**STATEMENTS**

1)

Comparative genomics provides a means to answer important questions about evolutionary history, including origin of species, adaptation to biotic and abiotic factors and survival from extinction.

2)

Knowledge about adaptation to biotic and abiotic factors constitutes a basis for genetic therapy/remediation in a broad sense comprising areas such as agriculture, ecology, medicine and veterinary.

3)

*In situ* (on-farm) conservation of genetic resources of wild species and crop landraces is essential to facilitate the evolutionary processes underlying biodiversity and could be strengthened in the future by the development of programs coordinating the activities of farmers, governments and the biotechnology industries.

4)

Bio-molecular techniques offer a fast and reliable way to control the quality of export/import agricultural products within the international market sanitary standards.

5)

Genetic engineering of forage legumes including the introduction of resistance to pathogens, adaptation to environmental conditions such as drought, enhanced nutritional feed-value, and preservation of feed-value under storage conditions will result in increased livestock production.

6)

Of the agricultural important mammals cow has the highest density of markers associated with QTLs and economic trait loci (ETLs). This offers the possibility for positional cloning of these genes by comparative genomics using the gene-rich human, mouse and rat genetic maps, which will markedly increase the opportunities to improve cattle production.

7)

Biotechnology should be criticized, by similar standards as any other scientific, cultural or social activity that drove us in time from the prehistoric caves into the modern contemporary society.

8)

" Y sé que en nuestras vidas se produjo
El milagro inefable del reflejo...
En el silencio de la noche mi alma
Llega a la tuya como un gran espejo."

Delmira Agustini (1886-1914, Uruguay).
Fragment from the poem "Intima" in *El libro Blanco*
(O.M. Bertani ed.), Montevideo 1907.

9)

In this thesis, the *Medicago truncatula* Sym2-orthologous region was cloned and delimitated to 350 Kbp, providing a solid basis for the cloning of the pea Sym2 gene by comparative genomics.

10)

It is highly probable that the leucine-rich Cf4/Cf9-like pea cDNA RFLP marker *PscLRR52*-isolated in this thesis- is involved in Nod factor perception or /and signal transduction, and represents the pea Sym2 gene.

# Medicago truncatula, an intergenomic vehicle for the map-based cloning of pea (*Pisum sativum*) genes

## Comparative structural genomic studies of the pea *Sym2-Nod3* region

Gustavo S. Gualtieri González-Latorre

Promotor: prof. dr. A.H.J. Bisseling, hoogleraar Moleculaire Biologie

Samenstelling promotiecommissie:
        prof. D. Cook (University of California, USA)
        dr. J.H. de Jong (Wageningen Universiteit)
        prof. dr. ir. M. Koornneef (Wageningen Universiteit)
        prof. dr. W.J. Stiekema (Plant Research International;  Wageningen Universiteit)
        prof. dr. S.C. de Vries (Wageningen Universiteit)

# Medicago truncatula, an intergenomic vehicle for the map-based cloning of pea (*Pisum sativum*) genes

## Comparative structural genomic studies of the pea *Sym2-Nod3* region

**Gustavo S. Gualtieri González-Latorre**

# CONTENTS

# OUTLINE

Biological nitrogen fixation is a unique process that is only carried out by prokaryotes. In some cases microbes fix nitrogen in symbiotic association with specific groups of dicot plants. These prokaryotes produce the enzyme nitrogenase that reduces nitrogen gas into ammonia. Chapter 1 presents recent molecular phylogenetic studies of seed plants showing that all nodule-forming plants belong to a single clade named Rosid I. Within this clade four subclades out of six are able to form nodular symbioses. One of them contains the family Leguminosae that is nodulated exclusively by bacteria of the gram-negative genera *Azo-, Brady-, Meso-*and *Sino-Rhizobium*, while the other three clades contain the families Rosaceae, Ulmaceae, Elaeagnaceae, Rhamnaceae, Betulaceae, Casuarinaceae, Myricaceae, Coriariaceae and Datiscaceae that interact with the gram-positive actinomycete *Frankia* and are collectively called actinorhizal plants. However, the tropical tree *Parasponia* belongs to the Ulmaceae, but can only be nodulated by rhizobia. Thus, molecular phylogenetic studies show that plants able to establish an endosymbiotic association leading to nitrogen fixing nodules are more related than what their morphological features indicate. Chapter 1 describes and compares legume and actinorhizal nodule ontogeny. It seems likely that the prenodule formed in actinorhizal symbioses is related to the nodule primordium formed in legume symbioses. Moreover, this comparison suggests that the mechanisms underlying nodule development are common to all nodule-forming plants. Furthermore, it is discussed that the nodulation process is (in part?) derived from processes that are widespread among higher plants.

One of the main questions addressed in this thesis, is whether microsynteny exists between legume species and this is analyzed by comparative genomics studies. Chapter 2 is a structural comparative genomic study on two plants belonging to the Leguminosae: *Pisum sativum* bv *viciae* (pea) and *Medicago truncatula* A17 line. The level of microsynteny of the pea "*Sym2* region" and the orthologous *Medicago* region is studied. This chapter represents the beginning of a microsynteny-based positional cloning approach for the pea *Sym2* gene, which controls Nod factor structure-dependent strain-specific nodulation. A marker which is tightly linked to *Sym2* was used to screen a *M. truncatula* A17 BAC library and three physically unlinked contigs (named c1, c2 and c3) were constructed and extended by chromosome walking to a final total size of about 600 Kbp. These contigs were mapped genetically and by FISH on *M. truncatula* A17 chromosome 5. RFLP analysis demonstrated that c1/c2 is the *Medicago Sym2*-orthologous genomic region. Furthermore, some of these RFLP markers

revealed recombinations around the *Sym2* locus and therefore delimitated the *Sym2*-orthologous region within c1/c2. Moreover, the potential of *Medicago truncatula* A17 as intergenomic cloning vehicle of pea genes located at microsyntenic regions is demonstrated by the isolation-from a pea root hair library-of the cDNA *PscLRR52*, which contains a LRR motif that is highly homologous to the LRR motives of the Cf4 and Cf9 [*Licopersicon esculentum*] disease resistance proteins.

In Chapter 3, the distribution of several sequences in c1/c2 and c3 is studied through the construction of detailed contig molecular maps. This showed that the majority of the sequences located in c1/c2 do not occur c3. This demonstrates again that c1/c2 is the *Sym2*-orthologous region. However, c3 contains a sequence that is highly homologous to *PscW62-1*. In addition, the *Sym2*-linked RFLP markers Mtg3552 and *Mtc411* from c1/c2, and the c2 genomic subclone Mtg2.4, are also present in both c1/2 and c3. Moreover, it was found that Mtg3552 and Mtg2.4, and *PscW62-1* and *Mtc411*, are physically clustered and that these clusters have a similar order in c1/c2 and c3, respectively. Thus, the colinearity of these marker clusters indicates that (a part of?) c1/c2 and c3 arose by genome duplication, rather than by independent evolution. Similar duplications have been described in grasses and in Brassicaceae.

In Chapter 4 the RNA differential display mediated the isolation of the pea RFLP markers *dd21.5* and *Psc2.6* that are tightly linked to each other and located close to the *nod3* hypernodulation locus that maps at a 2 cM distance from *Sym2* in pea linkage group I. These two markers are probably as close to *Nod3* as the c1 markers *Mtc831* and *Mtc923* are from *Sym2*, because two recombinations map between *dd21.5/Psc2.6* and *Nod3*, and *Mtc831/Mtc923* and *Sym2*. Moreover, since the *Medicago Sym2*-orthologous region was delimitated to about 350 c1/c2 Kbp (by markers *Mtc831*, *Mtc923* and Mtg63EB4), it is probable that the size of the *Medicago* physical region containing the *Nod3*, *dd21.5*, and *Psc2.6* homologues would be about this size range. Thus, the identification of BACs composing this *Medicago* orthologous region was started in Chapter 4. By screening a *M. truncatula* A17 BAC library with marker *dd21.5*, BAC clone 21F22 was isolated. This BAC contains the complete *Medicago dd21.5*-true orthologue Mtg21.5, which is a single copy sequence. However, a *Medicago dd21.5*-distant homologue, not located in 21F22, is revealed by low stringency Southern blots. BAC library screenings also revealed five BACs that co-hybridize with both *dd21.5* and *Psc2.6*. Thus, the tight linkage of these two pea markers is

also observed for their *Medicago* homologues, demonstrating that (submegabase) microsynteny exists between the genomic regions containing these sequences in the two legume species. However, at the chromosome level a translocation (disturbance of synteny) was observed. BAC 21F22 was mapped by a strong FISH signal on *M. truncatula* A17 chromosome 4. Additional weak signals were observed close to both telomeres of chromosome 5. The mapping of 21F22 on chromosome 4 demonstrates a local loss of chromosomal synteny-at the "*dd21.5* region"-between *Medicago* chromosome 5 and pea linkage group I that are otherwise syntenic for the "*Sym2* region" and "*Eil2* region". However, it is not known yet whether the five BACs that contain homologues of both *dd21.5* and *Psc2.6* form a contig with 21F22 and map in chromosome 4, or if they contain the *Medicago dd21.5*-distant homologue and map somewhere else in the genome, e.g. on chromosome 5 where 21F22 reveals weak signals. In addition, it cannot yet be concluded if the *Medicago Nod3*-orthologous locus is, as Mtg21.5, located on chromosome 4.

Thus, both Chapter 4 that focuses on microsynteny and chromosomal synteny around the *Nod3* locus, and Chapter 3 that focuses on microsynteny between the c1/c2 and c3 regions, revealed the genome dynamics during the evolution of these two Leguminosae species.

In Chapter 5 a pachytene karyogram was established for *M. truncatula* A17, in which all chromosomes can be identified based on chromosome length, centromere position, heterochromatin patterns, and the position of three repetitive sequences (5S rDNA, 45S rDNA and the *Mt*R1 tandem repeat) visualized by FISH. The correlation between linkage groups and chromosomes was determined by FISH of 20 BACs on pachytene chromosomes (with two to five BACs per linkage group) and 3 repetitive sequences. The physical and genetic distances between these markers were compared. The FISH mapping resolution was studied in the euchromatic c1/c2 *Sym2*-orthologous region of chromosome 5 and determined as 60 Kbp. This high resolution creates the basis for the integration of molecular and cytogenetic maps.

# CHAPTER 1

## The evolution of nodulation

Gustavo Gualtieri* and Ton Bisseling

Department of Molecular Biology, Agricultural University, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands.

* author for correspondence.

**Abstract**

In this review we will first describe the different steps leading to nodule formation, and these will be compared with processes of non-symbiotic plant development and growth. In general, aspects of both actinorhizal as well as rhizobial symbioses are described, but in several cases, the emphasis will be on the *Rhizobium*-legume symbiosis because more knowledge on this system is available. Subsequently, the phylogeny of nodulating plants is described and a comparison is made between several aspects of legume and actinorhizal nodulation. At the end of this chapter the relationship between nodule symbiosis and endomycorrhizal symbiosis is described, and it is discussed to what extent the development of root nodules involves unique properties, or whether processes and genes have been recruited from common plant development and the endomycorrizal symbiosis.

**Nitrogen-fixing root nodules**

Biological nitrogen fixation is a process that can only be performed by prokaryotes. In some cases these microbes fix nitrogen in symbiosis with plants. In several of these symbioses a new organ is formed, the so-called root nodules, where the bacteria are hosted in an intracellular manner. There, the bacteria make the enzyme nitrogenase by which they can reduce nitrogen gas into ammonia. The bacteria that can establish a nodule symbiosis belong to two phylogenetic groups namely (*Azo-*, *Brady-*, *Sino-*) *Rhizobium* (here collectively called rhizobia) [68] and *Frankia* [5].

Rhizobia are gram-negative soil bacteria that have the ability to interact with plant species belonging to the Leguminosae family. In addition to these leguminous plants, the tropical tree *Parasponia* belonging to the Ulmaceae has the ability to form nodules with rhizobia [3]. In contrast, plant species belonging to eight families of angiosperms, collectively called actinorhizal plants, can form root nodules with the gram-positive actinomycete *Frankia* [5].

In addition to these root nodule endosymbiotic symbioses, certain cyanobacteria (e.g. *Anabaena*, *Nostoc*) can establish a nitrogen fixing symbiosis with plants belonging to the Gunneraceae [6]. However, in this interaction already existing glands of the plant are used to host the nitrogen-fixing microbe.

**Infection**

The formation of a nitrogen-fixing nodule requires that the bacteria mitotically activate certain root cells, which form the nodule primordia. Which root cells will divide is controlled by the host plant (see Nodule ontogeny and tissue organization). In addition to the formation of primordia the bacteria have to enter the plant to infect primordial cells after which a nitrogen-fixing nodule can develop.

To enter the root, both intracellular and intercellular mechanisms of infection have evolved. The intracellular mode of infection is used in most *Rhizobium*- legume interactions as well as in several of the *Frankia* symbioses. The process begins with curling of root hairs and in these curls the cell wall is locally hydrolyzed, the plasma membrane invaginates, and new cell wall material is deposited at the site of infection [26, 34, 45]. In this way, a tube like structure, called an infection thread, is formed through which the bacteria can enter the plant. Most probably, infection threads grow by vesicle incorporation at their tip [13] as in growing pollen tubes.
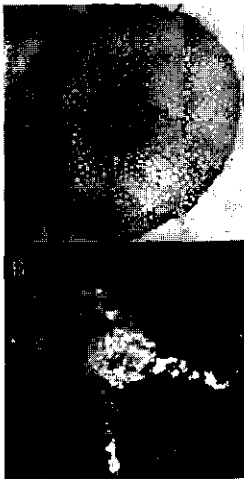


**Figure 1**
Cross-section of a pea root one day after inoculation with *Rhizobium leguminosarum* bv *viciae*. By *in situ* hybridization histone H4 mRNA is visualized in panel A, while in panel B (dark-field illumination) it is visible as white dots. H4 expression is specific for cells in the S-phase of the cell cycle. Only the inner cortical cells progress through the cycle and will divide. The outer cortical cells become arrested at the G2-phase. In these later cells a phragmoplast-like structure is formed that is used to support the growth of the infection thread. Arrows indicate xylem poles.

A            B            C

▬▬▬▬  vascular system

▬▬▬▬  area with infected cells containing nitrogen-fixing bacteria

▭▭▭▭  expanded pericycle
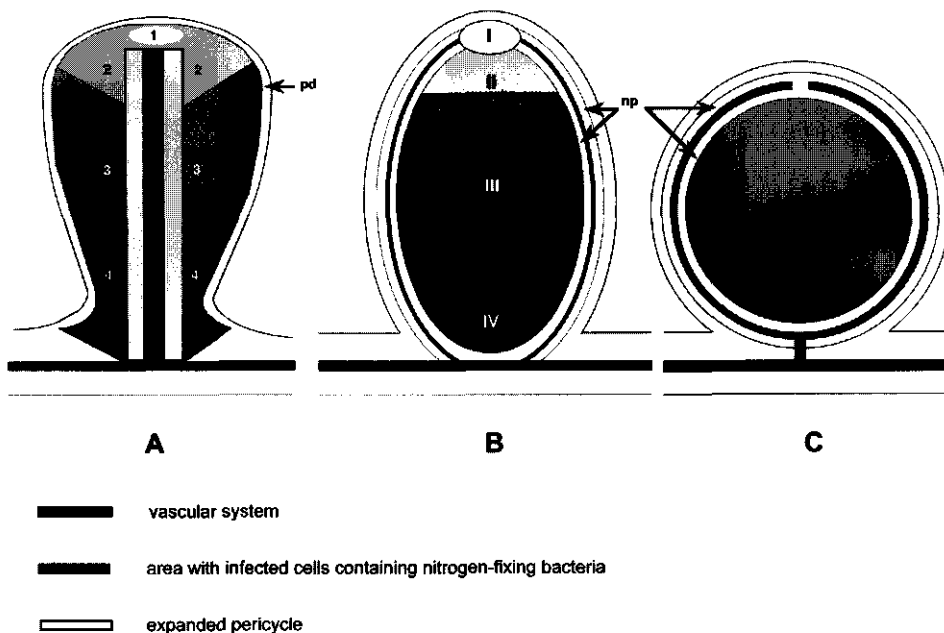
**Figure2**
A: Lobe of an actinorhizal nodule. The lobe is surrounded by a periderm (pd). Analogous to indeterminate legume nodules, a zonation of the cortex can be defined [51]. Zone 1 is formed by the meristem. Zone 2 contains cells that become infected by and gradually filled with *Frankia* hyphae, which subsequently form vesicles [56]. In the course of vesicle differentiation, *nif* gene expression and nitrogenase production are induced [30, 51]. The onset of *nif* gene expression marks the shift to nitrogen fixation in zone 3. In the senescence zone 4, nitrogen fixation has ceased, and plant cytoplasm and bacteria are degraded.

B: Indeterminate legume nodule. The central tissue can be divided into five zones [72]. Directly below the meristem (I), in the prefixation zone (II), cells become infected. Rhizobia are enclosed by peribacteroid membranes (PBMs) and start to differentiate into their symbiotic form, the bacteroids. In the interzone (II-III), bacterial nitrogen fixation starts [77] and takes place throughout the nitrogen fixation zone (III). In the senescent zone (IV), bacteria are degraded. The oxygen diffusion barrier is formed by the nodule parenchyma (np).

C: Determinate legume nodule. The central tissue is surrounded by a nodule parenchyma (in which the vascular bundles are located), an endodermis and the outermost tissue is the nodule cortex. The central tissue is surrounded by the same tissues as in the indeterminate nodules. All cells of the central tissue are more or less in the same developmental stage. This figure was kindly provided by K. Pawlowsky [published in 47].
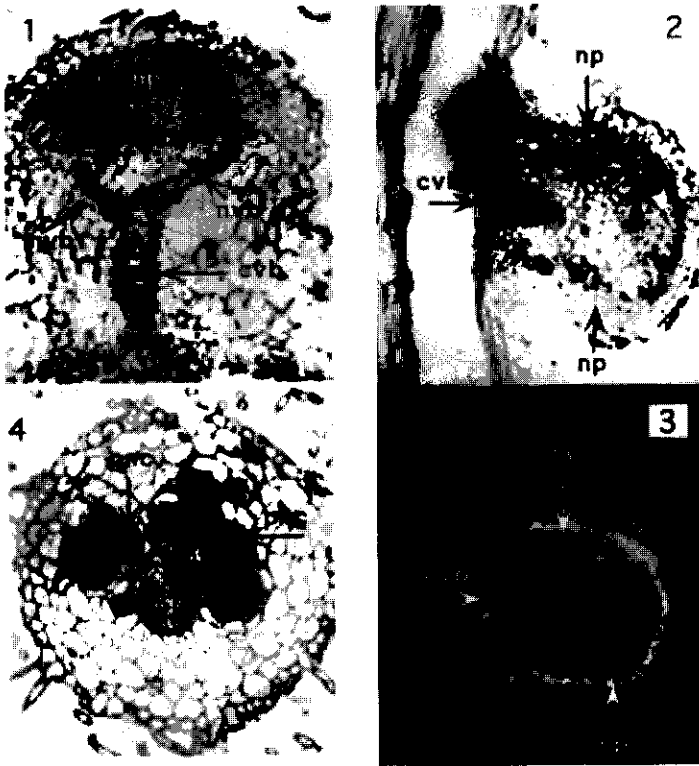
**Figure 4**
1-3. Comparisons of wild-type soybean nodule primordium and the response induced by *Bradyrhizobium japonicum* mutant 3160 [78].
1. Soybean nodule primordia (np) are formed in the outer cortex. When the primordium differentiates, nodule vascular bundles (nvb) are formed that are located surrounding the central tissue. In the root inner cortex a vascular bundle is formed that "connects" the nodule (primordium) with the central stele (cvb).
2, 3. Nodule-like structure induced by inoculation with *B. japonicum* 3160. The nodule primordium is visualized by hybridization *ENOD40*, and visible as white dots in panel 3. The nodule primordium (np) forms a thin layer of cells surrounding a prominent "connecting vascular bundle"(cvb).
4. Prenodule of *Alnus glutinosa* inoculated with *Frankia*. A few cells are infected and these form the prenodule. *Frankia* has amplified in the infected cells. The small globular structures in these infected cells are the vesicles occupied by *Frankia*. This picture was kindly provided by Emile Duhoux. cvb, connecting vascular bundle; np, nodule primordia; nvb, nodule vascular bundle; rvb, root vascular bundle; pnc, prenodule cell.

The infection thread must transverse several layers of root cortical cells when the nodule primordia are formed in the inner layers of the root. Cortical cells do not have a polar-organized cytoplasm and, therefore, the cytoarchitecture of these cells has to be changed in order to support the polar growth of infection threads. Before infection threads enter cortical cells, the nucleus moves to the center of the cell and cytoplasm and microtubules rearrange to form a phragmosome-like structure, that is called the cytoplasmic bridge [69]. These structures have a polar organization, with the bulk of the cytoplasm and the endomembranes located at the outer sides. The cytoplasmic bridges are more or less radially aligned in the

cortex, and in this way can guide the infection thread to the nodule primordium. These cytological changes indicate that the cortical cells preparing for infection thread penetration enter the cell cycle although they do not divide. By using a histone H4 probe it was demonstrated that narrow rows of cortical cells enter the S phase, but since a mitotic cyclin gene is not activated it was concluded that the cytoplasmic bridge-forming cells arrest at the G2 phase (Figure 1) [79]. Thus, intracellular infection is facilitated by a modification in the cell division process in cortical cells.

Rhizobia and *Frankia* can also penetrate the plant intercellularly via gaps in the epidermis or by penetration of the middle lamella between intact epidermal cells of the root [10, 14, 32, 43, 49]. Such infections mostly continue via an intercellular pathway in the cortex [43, 49]; however, in some cases, combined mechanisms developed in which infection starts via an intercellular pathway, and then switches to an intracellular infection thread in the root cortex [62, 67]. The intracellular mode of infection seems to provide the plant a better way to control and guide the growth of the microsymbiont, since the direction and speed of infection thread growth are determined by the host.

When the infecting bacteria reach the nodule primordium, they infect a number of primordium cells. In the legume subfamilies Mimosoideae and Papilionoideae, rhizobia are released into the cytoplasm by a process resembling endocytosis [4]. In this process, the bacteria become surrounded by a plant membrane and together they form the symbiosome [53]. These symbiosomes divide and the infected cells become fully packed with the microsymbiont. On the other hand, in all actinorhizal symbiosis and in several legume symbioses, bacteria are not released from the infection threads [7, 35, 44]. In these symbioses, the infected cells are filled with branching infection threads containing the microsymbionts. After infection the primordium develops into a nodule.

### Nodule ontogeny and tissue organisation

Legume and actinorhizal nodules are ontogenetically and morphologically different (Figure 2). Legume nodules have a stem-like anatomy with peripheral vascular bundles and infected cells in the central tissue, and they develop from primordia that are initiated in the cortex. In contrast, actinorhizal nodules have a root-like anatomy and develop from primordia formed in the pericycle.

Two types of legume nodules have been defined: the determinate and indeterminate nodules. The indeterminate nodules originate from cell divisions in the inner cortex and have a persistent meristem at their apex [26]. Due to the continuous activity of this meristem the nodule cells form a developmental gradient from the distal meristem to the proximal root attachment site [72]. The different developmental zones can be characterized by the expression of specific plant genes that can serve as molecular markers [22].

The determinate nodule type originates from cell divisions in the outer cortex and does not have a persistent meristem. The meristem of this nodule type ceases to divide at an early stage of development. As a result, the nodule cells are at a rather similar developmental stage, although some developmental zonation could be demonstrated by using molecular markers [46].

Actinorhizal nodules have an indeterminate growth pattern, like the indeterminate legume nodule. However, in contrast to these legume nodules, they are composed of several lateral root-like structures (lobes) that lack a root cap. These lobes contain a central vascular bundle and infected cells in their cortex [47]. Like lateral roots, the nodule lobes originate from primordia that are formed in the pericycle. The only non-legume that is able to form nodules with rhizobia, *Parasponia*, forms nodules that have a tissue organization and ontogeny that is different from that of legumes, but it is very similar to that of actinorhizal nodules. Because the formation of legume and *Parasponia* nodules involves the same rhizobial signal molecules, the so called "Nod factors" [42] (see Rhizobial Nod factors), that yet can produce different nodule types, this shows that the host plant controls which nodule type will be formed. Furthermore, this shows that actinorhizal-like nodules and legume nodules can be induced by the same bacterial signal molecule. This suggests that these nodules might be more related than their morphology and ontogeny, at first sight, suggest. Studies on the phylogenetic relationships of nodulating plants gives further support to this idea.

## Phylogeny of nodulating plants

Whereas rhizobia almost exclusively form nodules with leguminous plants, *Frankia* bacteria can interact with plant species of 25 genera belonging to 8 different families of dicotyledonous plants. Traditional taxonomic studies based on morphological traits, indicated that these families were rather unrelated and distributed in four of Cronquist's six major subclasses of dicotyledons: Magnoliidae, Dilleniidae, Rosidae and Hamamelidae [12] (Figure

3 I). However, recent molecular phylogenetic studies of seed plants, based on the sequences of the *rbcL* gene, have provided an alternative view on the relationship of some plant groups [11, 60, 63]. These studies suggest that all plant families involved in actinorhizal symbiosis are rather closely related and belong to the same clade, called Rosid I (Figure 3 II and 3III). Furthermore, also the Leguminosae family as well as *Parasponia* belong to this clade. This means that all nodulating plants are more closely related than previously thought and belong to a single clade, where they are divided among 4 subclades (Figure 3 III) [60]. One of these subclades contains legumes and non-symbiotic relatives, while the other three subclades contain actinorhizal plants along with other non-symbiotic related taxa. *Parasponia* is a member of the subclade that also includes actinorhizal plants belonging to the Rosaceae, Rhamnaceae and Elaeagnaceae. Such phylogenetic studies indicated that nodulating plants are rather related. However, the division of these plants among four subclades, as shown in Figure 3, is still subject of discussion since other molecular phylogenies suggest that actinorhizal plants belong to four subclades instead of three [63].

It is worth noting that Gunneraceae, that interact with nitrogen-fixing cyanobacteria [6], do not belong to the Rosid I clade (Figure 3 II). As described above, in this interaction no new organ is formed but existing glands become occupied by the microbes. This together with the phylogenetic data shows that this endosymbiosis is unrelated to nodulation.

The fact that all nodule-forming plants belong to a single clade suggests that a plant predisposition to nodulation by rhizobia or *Frankia* has originated in a common ancestor of the Rosid I plants. Only a minority of the genera within this clade is able to establish a nodule symbioses and each subclade that contains nodule-forming species, has non-symbiotic members as well. Therefore it has been postulated that some predisposition to nodulation - which nature remains unknown- evolved in the ancestor of Rosid I clade plants [60]. Furthermore, the ability to form nodules would have independently developed several times within the subclades, followed by the loss of nodulation capacity in the ancestors of the non-nodulating relatives that are scattered among nodulating plants [18, 60, 63]. The number of putative independent origins of nodulation has been extensively analyzed in the actinorhizal symbiosis [63]. Phylogenies constructed by combining *rbcL* gene sequence data with symbiotic morphological features, suggests that actinorhizal symbiosis has evolved at least four times, and possibly as many as six times, during angiosperm evolution [63].
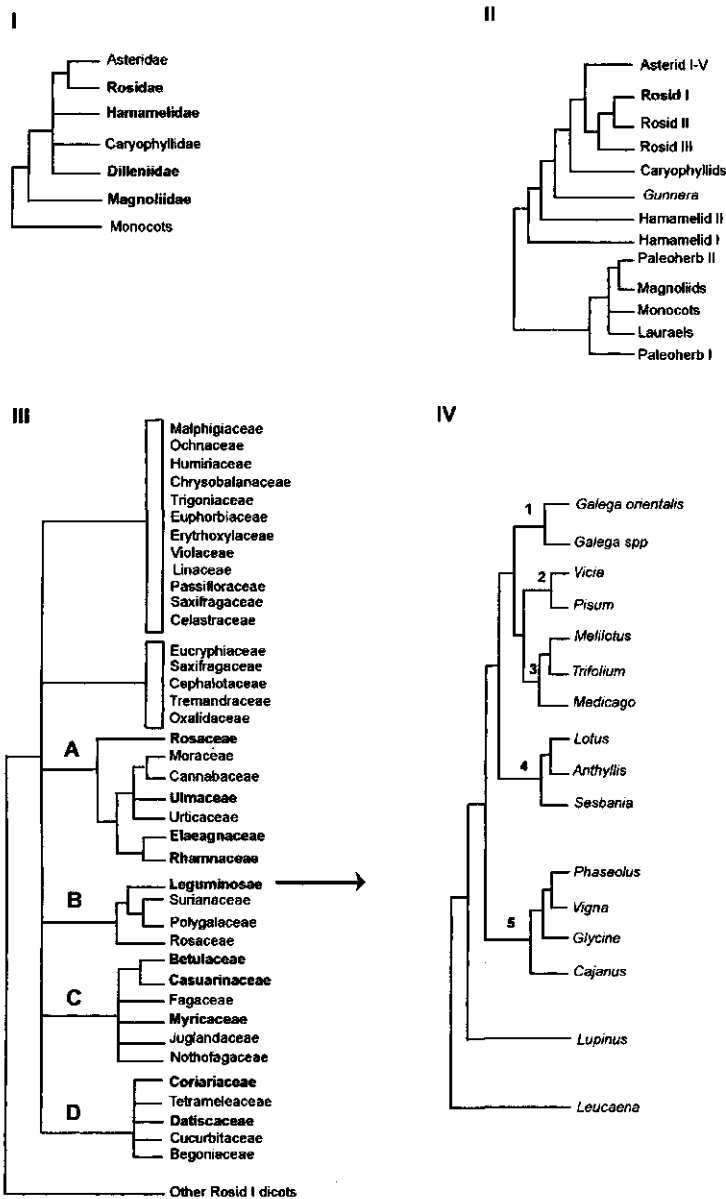
**Figure 3**

**Figure 3**
I. Traditional classification scheme for flowering plants. Groups (subclasses) shown in bold are involved in nodular symbiosis. **II, III, IV.** Phylogenies deduced from *rbcL* gene sequence data. **II.** Major angiosperm clades identified. All plants involved in nodular symbiosis group within the Rosid I clade (shown in bold). **III.** Phylogeny of angiosperm families comprising the Rosid I clade. Families indicated in bold are involved in nodule symbiosis, other families represent non-symbiotic relatives. The Rosid I clade is divided into six subclades, four of which (indicated as A, B, C, and D) contain families that are able to form root nodules. **IV.** Phylogeny of the genera comprising the Leguminosae family. The numbers represent the different tribes mentioned in the text. Modified after Doyle [18].

**Actinorhizal prenodules**

As described above, the morphology of mature legume and actinorhizal nodules is markedly different. However, in some actinorhizal plants an additional symbiotic structure is formed that seems more related to legume nodules and therefore might give some insight in the mechanism underlying the evolution of the two nodule types. In actinorhizal plants that are infected intracellularly (see "Infection"), *Frankia* induces cell divisions in the root cortex, before they mitotically activate root pericycle cells. The dividing cortical cells form a small protuberance called the prenodule (Figure 4)[7]. Later on, the pericycle is mitotically activated to form nodule primordia, from which lateral root-like nodule lobes are formed that absorb the cortical prenodule cell aggregates. Strikingly, a prenodule is also induced during the infection of *Parasponia* by rhizobia.

Recent studies involving molecular markers for stages of actinorhizal nodule development show that the prenodule is a fully differentiated structure composed of two cell types, infected and uninfected cells. Laplaze *et al.* [37] studied prenodule formation in the actinorhizal plant *Casuarina*. Two marker genes for infected nodule cells were used, namely *cg12*, encoding a serine protease, that is expressed in infected nodule cells when infection proceeds [51], and *cghb*, a Class 2 haemoglobin gene that is active in mature infected cells. In prenodules and nodules these two genes are expressed in a similar way; *cg12* is expressed in the cells that just have been infected and *cghb* is active in fully infected cells. In addition, the bacterial *nifH* gene is highly expressed in the fully infected prenodule cells indicating that nitrogen fixation takes place in the prenodule. Furthermore, as in nodules, the uninfected cells of the prenodule accumulate starch granules. Thus the prenodule appears to be an independent symbiotic organ composed of two cell types that display the characteristics of the infected and uninfected actinorhizal nodule cell types, respectively.

Both legume nodule primordium and actinorhizal prenodule formation involve the induction of cortical cell divisions. The legume nodule primordia develops into an organ with a central tissue surrounded by several peripheral tissues including vascular bundles. These vascular bundles connect the nodule with the root vascular bundle (Figure 2). The formation of part of the vascular bundle that connects the nodule with the root vascular system involves cell divisions in the pericycle as well as in the cortex. Therefore this process seems more similar to nodule lobe formation in actinorhizal plants.

Studies with rhizobial mutants show that the formation of a nodule primordium and the development of the connecting vascular bundle are two independent processes. For example, soybean plants inoculated with *Bradyrhizobim* mutant 3160, which is unable to infect its host, form nodules with a prominent central vascular bundle (Figure 4). Thus, these soybean nodules are more similar to actinorhizal nodules. The early nodulin gene *ENOD40*, which encodes a peptide and might also be active at the RNA level [78], has been used as a molecular marker for the nodule primordium. These studies show that the primordium is present in these altered nodules, but it is very small and surrounds the distal part of the central vascular bundle (the vascular bundle that connects the nodule with the root vascular system). Thus, the formation of the "connecting" vascular bundle is unaffected. In contrast a nodule primordium is formed but becomes arrested at a very early stage of development. Thus, the phenotype of the nodule formed by the infection mutant of *Bradyrhizobium* shows that an actinorhizal-like nodule can be induced in a legume by a single mutation in the infecting bacteria. Moreover, the formation of legume primordia and connecting vascular bundle can be uncoupled to a certain extent, indicating that these are independent processes. Thus, in legumes as well as actinorhizal plants, two processes are induced that might be related to one another. The prenodules might be related to legume nodule primordia and actinorhizal nodule lobes could be homologous to the "connecting" vascular bundle of the legume nodule. In legumes the two processes merged and together form the nodule development program. In actinorhizal plants the two processes remained separate and both independently resulted in the formation of a symbiotic organ, the prenodule, and the nodule lobes, respectively.

**Positioning of nodule primordia**

Nodule primordia are preferentially formed opposite to protoxylem poles, implying that the plant provides some sort of positional information during the process of nodulation.

In pea root explants cortical cell divisions opposite the protoxylem poles are specifically induced by phytohormones. However, if the stele has been removed, cell divisions are induced throughout the cortex [40, 58]. Such studies show that all cortical cells have the potential to dedifferentiate and to divide. Furthermore, it suggests that endogenous transverse gradients of positive and/or negative cell division factors from the stele control the positioning of nodule primordia [41].

In situ hybridization experiments have shown that a 1-aminocyclopropane-1-carboxylate (ACC) oxidase gene is specifically expressed in the regions of the root pericycle

opposite phloem poles [29]. ACC oxidase catalyses the last step in the biosynthesis of ethylene. Knowing that ethylene is a potent blocker of cortical cell division [39] and that it is locally produced opposite to phloem poles, ethylene could be a negative regulator in controlling the positioning of nodule primordia. Studies with ethylene antagonists show that this is indeed the case [29]. Furthermore, Penmetsa *et al.* have shown that a *Medicago* mutant insensitive to ethylene [48] forms nodules opposite to the phloem poles (pers. comm.). Thus genetic as well as biochemical studies show that local production of ethylene provides a major positional cue for nodule primordium formation.

Like root nodules, lateral roots are also formed opposite protoxylem poles. The mechanisms controlling the position of lateral root primordia are unknown; however, the formation of both nodule and lateral root primordia involve the mitotic activation of root cells. Hence it seems probable that the local production of ethylene plays a role in the positioning of both primordia.

**Origin of nodulins**

Research on plant genes involved in nodulation has focused especially on genes that are expressed in root nodules but not in other parts of the plant. The legume nodule specific proteins have been named nodulins. This name was first proposed in 1984 [70]. Since then many nodulin genes have been identified in legume nodules, and more recently nodule-specific genes from actinorhizal nodules have been cloned. Since the discovery of the first nodulin genes, methods to detect transcripts, for example PCR-based methods, have become markedly more sensitive. By applying such methods, it has become clear that many of the "nodule specific" genes are in fact also expressed in other parts of the plants. This suggests that these genes have been recruited from other developmental processes. The best information on the origin of "nodule specific" genes comes from studies on leghaemoglobins.

**Haemoglobins**

Leghaemoglobins were the first nodule-specific proteins identified in legumes, and haemoglobins have also been found in several actinorhizal nodules. These proteins play an important role in the regulation of the free oxygen concentration inside the nodule. The nitrogenase enzyme of the bacteria is very oxygen-sensitive and for this reason the free oxygen concentration inside nodules has to be low. However, the high energy requirement of the nitrogen fixation process is generated by oxidative phosphorylation. Leghaemoglobins play a key role in this oxygen paradox since they facilitate the flux of oxygen to the bacteria at

a low free oxygen concentration. In this way nitrogenase is protected from oxigen damage while sufficient ATP can be formed to support the nitrogen fixation biochemistry.

For a long time, plant haemoglobins were considered to be nodule-specific proteins. However, haemoglobin genes have now been identified in several plants that are unable to form nitrogen-fixing nodules [8, 64, and 66]. The comparison of their DNA sequences provides some insight into their phylogeny.

Plants that form nodules often have haemoglobin genes that are exclusively expressed in non-symbiotic organs, whereas other haemoglobin genes are expressed in the root nodules. Based on their DNA sequence these "non-symbiotic" and "symbiotic" haemoglobin genes can be classified as class 1 and class 2 haemoglobin genes, respectively [66]. An exception is the *Parasponia* haemoglobin gene that is expressed at high level in nodules [36]. This gene is structurally related to the class 1 genes. *Parasponia* nodules do not contain (at detectable level) a Class 2 haemoglobin "symbiotic" protein [3] and whether or not *Parasponia* even has a Class 2 haemoglobin gene is unknown.

Class 1 haemoglobins have been found in a wide variety of non-nodulating plants including monocots (e.g. rice) as well as dicots (e.g. *Arabidopsis*). The class 1 genes are highly conserved. Class 1 genes of dicots and monocots are even more similar in sequence than class 1 and class 2 haemoglobins within a dicot species. This suggests that class 1 haemoglobins originated before monocots and dicots separated [65]. It is unknown whether this is also the case for class 2 haemoglobins, since these haemoglobins have not yet been identified in monocots. A class 2 haemoglobin has recently been identified in *Arabidopsis*, a non- nodulating dicot, demonstrating that the class 2 ("symbiotic") haemoglobins are indeed not restricted to plants of the Rosid I clade. Although *Arabidopsis* belongs to the Rosid II clade that is relatively closely related to the Rosid I clade, this finding supports the idea that class 2 haemoglobins might be widespread in the plant kingdom.

Thus, in general, among the plants that have been studied, nodulating plants use class 2 haemoglobins in the symbiotic interaction, but in one case (*Parasponia*) a class 1 gene is used to produce high amounts of haemoglobin in nodules. This suggests that both haemoglobin classes can facilitate oxygen flux to the bacteria at a low free oxygen concentration, and it also suggests that properties of the promoters determine which haemoglobin is used.

Haemoglobin genes of several legumes, *Parasponia* and *Casuarina* (an actinorhizal plant) have been extensively studied. Studies on haemoglobin expression in heterologous systems showed that the promoter of an actinorhizal (*Casuarina*) class 2 haemoglobin maintains its nodule-specific expression when introduced in legumes [31]. Conversely: legume class 2 leghaemoglobin promoters are specifically active in actinorhizal nodules [21]. In both cases the haemoglobin promoters were also expressed in the proper cell type, the infected cell. This suggests that the transcription machinery leading to nodule-specific haemoglobin expression in actinorhizal and legume nodules is conserved.

This further suggests that the promoter of the *Parasponia* class 1 haemoglobin gene, which is highly expressed in nodules, may be similar to the promoters of actinorhizal and legume class 2 haemoglobin genes. The *Parasponia* (class 1) haemoglobin promoter shows a high level of expression in legume as well as actinorhizal nodules [2, 21]; however, when the *Parasponia* haemoglobin gene was introduced into *Lotus corniculatus*, most of the expression occurred in uninfected nodule cells. Thus, it seems that while the infected cell- specific transcription machinery of *Lotus corniculatus* is able to recognize actinorhizal promoters properly, it recognizes the *Parasponia* class 1 haemoglobin promoter less efficiently.

Trema is a very close relative of *Parasponia* but is unable to form root nodules. A comparison of the *Parasponia* haemoglobin gene with that of *Trema* indicates that the nodule enhanced expression of the Parasponia promoter is a recently acquired property. The *Trema* class 1 haemoglobin promoter is not expressed at elevated levels in nodules of transgenic legumes or actinorhizal plants. Taken together, these data suggest that in *Parasponia* a class 1 globin gene acquired the ability to be expressed at an elevated level in nodules. Thus, nodule specific expression of haemoglobin genes seems to have evolved more than once.

On the other hand, the fact that legume, actinorhizal and *Parasponia* symbiotic haemoglobin genes are certainly expressed in heterologous nodules, suggests that in these plant groups, similar transcription factors are used to regulate these genes. The identification of these transcription factors, and the study of their distribution and action within and beyond the Rosid I clade species, will give insight as to whether they are part of a plant's predisposition for nodulation.

**Rhizobial Nod factor structure and legume phylogenies.**

Host specificity is a striking characteristic of nodule symbiosis, and since it has especially been studied for the interaction of legumes and rhizobia, we will focus on this specific system in the following discussion.

By using 16S rRNA sequences it has been shown that rhizobia interacting with members of the Leguminosae are very diverse and do not form a discrete clade [80] (Figure 5). Several closely related legume species are infected by distantly related rhizobia and conversely closely related rhizobia can interact with legume species belonging to different tribes. This is illustrated by the rhizobia that nodulate plants belonging to the tribe Phaseoleae (Figure 3IV, tribe 5) which contains the genera *Phaseolus*, *Vigna*, *Glycine*, and *Cajanus*. *Phaseolus* species are nodulated by *Rhizobium leguminosarum* biovar *phaseoli*. The other biovars of this *Rhizobium* species, *R.l.* bv *triflolii* and *R.l.* bv *viceae*, nodulate legume genera belonging to completely different tribes, namely Trifolieae (Figure 3IV, tribe 3) and Vicieae (Figure 3IV, tribe 2), respectively. Moreover, *Glycine* (belonging to the same tribe as Phaseolus) cannot be nodulated by bacteria of the genus *Rhizobium*, but only by *Bradyrhizobium* and *Sinorhizobium* species. How is it possible that such diverse rhizobial species can infect closely related plant species? Since the ability to interact with a certain legume host is determined by the nature of the rhizobial signal molecules [17], it seems probable that a comparison of the structure of rhizobial signal molecules, could provide insight into why unrelated rhizobia can nodulate closely related legumes. Key signal molecules in legume nodulation are the rhizobial Nod factors, which are sufficient to induce early steps of nodulation [28].
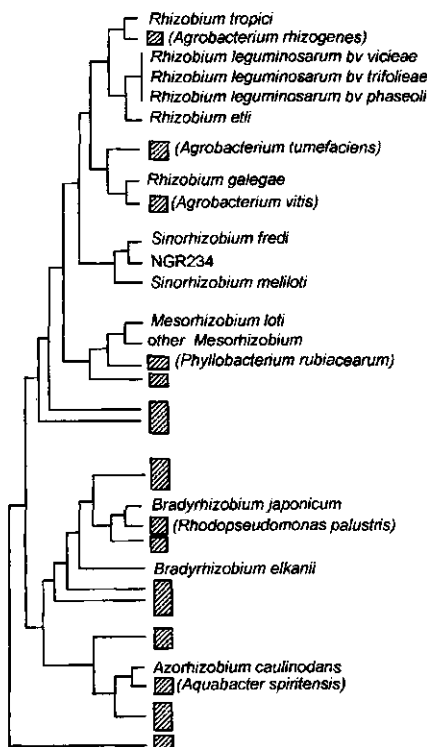
**Figure 5**

Phylogeny of rhizobia as deduced from 16S ribosomal RNA gene sequences. Linages of non-symbiotic bacteria are shown by boxes, with a representative given in parentheses for linages closely related to symbiotic groups. Modified after Doyle [18].

Nod factors are lipochito-oligosaccharides, consisting of a backbone of three to five B-1,4-linked N-acetylglucosamines bearing a fatty acid on the non-reducing sugar residue. In addition, Nod factors can have various substitutions on the reducing and/or non-reducing terminal sugar residue [for review on Nod factor structure and biosynthesis see 9, 16, 20, 61]. The rhizobial genes involved in the biosynthesis of the Nod factor core are the common *nodABC* genes [24, 33, 52]. Other *nod* genes determine the nature of the acyl moiety and the presence of substitutions at the terminal sugar residues, respectively.

Rhizobia nodulating the related tribes Trifolieae (Figure 3IV, tribe 3), Vicieae (Figure 3IV, tribe 2) and Galegeae (Figure 3IV, tribe 1) belong to unrelated groups, namely *Rhizobium* senso stricto, *Sinorhizobium* and *Mesorhizobium*. However, only the rhizobia interacting with these legumes produce Nod factors containing polyunsaturated fatty acids. Thus, although these bacteria are phylogenetically unrelated, they produce Nod factors with a

similar structure [76]. The *nodE* and *nodF* genes are involved in the production of polyunsaturated acyl chains. These three groups of rhizobia also contain rhizobial species that do not produce Nod factors containing polyunsaturated fatty acids and these species lack the *nodE* and *nodF* genes. These findings strongly suggest that the bacteria acquired genes determining Nod factor structure by intergeneric horizontal gene tansfer [76]. This can explain why distantly related rhizobia posses similar genes whereas some closely related species contain different subsets of genes. In contrast with the proposed intergeneric horizontal transfer of the *nodE* and *nodF* genes, recent studies have shown that the common *nod* genes, *nodB* and *nodC*, are most likely transfered among congeneric rhizobial species [74] . These studies indicate that the common *nod* genes have been horizontally transferred between rhizobia species belonging to the same genera, whereas transfer between rhizobia of different genera does not seemed to have occured [74].

Thus, the variations in Nod factor structure are clearly correlated with the phylogeny of the legume hosts, since distant rhizobia that can nodulate closely related legume species do secrete Nod factors with a similar structure. Moreover, the lack of correlation between rhizobia phylogenies with Nod factor structure and plant phylogenies suggest that rhizobia obtained *nod* genes by horizontal gene transfer.

**Endomycorrhizae and nodule formation**
In addition to nitrogen fixing soil bacteria, mycorrhizal fungi belonging to the order Glomales can also estabish endosymbiotic relationships with plant roots. In contrast to root nodules, arbuscular endomycorrhiza (AM), an endosymbiotic association between plant roots and mycorrhizal fungi, can be formed by the vast majority of higher plants (about 80%). In this symbiotic interaction, the fungus colonizes the root surface and, after the attachment of hyphae to the root surface, a swollen structure, the appressorium, is formed at the tip of the hypha. New hyphae develop from the appressorium, and these enter the root in either an intra- or intercellular way depending on the host plant. Inside the root, the hyphae grow towards the inner cortex where cells are infected, and the hyphae differentiate into highly branched structures called arbuscules [for review see 25, 27, 59]. During arbuscle formation, the morphology of the infected cortical cells markedly changes. However, in contrast to the cortical cells forming nodule primordia, they do not divide. The hyphae that are outside the plant roots form a continuum with the arbuscles, and in this way the fungus connects the inner root and the soil, facilitating the uptake of nutrients such as phosphate.

The morphology of the two endosymbiotic structures, root nodules and arbuscular endomycorrhiza, is different. Furthermore, the level of host specificity controlling the endosymbiotic interactions is markedly different. Whereas rhizobia, in general, only interact with species of a single family, endomycorrhizal fungi can interact with most higher plants. Therefore, it seemed probable that nodulation and endomycorrhiza formation were unrelated processes, and thus the discovery of Duc *et al.* that common host genes are essential for both interactions was a great surprise [19]. This group showed that some pea mutants that are blocked in *Rhizobium*-induced nodulation (Nod$^-$) have also lost the ability to interact with mycorrhizal fungi (Myc$^-$) [19]. Other studies have now shown that several nodulation mutants of *Medicago* [54], *Lotus* [73] and *Phaseolus* [57] are also Myc$^-$. Among the best studied Nod$^-$Myc$^-$ legume mutants are the *Sym8* of pea [discussed in 25], and the *Medicago* mutant MN NN 1008 [discussed in 27]. In both mutants, the mycorrhizal fungi can still form appressoria, but they are unable to form hyphae that penetrate the root. As far as has been studied, rhizobia do not induce responses in these mutants.

After the discovery that common host genes are essential in nodulation and endomycorrhiza formation, several groups have studied whether nodulin genes are also activated during the interaction with mycorrhizal fungi. Such studies showed that, for example, the early nodulin genes *ENOD2*, *ENOD40* [71], *ENOD5*, *ENOD12* [1], the leghaemoglobin gene *VFLb29* [23] and the aquaporin-encoding *NOD26* [75] are activated during endomycorrhiza formation. The early nodulin genes *ENOD12* and *ENOD5* are activated in cells that are or will become infected by rhizobia. Furthermore, rhizobial Nod factors are sufficient to induce the expression of these genes. When plants become infected by mycorrhizal fungi, these genes are transiently activated during the infection stage. Thus, it is probable that in both interactions these early nodulin genes have a function in the infection process. The induction of *ENOD5* and *ENOD12* by either microsymbiont requires a functional *sym8* gene [1]. Thus, it was shown that the mechanism by which these nodulin genes are activated is probably similar, and at least has *Sym8* in common.

The studies reviewed above have made it clear that common host genes are involved in the rhizobial and mycorrhizal interaction. This finding has important implications since, in contrast to *Rhizobium*, arbuscular mycorrhizal (AM) fungi have the ability to interact with a wide range of higher plants. Assuming that the mechanisms by which AM fungi infect their various hosts are similar, it implies that *sym* and *ENOD* genes, involved in the interaction of

legumes with both micro-symbionts, are most probably widespread in the angiosperm. Furthermore, it becomes probable that several nodulation functions have been recruited from the more ancient endomycorrhizal symbiosis [38]. Considering that endomycorrhiza symbiosis is widespread within the terrestrial flowering plants, it can be postulated that the mechanism of endomycorrhiza formation evolved in the ancestors of higher plants and, later on, a Rosid I ancestor recruited part of these mechanisms for nodule endosymbiosis.

**Nod factor recognition by non-legumes**

The involvement of common host genes in nodulation and mycorrhizae formation suggests that several genes that are essential for nodulation are widespread in the plant kingdom. This together with the fact that the non-legume *Parasponia* can recognize the rhizobial Nod factors, raises the question as to whether these rhizobial signal molecules are recognised by other non-legumes as well.

The first indication that Nod factor-like molecules are recognized by non-legumes came from studies of de Jong *et al.* [15] who studied a mutant carrot suspension cell line that is blocked in somatic embryogenesis. This mutant line can be rescued by the addition of a chitinase that is secreted by wild type suspension cells. It was postulated that this chitinase could hydrolyze an unknown inactive precursor containing an N-acetyl glucosamine residue, resulting in an active growth factor. Since Nod factors contain a chitin backbone they were tested for the ability to rescue the embryo mutant phenotype, and this was indeed the case. After this discovery, a few reports were published showing that Nod factor-like molecules are recognized by tobacco protoplasts but these observations were not reproducible [55].

The most convincing data showing that Nod factors are recognized by non-legumes came from recent studies with transgenic rice (*Oryza sativa*)[50]. These studies showed that a *Medicago ENOD12* promoter in transgenic rice can be activated by rhizobial Nod factors. This demonstrates that the perception and transduction machinery required for the activation of this leguminous promoter by Nod factors is present in rice. Although the function of this non-leguminous perception/transduction mechanism is not clear, it seems probable that it has a widespread occurrence, and that the Nod factor perception/transduction mechanism of legumes has evolved from it.

## Concluding remarks

Only a small group of plants have the ability to make nitrogen-fixing root nodules, which suggests that this ability involves rather unique properties of the host plant. This idea is further strengthened by the fact that all nodule-forming plants belong to a single clade.

The relatively close relationship of the nodule-forming plants contrasts with the fact that rhizobia and *Frankia* induce nodules that have a different ontogeny and tissue organization, suggesting that different mechanisms underlie the development of these two nodule types. Most actinorhizal plants form two different symbiotic organs; most prominent is the nodule that is composed of lateral root-like lobes, whereas the prenodule that is formed in the cortex only forms a small protrusion. Based on the expression of marker genes, the prenodule appears to be an organ on its own, and its formation seems to be independent of the formation of the lateral root-like nodule. Rhizobia induce the formation of a single symbiotic organ, however the formation of this organ appears to involve two processes that to a certain extent can be uncoupled: the formation of a nodule primordium and a "connecting" vascular bundle, respectively. It is possible that the formation of a legume nodule primordium is related to the prenodule formation, whereas the "connecting" vascular bundle results from a process more related to the actinorhizal nodule lobe formation. If this hypothesis is correct, then these two nodule types are closely related, despite their different morphology.

Although nodulation is a unique property of a small closely related group of host plants, several data show that genes and processes underlying nodulation are in fact recruited from processes that are widespread among the higher plants. Several plant genes have been identified that are expressed at elevated levels in root nodules. Many of these genes were considered to be nodule-specific. However, more sensitive detection methods showed that most of these genes are also expressed in other (non-symbiotic) organs of the host plant. Some of these nodule specific genes were also considered to be legume/actinorhizal plant-specific, but now their homologous genes have been found in plants unable to establish a nodule symbiosis. The idea that the nodulation process is (in part) derived from processes that are widespread among the higher plants is furthermore illustrated by the facts that: infection involves steps derived from the cell cycle machinery; nodulation and the widespread endomycorhizal symbiosis are related processes; and the presence of a Nod factor perception and transduction machinery in rice plants.

So, on the one hand, the phylogenetic data show that the nodule-forming plants form a closely related group, stressing the uniqueness of nodule-forming ability. On the other hand, it has become clear that processes that contribute to nodule development have been recruited from processes that are common to most if not all higher plants. This still leaves us with the question: what properties distinguish the nodule forming Rosid I clade plants from the other higher plants? Perhaps a few unidentified symbioses specific genes determine the predisposition to nodulation. Genome projects might reveal such genes. At present genome projects have been started on several model plants, and it is to be expected that the sequences of the genomes of *Arabidopsis* and rice will be available within a few years. More recently, a genome project has been initiated on a model legume, *Medicago truncatula*. The comparison of such genomes might give us the ultimate answers as to whether genes unique to nodulating plants indeed exist.

## Acknowledgements

**References**

1. Albrecht C, Geurts R, Lapeyrie F, Bisseling T: Endomycorrhizae and rhizobial Nod factors both require SYM8 to induce expression of the early nodulin genes *PsENOD5* and *PsENOD12A*. Plant J 15(5):605-614 (1998).

2. Andersson CA, Llewellyn DJ, Peacock WJ, Dennis ES: Cell-specific expression of the promoters of two nonlegume haemoglobin genes in transgenic legume, *Lotus corniculatus*. Plant Physiol 113: 45-57(1997).

3. Appleby CA, Tjepkema JD, Trinick MJ: Haemoglobin in a Non-leguminous Plant, *Parasponia*: Possible genetic Origin and Function in Nitrogen Fixation. Science 220: 951-953 (1983).

4. Basset B, Goodman RN, Novacky A: Ultrastructure of soybean nodules. I. Release of rhizobia from the infection thread. Can J Microbiol 23:573-582 (1977).

5. Benson DR, Silvester WB: Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol Rev 57:293-319 (1993).

6. Bergman B, Rai AN, Johansson C, Soderback E: Cyanobacterial-plant symbiosis. Symbiosis 14: 61-81 (1992).

7. Berry AM, Sunell LA: The infection process and nodule development. In Schwintzer CR, Tjepkema JD (eds) The biology of *Frankia* and Actinorhizal plants, pp 61-68 Academi Press, New York (1990).

8. Bogusz D, Appleby CA, Landsmann J, Dennis ES, Trinick MJ, Peacock WJ: Functioning haemoglobin genes in non-nodulating plants. Nature 331: 178-180 (1988).

9. Carlson RW, Price NPJ, Stacey G: The biosynthesis of rhizobial lipo-oligosaccharide signal molecules. Mol Plant-Microbe Interact 7:684-695 (1994).

10. Chandler MR, Date RA, Roughley RJ: Infection and root nodule development in *Stylosanthes* species by *Rhizobium*. J Exp Bot 33:47-57 (1982).

11. Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD: Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Annals of the Missouri Botanical Garden 80:528-580 (1993).

12. Cronquist, A: An integrated System of Classification of Flowering Plants. Columbia University Press, New York (1981).

13. Dart PJ: The infection process. In A. Quispel (ed) The Biology of Nitrogen Fixation, pp 381-429 Amsterdam: North Holland Publishing Co (1974).

14. de Faria SM, Hay GT, Sprent JI: Entry of rhizobia into roots of *Mimosa scabrella* Bentham occurs between epidermal cells. J Gen Microbiol 134: 2291-2296 (1988).

15. de Jong AJ, Heidstra R, Spaink HP, Hartog MV, Meijer EA, Hendriks T, Lo Schiavo F, Terzi M, Bisseling T, van Kammen A, de Vries SC: *Rhizobium* lipo-oligosaccharides rescue a carrot somatic embryo variant. The Plant Cell 5: 615-620 (1993).

16. Dènariè J, Cullimore J: Lipo-oligosaccharide nodulation factors: A new class of signaling molecules mediating recognition and morphogenesis. Cell 74: 951-954 (1993).

17. Dènariè J, Debelle F, Prome JC: *Rhizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. Annu Rev Biochem 65:503-535 (1996).

18. Doyle JJ: Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. Trends Plant Sci 3(12): 473-478 (1998).

19. Duc G, Trouvelot A, Gianinazzi-Pearson V, Gianinazzi S: First report of non-mycorrhizal plant mutants (myc⁻) obtained in pea (*Pisum sativum*) and Fababean (*Vicia Faba* L.). Plant Sci 60:215-222 (1989).

20. Fisher RF, Long SR: *Rhizobium*-plant signal exchange. Nature 357: 655-660 (1992).

21. Franche C, Diouf D, Laplaze L, Auguy F, Frutz T, Rio M, Duhoux E, Bogusz D: Soybean (*lbc3*), *Parasponia*, and *Trema* Haemoglobin Gene Promoters Retain Symbiotic and Nonsymbiotic Specificity in Transgenic Casuarinaceae: Implications for Haemoglobin Gene Evolution and Root Nodule Symbiosis. Mol Plant-Microbe Interact 11(9): 887-894 (1998).

22. Franssen HJ, Vijn I, Yang WC, Bisseling T: Developmental aspects of the *Rhizobium*-legume symbiosis. In Schilperoot R, Dure L (eds) 10 Years Plant Molecular Biology Vol 19 pp 89-107 Kluwer Academic Publishers (1992).

23. Fruhling M, Roussel H, Gianinazzi-Pearson V, Puhler A, Perlick AM: The *Vicia faba* leghaemoglobin gene *VfLb29* is induced in root nodules and in roots colonized by the arbuscular micorrhizal fungus *Glomus fasciculatum*. Mol Plant-Microbe Interact 10:124-131 (1997).

24. Geremia RA, Mergaert P, Geelen D, van Montagu M, Holsters M: The NodC protein of *Azorhizobium caulinodans* is an N-acetylglucosaminyltransferase. Proc Natl Acad Sci USA 91:2669-2673 (1994).

25. Gianinazzi-Pearson V: Plant cell responses to arbuscular mycorrhizal fungi: getting to the roots of the symbiosis. Plant Cell 8:1871-1883 (1996).

26. Hadri A-E, Spaink HP, Bisseling T, Brewin NJ: Diversity of root nodulation and rhizobial infection processes. In: Spaink HP, Kondorosi A, Hooykaas PJJ (eds) The Rhizobiaceae pp 347-359 Kluwer Academic Publishers (1998).

27. Harrison MJ: The arbuscular mycorrhizal symbiosis: an underground association. Trends Plant Sci 2:54-60 (1997).

28. Heidstra R, Bisseling T: Nod factor-induced host responses and mechanism of Nod factor perception. New Phytol 133: 25-43 (1996).

29. Heidstra R, Yang WC, Yalcin Y, Peck S, Emons AM, Van Kammen A, Bisseling T: 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: 1781-1787 (1997).

30. Huss-Danell K, Bergman B: Nitrogenase in *Frankia* from root nodules of Alnus incana (L.) Moench: Immunolocalization of the Fe- and MoFe-proteins during vesicle differentiation. New Phytol 116: 443-455 (1990).

31. Jacobsen-Lyon K, Jensen EO, Jorgensen J, Marker KA, Peacock WJ, Dennis ES: Symbiotic and non-symbiotic haemoglobin genes of *Casuarina glauca.* Plant Cell 7: 213-222 (1995).

32. James EK, Sprent JI, Sutherland JM, McInroy SG, Minchin FR: The structure of nitrogen fixing root nodules on the aquatic mimosoid legume *Neptunia plena.* Ann Bot 69:173-180 (1992).

33. John M, Rôhring H, Schmidt J, Wieneke U, Schell J: *Rhizobium* NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. Proc Natl Acad Sci USA 90:625-629 (1993).

34. Kijne JW: The *Rhizobium* infection process. In: Stacey G, Burris RH,Evans HJ (eds) Biological Nitrogen Fixation pp 349-398, New York: Chapman and Hall (1992).

35. Lancelle SA, Torrey JG: Early development of *Rhizobium*-induced root nodules of *Parasponia* rigida. II. Nodule morphogenesis and symbiotic development. Can J Bot 63: 25-35 (1984)

36. Landsmann J, Dennis ES, Higgins TJ, Appleby CA, Kortt AA, Peacock WJ: Common evolutionary origin of legume and non-legume plant haemoglobins. Nature 324: 166-168 (1986)

.
37. Laplaze L., Ribeiro A., Franche C., Duhoux E., Auguy F., Bogusz D., Pawlowski K: Characterization of a Casuarina glauca nodule-specific subtilisin-like protease gene, a homolog of Alnus glutinosa ag12, Mol. Plant Microbe Interact. 13 (1), 113-117 (2000)

38. LaRue TA, Weeden NF: The Symbiosis Genes of the Host. In: Kiss GB, Endre G (eds) Proceedings of the 1st European Nitrogen Fixation Conference 147- 151 Officina Press, Szeged, Hungary 147- 151 (1994)

39. Lee KH, LaRue TA: Exogenous ethylene inhibits nodulation of *Pisum sativum L.* cv Sparckle. Plant Physiol 100: 1759-1763 (1992).

40. Libbenga KR, Harkes PAA: Initial proliferation of cortical cells in the formation of root nodules in *Pisum sativum L.* Planta 114:17-28 (1973a).

41. Libbenga KR, van Iren F, Bogers RJ, Schraag-Lamers MF: The role of hormones and gradients in the initiation of cortex proliferation and nodule formation in *Pisum sativum L.* Planta 114: 29-39 (1973b).

42. Marvel DJ, Torrey JG, Ausubel FM: *Rhizobium* symbiotic genes required for nodulation of legume and nonlegume hosts. Proc Natl Acad Sci USA 84: 1319-1323 (1987).

43. Miller IM, Baker DD: The initiation, development and structure of root nodules in *Elaeagnus angustifolia L.* (Elaeagnaceae). Protoplasma 128: 107-119 (1985).

44. Naisbitt T, James EK, Sprent JI: The evolutionary significance of the genus *Chamaecrista*, as determined by nodule structure. New Phytol 122: 487-492 (1992).

45. Newcomb W: Nodule morphogenesis and differentiation. Int Rev Cytol 13 (suppl): S247-S297 (1981).

46. Patriarca EJ, Tate R, Riccio A, Fedorovoa E, Defez R, Iaccarino M: Down-regulation of the *Rhizobium ntr* system in the determinate nodule of *Phaseolus vulgaris* identifies a specific developmental zone. Mol Plant-Microbe Interact 9:243-251 (1996).

47. Pawlowsky K, Bisseling T: Rhizobial and Actinorhizal symbioses: What are the Shared Features? Plant Cell 8:1899-1913 (1996).

48. Penmetsa RV, Cook DR: A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. Science 275: 527-530 (1997).

49. Racette S, Torrey JG: Root nodule initiation in *Gymnostoma* (Casuarinaceae) and *Shepherdia* (Elaeagnaceae) induced by *Frankia* strain HFPGpl1. Can J Bot 67: 2873-2879 (1989).

50. Reddy PM, Ladha JK, Ramos MC, Maillet F, Hernandez RJ, Torrizo LB, Oliva NP, Datta SK: Rhizobial lipochitooligosaccharide nodulation factors activate expression of the legume early nodulation gene *ENOD12* in rice. Plant J 14: 693-702 (1998).

51. Ribeiro A, Akkermans ADL, van Kammen A, Bisseling T, Pawlowsky K: A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development. Plant Cell 7: 785-794 (1995).

52. Rôhring H, Schmidt J, Wieneke U, Kondorosi E, Barlier I, Schell J, John M: Biosynthesis of lipooligosaccharide nodulation factors: *Rhizobium* NodA protein is involved in N-acylation of the chitooligosaccaride backbone. Proc. Natl. Acad. Sci USA 91:3122-3126 (1994).

53. Roth LE, Stacey G: Bacterium release into host cells of nitrogen-fixing soybean nodules: The symbiosome membrane comes from three sources. Eur J Cell Biol 49:13-23 (1989).

54. Sagan M, Morandi D, Tarenghi E, Duc G: Selection of nodulation and mycorrhizal mutants in the model plant *Medicago truncatula* (Gaertn.) after gamma-ray mutagenesis. Plant Sci 111:63-71 (1995).

55. Schell J, Schmidt J, Walden R: Notice. Trends in Plant Sci 3:130 (1998).

56. Schwintzer CR, Berry AM, Disney LD: Seasonal patterns of root nodule growth, endophyte morphology, nitrogenase activity and shoot development in *Myrica gale*. Can J Bot 60:746-757 (1982).

57. Shirtliffe SJ, Vessey JK: A nodulation (Nod+/Fix-) mutant of *Phaseolus vulgaris L.* has nodule like structures lacking peripheral vascular bundles (Pbv⁻) and is resistant to mycorrhyzal infection (Myc⁻). Plant Sci 118:209-220 (1996)

58. Smit G, de Koster CC, Schripsema J, Spaink HP, Van Brussel AAN, Kijne JW: Uridine, a cell division factor in pea roots. Plant Mol Biol 29: 869-873 (1995).

59. Smith FA, Smith SE: Structural diversity in (vesicular)-arbuscular mycorrhizal symbioses. New Phytol 137:373-388 (1997).

60. Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG: Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc Natl Acad Sci USA 92:2647-2651 (1995)

61. Spaink HP: Rhizobial lipo-oligosaccharides: Answers and questions. Plant Mol Biol 20:977-986 (1992)

62. Subba-Rao NS, Mateos PF, Baker D, Pankratz HS, Palma J, Dazzo FB, Sprent JI: The unique root-nodule symbiosis between *Rhizobium* and the aquatic legume, *Neptunia natans* (L.F.) Druce. Planta 196, 311-320.

63. Swensen SM: The evolution of actinorhizal symbiosis: evidence for multiple origins of the symbiotic association. Am J Bot 83:1503-1512 (1996).

64. Taylor ER, Nie XZ, MacGregor AW, Hill RD: A cereal haemoglobin gene is expressed in seed and root tissues under anaerobic conditions. Plant Mol Biol 24: 853-862 (1994).

65. Trevaskis JB: The characterization of haemoglobins from *Arabidopsis thaliana*. PhD Thesis, Australian National University (1997).

66. Trevaskis B, Watts R, Andersson CR, Llewellyn DJ, Hargrove MS, Olson JS, Dennis ES, Peacock WJ: Two haemoglobin genes in *Arabidopsis thaliana*: The evolutionary origins of leghaemoglobins. Proc Natl Acad Sci USA 94: 12230-12234 (1997).

67. Tsien HC, Dreyfus BL, Schmidt EL: Initial stages in the morphogenesis of nitrogen-fixing stem nodules of *Sesbania rostrata*. J Bacteriol 156:888-897 (1983).

68. Van Berkun P, Eardly BD: Molecular evolutionary systematics of the Rhizobiaceae. In: Spaink HP, Kondorosi A, Hooykaas PJJ (eds) The Rhizobiaceae, pp. 9-20. Kluwer Academic Publishers (1998).

69. Van Brussel AAN, Bakhuizen R, van Spronsen PC, Spaink HP, Tak T, Lugtenberg BJJ: Induction of pre-infection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of *Rhizobium*. Science 257:70-71 (1992).

70. Van Kammen A: Suggested nomenclature for plant genes involved in nodulation and symbiosis. Plant Mol Biol Rep 2:43-45 (1984).

71. Van Rhijn P, Fang Y, Galili S, Shaul O, Atzmon N, Wininger S, Eshed Y, Lum M, Li Y, To V, Fujishige N, Kapulnik Y, Hirsch A: Expression of early nodulin genes in alfalfa mycorrhizae indicates that signal transduction pathways used in forming arbuscular mycorrhizae and *Rhizobium*-induced nodules may be conserved. Proc   Natl   Sci USA 94:5467-5472 (1997).

72. Vasse J, De Billy F, Camut S, Truchet G: Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. J Bacteriol 172: 4295-4306 (1990).

73. Wegel E, Schauser L, Sandal N, Stougaard J, Parniske M: Mycorrhiza mutants of *Lotus Japonicus* define genetically independent steps during Symbiotic infection. Mol Plant-Microbe Interact 11(9): 933-936 (1998).

74. Wernegreen JJ, Riley MA: Comparison of the evolutionary dynamics of symbiotic housekeeping loci: A case for the genetic coherence of rhizobial linages. Molecular Biology and Evolution. 16(1):98-113 (1999).

75. Wyss P, Mellor RB, Wiemken A: Vesicular-arbuscular mycorrhizas of the wild-type soybean and non-nodulating mutants with *Glomus mossae* contain symbiosis-specific polypeptides (micorrhizins), immunologically cross-reactive with nodulins. Planta 182:22-26 (1990).

76. Yang GP, Debelle F, Ferro M, Maillet F, Schiltz O, Vialas C, Savagnac A, Prome JC, Dènariè J: *Rhizobium* Nod factor structure and the phylogeny of temperate legumes. In: Elmerich C. (eds) Biological Nitrogen Fixation for the 21st Century pp. 185-188 Kluwer Academic Publishers (1998).

77. Yang WC, Horvath B, Hontelez J, van Kammen A, Bisseling T: In situ localization of *Rhizobium* mRNAs in pea root nodules: *nifA* and *nifH* localization. Mol Plant-Microbe Interact 7:276-281 (1991).

78. Yang WC, Katinakis P, Hendriks P, Smolders A, de Vries F, Spee J, van Kammen A, Bisseling T, Franssen H: Characterization of *GmENOD40*, a gene showing novel patterns of cell-specific expression during soybean nodule development. Plant J 3:573-585 (1993).

79. Yang WC, de Blank C, Meskiene I, Hirt H, Bakker J, van Kammen A, Franssen H, Bisseling T: *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 (1994).

80. Young JPW: Phylogeny and taxonomy of rhizobia. Plant and Soil 186:45-52 (1996)

# CHAPTER 2

**Towards the cloning of the pea *Sym2* gene by using *Medicago truncatula* as intergenomic cloning vehicle: A region on *M. truncatula* chromosome 5 is highly microsyntenic to the pea *Sym2*-containing region**

Gustavo Gualtieri[1], Olga Kulikova[1], Douglas Cook[2] and Ton Bisseling[1]

1) Molecular Biology, Agricultural University Wageningen. Dreijenlaan 3, 6703HA Wageningen, The Netherlands.
2) Department of Plant Pathology, University of California at Davis. One Shields Avenue. Davis, CA 95616-8680. U.S.A.

## ABSTRACT

The pea *Sym2* gene controls the Nod factor structure-dependent infection by the symbiotic *Rhizobium* bacteria. The previously isolated cDNA RFLP marker *PscW62-1* that is tightly linked to *Sym2* was used to screen a *M. truncatula* A17 BAC library and 11 clones were isolated. These clones formed 3 physically unlinked contigs named c1, c2, and c3, that were further extended by chromosome walking. The three contigs are located on *M. truncatula* A17 chromosome 5. c1 and c2 are tightly linked, while genetic mapping showed that c3 is located on the same arm of chromosome 5 at a distance of 9 cM from the c1/c2 region. Microsynteny between c1/c2 and the pea *Sym2*-containing genomic region was studied by RFLP analysis of cDNAs and genomic subclones isolated from these contigs. This showed a high level of microsynteny between these genomic regions of pea and *M. truncatula*. In addition, a genomic subclone from c2, highly homologous to the LRR motif of *Cf4* and *Cf9 L. esculentum* disease resistance genes, was used to isolate the cDNA *PscLRR52* from a pea root hair cDNA library. RFLP mapping showed that *PscLRR52* is tightly linked to *Sym2*, demonstrating the value of *M. truncatula* as intergenomic gene cloning vehicle. Moreover, one RFLP marker from c2 and two RFLP markers from c1 that were isolated from BACs obtained by chromosome walking have shown recombinations between the corresponding pea loci and *Sym2*, thus delimitating the *Medicago Sym2*-orthologous region to about 350 Kbp and confirming that *PscW62-1* was indeed tightly linked to *Sym2*. In addition, the delimitation of this orthologous region shows that the order of clusters of studied genomic sequences and genes is conserved between c1/c2 and the pea *Sym2*-containing region. The cloning and delimitation of the *M. truncatula Sym2*-orthologous region, provides a promising start for the microsynteny-based positional cloning of the pea *Sym2* gene.

## INTRODUCTION

The legume sub-family Papilionoidae contains many economically important species like soybean, alfalfa and pea. A unique property that contributes to their agronomic importance is their capacity to establish a symbiosis with rhizobia that results in the formation of nitrogen fixing root nodules. Several of these crop legumes are genetically well characterized, but due to their large genome or complex ploidy, they are not amenable for positional cloning strategies. An example is pea (*Pisum sativum*) of which many genes have been genetically mapped. Among these are about 30 *Sym* genes which are essential for the formation of nitrogen fixing root nodules (Borisov et al., 2000). However, at present none of these pea *Sym* genes has been cloned, mainly because pea has a rather large genome. For such reasons, legume species have been selected that could serve as a molecular-genetic model system. The best developed systems are at present *Medicago truncatula* (Cook, 1999) and *Lotus japonicus* (Jiang and Gresshoff, 1997).

It was shown that closely related plant species have genomes with conserved gene order (For reviews see: Bennetzen, 2000; and Schmidt, 2000). *M. truncatula* is a close relative of pea and it has a genome size of about $5x10^8$, which is only about 10% of that of the pea genome. In case the genomes of pea and *M. truncatula* would be colinear in the region of interest, then *M. truncatula* could be used as an intergenomic cloning vehicle for pea genes. The latter is of interest especially if mutations in orthologous genes have not been identified in a model legume.

Linkage group I of pea contains many symbiotic genes located within a region of about 20 cM; e.g. *Nod3*, *Sym2*, *Sym5*, *Sym18* and *Sym19* as well as a cluster of leghaemoglobin genes, *Enod7* and *Enod40* (Weeden *et al.*, 1990; Temnykh *et al.*, 1995a,b; Schneider et al., 1999). This is the region with the highest density of legume genes involved in *Rhizobium* symbiosis. Mutations in some of these genes have unique phenotypes that are not observed in model legume mutants. Thus, in case *M. truncatula* and pea would show colinearity around the loci of interest, the map-based positional cloning of such genes might be feasible. In this study, we focused on the pea *Sym2* locus. *Sym2* was first identified in the pea accession "Afghanistan" (Lie, 1984), where it controls the Nod factor structure-dependent infection process (Firmin et al., 1993; Kozik et al., 1995; Geurts et al., 1997; Ovtsyna et al., 1998). *Sym2*-containing peas are only nodulated by *Rl* bv *viciae* strains containing the *nodX* (or *nodZ*) gene that modifies

the Nod factor structure (Firmin et al., 1993; Ovtsyna et al., 1998), but peas without the Afghanistan *Sym2* allele are nodulated by bacterial strains both with and without the *nodX* gene (Kozik, 1995). Nod factors are the rhizobial signal molecules that are involved in the induction of various host responses (Heidstra et al., 1996). To understand the molecular mechanism by which *Sym2* regulates infection thread formation in relation to Nod factor structure, the cloning of this gene is essential.

Previous attempts to clone the *Sym2* gene relied on the direct isolation of this gene from pea, rather than by intergenomic microsynteny studies. To clone *Sym2* directly in pea, *Sym2*-containing introgression lines and their recurrent backcross parental line were compared by methods such as RAPDs, RNA differential display, and AFLP (Kozik, 1996; Geurts, 1998; Gualtieri, unpublished; respectively). These approaches relied on the identification and isolation of markers specific for the *Sym2*-containing introgressed regions. However, since the pea genome is relatively large, $3.8$-$4.8 \times 10^9$ bp per haploid genome (Ellis, 1993), it is probable that RAPDs (Williams et al., 1990) and AFLP (Vos et al., 1995) markers will mostly generate repetitive DNA markers, which cannot be used for the construction of a contig containing *Sym2*. Thus, methods such as the RNA differential display and its variations (Liang et al., 1992; McClelland et al., 1995) or cDNA-AFLP (Bachem et al., 1996; Money et al., 1996), would be more suitable to isolate single or low copy cDNA markers. Indeed, by using the RNA differential display, two cDNA RFLP markers linked to the pea *nod3* hypernodulation mutant locus were isolated (Gualtieri et al., submitted, c). Moreover, a cDNA named *PscW62-1*(Geurts, 1998; in Gualtieri et al., submitted, a) was isolated by comparing root hair RNA of two nearly isogenic lines that genetically differ (in theory) only at and around the *Sym2* locus. *PscW62-1* was present in the RNA fingerprints of the backcross-parental line but absent in the *Sym2*-containing introgression line (Geurts, 1998; in Gualtieri et al., submitted, a). RFLP mapping showed that *PscW62-1* is tightly linked to *Sym2* (Geurts, 1998; in Gualtieri et al., submitted, a). Therefore, *PscW62-1* was used in the work presented here as a basis to study microsynteny around this locus between *M. truncatula* and pea.

This manuscript describes the isolation of 11 *M. truncatula* BAC clones by using *PscW62-1* as a probe, that correspond to 3 physically unlinked BAC contigs (named c1, c2, and c3) that were further extended by chromosome walking. RFLP mapping in pea showed that several *M. truncatula* genes and genomic subclones located in c1 and c2 are linked to *Sym2*,

demonstrating that the c1/c2 *Medicago* genomic region is highly microsyntenic with the pea *Sym2*-containing region and represents the *M. truncatula Sym2*-orthologous region. Moreover, one RFLP marker (in Gualtieri et al., submitted, a) from c2 and two RFLP markers (this manuscript) from c1, that were isolated from BACs obtained by chromosome walking, have delimitated the *Medicago Sym2*-orthologous region and the pea *Sym2*-region. This confirmed that *PscW62-1* was indeed tightly linked to *Sym2*. Furthermore, in addition to the previously isolated pea *PscW62-1* (Geurts, 1998; in Gualtieri et al., submitted, a), this work resulted in the isolation of a second pea RFLP marker, named *PscLRR52*, which is also tightly linked to *Sym2*, demonstrating the usefulness of *M. truncatula* as intergenomic cloning vehicle of pea genes located within microsyntenic regions.

## RESULTS

### Isolation of BACs containing *PscW62-1* orthologues, and assembly of contigs.

A *M. truncatula* A17 BAC library (Nam et al., 1999) was screened, with the previously isolated *PscW62-1* (Geurts, 1998; in Gualtieri et al., submitted, a), and 11 clones were isolated (in Gualtieri et al., submitted, a). These BAC clones formed 3 physically unlinked contigs that were constructed by restriction/PCR (Fig 1A) and restriction/hybridization (Fig 1B) fingerprinting of individual BAC clones. Contig 1 (c1) had a size of 150 kbp, and contained the BAC clones 58F01, 38C14, 63O10, 18L02, 11I01, 20K04, 03I09, 41N12 (Fig 1C, shown as black horizontal bars). Contig 2 (c2) included BAC clones 15B03 and 45I09, and its total size was 110 kbp (Fig 1C, shown as black horizontal bars). "Contig" 3 (c3) was the single 65 kbp BAC clone 46I13 (Fig 1C, shown as a black horizontal bar). The position of contig sequences homologous to *PscW62-1* is shown in Fig 1C as vertical black bars above the overlapping BAC clones. Sequences with a high homology to pea *PscW62-1* are detected in c1 and c3, whereas c2 contains a sequence that weakly hybridizes with *PscW62-1*.

### Genetic mapping of contigs.

To determine the map position of the 3 BAC contigs in *M. truncatula*, CAPS (cleaved amplified polymorphic sequences) were generated from BAC-end sequences of clone 58F01 of c1 (marker DK6R), clone 45I09 of c2 (marker DK39R) and clone 46I13 of c3 (marker DK3R). Mapping was done using a F2 progeny of 93 individuals (see material and methods). This showed that all 3 contigs are located on linkage group V. DK6R and DK39R are tightly linked and located 3.6 cM and 3.0 cM respectively, to the south of the *Medicago EiI2* gene,

whereas c3 is located 9 cM to the south of c1/c2 (in Gualtieri et al., submitted, a). The chromosomal order of BACs containing these c1 and c2 genetic markers was confirmed by FISH when the orientation of contigs c1 and c2 was determined (see below).

### Extension of the contigs by chromosome walking.

The three contigs were extended by chromosome walking using the *M. truncatula* A17 BAC library. Fig 1C shows the BACs isolated by chromosome walking (in gray), and the original BACs composing the contigs that were isolated by hybridization with *PscW62-1* (in black). c3 was extended in both directions by three bi-directional chromosome walking steps to a final size of 150 Kbp. Both c1 and c2 could be extended on one end. c1 was extended from BAC clone 18L02 by two chromosome walking steps to a final size of 300 Kbp. However, the end of this contig represented by BAC 63010 could not be extended. An end subclone of 63010, named Mtg2511 was used to screen the BAC library and 3 BACs with high homology to the probes were identified, but these were not linking to c1. c2 was extended from BAC clone 15B03 by two chromosome walking steps to a final size of 170 Kbp. Attempts to extend the other end of c2 from 15B03, by using the end subclone Mtg3556, have failed. These results suggest that BAC clones overlapping with the ends of c1 and c2 that could not be extended would be absent in this BAC library. Indeed, when a recently constructed new *M. truncatula* A17 BAC library (Cook et al., unpublished) was screened with Mtg2511 and Mtg3556, both contig ends were extended by the same BAC clone, showing that the gap between c1 and c2 was only 10 Kbp (Geurts, pers comm).

**Fig 1:**
**A and B**: Amplified-restriction fragment fingerprinting (**A**) and restriction/hybridization fingerprinting (**B**) of BAC clones composing c1. Fingerprint bands that are common in between different BACs indicate the overlapping of these BACs in the contigs. These patterns of overlap are used to construct the contig molecular maps. The black arrow head in **A** shows a unique faint band in BAC 03109 that corresponds to the c1 *Medicago* RFLP marker MtgG28. "neg ctrl" in **A** represents a BAC which does not belong to c1 and has a small insert. This BAC was useful to recognize the bands in the fingerprint that correspond to the BAC vector (V). The faint background in all lanes corresponds to the amplification of bacterial DNA that is present in very low amounts in these BAC DNA preparations. "V" in **A** and **B** corresponds to BAC vector DNA.
**C**: Contigs of BAC clones isolated either by using *PscW62-1* as a probe (shown in black), or by chromosome walking done on the 3 contigs (shown in gray). The double arrow lines indicate the size in Kbp of the new sequence obtained by chromosome walking. Black arrowheads show the position of the contig-end subclones used for chromosome walking and the overlapping BACs isolated with these probes are drawn below these arrows. Gray arrowheads show the position of contig-end subclones that did not result in contig sequence extensions, and in c2 these end-subclones resulted in the isolation of additional BAC clones containing parts of the original sequence of this contig. However, two of these contig-end subclones from c1 and c2 (shown by gray arrow-heads) recently resulted in the isolation of a BAC clone linking these two contigs, when a new *M. truncatula* BAC library (Cook et al., unpublished) was used (see Results section). The position of the sequences that strongly or weakly hybridize with *PscW62-1* in c1 and c3, or with c2, respectively, are shown as black vertical lines above the contig BACs.
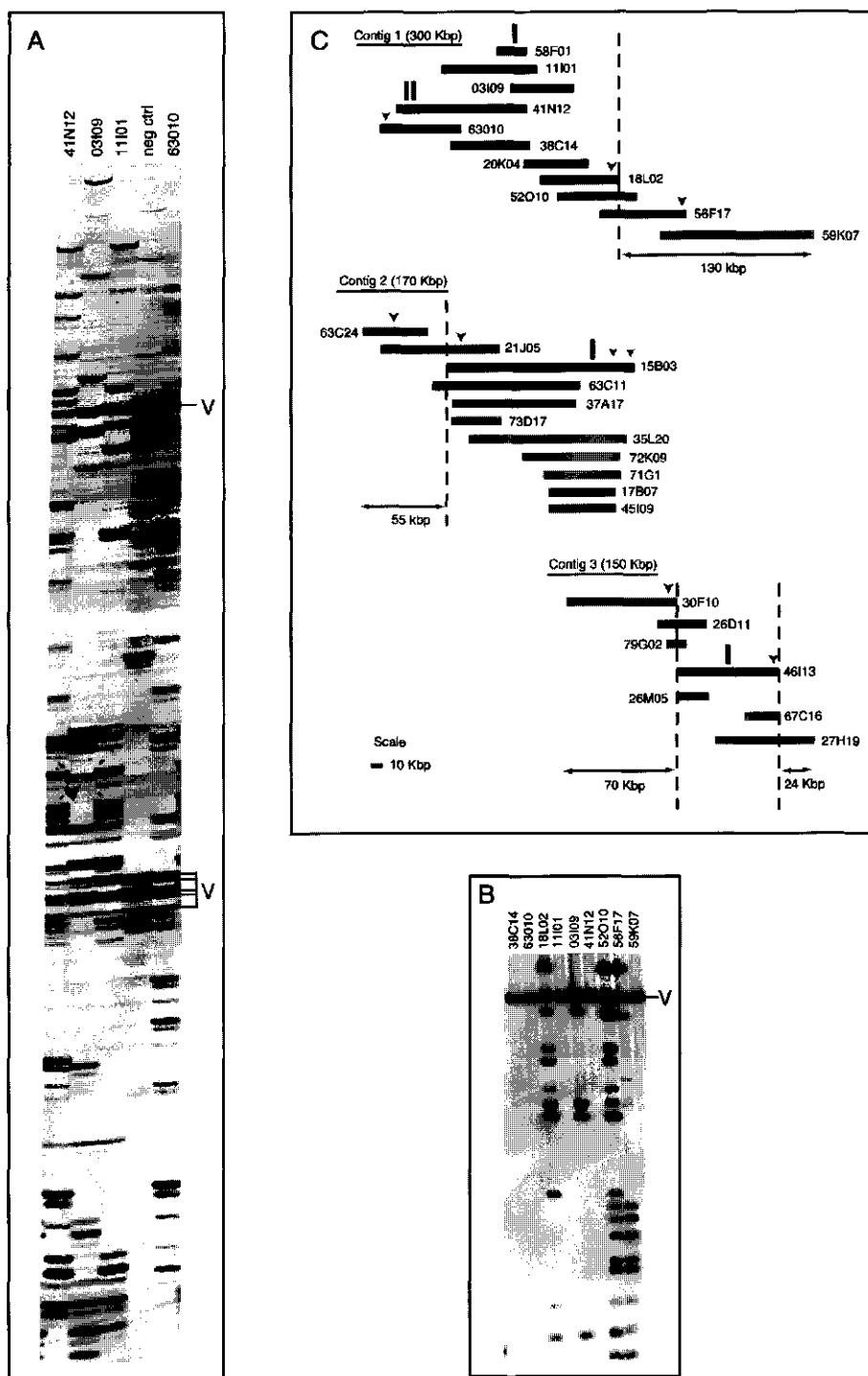
**Fig 1**

**Orientation of contigs in *Medicago* chromosome 5.**

The orientation of c1 and c2 on chromosome 5 was determined by 2-colour fluorescence in situ hybridization (FISH) on pachytene chromosomes (Fig 2). BAC clones 58F01 and 59K07 (obtained by chromosome walking) of c1, and 45I09 and 63C24 (obtained by chromosome walking) of c2, were used as probes. This showed that 58F01 of c1 and 45I09 of c2, are located close to each other, while 59K07 is the c1 clone that is closest to the centromere, and 63C24 is the clone from c2 that is closest to the telomere (Fig 2). In addition, these experiments confirmed the chromosome walking extensions of these contigs, as the FISH signals of the clones obtained by chromosome walking overlap with the signals of the BACs from the original contigs.
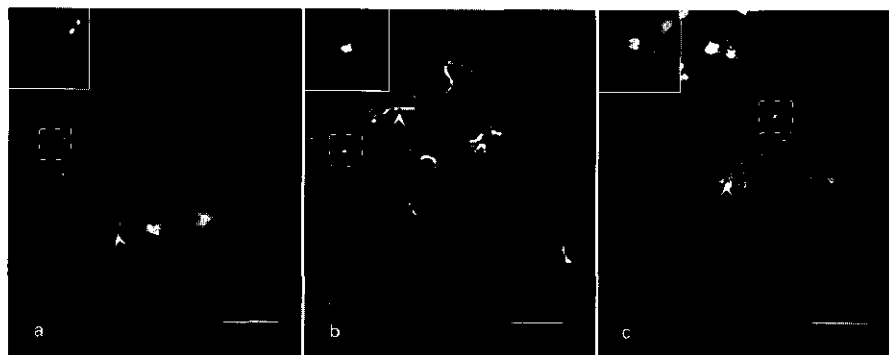


**Fig 2:**
Positioning of BAC contigs on pachytene chromosomes of *M. truncatula* A17 by two-color FISH. The DAPI stained chromosomes have bright fluorescent heterochromatin around the centromeres (pericentromeric heterochromatin). The chromosomes were hybridized with BACs of the 3 contigs. **a.** Hybridization of BAC58F01 (green) of c1 and BAC 45I09 (red) of c2. The two BACs are tightly linked and the signals partly overlap. The chromosomes were subsequently hybridized with 27H19 (red) of c3 which is located on the same chromosome arm. **b.** Hybridization with BAC59K07 (red) and BAC58F01 (green) of c1. **c.** Hybridization with BAC45I09 (green) and BAC63C24 (red) of c2. Yellow fluorescence is caused by overlap of green and red signals. The position of the centromere of chromosome 5 is indicated with an arrowhead, the telomere closest to c1/2 with an asterisk. The parts that have been magnified are indicated. Bar = 10 μm.

**Identification of the *Sym2* orthologous region in *M. truncatula*.**

We studied by RFLP mapping, whether c1/c2 or c3 represents the *Medicago* genomic region that is orthologous to the pea *Sym2* region. For this purpose, we isolated several cDNAs encoded by c1/c2 by screening a *M. truncatula* root hair cDNA library (Covitz et al., 1998) with labeled BAC inserts. The c1 cDNA clones *Mtc831*, and *Mtc411*, and the c2 cDNA clone *Mtc7x-1*, strongly hybridized with 1 to 3 genomic restriction fragments in pea and in *Medicago*, except for *Mtc923* that hybridizes with 4 genomic fragments in *Medicago* (Table 1). A comparison of HindIII restricted *Medicago* genomic DNA, c1/c2 DNA, and c3 DNA,

showed that all the genomic DNA bands hybridizing with *Mtc831*(Fig 3), *Mtc7x-1* and *Mtc411*, are present in c1/c2. In contrast, although *Mtc411* is located on c1, it weakly hybridizes with a fragment of c3. On the other hand, *Mtc923* (Fig 3) hybridizes strongly with 3 and weakly with 1 HindIII-restricted *Medicago* genomic DNA bands, out of which only 2 strongly hybridizing bands are located on c1. Therefore in all cases, except for *Mtc923*, it can be concluded that the pea genomic restriction fragments strongly hybridizing with these *Medicago* cDNAs, represent the pea sequences truly orthologous to these cDNAs located on c1/c2.
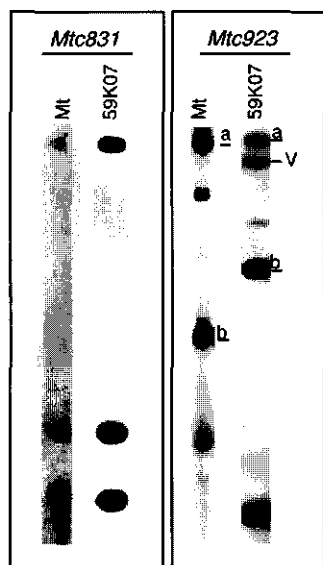


Figure 3: Hybridization of *Mtc831* and *Mtc923* with HindIII restricted *M. truncatula* genomic DNA (Mt) and BAC 59K07 DNA. *Mtc831* is entirely located in 59K07 since all the fragments observed in Mt genomic DNA are present in this BAC. On the other hand, *Mtc923* is not entirely located in 59K07: the strongly hybridizing fragments in the Mt and the 59K07 lanes named as "a" and "b" have an identical size and correspond to the same HindIII fragment (these two lanes were aligned at band "a" (9.1Kbp) but originate from independently electrophoresed gels were the Mt sample migrated a larger distance; therefore, fragment "b" has migrated a larger distance in Mt). Apart from bands "a" and "b" Mt contains a weakly (between "a" and "b") and a strongly (below "b") hybridizing fragments that are not present in 59K07, demonstrating that *Mtc923* has homologous sequences in the genome that are not comprised in 59K07. In addition to fragments "a" and "b", *Mtc923* weakly hybridizes in 59K07 with a fragment between "a" and "b" and a fragment far-below "b", which in Mt is not detected and has run out of the gel, respectively. "V" represents BAC vector DNA (7.3 Kbp) that hybridized with the vector-containing *Mtc923* probe.

**Fig 3**

These 4 cDNAs showed an RFLP between Afghanistan and Rondo/Sparkle DNA. RFLP mapping using the RILs and *Sym2* introgression lines showed that *Mtc7x-1* and *Mtc411* strongly hybridized with one fragment that is tightly linked to *Sym2* (Table 1). In contrast, two recombination events map in between *Sym2*, and *Mtc831* and *Mtc923*, respectively (Fig 4, Fig 5, Table 1). Although *Mtc923* hybridizes to *Medicago* genomic restriction fragments that are not located on c1, RFLP studies in pea show that 1 out of 2 strongly hybridizing restriction fragments (highly homologous fragments) is linked to *Sym2* (Fig 4). *Mtc831* is entirely located on BAC 59K07 (Fig 3), and *Mtc923* is probably mostly located in BAC 59K07 although 1 out of 3 *Medicago* genomic restriction fragments that strongly hybridize with this cDNA is not located on this BAC (Fig 3) or in c1/c2. Thus, RFLP mapping with all

these cDNAs shows that the studied sequences occurring in c1/c2 and the pea *Sym2* region are conserved. Moreover, the two recombinations in between *Sym2*, and *Mtc831* and *Mtc923*, have delimitated both the *Medicago Sym-2*-orthologous genomic region in BAC 59K07 (obtained by chromosome walking) of c1 (Fig 1C) and the pea *Sym2* region in the pea RILs D23.3 and D27.2 (Fig 5).

To extend microsynteny studies and to find recombinations that delimitate the *M. truncatula Sym2*-orthologous region in c2, genomic subclones were generated from c1/c2. The MtgG28, Mtg3552, Mtg3556 and Mtg63EB4 (the isolation of the last clone described in Gualtieri et al., submitted, a) genomic subclones are single copy in *Medicago* (Table 1) as shown by genomic Southern blots of DNA restricted with the same enzyme used to generate these subclones from c1/c2. However, when these clones were hybridized to blots containing restricted BAC DNA (with a sequence molarity 100 times higher than in genomic blots), some of the subclones weakly hybridized with additional fragments on c1/c2 DNA blots (Gualtieri et al., submitted, b), and Mtg3552 weakly hybridized with a fragment in c3 (Gualtieri et al., submitted b).

The four genomic subclones from c1 and c2 were used for RFLP mapping in pea using the RILs and *Sym2* introgression lines. The *Medicago* c1 subclone MtgG28 strongly hybridizes with a single polymorphic fragment in blots containing HaeIII restricted pea genomic DNA (Fig 4). This polymorphic pea fragment is linked to *Sym2* and has the same size as the *Sym2*-linked polymorphic HaeIII restriction fragment hybridizing with *PscW62-1*(in Gualtieri et al., submitted, a). The hybridization pattern of c1 digested DNA hybridized with MtgG28 and *PsW62-1* results in a similar band pattern; however, in contrast with *PscW62-1*, MtgG28 does not hybridize with c2 and c3, and it is located on a c1 HindIII restriction fragment (shown by an asterisk in Fig 6) that does not hybridize with *PscW62-1* (Fig 6) (Gualtieri et al., submitted, b). These hybridization data in pea and *Medicago*, suggest that MtgG28 and *PscW62-1*, and their orthologues, are physically very tightly linked in both species.
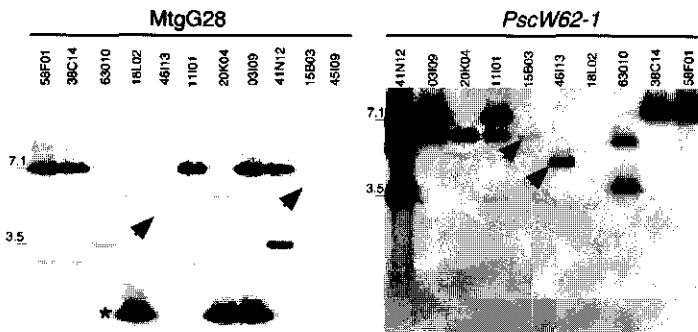
**Fig 6**

Blots of BAC DNA restricted with HindIII showing the hybridization of BACs from c1, c2 and c3 with MtgG28 and *PscW62-1*. The 7.1 and 3.5 Kbp fragments hybridize with both probes. The fragments located at the position indicated by an asterisk contains MtgG28 and do not hybridize with *PscW62-1*. *PscW62-1* hybridizes with BACs from the three contigs, while MtgG28 does not hybridize with c2 and c3. Arrowheads show the position of the fragments in BACs 46I13 and 15B03 that hybridize with *PscW62-1*, but that do not hybridize with MtgG28.

The c2 subclone Mtg3556 strongly hybridizes with 2 polymorphic fragments that are tightly linked to *Sym2* on pea genomic blots, while the c2 subclone Mtg3552 strongly hybridizes to several polymorphic fragments that are linked to *Sym2* (Fig 4). One end sequence of Mtg3552 (named MtgLRR52) revealed the presence of a LRR (Leucine Rich Repeat) region with homology to the LRR domain of the *Cf4* and *Cf9 L. esculentum* disease resistance genes (Parniske et al., 1997). Therefore, MtgLRR52 was used as a probe to screen a pea root hair cDNA library (see material and methods). A cDNA named *PscLRR52* was isolated. When genomic blots are hybridized with *PscLRR52* or Mtg3552 very similar hybridization patterns were obtained (see next section).

In contrast with the genomic subclones MtgG28, Mtg3556, and Mtg3552, that are tightly linked to *Sym2*, a recombination event in the *Sym2*-containing introgression line Sparkle-*Sym2*, maps in between *Sym2* and a pea restriction fragment hybridizing with Mtg63EB4 (Limpens, pers. comm.) (Table 1) (Fig 5). Mtg63EB4 is a genomic subclone of the c1 BAC 63C24 (obtained by chromosome walking) (Fig 1C). Thus, *Mtc831* and *Mtc923*, and Mtg63EB4, have delimitated the *Medicago* c1/c2 *Sym2*-orthologous region, at the BACs 59K07 and 63010, respectively (Fig 1C, Fig 5). The size of this *Sym2*-orthologous region is delimitated to about 350 Kbp. In pea, *Mtc831* and *Mtc923*, and Mtg63EB4, are separated from *Sym2* by recombinations that occurred in the RILs D23.3 and D27.7, and in the Sparkle-

*Sym2* containing introgression line, respectively (Fig 5). We refer the map position on pea linkage group I of *Mtc831* and *Mtc923*, and Mtg63EB4, to as the "south" and "north" of *Sym2*, respectively.

Thus, the combination of RFLP mapping data of 4 cDNAs and 4 genomic subclones from c1/c2, not only demonstrate that c1/c2 is the *Medicago Sym2*-orthologous region, but also 3 RFLP markers have delimitated this region.

**Table 1**: summary of the hybridization behaviour of the RFLP markers described in this manuscript.

| cDNA/genomic subclone (size in Kbp) | Number of bands in *M. truncatula* (HindIII digestion) | Restriction enzyme in pea RFLP analysis | Number of pea bands in RFLP analysis [1] | Number of pea bands linked to *Sym2* [1] | Author/ citation |
|---|---|---|---|---|---|
| *Mtc411* | ND | HindIII | 3 | 1 | Kulikova/ Gualtieri et al., submitted, a. |
| *Mtc7x-1* | 3 (determined by Gualtieri) | HindIII | 3 | 1 | Kulikova/ Gualtieri et al., submitted, a. |
| *Mtc831* | 3 | HindIII | 2 | 2 | Gualtieri/ Gualtieri et al., submitted, a. |
| *Mtc923* | 4 | BamHI | 3 | 2 | Gualtieri/ Gualtieri et al., submitted, a. |
| MtgG28 (0.18) | 1 | HaeIII | 1 | 1 | Gualtieri/ Gualtieri et al., submitted, a. |
| Mtg3552 (5.2) | 1 | EcoRI | 9[2] | 6 | Gualtieri/ Gualtieri et al., submitted, a. |
| Mtg3556 (5.6) | 1 | EcoRI | 2 | 2 | Gualtieri/ Gualtieri et al., submitted, a. |
| Mtg63EB4 (4.0) | 1 | BamHI | 1 | 1 | Limpens/ Gualtieri et al., submitted, a. |
| *PscLRR52* | ND | EcoRI | 7 | 5 | Gualtieri/ Gualtieri et al., submitted, a. |

[1] This total number of bands includes the unique and common "strongly" hybridizing bands from Afghanistan, Rondo, Sparkle and *nod3* lines.
[2] Includes both "strongly" and "weakly" hybridizing bands from Afghanistan, Rondo, and *nod3* lines.
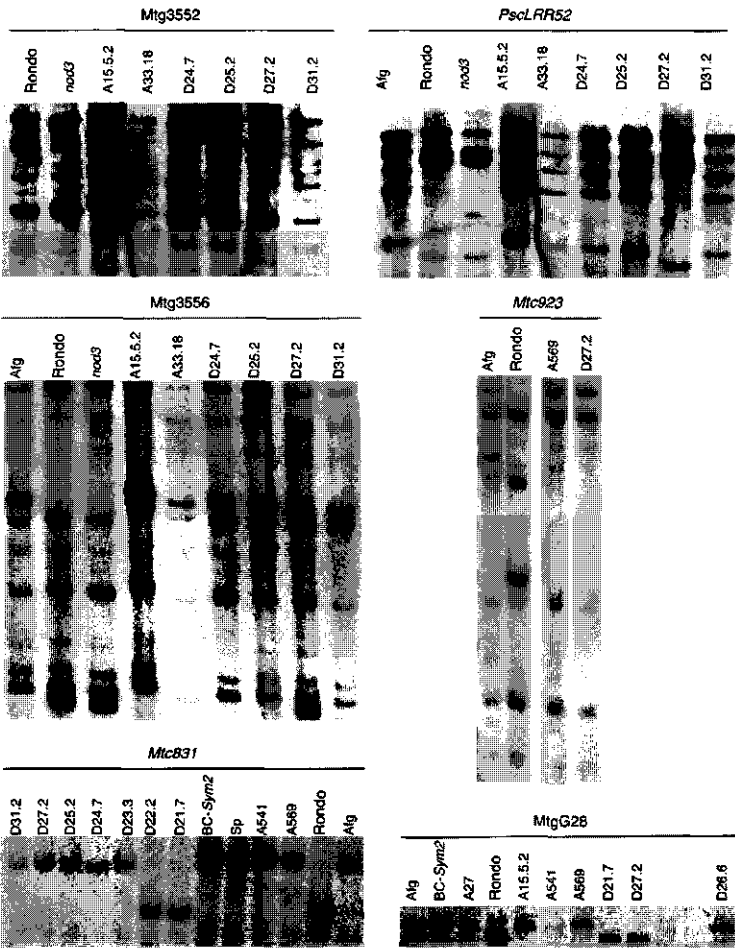
**Fig 4**

Pea genomic blots showing the different *Sym2*-linked RFLP markers. The restriction enzymes used to detect these polymorphisms are listed in Table 1. The genetic characteristics of the RILs and introgression lines used are given in Figure 5. The probes hybridized to each blot are indicated on top of the different pictures. The RFLP pattern of A15.5.2 after hybridization with Mtg3552 is identical to that of Afghanistan (data not showed).
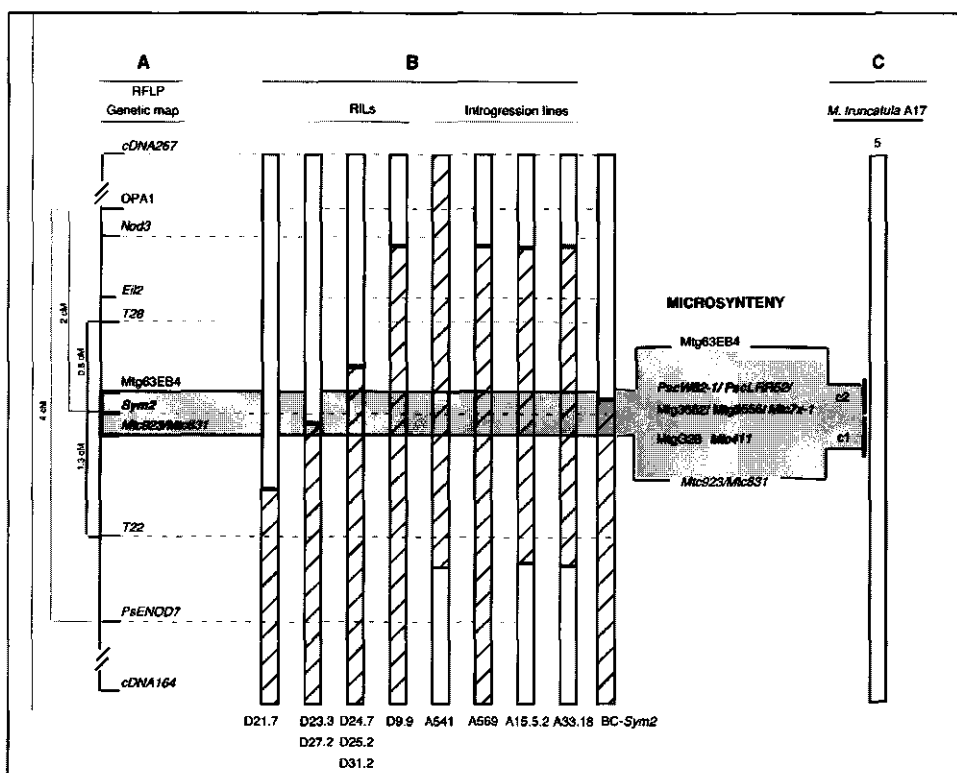
**Fig 5**

A: Genetic map of a 4 cM region spanning from OPA1 to *PsENOD7* on pea linkage group I. The genetic distance between *cDNA164* and *PsENOD7*, and OPA1 and *cDNA 267* (Kozik, 1996), is not presented in this figure. The genetic distance between markers obtained by recombination analysis within a segregating F2 population is indicated at the left side of the map (Geurts, 1998), while only the relative linear order is known for other markers mapped by using the RILs. The pea markers tightly linked to *Sym2* *PscW62-1* (Geurts, 1998) and *PscLRR52* (this thesis), and the markers isolated by using the *M. truncatula* A17 c1/c2 *Sym2*-orthologous region (this thesis), are listed (in between B and C) together with the *Sym2* locus in the horizontal shadowed area. Markers *Mtc923/Mtc831* and Mtg63EB4 were isolated from c1/c2 BACs obtained by chromosome walking and delimitated the pea *Sym2* region, since recombination events are located between the hybridizing pea sequences and the *Sym2* locus. The pea sequences hybridizing with *Mtc923/Mtc831* and Mtg63EB4 showed 2 and 1 recombinations with the *Sym2* locus, in the pea RILs D23.3 and D27.2, and in the pea cv Sparkle BC-*Sym2* introgression line, respectively. Therefore, these three markers are at the "north" and "south" border, respectively, of the horizontal shadowed area that represents the pea *Sym2*-containing region.

B: The position of the introgressed region in the RILs ("D" prefix) and introgression lines ("A" prefix) is shown by hatched/white boxes. Hatched boxes are Afghanistan DNA and white boxes are Rondo DNA. The lines that have Afghanistan DNA (hatched box) at the position where *Sym2* maps show the *Sym2* phenotype.

C: *M. truncatula* A17 chromosome 5 (white box). Small black vertical lines at the left side of chromosome 5 indicate the location and orientation of contigs c1 and c2 that define the A17 *Sym2*-orthologous region. The horizontal shadowed area in A, B and C, indicates the pea *Sym2*-containing region and the microsyntenic *Medicago* chromosome 5 c1/c2 *Sym2*-orthologous region.

**"From *M. truncatula* back to pea": isolation of orthologous LRR sequences from c1/c2, and the pea *Sym2* region.**

Since one end sequence of the genomic subclone RFLP marker Mtg3552 revealed an interesting high homology with the LRR-containing motif of *Cf4* and *Cf9 L. esculentum* resistance genes, primers were designed to specifically amplify this LRR-encoding sequence (named MtgLRR52), which was used as a probe to screen a pea root hair cDNA library. A pea cDNA was isolated and named *PscLRR52*. This cDNA encodes a protein containing a LRR motif that is also highly homologous to that of the Cf4 and Cf9 proteins. When hybridized to blots containing c1/c2 and c3 restricted DNA, this cDNA hybridized with several fragments of c2 (Gualtieri et al., submitted, b). When hybridized to genomic Southern blots, *PscLRR52* showed a similar hybridization pattern as Mtg3552 (Fig 4). Interestingly, *PscLRR52* strongly hybridizes to 7 pea restriction fragments, 5 of which map in the *Sym2* region (Fig 4, Fig 5). This clearly shows that the pea *Sym2* region is rich in *Cf4*/*Cf9* LRR motif-like sequences.

Since the pea *Sym2* region is rich in sequences highly homologous to Mtg3552, including the *Cf4*/*Cf9* LRR motif-like sequences, Mtg3552 was used to screen a pea genomic cosmid library in order to identify and isolate contigs containing these pea homologous sequences that map in the *Sym2* region. In this way, 17 cosmids were isolated. To study the presence and distribution of sequences homologous to *PscLRR52* and MtgLRR52, the cosmids were digested with HindIII and hybridized with these two clones (Fig 7). Hybridization with *PscLRR52* showed that 10 of the 17 cosmids hybridized with the probe. Further, their restriction pattern shows that they are physically unlinked except for three cosmids (52/1.2, 52/3.1 and 52/5.2) that form a contig and contain the same fragment hybridizing with the *PscLRR52* probe (Fig 7). Thus, 8 pea genomic regions are homologous to *PscLRR52*. This is comparable with the 7 bands (out of which 5 are linked to *Sym2*) observed on pea genomic blots hybridized with *PscLRR52* (Fig 4). In contrast, only 3 physically unlinked cosmids (52/15.2, 52/16.1, 52/5.1) out of these 8 genomic regions, hybridized with the *Medicago* LRR-containing sequence MtgLRR52 (Fig 7). This shows that the 4 remaining cosmids and the contig that only hybridize with *PscLRR52*, contain homology with the non-LRR encoding sequences composing this cDNA. The future subcloning and sequencing of the cosmids and contig sequences that hybridize with *PscLRR52* and MtgLRR52, will be useful to study sequence conservation and divergence within these sequence families, and to design primers for PCR-based fine mapping of individual sequences in relation to *Sym2*. This last would be

useful to narrow down the number of LRR sequences that should be functionally analyzed to test whether one of them is responsible for the *Sym2* phenotype. In addition, these numerous unlinked cosmids and contig could be useful to construct a large pea contig including the *Sym2* gene.
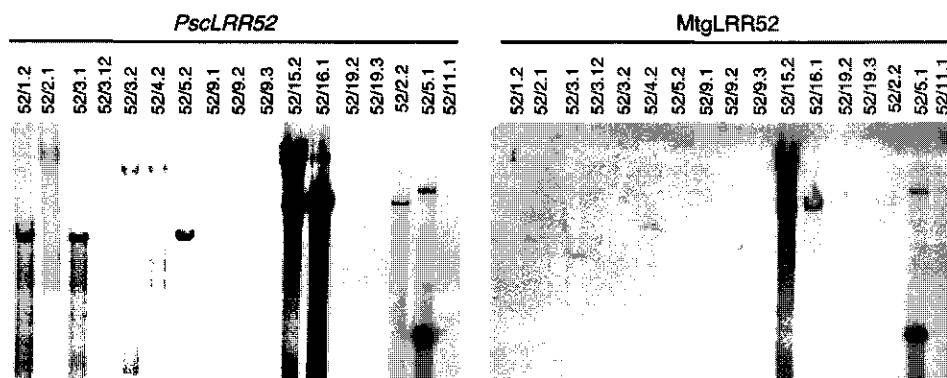


**Fig 7**

Hybridization of pea cosmids isolated by screening a pea genomic library with Mtg3552 (containing MtgLRR52). Cosmids were restricted with HindIII and hybridized with *PscLRR52* (A), or with MtgLRR52 (B). Many pea cosmids contain fragments that are homologous with *PscLRR52*, while only three cosmids contain sequences homologous with the LRR-encoding c2 sequence MtgLRR52. Cosmids 52/1.2, 52/3.1 and 52/5.2 form a contig as shown by their common band that hybridizes with *PscLRR52*.

**Isolation of pea genomic sequences orthologous to sequences in c1/c2.**

During the process of identification and subcloning of c2 single and low copy sequences to use for RFLP analysis, the *Medicago* sequence Mtg2114 (also described in Gualtieri et al., submitted, b) was isolated from BAC 21J05 (obtained by chromosome walking). However, this subclone did not result in a RFLP between Afghanistan, Rondo and Sparkle with several tested restriction enzymes. Therefore, Mtg2114 was used to screen a pea genomic cosmid library, in order to isolate orthologous pea sequences that could be used to design primers for PCR-based fine mapping in pea.

On pea Southern blots, Mtg2114 hybridizes with 2 non-polymorphic HaeIII fragments, the smaller of them (strongly hybridizing) being the true pea orthologue of this *Medicago* c2 subclone (Fig 8A). The genomic cosmid library screenings using Mtg2114 resulted in the isolation of 6 genomic clones that formed 2 physically unlinked contigs as shown by HindIII restriction, blotting and hybridization with this subclone (Fig 8B) and with the cosmid 2114/13.2 (Fig 8C). The strong hybridization of both contigs with Mtg2114 demonstrates that they contain the pea true orthologous sequence identified by Southern blot (Fig 8A and 8B).

Thus, the future subcloning and sequencing of the Mtg2114-pea orthologous sequences present in these two cosmid contigs will be useful to design PCR-based markers for fine mapping using a large pea segregating population. In addition, both Mtg2114-orthologous pea contigs could be used during the construction of a pea physical region containing the *Sym2* gene.
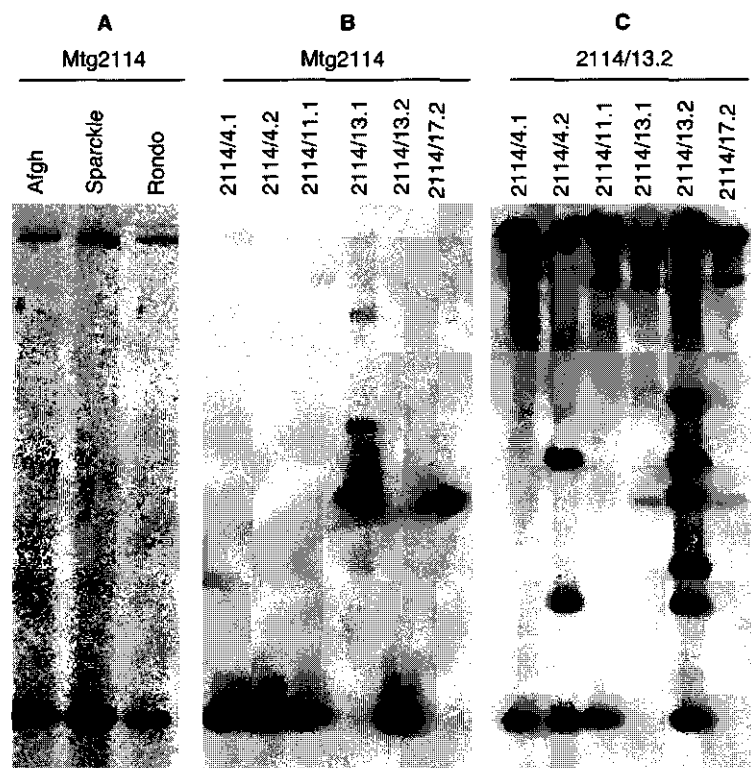


**Fig 8**

A: Southern blot containing pea genomic DNA restricted with HaeIII and hybridized with the c2 subclone Mtg2114. "Afgh" means Afghanistan.
B: Southern blot of pea cosmids restricted with HindIII and hybridized with Mtg2114.
C: Southern blot of pea cosmids restricted with HindIII and hybridized with the labeled cosmid 2114/13.2.
The common fragments present in cosmids in both B and C, indicate that they form two physically unlinked contigs composed by cosmids 2114/4.1, 2114/4.2, 2114/11.1, 2114/13.2; and 2114/13.1 and 2114/17.2.

## DISCUSSION

This study reveals a high conservation of the gene and genomic sequence composition between the *Medicago* c1/c2 and the pea *Sym2* genomic region. The combination of RFLP mapping data of 4 cDNAs and 4 genomic subclones from c1/c2, not only demonstrate that

c1/c2 is the *Medicago Sym2*-orthologous region, but 3 RFLP markers have delimitated this region to a size of about 350 Kbp.

Although the order of RFLP markers was determined in c1/c2 (Gualtieri et al., submitted, b), the order of these markers was not analyzed in the pea *Sym2* region. However, the order of three "clusters" of pea sequences orthologous to Mtg63EB4; MtgG28/Mtg3556/Mtg3552/*Mtc7x1*/*Mtc411*; and *Mtc831*/*Mtc923*, respectively, is the same as in *M. truncatula*. Altogether, both sequence composition and sequence cluster order reveal a high level of microsynteny between the pea *Sym2*-containing region and the c1/c2 *Medicago* region located on chromosome 5.

In general, this work shows that it is possible to isolate and delimitate a *M. truncatula* region that might contain a gene orthologous to a pea gene of interest for cloning. We show here that the c1/c2 *Medicago* orthologous region allowed the isolation of 8 RFLP markers around the pea *Sym2* gene. Moreover, the cloning of *PscLRR52* by using MtgLRR52, confirms that RFLP markers obtained from *Medicago* orthologous regions are useful to clone the pea orthologous gene of interest. This demonstrates the use of *M. truncatula* as intergenomic cloning vehicle of pea genes located within microsyntenic regions. However, even in highly microsyntenic species like grasses, although about 90% of the genes are conserved, gene order is somewhat more variable (Bennetzen, 2000). Hence although the probability that a gene of interest, like *Sym2*, is present in an orthologous region could be high, it remains to be demonstrated that this is indeed the case. Nevertheless, the strategy presented here is useful to saturate the area around the gene of interest with molecular (RFLP) markers that can be used to isolate a pea contig containing this gene. However, the success of the latter, depends on the local microsynteny between *Medicago* and pea, which should be studied in each individual target locus. For the cloning of the pea *Sym2* gene, we demonstrated that the isolation and delimitation of the c1/c2 *Medicago* orthologous region provides a good and promising start for the microsynteny-based positional cloning of this gene.

## MATERIAL AND METHODS

### BAC library screenings.

The BAC library of *M. truncatula* genotype A17 was screened according to Nam *et al.* (1999), and high density BAC library filters and BAC clones were obtained from the Clemson

University Genomics Institute (http//www.genome.clemson.edu). DNA probes for cDNA and BAC library screenings were labeled as described below.

## BAC DNA isolation and construction of contigs.

Bacterial Artificial Chromosome (BAC) clones of *M. truncatula* genotype A17 belonging to c1/2 and c3 (Gualtieri et al., submitted, a), were grown and BAC DNA was isolated according to Nam *et al.* (1999). Contigs were constructed by a combination of BAC DNA fingerprinting, and restriction/hybridization analysis. These studies resulted in arrays of overlapping BAC clones that formed contigs, and defined different contig regions within each contig. The sequences (markers) studied in this manuscript were assigned to different contig regions by hybridization (for detailed contig maps see Gualtieri et al., submitted, b).

## Genetic mapping in *M. truncatula*.

The mapping of the BAC-end sequences in *M. truncatula* was performed as described in Gualtieri *et al.* (submitted, a). Briefly, CAPS (Cleaved Amplified Polymorphic Sequences) or amplified length polymorphism markers, were mapped on a F2 population of 93 plants from a cross between genotype A17 x A20. Polymorphic DNAs were resolved on a 1.5% agarose gel and visualized by ethidium bromide staining.

## Chromosome Walking.

Chromosome walking on the *M. truncatula* A17 BAC library was done both by hybridization and PCR screenings. For the hybridization screenings, high density filter arrays obtained from the Clemson University Genomics Institute (http://www.genome.clemson.edu), were hybridized with end subclones obtained form contig-end BACs identified after construction of BAC contigs. The PCR-screenings were done on a multiplex DNA copy of the *M. truncatula* A17 BAC library as described by Nam *et al.* (1999). PCR primers were designed from BAC-end sequences.

## FISH mapping in *M. truncatula*.

Fluorescent *in situ* hybridization (FISH) was performed as described in detail by Zhong *et al.* (1996), and Kulikova *et al.* (in press).

**cDNA and cosmid library screenings.**

The pea genomic library was contructed in the Lambda ZAPII vector system, by using DNA from the pea cultivar Alaska. The pea root hair cDNA library was constructed by Stratagene in Lambda ZAPII vector system, using equal amounts of poly (A)+ RNA isolated from a mixture of root hairs of 6 days old cv Finale pea plants, uninoculated and inoculated for 48 hours with *R. leguminosarum* bv *viciae* 248 (Josey et al., 1979). The *Medicago truncatula* A17 root hair cDNA library was kindly provided by S. Long (Covitz et al., 1998). In order to (partially) remove repetitive DNA from the probe, the screening of the *M. truncatula* root hair cDNA library was done with 50 ng of BAC inserts that were divided in four pools of 12.5 ng and independently digested with 4-cutter restriction enzymes (HaeIII, RsaI, AluI and Sau3AI). Pools were mixed and hybridized for two rounds to a Cot100 on a 0.5 cm membrane disk (Hybond-N+) saturated with denatured *M. truncatula* genomic DNA (previously restricted independently with HaeIII and RsaI). The cDNA library screenings were done with Hybond-N+ membranes using the conditions recommended by Stratagene.

**Genomic DNA and BAC DNA blots.**

Pea and *Medicago* DNA was isolated according to A. Kozik (Kozik, 1996) and a similar method using 2 times more B-mercaptoethanol in the DNA CTAB-containing extraction buffer (Cook et al., pers comm), respectively. In the *Medicago* DNA isolation method, chloroform was used instead of phenol in order to extract proteins. Bacterial Artificial Chromosome (BAC) DNA of *M. truncatula* A17 was isolated according to Nam *et al.* (1999). All hybridizations were done with Hybond-N+ membranes (Amersham) under standard conditions (Sambrook et al., 1989). DNA probes for Southern blots and northern blots were labeled with [32P]dATP using the random priming method (Feinberg and Fogelstein, 1983).

**Plant material used for RFLP mapping in pea.**

A set of *Sym2*-containing introgression lines and RILs was used to determine by RFLP mapping whether the c1/c2 genomic subclones and cDNAs, and *PscLRR52*, map in the vicinity of *Sym2*. The Sparkle-*Sym2* introgression line used was BC-Sym2, derived from an F8 of the cross Afghanistan x Sparkle followed by 7 back crosses (Kneen at al., 1984; Temnykh et al., 1995). The introgression lines used were the Rondo-*Sym2* intogression lines A5.4.3 and A5.6.9 (both Kozik et al., 1995), and A15.5 and A33.18 (Kozik, 1996). The last two lines were made by crossing A5.4.3 with Rondo and selection among the F2 for a cross-over event between *Sym2* and the RAPD marker OPA-1 (Kozik, 1996). The RIL population

was constructed by R. Geurts (Geurts, 1998; in Gualtieri et al., submitted, a) as follows: The Rondo-*Sym2* introgression line A54 (Kozik et al., 1995) was crossed with the Rondo hypernodulating EMS mutant *nod3* (Jacobsen et al., 1984). A54 was generated by crossing Rondo with Afghanistan and subsequently using Rondo as the recurrent parent for 4 back crosses, and selecting for the *Sym2* phenotype. A54 contains an introgressed region spanning from marker cDNA267 to cDNA164, therefore including the Afghanistan *Sym2* gene. The *nod3* mutant was chosen for constructing a RIL population since it was previously shown that the *Nod3* locus is, like *Sym2*, located in linkage group I of pea (Temnykh et al., 1995a, 1995b). From the A54 x *nod3* F2 population of 882 plants, 190 hypernodulated individuals were selected upon inoculation with *R.l.* bv. *viciae* strain 248.pMW1071(*nodX*), a strain that nodulates both plants with and without the *Sym2* gene (Kozik et al., 1995). Marker *ENOD7*, that is located to the south of the *Sym2* locus, was analyzed and 14 out of these 190 plants contained the Afghanistan *ENOD7* allele. These plants were selfed and F3 plants homozigous for the Afghanistan *ENOD7* allele were inoculated with *R. leguminosarum* bv *viciae* 248, a strain that does not nodulate plants homozygous for the Afghanistan *Sym2* gene (Geurts et al., 1997). Six lines were hypernodulated (numbers: D4.3, 21.7, D23.3, D27.2, and D30.1), whereas the others (numbers: D9.9, D9.11, D21.9, D24.7, D25.5, D26.6, D31.2 and D53.8) formed only a few nodules (0-10 nodules). The last lines are homozygous for *nod3* as well as for *Sym2*. These lines were crossed with the *Sym2*-containing introgression line A569 (Kozik et al., 1995) and the nodulation phenotype of the F1 plants confirmed that the lines are homozygous for *Sym2*.

# CHAPTER 3

**Microsynteny between the *Medicago truncatula Sym2*-orthologous genomic region and another genomic region both located on the long arm of chromosome 5**

Gustavo Gualtieri[1]*, Douglas Cook[2], and Ton Bisseling[1].

1) Molecular Biology, Agricultural University Wageningen. Dreijenlaan 3, 6703HA Wageningen, The Netherlands.
2) Department of Plant Pathology, University of California at Davis. One Shields Avenue. Davis, CA 95616-8680. U.S.A.

*author for correspondence

## ABSTRACT

A synteny based positional cloning approach was started to clone the pea *Sym2* gene by using the model plant *Medicago truncatula*, as an intergenomic cloning vehicle. We reported that a marker tightly linked to *Sym2* was used to screen a *Medicago truncatula* BAC library, and three contigs were isolated and extended by chromosome walking. RFLP mapping showed that contig c1 (300 Kbp) and c2 (170 Kbp) are highly microsyntenic to the pea *Sym2* genomic region; however, some RFLP markers showed hybridization with the third contig c3 (150 Kbp). In this manuscript we analyze the distribution of 22 sequences among the three contigs, including single and low copy genomic sequences and cDNAs from c1/c2 and all the *Medicago* and the pea RFLP markers that are linked to *Sym2*. The three contigs contain (receptor) kinase coding sequences and c1 contains three distinct separate regions encoding them. c2 has two main classes of LRR-containing resistance gene-like coding sequences: one of them is distributed along three separate regions of this contig and the other is a single copy sequence unique in the genome. From all the studied sequences only 4 detected homologous sequences in c3 and these sequences are organized in clusters which linear order is comparable between c1/2 and c3, indicating that these two *Medicago* genomic regions could arise through a chromosomal duplication. Implications of these findings for the cloning of *Sym2* are discussed.

## INTRODUCTION

We have reported on the isolation of two tightly linked *M. truncatula* BAC contigs, named c1 and c2, which are highly microsyntenic with the pea *Sym2* genomic region as shown by 8 RFLP markers isolated from these contigs (Gualtieri et al, submitted, a). Therefore c1/c2 was concluded to be the *M. truncatula Sym2*-orthologous genomic region. These two contigs were isolated by using a pea RFLP marker that is tightly linked to *Sym2, PscW62-1* (Geurts, 1998). In addition to these two contigs a third contig was isolated, namely c3.

We also showed, both genetically and by FISH, that c1 and c2 were located on the long arm of chromosome 5 of *M. truncatula* and separated by a small gap of about 10 Kbp. c3 was also mapped on the long arm of chromosome 5 at a genetic distance of 9 cM to c1/2 (Gualtieri et al., submitted, a).

Contig c1 contains the RFLP markers *Mtc923*, *Mtc831*, *Mtc411* and MtgG28; while c2 contains the RFLP markers Mtg3556, *Mtc7x1*, Mtg3552 and Mtg63EB4. RFLP mapping demonstrated that pea sequences that are orthologous to these *Medicago* markers are tightly linked to *Sym2*, except for the *Mtc923* and *Mtc831* pea orthologues that are separated from *Sym2* by two recombinations located to the "south", and the Mtg63EB4 pea orthologue that is separated from *Sym2* by one recombination located at the "north". Therefore, markers *Mtc831*, *Mtc923* and Mtg63EB4, isolated from BACs obtained by chromosome walking, have delimitated the c1/c2 *M. truncatula Sym2*-orthologous genomic region to about 350 Kbp (Gualtieri et al., submitted, a). Marker Mtg3552 was used for the isolation of the pea cDNA clone *PscLRR52* that on pea EcoRI genomic blots hybridized to 7 fragments out of which 5 are polymorphic and tightly linked to *Sym2* (Gualtieri et al., submitted, a). This demonstrated that the isolation of pea markers located at the *Sym2* region is possible by using *M. truncatula* as an intergenomic vehicle.

We reported that 6 of the RFLP markers only hybridize with sequences located within c1/c2. However, Mtg3552 and *Mtc411*, weakly hybridize to single restriction fragments in c3. In addition, *PscW62-1* strongly hybridizes to both c1 and c3 and weakly to c2. Thus, it would be possible that c1/c2 and c3 are duplicated genomic regions. Saturated genetic maps of *Arabidopsis* and rice indicated that whole chromosome arms and small segments are duplicated (Kishimoto et al., 1994; Nagamura et al., 1995). Further, the *Arabidopsis* genome

sequencing shows that 60% of the sequences within the genome have been duplicated (Blanc et al., 2000; "The *Arabidopsis* Genome Initiative", 2000). Several other gene and chromosomal segment duplications were described in grasses and the family Brassicaceae (Ahn et al., 1993; Chen et al., 1997; Conner et al., 1998; Foote et al., 1997; Grant et al., 2000; Helentjaris et al., 1998; Ku et al, 2000; Langercrantz, 1998). At the gene level, *Arabidopsis* disease resistance genes encoding for LRR (Leucine Rich Repeat)-containing proteins arose by numerous duplications of LRR-encoding segments as well as intragenic unequal recombination events (Noel et al., 1999).

In this manuscript, we analyzed the possibility that c1/c2 and c3 are duplicated regions by studying the sequence composition and the order of conserved sequences within these contigs. We constructed molecular maps of the contigs showing the position of 22 single and low copy genomic sequences and cDNAs from c1/c2, including all the *Medicago* RFLP markers. In addition, we studied whether all these sequences and *PscW62-1* and *PscLRR52* hybridize with homologous sequences present in c1/c2 and c3, and we studied the distribution of these homologues in the contigs.

## RESULTS AND DISCUSSION

### Distribution of studied sequences and their homologues in c1/c2 and c3

The overlapping BACs of the three contigs are shown in Fig 1 A,B. In Fig 1B the BACs isolated with *PscW62-1* are indicated by a black line, whereas the BACs obtained by PCR- or hybridization-based chromosome walking (BACs 52010, 56F17, and 59K07, for c1; BACs 21J05, 63C24, 63C11, 37A17, 73D17, 35L20, 72K09, 71G01, and 17B07, for c2; and BACs 26D11, 30F10, 27H19, 67C16, 26M05, and 72G02, for c3) are indicated by gray lines. Only some of the BACs (52010, 56F17, and 59K07, for c1; 63C24 and 21J05, for c2; and 30F10, 26D11, 79G02, and 27H19, for c3) isolated by chromosome walking resulted in chromosome walking extensions of the contig sequence that was initially isolated by screening the BAC library with *PscW62-1* (represented by black lines in Fig 1A), and this extended sequence is represented by gray lines in Fig 1A. The position of 22 sequences was determined in contigs c1/c2 and c3 (Fig 1A) and their characteristics are given in Table 1.

Several cDNA clones were previously isolated from a *Medicago* and a pea root hair library (Gualtieri et al., submitted a). Other cDNAs are described in this manuscript. The cDNA

clones have a prefix *Mtc* or *Psc*, for *Medicago truncatula* and *Pisum sativum*, respectively. It has been demonstrated that the genes corresponding with these cDNA clones are located in c1/c2 by comparing their hybridization to *Medicago* genomic DNA and contig DNA blots. In addition to these cDNAs, several genomic subclones were made of c1/c2 and these obtained the prefix Mtg.

Fig 1A represents a HindIII restriction map of the contigs in which the positions of several HindIII restriction sites have been indicated. These restriction sites divide the contigs in so-called "contig regions" which have been numbered (bold numbers in Fig 1A). Most contig regions contain several HindIII fragments. However, the order of these fragments within the region has not been determined. The HindIII fragments within a contig region that hybridized with any of the 22 sequences studied in this manuscript are indicated as a bar above the contig in Fig 1A.

Fig1A

Fig1B

Contig 1 (300 Kbp)

Contig 2 150 (Kbp)

Contig 3 (170 kbp)
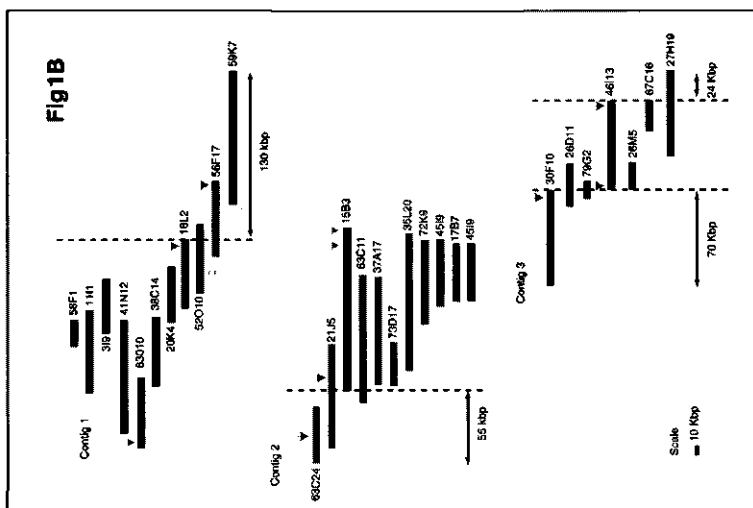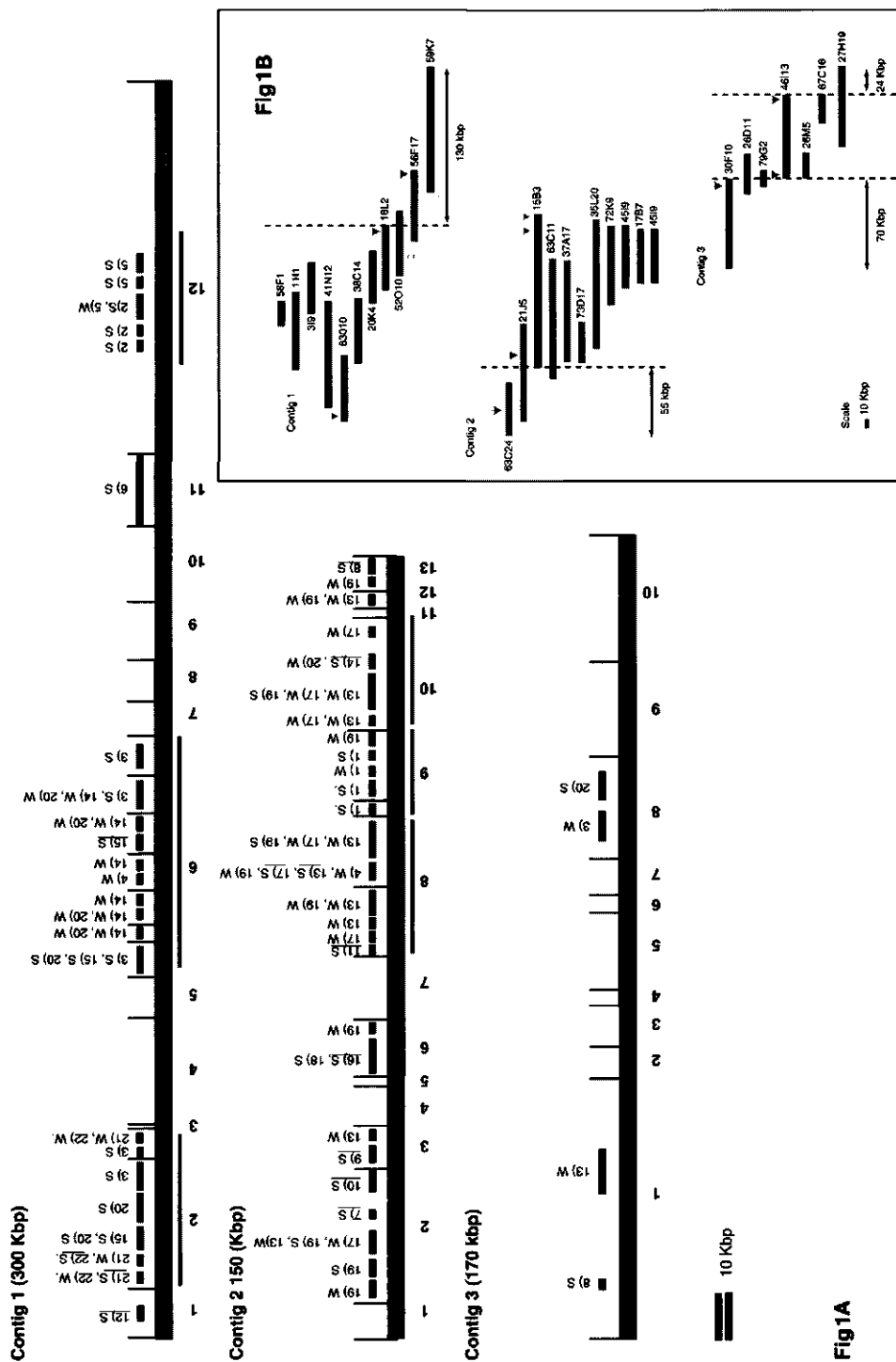
**A:** HindIII restriction maps of c1, c2 and c3. The thick horizontal black/gray line represents a contig. The black thick line shows the original contig length, and the gray thick line is the sequence gained by chromosome walking. The vertical thin lines are HindIII restriction sites (HindIII sites that have not been mapped are not indicated). The contig segment located in between two HindIII restriction sites is defined as a "contig region", and is denoted by bold numbers under each contig. In order to simplify the discussion in the text, some contiguous contig regions, are grouped and referred to as a single contig region indicated by a thin line under the contigs and denoted by a common bold number. In general, a contig region is composed of several HindIII fragments, but in some cases it contains a single HindIII fragment. Note that these contig maps show the linear order of regions within each contig, but the linear order of HindIII fragments within each region has not been determined. The HindIII fragments that hybridized with the 22 sequences studied in this manuscript are indicated as a horizontal black line located above their respective contig region, and the numbers above these lines correspond with the sequences (genomic, cDNAs) that hybridize with it (see Table 1). "S" and "W" following the clone numbers indicate a strong or weak hybridization signal, respectively. Underlined clone numbers indicate that the HindIII restriction fragment is the source of the corresponding genomic contig subclone (underlining is not used for cDNAs). Although a 10 Kbp scale bar is given, due to drawing constraints, the size of the HindIII fragments hybridizing with the 22 sequences studied here is (rather) out of scale.

**B:** Contigs of BAC clones isolated either by using *PscW62-1* as a probe (shown in black), or by chromosome walking extensions of the 3 contigs (shown in gray). The double arrow lines indicate the size in Kbp of the original contig sequence and the new sequence obtained by chromosome walking. Black arrowheads show the position of the contig-end subclones used for chromosome walking, and the overlapping BACs isolated with these probes. Gray arrowheads show the position of contig-end subclones, which did not result in a chromosome walking extension. However, one of these subclones from c1 and one from c2 have recently been used to screen a new *M. truncatula* BAC library (Cook et al., unpublished) and resulted in the isolation of a BAC clone that linked these two contigs.

## Table 1: Properties of the 22 single and low copy studied sequences

The first column shows the numbers given to the clones in Fig 1, and the second column gives the identity of these clones (prefixes *Mtc* and *Psc* indicate *M. truncatula* or *P. sativum* cDNAs, while Mtg means *M. truncatula* genomic subclone). All the M. truncatula cDNAs were isolated from a root hair library kindly provided by Sharon Long (Covitz et al., 1998). The third and fourth column indicates which clones hybridize with other contigs apart from the contig where they are located. The fifth column gives the sequence homologies as determined by BLAST searches. ND: not determined

| Clone number in Fig 1 | Contig subclone (size Kbp)/ cDNA | Location | Hybridization with contig/s | Author/ citation | Blast homologies |
|---|---|---|---|---|---|
| 1) | *Mtc7x1* | c2 | c2 | Kulikova O./ Gualtieri et al., submitted, a. | Putative membrane spanning protein [*A. thaliana*] |
| 2) | *Mtc156* | c1 | c1 | Gualtieri G./ This manuscript. | ESTs from *M. truncatula* [5] Homology with receptor kinases/kinases |
| 3) | *Mtc411* | c1 | c1/c3 | Kulikova O./ Gualtieri et al., submitted, a. | ND |
| 4) | *Mtc424* [1] | ? | c1/c2 | Gualtieri G./ This manuscript. | ESTs from *M. truncatula* [5] Homology with non- cyanogenic and cyanogenic beta glucosidase precursors, and beta glucosidases |
| 5) | *Mtc831* | c1 | c1 | Gualtieri G./ Gualtieri et al., submitted, a. | ESTs from *M. truncatula* [5] Homology with receptor kinases/kinases |
| 6) | *Mtc923* | c1 | c1/ two additional genomic locus | Gualtieri G./ Gualtieri et al., submitted, a. | ESTs from *M. truncatula* [5], and several other organisms. Phosphatase-like protein [*A. thaliana*]; Putative membrane trafficking factor [*A. thaliana*]/ Similarity to rat P47 protein [*Arabidopsis thaliana*] |
| 7) | Mtg2114 (1.4 Kbp) | c2 | c2 | Gualtieri G./ This manuscript | No homology with ESTs homology with putative protein [*A. thaliana*]. |
| 8) | Mtg2.4 (2.4 Kbp) | c2 | c2/c3 | Gualtieri G./ This manuscript | ND |
| 9) | Mtg2128 (2.8 Kbp) | c2 | c2 | Gualtieri G./ This manuscript | Low homology with *M. truncatula* ESTs [5] Putative non-LTR retroelement reverse transcriptase[*Arabidopsis thaliana*] |
| 10) | Mtg63EB4(4 Kbp) | c2 | c2 | Limpens E./ Gualtieri et al., submitted, a. | Toosled [*A. thaliana*] |
| 11) | Mtg1957 (2 Kbp) [2] | c2 | c2 | Gualtieri G./ This manuscript | No homology to ESTs putative disease resistance protein TMV N-like [*Arabidopsis thaliana*]; disease resistance protein RPS4 [*Arabidopsis thaliana*]; disease resistance RPP5 like protein [*Arabidopsis thaliana*] |
| 12) | Mtg2511 (2.5 Kbp) | c1 | c1/other genomic loci | Gualtieri G./ This manuscript | ND |
| 13) | Mtg3552 (5.2 Kbp) | c2 | c2/c3 | Gualtieri G./ Gualtieri et al., submitted,a. | (end sequence) No homology to ESTs LRR motif of *Cf-4* [*Lycopersicon hirsutum*]; putative disease resistance protein [*A. thaliana*]. |

| 14) | Mtg3556 (5.6 Kbp) | c2 | c1/c2 | Gualtieri G./ Gualtieri et al., submitted, a. | Homology *to M. truncatula* [5] *and Glycine max* (protein kinase isolog) ESTs. Similarity to receptor-like protein kinase [*Arabidopsis thaliana*]; putative protein *kinase* [*Arabidopsis thaliana*]. |
| 15) | MtgG28 (0.18 Kbp) [3] | c1 | c1 | Gualtieri G./ Gualtieri et al., submitted, a. | Homology with *M. truncatula* ESTs [5] Probable Transport ATP-binding protein MSBA, *E. coli* |
| 16) | MtgHC2.8 (2.8 Kbp) | c2 | c2 | Gualtieri G./ This manuscript. | ND |
| 17) | MtgLRR52 (0.43 Kbp) | c2 | c2 | Gualtieri G./ Gualtieri et al., submitted, a. | No homology to ESTs Cf-4 and Cf-9- like disease resistance proteins [*Lycopersicon*] (including LRR motif) |
| 18) | MtgX3.2 (3.2 Kbp) | c2 | c2 | Gualtieri G./ This manuscript. | ND |
| 19) | *PscLRR52* [4] | tightly linked to pea sym2 | c2 | Gualtieri G./ Gualtieri et al., submitted, a. | No homology to ESTs Cf-4 and Cf-9- like disease resistance proteins [*Lycopersicon*] (including LRR motif) |
| 20) | *PscW62-1* | tightly linked to pea sym2 | c1/c2/c3 | Geurts R./ Geurts R., 1998; Gualtieri et al., submitted, a. | Receptor-like protein kinases |
| 21) | MtgG5 | c2 | c2 | Gualtieri G./ This manuscript | No homology to ESTs Polyproteins from retrotransposons, reverse transcriptase, polyproteins [*Sorgum bicolor*] [*Oryza sativa*] |
| 22) | MtgG6 | c2 | c2 | Gualtieri G./ This manuscript | No homology to Ests Putative sulfite reductase [*E. coli*] |

[1] the weak hybridization of *Mtc424* with a single fragment in c1 and c2 suggests that this cDNA is probably not encoded in any of these two contigs.

[2] no cDNA was found hybridizing with this sequence in the *M. truncatula* root hair library.

[3] a single transcript hybridized with this clone on northern blots done with 4 days old uninnoculated root total RNA.

[4] cDNA *PscLRR52* was isolated by screening the pea root hair cDNA library with MtgLRR52.

[5]This sequences had homology with cDNAs from the following *Medicago* EST libraries:

Expressed sequence tags from a root hair-enriched *Medicago truncatula* cDNA library. Covitz,P.A., Smith,L.S. and Long,S.R. 1997, Unpublished.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation *Medicago truncatula* drought library Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. 2000, Unpublished.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation *Medicago truncatula* root library. Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D., Paiva,N.L. 2000, Unpublished.

ESTs from phosphate-starved roots of *Medicago truncatula*. Harrison,M.J., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E., Cho,J., Fraser,C.M. 2000, Unpublished

ESTs from roots of *Medicago truncatula* 24 hours after inoculation with *Sinorhizobium meliloti*. VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E., Fraser,C.M. 1999, Unpublished

ESTs from roots of *Medicago truncatula* after colonization with *Glomus versiforme*. Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E., Fraser,C.M. 2000, Unpublished.

*Medicago truncatula* ESTs from endomycorrhizal roots. Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V., Gamas,P. 2000, Unpublished.

*Medicago truncatula* ESTs from nitrogen-starved roots. Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V., Gamas,P. 2000, Unpublished.

The majority of the sequences only hybridize with the contig fragments from which they originate. The c1 *Mtc831*, *Mtc923*, *Mtc156*, Mtg2511, and the c2 Mtg63EB4, Mtg2114, Mtg2128, Mtg1957 (Fig 2F), MtgHC2.8, MtgX3.2, are located on a single contig region (Fig 1A). The c2 *Mtc7x*-1, MtgG5 and MtgG6, are located on two adjacent regions of this contig (Fig 1A). MtgG28, is a subclone of c1 (region 6) that hybridized with two additional fragments within the same contig and these fragments also hybridizes with *PscW62-1* (Fig 1A). The 5.2 Kbp c2 genomic subclone Mtg3552 (Gualtieri et al., submitted, a) contains an end sequence, MtgLRR52 (0.43 Kbp), which has homology to the LRR domain of the *Cf4* and *Cf9 L. esculentum* resistance genes (Parniske et al., 1997). MtgLRR52 was used to isolate the pea cDNA *PscLRR52* (Gualtieri et al, submitted, a) (also containing a sequence homologous to the LRR domain of *Cf4* and *Cf9*) from a pea root hair cDNA library. Mtg3552 (Fig 2A), MtgLRR52 (Fig 2B) and *PscLRR52* (Fig 2C) hybridize with several fragments in c2 (see next section). A comparison of *Medicago* genomic and contig DNA blots shows that Mtg2511 (c1 end subclone) and *Mtc923* (from c1), hybridize with three and two genomic fragments, respectively, in addition to the fragments present in c1. However, these additional fragments are not present in c2 and c3 but somewhere else in the genome.

**Fig 2:**
BAC clones with minimal overlap and representing the complete sequence from the 3 contigs, were digested with HindIII. A size marker in Kbp units is included at the left side. The band of 7.3 Kbp that is observed in all the samples of some blots (2A, 2C, 2F), corresponds to linear BAC vector DNA and is visible in cases when the probe contained some vector DNA. Lanes 1-6 corresponds to c1 DNA, 7-8 corresponds to c2 DNA, 9-11 corresponds to c3 DNA. The probes used were Mtg3552 in 2A, MtgLRR52 in 2B, *PscLRR52* in 2C, *PscW62-1* in 2D, Mtg3556 in 2E, and Mtg1957 in 2F.
Black arrows in 2A, 2B, 2E, and 2F indicate the fragment from which the probe originates. In 2D the white arrowheads show the sequences in c1 most homologous to *PscW62-1* (note that in lane 1 the hybridization of these two bands is weak due to a lower concentration of DNA in this lane), and in c3. The black arrow in 2D points to the sequence Mtg3556 that shows a weak hybridization with *PscW62-1*. In 2E the white arrowheads show that Mtg3556 does not hybridize with most homologous sequences to *PscW62-1* (indicated by a white arrowhead in 2D). The asterisks in 2E show the fragments that hybridize with Mtg3556 but not with *PscW62-1*. The BAC clones analyzed are 63010 (lane 1), 11101 (lane 2), 20K04 (lane 3), 52010 (lane 4), 56F17 (lane 5), 59K07 (lane 6) for c1; 21J05 (lane 7), 15B03 (lane 8) for c2; and 27H19 (lane 9), 30F10 (lane 10) and 46I13 (lane 11) for c3.
Note that while all the BACs were loaded in a comparable concentration, lanes 1 and 11 have a lower concentration, and conclusions about homology of fragments to the probes is based on several additional experiments not shown in this figure.
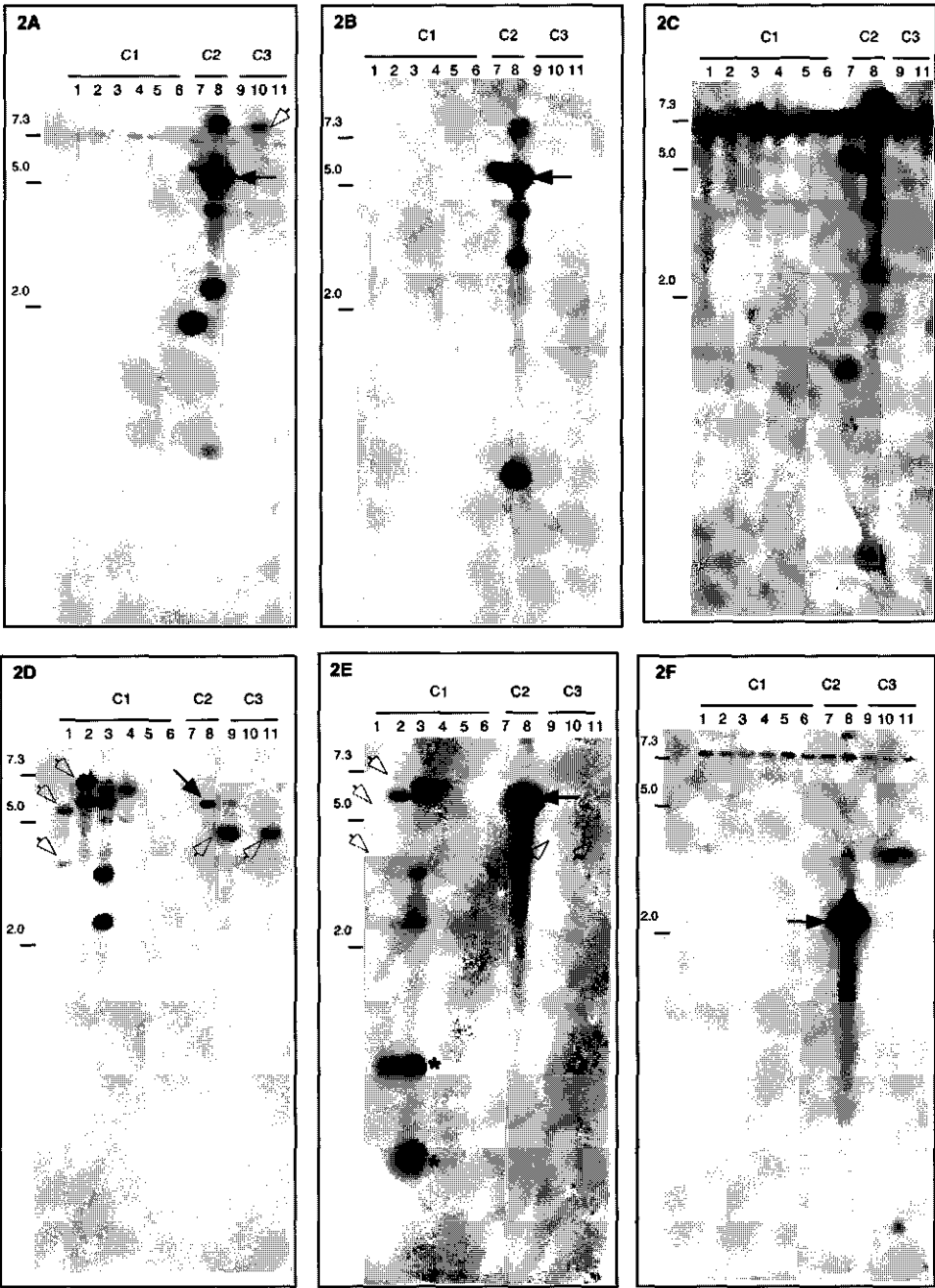
**Fig 2**

Some sequences hybridize with other contigs apart from the one from which they originate. In addition to MtgLRR52, Mtg3552 contains a sequence homologous to *Mtc424*, a cDNA with high homology to beta glucosidases that was isolated by screening a *M. truncatula* root hair cDNA library with Mtg3552. *Mtc424* hybridizes weakly with a fragment in c1 as well as in c2 (to the Mtg3552 fragment) (Fig 1A). This weak hybridization indicates that the gene corresponding with *Mtc424* is probably not located on these two contigs, but sequences with some homology to *Mtc424* are present in these contigs. The c2 5.6 Kbp subclone Mtg3556 hybridizes with a sequence family located in c1 (Fig 2E).

Out of the 22 studied sequences only four hybridized with c3: *Mtc411* and Mtg3552 (Figure 2A) with low intensity, and Mtg2.4 (c2-end genomic subclone) with high intensity. The pea cDNA clone *PscW62-1* hybridizes with similar high specificity to restriction fragments of c1 and c3, and with low specificity to a fragment in c2 (Fig 2D). These four sequences reveal homology between c1/2 and c3; however, the rest of the studied sequences are specific for the *Sym2*-orthologous *M. truncatula* c1/c2 genomic region.

## Distribution of kinase and LRR sequence families in the contigs

The LRR-containing sequences MtgLRR52 and *PscLRR52*, and the kinase-like sequences *PscW62-1* and Mtg3556, hybridized to many fragments that occur within several regions of the same contig and of different contigs (Fig 1A, Fig 2B,C,D,E). On the other hand, the LRR-containing subclone Mtg1957 is a single copy sequence in the *M. truncatula* genome and hybridizes to a single c2 contig region (Fig 1A, Fig 2F).

Most kinase-like sequences are localized on c1 (Fig 3A) and include the *PscW62-1* weak and strong hybridizing homologues (Fig 2D), the Mtg3556 weak hybridizing homologues (Fig 2E), and the fragments containing the genes of the cDNAs *Mtc156, Mtc831* (Fig 4). In c1, the *PscW62-1* homologues are located on two contig regions separated by about 50 Kbp (Fig 1A region 2, and region 6; and Fig 3A). The c1 fragments that weakly hybridize with *PscW62-1* also hybridize with the Mtg3556 (c2 subclone containing kinase-like sequences), but this *Medicago* subclone hybridizes with two additional small c1 HindIII fragments located in region 6 (Fig 1A, indicated by asterisks in Fig 2E) that are not detected by *PscW62-1*. This reveals that Mtg3556 (which sequence is fully homologous to kinase sequences. Geurts, pers. comm.) might contain unique kinase motives with homologues only in c1. Mtg3556 is homologous to *PscW62-1*, but their hybridization is very weak (Fig 2D, black arrow), and

Mtg3556 does not hybridize with the sequences most homologous to *PscW62-1* in c1 and in c3 (Fig 2D and 2E, white arrows). Thus, Mtg3556 seems to be more related to the c1 *PscW62-1* distant homologues than to the sequences in c1 and in c3 with high homology to *PscW62-1*. In addition to these kinase sequences from the three contigs that hybridize with *PscW62-1* and with Mtg3556, the c1 *Mtc831* and *Mtc156* (Fig 1A, region 12; Fig 3A, Fig 4) form a separate group of kinase sequences that does not cross-hybridize with *PscW62-1* and Mtg3556. One HindIII fragment of BAC59K07 hybridizes with both *Mtc831* and *Mtc156*, indicating the presence of common sequence motives in these two cDNAs (Fig 4, dashed line).
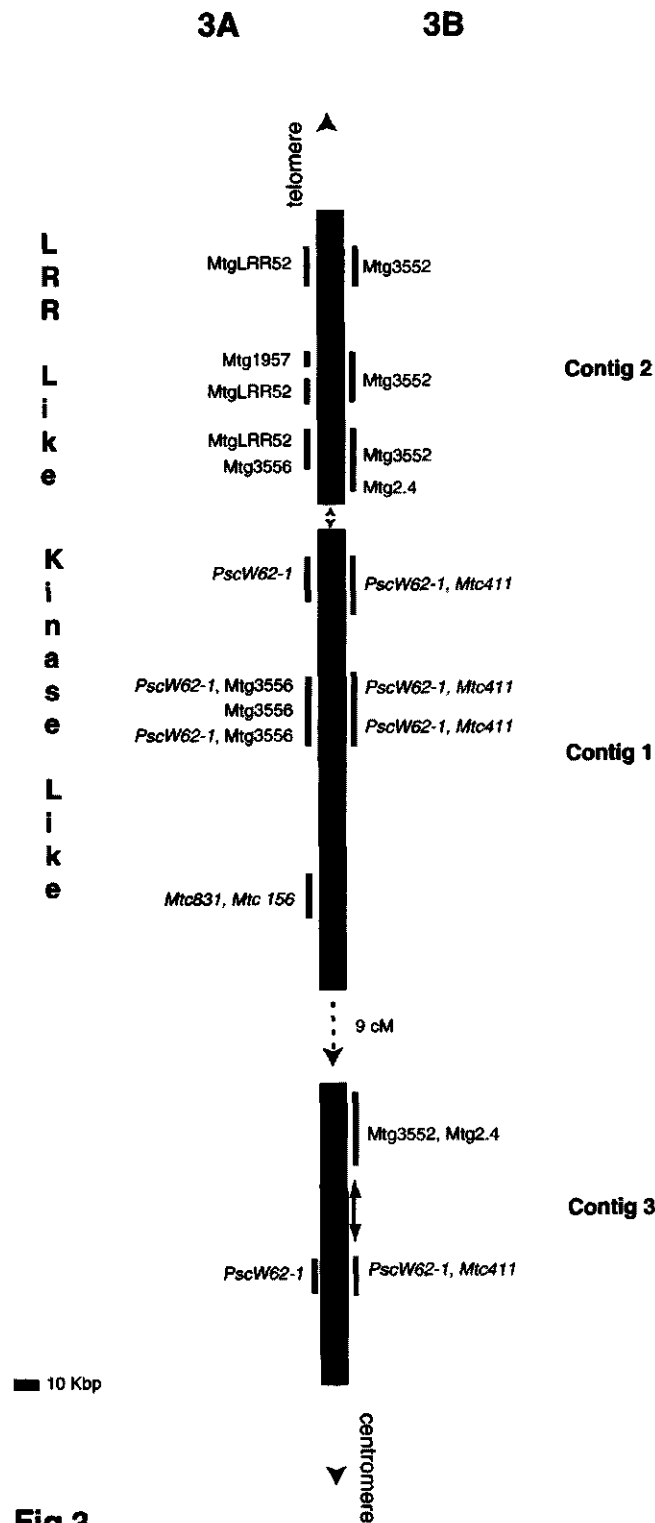
**Fig 3**

**Fig 3**: Scheme showing the distribution of sequences specified in 3A (left side of each contig) and 3B (right side of each contig). Only the contig regions hybridizing to these sequences are indicated as black or gray thin lines aside of the thick black/gray lines representing the contigs. Hybridizing contig regions that are contiguous or close to each other are drawn as a single line. c1 and c2 are drawn in the orientation as they occur in *M. truncatula* chromosome 5 (Gualtieri et al, submitted, a) with the chromosome walking sequence of c2 (gray thick line) directed towards the telomere and the chromosome walking sequence of c1 (gray thick line) towards the centromere. The orientation of c3 in *M. truncatula* chromosome 5 is unknown and therefore the orientation given in this figure is arbitrary. The double arrowhead line in between c1 and c2 represents the gap (10 kbp) between these two contigs in this chromosome.

**3A (left side)**: Distribution of kinase and LRR-containing sequences in the three contigs, indicated as black thin lines at the left side of the contig line.

**3B (right side)**: Comparison of the position of contig regions of c1/c2 and c3 containing Mtg2.4, Mtg3552, *Mtc411* and their homologues, and homologues of *PscW62-1*. Gray thin lines indicate sequence clusters that are colinear between c1/c2 and c3. Black thin lines represent other contig regions in c1/c2 containing additional homologues of these four sequences. The double arrowhead line in c3 indicates the distance (of about 50 kbp) separating the cluster groups 1 and 8, containing Mtg3552 and Mtg2.4, and *PscW62-1 Medicago* homologues and *Mtc411*, respectively.
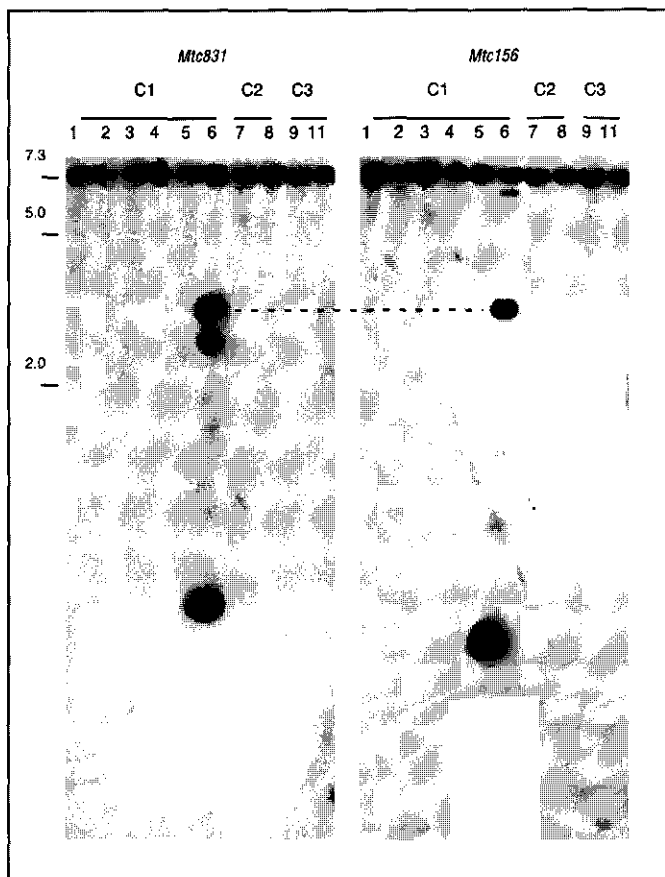


**Fig 4**

Southern blot of BACs from c1, c2, and c3 hybridized with *Mtc831* and *Mtc156*. The numbering of the lanes in this figure are the same as in Fig 2, except that the c3 BAC 27H19 is absent (*Mtc831* and *Mtc156* do not hybridize with BAC 27H19, data not showed). This figure shows how *Mtc831* and *Mtc156* represent a c1 specific kinase that does not hybridize with the *PscW62-1* homologs in c1 and c3, or with Mtg3556. In addition, these two cDNAs co-hybridize with a HindIII fragment of BAC 59K07 (indicated by a dashed line).

LRR-like sequences exclusively occur in c2 (Fig 1A; Fig 2B,F; Fig 3A) and three distinct LRR-containing regions can be distinguished in this contig. Region 2 is separated by about 55 Kbp from region 8, which is separated by about 15 Kbp from region 10 (Fig1A, Fig3A). These three contig regions contain restriction fragments that hybridize with the *Medicago* genomic sequence MtgLRR52 that is homologous to the LRR domain of *Cf4/Cf9*. Figure 2B shows that 7 HindIII fragments in c2 hybridize with MtgLRR52. Some of these bands also hybridize with the LRR-containing clones Mtg3552 (containing MtgLRR52) (Fig 2A) and *PscLRR52* (Fig 2C), while additional c2 bands that only hybridize with these two clones are homologous to non- MtgLRR52-like sequences present in Mtg3552 (5.2 Kbp) and *PscLRR52* (compare Fig 2A, 2B, and 2C).

In addition to these *Cf4/Cf9*-LRR domain like sequences in c2, the single copy sequence Mtg1957 (Fig 2F) is highly homologous to the LRR motif of the TMV resistance gene (Table 1), and is located on a single HindIII restriction fragment of c2 region 8 that does not hybridize with MtgLRR52.

Thus, two main types of LRR-like sequences that do not cross hybridize can be distinguished: the *Cf4/Cf9* LRR motif-like sequences that hybridize with MtgLRR52, and the TMV LRR motif-like sequence of Mtg1957.

## Order of sequences located in both c1/c2 and c3

In Fig 1A, the distribution of c3 HindIII restriction fragments that hybridize strongly with *PscW62-1* and Mtg2.4, and weakly with Mtg3552 and *Mtc411* is indicated. In c3, no LRR-like sequences and only a single genomic fragment highly homologous to *PscW62-1* is found. The c3 region 1 contains the homologues of Mtg2.4 and Mtg3552, and this is separated by about 50 Kbp from region 8 containing the homologues of *PscW62-1* and *Mtc411* (Fig 1A, Fig 3B). The distribution of these sequence clusters (Mtg2.4/Mtg3552 and *PscW62-1* homologues/*Mtc411*) in c3 is similar to that in c1/c2 (Fig1A, Fig 3B). In c1, *Mtc411* as well as *PscW62-1* homologues are localized in region 2; while in c2, Mtg3552 homologues are localized in region 10 that is only 5 Kbp apart from region 13 containing Mtg2.4 (Fig 1A; Fig 3B). Considering that the gap separating c1 and c2 is about 10 Kbp (Gualtieri et al., submitted, a), the c2 region 10 and region 13, are separated by about 20 Kbp from the c1 region 2 (Fig 1A, Fig 3B). All together, these observations show that the order of clusters of these four studied sequences is very similar in c1/c2 and c3, indicating that these genomic regions arose through duplication of a chromosomal segment. It is worth noting however that,

as already mentioned, additional homologues of three of these four sequences are present in c1/c2, which probably arose by duplications within this region (Fig 1A, Figure 3B).

## CONCLUSIONS

The synteny-based positional cloning of genes of a species of interest requires the identification of orthologous regions in a model species that can be used as intergenomic cloning vehicle. *M. truncatula* has several features making it useful as a model system (Cook, 1999). For microsynteny-based positional cloning studies, the identification of an orthologous genomic region in a model plant can be complicated by the existence of duplicated regions within the genome. Our studies show that although in general the sequence composition of c1/c2 is different from c3, the sequences Mtg3552, Mtg2.4, *Mtc411* and their homologues, as well as *PscW62-1* homologues, are present both in c1/c2 and c3. Moreover, their order is comparable in the *M. truncatula* chromosome 5 c1/2 and c3 genomic regions. Since the probability that these clusters of sequences originated independently is very low, it is quite likely that c1/c2 and c3 arose through a chromosomal duplication. However, since the orientation of c3 in chromosome 5 has not been determined by FISH it is not known whether these duplicated genomic regions have a direct or inverse chromosomal orientation. Similarly to our findings, in grasses and in *Arabidopsis* whole chromosome arms and chromosomal segments have been found duplicated (Ahn et al., 1993; Chen et al., 1997; Conner et al., 1998; Foote et al., 1997; Grant et al., 2000; Helentjaris et al., 1998; Ku et al, 2000; Langercrantz, 1998).

Our studies confirm that c1/c2 represents the *Medicago Sym2*-orthologous genomic region, because most of the 22 studied sequences that originate from c1/c2 are absent in c3. However, these sequences represent only about 2% of the total sequence of the three contigs, but after their hybridization to HindIII contig DNA blots about 50 % in c2 and 25% in c1 of the total contig HindIII fragments revealed to contain homologues to these sequences. Thus, the complete sequencing of the contigs will probably reveal additional sequence homologues present both in c1/c2 and c3, which might be linked to the pea *Sym2* gene. The latter becomes relevant if *Medicago* is used as a functional test-system (by reverse or direct genetics) to identify the *Medicago Sym2*-orthologue, because *Sym2* itself could belong to a family of homologous sequences.

## MATERIALS AND METHODS

### BAC clone isolation and construction of contigs

Bacterial Artificial Chromosome (BAC) DNA of *M. truncatula* genotype A17 belonging to c1/2 and c3 (Gualtieri et al., submitted, a) was isolated according to Nam *et al.* (1999). Contigs were constructed by a combination of BAC DNA fingerprinting and restriction/hybridization analysis. This resulted in arrays of overlapping BAC clones that formed contig. The sequences studied in this manuscript were assigned to different contig regions by hybridization.

### Contig DNA blots and *Medicago* genomic blots

To prepare contig DNA samples containing the total sequence from each contig, a set of BAC clones with a minimal overlap and representing the complete contig sequence was selected. Contig 1 DNA samples included BAC clones 63010, 11I01, 20K04, 52010, 56F17 and 59K07. Contig 2 DNA samples included BAC clones 21J05 and 15B03. Contig 3 DNA samples included BAC clones 27H19, 30F10 and 46I13. For Southern blot analysis 2 μg of BAC DNA of each selected BAC clone was used, whereas 5 μg of *M. truncatula* A17 genomic DNA was used per lane. The individual BAC DNA samples and *Medicago* genomic DNA were digested to completion with HindIII (Gibco BRL), and analyzed by electrophoresis on 20 cm long 1% agarose gels prepared and run in 1x TAE at 45 V. Electrophoreses were done over night. DNA was stained with ethidium bromide 0.05 % (w/v) in running buffer and visualized by UV illumination.

### Blot transfers and hybridizations

After agarose gel electrophoresis, DNA was transferred to a Hybond N+ (Amersham) nylon membrane. The conditions for DNA transfer and for prehybridization and hybridization with probes were as recommended by the membrane manufacturers (Amersham).

Probes were prepared by random-priming labeling (Feinberg et al., 1983) of 50 ng of template DNA, including 4 units of Klenow DNA polymerase and 2 μCi of [$^{32}$P] dATP in the reactions. The reactions were incubated for 1 hour and probes were purified from non-incorporated nucleotides through a Sephadex G50-medium column prepared in a 1 ml syringe.

Prehybridizations and hybridizations were carried out at 60-65 $^{o}$C, and washings were generally done in 5x SSPE 0.1 % SDS for 10 min, 2x SSPE 0.1 % SDS for 20 min, 1x SSPE

0.1% SDS for 15 min, and in 0.5 SSPE 0.1 % SDS for 10 min. In order to visualize the different hybridization specificities, hybridizations were done at 60 $^{o}$C-65 $^{o}$C, and blots were usually washed and exposed after both low and high stringency washings. Blots were analyzed with a Storm840phosphor imager.

**cDNA libraries and screenings**
The pea root hair cDNA library was constructed by Stratagene in Lambda ZAPII vector system using equal amounts of poly (A)+ RNA isolated from a mixture of root hairs of 6 days old cv Finale pea plants. These root hairs were isolated from uninoculated seedlings as well as seedlings inoculated with *R. leguminosarum* bv *viciae* 248 (Josey et al., 1979) 48 hrs before harvest, respectively. Equal amounts of these two root hair preparations were used for the construction of the library. The *Medicago truncatula* A17 root hair cDNA library was kindly provided by S. Long (Covitz et al., 1998). The cDNA library screenings were done with Hybond-N+ membranes using the conditions recommended by the library manufacturers (Stratagene). DNA probes for the cDNA library screenings were labeled as described above.

**DNA sequencing**
Automated sequencing was done on an ABI 377 DNA sequencer.

# CHAPTER 4

**Isolation of RFLP markers linked to the *nod3* hypernodulation locus on pea linkage group I, and identification of orthologous genomic regions in *Medicago truncatula* A17**

Gustavo Gualtieri [1][#], Olga Kulikova [1], Douglas Cook [2] and Ton Bisseling [1]

1) Molecular Biology, Agricultural University Wageningen. Dreijenlaan 3, 6703HA Wageningen, The Netherlands.
2) Department of Plant Pathology, University of California at Davis. One Shields Avenue. Davis, CA 95616-8680. U.S.A.

# Author for correspondence

## ABSTRACT

The pea (cultivar Rondo) mutant *nod3* has a hypernodulation phenotype determined by a monogenic and recessive allele. *Nod3* maps in pea linkage group I 2 cM above *Sym2*. In this manuscript we describe the isolation by RNA differential display of the cDNA RFLP markers *dd21.5* and *Psc2.6*. These markers map between *PsEil2* and *Nod3*, as was shown by using a set of RILs and introgression lines. Moreover, *dd21.5*, encoding a protein that is highly homologous to subtilisin proteases, is not expressed in root hairs of a *Sym2* containing introgression line. *Psc2.6* encodes for a protein homologous to that encoded by the ethylene-regulated gene *ER6* and is located close to the *PsEil2* locus.

A *Medicago truncatula* A17 BAC library was screened with the pea single copy sequence *dd21.5*, which *Medicago* orthologue (Mtg21.5) is also a single copy sequence in line A17. BAC clone 21F22 was isolated. Southern blot analysis of BAC and *Medicago* genomic DNA demonstrated that BAC 21F22 contains the complete sequence of the *Medicago* orthologue Mtg21.5. In addition to this BAC, five BAC clones were identified that co-hybridize both with *dd21.5* and *Psc2.6*, demonstrating that as in pea, a *Medicago* genomic region contains orthologues of these two markers that are tightly linked in both species. This reveals the existence of microsynteny between these genomic regions of pea and *Medicago*. However, at the chromosome level a disturbance of overall synteny was observed between the two species. BAC 21F22 was mapped by FISH on *M. truncatula* A17 chromosome 4. Additional weak signals were observed close to the telomeres of chromosome 5. The mapping of 21F22 on chromosome 4 demonstrates a local loss of chromosomal synteny-at the "*dd21.5* region"-between *Medicago* chromosome 5 and pea linkage group I that are otherwise syntenic for the "*Sym2* region" and "*Eil2* region". However, it is not known yet whether the five BACs that contain homologues of both *dd21.5* and *Psc2.6* form a contig with 21F22 and map on chromosome 4, or if they contain the *Medicago dd21.5*-distant homologue identified by low stringency Southern blots and map somewhere else in the genome, e.g. on chromosome 5 where 21F22 reveals weak signals. In addition, it cannot yet be concluded whether the *Medicago Nod3*-orthologous locus is, as Mtg21.5, located on chromosome 4. The construction of a larger contig encompassing this chromosome 4 region and the five BACs (that might or might not link to 21F22 and map on chromosome 4) and further microsynteny analysis of this/ese *Medicago* genomic region/s and the pea *Nod3* region, could result in the cloning and delimitation of the *Medicago* genomic region orthologous to the pea *Nod3* region.

This could lead to the cloning of pea *Nod3* by using *M. truncatula* as intergenomic gene-cloning vehicle.

## INTRODUCTION

Comparative molecular genomics opens the possibility to clone genes across species. This requires however, that the marker composition of the homologous genomic regions comprising the target locus is syntenic over a few hundred kilobase pairs.

In plants, most of the comparative genomics knowledge comes from studies on grasses (Family of Poaceae or Gramineae). Here, long-range as well as local colinearity were reported, at the whole-genome, and at the megabase/submegabase levels, respectively (Bennetzen et al. 1997, Foote et al. 1997, Gale et al. 1998, Hulbert et al. 1990, Dunford et al. 1995, Kilian et al. 1997, Chen et al. 1997, Ahn et al. 1993, Tikhonov et al. 1999). Exceptionally, small-scale chromosomal rearrangements, such as deletions, duplications and sequence displacements, have locally disrupted regions of high colinearity (Chen et al. 1997, Conner et al. 1998, Foote et al. 1997, Kilian et al. 1997, Tikhonov et al. 1999). In addition, gene and chromosomal segment duplications were described in grasses and the family Brassicaceae (Ahn et al., 1993; Chen et al., 1997; Conner et al., 1998; Grant et al., 2000; Helentjaris et al., 1998; Ku et al, 2000; Langercrantz, 1998). When grasses were compared with *Arabidopsis* a low level or no colinearity was observed (Devos et al., 1999; van Dodeweerd et al., 1999). However, within the family of Brassicaceae a high level of colinearity can be found (Langercrantz et al., 1996; Conner et al, 1998). We have recently reported on a synteny-based positional cloning approach of the pea *Sym2* locus that resulted in the isolation of an orthologous genomic region on the long arm of *Medicago truncatula* A17 chromosome 5 that is highly microsyntenic to the *Sym2* genomic region on pea linkage group I (Gualtieri et al., submitted, a). Thus, in general, species within a plant family appear to be (micro)syntenic.

We are testing the possibility to use microsynteny for the isolation of genes of legumes with a large genome. This is possible thanks to the generation of molecular and biological resources of the legume model plant *Medicago truncatula* (Cook, 1999; Covitz et al., 1998; Nam et al., 1999). Microsynteny would make it possible to clone genes which phenotypes have been described in species with a complex genome, but not in the model legumes. This is well

exemplified by pea (*Pisum sativum*), a legume plant with a relatively large genome ($\pm 4.10^9$ base pairs per haploid genome) (Ellis, 1993) that has a long history of studies, at both a genetic and a biochemical level.

Pea linkage group I has been extensively characterized at a genetic level since it contains many interesting symbiotic genes located in a region of about 20 cM in linkage group I; e.g. *Sym2*, *Sym5*, *Sym18*, *Sym19*, *Nod3*, leghaemoglobin (*Lb*), *Enod7* and *Enod40* (Weeden et al., 1990; Temnykh *et al.*, 1995, Kozik, 1996; Geurts, 1998). We have a special interest in two loci: *Sym2* and *Nod3*. *Sym2*, controlling Nod factor structure-dependent infection (Firmin et al., 1993; Kozik et al., 1995; Geurts et al., 1997; Ovtsyna et al., 1998) is a locus present in Afghanistan pea (Lie, 1984). *Sym2*-containing peas are only nodulated by *Rl* bv *viciae* strains containing the *nodX* (or *nodZ*) gene that modifies the Nod factor structure (Firmin et al., 1993; Ovtsyna et al., 1998), but peas without the Afghanistan *Sym2* allele are nodulated by bacterial strains both with and without the *nodX* gene (Kozik, 1995). *Nod3* (Jacobsen et al. 1984; Postma et al., 1988; Temnykh et al., 1995 a, b) leads to a hypernodulation phenotype when it is mutated. The distance between these two loci on pea linkage group I is about 2 cM (Geurts, 1998; Kozik, 1996).

In this manuscript, we report the isolation of the cDNA RFLP markers *Psc2.6* and *Psc21.5*, that are located tightly linked to each other on pea linkage group I, at a position in between *Eil2* (Kim et al., unpublished) and *Nod3*. These markers were isolated by comparing the RNA differential display profiles of root and root hair RNA from Recombinant Inbreed Lines (RILs), introgression lines and their respective parental lines. RFLP was used to map these markers in pea. In addition, a *M. truncatula* A17 BAC library was screened both with *Psc2.6* and *Psc21.5* in order to identify the *Medicago* genomic region that is orthologous to the pea segment of linkage group I containing these two markers. BAC clone 21F22 containing the orthologue of *dd21.5* was isolated and 5 BACs that co-hybridize both with *dd21.5* and *Psc2.6* were identified. The map position of BAC 21F22 was determined by FISH.

## RESULTS

**Isolation of markers linked to *Nod3* by RNA differential display.**
We have been using the RNA differential display to isolate genes that might be linked to *Sym2* or *Nod3*.

For the isolation of genes linked to *Sym2*, root and root hair total RNA of Rondo and the *Sym2*-containig introgression line A15.5.2, were compared by RNA differential display. Fig 3 shows the *Sym2*-containing introgressed region of A15.5.2 that originates from Afghanistan pea (the construction of A15.5.2 is described in "Material and Methods"). Since it is not known whether *Sym2* is a constitutively expressed or a *Rhizobium*-induced gene, these two lines were inoculated two days after germination with *Rhizobium leguminosarum* bv *viciae* strain 248pMW107 (Kozik et al., 1995) containing the *nodX* gene (Firmin et al., 1993). Roots and root hairs were harvested four days after germination and total RNA was isolated and compared by RNA differential display. In this way, one differentially occurring band, named *dd21.5*, was cloned and sequenced. *dd21.5* has a poly-A tail and a size of 219 bp. The sequence of this cDNA fragment has homology with subtilisin proteases. The RNA differential display analysis shows that *dd21.5* mRNA is present at a similar and relatively high level in roots of Rondo and the *Sym2*-containing line A15.5.2 (Fig 1A). This relatively high level of *dd21.5* RNA also occurs in root hairs of Rondo; however, the *dd21.5* band is completely absent in the root hair RNA fingerprint of the *Sym2*-containing A15.5.2 introgression line (Fig 1A). Since the *dd21.5* cDNA band is present in the root preparations of Rondo and A15.5.2, the *dd21.5* gene was differentially displayed not due to a PCR polymorphism. Therefore the difference observed between the root hair preparation of Rondo and A15.5.2 is most likely due to a difference in the expression level of the *dd21.5* gene. This indicates that either *Sym2* itself, or another gene present in the *Sym2*-containing introgressed region, or a gene present in an Afghanistan introgressed region other than the *Sym2*-containing region, is responsible for the repression of the *dd21.5* gene in root hairs of A15.5.2. Thus, because the differential display of *dd21.5* is caused by an expression difference rather than a PCR polymorphisms, it was not possible to predict from the display pattern whether *dd21.5* was located on linkage group I and therefore this cDNA was mapped by RFLP by analysis of a set of RILs and introgression lines. This showed that *dd21.5* is linked to *Nod3* (see RFLP section).

For the isolation of markers linked to *Nod3*, the Rondo line (containing the wild type *Nod3*) and the *nod3*-containing RILs (Geurts, 1998) D22.2, D25.2 (also *Sym2*-containing), D41.6 and D26.6 (also *Sym2*-containing), were compared by RNA differential display. The RNA differential display was carried out with total RNA isolated from 4-days old uninoculated roots of the *nod3*-containing RILs and Rondo. When *Nod3* is constitutively expressed (i.e. not dependent on bacterial inoculation), these experimental conditions would enable the isolation
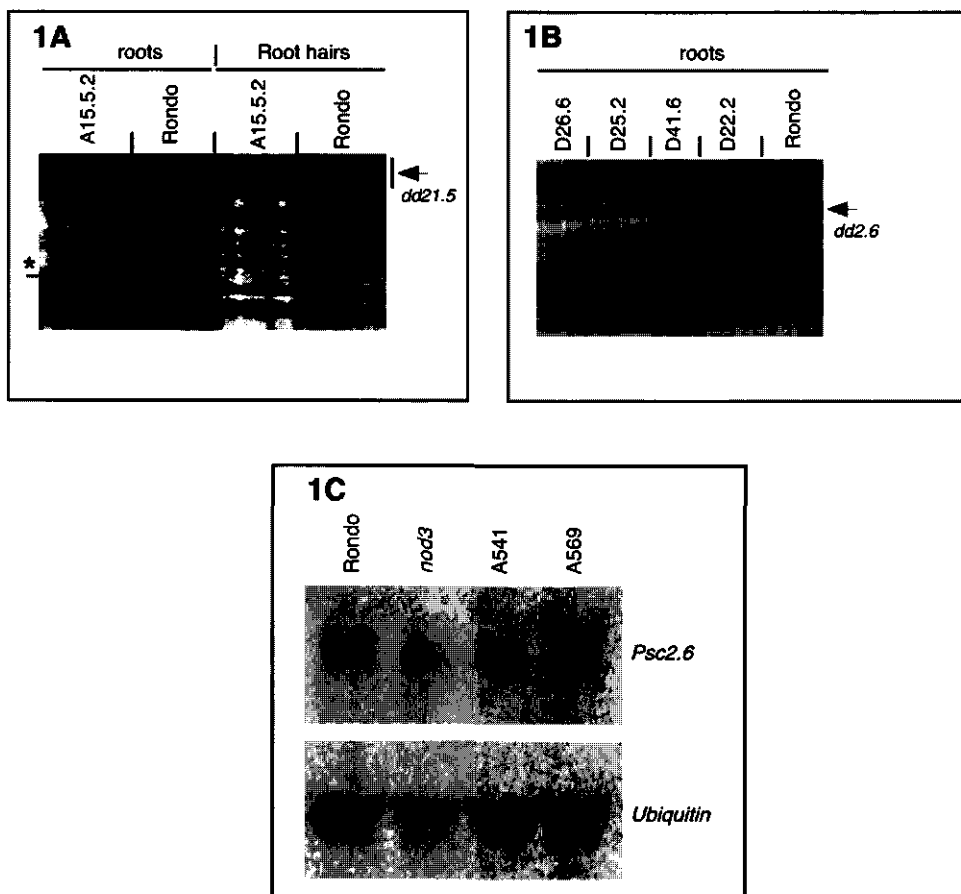
**Fig 1**

**1A**: Expression of *dd21.5* studied in roots and root hairs by DDRT-PCR. *dd21.5* (black arrow) is expressed in a comparable high level in roots from all plants. However, this transcript is absolutely and specifically abolished in root hairs of the *Sym2*-introgression line A15.5.2 (white arrowhead), in contrast with the high level of the transcript in root hairs of cv Rondo, which is similar to the level in roots. Other transcripts in the pattern, demonstrate that the concentrations of total RNA and first strand cDNA was similar for the different samples, and that degradation of transcripts has not occurred. The asterisk indicates a transcript, which is specifically present in roots (but not in root hairs) of line A15.5.2.
**1B and 1C**: Expression of *dd2.6* in roots. **1B**: DDRT-PCR showing the *dd2.6* band (black arrow) present in the RNA display of cv Rondo. However, this band is completely absent in the *nod3*-RILs (white arrowhead). **1C**: Northern blot showing that *dd2.6* hybridizes to a single transcript and that the expression of *dd2.6* is not affected in the *nod3* line. This indicates that the absence of *dd2.6* in the RNA differential display is due to a sequence polymorphism between the *nod3* RILs and the Rondo lines.

of *Nod3* itself by detecting polymorphisms (modified transcript) or the absence of a transcript caused by the mutagenesis. In addition, it would enable the isolation of genes whose expression could be regulated by *Nod3*. With these experimental conditions a 3' partial cDNA

named *dd2.6* was isolated, cloned and sequenced. *dd2.6* has a poly-A tail and a size of 294 bp. *dd2.6* is highly homologous to the ethylene responsive *ER6* gene *of L. esculentum* (Zegzouti et al., 1999). *dd2.6* is present at relatively high levels in the RNA fingerprint of the Rondo line, but is absent in those of the *nod3*-containing RILs that were analyzed (Fig 1B). This could imply that *dd2.6* is mutated in the *nod3* mutant or it might be a part of the *Nod3* gene itself. Alternatively, *dd2.6* could be positively regulated by the wild type *Nod3* gene. This later possibility was studied by northern blot analysis.

**Northern blot expression data**

To understand the reason for the absence of the *dd2.6* band in the RNA fingerprints of the *nod3*-containing RILs, the expression of the *dd2.6* gene in Rondo, A569 (contains *Nod3*), A541 (contains *Nod3*), and the *nod3* mutant was studied by northern blot. For this, total RNA of uninoculated roots harvested 4 days after germination was isolated.

When *dd2.6* is used to hybridize pea genomic blots of DNA digested with HaeIII, it hybridizes to a single band (Fig 4A). When *dd2.6* is used as a probe on northern blots, it hybridizes to a single transcript present at comparable level in the four lines studied (Fig 1C). The same single transcript is also observed when *Psc2.6* (that includes *dd2.6*, see below) is used as a probe on the same northern blots (data not shown). This demonstrates that the expression of *dd2.6* (and *Psc2.6*, see below) is not affected in the *nod3* mutant, and demonstrates that the absence of the *dd2.6* band in the RNA display of the *nod3*-containing RILs is caused by a sequence polymorphism. Therefore, the RNA differential display pattern of *dd2.6* should reflect its map position in the genome that should be linked to *Nod3*. This has been analyzed by RFLP mapping (see next section).

**RFLP mapping in pea**

DNA from the lines Afghanistan; Rondo; *nod3*; the introgression lines (named with an "A" prefix) A569 (Kozik et al., 1995), A541 (generated by an additional backcross of A54 with Rondo)(Kozik et al., 1995), and A15.5.2 and A33.18 (Kozik, 1996); and the RILs (Geurts, 1998) derived from a cross between A54 and *nod3* (named with an "D" prefix) were used for RFLP mapping (Fig 3) (see "Material and Methods").

On HaeIII genomic blots, *dd21.5* hybridizes weakly to two non-polymorphic bands, and strongly with a fragment that is polymorphic between the Rondo/*nod3* and Afghanistan lines. This polymorphic band maps in pea linkage group I, between *Nod3* and *PsEil2* (Fig 2, Fig 3).
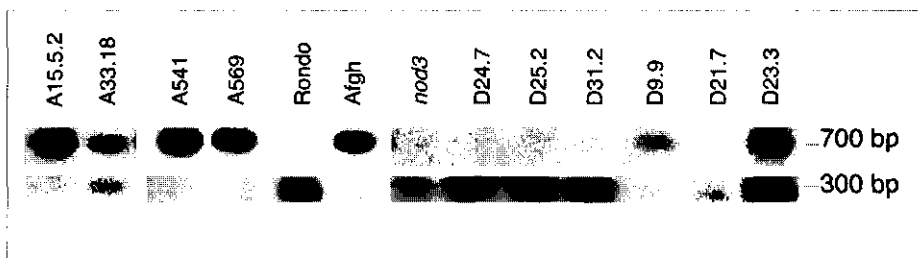


**Fig 2**

RFLP mapping of *dd21.5*. DNA was digested with HindIII. The number of introgression lines, RILs and parental lines is indicated above each lane. Note that the DNA concentration of lanes containing the lines A33.18, *nod3*, D9.9, and D21.7 is considerably lower than for the other lines tested. Because a real line cannot be heterozygotic, A33.18 and D23.3 are a mixture of two groups of individuals containing the Rondo/*nod3* allele or the Afghanistan allele at the *dd21.5* locus, respectively. The map position of *dd21.5* as deduced from this blot is represented in figure 3. "Afgh" means cv Afghanistan.

The introgression line A33.18 contains both the Rondo/*nod3* and the Afghanistan alleles of *dd21.5* (Fig 2), and since it is homozygotic at the *PsEil2* locus and all the markers from the *Sym2* region (Fig 3), A33.18 must be a mixture of two groups of individuals with a different allelic composition at the *dd21.5* locus. Therefore in a group of these individuals, a cross over has occurred in between *PsEil2* and *dd21.5* (Fig 2, Fig 3). Moreover, since A15.5.2 is homozygotic for the Afghanistan allele, the introgressed region of this plant extends further to the *Nod3* locus, than that of the A33.18 group of individuals referred to as G2 in Fig 3. The RIL D23.3 also must be a mixture of two groups of individuals with different alleles at the *dd21.5* locus, and in a group of these individuals a double cross over occurred in between *PsEil2* and *Nod3*. Since the intensity of hybridization of the two alleles is comparable in D23.3 (Fig 2), these two groups of individuals appear to be equally represented. Thus, a group of individuals (referred to as G2 in Fig 3) of the D23.3 plant mixture contains a segment of introgressed Afghanistan DNA at and around the *dd21.5* locus, in addition to the *Sym2*-containing Afghanistan introgressed region (Fig 2, Fig3). This introgressed segment is also revealed by the *Psc2.6[AH1]* band, which is tightly linked to *dd21.5* (see below).
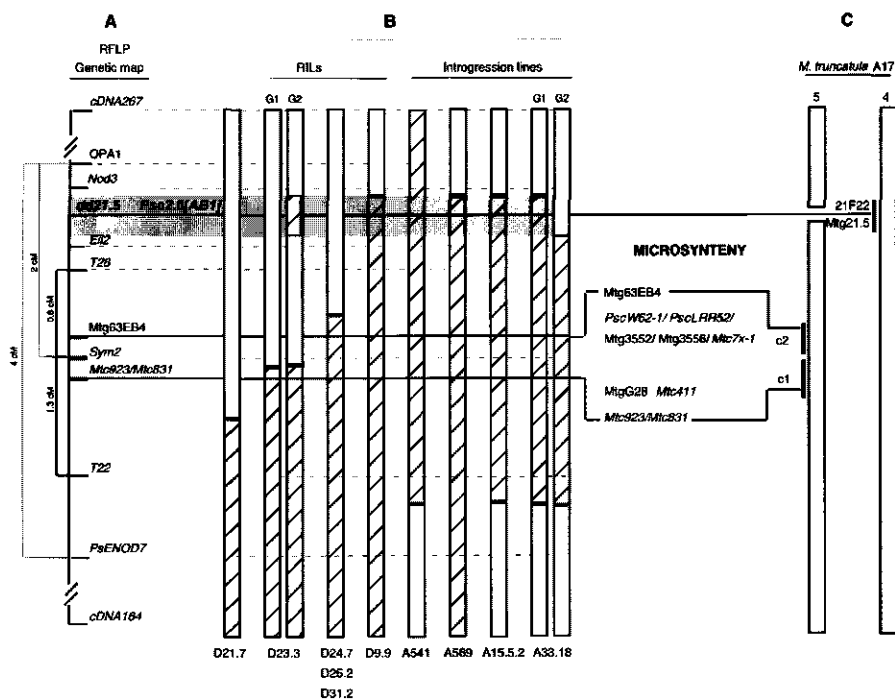
**Fig 3**

**A**: Genetic map of a 4 cM region that separates OPA1 from *PsENOD7* on pea linkage group I. The genetic distance between *cDNA164* and *PsENOD7*, and OPA1 and *cDNA 267* (Kozik, 1996), is not presented in this figure. The genetic distance between markers is indicated at the left of the map (Geurts, 1998), while only the linear order is known for other markers mapped by using the RILs. Markers tightly linked to *Sym2*, such as *PscW62-1* (Geurts, 1998) and the markers isolated by using the *M. truncatula* A17 *Sym2*-orthologous region (Gualtieri et al., submitted, a) are listed together with the *Sym2* locus. Markers *Mtc923/Mtc831* and Mtg63EB4, were isolated from the *Medicago Sym2*-orthologous region and  2 and 1 recombinations, respectively, were identified between *Sym2* and these markers (Gualtieri et al., submitted, a). The horizontal shadowed area indicates the map position of the new markers presented here, *dd21.5* and *Psc2.6[AB1]*(shown in bold), that map in between pea *Eil2* and *Nod3*.
**B**: Hatched boxes are Afghanistan DNA and white boxes Rondo DNA (lines with prefix "D" and "A" are *nod3* and *Nod3*, respectively). Since lines A33.18 and D23.3 are a mixture of two groups of individuals "G1" and "G2" indicate these two groups. The lines that have Afghanistan DNA (hatched box) at the position where *Sym2* maps show the *Sym2* phenotype.
**C**: representation of *M. truncatula* A17 chromosome 5 and 4 (white boxes). The chromosome number is indicated at the top of each box. Small black lines at the left side of the chromosomes indicate the position of contig 1 (c1) and contig 2 (c2) that define the *Sym2*-orthologous region on chromosome 5, and BAC 21F22 located on chromosome 4. Horizontal lines connecting figures **A** and **B** with **C** show regions of microsynteny between pea linkage group I and *Medicago* chromosomes 5 and 4.

cDNA clone *dd2.6* strongly hybridized to a single non-polymorphic band on Southern blots containing DNA digested with HaeIII (Fig 4A) and HindIII (not shown), respectively, and polymorphisms were also not detected with other restriction enzymes. To increase the chance to find polymorphisms, a pea root hair cDNA library was screened with *dd2.6* and a cDNA

clone with an insert of about 900 bp, named *Psc2.6*, was isolated and used for RFLP mapping. Both *dd2.6* (294 bp) and *Psc2.6* have a poly-A tail and sequence comparisons revealed that they are 100% identical. Thus, *dd2.6* is included in the larger cDNA sequence *Psc2.6* and corresponds to the same transcript. On genomic blots containing DNA digested with HaeIII, *Psc2.6* hybridizes to several DNA fragments in addition to the band detected by *dd2.6* (Fig 4A). One polymorphic band-named "*Psc2.6 Afghanistan Band 3*" (*Psc2.6[AB3]*)-is only present in Afghanistan and not in Rondo, *nod3*, the Rondo *Sym2*-containing introgression lines and the *nod3*-containing RILs, and therefore it does not map in pea linkage group I between *cDNA267* and *cDNA164* (Fig 3, some of the plants tested are shown in Fig 4A), but elsewhere in the genome. In addition, one weakly and one strongly hybridizing polymorphic band-both named "*Psc2.6 Rondo/nod3 Band 2*" (*Psc2.6[RnB2]*)-present in Rondo/*nod3*, in the Rondo *Sym2*-containing introgression lines and in the *nod3*-containing RILs, but absent in Afghanistan, do not map between *cDNA267* and *cDNA164* (Fig 3, some of the plants tested are shown in Fig 4A). It is unknown if *Psc2.6[AB3]* and *Psc2.6[RnB2]* are allelic.
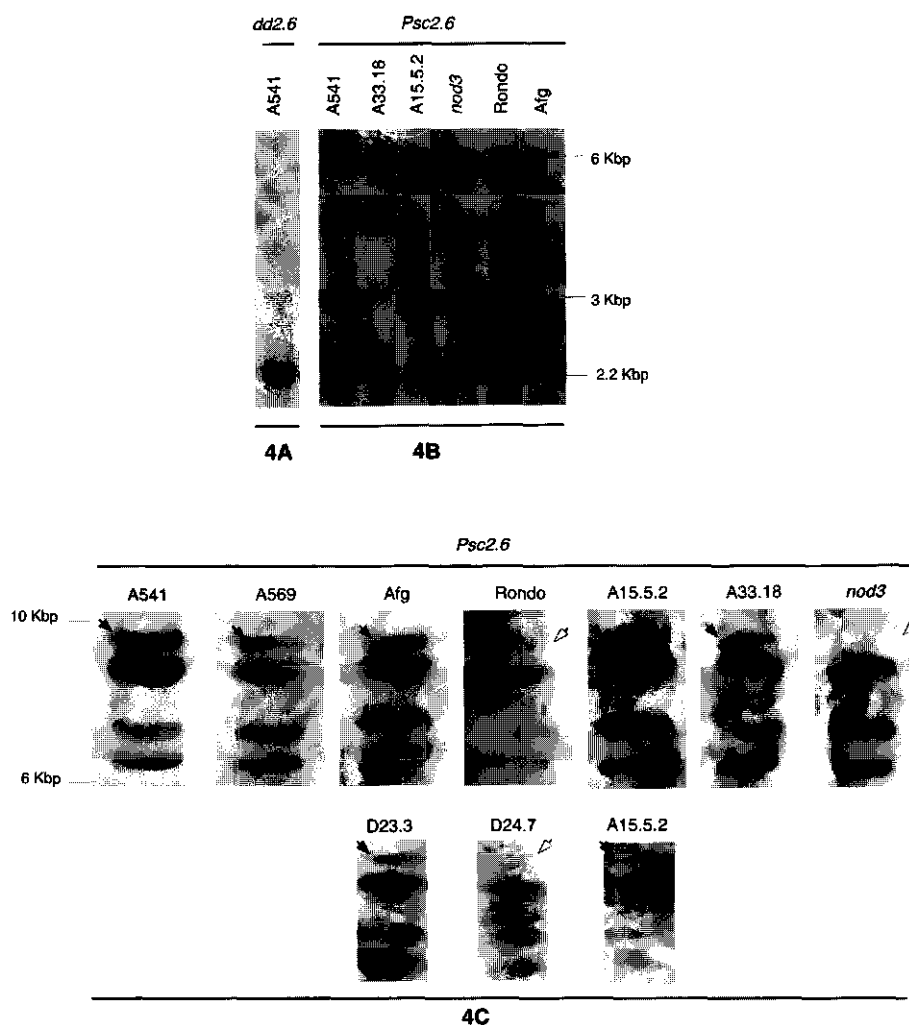
**Fig 4**

RFLP mapping of *Psc2.6* using RILs and introgression lines. Some of the hybridization data of the plants analyzed are shown in this figure. 4A and 4B: HaeIII RFLP. 4A shows the hybridization of the single copy cDNA clone *dd2.6*. 4B shows the hybridization of cDNA *Psc2.6*. The different bands referred in the text as *Psc2.6[AB3]* (white arrow head) and *Psc2.6[RnB2]* (black arrow head) are indicated. Both copies do not map between *cDNA267* and *cDNA164* (see Fig 3). 4C: HindIII RFLP showing the hybridization of *Psc2.6*. The copy referred in the text as *Psc2.6[AB1]* (indicated by a black arrow) is unique of Afghanistan (Afg) and is tightly linked to *dd21.5* and located in between *Eil2* and *Nod3* (see Fig 3).

On the other hand, when Hind III blots are hybridized with *Psc2.6* four bands are observed. The strongest band of the pattern also hybridizes with *dd2.6*. One weakly hybridizing polymorphic band-named "*Psc2.6 Afghanistan Band 1*" (*Psc2.6[AB1]*)-present in Afghanistan but absent in Rondo/*nod3*, shows the same map position as *dd21.5*, between *PsEil2* and *Nod3*. *Psc2.6[AB1]* is present in A15.5.2, A33.18, A541, A569, D9.9 and D23.3,

and absent in Rondo, *nod3*, D21.7, D24.7, D25.2, and D31.2 (Fig 3, some of these plants are shown in Fig 4C). Therefore, the *Psc2.6[AB1]* band revealed that the D23.3 RIL contains introgressed Afghanistan DNA in between *PsEil2* and *nod3*, as it was also shown by the hybridization with *dd21.5*. However, since in this case the polymorphism is due to the "presence" of a band in Afghanistan and the "absence" in Rondo/*nod3*, *Psc2.6[AB1]* cannot reveal as *dd21.5* that D23.3 is a mixture of individuals with different allelic composition at the *Psc2.6[AB1]* locus.

Thus, the HaeIII and HindIII RFLP analysis shows that several pea genomic fragments hybridize with *Psc2.6* and *Psc2.6[AB1]* is tightly linked to *dd21.5*, mapping in between *PsEil2* and *Nod3* (Fig 3). Therefore, when in this manuscript the RFLP marker *Psc2.6* is quoted as linked to *Nod3*, it is with reference to the *Psc2.6[AB1]* band (since polymorphic bands revealed by the HaeIII RFLP analysis are not linked to *Nod3*).


**Identification of *M. truncatula* A17 orthologous genomic regions**

A *M. truncatula* A17 BAC library (Nam et al. 1999) was screened with *dd21.5* and *dd2.6*, to identify genomic regions containing the orthologues of these two pea RFLP markers. BAC clone 21F22 was isolated by using *dd21.5* as a probe. This BAC clone has an insert of about 70 Kbp and a HindIII fragment of 700 bp that strongly hybridizes with the pea cDNA *dd21.5* (Fig 5). Blots of *M. truncatula* genomic DNA and BAC 21F22 DNA digested with HindIII and hybridized with *dd21.5* show that the *Medicago* orthologue (named Mtg21.5) of *dd21.5* is single copy and that it is completely located within BAC 21F22, since only the 700 bp band hybridizes in genomic DNA (Fig 5A). However, an additional band is visible at low stringency hybridization and washings (2 times in 5x SSPE, 0.1 % SDS for 15 min, and 1 time in 2x SSPE for 5 minutes) and this band is not present in 21F22, indicating the presence of a *dd21.5*-distant homologue not located in the *Medicago* genomic region represented by this BAC (Fig 5A). On the other hand, the same blot hybridized with *Psc2.6* (containing *dd2.6*) shows that, as expected, none of the bands detectable in *Medicago* genomic DNA are present in BAC 21F22 (Fig 5B). Thus, BAC 21F22 represents the *M. truncatula* A17 genomic region that contains the true orthologous sequence of the pea cDNA *dd21.5* that maps between *Nod3* and *PsEil2*.
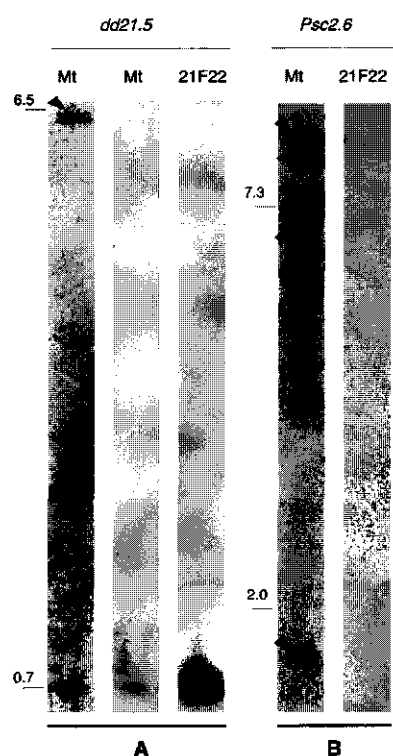
Southern blots of *M. truncatula* A17 (indicated as Mt) and BAC clone 21F22 hybridized with *dd21.5* and *Psc2.6*. In all cases DNA was digested with HindIII. **A.** *dd21.5* is a single copy sequence in *M. truncatula* and completely located on BAC 21F22 (compare second and third lane of A). A low stringency hybridization and washing reveals a band of about 6.5 Kbp weakly hybridizing with *dd21.5* (first lane of A). This band is not present in BAC 21F22. **B.** *Psc2.6* hybridizes with several *Medicago* fragments, but none of them are present in BAC 21F22. All hybridizing bands in this figure are indicated by a black arrowhead.

Fig 5

A recent *M. truncatula* A17 BAC library screening revealed 5 BAC clones (77011, 77J4, 72K10, 75L5 and 62F17) that hybridize with both *dd21.5* and *Psc2.6*. It is probable that these BACs form a contig that links to BAC 21F22. These 5 BAC clones demonstrate that *M. truncatula* A17 contains a genomic region that is microsyntenic and orthologous to the pea linkage group I segment containing the tightly linked pea markers *dd21.5* and *Psc2.6*. These 5 BACs are presently being analyzed.

**Mapping by FISH of BAC clone 21F22 in *M. truncatula***

The location of BAC clone 21F22 was determined by FISH on *M. truncatula* A17 prometaphase chromosomes. To facilitate the identification of chromosomes, the preparations were also hybridized with the pericentromeric repeat *Mt*R1 (Kulikova et al., in press).

BAC 21F22 was labeled with a red fluorophore and *Mt*R1 with a green fluorophore. BAC 21F22 mapped on *Medicago* chromosome 4 (Fig 6). In addition, *Medicago* chromosome 5 has regions close to both telomeres that weakly hybridizes with BAC 21F22 (Fig 6).
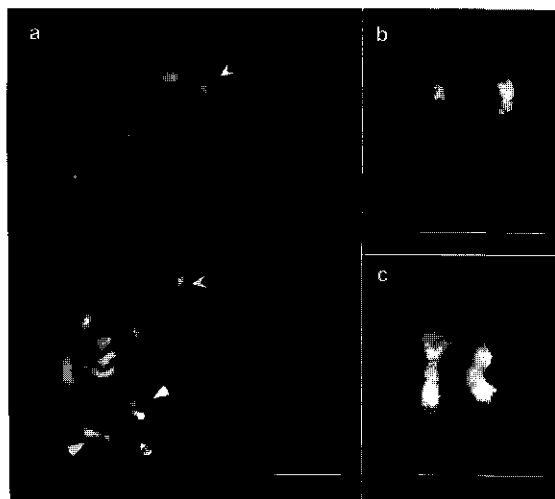
**Fig 6: a.** A complete complement of prometaphase chromosomes of *M. truncatula* hybridized with *Mt*R1(green) and BAC 21F22(red). Chromosomes 4 and 5 are indicated by an arrowhead and an arrow, respectively. **b.** Magnification of chromosomes 4. Chromosome 4 is the largest chromosome and contains a MtR1 region on its short arm.

A strong FISH signal shows the location of 21F22 on this chromosome. **c.** Magnifications of chromosome 5 showing the weak hybridization of BAC21F22 near the telomeres. Chromosome 5 is the only chromosome containing a secondary constriction and it lacks a region containing the MtR1 repeat.

## DISCUSSION

By using a series of RILs (Geurts, 1998) and introgression lines (Kozik et al., 1995), the RNA differential display resulted in the isolation of the cDNA RFLP markers *dd21.5* and *Psc2.6* (through *dd2.6*) that are tightly linked to each other and located between *PsEil2* and *Nod3* in pea linkage group I. Thus, *dd21.5* and *Psc2.6* become the closest RFLP markers to the *nod3* hypernodulation locus at this chromosomal side of the mutation that we refer to as to the "south" of *Nod3*. RFLP analysis revealed two recombinations between *dd21.5/Psc2.6* and *Nod3*. The distance between these two markers and *Nod3* can be estimated by comparison with the two recombinations that were also observed in pea between the c1 markers *Mtc831/Mtc923* and *Sym2* (Gualtieri et al., submitted, a). Thus, *dd21.5* and *Psc2.6* are probably as close to the *Nod3* locus, as *Mtc923* and *Mtc831* are from *Sym2*. Moreover, the size of the *Medicago Sym2*-orthologous region (delimitated by markers *Mtc831*, *Mtc923* and Mt63EB4) was determined to be of about 350 Kbp. Therefore the size of the *Medicago* region that would contain *Nod3*, *dd21.5* and *Psc2.6* would be within the same range. Although mapping of these two markers in a segregating population will be essential to determine more precisely their genetic distance to *Nod3*, the RFLP data presented here give a reliable idea of

the map position of these markers since plants that were generated independently (*Sym2*-containing introgression lines, and *nod3*-containing RILs) were used.

Northern blots demonstrated that *dd2.6* (included in the RFLP marker *Psc2.6*) was differentially displayed by an allelic PCR polymorphism between the Rondo transcript and the transcript found in the *nod3*-containing RILs, since no significant differences in the level of this transcript were observed between the plants tested. Consequently, this PCR polymorphism should predict the map position of the RFLP marker *Psc2.6* (that includes *dd2.6*). However, due to the complexity of the RFLP pattern of *Psc2.6*, it is not possible to establish a direct correlation between the differentially displayed *dd2.6* band and the *Nod3*-linked *Psc2.6[AB1]* band (revealed by Southern blot of DNA restricted with HindIII and hybridized with *Psc2.6*). In other words, it is difficult to assure whether *Psc2.6[AB1]* contains a part of *Psc2.6*, or whether this *Nod3*-linked band is simply homologous to the 5` part of *Psc2.6* that is not present in *dd2.6*. Although the first cannot be discarded, the latter option seems more probable due to the weak hybridization of *Psc2.6[AB1]* with *Psc2.6*, meaning that *Psc2.6* is probably entirely located within the strongly hybridizing non-polymorphic HindIII band that contains *dd2.6*. Thus, it is not known whether the map position of the band *Psc2.6[AB1]* is indeed representing the position of *dd2.6* (that is included in *Psc2.6*). However, the "relationship" and linkage to *Nod3*, of *dd2.6* and *Psc2.6[AB1]*, respectively, suggests that if *Psc2.6[AB1]* is not containing part of *Psc2.6* they should be located close to each other and to *Nod3* on pea linkage group I. This because it is very unlikely that the differential display of *dd2.6* and its relationship with *Nod3* would be a mere coincidence, since *dd2.6* has resulted in the isolation of the RFLP marker *Psc2.6* revealing the *Psc2.6[AB1]* band that is linked to *Nod3* . Thus, both the differential display and the RFLP analysis indicate that *dd2.6* (included in *Psc2.6*) and *Psc2.6[AB1]* map close to each other and to *Nod3*. The tight linkage of homologous sequences in pea and *Medicago* has been observed for several low copy sequences present at the pea *Sym2* and *Medicago Sym2*-orthologous regions, respectively (Gualtieri et al., submitted, a; Gualtieri et al., submitted, b). When *dd2.6* would indeed be-*as Psc2.6[AB1]*-tightly linked to *Nod3*, it is possible that the PCR polymorphism differentially displaying *dd2.6* is caused by an additional mutation in the region neighboring *nod3*.

In general, when the differential display of a cDNA is due to a polymorphism, this technique can be used to predict whether the cDNA maps at a given chromosomal region defined by the introgressed regions. However, such data should be confirmed by genetic mapping, because

the differential display can be due to differences in expression level of a transcript and not be caused by a sequence polymorphisms. This was illustrated by *dd21.5* that was differentially displayed as a result of a root-hair specific transcriptional repression effected by an Afghanistan introgressed region of A15.5.2. The reliability of the RNA differential display to show the expression pattern of *dd21.5* is evident from the fact that other fragments in the display profile have a similar concentration in the compared samples and therefore serves as internal controls. Moreover, in other studies it was shown that the RNA differential display is a semi-quantitative method (Holsters, pers. comm.). The cDNA-AFLP is also useful as a semi-quantitative method to directly study gene expression (Bachem et al., 1996).

The root hair specific repression of *dd21.5* could be caused by *Sym2*, or by any other gene from the *Sym2*-containing introgressed region, or by a gene present in another introgressed region that A15.5.2 might contain. The probability that A15.5.2 contains additional introgressed regions other than the *Sym2*-containing region is quite high. After the four backcrosses that generated A15.5.2, this line should contain in theory about 3% of Afghanistan DNA in its genome, what makes about $1.2 \times 10^8$ bp or 60 cM of the 2000 cM pea genetic map. Since the *Sym2*-containing introgressed region of A15.5.2 is about 4 cM it is probable that additional Afghanistan introgressed regions are present. Thus, that the root hair specific repression of *dd21.5* is mediated by a regulatory mechanism effected by the *Sym2*-containing introgressed region of A15.5.2, could be proven by analyzing the expression of this gene in several other independently generated plants, but selected for the presence of a *Sym2*-containing introgressed region similar to that of A15.5.2 (containing Afghanistan DNA from *Psc2.6[AB1]* to *T22*), such as A33.18. cDNA *dd21.5* maps in the Afghanistan-introgressed region of A15.5.2. Hence, in case the level of transcription of *dd21.5* is indeed effected by the *Sym2*-containing introgressed region, this would also demonstrate that the RNA differential display would be useful to detect a functional interaction between genes located within an introgressed genomic region.

The search for the *M. truncatula* A17 genomic region orthologous to the pea linkage group I segment containing *dd21.5*, resulted in the isolation of BAC clone 21F22, containing the *Medicago* orthologue Mtg21.5. Although *dd21.5* is completely present within BAC 21F22, this BAC does not contain the *Medicago Psc2.6*-orthologue, even though *dd21.5* and *Psc2.6* are tightly linked in pea linkage group I. However, recent BAC library screenings identified 5 BACs that co-hybridize with both *dd21.5* and *Psc2.6*, and since it is quite likely that these 5

BACs form a large contig linking to BAC 21F22, this is being presently analyzed. Thus, the finding of these 5 BACs, combined with the single copy nature of Mtg21.5 demonstrates that *M. truncatula* A17 has a genomic region that contains both the orthologues of *dd21.5* and *Psc2.6*. Therefore, this *Medicago* region is microsyntenic to the pea linkage group I segment containing the tightly linked markers *dd21.5* and *Psc2.6*. Further analysis of these 5 BACs and BAC 21F22 through the construction of contigs and the isolation of other RFLP markers in addition to the *dd21.5* and *Psc2.6* orthologues, is essential to obtain a more complete picture of the microsyntenic relationship of these two regions. This will ultimately reveal the level of microsynteny and orthology between this *Medicago* region and the pea linkage group I segment containing *dd21.5* and *Psc2.6* between *Nod3* and *Eil2*. When microsynteny extends along this *Medicago* region, chromosome walking can be used to extend this region towards and beyond the *Nod3*-orthologous locus, and molecular marker technology can be later used to delimitate the minimal *Medicago* orthologous region putatively containing the *Nod3*-orthologous locus. A similar strategy was already successfully applied for the subcloning and delimitation of the *Medicago Sym2*-orthologous region (Gualtieri et al., submitted, a).

Studies were initiated to map the *M. truncatula* A17 genomic region orhtologous to the pea region containing *dd21.5* and *Psc2.6*. BAC 21F22 (containing the single copy orthologous sequence Mtg21.5) was mapped by FISH on *Medicago* chromosome 4, reflecting a disturbance in "chromosomal" synteny between *Medicago* chromosome 5 and pea linkage group I that were otherwise syntenic at the "*Sym2* region" and at the "*Eil2* region". Thus, while other markers from pea linkage group I are syntenic with *Medicago* chromosome 5, 21F22 reveals that *dd21.5* is located on *Medicago* chromosome 4, indicating a translocation event in one of these two legume species. Whether this synteny with chromosome 4 extends beyond *dd21.5* and includes the *Nod3* locus is unknown, but could be analyzed by studying markers located to the "north" of *Nod3*. In addition to the strong FISH signal given by 21F22 on chromosome 4, 21F22 gives weak but reproducible FISH signals on chromosome 5 close to both telomeric regions of the two homologous chromosomes. One of these locations is similar to that of the c1/c2 *Sym2*-orthologous region and so it cannot be excluded that this sequence weakly hybridizing with BAC21F22 has a position syntenic with *dd21.5*. However it should be kept in mind that the resolution of FISH on metaphase chromosomes is low (Kulikova et al., in press). Thus, although BAC 21F22 is located on chromosome 4 it hybridizes with homologous regions on chromosome 5. Future experiments will be essential to test whether the five BACs that contain both the *dd21.5* and *Psc2.6* homologues map in

chromosome 4 and form a contig with 21F22, or whether they contain the *dd21.5*-distant homologue (revealed by low stringency Southern blot) and might map in chromosome 5 at the location where the *Medicago Nod3*-orthologue may map when synteny extends along *Medicago* chromosome 5 beyond the non-syntenic locus Mtg21.5.

## MATERIALS AND METHODS

### Plant material

The original parental lines of the RILs and introgression lines used in this study were the European Rondo cultivar (Lie, 1984), the wild type Afghanistan pea carrying *Sym2* (Lie, 1984), and the *nod3* hypernodulation EMS mutant of the Rondo European cultivar (Jacobsen et al., 1984). The Rondo-*Sym2* introgression lines A569 (Kozik et al., 1995) and A541-generated by an additional backcross on the introgression line A54 (Kozik *et al*, 1995), and A15.5.2 and A33.18 (Kozik, 1996) were used for RFLP mapping, in combination with a series of RILs constructed by crossing the *Sym2*-introgression line A54 with the *nod3* line (Geurts, 1998; published in Gualtieri et al., submitted, a). Rills D21.7 and D23.3 are hypernodulated but do not contain *Sym2*, whereas the RILs D24.7, D25.2, D31.2 and D9.9 are hypernodulated and contain *Sym2*.

The RNA differential display was carried out by using the European Rondo cultivar line, the *nod3* line, the A15.5.2 introgression line, and the RILs D25.2, D26.6 (containing both *nod3* and *Sym2*), and D22.2 and D41.6 (containing *nod3* but not *Sym2*).

### RNA differential display

The RNA differential display was done on roots and root hairs of 4 days old seedlings that were non-inoculated, or inoculated two days after germination with *Rhizobium leguminosarum* bv *viciae* 248.pMW1071 (containing the *nodX* gene), a strain that nodulates plants with and without *Sym2* (Kozik et al., 1995). Root hairs were harvested as previously described (Gloudemans et al., 1989). Total RNA was isolated according to Pawlowski *et al.* (1984), followed by a DNaseI (Promega) treatment. The RNA differential display was done according to Liang *et al.* (1992). Differential cDNAs were cut from the acryl amide gel and collected in eppendorf tubes. The fragments were incubated in 100 ul TE (pH9.0) for 10 min at room temperature and subsequently boiled for 20 min. The supernatant was filtered through a glass wool filter. The DNA was ethanol precipitated in the presence of 0.3M Na-acetate (pH5.2) and 50 ng glycogen, and subsequently re-dissolved in 10 μl of distilled water. This

DNA was used as a template in a PCR reaction using identical conditions as for the RNA differential display. The PCR reaction was run on a 1.5% agarose gel. DNA fragments were isolated from the gel using a gel extraction kit (MBI Fermentas) and cloned in the pGEM-T vector (Promega).

**Genomic DNA, BAC DNA, and total RNA blots**

Pea and *Medicago* DNA was isolated according to A. Kozik (Kozik, 1996) and D. Cook (pers. comm.), respectively. RNA was isolated according to Pawlowski *et al.* (1984). Bacterial Artificial Chromosome (BAC) DNA of *M. truncatula* A17 was isolated according to Nam *et al.* (1999). All hybridizations were done with Hybond-N+ membranes (Amersham) under standard conditions (Sambrook et al., 1989). DNA probes for Southern blots and northern blots were labeled with $[^{32}P]dATP$ using the random priming method (Feinberg and Fogelstein, 1983).

**cDNA and BAC library screenings**

The pea root hair cDNA library was constructed by Stratagene in Lambda ZAPII vector system, using equal amounts of poly (A)+ RNA isolated from a mixture of root hairs of 6 days old cv Finale pea plants, uninoculated and inoculated for 48 hours with *R. leguminosarum* bv *viciae* 248. The cDNA library screenings were done with Hybond-N+ membranes using the conditions recommended by the library manufacturers (Stratagene). The BAC library of *M. truncatula* genotype A17 was screened according to Nam *et al.* (1999), and high-density BAC library filters and BAC clones were obtained from the Clemson University Genomics Institute (http//www.genome.clemson.edu). DNA probes for cDNA and BAC library screenings were labeled as described above.

**DNA sequencing**

Automated sequencing was done on an ABI 377 DNA sequencer.

# CHAPTER 5

## Integration of the FISH-pachytene and genetic maps of *Medicago truncatula*

Olga Kulikova[1,2*], Gustavo Gualtieri[1], René Geurts[1], Dong-Jin Kim[3], Douglas Cook[3], Thierry Huguet[4], J. Hans de Jong[5], Paul F. Fransz[6] and Ton Bisseling[1]

1. Wageningen University, Department of Plant Sciences, Laboratory of Molecular Biology, the Netherlands
2. Laboratory of Biotechnology, Institute for Agricultural Microbiology, St.-Petersburg-Pushkin, Russia
3. Department of Plant Pathology, University of California, Davis, CA, USA
4. INRA-CNRS, UMR215, BP 27, 31326 Castanet-Tolosan Cedex, France
5. Wageningen University, Department of Plant Sciences, Laboratory of Genetics, the Netherlands
6. Swammerdam Institute for Life Sciences, University of Amsterdam, the Netherlands

For correspondence: Laboratory of Molecular Biology, Dreijenlaan 3, 6703 HA Wageningen, the Netherlands. Fax +31 (0)317 48 35 84; e-mail olga.kulikova@mac.mb.wau.nl

## SUMMARY

A molecular cytogenetic map of *Medicago truncatula* (2n=2x=16) was constructed on the basis of a pachytene DAPI karyogram. Chromosomes at this meiotic prophase stage are 20 times longer than at mitotic metaphase and display a well-differentiated pattern of brightly fluorescing heterochromatin segments. We describe here a pachytene karyogram in which all chromosomes can be identified based on chromosome length, centromere position, heterochromatin patterns and the positions of three repetitive sequences (5S rDNA, 45S rDNA and the *Mt*R1 tandem repeat), visualised by FISH (Fluorescence *in situ* Hybridization).

We determined the correlation between genetic linkage groups and chromosomes by FISH mapping of BAC (Bacterial Artificial Chromosome) clones, with two to five BACs per linkage group. In the cytogenetic map chromosomes were numbered according to their corresponding linkage groups. We determined the relative positions of the twenty BACs and 3 repetitive sequences on the pachytene chromosomes and compared the genetic and cytological distances between markers. The mapping resolution was determined in a euchromatic part of chromosome 5 by comparing the cytological distances between FISH signals of clones of a BAC contig with their corresponding physical distance and showed that resolution in this region is about 60 kb. The establishment of this FISH pachytene karyotype, with a far better mapping resolution and detection sensitivity compared to those in the highly condensed mitotic metaphase complements, has created the basis for the integration of molecular, genetic and cytogenetic maps in *M. truncatula*.

**Key words**: *Medicago truncatula* Jemalomg A17, Fluorescence *in situ* Hybridization, cytogenetic map, pachytene karyotype, linkage group

## INTRODUCTION

The *Fabaceae*, or legumes, comprise the third largest family of flowering plants and display a striking variety of plant types, ranging from small annual herbs to massive tropical trees. Among the legumes, the sub-family Papilionoideae contains the majority of agronomic species, including the galegoid species, such as *Pisum sativum* (pea) and *Medicago sativa* (alfalfa), and beans, such as *Phaseolus*, *Vigna* and *Glycine* spp. In addition to their agronomic importance as producers of oil and protein for the diets of humans and livestock, legumes have the distinction of being one of the major sources of biologically available nitrogen in the biosphere, which is caused by their unique ability to establish a symbiosis with several genera of bacteria that are collectively called rhizobia. This symbiotic interaction results in the

formation of root nodules where the bacteria are capable of reducing atmospheric nitrogen. Due to the narrow host range, none of these micro-symbionts can interact with model plants like *Arabidopsis* and rice, and hence legume species have been proposed as a model system for unravelling the molecular mechanisms that control the development and functioning of this beneficial plant-microbe interaction. *Medicago truncatula* Gaerth. (Barrel medic) has been selected for this purpose (Barker et al., 1990). This annual, autogamous diploid forage crop has 16 chromosomes and a relatively small genome (1.15 pg/2C, Blondon et al., 1994) of about the same size as that of rice and 4 times larger than that of *Arabidopsis*. In addition, molecular and genetic studies of *M. truncatula* became possible with the recently developed efficient transformation and regeneration procedures and the establishment of BAC libraries and genetic maps (Cook, 1999).

In order to position BACs and other DNA sequences along the chromosomes, molecular cytogenetic maps are needed. Among the tools for such maps, fluorescence *in situ* hybridization (FISH) is undoubtedly the most versatile and accurate one for ordering repetitive and single copy DNA sequences on the chromosomes with respect to centromeres, telomeres and heterochromatic regions (Jiang and Gill, 1994; De Jong et al., 1999). Cytogenetic maps can be highly informative to support the construction of physical maps, map-based cloning projects and to position genes in large heterochromatic regions where linkage distances are inaccurate by the low levels of meiotic recombination (Roberts, 1965; Lambdie and Roeder, 1986; Zhong et al., 1999).

Microscopic preparations with mitotic metaphase complements for FISH are mostly prepared from root tips. Chromosomes at this stage are highly condensed, which limits the optical resolution of adjacent FISH targets to Mbs rather than kbs. Where higher resolution is required, pachytene chromosomes are more appropriate, because their complements measure 10-40 x longer and display a differentiated pattern of heterochromatic and euchromatic regions (Shen et al., 1987; Albini and Schwarzacher, 1992; Zhong et al., 1996; Fransz et al., 1998; Fransz et al., 2000). This heterochromatin banding along with chromosome length and centromere position is important for the identification of individual chromosomes. However, the production of high-quality pachytene spreads in comparison to metaphase chromosomes is technically more demanding and the success for satisfactory results may differ between related species and even between genotypes of the same species (review by De Jong et al., 1999).

The published karyotypes of *M. truncatula* were based on simple morphometric measurements of metaphase plates from accession V357 (Agarwal and Gupta, 1983) and the accessions R-108-1 and Jemalong J5 (Gerbah et al., 1999). However, the morphology of metaphase chromosomes was found to be too similar to allow reliable identification of all

chromosomes on the basis of relative arm lengths and centromere positions. FISH of 5S and 45S rDNAs, which were applied as additional markers in the R-108-1 and Jemalong J5 accessions (Gerbah et al., 1999), showed that both genotypes have a single NOR (nucleolar organizing region), and that Jemalong J5 had three 5S rDNA loci, whereas R-108-1 has only two.

Pachytene karyotype of several species of genus *Medicago* were described including tetraploid and diploid *Medicago sativa* L. (Gillies, 1970; Gillies, 1968; Buss and Cleveland, 1968), and species closely related to *M. sativa* such as *M. falcata* L. (Gillies, 1970), *M. glomerata* Balb (Gillies, 1971), *M. coerulea* Less, *M. glandulosa* David, *M. glutinosa* M.B. and *M. prostrata Jacq.* (Gillies, 1972) . Chromosomes at the pachytene stage of microspoocytes were measured and characterized by average total chromosome length, arm length, arm length ratio and number, position and size of chromatic area.

This study is the first pachytene chromosome map of *M. truncatula* based on the Jemalong A17 genotype, which is the standard line in most current molecular genetic studies (Nam et al., 1999; Trieu et al., 2000; Kim et al., in preparation). The chromosomes were stained with DAPI and their fluorescence images were used for karyotype analysis on the basis of morphometric data and heterochromatin patterns. In addition, the pericentromeric repeat *MtR*1 and 5S rDNA were tested as probes in FISH experiments to facilitate chromosome identification. For assigning linkage groups to the cytogenetic map, we hybridized BAC clones positioned on the genetic map of *Medicago truncatula* (Kim et al., in preparation) to the pachytene complements and mapped their positions with respect to centromeres and heterochromatin segments.

## RESULTS

### The pachytene karyotype

Pollen mother cells at late pachytene clearly displayed eight fully paired bivalents with lengths varying from 29 to 68 μm and a total complement length of 406 μm. Chromosomes were submetacentric or metacentric, with centromere indexes between 27 and 47% (Fig. 1a; table 1). They were numbered according to their corresponding linkage maps as was decided during the 2nd *Medicago truncatula* Workshop (Amsterdam, The Netherlands, July 22-23, 1999), with their two arms denoted as S (short) and L (long), respectively. DAPI staining of pachytene chromosomes demonstrated striking differences in chromatin density. Brightly fluorescing heterochromatic blocks were detected in the pericentromeric regions of all chromosomes, although their lengths were different for each chromosome (Fig.1a, table 1).

The centromere itself appeared as a primary constriction, in which the chromatin fluoresces far weaker than the flanking heterochromatic regions (Fig.1a). Distal regions of the chromosome arms generally consist of weakly fluorescing euchromatin. In addition to the pericentromeric heterochromatin, smaller heterochromatic knobs were observed on the short arms of the chromosomes #3, #4 and #7, and both arms of chromosome #6, but their number was variable and could be observed in few cells only. The total length of all heterochromatic areas is 59.3 μm, which is about 15% of the length of the complement.
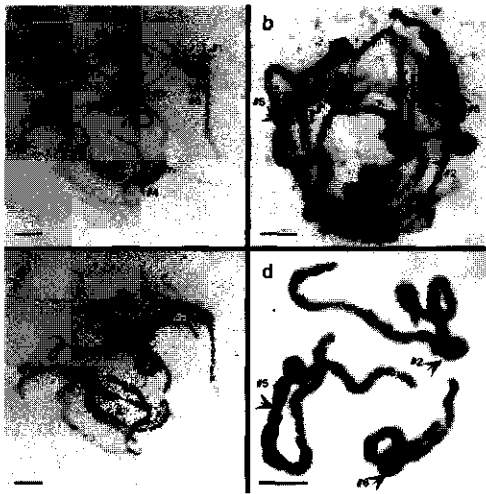


**Figure 1.** Pachytene chromosomes of *M. truncatula* Jemalong A17. (a) The complement of pachytene chromosomes. The DAPI stained chromosomes have brightly fluorescent heterochromatin around of centromeres (pericentromeric heterochromatin). Chromosomes are numbered according to corresponding linkage groups and indicated by arrowheads at centromere position. (b) Three chromosomes (#3, #5 and #6) give FISH signals with a 5S rDNA probe (red). The centromeres of the chromosomes containing a 5S rDNA region are indicated by arrowheads. Chromosome #5 carrying the major 5S rDNA cluster also contains the 45 rDNA region (green). (d) Three 5S rDNA-carrying chromosomes are dissected from the complement. (c) Hybridization with 5S rDNA (red) and *Mt*R1 (green) allows the identification of all eight pachytene bivalents (see a). Centromeres are indicated by arrowheads Bar = 5 μm in all figures.

**Table 1.** Absolute and relative lengths of individual chromosomes and chromosome regions, positions of 5S rDNA, NOR and *Mt*R1 on chromosomes. The chromosomes #3 and #4 can only be distinguished when FISH with *Mt*R1 was used as diagnostic maker.

| chromosome[1] | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | total |
|---|---|---|---|---|---|---|---|---|---|
| average length [2] | 60.3±6.2 | 49.3±4.1 | 68.1±5.9 | 66.0±4.3 | 49.5±5.2 | 29.2±3.8 | 50.4±5.1 | 33.4±5.4 | 406.2 |
| total cell complement [3] | 14.9 | 12.1 | 16.8 | 16.2 | 12.2 | 7.2 | 12.4 | 8.2 | 100% |
| centromere index [4] | 36.4 | 46.2 | 27.1 | 30.2 | 47.4 | 46.3 | 30.0 | 41.2 | |
| % heterochromatin | 15 | 14 | 8 | 8 | 21[5] | 25 | 15 | 22 | 14.6%[7] |
| FISH signals [6] | | | | | | | | | |
| 5S rDNA | - | L | - | - | S | S | - | - | |
| 45S rDNA | - | - | - | - | L | - | - | - | |
| *Mt*R1 | L | L | - | S | - | - | L + S | L | |

1) Chromosomes were ordered and numbered according to their corresponding linkage groups 2) Chromosome length in μm ± SD. 3) Total cell complement is percentage chromsome length / total length of all chromosomes 4) Centromere index is percentage of short arm / total chromosome length (Levan et al., 1964). 5) Value includes short arm heterochromatin + NOR. 6) Positions of the repeats on short arm (S) or long arm (L). 7) % heterochromatin in cell complement

Chromosome identification was based on their length, centromere position, size of pericentromeric heterochromatin and the presence of diagnostic heterochromatic knobs. Eight cells were selected in which all 8 bivalents could be discerned and fully traced along their length. Based on this morphometric characterisation the chromosomes could be distinguished as follows (Table 1):

- *The group of the three longest chromosomes* (#1, #3 and #4). These chromosomes measure 60 - 68 µm and have submedian centromere positions with centromere index (CI) values of 36%, 27% and 30%, respectively. Based on centromere position chromosome #1 can be distinguished from chromosomes #3 and #4. Chromosomes #3 and #4 have similar symmetric heterochromatic regions and differ only slightly in arm lengths and centromere positions. Distinction between them without additional diagnostic markers was therefore doubtful.

- *The group of the three medium-sized chromosomes* (#2, #5 and #7). These chromosomes measure about 50 µm. The former two have median centromere positions (centromere indexes of 47% and 46%, respectively), whereas chromosome #7 has a submedian centromere (CI = 30%). Furthermore, chromosome #5 has a characteristic pattern of four conspicuous pericentromeric heterochromatic knobs. In contrast, the chromosomes #2 and #7 have only one knob on each arm. Furthermore, chromosome #5 contains the secondary constriction (nucleolar organizer region). This weakly fluorescing region is located on the long arm, close to the centromere and is often clumped together with heterochromatic blocks of other chromosomes.

- *The group of the two smallest chromosomes* (#6 and #8). These chromosomes are 29 and 33 µm long, respectively and can easily be distinguished by differences in chromatic patterns. Chromosome #6 has several heterochromatic chromomeres on both arms, whereas chromosome #8 has two larger heterochromatic blocks on either side of the centromere.

The total length of the pachytene chromosomes is 406 µm, which is about 20x longer than that of the mitotic metaphase chromosomes (data not shown). The morphological features of the eight bivalents have been used to construct an ideogram (Fig. 2A). We selected the 5S rDNA, 45S rDNA and the *Mt*R1 tandem repeats as additional diagnostic markers to facilitate the identification of the individual chromosomes in the cell complement. FISH hybridization revealed that the 5S rDNA loci are located on the chromosomes #2, #5 and #6 (Fig. 1 b, d). A major 5S rDNA region occurs on the distal part of the pericentromeric heterochromatin of chromosome #5, on the arm containing a single heterochromatic knob. A second, smaller 5S rDNA region is located on the long arm of chromosome #2, close to the border of the
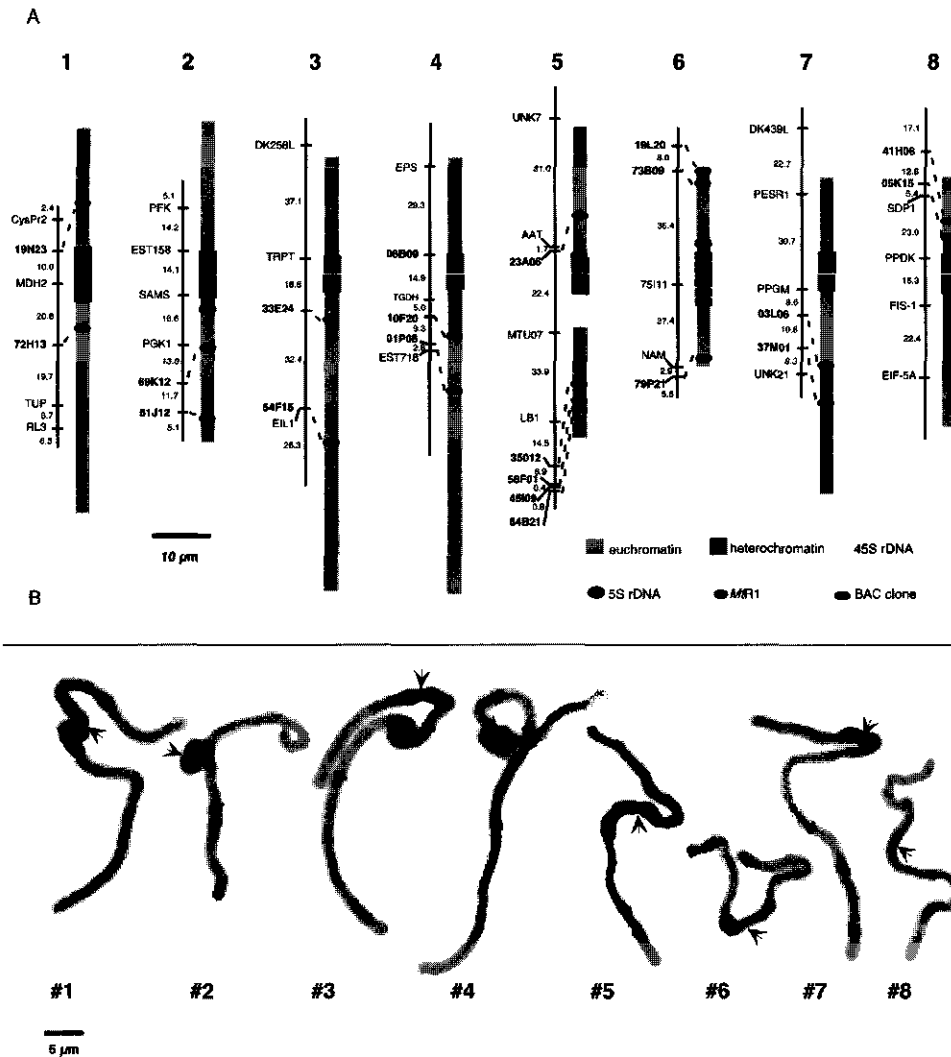
**Figure 2. Correlation between chromosomes and linkage groups. (A)** Ideogram of pachytene chromosomes and genetic linkage maps of *M. truncatula* Jemalong A17. BAC clones are positioned on ideogram according to their relative positions from centromeres. **(B)** Assignment of linkage group to pachytene chromosomes by FISH with linkage group specific BAC clones. Presented chromosomes are digitally sorted out of pachytene complements after hybridization with BAC clones, indicated in bold on ideogram. For the identification of chromosomes pachytene preparations were reprobed with 5S rDNA (red) and *Mt*R1 (green). Centromeres are indicated by arrowheads.

pericentromeric heterochromatin. A third 5S rDNA site is present on chromosome #6, 17% of the arm length distally from the centromere. FISH with the 45S rDNA probe demonstrated a bright spot on the secondary constriction of chromosome #5, between two proximal heterochromatic knobs (Fig. 1b, d). The same number of 5S and 45S rDNA loci were observed by FISH studies on metaphase chromosomes (Gerbah et al., 1999). The *Mt*R 1 tandem repeat has a 166 kb long motif and was identified in two randomly isolated BAC

clones, BAC75N01 and BAC53F10 (Table 2) and will be described elsewhere in more detail (Kulikova et al., in prep). This repeat is located in the pericentromeric regions of the chromosome arms #1-L, #2-L, #4-S, #7-S and #7–L, and #8-L (Fig. 1c). The *Mt*R1 signal on #8-S is weaker than that of the other *Mt*R1 signals. Thus *Mt*R1 is a good marker to distinguish chromosomes #3 and #4.

**Table 2.** Characteristics of BAC clones and DNA markers

| BAC clone | Linkage group | Genetic Marker | Marker template Accession | Marker type and homology of marker template hsp[1] | Distance (whole arm is 100%) from centromere ± SD |
|---|---|---|---|---|---|
| 75N01 | | | AQ841077 | BEST[2]; pericentromeric repeat *Mt*R1 | |
| 53F10 | | | AQ841071 | BEST "-----" | |
| 53F10 | | | AQ841072 | BEST "-----" | |
| 72H13 | 1 | DK049R | AQ841103 | BEST; putative beta-fructofuranosidase | 18.3 ± 1.2 |
| 19N23 | 1 | ENOD8 | n/a | BEST; BAC contains ENOD8 gene | 52.0± 2.5 |
| 69K12 | 2 | DK020R | AQ841084 | BEST; similar to putative Arabidopsis proteinase | 45.0± 2.2 |
| 51J12 | 2 | DK045R | AQ841099 | BEST; no known homology for marker template. | 90.1± 1.4 |
| 54F15 | 3 | DK123R | AQ841744 | BEST; homology to Arabidopsis hypothetical protein Z97335. BAC 54F15 survey sequencing reveals homology to multiple genes[3]. | 57.2± 1.1 |
| 33E24 | 3 | DK417L | AQ917383 | BEST; similar to NBS-LRR disease resistance protein (AB019186) RPR1 of *Oryza sativa*. | 16.2± 0.1 |
| 10F20 | 4 | DK043R | AQ841087 | BEST; no known homology for marker template. BAC 10F20 survey sequencing reveals homology to multiple genes. | 26.3± 1.0 |
| 01P05 | 4 | DK264L | AQ917083 | BEST; no known homology for marker template. BAC 01P05 contains putative MYB family transcription factor. | 43.7± 1.2 |
| 64B21 | 5 | EIL2-1 | n/a | BEST; BAC 64B21 is contiguous with BAC 42H09 from which survey sequencing reveals homology to multiple genes, including Arabidopsis ein3-like family. | 81.0± 1.0 |
| 23A06 | 5 | ENOD40 | n/a | PCR marker 3' of ENOD40 coding region. BAC 23A06 contains ENOD40 gene and putative receptor protein kinase. | 40.1± 1.6 |
| 35O12 | 5 | DK139L | AQ841733 | BEST; no known homology for marker template. BAC 35O12 survey sequencing reveals homology to multiple genes. | 64.2± 1.1 |
| 58F01 | 5 | DK006R | AQ841074 | BEST; no known homology for marker template. | |

| 45I09 | 5 | DK039R | AQ841114 | BEST; ; no known homology for marker template. BAC45I09 survey sequencing reveals homology to Mt ESTs | |
| 19L20 | 6 | DK125R | AQ841732 | BEST; similar to beta-transducin. BAC 19L20 contains Medicago truncatula cycloartenol synthase gene (Y15366.1). | 100.0± 0.0 |
| 73B09 | 6 | DK229L | AQ917196 | BEST; homology to tomato callus EST AW029689. | 83.1± 1.0 |
| 79P21 | 6 | 79P21R | n/a | BEST; BAC contains homology to LBS-LRR-TIR family of resistance genes. | 89.3± 2.1 |
| 37M01 | 7 | DK427R | AQ917398 | BEST; no known homology for marker template. | 57.3±1.1 |
| 03L06 | 7 | DK274L | AQ917096 | BEST; no known homology for marker template. | 44.8± 1.1 |
| 05K15 | 8 | DK505R | AQ917527 | BEST; similar to peptide transporter. BAC contains homology to Arabidopsis genomic DNA by tblastX. | 37.3± 0.1 |
| 41H08 | 8 | DK455L | AQ917442 | BEST; similar to Arabisopsis hypothetical protein. BAC 41H08 contains *Medicago truncatula* EST AW775698. | 53.1± 0.1 |

1. hsp, high-scoring sequence pair
2. BEST, BAC end sequence tag
3. Additional BAC survey sequence information is available by querying with BAC ID in "CloneID" field at http://chrysie.tamu.edu/medicago.

All together, as shown in Fig. 2A and Table 1, the karyotype analysis of the pachytene chromosome morphology and FISH patterns of the 5S rDNA, 45S rDNA, and *Mt*R1 repeats allows identification of all eight chromosomes.

## Integration of cytogenetic map and linkage groups

The numbering convention for the eight genetically identified linkage groups of *M. truncatula* was adopted from *M. sativa* (Kaló et al., 2000), as determined by comparative map analysis (Kim et al., unpublished data). For assigning individual linkage groups to the chromosomes, we selected BACs of each linkage group and used these as probes for FISH mapping to the pachytene chromosomes (Table 2). Concurrent hybridizations with the 5S rDNA and *Mt*R1 repeats were used to assist chromosome identification. Only FISH signals that occurred in at least 90% of the cell complements were considered for quantitative analysis. Examples of representative FISH patterns are shown in Fig. 2B. We measured the distance of the FISH signal in a relative scale from centromere to telomere in seven to ten cells (Table 2). Their distance values were averaged for drawing their positions on the cytogenetic map (Fig. 2A).

To provide a frame of reference to the genetic map of *M. truncatula*, five or more sequence-characterised genetic markers are indicated for each linkage group (Fig. 2A). These markers were developed on the basis of BAC and sequence information (see "Material and methods"). 2-5 BAC clones were used for mapping on pachytene chromosomes. Fig. 2A shows their genetic map position along with their corresponding position on the pachytene FISH map. Detailed information on every individual marker is given in Table 2 and at "http://chrysie.tamu. edu/perl-bin/mt_marker_query.2.pl".

**Table 3.** Comparison of genetic and cytogenetic distances between neighbouring BAC clone pairs

| Chromosome | BAC clones | Genetic distance (cM) | microscopic distance (μm) | cM /μm |
|---|---|---|---|---|
| 2 | 51J12- 69K12 | 11.7 | 12.7 | 0.9 |
| 3 | 33E24- 54F15 | 32.4 | 20.2 | 1.6 |
| 4 | 10F20- 01P05 | 9.3 | 8.2 | 1.1 |
| 5 | 64B21- 35O12 | 8.1 | 4.2 | 1.9 |
| 6 | 19L20- 73B09 | 8.0 | 2.6 | 3.1 |
| 7 | 37M01- 03L06 | 10.8 | 5.8 | 1.9 |
| 8 | 41H08- 05K15 | 12.6 | 4.2 | 3.0 |

As the distal parts of most chromosomes are euchromatic, it is expected that the genetic and FISH map distances be directly correlated. Table 3 gives a comparison of genetic map distances between six marker pairs and their corresponding microscopic distances on the pachytene FISH map. The ratio of genetic and cytogenetic distance values ranges from 1-2 cM/μm for most marker pairs. However, this value is markedly higher (3.0-3.1 cM/μm) for the pairs located on the small chromosomes #6 and #8.

## Resolution of FISH mapping

The BAC clones 58F01 and 59K07 from a BAC contig of chromosome 5 (Gualtieri et al., in prep.) were selected for estimating the resolution of FISH mapping in a euchromatic region of a pachytene chromosome 5 (Fig. 2A). The BACs were hybridized and detected with digoxigenin-FITC (green signal) and biotin-Texas Red (red signal), respectively (Fig. 3a, b).
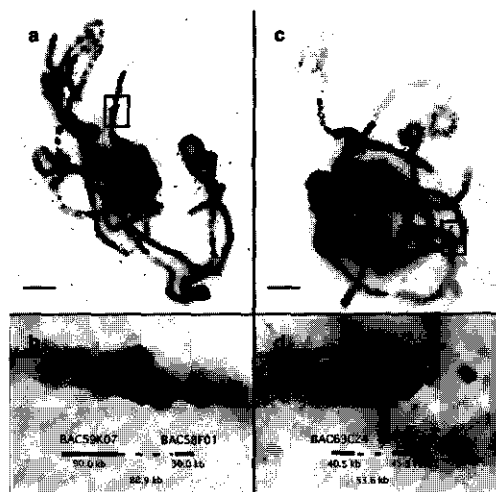
**Figure 3. Chromatin condensation degree in euchromatic region of chromosome #5. (a, c)** Two-color FISH with two pairs of BAC clones, BAC59K07 (red) and BAC58F01 (green), BAC63C24 (red) and BAC45I09 (green). **(b, d)** 10– fold magnified images *( indicated as insets in a, d)*. Yellow fluorescence indicates co-localization of green and red signals. White dash line indicates the distance between BAC pairs in their physical contigs.

A small band of yellow fluorescence in between the red and green spot represents the region of signal overlap. The midpoints of the sequences covered by these BACs are separated by 150 kb. This corresponds to a microscopic distance of about 0.5 µm between the centres of the red and green spots on the pachytene chromosome, implying a chromatin density of 300 kb/µm for that euchromatin region. With the spatial resolution limit of 0.2 µm for the fluorescence microscope, mapping resolution in this chromosome segment can be estimated at about 60 kb. This is confirmed by a comparable FISH experiment with the BAC clones 45I09 and 63C24 located in the same region of chromosome 5 (Fig.2A). These BACs are about the same size that is ~40 kb, and they are separated by about 55 kb (Fig. 4c, d). Fluorescence microscopic observation, revealed a prominent yellow spot, flanked by small green and red regions, thus confirming that mapping resolution of adjacent targets in this region is about 60 kb.

## DISCUSSION

We showed here that DAPI stained pachytene chromosomes of *M. truncatula* are very suitable to construct a detailed karyotype. The fully paired chromosomes at this meiotic prophase stage measure 406 µm (table 1), which is twenty times longer than their counterparts at mitotic metaphase and allows mapping with a high resolution (60 kb). In general, pachytene

chromosomes clearly display the differentiation of large heterochromatic blocks around centromere, whereas the distal parts of the arms are euchromatic. This pattern of well-defined heterochromatic areas is reminiscent of the conspicuous heterochromatin blocks on *Arabidopsis thaliana* pachytene chromosomes (Ross et al., 1996; Fransz et al., 1998, de Jong et al. 1999), but strongly differs from other small genome species. For example, rice exhibits numerous smaller heterochromatic knobs distributed along all chromosome arms (Khan, 1975). With this simple organization of solid heterochromatin blocks flanking the centromeres and long stretches of euchromatin in the distal areas, *M. truncatula* now becomes an attractive model species for cytogenetic analyses.

The combination of chromosome length, centromere position and heterochromatin patterns of the DAPI stained pachytene complements proved to be sufficient to identify all chromosomes except #3 and #4. However, with the *Mt*R1 repeat it became possible to distinguish these two chromosomes as well. In general, chromosomes were hybridized with 5S rDNA and *Mt*R1 to facilitate the recognition of the chromosomes, even when the chromosomes were clustered or partly overlapping.

Our second goal was to assign the eight chromosomes to the linkage groups of *M. truncatula*, which was recently established for this species (Kiss, Huguet, Kim and Cook, unpublished data). We selected two to five markers for each linkage group and mapped the corresponding BACs as probes in FISH experiments on pachytene cells. Markers belonging to a certain linkage group always hybridized to the same chromosome.

Heterochromatic regions are supposed to be rich in repeated sequences and to contain a low density of expressed genes (Dean and Schmidt, 1995). Euchromatin contains markedly fewer repeated sequences and has a higher density of genes. The repetitive nature of heterochromatic regions is consistent with the location of *Mt*R1 and the highly repeated ribosomal genes in heterochromatic pericentromeric regions. Furthermore, most of the linkage group specific BACs gave distinct FISH signals in euchromatic parts of the chromosomes and no background labelling was observed, suggesting that the interstitial segments of *M. truncatula* chromosomes contain a relatively low number of repeated sequences. In contrast, BAC paintings on pachytene chromosomes of maize, in which transcribed genes are separated by areas of repetitive DNA sequences (Benetzen et al., 1994; SanMiguel et al., 1996), blocking with $C_0t$-100 fraction of genomic DNA was required to suppress signals generated by these repeated sequences (Sadder et al. 2000). Results from whole BAC clone sequencing have indicated a gene density of approximately 1 predicted gene/6 kb in gene-rich regions of *M. truncatula* (Kim and Cook, unpublished data), which is consistent with the expectation that the arms of *M.truncatula* are rich in transcribed genes and that repeat sequences will be

underrepresented in the intergenic regions. This implies that positional cloning strategies for genes located in the euchromatic regions will not be hampered by a high density of repeated sequences such as might be expected in heterochromatic parts of chromosomes.

The degree of chromatin condensation in a euchromatic part of chromosome 5 was shown to be about 300 kb per μm, similar to that of Arabidopsis euchromatin which varies between 150 - 300 kb/μm (Fransz et al., 1998; Fransz et al., 2000). Assuming that the average degree of condensation in the euchromatic regions of *M. truncatula* is about 300 kb/μm, one can estimate the fraction of the *M. truncatula* genome contained within euchromatic and heterochromatic regions, respectively. As shown in Table 1, the total length of pachytene chromosomes is 406 μm, of which about 350 μm is euchromatic. Thus, the total euchromatic fraction of *M. truncatula* DNA is estimated at 105 Mb (i. e 300kb/μm x 350 μm). Since the genome size is about 500 Mb, 395 Mb, or almost 80% of the *M. truncatula* genome, is estimated to occupy heterochromatic regions. These values, of course, are the result of a rough calculation, since the degree of condensation within euchromatin has only been determined in one region. Nevertheless, it strongly indicates that the majority of the genome is located in heterochromatic region. Since this part of the genome is clustered in the pericentromeric regions which are predicted to contain few transcribed genes, it will be attractive and possible to focus a future *M. truncatula* genome sequencing program on the euchromatic parts of the chromosome arms.

The detailed pachytene karyotype along with 20 BACs, which could be positioned in the euchromatic region of the long or short arms of the chromosomes (Fig. 2), now provides the basis for a high resolution cytogenetic map and will be the first step in the integration of the physical, chromosomal and genetic maps. Once sufficient markers are mapped to cover all chromosome regions, an informative cytogenetic map will provide an indispensable tool for map based cloning studies where reliable estimation of the physical lengths between adjacent markers and their precise position with respect to centromere, telomere and heterochromatin areas is required. Furthermore, it can reveal interesting cytogenetic properties. For example our studies indicate that the small chromosomes #6 and #8 display a higher cM/μm ratio than the other chromosomes. A more detailed comparison of the genetic and cytogenetic map is therefore essential to show whether this difference holds true for the other euchromatic regions of these small chromosomes, and such comparison can provide the basis to study differential chromosomal behaviour.

## EXPERIMENTAL PROCEDURES

### Preparation of meiotic pachytene chromosomes

The method we developed for the preparation of *Medicago* pachytene spreads is adapted from the *Arabidopsis* procedure described by Ross et al., 1996. Immature flower buds of 1.5 – 1.8 mm in length were directly fixed in ethanol/acetic acid (3: 1) for at least 3 h and can be stored in this fixative at -20°C for several months. For cell wall digestion we rinsed the buds 3 times for 1 min in deionised water and transferred them to a pectolytic enzyme mixture [0.3% (w/v) pectolyase Y23 (Sigma), 0.3% (w/v) cytohelicase (Sepracor) and 0.3% (w/v) cellulase RS (Sigma) in citrate buffer (10 mM sodium citrate buffer, pH 4.5)] at 37 °C for 2 h. The vulnerable material was rinsed again with deionised water and each flower bud was transferred to a droplet of water on a microscope slide. Anthers were dissected from flower buds with fine needles and transferred to a grease-free slide. The resulting cell suspension was spread on a clean glass slide with 30 µl of 60% acetic acid at 45 °C for 1 min. Finally 1 ml of ice-cold ethanol/acetic acid (3: 1) was added in a circle around the suspension before leaving the slides to dry.

### Probe preparation and labelling

Clone pCT4.2, which contains a 5S ribosomal DNA repeat unit (~500 bp) of *Arabidopsis thaliana* in pBS (Campell et al., 1992), was PCR-labelled with biotin-16-dUTP (Boehringer Mannheim) according to the instructions of the manufacturer. Clone pTA71, which contains a 9.1 kb fragment of 18S-5.8S-26S rDNA of common wheat (Gerlach and Bedbrook, 1979), was labelled with digoxigenin-11-dUTP using the high primed labelling kit (Boehringer Mannheim). The *Mt*R1 repeat was identified in the randomly isolated BAC 75N01 (20 kb insert) and BAC 53F10 (30 kb) clones of *M. truncatula* (Table 2). The ends of these BAC clones are composed of a series of 166 bp long direct repeats, named *Mt*R1 (Fig. 1). Two oligonucleotide primers (5' AAAAATTCGAATGCACCAAAACTGG 3' and 5' TCAGGATCTCATGAAACTGCTCTTTT 3') were used to amplify a 307 bp fragment by PCR with DNA of BAC75N01. This fragment was subcloned in the pGEM-T (Promega). Clone p*Mt*R1 containing a 307 bp fragment of the 166 bp *Mt*R1 repeat motif was labelled by PCR with digoxigenin-11-dUTP (Boehringer Mannheim).

### BAC clone isolation and manipulation

Bacterial Artificial Chromosome (BAC) clones of *M. truncatula* genotype Jemalong A17 were identified either by means of hybridization to high density filter arrays obtained from the

Clemson University Genomics Institute (http://www.genome.clemson.edu), or by PCR-screening of a multiplexed DNA copy of the *M. truncatula* BAC library as described by Nam et al. (1999). BAC DNA was isolated according to the alkaline lysate method (Woo et al., 1994) and labelled with either biotin-dUTP or digoxigenin-dUTP using the nick translation mix (Boehringer Mannheim) for FISH. BAC end sequencing was performed on whole BAC clones using primers complementary to the pBeloBAC11 vector ("left primer": AACGCCAGGGTTTTCCCAGTCACGACG; "right primer": ACACAGGAAACAGCTATG ACCATGATTACG). The low pass survey sequence reported in Table 2 and at "http://chrysie.tamu.edu/medicago" was obtained by sequencing of fragmented sub-libraries from individual BAC clones. Briefly, BAC DNA was sheared to a range of 1 to 3 kb and subcloned into the *Sma*I site of pUC18. The template DNA for sequencing was obtained either by PCR using universal primers, or by isolation of plasmid DNA. Plasmid DNA was sequenced using a universal primer directed against the pUC18 poly-linker (i.e. CAGGAAACAGCTATGACCATGATTACGA).

## Fluorescence in situ hybridization (FISH)

FISH was performed as described in detail by Zhong et al. (1996) without pepsin treatment. For hybridizations with the BAC10F20 and BAC63C24 clones the addition of competitor DNA was required. We added a 100-fold excess of fragmented genomic DNA to the hybridization mixture, before denaturation at 80 °C for 10 min and pre-annealing at 37 °C for 1 h. After this treatment the mixture was applied to slides (Jiang et al., 1995). Genomic DNA was isolated from seedlings of *M. truncatula* according to the CTAB DNA extraction method (Rogers and Bendish, 1988) and subsequently fragmented by autoclaving at 15 lb/cm$^2$ for 5 min. Biotin-labelled probes were detected with Avidin - Texas Red and amplified with biotin-conjugated goat-anti-Avidin and Avidin-Texas Red. Digoxigenin-labelled probes were detected with sheep-anti-digoxigenin-fluorescein (FITC) and amplified with rabbit-anti-sheep-FITC. Double fluorescence detection of probes was performed according to Fransz et al. (1996). Chromosomes were counterstained with DAPI (4',6-diamidino-2-phenylindole) in Vectashield anti-fade solution (Vector Laboratories), 5 μg/ml. Some chromosome preparations were re-used for FISH with a new set of probes according to the method of Heslop-Harrison et al. (1992). Chromosome preparations were studied and photographed under a Zeiss Axioskop fluorescence microscope equipped with separate excitation filter sets for DAPI (01), FITC (09) and Texas Red (14). FISH signals with different colours were recorded on a single photograph by double exposure. The colour negatives were scanned at

1000 dpi and their digital images were optimised for contrast and brightness using Adobe Photoshop 5.0.2 (Adobe Inc.). To separate individual bivalents, each bivalent was digitally excised and copied into a new image. Chromosomes were measured with "MicroMeasure", a freeware software programme from Colorado State University (http://www.colostate.edu/Depts/Biology/MicroMeasure). Chromosome nomenclature was according to Levan et al. (1964).

## Genetic Mapping

Genetic markers were developed on the basis of BAC end sequence information (a so-called BAC End Sequence Tag or "BEST"), obtained from BAC clones listed in Table 2. Briefly, oligonucleotide primers were designed based on the corresponding BAC end sequence information, and used to PCR amplify and sequence genomic DNA from the *M. truncatula* A17 and A20 genotypes (Penmtesa and Cook, 2000). Restriction enzyme polymorphisms were identified by comparing sequence differences between the parental genotypes against known restriction sites. The resulting CAPS (cleaved amplified polymorphic sequences) or length polymorphism markers were mapped on a population of 93 F2 progeny from a genotype A17 X A20 cross (Kim et al., in preparation). Polymorphic DNAs were resolved on a 1.5% agarose gel and visualised by ethidium bromide staining. Primers, PCR conditions, and restriction enzyme information is given at "http://chrysie.tamu.edu/perl-bin/mt_marker_query.2.pl". DNA was extracted from the mapping population by means of the Nucleon Phytopure kit (Amersham Life Sciences, Inc.), according to manufacturer's instructions.

## ACKNOWLEDGEMENTS

## REFERENCES

Agarwal K. and Gupta P.K. (1983) Cytological studies in the genus *Medicago* Linn. *Cytologia*, **48**, 781-793.

Albini S.M. and Schwarzacher T. (1992) *In situ* localisation of two repetitive DNA sequences to surface-spread pachytene chromosomes of rye. *Genome* **35**, 551-559.

Barker D., Bianchi S., Blondon F., Dattee Y., Duc G., Essad S., Flament P., Gallusci Ph., Genier G., Guy P., Muel X., Tourneur J., Denarie J. and Huguet T. (1990) *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*- legume symbiosis. *Plant Mol Biol Rep.* **8**, 40-49.

Bennetzen J.L., Schrick K., Springer P.S., Brown W.E., SanMiguel P. (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome* **37**, 565-576.

Blondon F., Marie D., Brown S. and Kondorosi A. (1994) Genome size and base composition in *Medicago sativa* and *M. truncatula* species. *Genome* **37**, 264-275.

Buss G.R. and Cleveland R.W. (1968) Pachytene chromosomes of diploid *Medicago sativa* L. *Crop Science* **8**, 744-747.

Campell, B.R., Soung, Y., Posch, T.E., Cullis, C.A. and Town, C.D. (1992) Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. *Gene* **112**, 225-228.

Cook D. (1999*) Medicago truncatula* – a model in the making! *Current Opinion in Plant Biology* **2**, 301-304.

Dean C. and Schmidt R. (1995) Plant genome: a current molecular description. *Annu Rev Plant Physiol Plant Mol Biol.* **46**, 395-418.

De Jong H., Fransz P. and Zabel P. (1999) High resolution FISH in plants – techniques and applications. *Trends Plants Sci.* **4**, 258-263.

Fransz P., Stam M., Montijn B., Ten Hoopen R., Wiegant J., Kooter J.M., Oud O. and Nanninga N. (1996) Detection of single copy genes and chromosome rearrangement in *Petunia hybrida* by fluorescence *in situ* hybridisation. *Plant J.* **9**, 767-774.

Fransz P., Armstrong S., Alonso-Blanco C., Fischer T.C., Torrez-Ruiz R.A. and Jones G. (1998) Cytogenetics for the model system *Arabidopsis thaliana*. *The Plant J.* **13**, 867-876.

Fransz P., Armstrong S., De Jong J.H., Parnell L.D., van Drunen C., Dean C., Zabel P., Bisseling T. and Jones G.H. (2000) Integrated cytogenetic map of the short arm of chromosome 4 of *Arabidopsis thaliana*: structural organization of a heterochromatic knob and pericentromeric heterochromatin. *Cell* **100**, 367-376.

Gerbah M., Kevei Z., Siljak-Yakovlev S., Kondorosi E., Kondorosi A. and Trinh T.H. (1999) FISH chromosome mapping allowing karyotype analysis in *Medicago truncatula* lines Jemalong J5 and R-108-1. *Molecular Plant Microbe Interactions* **12**, 947-950.

Gerlach W.L., Bedbrook J.R. (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucl Acids Res.* **7**, 1869-1885.

Gillies C.B. (1968) The pachytene chromosomes of a diploid *Medicago sativa*. *Can. J. Genet. Cytol.* **10**, 788-793.

Gillies C.B. (1970) Alfalfa chromosomes. I. Pachytene karyotype of a diploid *Medicago falcata* L. and its relationship to *M. sativa* L. *Crop Science* **10**, 169-171.

Gillies C.B. (1970) Alfalfa chromosomes. II. Pachytene karyotype of a tetraploid *Medicago falcata* L. *Crop Science* **10**, 172-175.

Gillies C.B. (1971) Alfalfa chromosomes. III. *Medicago glomerata* Balb. pachytene karyotype. *Crop Science* **11**, 463-464.

Gillies C.B. (1972) Pachytene chromosomes of perennial *Medicago* species. I. Species closely related to *M. sativa*. *Hereditas* **72**, 277-288.

Heslop-Harrison J.S., Harrison G.E. and Leitch I.J. (1992) Reprobing of DNA: DNA *in situ* hybridization preparations. *Trends Genet.* **8**, 372-373.

Jiang J.M. and Gill B.S. (1994) Nonisotopic *in situ* hybridization and plant genome mapping: the first 10 years. *Genome* **37**, 717-725.

Jiang J., Gill B. S., Wang G-L, Ronald P.C. and Ward D.C. (1995) Metaphase and interphase fluorescence *in situ* hybridization of the rice genome with bacterial artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **92**, 4487-4491.

Kaló P., Endre G., Zimányi L., Csanádi G. and Kiss G.B. (2000) Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*). *Theor. Appl. Genet.* **100**, 641-657.

Khan S.H. (1975) A technique for staining rice chromosomes. *Cytologia* **40**, 595-598.

Lambie E.I., Roeder G.S. (1986) Repression of meiotic crossings over by a centromere (Cen30) in *Saccharomyces cerevisiae*. *Genetics* **114**, 768-769.

Levan A., Fredga K. and Sandberg A. (1964) Nomenclature for centromeric position on chromosomes. *Hereditas* **52**, 201-220.

Nam Y-W., Penmetsa R.V., Endre G., Uribe P., Kim D. and Cook, D.R. (1999) Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene-response genes. *Theor Appl Genet.* **98**, 638-646.

Roberts P.A. (1965) Difference in the behaviour of eu- and heterochromatin: crossing over. *Nature* **205**, 725-726.

Rogers S.O. and Bendish A.J. (1988) Extraction of DNA from plant tissues. In: *Plant Molecular Manual* (eds. S.B. Gelvin and R.A. Schilperoort). Dordrecht, Kluwer Academic publishers **A6**, 1- 10.

Ross K., Fransz P.F., and Jones G.H. (1996) A light microscopic atlas of meiosis in *Arabidopsis thaliana*. *Chromosome Res.* **4**, 507-516.

Sadder M.T., Ponelies N., Born U., and Weber G. (2000) Physical localization of single-copy sequences on pachytene chromosomes in maize (*Zea mays* L.) by chromosome *in situ* suppression hybridization. *Genome* (in press).

SanMiguel P.J., Tikhonov A., Jin Y-K, Mochoulskaia N., Zakharov D., Melake-Berhan A., Springer P. S., Edwards K.J., Lee M., Avramova Z. and Bennetzen J. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274,** 765-768.

Shen D.L., Wang Z.F. and Wu M. (1987). Gene mapping on maize pachytene chromosomes by *in situ* hybridization. *Chromosoma* 95, 311-314.

Trieu A.T., Burleigh S.H., Kardailsky I.V., Maldonado-Mendoza I.E., Versaw W.K., Blaylock L.A., Shin H., Chiou T-J., Katagi H., Dewbre G.R., Weigel D. and Harrison M.J. (2000) Transformation of *Medicago truncatula via* infiltration of seedlings or flowering plants with *Agrobacterium. The Plant J.* **22,** 531-542.

Zhong X-B., Fransz P.F., Wennekes-van Eden J., Zabel P., van Kammen A., de Jong J.H. (1996) High resolution mapping by fluorescence in situ hybridization to pachytene chromosomes and extended DNA fibres. *Plant Mol Biol Rep.* **14,** 232-242.

Zhong X-B., Bodeau J., Fransz P.F., Williamson V.M., van Kammen A., de Jong J.H. and Zabel P. (1999) FISH to meiotic pachytene chromosomes of tomato locates the root-knot nematode resistance gene *Mi-1* and the acid phosphatase gene *Aps-1* near the junction of euchromatin and pericentromeric heterochromatin of chromosome arms 6S and 6L, respectively. *Theor Appl Genet.* **98,** 365-370.

# CHAPTER 6

## CONCLUDING REMARKS

One of the most interesting questions of plant biology and comparative genomics is whether the ability of legume and actinorhizal plants to establish a nodular symbioses, is given by unique properties that can be observed at the genome level. Addressing this question requires the comparison of genomes of nodule-forming plants with the genomes of other non-nodule-forming plants, in order to test whether symbioses-specific genes and chromosomal segments do exists. These studies will promote not only the identification of such genes, but also the synteny-based positional cloning of these and other genes which phenotypes are described in species with a complex genome. With the latter aim in mind, comparative genomic studies on the legumes *Pisum sativum* and *Medicago truncatula* were initiated in this thesis. These studies were focused on genomic regions containing two pea loci, *Sym2* and *Nod3*, involved in the control of Nod factor-structure dependent infection and autoregulation of nodule number, respectively. It is expected that the results presented here will soon lead to the cloning of these two loci. Moreover, these results set the basis for the comparison with other nodule-forming and non-nodule-forming plants that could result in the discovery of novel genes and gene clusters, if any, that could be unique for the nodulation process.

Chapter 2 describes structural comparative genomic studies between two plants belonging to the Leguminosae: *Pisum sativum* bv *viciae* (pea) and *Medicago truncatula* A17 line. This

chapter represents the beginning of a microsynteny-based positional cloning approach for the pea *Sym2* gene. Previous positional cloning approaches were done directly in pea (Kozik, 1996; Geurts, 1998). In this thesis (Chapter 2), a marker which is tightly linked to *Sym2* (Geurts, 1998), was used to screen a *Medicago truncatula* A17 BAC library and three physically unlinked contigs (named c1, c2 and c3) were constructed and extended by chromosome walking to a final total size of about 600 Kbp of the A17 genome. RFLP markers were isolated from these contigs and mapped in pea by using a series of RILs and introgression lines. Through the isolation of 8 RFLP markers it was demonstrated that c1 and c2 are highly microsyntenic with the pea "*Sym2* (containing) region", and that therefore c1/c2 represents the *M. truncatula Sym2*-orthologous genomic region. Furthermore, three of these RFLP markers could be used to map recombinations in pea at both sides around the *Sym2* locus, delimitating the *Sym2*-orthologous region to about 350 Kbp of the c1/c2 sequence. The use of *Medicago truncatula* as intergenomic positional-cloning vehicle of pea genes located at microsyntenic regions is demonstrated by the isolation of the cDNA clone *PscLRR52* from a pea root hair library. *PscLRR52* encodes for a protein with a LRR motif that is highly homologous to the LRR motif of the Cf4 and Cf9 disease resistance proteins [*Licopersicon esculentum*]. RFLP mapping revealed that the pea *Sym2* region is rich in *Cf4* and *Cf9* [*Licopersicon esculentum*] homologous sequences (including the LRR motives), as the *Medicago* c1/c2 region is.

In Chapter 3, the distribution of several sequences is studied in c1/c2 and c3, through the construction of detailed contig molecular maps. It is shown that the general sequence composition of c1/c2 is different from c3, although c3 contains a sequence that is highly homologous to *PscW62-1*. This confirms that c1/c2 is the *Sym2*-orthologus region. However, Mtg3552 and *Mtc411* from c1/c2, and the c2 genomic subclone Mtg2.4, like *PscW62-1*, hybridize with sequences present in both c1/2 and c3. Moreover, it was found that Mtg3552 and Mtg2.4, and *PscW62-1* and *Mtc411*, are physically clustered in c1/c2 and c3, and that these clusters have a similar order. Thus, the clustering and linearity of these 4 markers indicates that c1/c2 and c3 arose as a consequence of a genome duplication, rather than by independent evolution. Similar to this, chromosomal segment duplications were described in the Poaceae and the Brassicaceae (Ahn et al., 1993; Chen et al., 1997; Conner et al., 1998; Grant et al., 2000; Helentjaris et al., 1998; Kishimoto et al., 1994, Ku et al, 2000; Langercrantz, 1998, Nagamura et al., 1995). *Arabidopsis* disease resistance genes encoding LRR-containing proteins arose by numerous duplications of LRR encoding segments as well

as intragenic unequal recombination events (Noel et al., 1999). It is possible that similar mechanisms have been involved in generating the various LRR sequences in c2.

In Chapter 4, the RNA differential display resulted in the isolation of the pea RFLP markers *dd21.5* and *Psc2.6* (as revealed by *Psc2.6[AB1]*) that are tightly linked and separated by two recombinations (in the used set of RILs) from the *nod3* hypernodulation locus that maps 2 cM above *Sym2*. These recombinations map between *Sym2* and *Nod3*, a position referred as to the "south" of *Nod3*. Although the distance between these two markers and *Nod3* has not been determined by mapping using a large F2 segregating population, an estimation can be made since two recombinations have occurred between these two markers and *Nod3*, and similarly two recombinations (within the same set of RILs) are observed between the c1 markers *Mtc831* and *Mtc923*, and *Sym2*. Thus, *dd21.5* and *Psc2.6* are probably as close to the *Nod3* locus, as *Mtc923* and *Mtc831* are from *Sym2*. The *Medicago Sym2*-orthologous region was determined (Chapter 2) to be about 350 c1/c2 Kbp, indicating that the size of the *Medicago* region that could contain *Nod3*, *dd21.5* and *Psc2.6* would be similar. Therefore the identification and construction of the *M. truncatula* A17 genomic region containing the orthologues of *dd21.5* and *Psc2.6* was initiated. This region might contain the *Medicago Nod3* orthologue. By screening a *M. truncatula* A17 BAC library with marker *dd21.5*, BAC clone 21F22 was isolated. BAC 21F22 fully contains the *Medicago dd21.5*-orthologue single copy sequence Mtg21.5, but it does not contain the *Psc2.6* orthologue. However, five additional BAC clones were identified that co-hybridize both with *dd21.5* and *Psc2.6*, and since Mtg21.5 is single copy in *Medicago*, it is likely that these five BACs form a contig with 21F22. Alternatively, these five BACs could be not linking to BAC 21F22, but contain the *dd21.5*-distant homologous sequence that was identified by low stringency Southern blotting. In any case, the simultaneous presence of the *dd21.5* and *Psc2.6* hybridizing sequences on these five BACs, demonstrates that *Medicago* contains genomic regions microsyntenic with the pea linkage group I region containing *dd21.5* and *Psc2.6*.

Synteny is often disrupted at both the chromosome and megabase/submegabase levels. Even for very closely related species, synteny is not absolutely conserved (e.g. sorghum and maize). This is caused by events like inversions, duplications, translocations and deletions. Chapter 4 describes a translocation either in *Medicago* or in pea. When BAC 21F22 (containing the single copy orthologous sequence Mtg21.5) was studied by FISH it mapped on *Medicago* chromosome 4. These reflects a change in synteny between *Medicago* chromosome 5 and pea

linkage group I that were syntenic at the chromosomal level (e.g. at the *Sym2* and *Eil2* loci) and microsyntenic at the *Sym2* region. In contrast with other markers from pea linkage group I that were syntenic with *Medicago* chromosome 5, 21F22 reveals that *dd21.5* is syntenic with *Medicago* chromosome 4. Whether this synteny with chromosome 4 extends beyond *dd21.5* and includes the *Nod3* locus is unknown, but could be analyzed by studying markers located to the "north" of *Nod3*. Apart from the strong FISH signal given by 21F22 on chromosome 4, 21F22 gives weak reproducible FISH signals on chromosome 5, close to both telomeric regions on the two homologous chromosomes. In one case, this location is similar to that of the c1/c2 *Sym2*-orthologous region, and so it cannot be excluded that this chromosome 5 sequence weakly hybridizing with BAC21F22 has a position "syntenic" with the pea *dd21.5* region. This could be well analyzed by FISH of a BAC from the c1/c2 *Sym2*-orthologous region in combination with BAC21F22. Thus, BAC 21F22, located on chromosome 4, hybridizes with homologous regions on chromosome 5. Future experiments will be essential to test whether the five BACs that contain both the *dd21.5* and *Psc2.6* orthologues map in chromosome 4 and form a contig with 21F22, or whether they contain the *dd21.5*-distant homologue and might map in chromosome 5 at the location where the *Medicago Nod3* orthologue may map when synteny exists with *Medicago* chromosome 5 beyond the non-syntenic Mtg21.5 locus.

The studies presented in this thesis demonstrate that microsynteny exists between two species belonging to the Leguminosae. Thus, one of the main questions addressed within this thesis, whether legumes are microsyntenic, is answered for two species belonging to two different but related tribes of the Leguminosae: *Pisum sativum* and *Medicago truncatula*, of the Vicieae and Trifolieae tribes, respectively. Microsynteny in plants was extensively studied within the Poaceae (grasses) and the Brassicaceae (e.g. *Arabidopsis*, *Brassica*). In grasses, long-range as well as local high colinearity were reported, at the whole-genome, and at the megabase / submegabase levels, respectively (Bennetzen et al. 1997, Foote et al. 1997, Gale et al. 1998, Hulbert et al. 1990, Dunford et al. 1995, Kilian et al. 1997, Chen et al. 1997, Ahn et al. 1993, Tikhonov et al. 1999). A high level of colinearity was also observed within the Brassicaceae (Conner et al., 1998). In contrast, when grasses were compared with *Arabidopsis*, a low level or no colinearity was observed (Devos et al. 1999, van Dodeweerd et al. 1999). Thus, some of these studies, together with the data presented in this thesis, indicate that submegabase microsynteny is conserved between species within plant families, while in general, microsynteny is no longer observed when more distantly related plants like monocots and

dicots are compared. At present, no comparisons have been reported yet for species belonging to different dicot plant families. The RFLP markers and additional sequences that were subcloned from c1/c2 will enable the (microsyntenic) comparison of the *Sym2* region with region(s) within the *Arabidopsis* genome. This will answer the question as to whether microsynteny exists between species belonging to closely related clades of dicots: *Pisum sativum* and *Medicago truncatula* from the Rosid I clade, and *Arabidopsis* from the Rosid II clade.

The pea and *Medicago* molecular resources presented in this thesis (RFLP markers, c1/c2 subclones and cDNAs, pea cosmid contigs, *Medicago* BACs), will be useful to readily extend comparative microsynteny studies, to other nodulating and non-nodulating plants, such as *Lotus* and actinorhizal plants, and *Arabidopsis* and grasses, respectively. These studies will be crucial to know whether genes specific of nodule-forming plants do exist at the *Sym2* region, and whether *Sym2* itself is one of such specific genes. Moreover, it will answer the question as to whether a symbioses-specific gene cluster exists at the *Sym2* region. Similar comparative studies can be done on genomic regions carrying other important loci involved in nodulation (e.g. *Nod3*), by comparing fully sequenced genomic regions, or when such sequences are not available by using the molecular genetic methodologies presented in this thesis. Moreover, these studies can be done at the whole genome level by comparing and integrating recombination maps of species. All these studies will enhance the identification, cloning and functional characterization of genes that are unique to nodule forming plants.

The findings reviewed in Chapter 1 indicate that (some of) the mechanisms controlling nodulation are widespread in the plant kingdom. Some genes and signaling pathways that were considered to be specific of nodulating plants were found in plants unable to establish a nodule symbiosis. For example, a mutant carrot suspension cell line was rescued by the addition of Nod factors in the medium (de Jong et al., 1993), and a Nod factor perception and transduction machinery was found in rice (Reddy et al., 1998). In addition, at least some common genes are essential for a signal transduction pathway leading to gene activation in both nodular symbiosis and endomycorrhizal symbiosis (Albrecht et al., 1998). Furthermore, genes that before were considered to be symbioses-specific, have been found to be expressed in other non-symbiotic plant tissues, and part of the infection steps of the symbiosis, are derived from the cell cycle machinery. All these data indicate that the nodulation process is (in part?) derived from processes that are widespread among higher plants. Thus, it is

probable that comparative genomic studies similar to the ones presented in this thesis, and the extension of these studies to other nodule-forming and non-nodule-forming species, could help to reveal whether genes specific of nodule-forming plants do exist and are responsible for the unique properties shared by these plants. In addition, comparative genomics in combination with fuctional genomics will help to reveal how features common to all plants have been recruited for specific functions in nodulation.

# REFERENCES

Ahn S., Tanksley S.D. 1993.Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. USA 90: 7980-7984.

Bachem C.W.B., van der Hoeven R.S., de Bruijn S.M., Vreugdenhil D., Zabeau M., Visser R.G.F.1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. Plant J. 9(5): 745-753.

Bennetzen J.L., Freeling M. 1997. The unified grass genome: synergy in synteny. Genome Res. 7: 301-306.

Bennetzen J.L. 2000. Comparative Sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12: 1021-1029.

Blanc G., Barakat A., Guyot R., Cooke R., Delseny M. 2000. Extensive duplication and reshuffling in the *Arabidopsis thaliana* genome. Plant Cell 12: 1093- 1101.

Borisov AY, Barmicheva EM, Jacobi LM, Tsyganov VE, Voroshilova, VA and Tikhanovich. 2000. Pea (*Pisum sativum* L.) mendelian genes controlling development of nitrogen-fixing nodules and arbuscular mycorrhizae. Czech J. Genet. Plant Breeding 36: 106-110.

Chen M., SanMiguel P., De Oliveira A.C., Woo S.-S., Zhang H., Wing R. A., J.L. Bennetzen. 1997. Microcolinearity in *sh-2* homologous regions of the maize, rice, and sorghum genomes. Proc. Natl. Acad. Sci. USA 94: 3431-3435.

Conner J.A., Conner P., Nasrallah M.E., and Nasrallah J.B. 1998. Comparative mapping of the *Brassica S* locus region and its homeolog in *Arabidopsis*: implications for the evolution of mating systems in the Brassicaceae. Plant Cell 10: 801-812.

Cook D.R. 1999. *Medicago truncatula* - a model in the making! Curr. Opin. Plant Biol. 2: 301-304.

Covitz PA, Smith LS, Long SR. 1998. Expressed sequence tags from a root-hair enriched *Medicago truncatula* cDNA library. Plant Physiol. 117: 1325-32

Devos K.M., Beales J., Nagamura Y., Sasaki T. 1999. *Arabidopsis*-rice: will colinearity allow gene prediction across the Eudicot-Monocot divide? Genome Res. 9: 825-829.

Dunford R.P., Kurata N., Laurie D.A., Money T.A., Minobe Y., Moore G. 1995. Conservation of fine-scale DNA marker order in the genomes of rice and the Triticeae. Nucleic Acids Res. 23: 2724-2728.

Ellis T.H.N. 1993. The nuclear genome. In *Peas: genetics, molecular biology and Biotechnology* (eds Casey R., Davies D.R. ), Biotechnology in Agriculture series, 10, CAB international Oxon UK, p 13-48.

Feinberg A., Vogelstein B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13.

Firmin JL, Wilson KE, Carlson RW, Davies AE and Downie J. 1993. Resistance to nodulation of c.v. Afghansitaan peas is overcome by *nodX*, which mediates an *O*-acetylation of the *Rhizobium leguminosarum* lipo-oligosaccharide nodulation factor. *Mol. Microbiol.* 10: 351-360.

Foote T., Roberts M., Kurata N., Sasaki T. and G. Moore. 1997. Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. Genetics 147: 801-807.

Fransz P., Armstrong S., Alonso-Blanco C., Fischer T.C., Torrez-Ruiz R.A., Jones G. 1998. Cytogenetics for the model system *Arabidopsis thaliana*. Plant J. 12: 867-876.

Gale M.D., Devos K.M. 1998. Comparative genetics in grasses. Proc. Natl. Acad. Sci. USA 95: 1971-1974.

Geurts R., Heidstra R., Hardi A.-Z., Downie A., Franssen H., van Kammen A., Bisseling T. 1997. *Sym2* of *Pisum sativum* is involved in a Nod factor perception mechanism that controls the infection process in the epidermis. Plant Physiol. 115: 351-359.

Geurts R. 1998. A genetic approach to study Rhizobial Nod factor and mycorrhizal activated signaling. PhD Thesis, Wageningen Agricultural University, The Netherlands, 110 p.

Gloudemans T., Bhuvaneswari TV., Moerman M., van Brussel AAN., van Kammen A., Bisseling T. 1989. Involvement of *Rhizobium leguminosarum* nodulation genes in gene expression in pea root hairs. Plant Mol. Biol. 12: 157-167.

Grant D., Cregan P., Shoemaker R. C. 2000. Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proc. Natl. Acad. Sci. USA 97(8): 4168- 4173.

Gualtieri G., Kulikova O., Limpens E., Kim D-J., Cook D.R., Bisseling T. and˙Geurts R. Characterization of the *Pisum sativum sym2* locus by means of conserved genome structure with the model legume *Medicago truncatula*. Submitted, a.

Gualtieri G., Cook D., Bisseling T. Microsynteny between the *Medicago truncatula Sym2*-orthologous genomic region and another genomic region both located on the long arm of chromosome 5. Submitted, b.

Gualtieri G., Kulikova O., Cook D., Bisseling T. Isolation of RFLP markers linked to the *nod3* hypernodulation locus on pea linkage group I, and identification of orthologous genomic regions in *Medicago truncatula* A17. Submitted, c.

Heidstra R., Bisseling T. 1996. Nod factor-induced host responses and mechanisms of Nod factor perception. New Phytol. 133: 25-43.

Helentjaris T., Weber D.L., Wright S. 1998. Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics 118: 353-363.

Hulbert S.H., Richter T.E., Axtell J.D., Bennetzen J.L. 1990. Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc. Natl. Acad. Sci. USA 87: 4251-4255.

Jacobsen E., Feenstra W.J. 1984. A new pea mutant with efficient nodulation in the presence of nitrate. Plant Sci. Letters 33: 337-344.

Jiang Q and Gresshoff PM.1997. Classical and molecular genetics of the model legume *Lotus japonicus*. Mol. Plant. Microbe Interact. 10: 59-68.

Josey D.P., Beylon J.L., Johnston A.W.B., Beringer J.E. 1979. Strain identification in *Rhizobium* using intrinsic antibiotic resistance. J Appl. Bacteriol. 46: 343-350.

Kilian A., Chen J., Han F., Steffenson B., Kleinhofs A. 1997. Towards the map-based cloning of the barley rust resistance genes *Rpg1* and *Rpg4* using rice as an intergenomic cloning vehicle. Plant Mol. Biol. 35: 187-195.

Kishimoto N., Higo H., Abe K., Arai S., Saito A., K. Higo. 1994. Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known funtions. Theor. Appl. Genet. 88: 722-726.

Kneen B.E., LaRue T.A., Weeden N. 1984. Genes reported to affect symbiotic nitrogen fixation by peas. Pisum News Letters 16: 31-34.

Kozik A., Heidstra R., Horvath B., Kulikova O., Tikhonovich I., Ellis THN., van Kammen A., and 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.

Kozik A. 1996. Fine mapping of the *SYM2* locus of pea linkage group I. PhD Thesis, Wageningen Agricultural University, The Netherlands, 111 p

Ku HM, Vision T., Liu J., Tanksley S.D. 2000. Comparative sequenced segments of the tomato and *Arabidopsis* genome: large-scale duplication followed by selective gene loss creates a network of synteny. Proc. Natl. Acad. Sci. USA 97(16): 9121-9126.

Kulikova O., Gualtieri G., Geurts R., Kim D-J., Cook D., Huguet, T., de Jong J.H., Fransz P.F., and Bisseling T. Integration of the FISH-pachytene and genetic maps of *Medicago truncatula*. Plant J., in press.

Langercrantz U., Putterill J., Coupland G., Lydiate D. 1996. Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J. 9(1): 13-20.

Langercrantz, U. 1988. Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150: 1217-1228.

Liang P., Pardee A.B. 1992. Differential display of eucariotic messenger RNA by means of the polymerase chain reaction. Science 257: 967-971.

Lie T.A. 1984. Host genes in *Pisum sativum* L. conferring resistance to European *Rhizobium leguminosarum* strains. Plant Soil 82: 415-425.

McClelland M., Mathieu-Daude F., Welsh J. 1995. RNA fingerprinting and differential display using arbitrary primed PCR. Trends Genet. 11:242-246.

Money T., Reader S., Qu L.J., Dunford., Moore G. 1996. AFLP-based mRNA fingerprinting. Nucleic Acids Res. 24(13): 2616-2617.

Nagamura Y., Inoue T., Antonio B.A., Shimano T., Kajiya H., Shomura A., Lin S.Y., Kuboki Y., Harushima Y., Kurata N., Minobe Y., Yano M., and Sasaki T. 1995. Conservation of duplicated segments between rice chromosomes 11 and 12. Breed. Sci. 45: 373-376.

Nam YW., Penmetsa RV., Endre G., Uribe P., Kim D., Cook DR. 1999. Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene-response genes. Theor. Appl. Genet. 98: 638-646.

Noel L., Moores T.L., van der Biezen E.A., Parniske M., Daniels M.J., Parker J.E., Jones J.D.G. 1999. Pronounced intraespecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. Plant Cell 11: 2099-2111.

Ovtsyna AO, Geurts R, Bisseling T, Lugtenberg BJJ, Tikhonovich IA, Spaink HP. 1998. Restriction of Host Range by the *sym2* Allele of Afghan pea is non-specific for the type of modification at the reducing terminus of nodulation signals. Mol. Plant-Microbe Ineract. 11: 418-422.

Parniske M., Hammond-Kosack KE, Golstein C., Thomas CM., Jones DA., Harrison K., Wulff BB., Jones JD. 1997. Novel disease resistane specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell 91: 821-832.

Pawlowski K., Kunze R., De Vries S., and Bisseling T. 1994. Isolation of total, poly(A) and polysomal RNA from plant tissues. In: Plant Mol. Biol. Manual (Gelvin SB, Schilperoort RA eds), Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 1-13.

Postma J.G., Jacobsen E., Feenstra W.J. 1988. Three pea mutants with an altered nodulation studied by genetic analysis and grafting. J. Plant Physiol. 132: 424-430.

Sambrook J., Fritsch E.F., and Maniatis T. 1989. Molecular Cloning, a laboratory manual, Second edition. Cold Spring Harbor Laboratorium Press.

Schmidt R. 2000. Synteny: recent advances and future prospects. Curr. Opin. Plant Biol. 3: 97-102.

Schneider A., Walker S.A., Poyser S., Sagan M., Ellis T.H., Downie J.A. 1999. Genetic mapping and functional analysis of a nodulation-defective mutant (*sym19*) of pea (*Pisum sativum* L.). Mol. Gen. Genet. 262:1-11.

Temnykh S.V., Kneen B.E., Weeden N.F., LaRue T.A. 1995(a). Localization of *nod-3*, a gene conditioning hypernodulation, and identification of a novel translocation in *Pisum sativum* L. cv. Rondo. J. Heredity 86: 303-305.

Temnykh S.V., Weeden N.F., LaRue T.A. 1995(b). *Sym-2* and *nod-3* are independent but closely linked genes influencing nodule development in pea. Pisum Genet. 27: 26-28.

Tikhonov A.P., SanMiguel, Nakajima Y., Gorenstein N.M., Bennetzen J.L., Avramova Z. 1999. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc. Natl. Acad. Sci. USA 96: 7409-7414.

"The Arabidopsis Genome Initiative". 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408(6814):796-815

van Dodeweerd A.M., Hall C.R., Bent E.G., Johnson S.J., Bevan M.W. and Bancroft I. 1999. Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. Genome 42:887:892.

Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23(21): 4407-4414.

Weeden NF, Kneen BE, LaRue TA.1990. Genetic analysis of *Sym* genes and other nodule-related genes in *Pisum sativum*. In: Nitogen Fixation: achievements and Objectives (Gresshoff, Roth, Stacey, Newton *eds.*), Chapman and Hall New York, 323-330.

Williams J.G.K., Kubelik A.R., Lavak K.J., Rafalski J.A., Tingey S.Y. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18: 6531-6535.

Zegzouti H, Jones B, Frasse P, Marty C, Maitre B, Latch A, Pech JC, Bouzayen M. 1999. Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. Plant J 18(6):589-600.

# SUMMARY

To determine the usefulness of *M. truncatula* as intergenomic vehicle for the positional cloning of pea genes it was studied whether these legumes are microsyntenic. These studies were focused on the pea *Sym2* and *Nod3* genomic regions. The *M. truncatula* orthologous genomic regions have been cloned and it was shown that these regions of the two legumes are microsyntenic. Both *Sym2* and *Nod3* play a key role in the pea-*Rhizobium* symbiosis, controlling Nod factor-structure dependent infection and autoregulation of nodule number, respectively.

A *M. truncatula* A17 BAC library was screened with a pea marker tightly linked to *Sym2* and 11 clones were isolated. These clones formed three different contigs c1, c2, and c3. These three contigs were extended to about twice their original size by chromosome walking. Genetic and FISH mapping in *Medicago* revealed that the three contigs map on chromosome 5, and that c1 and c2 are tightly linked while c3 maps at a distance of 9 cM from c1/c2 on the same arm of this chromosome. 8 RFLP markers (including cDNAs and contig subclones) were isolated from c1/c2. Mapping of these markers using pea RILs and introgression lines demonstrated that c1/c2 represents the *Medicago Sym2*-orthologous genomic region. Moreover, three markers showed recombinations between their pea homologous sequences and *Sym2*, delimitating the pea *Sym2* region in the RILs and introgression lines and the *Medicago Sym2*-orthologous region. The *Medicago Sym2*-orthologous region was delimitated to about 350 Kbp of c1/c2. In addition, by using a c2 subclone that encodes for a sequence highly homologous to the LRR-motif of the Cf4 and Cf9 tomato [*L. esculentum*] disease resistance proteins, a pea cDNA was isolated from a pea root hair library that also contains a LRR-domain highly homologous to that of *Cf4/Cf9*. The isolation of this pea RFLP marker demonstrates the use of *M. truncatula* as intergenomic positional-cloning vehicle of pea genes located within microsyntenic genomic regions. Furthermore, it was shown that the pea *Sym2*-region is rich in *Cf4/Cf9* LRR-like sequences. Detailed analysis of 22 sequences from c1/c2 (including all RFLP markers) showed that 4 of these sequences have homologues in c3, and that they are organized in clusters with a similar linear order in these contigs. This indicates that c1/c2 and c3 probably arose through duplication of a chromosomal segment.

By using the RNA differential display in combination with RILs and introgression lines, the tightly linked RFLP markers *dd21.5* and *Psc2.6* were isolated that are linked to the hypernodulating *Nod3* locus and represent the closest markers mapping to the "south" of this locus on pea linkage group I. These markers were used to identify the *M. truncatula* A17 orthologous region with the aim to start the microsynteny-based positional cloning of *Nod3*. The *M. truncatula* A17 BAC library was screened with these two markers. BAC clone 21F22 was isolated that contains the orthologue of *dd21.5*. In addition 5 BAC clones were identified that co-hybridize with both *dd21.5* and *Psc2.6*, demonstrating the existence of *Medicago* genomic regions microsyntenic with the pea genomic region containing these two markers. BAC 21F22 was mapped by FISH on *Medicago* chromosome 4, showing a local disruption of synteny at the *dd21.5* locus between pea linkage group 1 and *Medicago* chromosome 5 that were syntenic for other markers (e.g. the markers isolated from the *Sym2*-orthologous region, and the pea *Eil2* marker). This finding reveals a chromosomal translocation that took place either in pea or in *Medicago*. However, it is unknown whether this translocation extends beyond the *dd21.5* locus and includes the *Nod3* locus. In addition to the strong FISH signal given by 21F22 on chromosome 4, weak signals were observed close to the telomeres of chromosome 5 and the position of one of these signals is comparable to the position of the c1/c2 *Sym2*-orthologous region. Thus, it is possible that in spite of the translocation, chromosome 5 contains sequences with a low homology with those of the pea linkage group I segment containing *dd21.5*. It remains to be determined whether the five microsyntenic BACs map in these chromosome 5 regions or whether they map in chromosome 4 and form a contig with BAC 21F22.

The data presented in this thesis set the basis for the microsynteny-based cloning of the *Sym2* and *Nod3* genes by using *M. truncatula* as intergenomic positional-cloning vehicle. In addition, the molecular resources generated in this thesis are useful to extend microsynteny studies at the *Sym2* and *Nod3* regions to other legume and actinorhizal nodule-forming species (e.g. *Lotus japonicus*), and also to non-nodulating species. (e.g. *Arabidopsis*). This will enable the identification of genes within these genomic regions that might be unique to nodule-forming plants and to symbitic nitrogen fixation

# SAMENVATTING

Erwt is een belangrijk cultuur gewas maar omdat het genoom van deze plant groot is, is het zeer moeilijk om genetisch geïdentificeerde genen te kloneren. Medicago truncatula heeft daarentegen een veel kleiner genoom. Het is daarom aantrekkelijk om Medicago te gebruiken als 'intergenomic cloning vehicle' voor erwtengenen. Om te beoordelen of dit mogelijk zou zijn, is bepaald of de genenvolgorde van deze 2 vlinderbloemigen veel op elkaar lijkt (microsynteny). Deze studies zijn gericht geweest op de regio's rond de erwtengenen Sym2 en Nod3, twee genen die een rol spelen in de Rhizobium-plant interactie. De overeenkomstige Medicago regio's zijn gekloneerd en er is aangetoond dat de volgorde van de genen in deze regio's geconserveerd is.

M.b.v. een erwtenmerker die strikt gekoppeld is aan Sym2 werden uit een Medicago BAC bibliotheek klonen geïsoleerd die 3 contigs, c1, c2 c3, vormden. M.b.v. "chromosome walking" zijn deze contigs vervolgens verlengd. Genetische analyses lieten zien dat deze 3 contigs op chromosoom 5 gelegen zijn en dit werd m.b.v. FISH studies bevestigd. C1 en c2 zijn nauw gekoppeld en c3 ligt 9cM verwijderd van deze 2 contigs. 8 RFLP merkers werden geïsoleerd die op c1/c2 gelegen zijn. De genetische positie van deze merkers op het erwtengenoom werd bepaald m.b.v. RILs en introgressie lijnen en deze studies lieten zien dat c1/c2 het Medicago gebied ortholoog aan de Sym2 regio van de erwt omvat. Verder werden de overeenkomstige posities van recombinaties die het Sym2 gen in de erwt flankeren op c1/c2 bepaald. Op grond hiervan kan geconcludeerd worden dat het gebied waarbinnen een evt. Medicago Sym2 gen gelegen is maximaal 350 kb groot is. Verder werd met een fragment van c2, dat sterk homoloog is met het LRR-motief van de Cf4 en Cf9 resistentie genen van tomaat, een erwte cDNA kloon geïsoleerd die codeert voor een eiwit dat ook een dergelijk LRR domein bevat. De isolatie van deze erwtenkloon bevestigd dat Medicago een goede intermediair is voor de klonering van erwtengenen. Verder werd aangetoond dat de Sym2 regio van de erwt relatief veel LRR-achtige sequenties bevat. Hybridisatie studies met 22 subfragmenten van c1/2 (inclusief alle gebruikte RFLP merkers) lieten zien dat 4 fragmenten hybridiseren met c3 fragmenten. De clusters waarin deze sequenties voorkomen hebben in de beide contigs een zelfde volgorde. Dit geeft aan dat c1/c2 en c3 zijn ontstaan door duplicatie van een chromosoom segment.

M.b.v. RNA differential display zijn de RFLP merkers dd21.5 en Psc2.6 geïsoleerd die beide nauw gekoppeld zijn aan het Nod3 locus en hiervan ten zuiden liggen. De M. truncatula orthologe regio is vervolgens geïsoleerd met als uiteindelijke doel Nod3 te kloneren. Medicago BAC 21F22 bevat een ortholoog van dd21.5. Verder werden 5 BAC's geïsoleerd die met beide erwtenmerkers hybridiseren, hetgeen aangeeft dat de regio's waar deze genen gelegen zijn een geconserveerde genvolgorde hebben. FISH studies lieten zien dat BAC21F22 gelegen is op chromosoom 4, hetgeen betekent dat de synteny tussen koppelingsgroep 1van erwt en 5 van Medicago dus verstoord is. Dit wijst erop dat in erwt of Medicago een translocatie heeft plaatsgevonden. Het is echter niet duidelijk of deze translocatie ook Nod3 omvat. 21F22 hybridiseert sterk met chromosoom 4 maar ook met 2 gebieden, gelegen aan de einden van chromosoom5. De positie van een van deze gebieden is vergelijkbaar met die van c1/c2 en het is daarom niet uitgesloten dat chromosoom 5 van Medicago sequenties bevat die enige homologie vertonen met dd21.5.

De waarnemingen beschreven in dit proefschrift vormen een basis voor de klonering van de erwtengenen Sym2 en Nod3 mm. M.truncatula als 'intergenomic cloning vehicle'. Verder kunnen de gekloneerde fragmenten gebruikt worden om de Sym2 en Nod3 orthologe gebieden van andere vlinderbloemigen (b.v. Lotus japonicus), actinorhiza planten en soorten die geen wortelknolsymbiose kunnen aangaan (b.v. Arabidopsis) te analyseren. Dit zal het mogelijk maken genen op te sporen die uniek zijn voor planten die stikstofbindende wortelknollen kunnen vormen.

# RESUMEN

Con el objetivo de determinar el valor de *Medicago tuncatula* como vector intergenómico para el clonado posicional de genes en la arveja (*Pisum sativum*), se estudio el nivel de microsintonía entre estas dos especies de leguminosas. Estos estudios se enfocaron en dos regiones genómicas independientes de la arveja que contienen los loci *Sym2* y *Nod3*, respectivamente. Las regiones genómicas ortólogas de *M. truncatula* fueron clonadas, demostrándose la existencia de microsintonía entre estas regiones en ambas especies. *Sym2* y *Nod3* tienen un rol principal en la simbiosis arveja-*Rhizobium*, y controlan la infección bacteriana dependiente de la estrucutra del factor de nodulación ("Nod factor") y la autorregulación del número de nódulos, respectivamente.

Una biblioteca genómica BAC ("Bacterial Artificial Chromosome") fue investigada empleando una sonda correspondiente a un marcador molecular (RFLP) ligado al gen *Sym2*. Esto resultó en el aislamiento be 11 clones BAC que se organizaron en 3 contigs independientes: c1, c2 y c3. Los tres contigs fueron extendidos aproximadamente al doble de su dimensión por la tecnica de "caminando sobre el cromosoma" ("chromosome walking"), resultando en contigs de 300 Kbp, 170 Kbp y 150 Kbp para c1, c2 y c3, respectivamente. Mapeo genetico y por FISH ("Fluorescent *In Situ* Hybridization") en cromosomas meióticos de *M. truncatula* en estadio de paquiteno demostraron que los tres contigs mapean en el cromosoma 5 y que c1 y c2 se encuentran altamente ligados mientras que c3 mapea a una distancia de 9 cM de la región genómica c1/c2 en el mismo brazo del cromosoma. Ocho marcadores moleculares, incluyendo cDNAs y subclones de los contigs, fueron aislados de la región c1/c2. El mapeo de estos marcadores moleculares empleando RILs ("Recombinant Inbreed Lines") y lineas introgresivas ("introgression lines") demostraron que la región c1/c2 representa la región de *M. truncatula* que es ortóloga a la región de la arveja que contiene el locus *Sym2*. Mas aun, tres marcadores moleculares mostraron recombinaciones entre sus secuencias homologas en arveja y *Sym2*. Como consecuecia la región genómica c1/c2 de *M. truncatula*, ortóloga de la region *Sym2* de la arveja, quedó delimitada en 350 Kbp. Además, un subclon de c2 con alta homología al motivo LRR (Leucine Rich Motif) de los genes de resistencia a patógenos de tomate [*Licopersicon esculentum*] *Cf4* y *Cf9* se utilizó para aislar el clon cDNA *PscLRR52* de una biblioteca de cDNAs preparada a partir de pelos radiculares de arveja. *PscLRR52* también contiene un motivo LRR altamente homologo al de las proteínas Cf4 y Cf9. Cuando este cDNA fue mapeado por RFLP en

arveja, se encontró que hibridizaba a 7 fragmentos de restricción 5 de los cuales eran polimórficos y se encuentraron altamente ligados a *Sym2*. Esto demuestra el valor de *M. truncatula* como vector intergenómico de clonacion posicional de genes de arveja comprendidos en regiones genómicas microsinténicas. También se demostró que la región genómica de la arveja que contiene el gen *Sym2* es rica en motivos LRR de tipo Cf4 y Cf9, que por otra parte podrian corresponder al gen *Sym2*. Entre varios otros genes aislados candidatos a ser el gen *Sym2* se encuentran diferentes kinasas de tipo receptor y genes de resistencia que contiene motivos LRR de tipo TMV. El análisis detallado de 22 secuencias aisladas de c1/c2 (incluyendo todos los marcadores RFLP) y de la región *Sym2* de la arveja, mostó que 4 de estas secuencias presentan secuencias homologas en c3, que se encuentran organizadas en grupos ("clusters") que presentan un orden linear similar en estos contigs. Esto indica que las regiónes genómicas c1/c2 y c3 surgieron probablemente como resultado de una duplicación de un segmento cromosómico.

Empleando la tecnica de "RNA differential display" en RILs y lineas introgresivas, fueron aislados los marcadores moleculares de tipo RFLP *dd21.5* y *Psc2.6*, que se encuentran altamente ligados entre si y ligados al locus *Nod3* (*nod3* corresponde a la mutación supernoduladora). Estos dos marcadores constituyen los marcadores que presentan hasta la fecha el mas alto ligamiento al "sur" del locus *Nod3* que se encuentra en el grupo de ligamineto I de la arveja. Se emplearon estos marcadores para identificar la región genómica ortóloga de *M. truncatula* que contiene los homólogos de estos marcadores, con el objetivo de comenzar el clonado posicional de gen *Nod3* de arveja basándose en la microsintonía entre estas dos especies. Asi se investigó la biblioteca BAC de *M. truncatula* empleando ambos marcadores como sonda. El clon BAC 21F22 que contiene el ortólogo de *dd21.5* fue aislado. Ademas, 5 clones BAC fueron aislados que co-hibridizan con *dd21.5* y con *Psc2.6*, demostrándose la existencia de regiónes genómicas en *M. truncatula* que presentan microsintonía con la región de arveja que contiene estos dos marcadores en el grupo de ligamiento I. El clon BAC 21F22 fue mapeado por FISH en el cromosoma 4 de *M. truncatula*, demostrándose asi una disrupción local, a nivel del locus *dd21.5*, de la sintonía entre el cromosoma 5 de *M. truncatula* y el grupo de ligamiento I de la arveja que mostraron sintonía cromosómica para otros marcadores (p.e. marcadores de la región ortóloga de la región *Sym2*, y el marcador de la arveja *Eil2*). Este descubrimiento revela una translocación cromosómica que ocurrió en la arveja o en *M. truncatula*. Sin embargo, no es sabido aun si esta translocación se extiende mas alla del locus *dd21.5* e incluye el locus *Nod3*. Ademas de

la hibridización intensa de la BAC 21F22 reveladas por FISH en el cromosoma 4, hibridizaciones ténues se registraron cerca de los telómeros del cromosoma 5 siendo la posición de una de estas comparable con aquella de la región *Sym2*-ortóloga c1/c2. Por lo tanto, es probable que a pesar de esta translocación, el cromosoma 5 contenga secuencias con baja homología con aquellas del grupo de ligamiento I de arveja alrededor del locus *dd21.5*. Resta aun por determinar si los 5 clones BAC microsinténicos mapean en *M. truncatula* en estas regiones del cromosoma 5, o si mapean en el cromosoma 4 formado un contig con el clon BAC 21F22.

Los resultados presentados en esta tesis constituyen la "estructura" para el clonado basado en microsintonía de los genes *Sym2* y *Nod3* de la arveja, empleando a *M. truncatula* como vehículo intergenómico de clonado posicional. Ademas, los recursos moleculares generados en esta tesis son útiles para extender estudios de microsintonía de las regiones *Sym2* y *Nod3* a otras especies de leguminosas (p.e. *Lotus japonicus*) y actinorrizas formadoras de "nodulos", y también a especies relacionadas no formadoras de "nodulos" pertenecientes al clado Rosid I y a especies menos relacionadas pertenecientes al clado Rosid II (p.e. especies de los géneros *Arabidopsis* y *Brassica*). Esto permitirá la identificación de genes comprendidos en estas regiones genómicas que podrian ser exclusivos de plantas formadoras de nódulos y de esta fijación simbiótica de nitrógeno. Dicho de otra forma, estos estudios haran posible abordar una de las interrogantes mas importantes de la biología vegetal y de la ciencia de comparación de genomas: si existen propiedades únicas de las plantas capaces de establecer una simbiosis nodular y si estas propiedades han dejado su "firma" evidenciada a nivel de la estructura genómica.

# CURRICULUM VITAE

Gustavo Gualtieri was born in Montevideo, "República Oriental del Uruguay", on November 30, 1967. He studied at the President John Fischeral Kennedy School and carried out additional education at the Anglo-Uruguayan Cultural Institute and Mathematics Academy. His high-school specialization study was Biology and Medicine, concluding in 1985. From when he was 6 years old he intensively practiced sports. Until 1987 he was a gymnast of the "L'Avenir" Club tournament team, after what he engaged in artistic and acrobatic performances. Being a referee of gymnastics, in 1987 he was appointed Referee Principal and was elected by the Gymnastics League representatives of the Uruguayan Olympic Committee as President and Director of the Institute and School of Gymnastics Referees, respectively.

In 1986 he started studies at the University of the Republic of Uruguay and in March 1991, after a four-years study, he graduated as "Licenciado" in Biological Sciences. Since 1991 he worked in Uruguay at the Laboratory of Crop Protection of the Ministry of Agriculture and at the Department of Biochemistry of the Biological Research Institute of the Ministry of Education and Culture in collaboration with the University of the Republic. In September 1993 he started a Master of Science in Biotechnology at Wageningen Agricultural University in Netherlands and graduated in January 1995, after what he re-engaged work in Uruguay. In September 1996 he started Ph.D. studies towards this doctoral thesis. He carried out research with Prof. Dr. Ton Bisseling at the Laboratory of Molecular Biology of the Department of Experimental Plant Sciences of Wageningen Agricultural University. During August 4 to 20 1997, he participated in the theoretical/practical course "Genome diversity and genome expression in plants", organized by Prof. Dr. Marc Van Montagu at the Department of Genetics of the Flanders Interuniversity Institute for Biotechnology of the University of Gent in Belgium. Since March 1998 he did research for three months with Prof. Dr. Douglas Cook at the Crop Biotechnology Center of Texas A&M University in the United States of America.

# ACKNOWLEDGEMENT

the modifications and formatting of this thesis' text and figures, respectively. To Marie-José van Iersel and Maria Augustijn for kindly assisting me during my stay in Wageningen. To Jan Hontelez and Jan Verver for readily facilitating research chemicals. To Piet Madern and Gerrit Nellestijn for on-due-time provision of research materials.

I greatly acknowledge Wageningen Agricultural University and CSIC-Uruguay for the awarding of Ph.D.-fellowships to my person and to Wageningen University for funding the printing of this thesis. I am also grateful to the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (N.W.O.) and Human Frontiers Science Organization (H.F.S.O.) for financing the research presented in this thesis through grants to Dr. Ton Bisseling.

Wageningen University will always outstand in my career highlights.

I am very much grateful to the Embassy of Uruguay in Holland and the Embassy of Holland in Uruguay for their constant support.

I thank to all my friends with who I shared a great time in Nederland and to all those who were far away but close enough to make me laugh a lot!

I am deeply acknowledged to my family for educating me the way they did, thereby giving me the basic tools and skills that were largely required to carry out the research presented in this thesis.