431	Hypothetical	yes
B7-2	AJ311928				
B9-1	AJ307887	TC12312	AB025617	hypothetical.	yes
L6-1(RB)	AJ311929				
L6-1(LB)	AJ311930				
L8-1	AJ311931				
M9-8	AJ311932			Retrotransposon (copia)	no
M17-1	AJ311933				
N4-1	AJ311934	AW683943	A96753	threonine synthase	no
N27-2(RB)	AJ311935				
N27-2(LB)	AJ311936				

(a): *M. truncatula* TC and EST sequences are from the Tigr Medicago gene index (<http://www.tigr.org/tdb/mtgi>).

(b): The accession number of the most homologous *A. thaliana* sequence.

(c): Putative functions are deduced from sequence comparison. Hypothetical indicate that the function of the gene cannot be deduced from sequence comparison.

(d): yes or no indicates whether the transcription unit of the locus was disrupted by the T-DNA insertion or not.

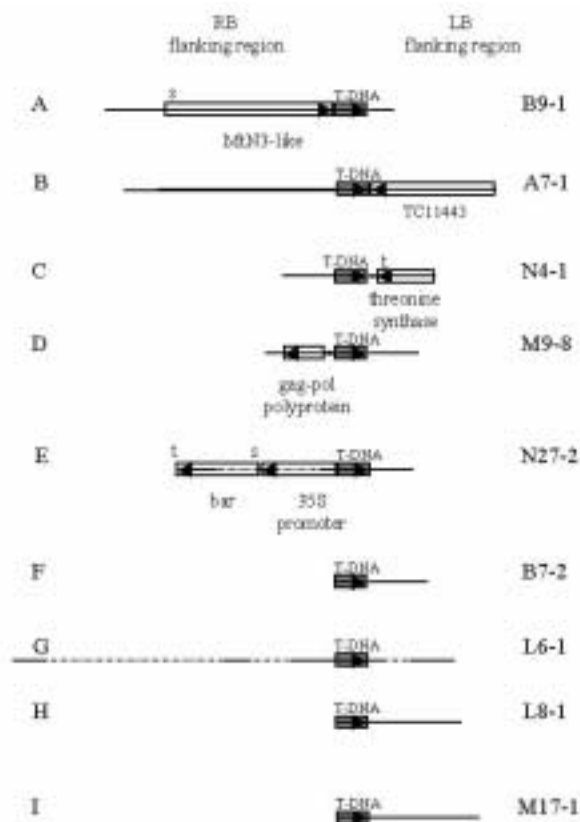


Figure 4.4 Molecular organization around the T-DNA in lines B9-1 (A), A7-1 (B), N4-1 (C), M9-8 (D), N27-2 (E), B7-2 (F), L6-1 (G), L8-1 (H) and M17-1 (I). Lines represent sequenced DNA, dotted lines represent unsequenced DNA. s: start codon, t: stop codon. Arrows indicate the direction of transcription.

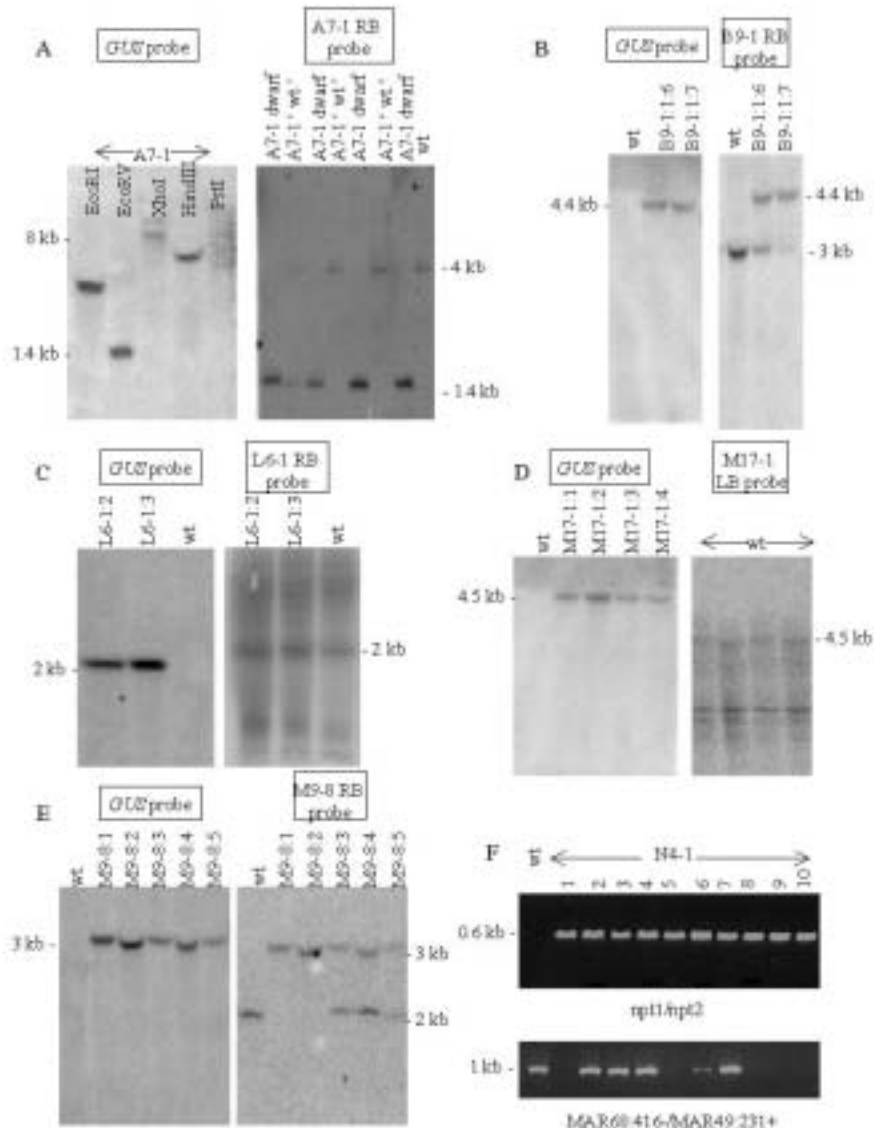


Figure 4.5. Molecular analysis of tagged plant lines. **A:** Southern blot analysis of line A7-1. For the left Southern the used restriction enzymes are marked above the lanes, for the right Southern the enzyme EcoRV was used. **B, C, D** and **E:** Southern blot analysis of lines B9-1, L6-1, M17-1 and M9-8. The DNA was digested with the enzymes BamHI and EcoRI. Hybridizations with the *GUS* probe were carried out at 65 °C, the other hybridizations at 55 °C. **F:** PCR on genomic DNA of line N4-1.

#### 4.2.5 Copy number of tagged loci

To determine whether the T-DNA of the tagged *M. truncatula* lines was inserted into unique DNA regions or repetitive sequences, the copy number of one of the T-DNA flanking regions of lines A7-1, B9-1, L6-1, M9-8, M17-1 and N4-1 was determined by Southern analysis.

Hybridizing the right border flanking regions of lines A7-1, B9-1, M9-8 (Figs 4.5 A, B, and E) and N4-1 (result not shown) resulted in a single band, and hybridizing the RB and LB flanking region of lines L6-1 and M17-1 respectively resulted in the appearance of numerous bands (Figs. 4.5 C and D). These results indicate, that in these tagged *M. truncatula* lines the T-DNA inserted preferentially into single copy loci.

#### 4.2.6 Search for homozygous insertion mutants

Plants of line A7-1 showed a dwarf phenotype. It was tested by Southern blot analysis whether the dwarf phenotype was linked to the T-DNA insertion. Genomic DNA of 7 T1 plants of line A7-1 was digested with

EcoRV and hybridized with the RB T-DNA flanking region as a probe. This revealed a 4 kb fragment in the plants with a wild-type phenotype that was absent in the dwarf plants and corresponded to the wild-type locus. A 1.4 kb fragment that corresponded probably to the tagged locus was present in all the dwarf plants and a heterozygous A7-1 plant but absent in the wild type (Fig. 4.5 A). This 1.4 kb fragment was also detected when *GUS* was used as a probe (Fig. 4.5 A). These data suggested linkage of the dwarf phenotype to the homozygous T-DNA insertion, which was confirmed by the analysis of a larger number of plants from this line (chapter 6).

Lines B9-1, L6-1, M17-1 and N4-1 did not exhibit phenotypes in spite of expression of the *GUS* gene. One explanation can be that aberrant phenotypes can only be detected in homozygous knock-out mutants. Therefore, a search for homozygous T-DNA insertion plants was carried out in the progenies of these lines.

In a preliminary experiment, for line B9-1 this was done by Southern analysis of 14 Basta-resistant T1 plants. The RB flanking region used as a probe hybridized to a 3 kb fragment in the wild-type plants and to a 3 kb and a 4.4 kb fragment in the B9-1 plants (a representative Southern hybridization is shown in Fig. 4.5 B). The 4.4 kb fragment probably corresponds to DNA that is linked to the T-DNA, since a band of the same size was found with the *GUS* probe (Fig. 4.5 B). These results suggested that line B9-1 contained only plants heterozygous for the T-DNA insertion.

Southern analysis of 5 M9-8 T1 plants using a T-DNA RB flanking region as a probe showed the existence of both homozygously and heterozygously tagged plants (Fig. 4.5 E). Again the high molecular weight fragment corresponded to the tagged locus, since this fragment had the same size as the band obtained when *GUS* is used as a probe (Fig. 4.5 E).

The numerous bands obtained in genomic DNA of line L6-1 when using the RB region as a probe made it difficult to identify homozygous insertions (Fig. 4.5 C). Although it was not determined, this is probably also true for M17-1 plants, for hybridization of the M17-1 LB probe to wild-type plants revealed numerous bands (Fig. 4.5 D).

In line N4-1 genomic DNA of 10 Basta-resistant T2 plants from the T1 plant N4-1:4 was used for a PCR approach to select homozygous insertion mutants. PCR with the primers npt1 and npt2 resulted in the amplification of the expected 1000 bp fragment of the *NPTII* gene from the T-DNA (Fig. 4.5 F). When the primers MAR68:416- and MAR49:231+ flanking the T-DNA insertion site were used, no band could be amplified in five out of ten tested N4-1:4 progeny plants (Fig. 4.5 F). This indicated that in these plants the fragment was too long to be amplified due to the homozygous T-DNA insertion.

## 4.3 Discussion

### 4.3.1 Amplification and sequence analysis of the T-DNA flanking regions

For the amplification of the flanking regions of nine tagged plants the PCR walk protocol (Siebert *et al.*, 1995; Devic *et al.*, 1997) was used.

As can be seen in Table 4.1, only in lines B7-2, L8-1, M17-1 and N4-1 the left border flanking sequence could be amplified by using the primers LBBAR1 and LBBAR2. The annealing sites of the primers LBBAR1 and LBBAR2 are located 373 and 407 bp downstream from the bar gene (Fig. 4.2). This implies that even though the plant lines are Basta-resistant, these annealing sites may be absent due to incomplete T-DNA transfer, which may have been the case for the plant lines where flanking sequences could not be amplified. In order to circumvent this problem, two new LB primers were designed (GKB5-A and GKB5-B). The primer GKB5-A anneals 18 bp outside the *BAR* gene (Fig. 4.2), thus increasing the chance that the primer annealing sites in the Basta-resistant plants are present. By using these new primers the LB T-DNA flanking region could be amplified for four additional lines (lines A7-1, L6-1, M9-8 and N27-2). Sequence analysis revealed, that line L6-1 indeed did not contain the annealing sites for LBBAR1 and LBBAR2 and that line A7-1 did not contain the annealing site for LBBAR2.

Despite several trials, the isolated and subcloned T-DNA border flanking regions of line L6-1 could not be sequenced entirely by using additional sequencing primers. Southern blot analysis showed that in this line the T-DNA was inserted in highly repetitive DNA (Fig. 4.5 C). It is therefore possible that the newly designed sequencing primers had also multiple annealing sites on the subcloned PCR fragment, which made sequencing impossible.

Although for each plant line all the restriction sites used for the construction of the libraries were present in the isolated T-DNA flanking sequences, not all the predicted restriction sites had been used to generate fragments of different sizes, probably due to incomplete digestion of the DNA.

Analysis of the junctions between the T-DNA and the plant DNA showed that maximally 20 and 202 bp were deleted from the T-DNA right and left border, respectively. In *Arabidopsis* integrated T-DNA had undergone similar deletions (Grevelding *et al.*, 1993; Tinland, 1996).

The plant lines, except B7-2, all displayed GUS staining and therefore it is not surprising that only small deletions at the RB could be detected for the *GUS* coding sequence starts 40 bp after the right border.

Comparison of the T-DNA tagged loci to the corresponding wild-type loci showed that up to 343 bp of the genomic DNA were deleted at the transgenic loci. Such rearrangements were previously described at T-DNA integration loci (Tinland, 1996). Thus, the modifications of both T-DNA ends and the rearrangements of the tagged loci indicate that integration of the T-DNA in *M. truncatula* uses the same mechanism of illegitimate recombination that was previously described for its integration in the *Arabidopsis* genome.

### 4.3.2 Can the T-DNA border regions explain the *GUS* expression?

Sequence analysis of eight tagged, *GUS* expressing plants revealed that the T-DNA inserted in the transcribed region of a gene only in lines A7-1 and B9-1. The *GUS* expression appeared to be the result of a gene fusion only in line B9-1, expressing the *GUS* reporter gene in the vascular tissues of leaves, roots, stems, flowers, seed pods and nodules (chapter 3). In this line, a translational gene fusion occurred between the *GUS* and a *MtN3*-related, so far uncharacterized gene (Gamas *et al.*, 1996).

In line A7-1, expressing the *GUS* reporter gene in the vascular tissues of leaves, roots, stem, flower and nodules (chapter 3), the T-DNA was inserted in opposite direction in the 3' end of a gene corresponding to the *M. truncatula* expressed sequence TC11443 (Table 4.4). Upstream of *GUS* no long open reading frame could be detected, although an AUG translation initiation site was found in frame with the *GUS* gene. This AUG is placed 20 codons before the *GUS* and the surrounding sequence (AAAAUGGUGA) was compatible with an efficient translation initiation in plants (Joshi *et al.*, 1997). Thus, although a gene was tagged in this line, the observed *GUS* expression pattern probably did not result from a fusion with this gene but may be the result of a small ORF-gus fusion.

In line N27-2 showing *GUS* expression in leaves, roots, nodules and cotyledons (chapter 3), the sequence flanking the right T-DNA border corresponds to the *CAMV 35S* promoter region indicating that two T-DNA copies are physically linked at the same locus in this transgenic line. Although the *CAMV 35S* promoter region is in opposite direction to the *GUS* gene, in this line the *GUS* gene expression is probably influenced by enhancer sequences from the *CAMV 35S* promoter region of the other T-DNA copy.

In line M9-8 the right border flanking sequence contained a part of the Gag-Pol protein from a copia type retrotransposon. Retrotransposon-like sequences are found in the vicinity of many plant genes and it was suggested that these sequences may influence the expression of adjacent genes (White *et al.*, 1994). Line M9-8 exhibited a variable *GUS* expression pattern (chapter 3). One may speculate that regulation of transcription by the transposon like sequence might be influenced by environmental conditions, which may have been different for each tested plant of line M9-8. Interestingly, only one copy of the retrotransposon-like fragment was present in the *M. truncatula* genome. It may be hypothesized that the retrotransposon was inactivated by a rearrangement before it could transpose further.

In the other *GUS* expressing lines, no open reading frames or promoter elements known to regulate gene expression were found in the vicinity of the *GUS* gene. The reason for *GUS* activity in these lines is unclear but one explanation can be that upon insertion of the 6.5 kb T-DNA into the genome, the genomic context may change allowing activation of cryptic promoters. This happened for example in tobacco, where a promoter-less *GUS* gene inserted next to a cryptic promoter, resulting in seed coat-specific *GUS* expression (Fobert *et al.*, 1994).

Alternatively or in addition, introduction of the *CAMV 35S* enhancer sequences present on the T-DNA might also increase the activity of cryptic promoters. This feature of *CAMV 35S* enhancers is used in the activation tagging strategy to produce gain of function mutants by inserting T-DNA carrying *CAMV 35S* enhancers in the plant genome. It was found, that in these mutants the distance between the enhancers and the overexpressed genes ranged from 380 bp to 3.6 kb (Weigel *et al.*, 2000). The distance between the *CAMV 35S* promoter and the RB of the pGKB5 T-DNA is more than 3.6 kb (ca. 4 kb) and also in our plants the enhancer sequences could have influenced the expression of the *GUS*.

An additional argument for the possible involvement of the *CAMV 35S* promoter in the *GUS* expression is the fact that a large percentage of the *GUS* expressing plants showed *GUS* activity in or around the vascular tissues, in addition to *GUS* activity in other tissue types. The presence of a *CAMV 35S* promoter was found to have a similar effect on the expression of a promoter *PsENOD12B-GUS* fusion in hairy roots of *V. hirsuta*: wild-type *PsENOD12B* is expressed at the nodule apex but in the case of a construct carrying both a *PsENOD12B-GUS* fusion and a *CAMV 35S* promoter, additional *GUS* staining was found in the stele (Hansen *et al.* 1999).

### 4.3.3 Do the analyzed lines represent knock-out mutants?

The insertion of the T-DNA in a gene may cause knock-out mutants in homozygously tagged plants. This was the case for the homozygous T-DNA insertion plants of line A7-1, which all exhibited a dwarf phenotype.

However, line B9-1 that was tagged in an *MtN3*-like gene, did not show a phenotype. Interestingly, line B9-1 contained only plants hemizygous for the T-DNA insertion, suggesting that the *MtN3*-like gene fulfills a vital function or is involved in male or female fertility. The results of a more detailed analysis of lines A7-1 and B9-1 will be described in chapters 5 and 6.

Since the intergenic regions and introns may contain regulatory sequences, also the insertion of T-DNA in such regions may cause a mutant phenotype in plants homozygous for the T-DNA insertion. Therefore, it was determined whether homozygous plants existed in lines L6-1, M9-8 and N4-1. For line L6-1 the result was not conclusive, but in lines M9-8 and N4-1 homozygous insertion mutants were found, suggesting that in these lines the T-DNA did not insert in regions that are indispensable under the used growth conditions.

#### 4.3.4 Conclusion

The T-DNA flanking regions of nine transgenic lines were isolated by the PCR walk method, demonstrating that this method can be used in *M. truncatula*. Sequence analysis of the T-DNA insertion sites showed that upon insertion in the plant genome, both the LB and RB had been subjected to moderate nucleotide deletions. Comparison of the T-DNA flanking sequences to the wild-type loci showed that small deletions had occurred upon T-DNA insertion without chromosomal rearrangements. In addition, it was found, that the majority of the tagged plants contained a single T-DNA insert. This relatively simple molecular organization at the T-DNA insertion sites is valuable for the T-DNA tagging project in *M. truncatula*, since it simplifies the isolation of the tagged loci.

Already in this small scale experiment two lines were found to be tagged in a gene and one of them showed a phenotype that may be linked to the T-DNA tag. This suggests that the T-DNA approach might be a powerful tool for gene discovery in *M. truncatula*.

However, the GUS staining observed in several plant lines was not due to obvious gene fusions or regulation by an upstream promoter. Although this may decrease the efficiency of gene discovery, the fusions are valuable as new markers to study various developmental problems such as symbiosis. In addition, regardless of GUS activity, screening of the T-DNA tagged population may result in the discovery of plant lines exhibiting, for example, nodulation phenotypes.

## 4.4 Materials and methods

### 4.4.1 Isolation of genomic DNA and Southern blot analysis

Plant genomic DNA was isolated as described by Dellaporta *et al.* (1983). Southern blot analysis was performed as described in chapter 2 using the following probes:

- € GUS probe: 702-bp PCR fragment (nucleotides 240 to 942 in the GUS ORF) amplified on pGKB5 using the primers gus1 (5'GGCCAGCGTATCGTGCTGCG3') and gus2 (5'GGTCGTGCACCATCAGCACG3').
- € A7-1 RB probe: 829 bp PCR fragment amplified on the A7-1 EcoRV-PvuII library with the primers RBGUS2 and AP2.
- € B9-1 RB probe: 1147 bp PCR fragment amplified on the B9-1 HpaI-ScaI library with the primers RBGUS2 and AP2.
- € L6-1 RB probe: 1000 bp PCR fragment amplified on the L6-1 DraI library with the primers RBGUS2 and AP2.
- € M17-1 LB probe: 1000 bp PCR fragment amplified on the M17-1 DraI library with the primers LBBAR2 and AP2.
- € M9-8 RB probe: 641 PCR fragment amplified on the M9-8 HincII library with the primers RBGUS2 and AP2.

### 4.4.2 Amplification and isolation of the T-DNA border flanking regions

T-DNA borders were isolated using a modification of the PCR walk protocol described by Siebert *et al.* (1995) and Devic *et al.* (1997).

The oligonucleotide adaptor consisted of oligonucleotide 5'CTAATACGACTCACTATAGGGCTCGAGCGGCCCGGGGAGGT3' annealed to the octamer oligonucleotide 5'ACCTCCCC3', 5'phosphorylated and 3' aminated. Oligonucleotides 5'GGATCCTAATACGACTCACTATAGGGC3' (AP1) and 5'CTATAGGGCTCGAGCGGC3' (AP2) were used in the first and second PCR reactions respectively, in combination with T-DNA specific oligonucleotides.

For the pGKB5 vector (Bouchez *et al.*, 1993; Accession number AX063413) the first and second oligonucleotides used for amplification of the right border from the *GUS* coding sequence are respectively 5'CCAGACTGAATGCCACAGGCCGTC3' (RBGUS1) and 5'TCACGGGTTGGGGTTTCTACAGGAC3' (RBGUS2). For the left border we used oligonucleotides 5'CAACCCTCAACTGGAAACGGGCCGGA3' (LBBAR1), 5'CGTGTGCCAGGTGCCACGGAATAGT3' (LBBAR2), ATCCGATGGATCCCCCGATGAGCTAA (GKB5A), and CGGCAATGTACCAGCTGAT (GKB5B).

1µg genomic DNA was digested over night by restriction enzymes DraI or SspI or HincII or EcoRV plus PvuII or ScaI plus HpaI or AatII or Eco47III or SmaI or NaeI or BbrPI in a final volume of 100 µl. The digestion mixtures of AatII, Eco47III and SmaI were pooled as well as those of NaeI and BbrPI. After digestions, DNA was purified by phenol/chloroform extraction and ligated overnight at 16°C in a 20µl final volume in the presence of the adaptor at a final concentration of 5 mM and 10 units of T4 DNA ligase. The ligation mixture was diluted 25-fold in water and 2 µl were used for the first PCR reaction (94°C: 45s; 72°C: 8mn, 7 cycles; 94°C: 45s; 67 °C: 8mn, 25 cycles). This first PCR reaction was diluted 100-fold in water and 1 µl was used for the second PCR reaction. In order to amplify fragments of large size (more than 1 kb), we used a mixture of the two enzymes EurobioTaq (Eurobio, France) and Elongase (Gibco BRL). For the first PCR reaction the reaction mix consisted of 0.3u Elongase with 0.1u EurobioTaq and for the second PCR reaction it consisted of 1u Elongase with 0.25u EurobioTaq. When the reaction yielded a single fragment of a size compatible with a T-DNA border amplification, this fragment was recloned in the pGEM-T vector (Promega) and sequenced.

#### 4.4.3 PCR

PCRs were carried out on 1 µl of plant genomic DNA (ca. 100 ng) using the Taq polymerase and buffer from Boehringer Mannheim (Mannheim, Germany). The PCR program was: 95°C: 90s, 1 cycle; 95 °C: 30s, 55 °C: 30 s 72°C: 90s, 25 cycles; 95 °C: 10 mn, 1 cycle.

For the amplification of the *NPTII* gene in N4-1 plants the primers npt1 5'GAGGCTATTCGGCTATGACTG 3' and npt2: 5'ATCGGGAGCGCGATACCGTA3' were used. For the isolation of the wild-type loci corresponding to the tagged plants the following primers were used:

##### line A7-1

A7-1RBB: 5'CACAGGGAGTCCGAGAAACA3'  
(nucleotides 1855-1874 of Acc. No. AJ311927).  
AP2: 5'CTATAGGGCTCGAGCGGC3'  
(present on adaptor sequence wild-type genomic library (see chapter 6))

##### line B9-1

MAR91:315+: 5'AATCTGCTCTTCTCTCATCAC3'  
(nucleotides 1390-1411 of Acc. No. AJ307887).  
B9-13'cDNA: 5'GTTTGACTGAGGGATACACTA3'  
(nucleotides 304-325 downstream the *MtN3-like* stop codon).

##### line M9-8

M50:344-: 5'GCTGCTGGTGCTCTTTTAT3'  
(nucleotides 362-380 of Acc. No. AJ311932).  
M114:131+: 5'CCCAAGGCAAAGAAATAACG3'  
(nucleotides 1014-1033 of Acc. No. AJ311932).

##### line N4-1

MAR49:231+: 5'CGTCAACCACAGCAGCCAATC3'  
(nucleotides 194-214 of Acc. No. AJ311934).  
MAR68:416-: 5'CTAGTGAAAGGGTTGTGGTTG3'  
(nucleotides 1212-1232 of Acc. No. AJ311934).

#### 4.4.4 Sequence analysis

Sequencing was performed with an automatic sequencer 373A (Applied Biosystems, Foster city, CA) with the dideoxy chain termination method (Pharmacia kit, Uppsala, Sweden). The primers used for sequencing were: RBGUS2, LBBAR2, GKB5B, AP2. When the fragments could not be amplified entirely with these primers additional plant line specific primers were used.

The obtained sequences were assembled and analyzed using a contig assembly program [http://www.infobiogen.fr/services/analyse/cgi-bin/cap\\_in.pl](http://www.infobiogen.fr/services/analyse/cgi-bin/cap_in.pl) (Huang, 1992) and the Genetics Computer Group Program (GCG). The sequences of the tagged loci were compared to sequences of the data base using the blast program <http://www.blast.genome.ad.jp/SIT/BLAST.html> or the Tigr Medicago gene index <http://www.tigr.org/tdb/mtgi>.



## 5 T-DNA insertion in line B9-1 disrupted the *MtN3*-like gene which encodes a putative seven transmembrane protein

### 5.1 Introduction

The aim of the T-DNA tagging project was to assess the possibilities of T-DNA tagging to discover genes that play a role in nodule formation in *M. truncatula* and to study their function. The T-DNA tagged line B9-1 seemed of particular interest, for this line showed GUS staining in the vascular tissue of the nodules. The sequence of the T-DNA right border flanking region showed an in frame fusion between *GUS* and a homologue of the nodulin *MtN3* (Gamas *et al.*, 1996) and therefore the gene that was tagged in line B9-1 was called *MtN3*-like (see chapter 4). This chapter will describe in more detail the analysis of line B9-1 in order to determine whether the *MtN3*-like gene is involved in nodulation.

### 5.2 Results

#### 5.2.1 The *MtN3*-like-*GUS* fusion is expressed in nodules but is not nodule-specific

In order to determine the expression of the *MtN3*-like-*GUS* fusion during nodulation, a time course experiment was carried out. B9-1 plants were grown under aeroponic conditions with a low concentration of nitrate (see materials and methods) and GUS staining was performed 0, 1, 2, 4, 6, 9, 11, and 16 days after inoculation with wild-type *S. meliloti*. In uninoculated plants the *MtN3*-like-*GUS* fusion was not, or very weakly, expressed in the roots. However, at 9 dpi when the nodule primordia were formed, GUS staining was observed at the nodule-root junction (Fig. 5.1 J). At 11 dpi the GUS staining was observed in the vascular bundles of the young nodules (Fig. 5.1 K) and this GUS activity was maintained in the mature, nitrogen-fixing nodules at 16 dpi (Fig. 5.1 L).

Growth in low-nitrogen nutritive solution or nodule formation have no effect on the *MtN3*-like-*GUS* expression pattern in the aerial part (see below) or the root, since the expression patterns found in plants grown under N-limiting conditions or nodulated plants was identical to that of plants watered with complete nutrient solution (results not shown).

To study whether nitrogen-fixation affects the expression pattern of *MtN3*-like-*GUS* in nodules, B9-1 plants grown in Perlite were inoculated with either of the two *Fix<sup>-</sup>* mutants of *S. meliloti*: Rm0540 and Rm8360. Both mutants were able to induce root nodule formation. Rm0540 is affected in exopolysaccharide synthesis (*Exo<sup>-</sup>*) and induces root nodule formation, but these nodules are not infected. In contrast, Rm8368 infects the nodule but is defective in early bacteroid development (*Bac<sup>-</sup>*). On roots inoculated with the *Exo<sup>-</sup>* mutant Rm0540 at 18 dpi small nodules were present and the GUS staining occurred in the root vascular tissue at the position where the empty nodule is formed (Fig. 5.1 M), comparable to the pattern observed in plants inoculated with wild-type *S. meliloti* 9 dpi. The nodules induced by the *Bac<sup>-</sup>* mutant Rm8360 showed a GUS staining pattern at the nodule basis and in the vascular bundles (Fig. 5.1 N), comparable to B9-1 plants inoculated with wild-type *S. meliloti* at 11 dpi.

To determine whether the expression of the *MtN3*-like-*GUS* fusion is nodule-specific, histological examination of different other plant organs was carried out. In seedlings strong GUS staining was detected in the hypocotyl, the cotyledons (Fig. 5.1 A) and in the junction between the hypocotyl and the root (data not shown). Leaves did not show *GUS* expression when they were still closed and young, but opened leaves showed GUS activity in the vascular tissue that increased when the leaves became older (Figs. 5.1 A, B and C). This leaf expression pattern was found in both the first formed, unifoliate leaves as well as in the trifoliate leaves. GUS

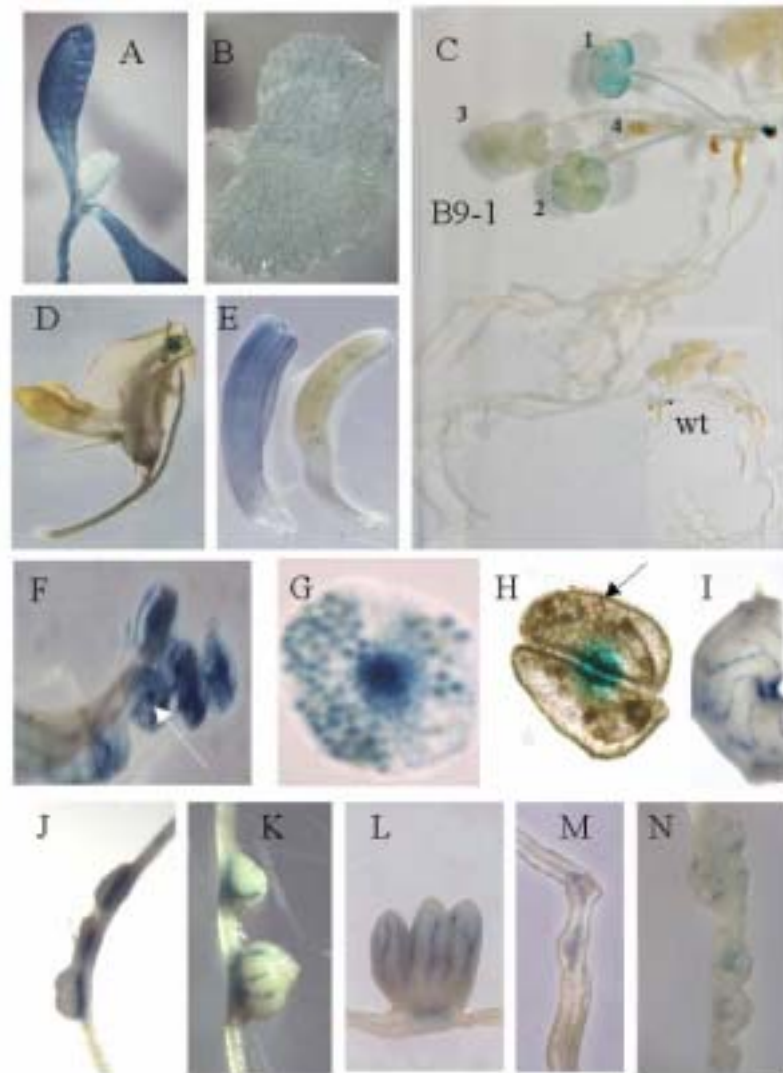


Figure 5.1 GUS staining in T-DNA tagged *M. truncatula* line B9-1. **A:** Plant at unifoliate stage. **B:** Fully expanded unifoliate leaf. **C:** GUS staining in a B9-1 plant. No GUS staining was present in a wild-type plant (bottom right). 1= unifoliate leaf, 2= oldest trifoliate leaf, 3= third trifoliate leaf, 4= youngest closed trifoliate leaf. **D:** Open flower. **E:** Anther filaments (left) and style (right). **F:** Anthers, the arrow indicates the junction of the anther to the anther filament. **G** and **H:** Anther before dehiscence containing immature pollens (**G**) and anther after dehiscence containing some left pollens (arrow). The GUS staining in the anther center corresponds to the attachment point to the anther filament. **I:** Seed pod turn. **J:** Nodules 9 dpi. **K:** Nodules 11 dpi. **L:** Nodule 18 dpi. **M:** Exo<sup>-</sup> nodule 18 dpi. **N:** Bac<sup>-</sup> nodule 18 dpi.

activity was also detected in the vascular bundles of stems and petioles, although less strong than in the leaves (Fig. 5.1 C). No GUS staining was found in the shoot apical meristem (Fig. 5.1 C). In open flowers, *GUS* was expressed in the anthers (Fig. 5.1 D), the developing seeds and in the vascular bundles of the stamen filament (Fig. 5.1 E). A more detailed analysis of the anthers showed that the GUS staining was most intense at the junction between the anther and the anther filament (Fig. 5.1 F). *GUS* expression in the pollen was transient: before dehiscence the immature pollen expressed *GUS* (Fig. 5.1 G), while after dehiscence this GUS staining disappeared (Fig. 5.1 H). In the seed pods, the GUS staining was located in the vascular bundles (Fig. 5.1 I).

These experiments showed that the *MtN3-like-GUS* fusion is expressed in the nodules, but is not nodule-specific.

## 5.2.2 Expression pattern of *MtN3-like-GUS* 'co-locates' with that of *MtN3-like*

In order to determine whether in line B9-1 the GUS activity coincides with the expression of *MtN3-like*, the GUS staining pattern was compared to hybridization signals obtained by Northern blot analysis using total RNA of wild-type plants from closed leaves, young open leaves, old leaves, apical meristems, stems, young flower

buds, flower buds, young flowers, open flowers, young seed pods, old seed pods, roots, and nodules. 480 bp of the 3' end of the shortest *MtN3-like* cDNA was used as a hybridization probe (see section 5.2.5). This probe showed only 39% identity to *MtN3* minimizing the risk of cross-hybridization to *MtN3* RNA (Gamas *et al*, 1996).

In accordance with the *GUS* expression, *MtN3-like* mRNA was detected in the old opened leaves, but not in closed or young open leaves (Fig. 5.2). The *MtN3-like* transcript was present at all stages of flower development (Fig. 5.2). The highest transcript level was found in young seedpods that still were attached to the wilted petals (Fig. 5.2). In old seed pods no mRNA was detected. For the old flowers and young seed pods the location of the mRNA was confirmed by *in situ* hybridization, showing strong hybridization signals in the vascular tissue next to the xylem vessels (data not shown). The shoot apical meristem contained *MtN3-like* mRNA which contrasts to the *GUS* staining experiment that did not show *GUS* staining in the apical meristem. In accordance with the weak *GUS* expression, only low amounts of *MtN3-like* mRNA were present in the stems (Fig. 5.2). Also the relative *MtN3-like* mRNA levels in the nodules and the roots reflected the *GUS* staining intensity: the nodules contained clearly more *MtN3-like* mRNA than the roots (Fig. 5.2). However, the absolute level of *MtN3-like* RNA in nodules was rather low. *GUS* activity was detected in the vascular bundles, representing only a small portion of the entire nodule, and since the cells of the central tissues have a high mRNA level compared to the other cells (T. Bisseling, personal communication) it is possible that the *MtN3-like* mRNA is only a relatively small fraction of the total nodule RNA sample. This can explain the relatively weak hybridization signal.

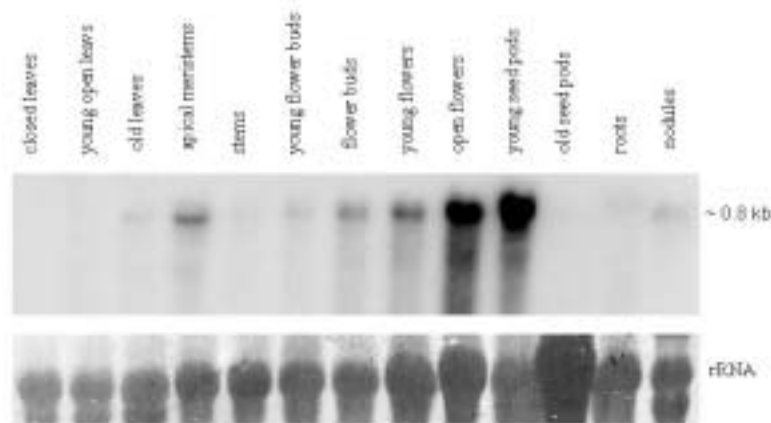


Figure 5.2 Northern blot analysis of *MtN3-like* expression in different organs. Northern blot hybridized with the *MtN3-like* probe (upper panel) and rRNA quantification by staining with methylene blue (lower panel).

The 'co-location' of the *GUS* staining and the *MtN3-like* transcripts indicates that the expression of the gene fusion in general reflects the *in vivo* *MtN3-like* gene expression.

### 5.2.3 In line B9-1 homozygous insertion mutants are rare

It was shown previously that line B9-1 contained a single T-DNA insertion and that *MtN3-like* is a single copy gene (Fig. 5.3 A and chapters 3 and 4). Therefore homozygous insertion mutants are knock-out mutants that might show an aberrant nodulation phenotype. Southern blot analysis of 14 Basta-resistant B9-1 plants showed that all these plants were heterozygous insertion mutants and surprisingly no homozygous insertion mutants were found (chapters 3 and 4). Further Southern blot analysis of BamHI-EcoRV digested genomic DNA of 42 randomly chosen plants that were descendants of a heterozygous parent revealed 29 heterozygous plants (containing both the wild-type (3 kb) and the T-DNA tagged *MtN3-like* gene (4.4 kb)) and 13 wild-type plants (Fig. 5.3 B). Like the B9-1 RB probe also the *GUS* probe hybridized to the 4.4 kb fragment, showing the linkage of *MtN3-like* to the T-DNA (chapter 4). The absence of homozygous insertion mutants is consistent with the fact that nodules of 86 T2 plants (progeny from a heterozygous insertion mutant) did not reveal any aberrant nodulation phenotype.

A further PCR-based search for homozygous insertion lines was carried out on 27 Basta-resistant plants which were the progeny of the 5 heterozygous T2 plants 14, 57, 60, 61 and 69 (Fig. 5.3 B). The presence of the T-DNA was confirmed by PCR using the primers npt1 and npt2 by which on genomic DNA a 600 bp fragment of the *NPTII* gene that is present in the T-DNA could be amplified (result not shown). In addition, PCR with the *MtN3-like* specific primer MAR91:315+ and the T-DNA specific primer RBGUS2 yielded a 900 bp fragment in each plant (Fig. 5.3 C). PCRs using the *MtN3-like* specific primers MAR91:315+ and MAR83:234- yielded in

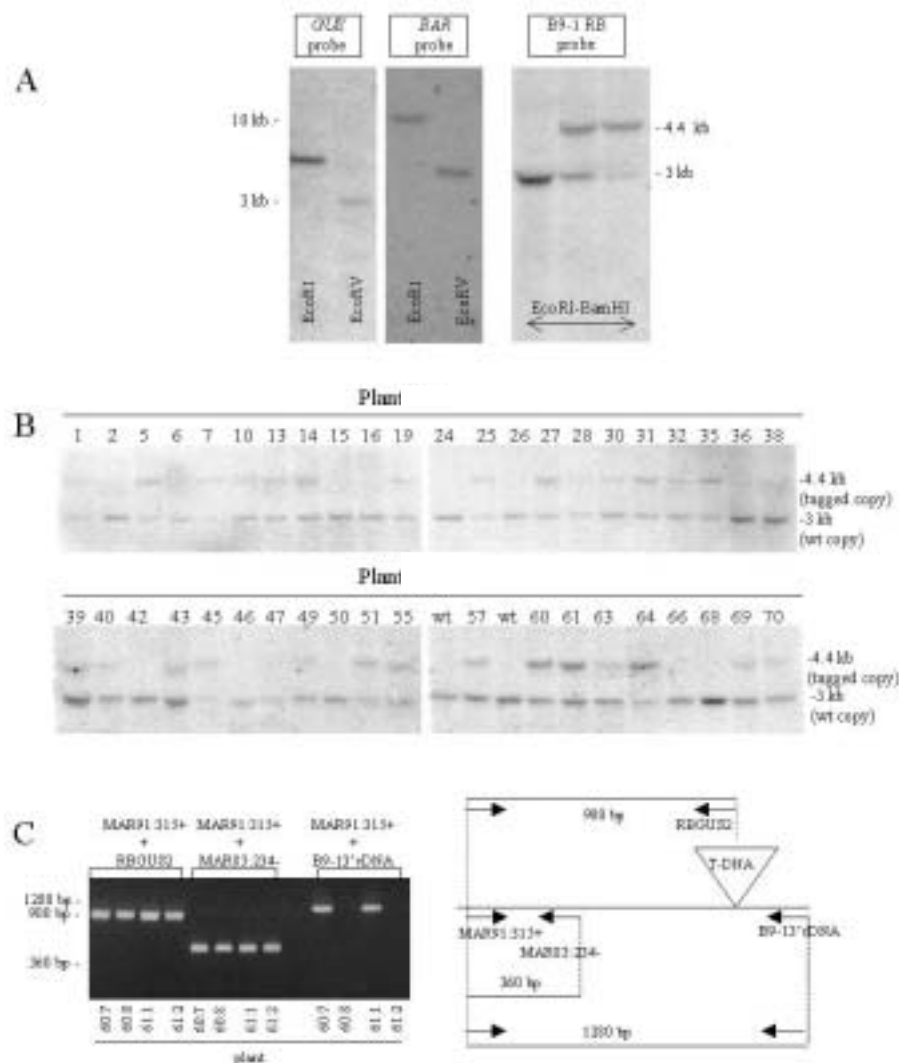


Figure 5.3 Molecular analysis of line B9-1. **A**: Southern blot analysis using the *GUS*, *BAR* or B9-1 RB probe. The restriction enzymes used for digestion of genomic DNA are marked in the figure. **B**: Southern blot analysis of *EcoRI*-*Bam*HI digested genomic DNA of 42 T2 B9-1 plants from the same heterozygous parent. The B9-1 RB T-DNA flanking region was used as a probe. **C**: Representative electrophoresis gel showing fragments obtained by PCR on genomic DNA of T3 plants of line B9-1 (left) and a schematic representation of the PCR amplifications (right).

each plant a 360 bp fragment (Fig. 5.3 C), indicating that in this region of the DNA no detectable recombinations had taken place. By using the primers MAR91:315+ and B9-13'cDNA, a 1280 bp fragment could be amplified on the untagged wild-type allele of *MtN3-like* (Fig. 5.3 C), indicating that these plants were heterozygous for the T-DNA insertion. However, this band was not obtained on plants 60:8 and 61:2, suggesting that these plants contained a T-DNA insertion in both alleles of the *MtN3-like* gene making the fragment too large to be amplified. Unfortunately, due to technical green house problems, all the plants of this series produced very few seeds of poor quality and no seeds of plants 60:8 and 61:2 were obtained. Therefore it was not possible to prove that these plants were indeed homozygous insertion lines. As described above, homozygous T-DNA tagged plants were not found or were very rare. To get a clue about the cause of the low frequency of homozygous mutants, the ratio of T-DNA containing B9-1 plants to B9-1 plants without T-DNA was determined by Southern blot analysis and GUS staining of the progeny from a heterozygous parent. Out of the 42 B9-1 plants used for the Southern blot analysis (Fig. 5.3 B) 13 (= 31%) plants did not contain the T-DNA and 29 (= 69%) did. This ratio was confirmed by the GUS staining of 175 B9-1 plants: 64 (=36%) were GUS-negative and 112 (=64%) were GUS-positive. The ratio of wild-type to T-DNA containing plants, being about 1:2, suggested embryo lethality as the cause of the low number of homozygous insertion mutants.

Although line B9-1 did not show a nodulation phenotype, the leaf form and the seed production of all the B9-1 plants (GUS-positive and GUS-negative) were different in comparison to wild-type plants. Since the leaf and seed pod phenotypes were independent of the presence of T-DNA, they are the result of an other

mutation for which line B9-1 is homozygous. To analyze the effect of the T-DNA insertion in the *MtN3-like* gene, T-DNA containing B9-1 plants should be crossed with wild-type *M. truncatula* plants to eliminate this additional mutation.

### 5.2.4 A reverse genetics approach to study the functions of *MtN3* and *MtN3-like*

We could not obtain *M. truncatula* plants homozygous for the T-DNA insertion. Therefore, the effect of a modified expression level of the *MtN3-like* gene on nodulation was studied in hairy roots of *M. truncatula* containing sense or anti-sense constructs of *MtN3-like* or *MtN3*, respectively under the control of the *CAMV 35S* promoter. These constructs are cloned in pG3.3 which contains a *GUS* gene driven by the *CAMV 35S* promoter, which makes screening for transformed roots by GUS staining possible. As a control the empty pG3.3 vector was used. One month after infection with *A. rhizogenes* the plants had formed hairy roots of about 8 cm long. From each series 10 plants were transferred to an aeroponic system and inoculated with *Sinorhizobium meliloti* strain Rm41. About three weeks after inoculation nodules with a wild type-like overall morphology had formed. The transgene expression levels were not determined at different stages of nodule development, but Northern blot analysis of total RNA from the nodulated hairy roots that were hybridized with gene specific probes showed that the hairy roots transformed with the sense and anti-sense constructs, respectively, overexpressed the transgenes (result not shown). Not all the hairy roots do have to contain the T-DNA and in order to determine if the nodules had formed on transgenic hairy roots that expressed the *GUS* (and thus likely the sense and anti-sense constructs), the nodulated hairy roots were GUS stained. Each plant contained 20-50% of GUS-positive roots that could not be distinguished from GUS-negative hairy roots. Equal numbers of nodules had formed on both GUS-negative and GUS-positive hairy roots. Macroscopically these nodules could not be distinguished from wild-type nodules and, for example, they had a reddish color which reflects the presence of leghemoglobin. With the results of these experiments the roles of *MtN3-like* or *MtN3* during nodulation did not become clear.

To study the effect of expression of *MtN3-like* and *MtN3* in a heterologous plant system, the same sense vectors and the empty pG3.3 vector used for the transformation of hairy roots were used for the transformation of *A. thaliana* by the flower dip method. The kanamycin resistant primary transformants showed all GUS staining. Northern blots of total RNA from leaves that were hybridized with gene-specific probes showed that the transgenic plants indeed overexpressed the transgenes (Fig. 5.4). In addition, Northern blot analysis showed that no detectable cross-hybridization with related *Arabidopsis* genes occurred and the probes thus appeared to be gene-specific. Although the transgenic plants expressed the transgenes, they had wild-type appearances.

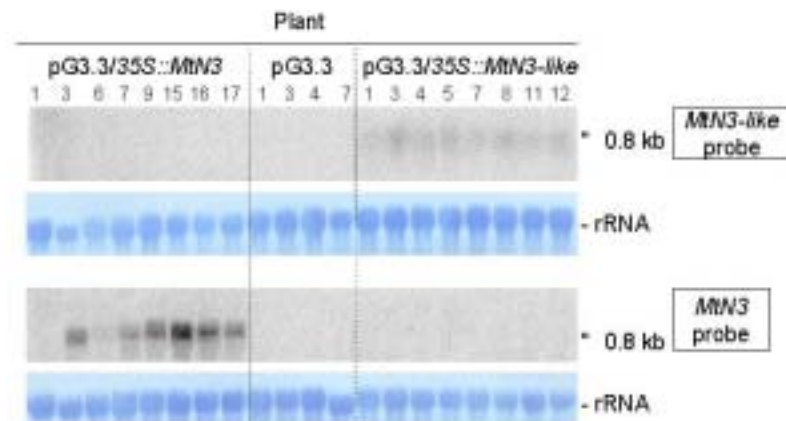


Figure 5.4 Northern blot analysis of transgenic *A. thaliana* plants overexpressing *MtN3* or *MtN3-like*. pG3.3 plants contained the empty vector. For rRNA quantification the membranes were stained with methylene blue.

Overexpression of an *MtN3-like* related gene from *Petunia*, the *NEC1* led to enlarged phloem bundles compared to wild-type *Petunia* plants (Ge *et al.*, 2000). To test if overexpression of *MtN3-like* or *MtN3* would have a similar effect on the phloem bundles of *A. thaliana*, leaf sections of progeny plants of the overexpressing primary *A. thaliana* transformants pG3.3/35S::*MtN3* 3, 7, 9 and 16 and pG3.3/35S::*MtN3-like* 1, 3, 7 and 12 were compared with wild type or plants containing the empty pG3.3. However, the leaf sections showed no obvious differences in the zone of the phloem bundles compared with those from wild types (data not shown).

### 5.2.5 *MtN3-like* has homologues in many species

As described in chapter 4, genomic DNA flanking the T-DNA RB border of line B9-1 was isolated by a PCR walk approach. Using the primers MAR91:315+ and MAR82:102+ that were designed on the putative open

reading frame of the flanking sequence, a 421 bp DNA fragment was amplified by PCR from a young nodule cDNA library of *M. truncatula* (Györgyey *et al.*, 2000). In order to obtain the full-length *MtN3-like* cDNA, this fragment was used as a probe to screen the same nodule cDNA library. This screen yielded 17 hybridizing clones. Five of the cDNA clones were sequenced and three of them contained the full-length ORF within 960, 1085 and 1312 bp-long cDNAs. The cDNA sequences were 100 percent identical to each other but they differed in the length of the 3' UTRs, indicating that this gene contains several polyadenylation sites. The deduced amino acid sequence revealed a protein of 255 amino acid residues (Fig. 5.5).

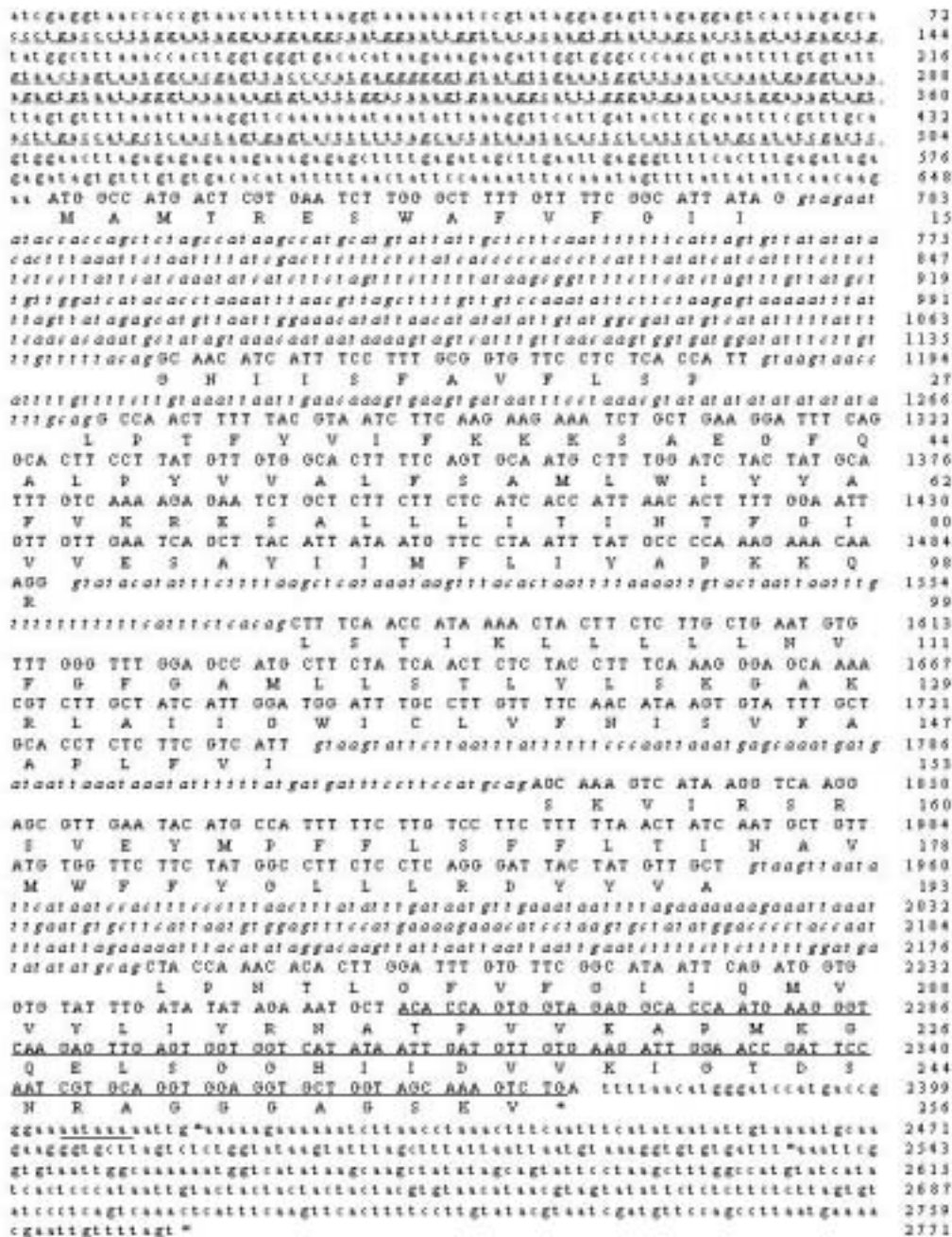


Figure 5.5 Wild-type locus of *MtN3-like*. Nucleotides that are not transcribed are underlined with a dotted line. The deduced amino acid sequence is indicated in capitals below the nucleotide sequence. Introns are shown in italics. In tagged plants of line B9-1 the underlined sequence was replaced by the T-DNA during T-DNA transfer. Double underlined sequences indicate putative polyadenylation signals (Rothnie, 1996). Stars indicate the position at which poly(A) tracks were added to the mRNA sequences.

The PCR walk method did not permit the isolation of the T-DNA LB flanking region. Instead, this was achieved by PCR on genomic DNA from line B9-1 using the *MtN3-like* specific primer B9-1 3' cDNA (designed on the cDNA) and the T-DNA LB specific primer LBBAR2, yielding an 426 bp fragment.



Comparison of the cDNA and the wild-type *MtN3-like* gene showed that the gene contains six exons, ranging in size from 37 bp (exon 2) to 214 bp (exon 3) and five introns, ranging in size from 90 (intron 3) to 451 bp (intron 1) (Fig. 5.5). Comparison of the wild-type *MtN3-like* gene and the sequences of the right and left T-DNA flanking regions of line B9-1 showed that the T-DNA inserted at 1607 bp downstream of the putative start codon of *MtN3-like* in the sixth exon and further a deletion of 119 bp was introduced by which 39 amino acids of the C-terminal end are deleted (underlined in Fig. 5.5).

In order to investigate whether other *MtN3-like* homologues are present in *M. truncatula*, Blast searches were carried out (<http://www.tigr.org/tdb/tgi.shtml> and <http://www.blast.genome.ad.jp>) using the deduced amino acid sequence of the *MtN3-like* gene. This search revealed that in *M. truncatula* at least 5 different *MtN3-like* ESTs with P values  $< e^{-10}$  were found. The deduced amino acid sequences of these ESTs varied in length between 246 (TC19246) and 268 (MtN3) amino acid residues. These proteins were homologous to each other in the region encompassing the first 220 amino acids, while the sequence of the C-terminus varied between the proteins. The protein MtN3 was the most homologous to MtN3-like and showed 56% identity to MtN3-like, while TC16865 showed only 33% identity to MtN3-like. Blast searches revealed that *N3* related ESTs were also present in other plant species and in at least 4 plant orders (Fabales, Solanales, Poales and Brassicales) full length ESTs were found. Each plant species expressed several *N3* homologues. Interestingly, the deduced amino acid sequences of genes of the *N3* family within a species are less homologous to each other than to *N3* orthologues in other species. For example the protein showing the highest homology to MtN3-like was found in *Glycine max* (TC62361, 83% identity) and not in *M. truncatula*. The most homologous *N3*-like protein of a non legume was found in *Arabidopsis thaliana* (TC118636, 57% identities). This indicates that the genes of the *N3* family are paralogues that diversified at a very early stage in evolution and may fulfill different functions. As was the case for the different *N3* family members of *M. truncatula*, also alignment of *N3* proteins of other species shows highly conserved amino acids in the N-terminal 220 amino acids, while the sequence of the C-terminus varies (Fig. 5.6 A). Expressed *N3* homologues were also found in animals such as *Drosophila melanogaster*, *Mus musculus* and *Xenopus laevis* and human. In contrast to plants these animal species seem to express only few *N3*-like homologues. In human for example, only 1 EST sequence with a P value  $< e^{-10}$  was found. The high number of *N3*-related genes thus seems to be characteristic for plants.

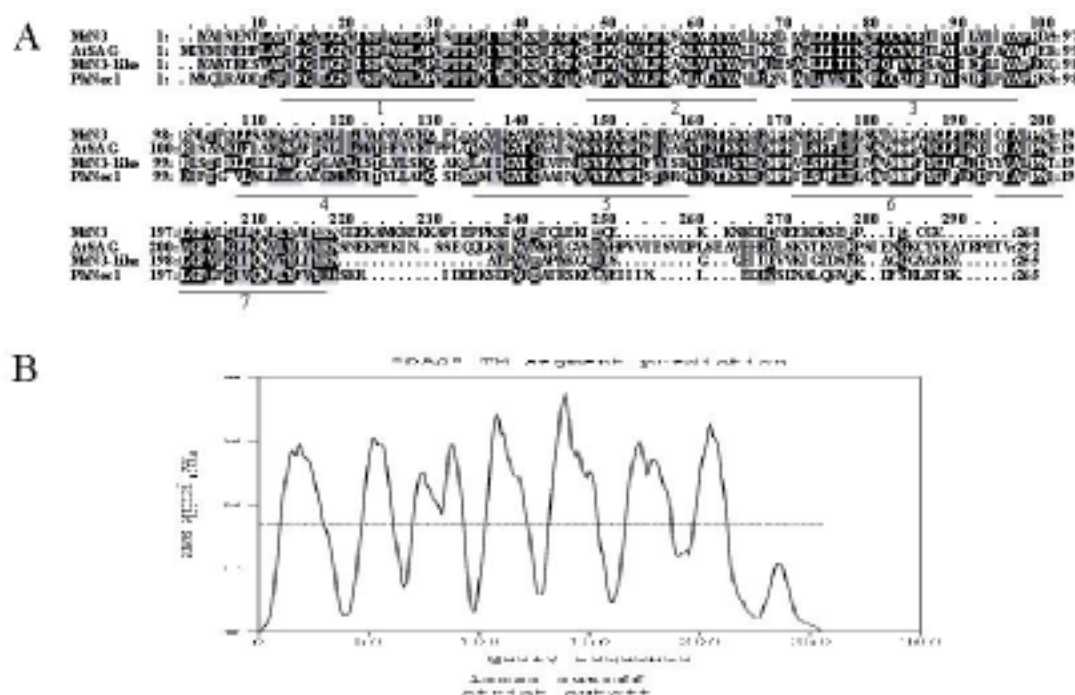


Figure 5.6 A: Alignment of the deduced amino acid sequences of the plant *N3* genes *MtN3*, *AtSAG29*, *MtN3-like* and *PhNec1*. Residues that are conserved or similar in at least three sequences are shown on a black and gray background, respectively. Hydrophobic regions are underlined. B: Hydrophobicity plot of *MtN3-like*.

### 5.2.6 Plant N3 proteins contain seven putative transmembrane regions

The hydrophobicity of the deduced MtN3-like protein was analyzed using the 'DAS' TM-segment prediction server (<http://www.sbc.su.se/~miklos/DAS>). The homologous region of the deduced amino acid sequences of the ESTs present in plants contains putative transmembrane regions (shown for MtN3-like in Fig. 5.6 B). The animal N3 genes also encode putative membrane spanning proteins, but these contain only 4-6 putative transmembrane regions (data not shown).

Sequence analysis suggests that the N3 family members are membrane proteins. To test this, the subcellular location of MtN3-like and MtN3 was determined. For this, the C-terminal terminus of *MtN3-like* and *MtN3*, respectively, was fused in frame with smGFP (soluble-modified GFP; Davis and Vierstra, 1998). These translational gene fusions were cloned behind the 35S promoter in the vector psmGFP (Davis and Vierstra, 1998). This vector was used for transient expression studies in the monolayer inner epidermis cells of onion bulb scales after particle bombardment. The GFP fluorescence was viewed after overnight incubation at room temperature. As a control the cells were bombarded with the psmGFP vector containing a 35S::smGFP fusion. In these cells the fluorescence was visible in the cytosol and the nucleus (Fig. 5.7 H) as was reported for *A. thaliana* cells bombarded with the same vector (Davis and Vierstra, 1998).

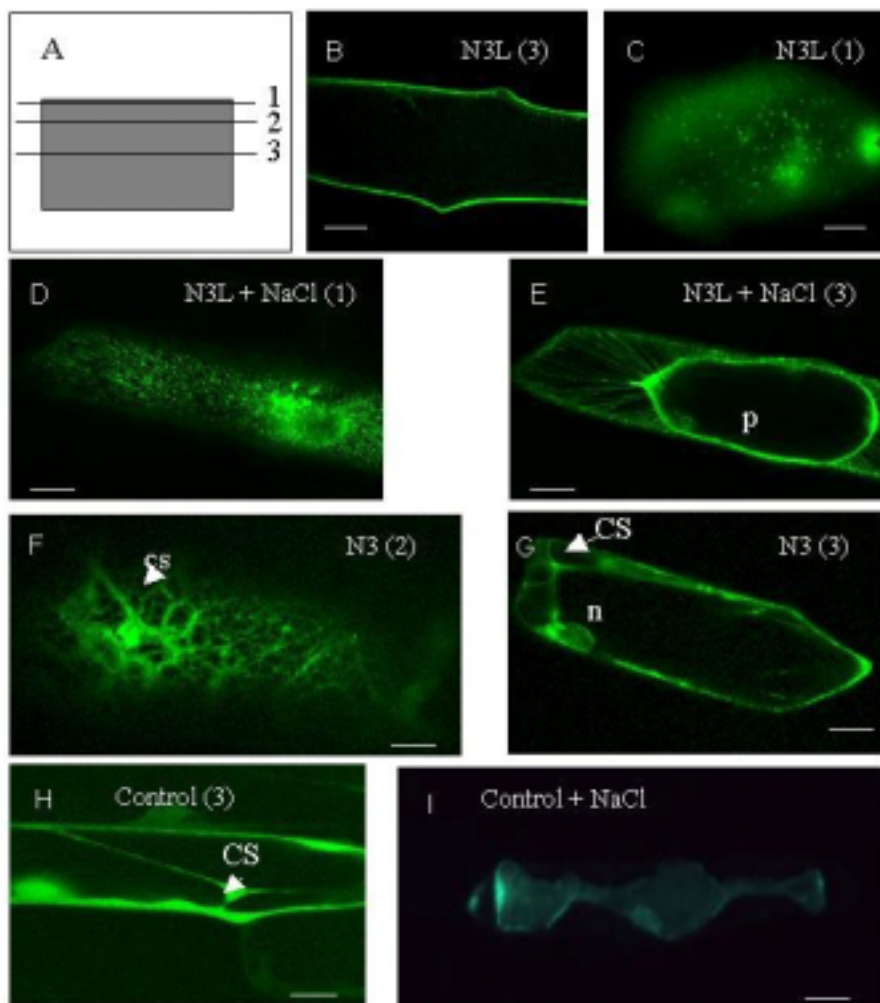


Figure 5.7 Location of MtN3-like-GFP and MtN3-GFP fusions in bombarded epidermal onion cells observed by confocal microscopy. The level is shown between brackets on each photograph. **A:** Schematic overview of an onion cell indicating the levels where the confocal images were taken. **B** and **C:** Onion cells bombarded with psmGFP/35S::MtN3-like-smGFP. **D** and **E:** Plasmolyzed onion cells bombarded with psmGFP/35S::MtN3-like-smGFP. **F** and **G:** Onion cells bombarded with psmGFP/35S::MtN3-smGFP. **H:** Onion cell bombarded with control construct psmGFP/35S::smGFP. **I:** Plasmolyzed onion cell bombarded with the control construct psmGFP/35S::smGFP. N3L = MtN3-like-smGFP protein fusion, N3 = MtN3-smGFP protein fusion, p = contracted protoplast, cs = cytoplasmic strand, n = nucleus. For the observation of the expression of the control GFP construct in a plasmolyzed onion cell a fluorescence stereomicroscope was used. Bar = 40  $\mu$ m.



Optical sections of cells bombarded with the *35S::MtN3-like-smGFP* fusion showed that GFP is present at the periphery of the cell which could correspond to the cell wall, plasmalemma or cortical cytoplasm (Fig. 5.7 B). Sections of the periphery of the cell showed that the GFP was not evenly distributed, but showed a punctuated pattern (Fig. 5.7 C). In order to determine whether the labeling was associated with the cell membrane/cortical cytoplasm or with the cell wall, the cells were plasmolyzed by incubation in 1 M NaCl for 15 to 30 minutes. After plasmolysis the GFP was located within the contracted protoplast (Figs. 5.7 D and E). This shows that the protein fusion is not located within the cell wall but is rather membrane or cytosol located. In addition, the plasmolyzed cells have fluorescent strands connecting the protoplast with the cell wall (Figs. 5.7 D and E). These strands were not visible in plasmolyzed cells expressing the control construct (Fig. 5.7 I). The fluorescent spots observed in Fig. 5.7 C may correspond to the attachment sites of these strands to the cell wall. These so called 'Hechtian strands' (Oparka *et al.*, 1994) are plasma membrane derived and therefore it is very probable that the MtN3-like-GFP protein fusion is located in the plasma membrane.

Onion cells bombarded with the *35S::MtN3 smGFP* fusion showed a fluorescent polygonal network (Fig. 5.7 F) and fluorescent cytoplasmic strands traversing the vacuole (Fig. 5.7 G). Further the nucleus is fluorescent (Fig. 5.7 G). These fluorescence patterns suggest that the MtN3-GFP fusion protein accumulates in the ER.

These results indicate that MtN3-like is located in the plasma membrane and MtN3 in the ER membrane, although the amino acid sequence of the latter does not contain a typical ER retention signal.

## 5.3 Discussion

### 5.3.1 B9-1 is the first characterized T-DNA-tagged *Medicago* line with a translational gene fusion.

This chapter described the analysis of the T-DNA tagged line B9-1. This line was selected because it showed GUS staining in nodules. However, here we showed that this expression is not nodule-specific, since this gene is also expressed in other organs. The sequence of the T-DNA flanking regions showed that in this line the T-DNA was inserted in *MtN3-like*, a homologue of the early nodulin *MtN3* (Gamas *et al.*, 1996). According to the sequence of the junction between the *MtN3-like* gene and the T-DNA RB, the GUS activity was the result of a translational fusion between *MtN3-like* and *GUS*. Previously (see chapter 4) the flanking sequence of seven other *GUS* expressing lines was obtained. However, in none of these lines open reading frames or plant promoter elements known to regulate gene expression were found in the vicinity of the *GUS* gene and the reason for the GUS activity in these lines is unclear. To determine whether the GUS activity in line B9-1 was caused by promoter activity of the *MtN3-like* promoter, the expression pattern of the *MtN3* gene was compared to that of the *MtN3-like-GUS* fusion. This comparison showed that the expression pattern of the wild-type *MtN3-like* gene and the *GUS*-tagged gene is similar. Thus B9-1 is the first characterized T-DNA tagged *Medicago* line carrying a translational fusion of *GUS* and a plant gene. T-DNA tagging using a promoter-less *GUS* gene thus can be used as a tool to discover *Medicago* (marker) genes.

### 5.3.2 MtN3-like is of vital importance for the plant

The aim of this chapter was to determine the role of MtN3-like in nodulation. However, surprisingly no homozygous insertion mutants could be obtained and therefore the phenotype of plants with a knock-out mutation of the *MtN3-like* gene could not be studied. The results described in this chapter showed that embryo lethality probably is the cause of the lack of knock-out mutants. Since the heterozygous insertion mutants do not show an obvious phenotype, this gene probably would not have been found with the use of 'classical' mutagens such as chemicals or radiation. Line B9-1 therefore nicely illustrates the advantage of the use of a promoter-less reporter gene to visualize the T-DNA insertion in genes that are vital for the plant.

The absence of mutants complicated the study of the function of MtN3-like. To study the role of MtN3-like in nodulation, hairy roots carrying sense or anti-sense *MtN3-like* constructs were made. To study the effect of over and underexpression of the nodulin MtN3 also hairy roots carrying sense or anti-sense *MtN3* constructs were obtained. However, macroscopically the nodules of the hairy roots containing the transgenes could not be distinguished from those of 'wild-type' hairy roots. It is possible that the gene silencing was not effective or that an altered transcript level did not influence the nodulation. Alternatively, or in addition, the obtained phenotypes may have been very subtle and only visible at microscopic or metabolic level. Therefore, from the described experiments no conclusion concerning the functions of MtN3-like or MtN3 in nodulation could be drawn and more detailed analysis of the nodulated hairy roots is necessary.

### 5.3.3 N3 proteins share their hydrophobicity, but probably fulfill different roles

Blast searches with the deduced amino acid sequence of MtN3-like revealed that N3-homologues are expressed in at least four plant orders and that they are probably paralogues having different functions. All the plant N3 proteins encode putative seven transmembrane proteins. MtN3-like-GFP fusion proteins indeed appeared to be located in the plasma membrane. However, the MtN3-GFP fusion protein is most likely located in the endoplasmic reticulum. This may indicate that MtN3 and MtN3-like have different roles. Another indication for the different roles of at least part of the family members of the N3 proteins are their differential expression patterns. For example, while *MtN3-like* from *Medicago* is expressed in numerous organs (this work) the transcription level of *SAG29* from *Arabidopsis* increased during *Arabidopsis* leaf senescence (Quirino *et al.*, 1999), the *Medicago MtN3* mRNA level was maximal in 4-days old nodules (Gamas *et al.*, 1996) and *NEC1* from *Petunia hybrida* was highly expressed in *Petunia* nectaries (Ge *et al.*, 2000). Alignment of the amino acid sequences of the plant N3 proteins showed that especially the C-terminal regions vary between the different proteins and it is possible that these diverse regions determine the function and/or the location of the proteins.

### 5.3.4 The hypothetical role for MtN3-like

Although *NEC1* from *Petunia hybrida* and MtN3-like are not expressed in the same organs, comparison of their expression patterns may shed light on the biological role of MtN3-like. *NEC1* was found to be highly expressed in *Petunia* nectaries actively excreting nectar (Ge *et al.*, 2000). In addition *NEC1* is expressed in the anther stomium cells and the upper part of the anther filament (Ge *et al.*, 2000). Leaves of *Petunia* that expressed the *NEC1* gene ectopically had enlarged phloem bundles compared to the wild-type plants and co-suppression of *NEC1* and a knock-out mutation of the gene resulted in an aberrant dehiscence phenotype. Based on these results the authors proposed a role for *NEC1* in the regulation of transport of sugar involved in the process of nectar secretion or sugars used to build up an osmotic pressure during anther dehiscence (Ge *et al.*, 2000; 2001).

Could MtN3-like also somehow be involved in sugar translocation? The *MtN3-like-GUS* fusion was expressed in the root vascular bundle at the site where nodule primordia were formed. When the nodule was formed, the *MtN3-like-GUS* fusion was active in the nodule vascular bundle. This expression pattern is consistent with a function of MtN3-like in the translocation of plant products (e.g. photosynthesis products) needed for nodule organogenesis and functioning. A transport function is further consistent with the *MtN3-like* expression pattern in the vascular bundles of the leaves, flowers and seed pods. The fact that in flowers the GUS staining is located at the connection between the anther and the anther filament, as was found for *NEC1* (Ge *et al.*, 2000; 2001), may suggest that MtN3-like is involved in the transport from sugars used as an osmoticum during anther dehiscence and opening, as was proposed for *NEC1* (Ge *et al.*, 2000; 2001).

Together, all the data described in this chapter are consistent with a possible role of MtN3-like in the regulation of transport of plant products. However, more detailed analysis of the *MtN3-like* gene and of the (nodulation) phenotype of plants or hairy roots over- or underexpressing the gene is necessary to clarify its exact function.

## 5.4 Materials and Methods

### 5.4.1 Plasmids

For the construction of the vectors psmGFP/35S::*MtN3-like-smGFP* and psmGFP/35S::*MtN3-smGFP*, XbaI and BamHI sites were introduced into the 5' and 3' ends of the *MtN3-like* and *MtN3* cDNA using the mutagenic primers Xba5'B9-1+ and BamH3'B9-1- for *MtN3-like* and XbaGAM79+ and BAMGAM895- for *MtN3* (for the sequences of the primers: see below). The PCR fragment were digested with XbaI and BamHI and cloned between the *CAMV 35S* promoter and the *smGFP* of the vector psmGFP (Vierstra *et al.*, 1998). The introduction of the BamHI site removed the stop codons of *MtN3-like* and *MtN3* and made *MtN3-like-smGFP* and *MtN3-smGFP* translational gene fusions possible.

The cloning steps for the construction of the binary vectors PG3.3/35S::*MtN3-like* sense, pG3.3/35S::*MtN3-like* anti-sense, pG3.3/35S::*MtN3* sense and pG3.3/35S::*MtN3* anti-sense are shown in Figure 5.8.

For the construction of the binary vector PG3.3/35S::*MtN3-like* sense the *MtN3-like* cDNA was excised from the vector pADGAL4 (Stratagene, La Jolla, CA) and cloned as an EcoRI-XmaI fragment in the vector pPR98 KS (P. Ratet, unpublished). Then the *MtN3-like* cDNA was cloned as an EcoRI-SstI fragment in the binary vector pG3.3 (P. Ratet, unpublished). pG3.3/35S::*MtN3-like* anti-sense was constructed by cloning the *MtN3-like* cDNA as an EcoRI-XbaI fragment from the vector pADGAL4 in pG3.3. pG3.3/35S::*MtN3* sense

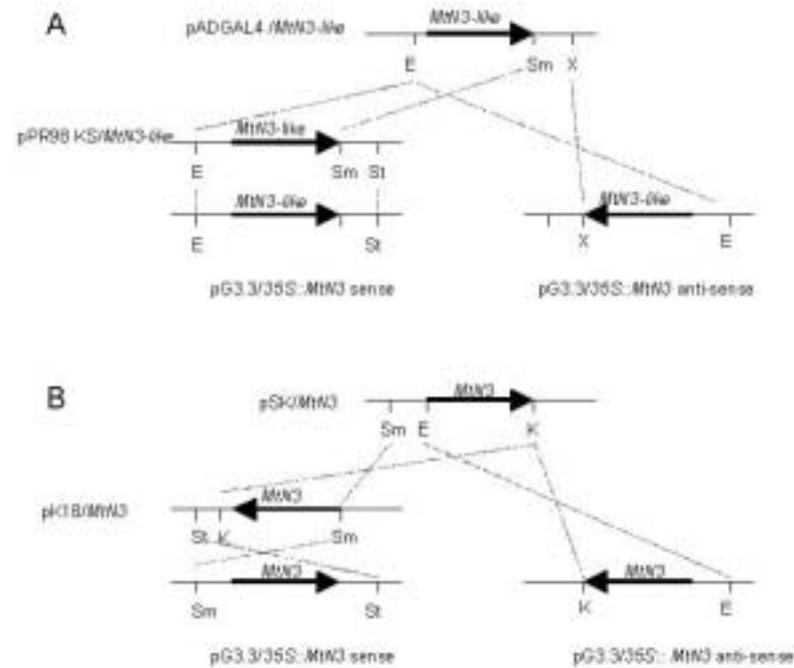


Figure 5.8 Construction of pG3.3/35S::MtN3-like sense and pG3.3/35S::MtN3-like anti-sense (A) and pG3.3/35S::MtN3 sense and pG3.3/35S::MtN3 anti-sense (B). E: EcoR1, Sm: Xma1, St: Sst, X: Xba1, K: Kpn1. Pictures are not to scale. See text for details.

was constructed by cloning the *MtN3* cDNA as a Xma1-Kpn1 from the vector pSK/*MtN3* (provided by P. Gamas) in the vector pK18 (Pridmore, 1987), followed by cloning the *MtN3* cDNA as a Xma1-Sst1 fragment in pG3.3. pG3.3/35S::MtN3 anti-sense was constructed by cloning the *MtN3* cDNA as an EcoR1-Kpn1 fragment in pG3.3.

#### 5.4.2 *Rhizobium* strains

*Medicago truncatula* plants were inoculated with the wild-type *Sinorhizobium meliloti* strain Rm41 or strain Rm2011. Fix<sup>-</sup> nodules were obtained by inoculation of the roots with the strain Rm0540 (*exoY*::Tn5, EPS<sup>-</sup>, Inf<sup>-</sup>; Mueller *et al.*, 1988) or the strain Rm8368 (*pho*-1, *bacA386*::Tn*phoA*; Glazenbrook *et al.*, 1993).

#### 5.4.3 Plant material, plant growth conditions and plant transformation

*Medicago truncatula* (Gaertn.) line R108-1 (c3), was cultured in the greenhouse and *in vitro* as described in chapter 2. Transgenic hairy roots were obtained by *Agrobacterium rhizogenes* strain Arqual (Quandt *et al.*, 1993) mediated transformation as described by Boissons-Dernier *et al.* (2001). For nodulation experiments wild-type and B9-1 plants were grown in Perlite or in an aeroponic system (for the time-course experiments) and watered with solution I (<http://www.isv.cnrs-gif.fr/embo2/manuels/index.html> and chapter 2).

*Arabidopsis thaliana* plants, cultivar Columbia, was grown in the greenhouse at 60% relative humidity under a 16h photoperiod at 25 °C and 15 °C during day and night respectively. *A. thaliana* was transformed by *Agrobacterium tumefaciens* strain EHA105 (Hood *et al.*, 1993) using the flower dip method described by Clough and Bent (1998). *A. thaliana* plants were watered with commercially available fertilizer (Soluplant, Duclos International).

#### 5.4.4 DNA isolation and Southern blot analysis

Plant genomic DNA was isolated as described by Dellaporta *et al.* (1983). Southern blot analysis was performed as described in chapter 2. The hybridization temperature was 65 °C. The *GUS* probe consisted of a 702-bp PCR fragment (nucleotides 240 to 942 downstream of the *GUS* translation initiation site; Jefferson *et al.*, 1986) amplified on the vector pGKB5 (Bouchez *et al.*, 1993) using the oligonucleotides 5'GGCCAGCGTATCGTGCTGCG3' and 5'GGTCGTGCACCATCAGCACG3'. The B9-1 RB probe consisted of an 1147 bp PCR fragment amplified on the B9-1 HpaI-ScaI library with the primers RBGUS2 and AP2. The hybridization temperature was 65 °C.

### 5.4.5 PCR

PCRs were carried out on 1  $\mu$ l of plant genomic DNA (ca. 100 ng) or 0.5 ng plasmid DNA using the Taq polymerase and buffer from Boehringer Mannheim (Mannheim, Germany).

The following primers were used:

MAR91:315+:	5'AATCTGCTCTTCTTCATCAC3' (nucleotides 1390-1411 of Acc. No. AJ307887).
MAR82:102+:	5'CCATCTGAATTATGCCGAACA3' (nucleotides 2208-2228 of Acc. No. AJ307887)
MAR83:234-:	5'AGAATACTTACAATGACGAA3' (nucleotides 1731-1750 of Acc. No. AJ307887)
B9-13'cDNA:	5'GTTTGACTGAGGGATACACTA3' (nucleotides 304-325 downstream of the <i>MtN3</i> -like stop codon).
Npt1:	5'GAGGCTATTCGGCTATGACTG3'
Npt2:	5'ATCGGGAGCGGCGATACCGTA3'
RBGUS2:	5'TCACGGGTTGGGGTTTCTACAGGAC3'
EcoB9-1 3' :	5'CAGGGATTACTATGTTGCTCTA 3' (nucleotides 359-337 upstream of the <i>MtN3</i> -like stop codon)
EcoGAM748+ :	5'TCTACAGGAATGGTGGGGAAAAAGC3' (nucleotides 131-106 upstream of the <i>MtN3</i> stop codon).
Xba5'B9-1+ :	5'GCTCTAGATAGAGAGATAGTGTGTTGTG3' (annealing to nucleotides 38-60 upstream of the <i>MtN3</i> -like start codon)
BamH3'B9-1- :	5'CGCGGATCCTGACTTTGCTACCAGCACCTCCAC3' (annealing to the 16 nucleotides before the <i>MtN3</i> -like stop codon)
XbaGAM79+ :	5'GATGTGTAGTTGTTGTTGTTGTTG 3' (annealing to nucleotides 35-68 upstream of the <i>MtN3</i> start codon)
BamGAM895- :	5'CGCGGATCCTAACTCCACAACCGATAGGTTCTTC3' (annealing to the 16 nucleotides before the <i>MtN3</i> stop codon)

### 5.4.6 cDNA library screening

The *MtN3*-like cDNA was obtained by screening of a young root nodule cDNA library (Györgyey *et al.*, 2000). As a probe a 421 bp DNA fragment that was obtained by PCR on the same nodule cDNA library with the primers MAR91:315+ and MAR82:102+ (see above for sequences) that were designed on the putative open reading frame of the T-DNA RB flanking sequence was used. Digestion with EcoRI and XhoI (the enzymes used to clone the cDNAs in the phage vector) showed that the hybridizing clones contained inserts ranging between 400 and 1700 bp.

### 5.4.7 RNA isolation and Northern blot analysis

RNA from *M. truncatula* and *A. thaliana* was extracted according to the method described in Kay *et al.* (1987). 15  $\mu$ g of total RNA were separated on a formaldehyde 1.2% agarose gel and transferred to a nylon membrane (Biotrans Nylon Membrane, ICN) according to the manufacturer's instructions. A 532 bp DNA template for the *MtN3*-like probe was obtained by PCR amplification on *MtN3*-like cDNA with the primers EcoB9-1 3' and B9-13' cDNA, a 360 bp template for the *MtN3* probe was obtained by PCR amplification on *MtN3* cDNA cloned in the vector pBluescript (provided by P. Gamas) with the primers EcoGAM748+ and T7. The probes were labeled with <sup>32</sup>P by random priming using the Megaprime DNA Labeling System (Amersham). The hybridization was carried out overnight at 65°C in Church and Gilbert buffer (Church and Gilbert, 1984) and the membranes were washed according to Sambrook *et al.* (1989). Blots were exposed on a Storm phosphorimager (Molecular Dynamics, Sunnyvale, CA). For RNA quantification the blots were stained with 0.3 % methylene blue in 0.3 M sodium acetate, pH 5.2.

### 5.4.8 Particle bombardment of onion epidermal cells

For the particle bombardment of onion cells the protocol described by Scott *et al.* (1999) was followed. For this the 'Biolistic PDS-1000/He' (Bio-Rad) system was used with rupture discs of 750 Psi and a 28 mm Hg vacuum. GFP was viewed using a confocal microscope Sarastro 2000 (Molecular Dynamics) equipped with an ImageSpace 3D software (Molecular Dynamics, Saclay, France) on SiliconGraphics workstation using a 40X NA 0.75 water immersion objective (Nikon). Excitation was at 488 nm and emission at > 510.

## 6 Disruption of the *NARF*-like gene in line A7-1 causes a dwarf phenotype

### 6.1 Introduction

Originally the T-DNA tagged line A7-1 was selected because the tagged plants exhibited GUS staining in the vascular tissues of several organs (chapter 3). However, this *GUS* expression seemed not to be the result of a gene fusion, since no long open reading frame was present upstream of the *GUS* reporter gene and the T-DNA was inserted in opposite direction downstream of a gene corresponding to an *M. truncatula* expressed sequence (TC11443). In principle, the insertion of T-DNA in a gene may cause knock-out mutants in homozygous insertion mutants. This seemed to be the case for plants of line A7-1 that displayed a dwarf phenotype. This chapter will describe in more detail the analysis of line A7-1, in order to examine whether the dwarf phenotype indeed was caused by the T-DNA insertion.

### 6.2 Results

#### 6.2.1 A7-1 plants are unaffected in nodulation but exhibit a dwarf phenotype

In order to determine whether plants of line A7-1 showed an aberrant nodulation phenotype, 88 T2 plants from a heterozygous A7-1 plant were inoculated with *S. meliloti* and tested after ca. 4 weeks for nodulation. All the plants formed nodules with a wild-type appearance. However, 23 (=26%) showed a dwarf phenotype in the aerial part (Figs. 6.1 A and B), while the size of the root system was unaltered. The dwarf plants produced remarkably fewer flowers and seeds than wild-type plants. These seeds germinated like wild-type seeds and all gave rise to dwarf plants. Analysis of the leaves of wild-type and dwarf plants by light microscopy did not reveal differences between the leaf cells of wild types and dwarfs (data not shown). However, scanning electron microscopy of the surface of leaves of wild-type and dwarf plants showed that all the cells of the wild-type plants were round-shaped, while a part of those of the dwarf plants had a collapsed appearance (Figs 6.1 C and D). The size of the epidermal leaf cells is the same in dwarfs and wild-type plants. This suggested that the dwarf phenotype was caused by a reduction in cell number rather than by a reduction in cell size, although the size of the cells in deeper cell layers could not be observed by scanning electron microscopy.

When A7-1 plants were grown under greenhouse conditions, the dwarf phenotype started to become visible when the first trifoliolate leaf appeared (after ca. 2 weeks). This was much smaller than that from wild-type or A7-1 plants with a wild-type phenotype. However, when the plants were grown on solid nutrient medium in slanted square petri-dishes, the dwarf phenotype did not appear. After one month of growth in petri-dishes none out of 58 plants that were descendant of a heterozygous T-DNA insertion A7-1 plant showed an aberrant phenotype. However, after being transferred to the greenhouse, newly formed leaves and petioles of ten of them exhibited the dwarf phenotype. This suggested that the dwarf phenotype is conditional and might be affected by the growth conditions, such as humidity that is much higher in the petri-dishes than in the green house.

#### 6.2.2 The dwarf phenotype is linked to the T-DNA

Previously (chapters 3 and 4) it was shown that line A7-1 contained a single T-DNA insertion and Southern blot analysis of three dwarf plants and four A7-1 line plants with wild-type phenotype showed that the three dwarf plants were homozygous insertion mutants while this was not the case for any of the plants with the wild-type phenotype. This, together with the Mendelian segregation of the dwarf phenotype (see §6.2.1) and the fact that all the dwarf plants displayed GUS staining (result not shown), suggested linkage of the presence of the T-DNA insertion and the dwarf phenotype. Southern blot analysis of 14 A7-1 plants with, and 13 A7-1 plants without the

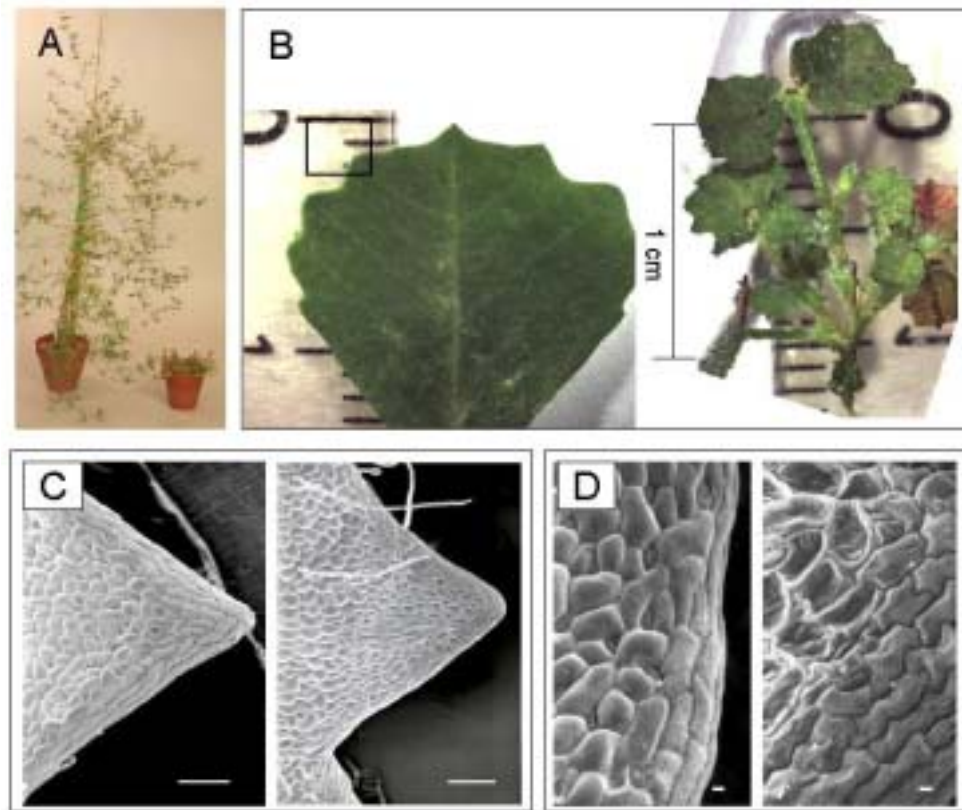


Figure 6.1 Macroscopic and microscopic phenotype of dwarf plants of T-DNA tagged *Medicago* plants of line A7-1. **A:** Two months old wild-type plant (left) and dwarf plant (right). **B:** Wild-type leaf (left) and dwarf shoot (right). **C:** Scanning electron microscopy of a wild-type leaf point (left, boxed in B) and a dwarf leaf point (right). **D:** Scanning electron microscopy of a leaf margin from a wild-type leaf (left) and dwarf leaf (right). Bars in C and D represent 1  $\mu$ m.

dwarf phenotype confirmed this linkage. The hybridization of EcoRV digested genomic DNA with the RB T-DNA flanking region revealed a 4 kb fragment in the plants with a wild-type phenotype that was absent in the dwarf plants and corresponded to the wild-type locus (Fig. 6.2).

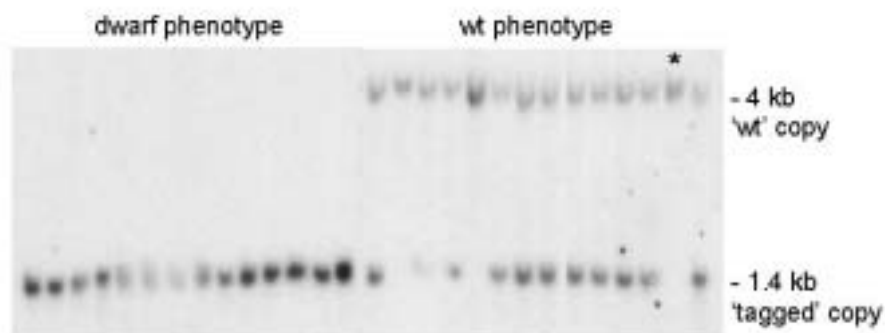


Figure 6.2 Southern blot analysis of EcoRV digested genomic DNA of 14 T2 A7-1 plants with a dwarf phenotype and 13 T2 A7-1 plants with a wild-type phenotype. The A7-1 RB T-DNA flanking region was used as a probe. The lane marked with an asterisk contained DNA from an untransformed plant.

In addition a 1.4 kb fragment hybridized with the probe. This 1.4 kb fragment that corresponds to the tagged locus was present in all the dwarf plants and in 11 heterozygous A7-1 plant but is absent in the wild type (Fig. 6.2). This 1.4 kb fragment is most likely linked to the T-DNA insertion since it was also detected when *GUS* was used as a probe (chapter 3). Analysis of additional A7-1 plants of later generations confirmed the linkage of the dwarf phenotype to the homozygous presence of the T-DNA insertion (result not shown).

### 6.2.3 In line A7-1 a *NARF-like* gene is tagged

Blast searches using the sequence of the T-DNA RB flanking region of line A7-1 (AJ3311927) that was isolated by PCR walk (chapter 4) revealed that in line A7-1 the T-DNA was inserted in a homologue of the human NARF protein (Nuclear prelamina A recognition factor; AF128406\_1), and therefore we named the tagged gene *MtNARF-like*.

While the sequence of the 5' end of the *Medicago* *NARF-like* gene was not available yet, the genomic sequence of the *Arabidopsis* gene homologous to *MtNARF-like* was known (AGI, 2000). Using the primers AtNarf5 and AtNarf3, that were designed on the putative 5' and 3' UTRs of the *Arabidopsis* *NARF-like* gene, a 2340 bp genomic fragment and a 1517 bp cDNA fragment could be amplified on genomic DNA and leaf cDNA of *A. thaliana* (ecotype WS), respectively. The deduced amino acid sequence of the *Arabidopsis* *NARF-like* cDNA revealed a protein of 476 amino acids and comparison of the genomic locus to the cDNA showed that the gene contains 10 exons ranging in size from 15 to 341 bp and 9 introns ranging in size from 73 to 159 bp (Fig. 6.3 A).

PCRs on wild-type genomic DNA that was digested with Hpa1/Sca1 and ligated to adaptor sequences (see chapter 4) using the primers A7-1RBA and A7-1RBB that were designed on the T-DNA right border flanking sequence of line A7-1 and the primers AP1 and AP2 that were adaptor specific (Fig. 6.3 B) permitted the isolation of an additional 1140 bp of the *M. truncatula* *NARF-like* genomic locus. Comparison of the sequence of this fragment to the *NARF-like* sequence of *A. thaliana* showed that this fragment contained the 3' exon of the *MtNARF-like* gene corresponding to the 10th *Arabidopsis* exon and part of the sequence corresponding to the 9th *Arabidopsis* exon. In this region of the gene the position of the intron was conserved between the two species. However, the intron was about 10-fold longer in *M. truncatula* than in *A. thaliana*. We supposed that also the other introns of the *Medicago* gene would be longer than those present in the *Arabidopsis* gene. Since a full-size *M. truncatula* clone was not yet available it was decided to use the *Arabidopsis* gene for complementation, overexpression and silencing experiments (see below). During this work additional *Medicago* sequences became available. Besides the known 3' *Medicago* EST (BB47G04) also two *Medicago* ESTs (BF637313 and BG450904) corresponding to the 5' end of the *NARF-like* gene were submitted to the gene bank, facilitating the isolation of the full-length genomic and cDNA sequences of the *Medicago* *NARF-like* gene. The *Arabidopsis* genome sequence shows that *Arabidopsis* contains a single *NARF-like* gene. Southern analysis showed that this was probably also the case for *M. truncatula* (Fig. 6.4 A, see below) and we supposed that, although the 5' and 3' ESTs present in the database did not overlap, they belonged to the same cDNA. Indeed, by using the primers MtNarf5 and MtNarf3 that were designed on the 5' and 3' UTRs of the 5' and 3' *M. truncatula* *NARF-like* ESTs respectively, a 1772 bp cDNA fragment could be amplified on leaf cDNA. The sequences of this fragment corresponding with the available EST sequences showed 100% identity. The deduced amino acid sequence revealed a protein of 479 amino acids. However, with this primer pair the corresponding genomic sequence could not be amplified. Since the length of the genomic locus might have been too long to be amplified with MtNarf5 and MtNarf3, the additional primers A7-1PLD, A7-1Y and A7-1X (Fig. 6.3 A) were designed on the *MtNARF-like* cDNA. PCRs with the primer pairs MtNarf5/A7-1Y and A7-1PLD/A7-1X resulted in the amplification of a 4367 bp and a 2900 bp fragment, respectively. The overlapping part of these sequences were 100% identical and assembling the sequences of these fragments with the already known sequence yielded the entire genomic sequence of the *NARF-like* coding region (AJ417911, Fig. 6.3 A). Comparison of the *NARF-like* cDNA to the genomic sequence showed that the *MtNARF-like* gene contains 11 exons, ranging in size from 50 to 486 bp, and 10 introns, ranging in size from 77 to 1378 bp (Fig. 6.3 A). Comparison of the gene structure of *MtNARF-like* to that of *AtNARF-like* revealed that the positions of the introns were remarkably well conserved between the two species (Fig. 6.3 C), except that equivalents of the first and fourth intron of *MtNARF-like* were absent in the *AtNARF-like* gene. In addition, the comparison showed that four introns of the *MtNARF-like* gene were markedly longer than the corresponding ones in the *Arabidopsis* *NARF-like* gene (Fig. 6.3 A).

To determine the T-DNA integration site of line A7-1, in addition to the T-DNA RB junction also the junction of the genomic DNA to the LB T-DNA was isolated. This was achieved by PCR on genomic DNA of a tagged A7-1 line with the primer pair Narf125/GKB5A resulting in a 600 bp fragment containing the junction between the T-DNA and the left border flanking region (Fig. 6.3 B). Comparison of the two T-DNA flanking regions to the corresponding wild-type locus showed that the T-DNA insertion resulted in a deletion of 223 bp, corresponding to a part of the last intron and 50 amino acids at the C-terminal end of the predicted protein (underlined in Fig. 6.3 C).

Southern blot analysis of genomic wild-type DNA of *M. truncatula* DNA digested with HindIII or EcoRI using a probe corresponding to the *MtNARF-like* gene confirmed that *M. truncatula* contains a single copy of the *NARF-like* gene (Fig. 6.4 A). By using the same probe, no *NARF-like* mRNA could be detected by Northern analysis of RNA isolated from closed leaves, leaves, apical meristems, flower buds, open flowers, young seed pods, roots or nodules (data not shown). However, expression of the gene was detected in these

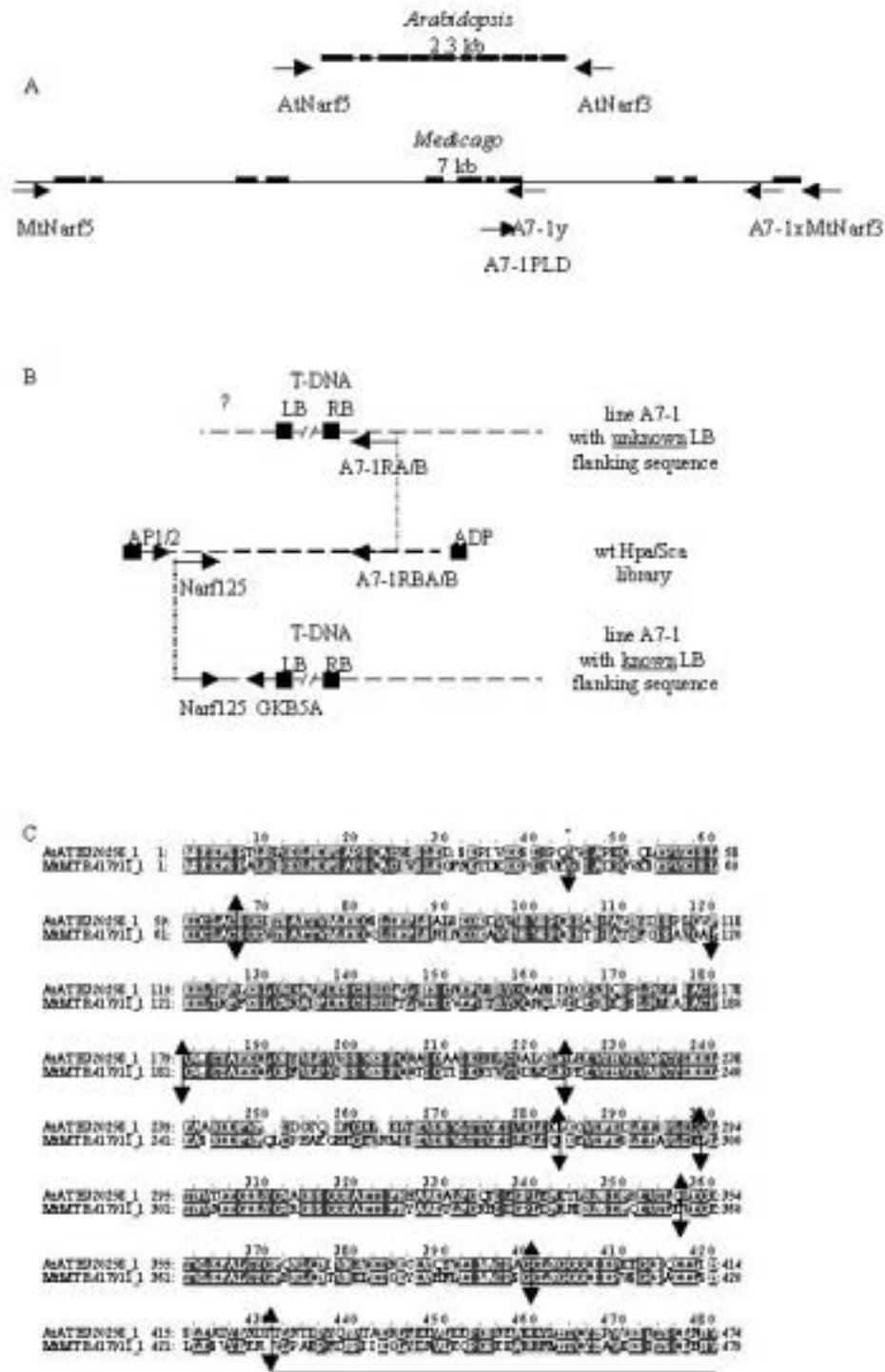


Figure 6.3 **A**: Genomic loci of the *A. thaliana* and *M. truncatula* *NARF-like* genes (picture is to scale). Boxes represent exons and lines introns. Arrows represent the primers used for the isolation of the genomic and cDNA sequences. **B**: PCR strategy to isolate the junction of the T-DNA LB to the T-DNA LB flanking sequence of line A7-1. Using the nested primers A7-1RBA and A7-1RBB, designed on the T-DNA RB flanking region, in combination with nested adaptor-specific primers AP1 and AP2, a PCR walk on genomic wild-type DNA digested with HpaI/ScaI yielded a PCR fragment of 1300 bp, which prolonged the known sequence of the 1957 bp T-DNA RB flanking region with 1140 nucleotides. PCR with the primers Narf125 and GKB5B permitted the isolation of a fragment containing the junction to the T-DNA LB. **C**: Comparison of the deduced amino acid sequences of the *Arabidopsis* and the *Medicago* *NARF-like* genes. The vertical arrows indicate the position of introns. The 55 amino acids that were deleted upon T-DNA insertion in the *Medicago* *NARF-like* gene are underlined.

organs by a RT-PCR experiment and in all organs a similar level of the mRNA was detected (Fig. 6.4 B). Thus, the *MtNARF-like* gene is expressed at low levels throughout the plant.



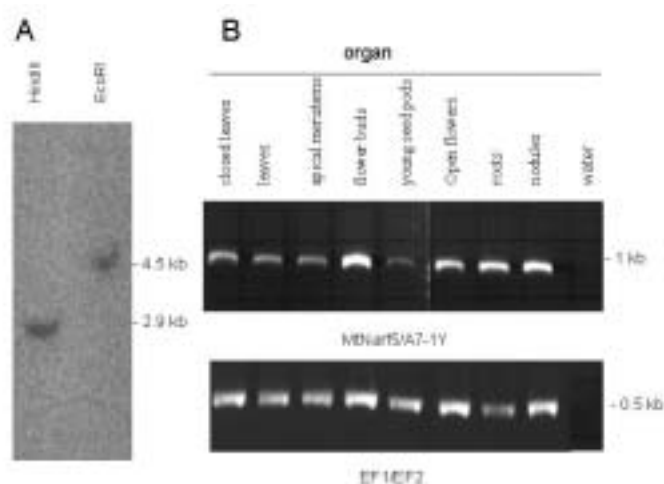


Figure 6.4 **A**: Southern blot analysis of wild-type genomic *M. truncatula* DNA using the *MtNARF*-like probe. The used restriction enzymes are shown in the figure. **B**: RT-PCR on *M. truncatula* cDNA from different organs. For the amplification of the *MtNARF*-like gene and the elongation factor 30 cycles were used.

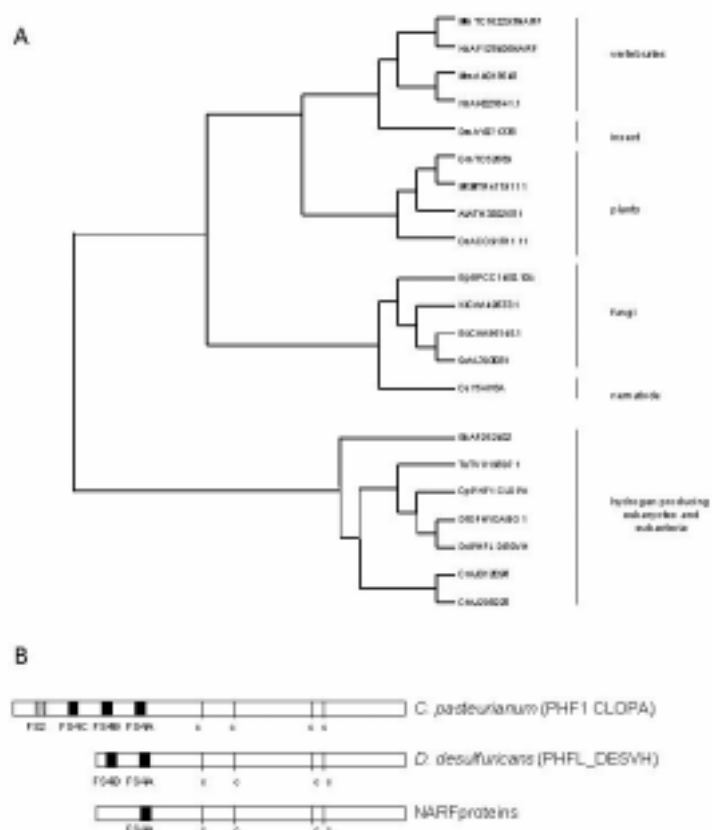


Figure 6.5 Relationship of the *M. truncatula* NARF-like to other NARF and NARF-like proteins. **A**: Dendrogram of deduced amino acid sequences of ESTs of *MtN3*-like genes generated with the PILE-UP program of GCG software package (At = *Arabidopsis thaliana*, Os = *Oryza sativa*, Mt = *Medicago truncatula*, Gm = *Glycine max*, Dm = *Drosophila melanogaster*, Mm = *Mus musculus*, Hs = *Homo sapiens*, Sp = *Schizosaccharomyces pombe*, Kl = *Kluyveromyces lactis*, Sc = *Saccharomyces cerevisiae*, Sr = *Zygosaccharomyces rouxii*, Ce = *Caenorhabditis elegans*, Sb = *Spironucleus barkhanus*, Cr = *Chlamydomonas reinhardtii*, Cf = *Chlorella fusca*, Tv = *Trichomonas vaginalis*, Cp = *Clostridium pasteurianum*, Dd = *Desulfovibrio desulfuricans* and Df = *Desulfovibrio fructosovorans*). **B**: Comparison of the conserved domain organization of the NARF proteins from plants, insects, vertebrates, fungi and nematodes (indicated with NARF proteins) with the hydrogenases of *C. pasteurianum* and *D. desulfuricans*. Below the sequences the FS2, FS4C, FS4A and FS4B Fe-S clusters and the cysteines that are involved in the coordination of the active site are shown. Figure B is adapted from Horner *et al.* (2000).

### 6.2.4 The *MtNARF*-like gene shows homology to iron-only hydrogenases

To determine whether besides human, *Medicago* and *Arabidopsis* also other species contain genes that are homologous to *MtNARF*-like, a TblastN search using the deduced amino acid sequence of the *MtNARF*-like gene was carried out. This search revealed the presence of ESTs homologous to *MtNARF*-like in plants, insects, vertebrates, fungi, nematodes, hydrogen producing protists and eubacteria. Phylogenetic analysis of the sequences showed that the NARF related proteins from plants, insects, vertebrates, fungi and nematodes probably have a common ancestor (Fig. 6.5 A). Surprisingly, for the vertebrates two *NARF* homologues were present in the database, while in plants, insects, fungi and nematodes only a single *NARF* homologue was found.

The Blast search revealed that the protein showing the highest homology with *MtNARF*-like was the human NARF protein (AF128406\_1). NARF was found to bind farnesylated prelamins A and it was hypothesized that this association is important for processing of prelamins A (Barton and Worman, 1999). The proteins from the hydrogen-forming anaerobic eubacteria *Clostridium pasteurianum* and *Desulfovibrio desulfuricans* that are homologous to *MtNARF*-like are iron-only hydrogenases. Iron-only hydrogenases catalyze the formation of hydrogen by combining protons and electrons. These protons are the terminal electron acceptor for electrons from anaerobic oxidation of pyruvate to carbon dioxide and acetate (Adams and Stiefel, 1998). The X-ray crystal structures of the iron-only hydrogenases from *Clostridium pasteurianum* and *Desulfovibrio desulfuricans* showed that the enzyme consists of two domains, one containing the active site domain and one containing four accessory Fe-S clusters termed FS2, FS4C, FS4B and FS4A (in the case of the *C. pasteurianum* hydrogenase PHF1 CLOPA; Peters *et al.*, 1998) or only the Fe-S clusters FS4B and FS4A (in the case of the *D. desulfuricans* hydrogenase PHFL\_DESVH, Nicolet *et al.*, 1999). In *Clostridium*, electrons are imported into the active site through these accessory Fe-S clusters. Comparison of the conserved domain organization of the NARF proteins from plants, insects, vertebrates, fungi and nematodes with the hydrogenases of *C. pasteurianum* and *D. desulfuricans* (Fig. 6.5 B) shows that the amino acids important for the FS2 cluster and the FS4C cluster are only present in the *C. pasteurianum* hydrogenase but not in the other proteins. The cysteines important for the FS4B and FS4A clusters are present in the two crystallized sequences while three of the four cysteines important for the FS4A cluster are present in the NARF and the NARF-like proteins. The cysteines forming the active site (or H-cluster) are present in all the proteins and thus may play a role in the functioning of all of them (Fig. 6.5 B).

### 6.2.5 Dwarfs can be rescued by complementation with *AtNARF*-like and silencing of the *NARF*-like gene in *Arabidopsis* results in a dwarf phenotype

To determine whether the disruption of the *NARF*-like gene caused the dwarf phenotype and whether the *Medicago* and *Arabidopsis* *NARF*-like genes have similar functions, two approaches were used: heterologous complementation of the dwarf phenotype with the *AtNARF*-like gene and silencing of the *AtNARF*-like gene in *Arabidopsis* by an RNAi approach.

For the heterologous complementation, leaf explants of *M. truncatula* dwarf plants were transformed using *Agrobacterium* transformation (Trinh *et al.*, 1998) with the T-DNA from the binary vector pSAM/35S::g*AtNARF*-like. This T-DNA contained the *Arabidopsis* *NARF*-like wild-type genomic locus under the control of the *CAMV* 35S promoter (Fig. 6.6 A) or the pSAM vector without the *AtNARF*-like genomic fragment, respectively. The regeneration of dwarf plants transformed with the empty vector pSAM appeared to be difficult and only 20 plants were obtained. All these plants had conserved the dwarf phenotype (result not shown). The transformation with the 35S::g*AtNARF*-like construct gave rise to numerous primary transformants, that at the moment of transfer from the *in vitro* conditions to the greenhouse could not be distinguished from each other. In total forty regenerated plants obtained from the dwarf leaf explants transformed with the 35S::g*AtNARF*-like construct were kept. They all had lost the dwarf phenotype and 22 of them showed a wild-type phenotype while eighteen plants had the wild-type size but an aberrant leaf form. The leaves of these plants had irregular, lobed forms (Fig. 6.6 A).

To study the effect of overexpression of the *AtNARF*-like gene in *M. truncatula*, wild-type plants were transformed with the T-DNA of pSAM/35S::g*AtNARF*-like or the empty vector. Overexpression of the gene did not influence the phenotype and the plants carrying the 35S::g*AtNARF*-like fusion could not be distinguished from wild types or plants carrying the T-DNA of pSAM (result not shown).

For the RNAi approach *Arabidopsis* plants carrying the 35S::g*AtNARF*-like-gDNA/cDNA construct (Fig. 6.6 B) in the binary vector pG3.3 were obtained by the flower dip method (Clough and Bent; 1998). In addition to a *CAMV* 35S promoter driving the RNAi construct, the vector pG3.3 contains a 35S::GUS fusion, which makes screening for transformed tissues by GUS staining possible. As a control, *Arabidopsis* plants carrying the pG3.3 T-DNA without the RNAi construct were obtained. From each transformation twelve plants were kept and these kanamycin resistant primary transformants all showed GUS staining. While two weeks after sowing the plants carrying the empty pG3.3 T-DNA all showed a wild-type phenotype, the progeny from the plants carrying the RNAi construct were of variable sizes (Fig. 6.6 B). Six weeks after sowing only in plant number 7 the dwarf phenotype persisted while the other plants could not be distinguished from the plants carrying the empty

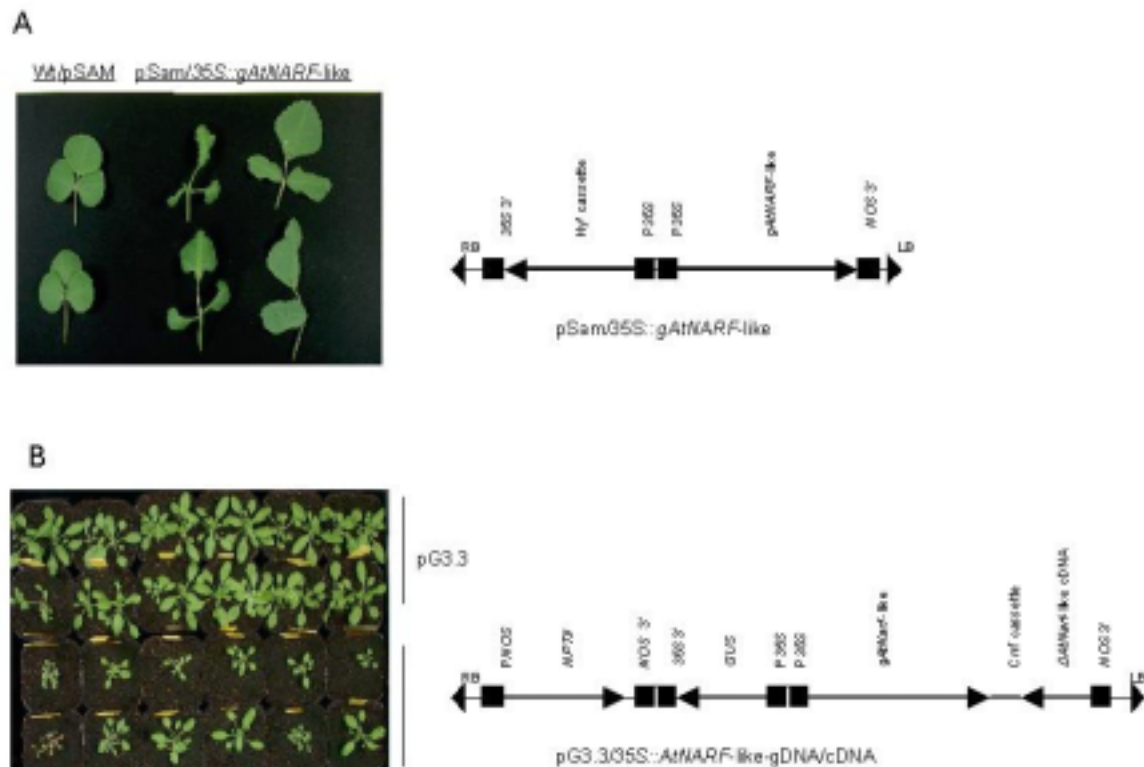


Figure 6.6 Phenotypes of *Medicago* dwarf plants complemented with the *AtNARF*-like gene (A) and *Arabidopsis* plants carrying the RNAi construct (B). Next to the photographs systematic overviews of the used constructions are shown. RB: T-DNA right border, 35S 3': 3' region of the 35S transcript of the Cauliflower Mosaic Virus, Hy<sup>r</sup>: hygromycin resistance cassette, P35S: promoter of the 35S gene of the Cauliflower Mosaic Virus, *gAtNARF*-like: genomic sequence containing the *Arabidopsis NARF*-like gene, NOS 3': 3' region of the nopaline synthase gene from pTiC58, PNOS: promoter region of the nopaline synthase gene, *NPTII*: neomycin phosphotransferase II, *GUS*: coding region of the  $\beta$ -glucuronidase (*GUS*) gene from *E. coli*, Cm<sup>r</sup>: chloramphenicol resistance cassette,  $\pm$ *AtNarf*-like cDNA: shortened *Arabidopsis NARF*-like cDNA.

pG3.3 T-DNA. Unfortunately, probably due to its small size, this plant produced very few seeds. However, three out of the four obtained progeny plants from plant 7 showed a dwarf phenotype and GUS staining, while the only plant with a wild-type phenotype was GUS negative. Although these results have to be confirmed using a larger number of plants, they suggested that the dwarf phenotype is linked to the presence of the RNAi construct and thus is the result of silencing of the endogenous gene.

These experiments showed the *Medicago* dwarfs can be rescued by the heterologous *AtNARF*-like and that in *Arabidopsis* the role of the *NARF*-like gene might be similar to that of the *Medicago NARF*-like gene.

## 6.3 Discussion

### 6.3.1 In line A7-1 the T-DNA insertion caused a dwarf phenotype

In the T-DNA tagged *M. truncatula* line A7-1 the T-DNA was inserted in a homologue of the human *NARF* gene. The homozygous T-DNA insertion mutants showed a pronounced dwarf phenotype. The aim of this chapter was to determine whether this dwarf phenotype was caused by the disruption of the *MtNARF*-like gene. The dwarf phenotype turned out to be tightly linked to the T-DNA and further it could be rescued by complementation with the *NARF*-like gene of *Arabidopsis*. In addition, silencing of the *Arabidopsis NARF*-like gene resulted in a dwarf phenotype. This showed that the dwarf phenotype indeed was caused by disruption of this gene and that the *NARF*-like genes of *Arabidopsis* and *Medicago* are probably orthologues fulfilling similar functions. While a knock-out mutation of the *NARF*-like gene caused a dwarf phenotype, overexpression of the *NARF*-like gene did not cause a mutant phenotype. This indicates that a high level of the *NARF*-like protein does not affect the plant.

Line A7-1 is the first T-DNA tagged *M. truncatula* line in which the mutant phenotype is proven to be caused by disruption of a gene by the T-DNA insertion. This line thus nicely illustrates that T-DNA tagging can be used as a tool to discover genes of *M. truncatula* and to study their functions.

### 6.3.2 What could be the function of the NARF-like proteins from *Arabidopsis* and *Medicago*?

The T-DNA tagged *NARF-like* gene in line A7-1 showed homology to the human NARF protein and, although less strong, to iron-only hydrogenases. The *Arabidopsis* and the *Medicago* NARF-like proteins contain the cystein residues found in the H-cluster active site of the iron-only hydrogenases, but lost the Fe-S accessory clusters. However, although the presence of the Fe-S accessory clusters is not a prerequisite for hydrogenase activity (e.g. the hydrogenase enzymes produced under anaerobic conditions in the green algae *Chlorella*, *Chlamydomonas* and *Scenedesmus* also lack the Fe-S clusters, but still produce hydrogen (Florin *et al.*, 2001)) it is not probable that the NARF-like proteins from *Arabidopsis* and *Medicago* act as hydrogenases, since hydrogen production has never been observed in 'higher' eukaryotes (Horner *et al.*, 2000).

Would the role of MtNARF-like be rather related to that of the human NARF protein? The human NARF studied in more detail, was found to bind farnesylated prelamins A and it was hypothesized that this association is important for processing of prelamins A (Barton and Worman, 1999). Lamins are part of the nuclear lamina that is located between the inner nuclear membrane and the peripheral chromatin. The nuclear lamina is required for proper nuclear organization, DNA replication, post-mitotic nuclear reassembly, cell differentiation and cell cycle regulation (reviewed in Hutchison *et al.*, 1994). Interestingly, the *Medicago* dwarf plants contained less, but not smaller cells. Therefore it might be possible that in the leaves of the dwarf plants the cell division activity is reduced.

The scanning electron microscopy showed that certain cells in the dwarf leaves had a collapsed appearance, in contrast to cells in wild-type leaves. It seems unlikely that such collapsed cells could be alive. However, the corresponding cells in the dwarf plants seemed to be alive and therefore this morphological difference is probably due to a lower resistance of the dwarf cell walls to the vacuum used during electron microscopy. Therefore, the dwarf plants also might have cell walls with an altered rigidity. It was shown that the dwarf phenotype is conditional, being visible in plants grown under greenhouse conditions, but absent from plants grown under *in vitro* conditions. The cause of this conditional character is however unclear.

It is clear that still many questions concerning the function of the NARF-like proteins from *Medicago* and *Arabidopsis* remain to be answered. However, the dwarf plants, the complemented dwarfs and the silenced *Arabidopsis* plants will be important tools to unravel the function of the NARF type proteins.

## 6.4 Materials and methods

### 6.4.1 Plant material

*Medicago truncatula* (Gaertn.) line R108-1 (c3), was cultured in the greenhouse and *in vitro* as described in chapter 3. *M. truncatula* plants were transformed and regenerated as described by Trinh *et al.*, 1998). The T-DNA tagged line A7-1 was obtained by insertional mutagenesis using the T-DNA vector pGKB5 (Bouché *et al.*, 1993; chapter 3). *M. truncatula* plants were nodulated by inoculation with the wild-type *Sinorhizobium meliloti* strain Rm41.

*Arabidopsis thaliana* plants, cultivar Columbia, were grown in the greenhouse and transformed by the flower dip method (Clough and Bent, 1998) as described in chapter 5.

### 6.4.2 Plasmids

Standard DNA manipulations were according to Sambrook *et al.* (1989). The binary vector pSAM/35S::gAtNARF-like (Fig. 6.6 A) was produced by cloning the *CAMV 35S* promoter of the vector pG3.3 (P. Ratet, unpublished) as a HindIII-EcoRI fragment in the binary vector pCambia 1390 (AF234307). By PCR amplification on subcloned *Arabidopsis* genomic NARF-like DNA using the primers Narf500 and Narf300, an EcoRI and XbaI restriction site were added to the 5' and 3' ends of the PCR fragment, respectively. The PCR fragment was digested with EcoRI and XbaI and cloned in the modified PCAMBIA 1390 vector yielding the vector pSAM/35S::gAtNARF-like (Fig. 6.6 A).

For the construction of the RNAi vector pG3.3/35S::AtNARF-like-gDNA/cDNA (Fig. 6.6 B) first 662 bp were removed from the 5' of the *Arabidopsis* NARF-like cDNA cloned in pGEM-T easy by restriction with NdeI and religation to remove an XbaI site that would complicate further cloning steps. Then this shortened

cDNA was cloned as an 855 bp EcoRI-SstI fragment in the vector pG3.3. The genomic *NARF-like* DNA was cloned as a 2124 bp XbaI-BamHI fragment in the pG3.3 vector containing the cDNA. Finally, the cDNA and genomic DNA fragments were separated by introducing the chloramphenicol resistance cassette from the vector pSKS114 as a BamHI fragment, yielding the vector pG3.3/35S::*AtNARF-like*-gDNA/cDNA (Fig. 6.6 B).

### 6.4.3 Southern and Northern blot analysis and cDNA synthesis

Plant genomic DNA was isolated as described by Dellaporta *et al.* (1983). Southern blot analysis was performed as described in chapter 3. The hybridization temperature was 65 °C. The A7-1 RB probe consisted of an 829 bp PCR fragment amplified on the A7-1 EcoRV-PvuII library with the primers RBGUS2 and AP2. The *MtNARF-like* probe consisted of a 1.3 kb fragment obtained by PCR with the primers Narf125 and A7-1RBB (Fig. 6.3 B). The hybridization temperature was 65 °C.

RNA was extracted according to the method described in Kay *et al.* (1987). cDNAs were synthesized by reverse transcription of 5 µg total RNA with the Superscript reverse transcriptase kit (Gibco-BRL/Life Technologies, Gaithersburg, MD) following the manufacturer's instructions. The cDNAs were dissolved in 100 µl water and for RT-PCRs 5µl of the cDNA solution was used.

### 6.4.4 PCR

For the isolation of genomic *AtNARF-like* DNA and *AtNARF-like* cDNA the following primers were used:

- AtNarf5: 5'TAGTTTCTTCGTCTCTAAAGATTAGA3' (nucleotides 1-25 of Acc. No. AJ320258)  
 AtNarf3: 5'CTTGTCAGTCAAAAAATTTGAACACT3' (nucleotides 2317-2340 of Acc. No. AJ320258).  
 Narf500: 5'AGTTTCTTCGTCTCTAGAGATTAGATT3' (annealing to nucleotides 1-27 of Acc. No. AJ320258, except the nucleotides that were changed to obtain the XbaI site).  
 Narf300: 5'TGTCAGTCAAAGAATTCGAACACTTCC3' (nucleotides 2314-2340 of Acc. No. AJ320258, except the nucleotides that were changed to obtain the EcoRI site).

For the isolation of genomic *MtNARF-like* DNA and *MtNARF-like* cDNA the following primers were used:

- MtNarf5: 5'CTCAGGTCCCTGCCTGTCCAT3' (nucleotides 1-21 of Acc. No. AJ417911)  
 MtNarf3: 5'GACCCACATGATCATCTGAGGAA3' (nucleotides 7378-7401 of Acc. No. AJ417911)  
 A7-1PLD: 5'GCTTAGAAGAGGCTCCGCTGGAT3' (nucleotides 4181-4203 of Acc. No. AJ417911)  
 A7-1y: 5'CGGAATATTGTTTCAGCGTAACCTCC3' (nucleotides 4344-4369 of Acc. No. AJ417911)  
 A7-1x: 5'TTCACGGCATGATATTGTGTATGCAT3' (nucleotides 7056-7081 of Acc. No. AJ417911)  
 A7-1RBA: 5'GACGACTCCACTCCAATAAGCCAACA3' (nucleotides 7216-7241 of Acc. No. AJ417911)  
 A7-1RBB: 5'CACAGGGAGTCCGAGAAACA3' (nucleotides 7192-7211 of Acc. No. AJ417911)  
 Narf125: 5'GCGTCGGTGTTTGGCTGTAT3' (nucleotides 6262-6281 of Acc. No. AJ417911)

PCR fragments to be sequenced and those to be used for further cloning strategies were subcloned in the vector pGEM-T easy (Promega).

For cDNA quantifying the *Medicago* elongation factor (AW688958) was amplified using the primers:

- EF1 5'AGTCTCTCTCTGCGGCTGAG3' (nucleotides 16-35 of Acc. No. AW688958)  
 EF2: 5'CGATTTTCATCGTACCTAGCCTT3' (nucleotides 5766597 of Acc. No. AW688958)

### 6.4.5 Scanning electron microscopy

For scanning electron microscopy the material was fixed by 2h incubation in cacodylate buffer (0.1 M, pH 7.4), followed by dehydration through an ethanol series.



## 7 Generation of an improved T-DNA tagging vector

### 7.1 Introduction

The binary vector pGKB5 (Bouchez *et al.*, 1993), designed for T-DNA tagging in plants was used successfully in *A. thaliana* (Bouchez *et al.*, 1993; Mollier *et al.*, 1995) and *M. truncatula* (this thesis). This T-DNA tagging vector contains the promoter-less *GUS* gene, a kanamycin-resistance gene under the control of the promoter of the *nos* gene and a Basta-resistance gene under the control of the Cauliflower Mosaic virus 35S promoter. The ATG start codon of the *GUS* gene is located close to the right border of the T-DNA and no in frame stop codons are present between the ATG of *GUS* and the right T-DNA border. Insertion of the T-DNA in a gene thus may result in a translational or a transcriptional gene fusion (Fig. 7.1) between *GUS* and an endogenous (tagged) gene, when the *GUS* inserts in the same orientation as the tagged gene. In all cases, upon insertion in a transcribed region of a plant gene, the presence of polyadenylation sites in the T-DNA will disrupt the transcription of this region.

A translational gene fusion is formed when, upon T-DNA insertion into an exon, the *GUS* gene is in frame with the open reading frame of the tagged plant gene (possibility I in Fig. 7.1).

Transcriptional gene fusions can be formed when the T-DNA inserts in the transcribed but untranslated leader of a plant gene (possibility II in Fig. 7.1). Occasionally plants can also express dicistronic transcription units (Angenon *et al.*, 1989; Wang and Wessler, 1998; Koncz *et al.*, 1989), therefore, transcriptional gene fusions may also be formed when the T-DNA inserts in an exon, but the *GUS* gene is out of frame with the plant gene (possibility IIIA) or when the T-DNA inserts in the 3' UTR before the plant gene's polyadenylation signal (possibility IIIB in Fig. 7.1). Due to the presence of two translation starts, possibility IIIA results in a dicistronic transcript encoding a truncated plant protein and GUS, while in possibility IIIB both the entire plant protein and GUS will be produced.

When the pGKB5 T-DNA inserts within an intron, the intron will not be spliced out because the transcription will stop in the T-DNA (possibility IV in Fig. 7.1 A). However, theoretically also in this case a dicistronic transcript can be formed resulting in the production of a truncated plant protein and the entire GUS enzyme. Although pGKB5 has been used successfully its effectiveness could probably be enhanced by several modifications in its T-DNA. This chapter describes the introduction of a plant intron in the coding sequence of the *GUS* marker gene, the optimization of the translational start site, and the introduction of intron splice sites upstream of the *GUS* start codon.

#### a) Introduction of an intron in the *GUS* gene

The T-DNA tagging vector pGKB5 does not contain an intron in the *GUS* marker gene. This implies that expression of *GUS* is also possible in the prokaryotic *Agrobacterium tumefaciens* that is used for the transformation of plants. The presence of bacteria containing the intronless *GUS* therefore can cause background GUS activity, which may complicate the analysis of transgenic plants. Prokaryotes do not possess the eukaryotic intron splicing apparatus and the introduction of an intron in the *GUS* marker gene thus avoids expression in bacteria. In addition, the presence of one or more introns in a gene construct has been found to increase the accumulation of mRNA and proteins in plants compared to similar fusions that lack introns (reviewed in Koziel *et al.*, 1996). For example, introduction of the *GUS* gene into the third exon of the *Arabidopsis PAT1* gene (encoding the enzyme phosphoribosylanthranilate transferase) increased GUS activity in transgenic *Arabidopsis* about 5-fold compared to plants containing the *GUS* cloned into the first *PAT1* exon (encoding a primary transcript lacking an intron sequence) (Rose and Last, 1997). The introduction of an intron in the promoter-less *GUS* gene used for promoter trapping thus may increase its expression level facilitating the discovery of genes expressed at low levels. This modification would increase the effectiveness for all four types of gene fusion shown in Fig. 7.1.

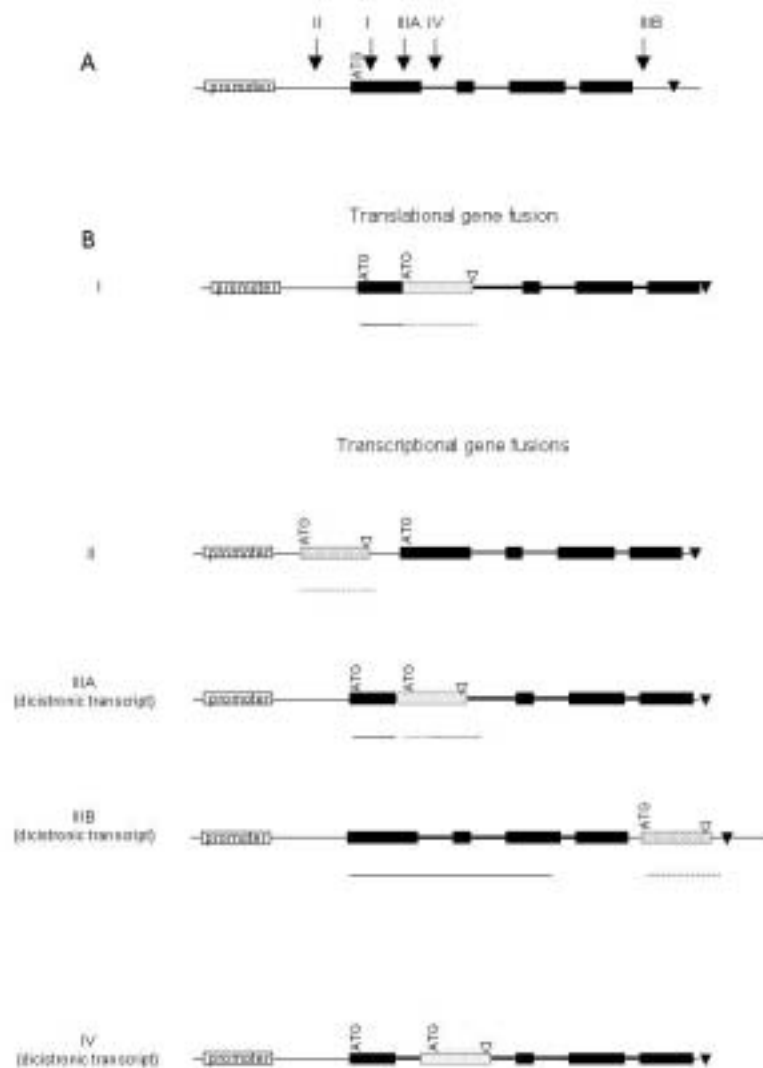


Figure 7.1 Theoretical gene fusions generated by insertional mutagenesis with a T-DNA carrying a promoter-less *GUS* gene. **A:** A plant gene with exons (boxes) and introns (thick lines) showing target sites of putative T-DNA insertions. **B:** Translational and transcriptional gene fusions yielding *GUS* expression resulting from insertions at the insertion sites shown in A. The black and white triangles represent the polyadenylation signal sequences of the plant and *GUS* gene, respectively. The dotted box represents the promoter-less *GUS* gene, dotted lines represent the *GUS* enzyme and lines represent the plant protein.

#### b) Modification of the sequence around the *GUS* AUG start codon

The initiation of translation of most eukaryotic mRNAs proceeds by the ribosome scanning model (reviewed in Kozak, 1989b). In this model, the 40S ribosomal subunit with associated factors (the preinitiation complex) engages the mRNA at or near the 5' cap and scans linearly in a 5'-to-3' direction for an AUG codon. When a codon in a favorable Kozak sequence context is encountered, the 60S subunit joins the 40S complex and initiates protein synthesis. An AUG codon in an unfavorable context may be bypassed by the preinitiation complex by a process called leaky scanning (Kozak, 1989a). The sequence around the AUG start codon of *GUS* in pGKB5 T-DNA is not optimal for translation in eukaryotes and therefore, modification of this sequence may improve the efficiency of translation in the case of transcriptional gene fusions (possibilities II, IIIA, IIIB and IV in Fig. 7.1).

#### c) Introduction of 3' intron splice sites upstream of the *GUS* AUG start codon

The introduction of 3' intron splice sites upstream of the start codon of the reporter gene may greatly enhance the efficiency of gene detection (Skarnes, 1990), because upon insertion in an intron the sequence between the 5' splice site of the plant and the introduced 3' splice site will form a functional intron that will be spliced out and the *GUS* can be fused to the host protein (Fig. 7.2 A). To guarantee a transcript that carries the *GUS* coding region in frame with that of the tagged gene, alternative 3' splice sites in the three reading frames are essential. In addition, to obtain a translational gene fusion no stop codons should be present between the 3' splice sites and



the start codon of *GUS*. In pGKB5 T-DNA the introduction of 3' splice sites upstream of the *GUS* ATG start codon may enhance the tagging efficiency when the T-DNA inserts within an intron (possibility IV in Fig. 7.1).

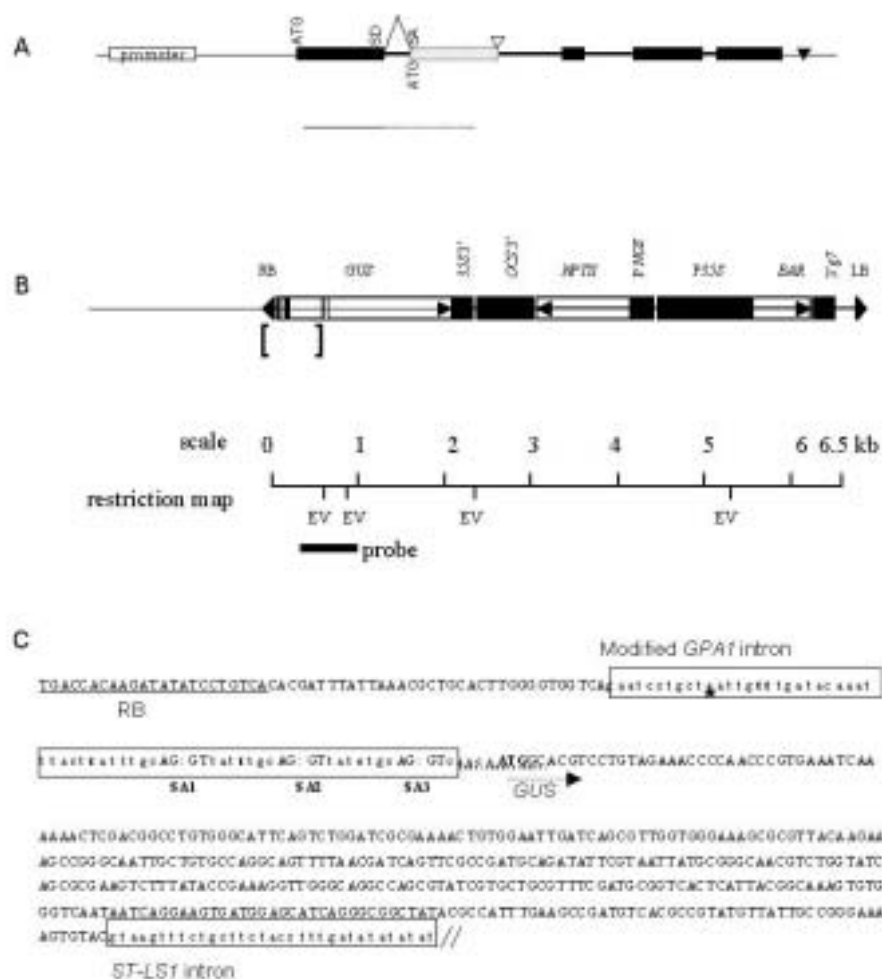


Figure 7.2 The tagging vector pMS1 **A**: Insertion of the pMS1 T-DNA in a plant intron resulting in a gene fusion between the plant gene and the *GUS* gene. Black boxes are plant exons, thick lines are plant introns, the dotted box is the *GUS* gene, the dotted line represents the *GUS* enzyme and the line represents the truncated plant protein. SD = 5' intron splice site, SA = 3' intron splice site. Black and white triangles represent the polyadenylation signals of the plant and *GUS* gene, respectively. **B**: Functional map of the pMS1 T-DNA. Arrows indicate coding sequences, and black boxes promoter/terminator regions. *GUS*: coding region of the  $\beta$ -glucuronidase gene from *E. coli*; *35S* 3': 3' region of the *35S* gene of the Cauliflower Mosaic Virus, *OCS* 3': 3' region of the octopine synthase gene from pTiAch5; *NPTII*: neomycin phosphotransferase II; *PNOS*: promoter region of the nopaline synthase gene; *P35S*: promoter of the *35S* gene of the Cauliflower Mosaic Virus; *BAR*: coding sequence of the Basta resistance gene from *Streptomyces hygroscopicus*; 3' *G7*: 3' region of the gene 7 from the T-DNA of pTi15955. Below the functional pMS1 T-DNA map the restriction map of the pMS1 T-DNA showing the restriction sites of the enzymes used for digestions of genomic DNA for Southern blot hybridizations. The black bar indicates the *GUS* probe. **C**: Sequence corresponding to the region between brackets in B. RB: T-DNA right border sequence. SA1, SA2 and SA3: intron splice sites. The modified *GPAl* intron and the *ST-LS1* intron are boxed. The star indicates a plant branch point consensus. Nucleotides that were changed in comparison to the modified *GPAl* intron sequence used by Sundaresan *et al.* (1995) are marked in bold and the changed sequence surrounding AUG is underlined with a dotted line.

## 7.2 Results

### 7.2.1 Construction of pMS1

To improve the expression of the *GUS* reporter gene and to avoid expression of *GUS* in *Agrobacterium* the second intron of the potato *ST-LS1* gene (Eckes *et al.*, 1986; Vancanneyt *et al.* 1990) was introduced into the *GUS* coding region of the tagging vector pGKB5 (Bouchez *et al.*, 1993). To adapt the sequence around the *GUS* AUG start codon to that of plants, the original sequence CCUUAUGUU was changed to AACAAUGGC (Joshi

*et al.*, 1997). Finally, part of the fourth intron of the *Arabidopsis* G-protein gene *GPA1* (Ma *et al.*, 1990), containing its 3' splice site and two alternative, artificial 3' splice sites (Sundaresan *et al.*, 1995) was modified by the replacement of three stop codons and three ATG codons by other codons (see material and methods). This intron construct was placed upstream of the *GUS* start codon (Figs. 7.2 B and C).

### 7.2.2 The intron-containing *GUS* is not expressed in *Agrobacterium*

To test whether the introduction of the intron in the *GUS* reporter gene indeed inhibited *GUS* expression in *Agrobacterium*, 10 µl of overnight *Agrobacterium* cultures containing either the original or the modified tagging vector were incubated overnight at 37°C in a buffer containing X-Gluc. The *Agrobacterium* carrying the intronless *GUS* showed strong GUS staining, whereas the *Agrobacterium* carrying the *GUS*-intron showed no GUS staining at all (data not shown). Introduction of the intron thus abolished *GUS* expression in *Agrobacterium*.

### 7.2.3 The plant intron splice machinery uses at least two of the introduced 3' splice sites

In order to test whether the three introduced 3' splice sites were recognized as such, the *GUS* with the modified upstream sequence was introduced into the sixth intron of the gene *AtARF* (Regad *et al.*, 1993), yielding the test vector pISV2678/35S35S::*AtArf*-*GUS*-new (Fig. 7.3 A). This gene was chosen because of its availability in the laboratory and the presence of restriction sites that were convenient for the construction of this test vector. Constitutive expression of the *AtARF*-gus-new fusion from the double *CAMV* 35S promoter of the binary vector pISV2678 in somatic pro-embryos of *M. truncatula*, obtained by *A. tumefaciens* mediated transformation, resulted in numerous *GUS*-expressing somatic-pro embryos (data not shown). This indicated that the intron introduced upstream of the *GUS* start codon was spliced out efficiently. To determine if besides the mRNA that contained the translational gene fusion between *AtARF* and *GUS* (possibility 2 in Fig. 7.3 B), also the other two possible mRNAs were formed (possibilities 1 and 3 in Fig. 7.3 B), junctions between *AtARF* and *GUS* were amplified by PCR on cDNA from *GUS*-expressing pro-somatic embryos using the primer pair RBGUS2/AtARF. The PCR fragments were subcloned in the vector pGEMTeasy and sequenced. In 15 out of 27 cDNA fragments the 3' splice site 2 had been used and in the 13 others splice site 1. In none of the sequenced cDNA fragments the 3' splice site 3 had been used.

The effect on the *GUS* expression level of the changed sequence around the start codon of the *GUS* and the introduction of an intron in the *GUS* coding sequence of pMS1 were not determined specifically, but surely had no negative effect, since the *GUS* was expressed correctly.

### 7.2.4 The use of pMS1 for T-DNA insertional mutagenesis

In order to test if insertion of the T-DNA of the improved tagging vector in the plant genome can result in gene fusions between a plant gene and *GUS*, primary transformants of *M. truncatula* carrying the pMS1 T-DNA were tested for GUS activity in the roots and leaves. A total of 124 primary transformants originating from 27 leaf explants were obtained. From these 27 leaf explants eleven (= 41%) yielded a total of 25 GUS-positive plants. From the 111 leaf explants used for the transformations with the pGKB5 T-DNA 17 (= 15%) yielded GUS-positive plants (chapter 3). Although in both tagging experiments different plant organs were tested for GUS staining (roots and leaves for the pMS1 T-DNA tagged plants and root and nodules for the pGKB5 T-DNA tagged plants) this preliminary experiment indicates that the efficiency of the improved tagging vector pMS1 was about 2.7-fold higher than that of the pGKB5 T-DNA.

Since this transformation experiment aimed at testing the tagging vector rather than at obtaining a large number of independent transformants, no special care was taken to avoid the regeneration of siblings. If several plants originating from the same explant showed GUS staining, in all cases the observed staining pattern was similar, indicating that the tested plants were siblings. Table 7.1 gives an overview of the observed staining patterns per leaf explant.

Table 7.1 GUS staining in primary transformants carrying the T-DNA of pMS1.

Explants	Organ
A11, B2	Root tip
A12, A13, A17, A19	Leaves
A16, B10	Root
B6, B8, B11	Root + leaves
A22	not tested

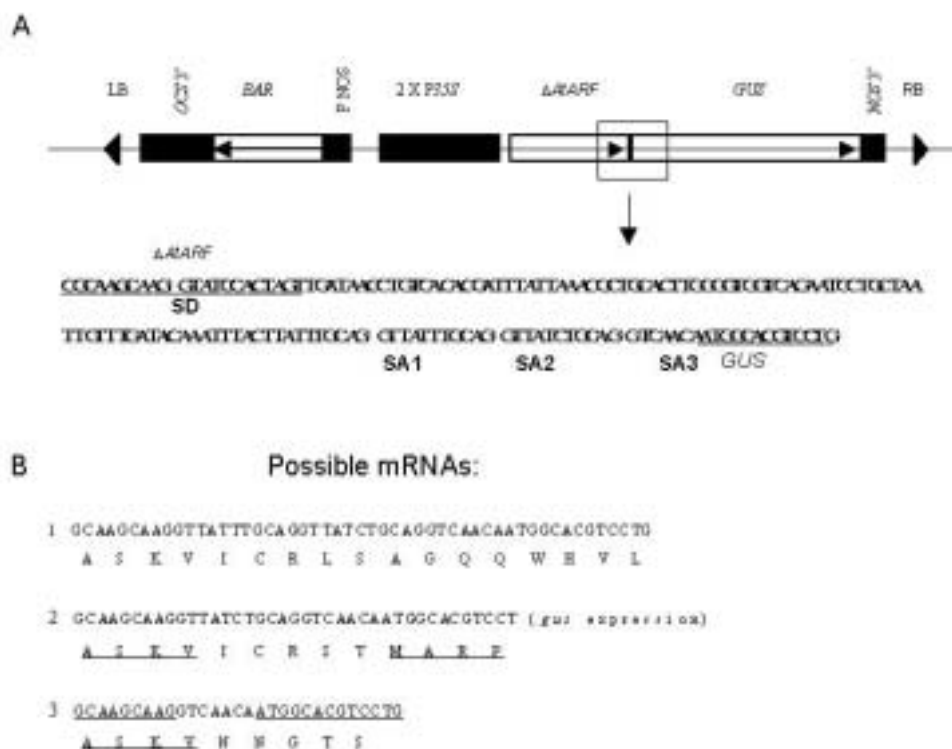


Figure 7.3 The test vector pISV2689/35S35S::*AtARF-GUSnew*. **A:** Functional map of the pISV2689/35S35S::*AtARF-GUSnew* T-DNA. Arrows indicate coding sequences, and black boxes promoter/terminator regions. *GUS*: coding region of the  $\beta$ -glucuronidase (*GUS*) gene from *E. coli*, *NOS* 3': 3' region of the nopaline synthase gene from pTiC58; *OCS* 3': 3' region of the octopine synthase gene from pTiAch5; 2 X P35S: two promoters of the 35S gene of the Cauliflower Mosaic Virus; *BAR*: coding sequence of the Basta resistance gene from *Streptomyces hygroscopicus*. The sequence corresponding to the boxed region is shown below the pISV2689/35S35S::*AtARF-GUSnew* map. SD: 5'intron splice site of the fifth intron from the gene *AtARF*, SA1, SA2 and SA3: 3'intron splice sites present in the modified *GPA1* intron. Underlined sequences correspond to *AtARF* and *GUS* **B:** Putative mRNAs and deduced amino acid sequences obtained after excision of the intron sequence using 3'splice sites SA1, SA2 or SA3. Only mRNA 2 results in a translational gene fusion between *AtARF* and *GUS*.

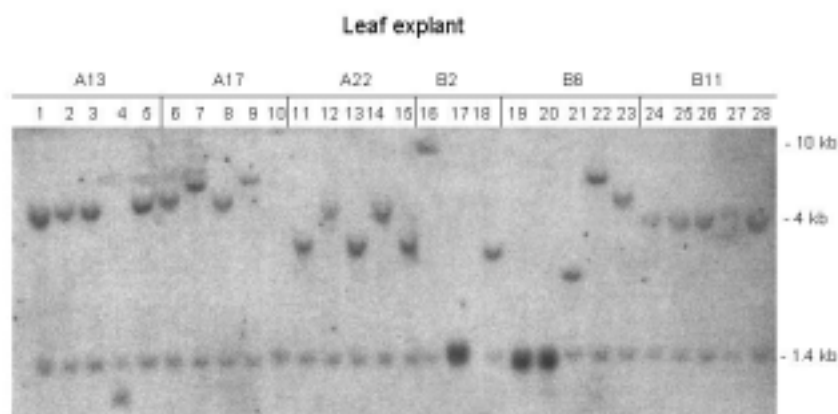


Figure 7.4 Southern blot hybridizations of genomic DNA of pMS1 T-DNA carrying plants digested with EcoRV. The sequence corresponding to the *GUS* probe is shown in Figure 7.2 A.

To determine if plants originating from the same explant were siblings or not and to analyze the T-DNA copy number and integration pattern, Southern blot analysis was carried out with primary transformants originating from explants A13, A17, A22, B2, B6 and B11 using a probe recognizing the *GUS* gene (Figs. 7.2 B and 7.4). The coding sequence of the *GUS* gene has two EcoRV restriction sites and there is one EcoRV site just downstream from the *GUS* gene. Therefore, when the DNA was digested with EcoRV this probe revealed 2

internal fragments of the T-DNA (230 and 1400 bp) and one fragment containing variable lengths of plant genomic DNA flanking the T-DNA right border. The Southern blot analysis showed that in all cases a single T-DNA was inserted into the plant genome and that often plants originating from the same explant appeared to be siblings (Fig. 7.4).

## 7.3 Discussion

### 7.3.1 Construction of an improved T-DNA tagging vector

This chapter described the introduction of an intron in the *GUS* reporter gene, the introduction of 3' intron splice sites in front of the *GUS* start codon and the modification of the sequence around the *GUS* AUG start codon in order to adapt it for plants. The introduction of an intron in the *GUS* of pMS1 T-DNA indeed eliminated *GUS* expression in *A. tumefaciens*. Also the introduction of 3' intron splice sites upstream of the *GUS* start codon seemed successful, for 3' splice sites 1 and 2 (SA1 and SA2, Fig. 7.3 A) both had been used to splice out the artificial intron between the fifth *AtARF* exon and the *GUS*. Although in *M. truncatula* splice site 3 (SA3, Fig. 7.3 A) was not used by the splice machinery for excision of the intron at a frequency equal or higher than 1/27, the tagging efficiency still will increase since the presence of the other two splice sites may provide the appropriate intron splice site to yield, together with an 5' intron splice site from the tagged gene a fusion between a plant gene and the *GUS* when the T-DNA inserts in an intron. For *Arabidopsis*, tobacco and rice, using the same intron sequence, it was also found that the third 3' intron splice site was not used (Nussaume *et al.*, 1995; Chin *et al.*, 1999). It was hypothesized that SA3 is not accessible for the splice machinery, because the distance between the branchpoint sequence and this splice site is too large. Branchpoints are positioned usually 18-40 nucleotides upstream of the splice site (reviewed in Brown and Simpson, 1998), while the distance between SA3 and the sequence TGCTAAT, corresponding to a plant branchpoint consensus, in the used modified *GPAL* intron is 50 nucleotides (Figure 7.2 C). Therefore, decreasing the distance between the branch point and the fused splice sites may further increase the efficiency of the tagging vector.

### 7.3.2 The use of pMS1 for T-DNA tagging of *M. truncatula*

In a small scale experiment different *GUS* expression patterns were obtained in *M. truncatula* plants that were stably transformed with the pMS1 T-DNA indicating that fusions between plant genes and the *GUS* gene were formed. From 27 leaf explants used for the transformation experiment 11 (= 41 %) yielded independent *GUS*-positive plants. This is roughly 2.7-fold more than obtained with the tagging vector pGKB. Also in other gene tagging studies the introduction of an artificial intron with splice sites in front of the promoter-less *GUS* gene resulted in an increase in tagging efficiency. For example in rice this modification roughly doubled the number of *GUS*-positive plants (from ca. 1 % to 2 %; Jeon *et al.*, 2000) and in *Arabidopsis* even 5-fold more *GUS*-positive plants (from 4 % to 20 %) were obtained (Bechtold *et al.*, 1993; Sundaresan *et al.*, 1995).

Previously (chapter 3) it was reported that the majority of the plants transformed with the pGKB5 T-DNA contained a single T-DNA insertion, which should facilitate the characterization of mutants and isolation of the sequences flanking the T-DNA. Southern blot analysis of pMS1 T-DNA carrying plants showed that also this T-DNA inserted preferentially as a single insert and the modifications introduced into the T-DNA thus did not change this advantageous feature.

### 7.3.3 Concluding remarks

Based upon the promising results described in this chapter and those obtained with the tagging vector pGKB5 (chapters 2-6) it was decided that the improved tagging vector pMS1 would be used for T-DNA tagging in *M. truncatula*. Currently a tagging project using, among others, the promoter tag vector pMS1 has been started in the framework of the European project 'Integrated structural, functional and comparative genomics of the model legume *Medicago truncatula*' in order to obtain a large *M. truncatula* T-DNA insertional mutant collection to be used for the identification of genes involved in endosymbiosis (both rhizobial and mycorrhizal) and other processes. So far, 250 independent T-DNA tagged lines were obtained and the goal is that at the end of the project the mutant collection will contain 2000 T-DNA tagged lines.

## 7.4 Materials and methods

### 7.4.1 Molecular techniques

Standard DNA manipulations were according to Sambrook *et al.* (1989). Genomic DNA was extracted from leaves according to Dellaporta *et al.* (1983) and digested with EcoRV for Southern analysis that was carried out as described in chapter 3. For the detection of the T-DNA a 702 bp PCR fragment (nucleotides 240 to 942 downstream of the *GUS* translation initiation site; Jefferson *et al.*, 1986) amplified using the oligonucleotides 5' GGCCAGCGTATCGTGCTGCG3' and 5'ATCGGGAGCGGCGATAACCGTA3' from the vector pGKB5 (Bouchez *et al.*, 1993) was used.

RNA from *M. truncatula* was extracted according to the method described in Kay *et al.* (1987). For RT-PCR, cDNAs were synthesized by reverse transcription of 5 µg total RNA with the Superscript reverse transcriptase kit (Gibco-BRL/Life Technologies, Gaithersburg, MD) following the manufacturer's instructions. The vector pGEM-T easy (Promega) was used for cloning of the PCR fragments obtained on the cDNA of the  $\div$ AtARF-new*GUS* fusion with the primers AtARF 5'GTGTGGGATGTTGGGGGTCA3' and RBGUS2 5'TCACGGGTTGGGGTTTCTACAGGAC3'.

### 7.4.2 Construction of pMS1

To facilitate subsequent cloning steps the origin of replication pRiA4, which is cloned upstream of the T-DNA in the tagging vector pGKB5 and contains numerous restriction sites, was removed by digestion with NotI and religation, yielding the pRiA4-less pGKB5 (Fig. 7.5 A). Next, the XbaI-EcoRI fragment containing the RB sequence and the *GUS* was cloned in pBluescript (Stratagene), yielding the vector pBS/*GUS* (Fig. 7.5 B). pBS/*GUS* was used as the basis for the modification of the AUG consensus, the introduction of 3' intron splice sites upstream of the *GUS*, and the introduction of an intron in the *GUS*.

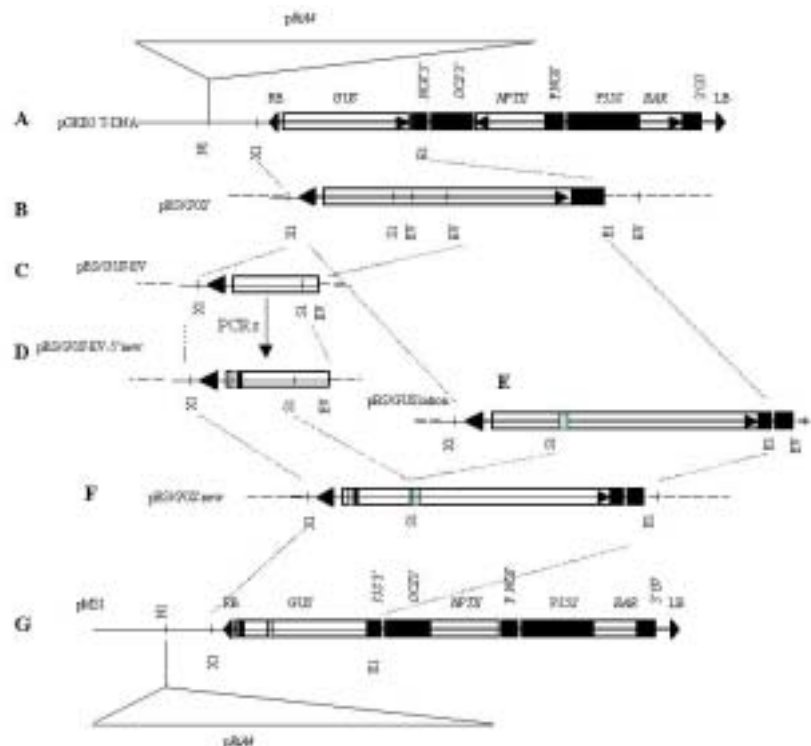


Figure 7.5 Schematic overview of the construction of pMS1. N1=NotI, X1=XbaI, E1=EcoRI, EV=EcoRV, S1=SnaBI. See text for explanation.

The adaptation for plants of the sequence around the *GUS* AUG start codon and the introduction of intron splice sites were carried out by PCR approaches using primers containing the modified sequence and the pBluescript primers T3 and T7. In order to reduce the distance between the region to be modified around the

*GUS* start codon and the T3 primer, 1570 bp were removed from pBS/*GUS* by digestion with EcoRV followed by religation yielding the vector pBS/*GUS*-EV (Fig. 7.5 C).

Two PCR reactions using the primer pairs T7/AUG- and T3/AUG+ were carried out to produce two fragments containing the modified sequence around the AUG. The 3' ends of the primers AUG- (5'TACAGGACGTGCCATTGTTTACTGACCACC3') and AUG+ (5'GTGGTCAGTAAACAATGGCACGTCCTGTAG3') were complementary to the pBS/*GUS*-EV template, whereas the 5' ends contained the modified sequence and were complementary to each other. In a second PCR reaction the overlapping PCR fragments were then fused and the resulting product amplified with T7 and T3 primers.

In a similar way the intron splice sites were introduced using the PCR fragment containing the changed AUG surrounding sequence as a template. Instead of AUG- and AUG+ the primers SA- (5'ACCTGCAAATAAGTAAATTTGTATCAAACAATTAGCAGGATTCTGACCACCCCAAGTGCA3') and SA+ (5'CAAATTTACTTATTTGCAGGTTATTTGCAGGTTATCTGCAGGTCAACAATGGCACGTCCT3') were used. Together, the 5' ends of these primers covered bp 54-119 of the fourth intron of the *Arabidopsis* G-protein gene *GPA1* (Ma *et al.*, 1990), modified by the addition of two consensus splice sequences (Sundaresan *et al.*, 1995). In order to ensure the expression of the *GUS* in case the T-DNA would insert in the right ORF and orientation within an exon, two stop codons in frame with the *GUS* start codon present in the modified *GPA1* intron fragment were eliminated by the replacement of tga by tgt and taa by tta and all the atg codons, except the *GUS* start codon were eliminated by replacement of atg by ttg, ttg and ctg respectively in order to advantage the use of the *GUS* start codon for the *GUS* expression (these nucleotide changes are underlined in Fig. 7.2 C). The PCR fragment with the modified AUG surrounding sequence and the 3' intron splice sites was cloned in pBS yielding the vector pBS/*GUS*-EV-5'new (Fig. 7.5 D). The fragments that were obtained by PCR were verified by sequencing.

In order to add an intron to the *GUS* gene originally present in pGKB5, the SnaB1-EcoR1 fragment from the vector pBS/*GUS* was replaced by the SnaB1-EcoR1 fragment from the *GUS* gene on the vector pPR96 (P. Ratet, unpublished), containing the 189 bp second intron of the potato *ST-LS1* gene (Eckes *et al.*, 1986) slightly modified as described by Vancanneyt *et al.* (1990), yielding the vector pBS/*GUS*-intron (Fig. 7.5 E).

The Xba1-SnaB1 fragment from pBS/*GUS*-EV-5'new and the SnaB1-EcoR1 fragment from pBS/*GUS*-intron were combined in pBS yielding pBS/*GUS*new (Fig. 7.5 F). The Xba-EcoR1 fragment of pBS/*GUS*new was then used to replace the Xba-EcoR1 fragment of pGKB5, before reinserting the pRiA4 fragment, yielding the improved T-DNA tagging vector pMS1 (Fig. 7.5 G).

### 7.4.3 Construction of the test vector pISV2689/35S35S::+*AtARF*-*GUS*new

For the construction of the test vector first a blunted Kpn1/Sst1 *AtARF* fragment was cloned in the EcoRV site of pBluescript and then cloned as a Sma1/Cla1 fragment behind the double *CAMV* 35S promoter of the binary vector pISV2689 (M. Schultze, unpublished). By PCR on pBS/*GUS*new using the primers splicetest 590+ 5'GCTCTAGACTAGTTGATAACCTGTCACACGATTTATTAAACG 3' and T3 a Spe1 site was introduced upstream of the new *GUS*. The PCR fragment was cloned in the EcoRV site of pBS and a Spe1/Cla1 fragment from this vector was used to replace the Spe1/Cla1 fragment from pISV2689/35S35S::+*AtARF* yielding the test vector pISV2689/35S35S::+*AtARF*-*GUS*new (Fig. 7.3 A).

### 7.4.4 Plant and bacterial material

*Medicago truncatula* (Gaertn.) line R108-1 (c3) was transformed according to Trinh *et al.* (1998) and cultured as described in chapter 3. For RNA extraction of cells expressing the test construct 35S35S::+*AtARF*-*GUS*new the transformed leaf explants were cultured until the pro-somatic embryo stages (5 weeks after vacuum infiltration of the leaf explants).

*Escherichia coli* strain DH5 $\zeta$  (Promega) was used for the amplification of the different constructs. *E. coli* was transformed by the use of the calcium chloride method as described by Sambrook *et al.* (1989) and grown at 37°C in liquid or on solid LB medium (Sambrook *et al.*, 1989) containing the appropriate antibiotics. *Agrobacterium tumefaciens* strain EHA105 (Hood *et al.*, 1993) was used to transform *M. truncatula*. *A. tumefaciens* was transformed with the binary constructs by electroporation as described by Mozo and Hooykaas (1991) and cultured at 30 °C in liquid or on solid YEB medium (Gartland and Davey, 1995) containing the appropriate antibiotics.

### 7.4.5 GUS enzymatic assays

For histological detection of the GUS activity transformed *M. truncatula* tissues were stained overnight at 37 °C in 2 mM X-gluc according to Trinh *et al.* (1998). Green tissues were cleared before observation by incubation in 50 % ethanol.

## 8 Concluding remarks

The aim of this thesis work was to generate and analyze a T-DNA tagged *M. truncatula* population in order to test T-DNA tagging for the isolation of plant genes involved in nodulation and the study of their functions in this model plant.

A low T-DNA copy number and a simple insertion pattern will facilitate the characterization of mutants and isolation of the sequences flanking the T-DNA. Molecular analysis of the T-DNA insertion sites of plants transformed with the binary vectors pPR97 (P. Ratet, unpublished), pSLJ8313, pSLJ8337 (Tissier *et al.*, 1999) and pGKB5 Bouchez *et al.*, 1993) showed that the T-DNAs of the pSLJ type and pGKB5 resulted mainly in single insertions without causing chromosomal rearrangements. The sequences of the isolated flanking regions of 21 lines that were tagged with the T-DNA of one of these vectors revealed that in 5 lines the T-DNA was inserted in the transcribed region of a gene and in 3 lines in the vicinity of a gene. T-DNA tagging thus can be used as a tool for gene discovery in *M. truncatula*.

The T-DNA of pGKB5 contained a promoter-less *GUS* reporter gene, which besides enabling studies on the expression pattern of the tagged gene, permits the discovery of lethal mutations or mutations that do not cause an easy detectable phenotype. In 19 out of 187 plants insertion of this T-DNA resulted in different, intriguing *GUS* expression patterns in the nodules. However, the *GUS* expression appeared to be the result of a gene fusion only in line B9-1. In this line, a translational gene fusion occurred between the *GUS* reporter gene and *MtN3*-like.

In line B9-1 homozygous insertion mutants did not occur and this gene probably would not have been found without the use of a promoter-less reporter gene to visualize the T-DNA insertion in heterozygous insertion mutants, that did not show any obvious phenotype.

In the other *GUS*-positive lines, the *GUS* staining was not due to obvious gene fusions or regulation by an upstream promoter. Using the pGKB5 T-DNA for T-DNA tagging thus could result in the discovery of new markers to study nodule development, but the selection of *GUS*-positive plants for the characterization of T-DNA flanking regions did not lead to the discovery of many novel *Medicago* genes.

For *M. truncatula*, it can be calculated that at least 120,000 independent tagged lines are required to tag 95% of the genes\*. The T-DNA insertions of the tagged population studied during this thesis work covered less than 1 % of the entire genome. This small number of tagged lines explains why we did not find nodulation phenotypes in our population. It is obvious that, using T-DNA as a mutagen, a tagged population saturating 95% of the genome with T-DNA insertions can not be obtained easily with the currently available transformation techniques. In contrast, mutagenesis techniques like irradiation or treatment with EMS can be used in a more efficient way to obtain saturating mutagenesis. The development of novel techniques such as TILLING (Targeting Induced Local Lesions IN Genomes; McCallum *et al.*, 2000) will facilitate the isolation of these mutations. However, at present the isolation of a mutation caused by a DNA-tag with a known sequence is still easier. The improvement of transformation techniques such as infiltration with *Agrobacterium* (Trieu *et al.*, 2000) or the use of transposons as a mutagen (I. d'Erfurth, unpublished) may facilitate the production of large numbers of independently tagged plants.

---

\* This number is calculated based on the following suppositions:

- 1) The T-DNA inserts preferentially in the euchromatic, gene-rich region of the genome.
- 2) In each tagged line one T-DNA inserts at random in this euchromatic region.
- 3) The euchromatic region is estimated to occupy about  $10^8$  bp, or 20% of the *M. truncatula* genome (Kulikova *et al.*, 2001). We assume that 1 insertion every 2500 bp. is needed to tag each gene and thus at least 40000 different insertions are needed.
- 4) The distribution of T-DNA insertions follows a Poisson distribution with:

$$P(\text{there are genes without insertion}) = e^{-\mu}$$

$$\text{and } P(\text{genes contain at least one insertion}) = 1 - e^{-\mu}$$

$$\text{where } \mu = (\text{mean number of insertions per genome} \times \text{the number of independent insertions}) / \text{potential genes number} = 1 \times n / 40000$$

$$\text{then the number of independent insertions (n) needed to tag 95 \% of the genes} = -\ln 0.05 \times 40000 = 119830$$

$$\text{and in the tagged population studied during this thesis work the T-DNA insertions cover } 1 - e^{-360/40000} = 8.9 \times 10^{-3}, = 0.9 \% \text{ of the genome-rich region.}$$

In addition, the value of the tagged population will increase when the plant lines will be screened for aberrant phenotypes or gene fusion expression during other processes than nodulation. For example, although no nodulation mutants were found in the small tagged population described in this thesis, in line B9-1 the T-DNA is inserted into a *MtN3*-like gene that may be involved in the transport of plant products (chapter 5). In addition, in line A7-1 the T-DNA disrupted a *NARF*-like gene that may be involved in the determination of plant size (chapter 6).

Thus T-DNA tagging is an efficient tool for the discovery of genes or the production of new markers in *M. truncatula*, but would be even more versatile when a larger mutant population would be available.



## 9 Bibliography

- Adams MW, Stiefel EI. 1998. Biological hydrogen production: not so elementary. *Science* 282:1842-3.
- Albrecht C, Geurts R, Bisseling T. 1999. Legume nodulation and mycorrhizae formation; two extremes in host specificity meet. *Embo J.* 18:281-8.
- Albrecht C, Geurts R, Lapeyrie F, Bisseling T. 1998. Endomycorrhizae and rhizobial Nod factors both require *SYM8* to induce the expression of the early nodulin genes *PsENOD5* and *PsENOD12A*. *Plant J.* 15:605-614.
- Allison LA, Kiss GB, Bauer P, Poirier M, Pierre M, Savoure A, Kondorosi E, Kondorosi A. 1993. Identification of two alfalfa early nodulin genes with homology to members of the pea *Enod12* gene family. *Plant Mol. Biol.* 21:375-80.
- Angenon G, Uotila J, Kurkela SA, Teeri TH, Botterman J, Van Montagu M, Depicker A. 1989. Expression of dicistronic transcriptional units in transgenic tobacco. *Mol. Cell. Biol.* 9:5676-84.
- Appleby CA. 1984. Leghemoglobin and *Rhizobium* respiration. *Annu. Rev. Plant Physiol.* 35:443-478.
- Ardourel M, Demont N, Debelle F, Maillet F, de Billy F, Prome JC, Denarie J, Truchet G. 1994. *Rhizobium meliloti* lipooligosaccharide nodulation factors: different structural requirements for bacterial entry into target root hair cells and induction of plant symbiotic developmental responses. *Plant Cell* 6:1357-74.
- Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J, Struhl K. 1989. *Current Protocols in Molecular Biology*. New York: Greene Publishing Associates and Wiley-Interscience.
- Barker D, Bianchi S, Blondon F, Dattée Y, Duc G, Essad S, Flament P, Gallusci P, Génier G, Guy P and others. 1990. *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*-Legume symbiosis. *Plant Mol Biol. Rep.* 8:40-49.
- Barton RM, Worman HJ. 1999. Prenylated prelamins A interacts with Narf, a novel nuclear protein. *J. Biol. Chem.* 274:30008-18.
- Bauer P, Poirier S, Ratet P, Kondorosi A. 1997. *MsEnod12A* expression is linked to meristematic activity during development of indeterminate and determinate nodules and roots. *Mol. Plant-Microbe Interact.* 10:39-49.
- Bauer P, Ratet P, Crespi M, Schultze M, Kondorosi A. 1996. Nod factor and cytokinins induce similar cortical cell divisions, amyloplast deposition and *MsEnod12A* expression patterns in alfalfa roots. *Plant J.* 10:91-105.
- Bechtold N, Pelletier G. 1998. *In planta Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Meth. Mol. Biol.* 82:259-266.
- Beck E, Ludwig G, Auerswald E, Reiss B, Schaller H. 1982. Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. *Gene* 19:327-336.
- Becker D, Kemper E, Schell J, Masterson R. 1992. New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol. Biol.* 20:1195-7.
- Benaben V, Duc G, Lefebvre V, Huguet T. 1995. *TE7*, an inefficient symbiotic mutant of *Medicago truncatula* (Gaertn.) cv Jemalong. *Plant Physiol.* 107:53-62.
- Bevan M. 1984. Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Res.* 12:8711-21.
- Bisseling T, Been C, Klugkist J, Van Kammen A, Nadler K. 1983. Nodule-specific host proteins in effective and ineffective root nodules of *Pisum sativum*. *Embo J.* 2:961-966.
- Blondon F, Marie D, Brown S, Kondorosi A. 1994. Genome size and base composition in *Medicago sativa* and *M. truncatula* species. *Genome* 37:264-270.
- Boisson-Dernier A, Chabaud M, Garcia F, Becard G, Rosenberg C, Barker DG. 2001. *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol. Plant-Microbe Interact.* 14:695-700.
- Bonfante P, Genre A, Faccio A, Martini I, Schauser L, Stougaard J, Web J, Parniske M. 2000. The *Lotus japonicus LjSym4* gene is required for the successful symbiotic infection of root epidermal cells. *Mol. Plant-Microbe Interact.* 13:1109-20.
- Bonnin I, Huguet T, Gherardi M, Prosperi J, I O. 1996. High level of polymorphism and spatial structure in a selfing plant species, *Medicago truncatula* (Leguminosae), shown using RAPD markers. *Am. J. Botany* 83:843-855.

- Bono JJ, Riond J, Nicolaou KC, Bockovich NJ, Estevez VA, Cullimore JV, Ranjeva R. 1995. Characterization of a binding site for chemically synthesized lipo-oligosaccharidic NodRm factors in particulate fractions prepared from roots. *Plant J.* 7:253-60.
- Boot KJM, van Brussel AA, Tak T, Spaik H, Kijne JW. 1999. Lipochitin oligosaccharides from *Rhizobium leguminosarum* bv. *viciae* reduce auxin transport capacity in *Vicia sativa* subsp. *nigra* roots. *Mol. Plant-Microbe Interact.* 12:839-844.
- Borisov AY, Rozov SM, Tsyganov VE, Morzhina EV, Lebsky VK, Tikhonovich IA. 1997. Sequential functioning of *Sym-13* and *Sym-31*, two genes affecting symbiosome development in root nodules of pea (*Pisum sativum* L.). *Mol. Gen. Genet.* 254:592-8.
- Bouchez D, Camilleri C, Caboche M. 1993. A binary vector based on Basta resistance for *in planta* transformation of *Arabidopsis thaliana*. *C.R. Acad. Sci. Paris* 316:1188-1193.
- Brewin NJ. 1998. Tissue and cell invasion by Rhizobium: the structure and development of infection threads and symbiosomes. In: Spaik H, Kondorosi A, Hooykaas P, editors. *Rhizobiaceae*. Dordrecht: Kluwer Academic Publisher. p 417-429.
- Brown JWS, Simpson CG. 1998. Splice site selection in plant pre-mRNA splicing. *Annu. Rev. Plant Physiol. Plant Mol. Biol. Genet* 49:77-95.
- Caetano-Anolles G, Bauer WD. 1988. Feedback regulation of nodule formation in alfalfa. *Planta* 175:546-557.
- Caetano-Anolles G, Gresshoff PM. 1990. Early induction of feedback regulatory responses governing nodulation in soybean. *Plant Sci.* 71:69-91.
- Caetano-Anolles G, Gresshoff PM. 1991. Plant genetic control of nodulation. *Annu. Rev. Microbiol.* 45:345-82.
- Campisi L, Yang Y, Yi Y, Heilig E, Herman B, Cassista AJ, Allen DW, Xiang H, Jack T. 1999. Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence. *Plant J.* 17:699-707.
- Catoira R, Galera C, de Billy F, Penmetsa RV, Journet EP, Maillet F, Rosenberg C, Cook D, Gough C, Denarie J. 2000. Four genes of *Medicago truncatula* controlling components of a nod factor transduction pathway. *Plant Cell* 12:1647-66.
- Catoira R, Timmers AC, Maillet F, Galera C, Penmetsa RV, Cook D, Denarie J, Gough C. 2001. The *HCL* gene of *Medicago truncatula* controls *Rhizobium*-induced root hair curling. *Development* 128:1507-18.
- Chang C, Shockey J. 1999. The ethylene-response pathway: signal perception to gene regulation. *Curr. Opin. Plant Biol.* 2:352-358.
- Charon C, Johansson C, Kondorosi E, Kondorosi A, Crespi M. 1997. *enod40* induces dedifferentiation and division of root cortical cells in legumes. *Proc. Natl. Acad. Sci. U S A* 94:8901-8906.
- Charon C, Sousa C, Crespi M, Kondorosi A. 1999. Alteration of *enod40* expression modifies *Medicago truncatula* root nodule development induced by *Sinorhizobium meliloti*. *Plant Cell* 11:1953-1965.
- Chin HG, Choe MS, Lee S-H, Park SH, Park SH, Koo JC, Kim NY, Lee JJ, Oh BG, Yi GH and others. 1999. Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system. *Plant J.* 19:615-623.
- Church GM, Gilbert W. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. U S A* 81:1991-5.
- Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16:735-43.
- Compaan B, Yang WC, Bisseling T, Franssen H. 2001. *ENOD40* expression in the pericycle precedes cortical cell division in *Rhizobium*-legume interaction and the highly conserved region of the gene does not encode a peptide. *Plant Soil* 230:1-8.
- Cook D. 1999. *Medicago truncatula*: A model in the making! *Curr. Opin. Plant Biol.* 2:301-304.
- Cook D, VandenBosch K, DeBruijn F, Huguet T. 1997. Model legumes get the Nod. *Plant Cell* 8:275-281.
- Cooper JB, Long SR. 1994. Morphogenetic rescue of *Rhizobium meliloti* nodulation mutants by *trans*-zeatin secretion. *Plant Cell* 6:215-225.
- Coronado C, Zuanazzi JAS, Sallaud C, Quirion J-C, Esnault R, Husson H-P, Kondorosi A, Ratet P. 1995. Alfalfa root flavonoid production is nitrogen regulated. *Plant Physiol.* 108:533-542.
- Crespi MD, Jurkevitch E, Poirer M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A. 1994. *enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *Embo J.* 13:5099-112.
- Davis SJ, Vierstra RD. 1998. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. *Plant Mol. Biol.* 36:521-8.
- Day RB, McAlvin CB, Loh JT, Denny RL, Wood TC, Young ND, Stacey G. 2000. Differential expression of two soybean apyrases, one of which is an early nodulin. *Mol. Plant-Microbe Interact.* 13:1053-70.
- De Buck S, De Wilde C, Van Montagu M, Depicker A. 2000. T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by *Agrobacterium*-mediated transformation. *Mol. Breeding* 6: 459-468.

- De Carvalho Niebel F, Lescure N, Cullimore JV, Gamas P. 1998. The *Medicago truncatula* *MtAnn1* gene encoding an annexin is induced by Nod factors and during the symbiotic interaction with *Rhizobium meliloti*. *Mol. Plant-Microbe Interact.* 11:504-513.
- De Lajudie P, Laurent-Fulele E, Willems A, Torck U, Coopman R, Collins MD, Kersters K, Dreyfus B, Gillis M. 1998. *Allorhizobium undicola* gen. nov., sp. nov., nitrogen-fixing bacteria that efficiently nodulate *Neptunia natans* in Senegal. *Int. J. Syst. Bacteriol.* 48:1277-90.
- Dellaporta SL, Wood J, Hicks JB. 1983. A plant DNA miniprep: Version II. *Plant Mol. Biol. Rep.* 1:19-21.
- Den Hartog M, Musgrave A, Munnik T. 2001. Nod factor-induced phosphatidic acid and diacylglycerol pyrophosphate formation: a role for phospholipase C and D in root hair deformation. *Plant J.* 25:55-65.
- Denarie J, Debelle F, Rosenberg C. 1992. Signaling and host range variation in nodulation. *Annu. Rev. Microbiol.* 46:497-531.
- Devic M, Albert S, Delseney M, Roscoe TJ. 1997. Efficient PCR walking on plant genomic DNA. *Plant Physiol. Biochem.* 35:331-339.
- Diaz CL, Melchers LS, Hooykaas P, Lugtenberg BJ, Kijne JW. 1989. Root lectin as a determinant of host-plant specificity in the *Rhizobium*-legume symbiosis. *Nature* 338:579-581.
- Ditta G, Stanfield S, Corbin D, Helinski DR. 1980. Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. U S A* 77:7347-51.
- Downie JA, Walker SA. 1999. Plant responses to nodulation factors. *Curr. Opin. Plant Biol.* 2:483-9.
- Doyle JJ. 1998. Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. *Trends in Plant Sci.* 3:473-478.
- Duc G, Messenger A. 1989. Mutagenesis of pea (*Pisum sativum* L.) and the isolation of mutants for nodulation and nitrogen fixation. *Plant Sci.* 60:207-213.
- Duc G, Trouvelot A, Gianinazzi-Pearson V, Gianinazzi S. 1989. First report on non-mycorrhizal plant mutants (Myc<sup>-</sup>) obtained in pea (*Pisum sativum* L.) and fababean (*Vicia faba* L.). *Plant Sci.* 60:215-222.
- Eckes P, Rosahl S, Schell J, Willmitzer L. 1986. Isolation and characterization of a light-inducible, organ-specific gene from potato and its expression after tagging and transfer into tobacco and potato shoots. *Mol. Gen. Genet.* 199:216-224.
- Ehrhardt DW, Atkinson EM, Long SR. 1992. Depolarization of alfalfa root hair membrane potential by *Rhizobium meliloti* nod factors. *Science* 256:998-1000.
- Ehrhardt DW, Wais R, Long SR. 1996. Calcium spiking in plant root hairs responding to *Rhizobium* nodulation signals. *Cell* 85:673-81.
- Endre G, Kereszt A, Kevei Z, Mihazea S, Kalo P, Kiss G. 2002. A receptor kinase gene regulating symbiotic nodule development. *Nature* 417:962-966.
- Etzler ME, Kalsi G, Ewing NN, Roberts NJ, Day RB, Murphy JB. 1999. A nod factor binding lectin with apyrase activity from legume roots. *Proc. Natl. Acad. Sci. U S A* 96:5856-61.
- Fang Y, Hirsch AM. 1998. Studying early nodulin gene *ENOD40* expression and induction by nodulation factor and cytokinin in transgenic alfalfa. *Plant Physiol.* 116:53-68.
- Fearn JC, LaRue TM. 1991. Ethylene inhibitors restore nodulation to *sym* 5 mutants of *Pisum sativum* L. cv Sparkle. *Plant Physiol* 96:239-244.
- Felle HH, Kondorosi E, Kondorosi A, Schultze M. 1995. Nod signal-induced plasma membrane potential changes in alfalfa root hairs are differentially sensitive to structural modifications of the lipochitoooligosaccharide. *Plant J.* 7:939-47.
- Felle HH, Kondorosi E, Kondorosi A, Schultze M. 1996. Rapid alkalization in alfalfa root hairs in response to rhizobial lipochitoooligosaccharide signals. *Plant J.* 10:295-301.
- Felle HH, Kondorosi E, Kondorosi A, Schultze M. 1998. The role of ion fluxes in Nod factor signalling in *Medicago sativa*. *Plant J.* 13:455-463.
- Felle HH, Kondorosi E, Kondorosi A, Schultze M. 2000. How alfalfa root hairs discriminate between Nod factors and oligochitin elicitors. *Plant Physiol.* 124:1373-80.
- Firmin JL, Wilson KE, Carlson RW, Davies AE, Downie JA. 1993. Resistance to nodulation of cv. Afghanistan peas is overcome by *nodX*, which mediates an O-acetylation of the *Rhizobium leguminosarum* lipooligosaccharide nodulation factor. *Mol. Microbiol.* 10:351-60.
- Florin L, Tsokoglou A, Happe T. 2001. A novel type of iron hydrogenase in the green alga *Scenedesmus obliquus* is linked to the photosynthetic electron transport chain. *J Biol Chem* 276:6125-32.
- Fobert PR, Labbe H, Cosmopoulos J, Gottlob-McHugh S, Ouellet T, Hattori J, Sunohara G, Iyer VN, Miki BL. 1994. T-DNA tagging of a seed coat-specific cryptic promoter in tobacco. *Plant J.* 6:567-77.
- Fobert PR, Miki BL, Iyer VN. 1991. Detection of gene regulatory signals in plants revealed by T-DNA-mediated fusions. *Plant Mol. Biol.* 17:837-51.
- Foster E, Hattori J, Labbe H, Ouellet T, Fobert PR, James LE, Iyer VN, Miki BL. 1999. A tobacco cryptic constitutive promoter, tCUP, revealed by T-DNA tagging. *Plant Mol. Biol.* 41:45-55.

- Foucher F, Kondorosi E. 2000. Cell cycle regulation in the course of nodule organogenesis in *Medicago*. Plant Mol. Biol. 43:773-86.
- Franssen JF, Nap JP, Gloude-mans T, Stiekema W, van Dam H, Govers F, Louwerse J, Van Kammen A, Bisseling T. 1987. Characterization of cDNA of nodulin-75 of soybean: A gene product involved in early stages of root nodule development. Proc. Natl. Acad. Sci. U S A 84:4495-4499.
- Frey M, Stettner C, Gierl A. 1998. A general method for gene isolation in tagging approaches: amplification of insertion mutagenised sites (AIMS). Plant J. 13:717-721.
- Furini A, Koncz C, Salamini F, Bartels D. 1997. High level transcription of a member of a repeated gene family confers dehydration tolerance to callus tissue of *Craterostigma plantagineum*. Embo J. 16:3599-608.
- Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P and others. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293:668-72.
- Gamas P, Niebel F, Lescure N, Cullimore J. 1996. Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. Mol. Plant-Microbe Interact. 9:233-42.
- Ge YX, Angenent GC, Dahlhaus E, Franken J, Peters J, Wullems GJ, Creemers-Molenaar J. 2001. Partial silencing of the *NECI* gene results in early opening of anthers in *Petunia hybrida*. Mol. Genet. Genomics 265:414-23.
- Ge YX, Angenent GC, Wittich PE, Peters J, Franken J, Busscher M, Zhang LM, Dahlhaus E, Kater MM, Wullems GJ and others. 2000. *NECI*, a novel gene, highly expressed in nectary tissue of *Petunia hybrida*. Plant J. 24:725-34.
- Geurts R, Heidstra R, Hadri A-E, Downie JA, Franssen H, Van Kammen A, Bisseling T. 1997. *Sym2* of pea is involved in a nodulation factor-perception mechanism that controls the infection process in the epidermis. Plant Physiol. 115:351-59.
- Glazebrook J, Ichige A, Walker GC. 1993. A *Rhizobium meliloti* homolog of the *Escherichia coli* peptide-antibiotic transport protein SbmA is essential for bacteroid development. Genes Dev. 7:1485-97.
- Goedhart J, Hink MA, Visser AJ, Bisseling T, Gadella TW, Jr. 2000. In vivo fluorescence correlation microscopy (FCM) reveals accumulation and immobilization of Nod factors in root hair cell walls. Plant J. 21:109-19.
- Goormachtig S, Valerio-Lepiniec M, Szczyglowski K, Van Montagu M, Holsters M, de Bruijn FJ. 1995. Use of differential display to identify novel *Sesbania rostrata* genes enhanced by *Azorhizobium caulinodans* infection. Mol. Plant-Microbe Interact. 8:816-24.
- Govers F, Gloude-mans T, Moerman M, Van Kammen A, Bisseling T. 1985. Expression of plant genes during the development of pea root nodules. Embo J. 4:861-867.
- Gressent F, Drouillard S, Mantegazza N, Samain E, Geremia RA, Canut H, Niebel A, Driguez H, Ranjeva R, Cullimore J and others. 1999. Ligand specificity of a high-affinity binding site for lipochitooligosaccharidic nod factors in *Medicago* cell suspension cultures. Proc. Natl. Acad. Sci. U S A 96:4704-9.
- Grevelding C, Fantes V, Kemper E, Schell J, Masterson R. 1993. Single-copy T-DNA insertions in *Arabidopsis* are the predominant form of integration in root-derived transgenics, whereas multiple insertions are found in leaf discs. Plant Mol. Biol. 23:847-60.
- Guinel FC, Sloetjes LL. 2000. Ethylene is involved in the nodulation phenotype of *Pisum sativum* R50 (*sym 16*), a pleiotropic mutant that nodulates poorly and has pale green leaves. J. Exp. Bot. 51:885-94.
- Gyorgyey J, Vaubert D, Jimenez-Zurdo JI, Charon C, Troussard L, Kondorosi A, Kondorosi E. 2000. Analysis of *Medicago truncatula* nodule expressed sequence tags. Mol. Plant-Microbe Interact. 13:62-71.
- Haseloff J, Siemering KR, Prasher DC, Hodge S. 1997. Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. Proc. Natl. Acad. Sci. USA 94:2122-7.
- Heidstra R, Geurts R, Franssen H, Spaik H, Van Kammen A, Bisseling T. 1994. Root hair deformation activity of nodulation factors and their fate on *Vicia sativa*. Plant Physiol. 105:787-797.
- Heidstra R, Nilsen G, Martinez-Abarca F, van Kammen A, Bisseling T. 1997b. Nod factor-induced expression of leghemoglobin to study the mechanism of  $\text{NH}_4\text{NO}_3$  inhibition on root hair deformation. Mol. Plant-Microbe Interact. 10:215-20.
- Heidstra R, Yang WC, Yalcin Y, Peck S, Emons AM, van Kammen A, Bisseling T. 1997a. Ethylene provides positional information on cortical cell division but is not involved in Nod factor-induced root hair tip growth in *Rhizobium*-legume interaction. Development 124:1781-7.
- Herman L, Jacobs A, Van Montagu M, Depicker A. 1990. Plant chromosome/marker gene fusion assay for study of normal and truncated T-DNA integration events. Mol. Gen. Genet. 224:248-56.
- Hirsch AM. 1992. Developmental biology of legume nodulation. New Phytol. 122:211-237.

- Hirsch AM. 1999. Role of lectins (and rhizobial exopolysaccharides) in legume nodulation. *Curr. Opin. Plant Biol.* 2:320-6.
- Hirsch AM, LaRue TA. 1997. Is the legume nodule a modified root or an organ *sui generis*? *Crit. Rev. in Plant Sc.* 16:361-392.
- Hoffman B, Trinh TH, Leung J, Kondorosi A, Kondorosi E. 1997. A new *Medicago truncatula* line with superior *in vitro* regeneration, transformation, symbiotic properties isolated through cell culture selection. *Mol. Plant-Microbe Interact.* 10:307-315.
- Hood EE, Gelvin SB, Melchters LS, Hoekema A. 1993. New *Agrobacterium* helper plasmids for gene transfer to plants. *Transgenic. Res.* 2:208-218.
- Horner DS, Foster PG, Embley TM. 2000. Iron hydrogenases and the evolution of anaerobic eukaryotes. *Mol. Biol. Evol.* 17:1695-709.
- Huang X. 1992. A contig assembly program based on sensitive detection of fragment overlaps. *Genomics* 14:18-25.
- Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD and others. 2000. Functional discovery via a compendium of expression profiles. *Cell* 102:109-26.
- Hutchison CJ, Bridger JM, Cox LS, Kill IR. 1994. Weaving a pattern from disparate threads: lamin function in nuclear assembly and DNA replication. *J. Cell Sci.* 107:3259-69.
- Imaizumi-Anraku H, Kawaguchi M, Koiwa H, Akao S, Syono K. 1997. Two ineffective-nodulating mutants of *Lotus japonicus*-different phenotypes caused by the blockage of endocytotic bacterial release and nodule maturation. *Plant Cell Physiol.* 38:871-881.
- Jacobs M, Rubery PH. 1988. Naturally occurring auxin transport regulators. *Science* 241:346-349.
- Jefferson R, Burgess S, Hirsh D. 1986. beta-glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc. Natl. Acad. Sci. USA* 83:8447-8451.
- Jeon JS, Lee S, Jung KH, Jun SH, Jeong DH, Lee J, Kim C, Jang S, Yang K, Nam J and others. 2000. T-DNA insertional mutagenesis for functional genomics in rice. *Plant J.* 22:561-70.
- Joshi CP, Zhou H, Huang X, Chiang VL. 1997. Context sequences of translational initiation codons in Plants. *Plant Mol. Biol.* 35:993-1001.
- Kalsi G, Etzler ME. 2000. Localization of a Nod factor-binding protein in legume roots and factors influencing its distribution and expression. *Plant Physiol.* 124:1039-48.
- Kamaté K, Rodriguez-Llorente ID, Scholte M, Durand P, Ratet P, Kondorosi E, Kondorosi A, Trinh TH. 2000. Transformation of floral organs with GFP in *Medicago truncatula*. *Plant Cell Reports* 19:647-653.
- Kannenberg EL, Brewin NJ. 1994. Host-plant invasion by *Rhizobium*: the role of cell-surface components. *Trends Microbiol.* 2:277-83.
- Kay R, Chan A, Daly M, McPerson J. 1987. Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. *Science* 236:1299-1302.
- Kieliszewski MJ, Lampert DT. 1994. Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J.* 5:157-72.
- Kijne JW. 1992. The *Rhizobium* infection process. In: Stacey G, Burris RH, Evans HJ, editors. *Biological Nitrogen Fixation*. New York: Chapman and Hall. p 349-398.
- Koncz C, Martini N, Mayerhofer R, Koncz-Kalman Z, Korber H, Redei GP, Schell J. 1989. High-frequency T-DNA-mediated gene tagging in plants. *Proc. Natl. Acad. Sci. USA* 86:8467-71.
- Kondorosi A, Barabas I, Svab Z, Orosz L, Sik T, Hotchkiss RD. 1973. Evidence for common genetic determinants of nitrogenase and nitrate reductase in *Rhizobium meliloti*. *Nat. New. Biol.* 246:153-4.
- Kondorosi E, Banfalvi Z, Kondorosi A. 1984. Physical and genetic analysis of a symbiotic region of *Rhizobium meliloti*: identification of nodulation genes. *Mol Gen. Genet.* 193:445-452.
- Kosslak RM, Bohloul BB. 1984. Suppression of nodule development of one side of a root system of soybeans caused by prior inoculation of the other side. *Plant Physiol.* 75:125-30.
- Kouchi H, Hata S. 1993. Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol Gen. Genet.* 238:106-19.
- Kozak M. 1989a. Context effects and inefficient initiation at non-AUG codons in eukaryotic cell-free translation systems. *Mol. Cell Biol.* 9:5073-5080.
- Kozak M. 1989b. The scanning model for translation-An update. *J. Cell Biol.* 229-241.
- Kozziel MG, Carozzi NB, Desai N. 1996. Optimizing expression of transgenes with an emphasis on post-transcriptional events. *Plant Mol. Biol.* 32:393-405.
- Kozik A, Heidstra R, Horvath B, Kulikova O, Tikhonovich I, Ellis THN, Van Kammen A, Lie TA, Bisseling T. 1995. Pea lines carrying *sym1* or *sym2* can be nodulated by *Rhizobium* strains containing *nodX*; *sym1* and *sym2* are allelic. *Plant Sci.* 108:41-49.
- Kulikova O, Gualtieri G, Geurts R, Kim D, Cook D, Huguet T, de Jong J, Fransz P, Bisseling T. 2001. Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. *Plant J.* 27:49-58.

- Laufs P, Autran D, Traas J. 1999. A chromosomal paracentric inversion associated with T-DNA integration in *Arabidopsis*. *Plant J.* 18:131-9.
- Leffell SM, Mabon SA, Stewart CN, Jr. 1997. Applications of green fluorescent protein in plants. *Biotechniques* 23:912-8.
- Legocki RP, Verma DP. 1980. Identification of 'nodule-specific' host proteins (nodulins) involved in the development of *Rhizobium*-legume symbiosis. *Cell* 20:153-63.
- Lerouge P, Roche P, Faucher C, Maillat F, Truchet G, Prome JC, Denarie J. 1990. Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* 344:781-4.
- Libbenga KR, Van Iren F, R.J. B, Schraag-Lamers MF. 1973. The role of hormones and gradients in the initiation of cortex proliferation and nodule formation in *Pisum sativum*. *Planta* 114:29-39.
- Lindsey K, Wei W, Clarke MC, McArdle HF, Rooke LM, Topping JF. 1993. Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants. *Transgenic Res.* 2:33-47.
- Liu YG, Mitsukawa N, Oosumi T, Whittier RF. 1995. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* 8:457-63.
- Ma H, Yanofsky MF, Meyerowitz EM. 1990. Molecular cloning and characterization of *GPA1*, a G protein subunit gene from *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 87:3821-3825.
- Markwei CM, LaRue TM. 1997. Phenotypic characterization of *sym21*, a gene conditioning shoot-controlled inhibition of nodulation in *Pisum sativum* cv. Sparkle. *Physiol. Plant.* 100:927-932.
- Martirani L, Stiller J, Mirabella R, Alfano F, Lamberti A, Radutoiu SE, Iaccarino M, Gresshoff PM, Chiurazzi M. 1999. T-DNA tagging of nodulation- and root-related genes in *Lotus japonicus*: expression patterns and potential for promoter trapping and insertional mutagenesis. *Mol. Plant-Microbe Interact.* 12:275-284.
- Mathesius U, Bayliss C, Weinman J, Schlaman HR, Spaink H, Rolfe BG, McCully M, Djordjevic MA. 1998a. Flavonoids synthesised in cortical cells during nodule initiation are early developmental markers in white clover. *Mol. Plant-Microbe Interact.* 11:1223-1232.
- Mathesius U, Schlaman HRM, Spaink HP, Sautter C, Rolfe BG, Djordjevic MA. 1998b. Auxin transport inhibition precedes root nodule formation in white clover roots and is regulated by flavonoids and derivatives of chitin oligosaccharides. *Plant J.* 14:23-34.
- Mathur J, Szabados L, Schaefer S, Grunenberg B, Lossow A, Jonas-Straube E, Schell J, Koncz C, Koncz-Kalman Z. 1998. Gene identification with sequenced T-DNA tags generated by transformation of *Arabidopsis* cell suspension. *Plant J.* 13:707-16.
- McCallum CM, Comai L, Greene EA, Henikoff S. 2000. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123:439-42.
- Meissner R, Chague V, Zhu Q, Emmanuel E, Elkind Y, Levy AA. 2000. Technical advance: a high throughput system for transposon tagging and promoter trapping in tomato. *Plant J.* 22:265-74.
- Miklashevichs E, Rohrig H, Schell J, Schmidt J. 2001. Perception and signal transduction of Rhizobial NOD factors. *Crit. Rev. Plant Sc.* 20:373-394.
- Mollier P, Montoro P, DeLarue M, Bechtold N, Bellini C, Pelletie G. 1995. Promoter-less *gusA* expression in a large number of *Arabidopsis thaliana* transformants obtained by the *in planta* infiltration method. *C.R. Acad. Sci. Paris* 318:465-475.
- Morzhina EV, Tsyganov VE, Borisov AY, Lebsky VK, Tikhonovich IA. 2000. Four developmental stages identified by genetic dissection of pea (*Pisum sativum* L.) root nodule morphogenesis. *Plant Sci.* 155:75-83.
- Mozo T, Hooykaas PJ. 1991. Electroporation of megaplasmids into *Agrobacterium*. *Plant Mol. Biol.* 16:917-8.
- Mueller P, Hynes MF, Kapp D, Niehaus K, Puehler A. 1988. Two classes of *Rhizobium meliloti* infection mutants differ in exopolysaccharide production and in coinoculation properties with nodulation mutants. *Mol. Gen. Genet.* 211:17-26.
- Muñoz JA, Coronado C, Perez-Hormaeche J, Kondorosi A, Ratet P, Palomares AJ. 1998. *MsPG3*, a *Medicago sativa* polygalacturonase gene expressed during the alfalfa-*Rhizobium meliloti* interaction. *Proc. Natl. Acad. Sci. USA* 95:9687-92.
- Muñoz JA, Palomares AJ, Ratet P. 1996. Plant genes induced in the *Rhizobium*-legume symbiosis. *World J. Microbiol. Biotechnol.* 12:189-202.
- Mylona P, Pawlowski K, Bisseling T. 1995. Symbiotic nitrogen fixation. *Plant Cell* 7:869-885.
- Nacry P, Camilleri C, Courtial B, Caboche M, Bouchez D. 1998. Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics* 149:641-50.
- Newcomb W. 1981. Nodule morphogenesis and differentiation. In: Giles KL, Atherly AG, editors. *International review of cytology*. New York: Academic Press. p 247-298.
- Nicolet Y, Piras C, Legrand P, Hatchikian CE, Fontecilla-Camps JC. 1999. *Desulfovibrio desulfuricans* iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. *Structure*

Fold. Des. 7:13-23.

- Niebel A, Bono JJ, Ranjeva R, Cullimore J. 1997. Identification of a high affinity binding site for lipooligosaccharidic NodRm factors in microsomal fraction of *Medicago* cell suspension cultures. *Mol. Plant-Microbe Interact.* 10:132-134.
- Nussaume L, Harrison K, Klimyuk V, Martienssen R, Sundaresan V, Jones JD. 1995. Analysis of splice donor and acceptor site function in a transposable gene trap derived from the maize element Activator. *Mol. Gen. Genet.* 249:91-101.
- Omer C, Lenstra R, Litle P, Dean C, Tepperman J, Leto K, Romesser J, O'Keefe D. 1990. Genes for two herbicide-inducible cytochromes P-450 from *Streptomyces griseolus*. *J. Bacteriol.* 172:3335-3345.
- Oparka K, Prior D, Crawford J. 1994. Behaviour of plasma membrane, cortical ER and plasmodesmata during plasmolysis of onion epidermal cells. *Plant Cell Environ.* 17:163-171.
- Pang SZ, DeBoer DL, Wan Y, Ye G, Layton JG, Neher MK, Armstrong CL, Fry JE, Hinchey MA, Fromm ME. 1996. An improved green fluorescent protein gene as a vital marker in plants. *Plant Physiol.* 112:893-900.
- Parinov S, Sevugan M, Ye D, Yang W, Kumuran M, Sundaresan V. 1999. Analysis of flanking sequences from Dissociation insertion lines: a database for reverse genetics in *Arabidopsis*. *Plant Cell* 11: 2263-2270.
- Pawlovski K, Bisseling T. 1996. Rhizobial and actinorhizal symbioses: what are the shared features? *Plant Cell* 8: 1899-1913.
- Penmetsa RV, Cook DR. 1997. A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. *Science* 275:527-530.
- Pérez-García A, de Vicente A, Canton FR, Cazorla FM, Codina JC, García-Gutiérrez A, Canovas FM. 1998. Light-dependent changes of tomato glutamine synthetase in response to *Pseudomonas syringae* infection or phosphinothricin treatment. *Physiol. Plant.* 102:377-384.
- Peters JW, Lanzilotta WN, Lemon BJ, Seefeldt LC. 1998. X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 angstrom resolution. *Science* 282:1853-8.
- Philip-Hollingsworth S, Dazzo FB, Hollingsworth RI. 1997. Structural requirements of *Rhizobium* chitolipooligosaccharides for uptake and bioactivity in legume roots as revealed by synthetic analogs and fluorescent probes. *J. Lipid Res.* 38:1229-41.
- Pichon M, Journet EP, Dedieu A, de Billy F, Truchet G, Barker DG. 1992. *Rhizobium meliloti* elicits transient expression of the early nodulin gene *ENOD12* in the differentiating root epidermis of transgenic alfalfa. *Plant Cell* 4:1199-211.
- Pingret JL, Journet EP, Barker DG. 1998. *Rhizobium* Nod factor signaling. Evidence for a g protein-mediated transduction mechanism. *Plant Cell* 10:659-72.
- Postma JG, Jacobsen E, Feenstra WJ. 1988. Three pea mutants with an altered nodulation studied by genetic analysis and grafting. *J. Plant Physiol.* 132:424-430.
- Postma JG, Jager D, Jacobsen E, Feenstra WJ. 1990. Studies on a non-fixing mutant of pea (*Pisum sativum* L.). I. phenotypical description and bacteroid activity. *Plant Sci.* 68:151-161.
- Pridmore 1987. New and versatile cloning vectors with kanamycin-resistance marker. *Gene* 56: 309-312
- Pueppke SG, Broughton WJ. 1999. *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Mol. Plant-Microbe Interact.* 12:293-318.
- Quandt HJ, Puhler A, Broer Y. 1993. Transgenic root nodules of *Vicia hirsuta*: a fast and efficient system for the study of gene expression in indeterminate-type nodules. *Mol. Plant-Microbe Interact.* 6:699-706.
- Quirino BF, Normanly J, Amasino RM. 1999. Diverse range of gene activity during *Arabidopsis thaliana* leaf senescence includes pathogen-independent induction of defense-related genes. *Plant Mol. Biol.* 40:267-78.
- Ratet P, Schell J, De Bruijn F. 1988. Mini-Mulac transposons with broad-host-range origins of conjugal transfer and replication designed for gene regulation studies in *Rhizobiaceae*. *Gene* 63:41-52.
- Regad F, Bardet C, Tremousaygue D, Moisan A, Lescure B, Axelos M. 1993. cDNA cloning and expression of an *Arabidopsis* GTP-binding protein of the ARF family. *FEBS Lett.* 316:133-6.
- Roche P, Debelle F, Maillet F, Lerouge P, Faucher C, Truchet G, Denarie J, Prome JC. 1991. Molecular basis of symbiotic host specificity in *Rhizobium meliloti*: nodH and nodPQ genes encode the sulfation of lipooligosaccharide signals. *Cell* 67:1131-43.
- Rose AB, Last R. 1997. Introns act post-transcriptionally to increase expression of the *Arabidopsis thaliana* tryptophan pathway gene *PAT1*. *Plant J.* 11:455-464.
- Roth LE, Stacey G. 1989. Bacterium release into host cells of nitrogen-fixing soybean nodules: the symbiosome membrane comes from three sources. *Eur. J. Cell Biol.* 49:13-23.
- Sagan M, Duc G. 1996. *Sym28* and *Sym29*, two new genes involved in regulation of nodulation in pea (*Pisum sativum* L.). *Symbiosis* 20:229-245.

- Sagan M, Huguet T, Barker D, Duc G. 1993. Characterization of two classes of non-fixing mutants of pea plants (*Pisum sativum* L.). Plant Sci. 95:55-66.
- Sagan M, Huguet T, Duc G. 1994. Phenotypic characterization and classification of nodulation mutants of pea (*Pisum sativum* L.). Plant Sci. 100:59-70.
- Sagan M, Morandi D, Tarengi E, Duc G. 1995. Selection of nodulation and mycorrhizal mutants in the model plant *Medicago truncatula* (Gaertn.) after v-ray mutagenesis. Plant Sci. 111:63-71.
- Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press.
- Schauser L, Handberg K, Sandal N, Stiller J, Thykjaer T, Pajuelo E, Nielsen A, Stougaard J. 1998. Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus*. Mol. Gen. Genet. 259:414-23.
- Schauser L, Roussis A, Stiller J, Stougaard J. 1999. A plant regulator controlling development of symbiotic root nodules. Nature 402:191-5.
- Scheres B, Van De Wiel C, Zalensky A, Horvath B, Spaink H, Van Eck H, Zwartkuis F, Wolters AM, Gloudemans T, Van Kammen A and others. 1990. The *ENOD12* gene product is involved in the infection process during the pea-*Rhizobium* interaction. Cell 60:281-94.
- Schlaman HR, Gisel AA, Quaadvlieg NE, Bloemberg GV, Lugtenberg BJ, Kijne JW, Potrykus I, Spaink HP, Sautter C. 1997. Chitin oligosaccharides can induce cortical cell division in roots of *Vicia sativa* when delivered by ballistic microtargeting. Development 124:4887-95.
- Schneider A, Walker SA, Poyser S, Sagan M, Ellis TH, Downie JA. 1999. Genetic mapping and functional analysis of a nodulation-defective mutant (*sym19*) of pea (*Pisum sativum* L.). Mol. Gen. Genet. 262:1-11.
- Schneider M, Ow DW, Howell SH. 1990. The in vivo pattern of firefly luciferase expression in transgenic plants. Plant Mol Biol 14(6):935-47.
- Schultze M, Kondorosi A. 1996. The role of lipochitooligosaccharides in root nodule organogenesis and plant cell growth. Curr. Opin. Genet. Dev. 6:631-8.
- Schultze M, Kondorosi A. 1998. Regulation of symbiotic root nodule development. Annu. Rev. Genet. 32:33-57.
- Schultze M, Kondorosi E, Ratet P, Buiré M, Kondorosi A. 1994. Cell and molecular biology of *Rhizobium*-plant interactions. Int. Rev. Cytol. 156:1-75.
- Schultze M, Quiclet-Sire B, Kondorosi E, Virelizer H, Glushka JN, Endre G, Gero SD, Kondorosi A. 1992. *Rhizobium meliloti* produces a family of sulfated lipooligosaccharides exhibiting different degrees of plant host specificity. Proc. Natl. Acad. Sci. USA 89:192-6.
- Scott A, Wyatt S, Tsou PL, Robertson D, Allen NS. 1999. Model system for plant cell biology: GFP imaging in living onion epidermal cells. Bio. Techniques 26:1125-1132.
- Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. 1995. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. 23:1087-8.
- Skarnes WC. 1990. Entrapment vectors: a new tool for mammalian genetics. Biotechnology (N Y) 8:827-31.
- Smit G, de Koster CC, Schripsema J, Spaink HP, van Brussel AA, Kijne JW. 1995. Uridine, a cell division factor in pea roots. Plant Mol. Biol. 29:869-73.
- Smith D, Yanai Y, Liu YG, Ishiguro S, Okada K, Shibata D, Whittier RF, Fedoroff NV. 1996. Characterization and mapping of Ds-GUS-T-DNA lines for targeted insertional mutagenesis. Plant J. 10:721-32.
- Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc. Natl. Acad. Sci. USA 92:2647-51.
- Spaink HP, Wijffelman CA, Pees E, Okker RJH, Lugtenberg BJJ. 1987. *Rhizobium* nodulation gene *nodD* as a determinant of host specificity. Nature 328:337-340.
- Springer PS. 2000. Gene traps: tools for plant development and genomics. Plant Cell 12:1007-20.
- Staehelin C, Charon C, Boller T, Crespi M, Kondorosi A. 2001. *Medicago truncatula* plants overexpressing the early nodulin gene *enod40* exhibit accelerated mycorrhizal colonization and enhanced formation of arbuscules. Proc. Natl. Acad. Sci. USA 98:15366-71.
- Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szczyglowski K and others. 2002. A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature 417:959-962.
- Sundaresan V. 1996. Horizontal spread of transposon mutagenesis: new uses for old elements. Trends Plant Sci. 1:184-190.
- Sundaresan V, Springer P, Volpe T, Haward S, Jones JD, Dean C, Ma H, Martienssen R. 1995. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev. 9:1797-810.
- Szabados L, Charrier B, Kondorosi A, De Bruijn FJ, Ratet P. 1995. New plant promoter and enhancer testing vectors. Mol. Breeding 1:419-423.



- Szczyglowski K, Hamburger D, Kapranov P, de Bruijn FJ. 1997. Construction of a *Lotus japonicus* late nodulin expressed sequence tag library and identification of novel nodule-specific genes. *Plant Physiol.* 114:1335-46.
- Szczyglowski K, Shaw RS, Wopereis J, Copeland S, Hamburger D, Kasiborski B, Dazzo FB, De Bruijn FJ. 1998. Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. *Mol. Plant-Microbe Interact.* 11:684-697.
- Takats ST. 1990. Early autoregulation of symbiotic root nodulation in soybeans. *Plant Physiol.* 94:865-869.
- Takken FL, Schipper D, Nijkamp HJ, Hille J. 1998. Identification and Ds-tagged isolation of a new gene at the Cf-4 locus of tomato involved in disease resistance to *Cladosporium fulvum* race 5. *Plant J.* 14:401-11.
- The Arabidopsis Genome Initiative. 2000. *Nature* 408:796-815.
- Thompson C, Movva N, Tizard R, Cramer R, Davies J, Lauwereys M, Botterman J. 1987. Characterization of the herbicide-resistance gene *bar* from *Streptomyces hygroscopicus*. *EMBO J.* 6:2519-2523.
- Timmers AC, Soupene E, Auriac MC, de Billy F, Vasse J, Boistard P, Truchet G. 2000. Saprophytic intracellular rhizobia in alfalfa nodules. *Mol. Plant-Microbe Interact.* 13:1204-13.
- Tinland B. 1996. The integration of T-DNA into plant genomes. *Trends Plant Sci.* 1:178-184.
- Tirichine L, de Billy F, Huguet T. 2000. *Mtsym6*, a gene conditioning *Sinorhizobium* strain-specific nitrogen fixation in *Medicago truncatula*. *Plant Physiol.* 123:845-51.
- Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JD. 1999. Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics. *Plant Cell* 11:1841-52.
- Trieu A, Burleigh S, Kardailsky I, Maldonado-Mendoza I, Versaw W, Blaylock L, Shin H, Chiou T, Katagi H, Dewbre G and others. 2000. Transformation of *Medicago truncatula* via infiltration of seedlings or flowering plants with *Agrobacterium*. *Plant J.* 22:531-541.
- Trinh TH, Ratet P, Kondorosi E, Durand P, Kamaté K, Bauer P, Kondorosi A. 1998. Rapid and efficient transformation of diploid *Medicago truncatula* and *Medicago sativa* ssp. *falcata* lines involved in somatic embryogenesis. *Plant Cell Reports* 17:345-355.
- Tsyganov VE, Morzhina EV, Stefanov SY, Borisov AY, Lebsky VK, Tikhonovich IA. 1998. The pea (*Pisum sativum* L.) genes *sym33* and *sym40* control infection thread formation and root nodule function. *Mol. Gen. Genet.* 259:491-503.
- Van Berkum P, Eardly BD. 1998. Molecular evolutionary systematics of the *Rhizobiaceae*. In: Spaink H, Kondorosi A, Hooykaas P, editors. *Rhizobiaceae*. Dordrecht: Kluwer Academic Publishers. p 1-24.
- Van Brussel AA, Bakhuizen R, van Spronsen PC, Spaink H, Tak T, Lugtenberg BJ, Kijne JW. 1992. Induction of pre-infection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of *Rhizobium*. *Science* 257:70-72.
- Van de Sande K, Pawlowski K, Czaja I, Wieneke U, Schell J, Schmidt J, Walden R, Matvienko M, Wellink J, van Kammen A and others. 1996. Modification of phytohormone response by a peptide encoded by *ENOD40* of legumes and a nonlegume. *Science* 273:370-3.
- Van de Wiel C, Scheres B, Franssen H, van Lierop MJ, van Lammeren A, van Kammen A, Bisseling T. 1990. The early nodulin transcript *ENOD2* is located in the nodule parenchyma (inner cortex) of pea and soybean root nodules. *Embo J.* 9:1-7.
- Van Kammen A. 1984. Suggested nomenclature for plant genes involved in nodulation and symbiosis. *Plant Mol. Biol. Reports* 2:43-45.
- Van Rhijn P, Goldberg RB, Hirsch AM. 1998. *Lotus corniculatus* nodulation specificity is changed by the presence of a soybean lectin gene. *Plant Cell* 10:1233-50.
- Van Spronsen PC, Bakhuizen R, van Brussel AA, Kijne JW. 1994. Cell wall degradation during infection thread formation by the root nodule bacterium *Rhizobium leguminosarum* is a two-step process. *Eur. J. Cell Biol.* 64:88-94.
- Vancanneyt G, Rosahl S, Willmitzer L. 1990. Translatability of a plant-mRNA strongly influences its accumulation in transgenic plants. *Nucleic Acids Res.* 18:2917-21.
- Vance C. 2001. Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. *Plant Physiol.* 127:390-397.
- Vasse J, de Billy F, Camut S, Truchet G. 1990. Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *J. Bacteriol.* 172:4295-306.
- Vijn I, Christiansen H, Lauridsen P, Kardailsky I, Quandt HJ, Broer I, Drenth J, Ostergaard Jensen E, van Kammen A, Bisseling T. 1995. A 200 bp region of the pea *ENOD12* promoter is sufficient for nodule-specific and nod factor induced expression. *Plant Mol. Biol.* 28:1103-10.
- Vlassak KM, Luyten E, Verreth C, van Rhijn P, Bisseling T, Vanderleyden J. 1998. The *Rhizobium* sp. BR816 nodO gene can function as a determinant for nodulation of *Leucaena leucocephala*, *Phaseolus vulgaris*, and *Trifolium repens* by a diversity of *Rhizobium* spp. *Mol. Plant-Microbe Interact.* 11:383-392.

- Wais RJ, Galera C, Oldroyd G, Catoira R, Penmetsa RV, Cook D, Gough C, Denarie J, Long SR. 2000. Genetic analysis of calcium spiking responses in nodulation mutants of *Medicago truncatula*. Proc. Natl. Acad. Sci. USA 97:13407-12.
- Walbot V. 1999. Genes, genomes, genomics. What can plant biologists expect from the 1998 National Science Foundation plant genome research program? Plant Physiol. 119:1151-1155.
- Walker SA, Viprey V, Downie JA. 2000. Dissection of nodulation signaling using pea mutants defective for calcium spiking induced by nod factors and chitin oligomers. Proc. Natl. Acad. Sci. USA 97:13413-8.
- Wang L, Wessler SR. 1998. Inefficient reinitiation is responsible for upstream open reading frame-mediated translational repression of the maize R gene. Plant Cell 10:1733-46.
- Webb KJ, Skot L, Nicholson MN, Jorgensen B, Mizen S. 2000. *Mesorhizobium loti* increases root-specific expression of a calcium-binding protein homologue identified by promoter tagging in *Lotus japonicus*. Mol. Plant-Microbe Interact. 13:606-16.
- Weeden NF, Kneen BE, LaRue TA. 1990. Genetic analysis of *sym* genes and other nodule-related genes in *Pisum sativum*. In: Gresshoff, Roth, Stacey, editors. Nitrogen fixation: Achievements and objectives. New York-London: Chapman and Hall. p 323-330.
- Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM and others. 2000. Activation tagging in *Arabidopsis*. Plant Physiol. 122:1003-13.
- Wilson KJ, Giller KE, Jefferson RA. 1991.  $\square$ -glucuronidase (GUS) operon fusion as a tool for studying plant-microbe interactions. In: Hennecke H, Verma DPS, editors. Advances in Molecular Genetics of Plant-Microbe Interactions. Dordrecht: Kluwer Academic publishers. p 226-229.
- Winzer T, Bail A, Linder M, Linder D, Werner D, Muller P. 1999. A novel 53-kDa nodulin of the symbiosome membrane of soybean nodules, controlled by *Bradyrhizobium japonicum*. Mol. Plant-Microbe Interact. 12:218-26.
- Wisman E, Cardon G, Fransz P, Saedler H. 1998. The behaviour of the autonomous maize transposable element *En/Spm* in *Arabidopsis thaliana* allows efficient mutagenesis. Plant Mol. Biol. 37:989-999.
- Wopereis J, Pajuelo E, Dazzo FB, Jiang Q, Gresshoff PM, De Bruijn FJ, Stougaard J, Szczyglowski K. 2000. Short root mutant of *Lotus japonicus* with a dramatically altered symbiotic phenotype. Plant J. 23:97-114.
- Wycoff KL, Hunt S, Gonzales MB, VandenBosch KA, Layzell DB, Hirsch AM. 1998. Effects of oxygen on nodule physiology and expression of nodulins in alfalfa. Plant Physiol. 117:385-95.
- Yang WC, De Blank C, Meskiene I, Hirt H, Bakker J, Van Kammen A, Franssen H, Bisseling T. 1994. *Rhizobium* Nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation. Plant Cell 6:1415-1426.
- Yang WC, Horvath B, Hontelez J, Van Kammen A, Bisseling T. 1991. *In situ* localization of *Rhizobium* mRNAs in pea root nodules: *nifA* and *nifH* localization. Mol. Plant-Microbe Interact. 7:276-281.
- Yang WC, Katinakis P, Hendriks P, Smolders A, De Vries F, Spee J, Van Kammen A, Bisseling T, Franssen H. 1993. Characterization of *GmEnod40*, a gene showing novel patterns of cell-specific expression during soybean nodule development. Plant J. 3:573-585.
- Zhang K, Letham DS, John PC. 1996. Cytokinin controls the cell cycle at mitosis by stimulating the tyrosine dephosphorylation and activation of p34cdc2-like H1 histone kinase. Planta 200:2-12.

## Samenvatting

Vlinderbloemigen kunnen onder stikstoflimiterende groeiomstandigheden een symbiose aangaan met *Rhizobium* bacteriën. De totstandkoming van de symbiose wordt gecontroleerd door moleculaire signalen van zowel de gastheerplant als de bacteriën. De door de plant uitgescheiden flavonoïden induceren in de bacterie de productie en de uitscheiding van een signaalmolecuul, de zogenaamde nodulatiefactor (Nod factor). De bacteriën dringen de wortels van de plant binnen en induceren daar de vorming van een nieuw plantenorgaan: de wortelknol. In de wortelknollen zijn de fysiologische omstandigheden ideaal voor de werking van het bacteriële enzym nitrogenase dat de reductie van moleculaire stikstof tot ammonia katalyseert. De plant kan deze stikstofvorm assimileren en voorziet de bacteriën in ruil hiervoor van suikers.

Bij de symbiose zijn zowel *Rhizobium*- als plantengenen betrokken. Inmiddels zijn de meeste bacteriële genen betrokken bij wortelknolvorming en hun rol hierbij bekend. Veel minder is echter bekend over de exacte functie van de plantengenen betrokken bij de symbiose. De studie van mutanten kan inzicht verschaffen in de functie van een gen. Hoewel er enkele tientallen symbiotische mutanten zijn geïsoleerd, heeft er niet één een mutatie in een van de reeds geïsoleerde genen betrokken bij de knolvorming. Wel zijn er veel andere genen betrokken bij de symbiose geïdentificeerd en op genetische kaarten geplaatst. De isolatie van deze genen kan gebeuren via positionele klonering. Dit is echter een tijdrovende klus. De isolatie van een gemuteerd gen kan vergemakkelijkt worden door het gebruik van T-DNA als mutagen. Hierbij kan de bekende basenvolgorde van het T-DNA namelijk dienen als uitgangspunt. Het T-DNA van *Agrobacterium tumefaciens* inserteert op willekeurige plaatsen in vooral het euchromatische gedeelte van het genoom van de plant en de kans dat een bepaald gen geraakt wordt neemt toe met de grootte van de gemuteerde populatie en vereist dus een efficiënt transformatiesysteem. Sinds kort is er een efficiënte methode voor de transformatie en *in vitro* regeneratie van *Medicago truncatula* beschikbaar, zodat T-DNA tagging in deze modelplant mogelijk werd. Het in dit proefschrift beschreven onderzoek had tot doel de toepasbaarheid van T-DNA tagging als methode voor het isoleren van genen betrokken bij de knolvorming in *M. truncatula* te onderzoeken.

Een laag T-DNA insertieaantal en een eenvoudig insertiepatroon zullen de karakterisatie van mutanten en de daaropvolgende klonering van het T-DNA flankerende genomische DNA vergemakkelijken. Moleculaire analyse van planten die getransformeerd waren met de vectoren pPR97, pSLJ8313, pSLJ8337 of pGKB5 lieten zien dat transformatie met de van pBin19 afgeleide vector pPR97 vooral in complexe T-DNA insertiepatronen resulteerde, terwijl de meeste planten die het T-DNA van de vectoren pSLJ8313, pSLJ8337 of pGKB5 bevatten slechts één T-DNA insertie hadden. pGKB5- en pSLJ-type vectoren bleken dus het meest geschikt voor T-DNA tagging in *M. truncatula* te zijn.

De vector pGKB5 bevat een promoterloos *GUS* reportergen aan het right border (RB) uiteinde van het T-DNA. Als het T-DNA in de juiste oriëntatie in een gen inserteert kunnen genfusies ontstaan. Hierdoor kan het expressiepatroon van het getagde gen gevolgd worden, maar kunnen ook mutaties die in homozygote situatie tot letaliteit leiden of mutaties die geen duidelijk zichtbaar fenotype veroorzaken zichtbaar worden gemaakt. Hoewel het gebruik van de reporter GFP (Green Fluorescent Protein) voordelen heeft boven dat van het enzym *GUS* (in tegenstelling tot *GUS* kan de aanwezigheid van GFP op niet-destructieve wijze vastgesteld worden) bleek door de hoge achtergrondfluorescentie in de wortelknollen van *M. truncatula* deze reporter niet geschikt voor de detectie van genexpressie in dit orgaan. Aangezien de reporter *GUS* daarentegen al met succes gebruikt was als reporter in de wortelknollen werd besloten om *GUS* voor de fusies te gebruiken. Van 187 transgene plantenlijnen die het pGKB5 T-DNA bevatten lieten er negentien *GUS*-activiteit in de wortels of de wortelknollen zien.

De basenvolgordes van T-DNA flankerend *Medicago* DNA van twaalf planten die het pSLJ-type T-DNA bevatten en van negen planten die met pGKB5 waren getransformeerd lieten zien dat in vijf plantenlijnen het T-DNA in het getranscribeerde deel van een gen, in drie lijnen in de buurt hiervan en in drie lijnen in transposon sequenties geïntegreerd was. Hierbij viel op dat hoewel het gebruik van pGKB5 interessante nieuwe markers opleverde, de selectie van *GUS*-positieve planten niet resulteerde in de identificatie van meer nieuwe *Medicago* genen. Geen van de getagde lijnen had een macroscopisch zichtbaar afwijkend nodulatiefenotype en alle planten maakten wortelknollen die niet van wild type wortelknollen te onderscheiden vielen.

Het nodulatiefenotype van homozygote T-DNA insertie mutanten van de *GUS*-positieve lijn B9-1 kon niet bepaald worden, aangezien deze afwezig waren. Van de nakomelingen van een heterozygote B9-1 plant was tweederde heterozygoot voor het T-DNA, terwijl dit in eenderde afwezig was. Deze segregatie van het T-DNA

duidde op embryoletaliteit veroorzaakt door de T-DNA insertie in het *MtN3*-like gen. Lijn B9-1 is een goede illustratie van het nut van een reporter-gen om letale mutaties die in heterozygoten geen duidelijk fenotype veroorzaken op te sporen. In lijn B9-1 bleek de *GUS*-expressie het gevolg te zijn van een translationele fusie van *GUS* en een aan de nodulin *MtN3* (Gamas *et al.*, 1996) verwant gen. Het complete cDNA van het wild-type *MtN3*-like gen werd geïsoleerd uit een cDNA bank en bleek, gebaseerd op de hydrofobie van de afgeleide aminozuurvolgorde, voor een membraaneiwit met zeven transmembraansegmenten te coderen. De transiënte expressie van een C-terminale *MtN3*-like GFP-fusie in gebombardeerde uiencellen bevestigde dit aangezien het GFP in de plasmalemma gelokaliseerd bleek te zijn. Ondanks de afwezigheid van een ER retentie sequentie bleek een *MtN3* GFP-fusie ER-gelokaliseerd te zijn. Om inzicht te verkrijgen in de functie van *MtN3*-like tijdens de wortelknolvorming werden respectievelijk sense en antisense constructen tot expressie gebracht in hairy roots. Deze experimenten gaven echter geen opheldering wat betreft de functie van dit eiwit tijdens de nodulatie.

In de wortels van B9-1 planten kwam de *MtN3*-like *GUS*-fusie tot expressie in de vaatbundels op de plaats waar een wortelknol zich ontwikkelde en ook in de vaten van ontwikkelde knollen. Verder was er *GUS* activiteit in de vaten van groene (kiem)bladeren, peulen en in de bloemen. Uit Northern blot analyse bleek dat het *MtN3*-like mRNA zich in dezelfde organen bevond als waar de *MtN3*-like *GUS* fusie tot expressie kwam. Hiermee is lijn B9-1 de eerste gekarakteriseerde getagde *Medicago* lijn met een translationele fusie tussen een planteneiwit en *GUS*. T-DNA tagging kan dus gebruikt worden om in *M. truncatula* (marker) genen op te sporen. Hoewel de *GUS* expressie patronen geen uitsluitel geven omtrent de functie van *MtN3*-like, duiden de *GUS* expressie patronen op een rol in het transport van plantaardige producten, zoals bijvoorbeeld suikers.

Lijn A7-1 had een uitgesproken dwergfenotype in de spruit, terwijl het wortelstelsel en het nodulatiefenotype normaal waren. Het dwergfenotype was gecorreleerd met homozygote aanwezigheid van het T-DNA dat in een homoloog van het menselijke NARF (nuclear prelamin A recognition factor) gen (Barton and Worman, 1999) geïnserteerd is. Het dwergfenotype verdween door complementatie met *AtNARF*-like en silencing van *AtNARF*-like in *Arabidopsis* resulteerde in een dwergfenotype. Lijn A7-1 is hiermee de eerste getagde *Medicago* lijn waarvoor aangetoond is dat het mutante fenotype inderdaad door de T-DNA insertie veroorzaakt is. Verder zijn de *Medicago* en *Arabidopsis* NARF-like genen waarschijnlijk ortologen met dezelfde functie.

Het menselijke NARF eiwit wordt verondersteld een rol te spelen bij de maturatie van lamin A. Lamines zijn onder andere belangrijk voor het doorlopen van de celcyclus. De cellen van de bladepidermis van dwergen waren even groot als die van wildtypes en het fenotype leek dan ook eerder veroorzaakt door een kleiner aantal cellen dan door kleinere cellen. Men zou zich kunnen voorstellen dat het lagere aantal cellen en het dwergfenotype het gevolg zijn van een verstoorde celcyclus waarbij minder celdeling optreden. Elektronenmicroscopie liet zien dat sommige cellen van dwergblaadjes, waarschijnlijk als gevolg van het gebruikte vacuüm, ingeklapt waren, terwijl overeenkomende cellen van bladeren van wildtypes bolvormig leken. NARF-like zou dus ook betrokken kunnen zijn bij het bepalen van de stevigheid van de celwand.

De tagging efficiëntie van het T-DNA van pGKB5 kon waarschijnlijk verhoogd worden door: een intron in *GUS* te kloneren, in elk van de open leesramen een 3' intron splice site voor het *GUS* start codon aan te brengen en de basenvolgorde rond het *GUS* start codon te optimaliseren voor planten. Het intron bevattende *GUS* kwam inderdaad niet tot expressie in *Agrobacterium tumefaciens*, waardoor eventuele achtergrond *GUS* activiteit van nog in de getransformeerde planten aanwezige bacteriën voorkomen zal worden. Ook bleken twee van de drie 3' intron splice sites als zodanig door de plant herkend te worden en een functioneel intron te vormen met de 5' intron splice site van het vijfde intron van het *AtARF* gen in een testconstructie. Hierdoor zal ook insertie van het T-DNA in een intron *GUS* activiteit tot gevolg kunnen hebben, omdat door het verwijderen van het intron eventuele stopcodons tussen het plantengen en *GUS* verwijderd worden en zo een translationele fusie gevormd kan worden. Hoewel dit niet getest is zouden de verandering van de basenvolgorde rond het *GUS* start codon en de introductie van het intron in *GUS* de expressie van *GUS* moeten verbeteren, waardoor getagde genen die een lage expressie hebben toch gedetecteerd kunnen worden. Uit een klein T-DNA tagging experiment met de vernieuwde vector bleek de effectiviteit ten opzichte van de originele vector ongeveer 2,7 keer hoger te zijn en deze vector zal dan ook voor toekomstige tagging experimenten gebruikt gaan worden.

Ten slotte kan geconcludeerd worden dat T-DNA tagging een goede methode is om in *M. truncatula* genen te muteren en te isoleren. Echter, om de kans dat het T-DNA genen raakt die betrokken zijn bij nodulatie te vergroten is een zeer grote gemutageniseerde *M. truncatula* populatie nodig. Het ontwikkelen van een nog efficiëntere transformatiemethode of het gebruik van transposons als tags lijken dan ook onontbeerlijk voor de realisatie van een grootschalig T-DNA tagging project.

## Curriculum vitae

Marije Scholte is op 9 juni 1971 te Stadskanaal geboren. Na het behalen van het VWO diploma aan het Röllingcollege in Groningen (diploma 1989), is zij milieukunde aan het Van Hall Instituut in Groningen (diploma 1995) en biologie aan de Rijksuniversiteit Groningen (diploma 1997) gaan studeren. Tijdens het doctoraalprogramma van de studie biologie heeft zij haar aandacht gericht op de moleculaire biologie en onder andere de vakken Moleculaire Genetica en Moleculaire Biologie van Planten gevolgd. Het tweede doctoraalonderwerp 'Preparation of transgenic *Medicago truncatula* and *Arabidopsis thaliana* plants for the analysis of the role of a Nod signal hydrolase in nodulation and plant development' heeft zij in de groep van dr. Eva Kondorosi onder leiding van dr. Michael Schultze uitgevoerd op het 'Institut des Sciences Végétales', onderdeel van het 'Centre National de la Recherche Scientifique (CNRS), in Gif-sur-Yvette in Frankrijk. Vervolgens heeft zij aan hetzelfde instituut, onder leiding van dr. Michael Schultze en na diens vertrek naar de universiteit van York onder leiding van dr. Pascal Ratet, een promotieonderzoek uitgevoerd. Het onderzoek, beschreven in dit proefschrift, werd gefinancierd door een TMR beurs van de Europese Unie. Momenteel werkt zij als post-doc bij het Nederlands Kanker Instituut in Amsterdam.



