Theor Appl Genet (2002) 105:771–779
DOI 10.1007/s00122-002-0949-6

G. Gualtieri · T. Bisseling

Microsynteny between the *Medicago truncatula SYM2*-orthologous genomic region and another region located on the same chromosome arm

Received: 8 June 2001 / Accepted: 7 January 2002 / Published online: 12 July 2002

Abstract A synteny based positional cloning approach was started to clone the pea *SYM2* gene by using locally conserved genome structure with the model plant *Medicago truncatula*. We reported that a pea marker tightly linked to *SYM2* was used to screen a *M. truncatula* BAC library, and two contigs named C1/C2 and C3 were constructed that are both located on the long arm of *M. truncatula* chromosome 5 and separated by 9 cM. C1/C2 is highly microsyntenic to the pea *SYM2* genomic region and corresponds to the *M. truncatula SYM2*-orthologous region, which is delimitated to 350 kbp. In this manuscript we analyze the distribution in the three contigs of 22 sequences and their homologues, including eight C1/C2 and two pea RFLP markers linked to *SYM2*. Among the analyzed sequences are several different (receptor) kinase-like gene sequences and two classes of LRR-containing resistance protein-like sequences. From all the studied sequences only four detected homologous sequences in C3, and their distribution is comparable in C1/C2 and C3, suggesting that a 70-kbp and a 120-kbp segments of these two contigs, respectively, arose through a duplication. The implications of these findings for the cloning of *SYM2* are discussed.

Keywords Comparative genomics · Duplication · *Medicago truncatula* · Microsynteny · *SYM2*

Introduction

A distinguishing property of legumes that contributes to their agronomic importance is their ability to establish a symbiosis with rhizobia, resulting in the formation of nitrogen-fixing root nodules. The legume sub-family of *Papilionoideae* contains many agronomically important species, including soybean (*Glycine* sp.), alfalfa (*Medicago sativa*) and pea (*Pisum sativum*). Although pea is genetically well characterized and about 30 *SYM* genes that are essential for the formation of nitrogen-fixing root nodules have been genetically mapped (Borisov et al. 2000), positional cloning of these genes is limited by the large genome of this species. Therefore, two legume species, *Medicago truncatula* (Cook 1999) and *Lotus japonicus* (Jiang and Gresshoff 1997), have been proposed as model systems for the study of legume biology. In particular, *M. truncatula* (Trifoleae tribe) is closely related to pea (Vicieae tribe) (Doyle 1995; Gualtieri and Bisseling 2000) and has a simple genome organization (Kulikova et al. 2001). Moreover, numerous genomic and genetic tools (e.g. Covitz et al. 1998; Nam et al. 1999; Cook 1999) have been developed for *M. truncatula*.

Recent studies have shown that *M. truncatula* chromosome 5 and pea linkage group I are syntenic, and that *M. truncatula* can be used as an intergenomic cloning vehicle for the pea *SYM2* gene (Gualtieri et al. 2002). In general, synteny was observed between species belonging to the same plant family and was mostly studied within the families Solanaceae, Poaceae and Brassicaceae (reviewed in Bennetzen 2000; Devos and Gale 2000; Paterson et al. 2000; Schmidt 2000), and to a less extent within the Leguminosae (Weeden et al. 1992; Menancio-Hautea et al. 1993; Boutin et al. 1995; Gualtieri et al. 2002). These studies have shown a conserved gene composition and order both at the chromosome and at the megabase/submegabase levels. In addition, they revealed that genome evolution is accompanied by gene and chromosomal segmental rearrangements such as deletions, insertions, translocations and duplications. These rearrangements result in synteny disruption and increased complexity in synteny studies aiming at the identification of orthologous genomic regions in the compared species. Rearrangements were also reveled after the completion and analysis of the *Arabidopsis* genome se-

Communicated by C. Möllers

G. Gualtieri (✉) · T. Bisseling
Molecular Biology, Agricultural University Wageningen.
Dreijenlaan 3, 6703HA Wageningen, The Netherlands
e-mail: ggualtieri@tifton.cpes.peachnet.edu

Present address:
G. Gualtieri, Coastal Plain Experimental Station,
Department of Horticulture, The University of Georgia,
P.O. Box 748, Tifton, GA 31793-0748, USA

quence (The *Arabidopsis* Genome Initiative 2000) that provided the first knowledge about the structure and composition of a plant genome. The *Arabidopsis* genome evolved through whole-genome duplication, accompanied by subsequent gene loss and extensive local gene duplications. In addition, 24 large duplications of at least 100 kbp are present within different chromosomes. The sequence conservation and the number of copies of genes within duplicated segments varies considerably.

Linkage group I of pea is the genomic region with the highest density of legume genes involved in the symbiosis with *Rhizobium leguminosarum* bv *viciae*, containing several symbiotic genes within a 20-cM interval (Weeden et al. 1990; Temnykh et al. 1995a, b; Kozik et al. 1996; Schneider et al. 1999). One of these genes is *SYM2* that was first identified in the pea accession 'Afghanistan' where it controls infection by rhizobia in a Nod factor structure-dependent manner (Firmin et al. 1993; Geurts et al. 1997). Nod factors are the rhizobial signal molecules that are involved in the induction of various host responses (Heidstra and Bisseling 1996).

Within a synteny-based positional cloning approach, the isolation and delimitation of the *M. truncatula* *SYM2*-orthologous region has been reported (Gualtieri et al. 2002). This region maps on *M. truncatula* chromosome 5 and is highly microsyntenic with the pea *SYM2* region located on linkage group I. This research was initiated by the isolation of two *M. truncatula* BAC contigs, named C1/C2 and C3, hybridizing with the pea RFLP marker *PscW62-1* that is tightly linked to *SYM2*. Microsynteny was demonstrated by the isolation of several RFLP markers (*Mtc831*, *Mtc923*, *Mtc411*, MtgG28, Mtg3556, *Mtc7x1*, Mtg3552 and Mtg63EB4) from the *M. truncatula* C1/C2 contig and the genetic mapping of these markers in pea. Markers *Mtc831* and *Mtc923*, and Mtg63EB4, have shown recombination between their hybridizing homologous pea sequences and *SYM2*, thereby delimiting the *M. truncatula* C1/C2 *SYM2*-orthologous region to 350 kbp. C3 was mapped at a genetic distance of 9 cM to C1/C2 on the long arm of chromosome 5 (Gualtieri et al. 2002). Apart from their hybridization to C1/C2, markers *PscW62-1*, and *Mtc411* and Mtg3552, showed a strong and very weak hybridization with C3, respectively. Therefore, to study the level of parasynteny between C1/C2 and C3, and to confirm that C1/C2 is indeed the *SYM2*-orthologous region, we constructed detailed contig physical maps and studied the distribution within these contigs of several sequences isolated from C1/C2 (including the previously described RFLP markers) and homologous sequences isolated from the pea *SYM2* region. These studies confirmed that C1/C2 is indeed the *SYM2*-orthologous region. Moreover, the presence and similar distribution of four sequences in both C1/C2 and C3 suggests that a segment of each contig arose through duplication.

Materials and methods

BAC clone isolation and construction of contig physical maps

The BAC clones composing C1, C2 and C3 belong to the *M. truncatula* A17 BAC library constructed by Nam et al. (1999), and were isolated by screening with the pea RFLP marker *PscW62-1* that is tightly linked to *SYM2* and by PCR- and hybridization-based chromosome walking (Gualtieri et al. 2002). BAC clone DNA was isolated according to Nam et al. (1999) and contigs were constructed by a combination of AFLP and restriction/hybridization BAC DNA fingerprinting. These studies resulted in arrays of overlapping BAC clones that formed contigs and defined different regions within each contig. Fingerprint DNA fragments were manually assigned to the different contig regions thereby generating a physical contig map. The sequences studied in this manuscript were assigned to different contig regions by hybridization. In *M. truncatula* chromosome 5, the tightly linked contigs C1 and C2 were initially separated by a small gap of about 10 kbp that was closed (Gualtieri et al. 2002) after screening an expanded version of the *M. truncatula* BAC library (Cook et al. unpublished) with the C1 and C2 end subclones Mtg2511 and Mtg3556, respectively. Although this 10-kbp sequence is not included in the present study, contigs C1 and C2 are sometimes referred here to as C1/C2.

Contig DNA blots and *M. truncatula* genomic blots

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 from each selected BAC clone and 5 µg of *M. truncatula* A17 genomic DNA were used per lane. The individual BAC DNA samples and the *M. truncatula* genomic DNA were digested to completion with *Hin*dIII (Gibco BRL) and analyzed by electrophoresis on 20-cm long 1% agarose gels prepared and run in 1 × TAE at 45 V. Electrophoreses were done overnight. 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 pre-hybridization and hybridization with probes were as recommended by the membrane manufacturer (Amersham).

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

In order to visualize the different hybridization specificities, pre-hybridizations and hybridizations were done at 60 °C-65 °C, and blots were usually washed and exposed after both low- and high-stringency washings. Washings were generally done in 5 × SSPE 0.1% SDS for 10 min, 2 × SSPE 0.1% SDS for 20 min 1 × SSPE 0.1% SDS for 15 min and in 0.5 SSPE 0.1% SDS for 20 min. Blots were analyzed with a Storm 840 phosphor imager.

cDNA libraries and screenings

The pea root-hair cDNA library was constructed by Stratagene in a Lambda ZAPII vector system using equal amounts of poly (A)+ RNA isolated from root hairs of 6-day old cv Finale plants non-in-

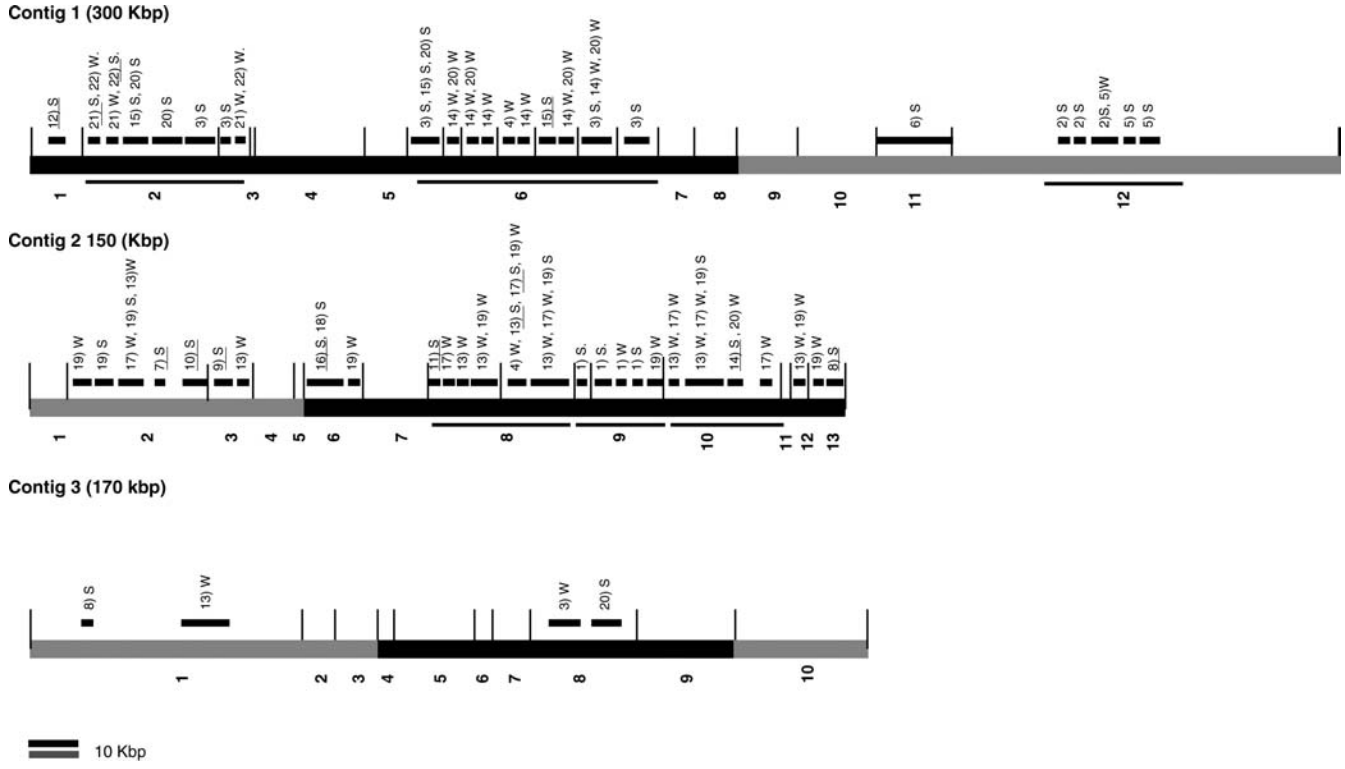


Fig. 1 Physical distribution of sequences within the contigs. *Hin*dIII restriction maps of C1, C2 and C3. The *thick horizontal black/gray line* represents a contig were *black lines* correspond to the contig sequence isolated by hybridization screening of the BAC library with *PscW62-1*, whereas *gray lines* correspond to the contig sequence obtained by PCR- or hybridization-based chromosome walking. The *vertical thin lines* are *Hin*dIII restriction sites (*Hin*dIII sites that have not been mapped are not indicated). The contig segment located in between two *Hin*dIII restriction sites is defined as a "contig region" and is denoted by *bold numbers* under each contig. 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 *Hin*dIII fragments, but in some cases it contains a single *Hin*dIII fragment. Note that contig maps show the linear order of regions within each contig, but the linear order of *Hin*dIII fragments within each region has not been determined. Only the *Hin*dIII fragments that hybridized with the 22 studied sequences are indicated as *black lines* located above their respective contig region, and the *numbers* above these lines correspond to hybridizing sequences (genomic, cDNAs) whose characteristics are given in Table 1. "S" and "W" following the clone numbers indicate a strong or weak hybridization signal, respectively. *Underlined* sequence numbers indicate that the *Hin*dIII restriction fragment is the source of the corresponding genomic contig subclone that always shows the strongest ("S") hybridization of the pattern; underlining is not used for cDNAs. Although a 10-kbp scale bar is given, due to drawing constraints, the size of contig regions and hybridizing *Hin*dIII fragments is in some cases out of scale

oculated and 48-h inoculated with *R. leguminosarum* bv *viciae* 248 (Josey et al. 1979). The *M. truncatula* A17 root-hair cDNA library (Covitz et al. 1998) was screened by using digested BAC clone inserts or contig subclones as probes. The cDNA library screenings were done with Hybond-N+ membranes using the conditions recommended by the library manufacturer (Stratagene). DNA probes for the cDNA library screenings were labeled as described above.

Results and discussion

Distribution of studied sequences and their homologues in C1/C2 and C3

A *M. truncatula* A17 BAC library (Nam et al. 1999) was used to isolate the overlapping BACs composing the three contigs, named C1, C2 and C3, as previously described (refer to Materials and methods). Figure 1 shows a *Hin*dIII restriction map of the contigs in which the position of several *Hin*dIII restriction sites have been indicated. These restriction sites divide the contigs into so-called "contig regions" which have been numbered (bold numbers below the contig lines). To simplify the discussion, some contiguous contig regions have been grouped and indicated by a numbered line below the contig. Only the *Hin*dIII fragments within a contig region that hybridized with any of the 22 numbered sequences studied in this manuscript are indicated as a bar above the contig and the characteristics of these sequences are given in Table 1. Many *Hin*dIII fragments hybridized with more than one sequence. The studied sequences include several cDNA clones isolated from a *M. truncatula* and a pea root hair cDNA library, respectively. Some of these sequences were previously described (Gualtieri et al. 2002), while others are described in this manuscript. The cDNA clones have a prefix *Mtc* or *Psc*, for *M. truncatula* and *P. sativum*, respectively. We demonstrated that the genes corresponding with the *M. truncatula* cDNA clones are located in C1/C2 by comparing the hybridization of these cDNAs to *M. truncatula* 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.

Table 1 Properties of the 22 single- and low-copy studied sequences. The first column shows the numbers given in Fig. 1 to the studied sequences, whose names appear in the second column (prefixes *Mtc* and *Psc* indicate *M. truncatula* or *P. sativum* cDNAs, while Mtg means the *M. truncatula* genomic subclone). The third and fourth column indicates which sequences 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

Sequence number in Fig. 1	Contig subclone (size kbp)/cDNA	Location	Hybridization with contig/s	Blast homologies
1	*Mtc7x1*[e]	C2	C2	ND
2	*Mtc156*	C1	C1	ESTs from *M. truncatula*[d] Homology with receptor kinases/kinases
3	*Mtc411*[e]	C1	C1/C3	ND
4	*Mtc424*[a]	?	C1/C2	ESTs from *M. truncatula*[d] Homology with non-cyanogenic and cyanogenic beta glucosidase precursors, and beta glucosidases
5	*Mtc831*[f]	C1	C1	ESTs from *M. truncatula*[d] Homology with receptor kinases/kinases
6	*Mtc923*[f]	C1	C1	ESTs from *M. truncatula*[d], and several other organisms Phosphatase-like protein (*A. thaliana*) Putative membrane trafficking factor (*A. thaliana*)/Similarity to rat P47 protein
7	Mtg2114 (1.4 kbp)	C2	C2	No homology with ESTs Homology with putative protein (*A. thaliana*)
8	Mtg2.4 (2.4 kbp)	C2	C2/C3	ND
9	Mtg2128 (2.8 kbp)	C2	C2	Low homology with *M. truncatula* ESTs[d] Putative non-LTR retroelement reverse transcriptase (*Arabidopsis thaliana*)
10	Mtg63EB4(4 kbp)[f]	C2	C2	ND
11	Mtg1957 (2 kbp)[b]	C2	C2	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	ND
13	Mtg3552 (5.2 kbp)[e]	C2	C2/C3	(clone end sequence) No homology to ESTs LRR motif of *Cf-4* (*Lycopersicon*); putative disease resistance protein (*A. thaliana*)
14	Mtg3556 (5.6 kbp)[e]	C2	C1/C2	Homology *to M. truncatula*[d] *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)[c,e]	C1	C1	Homology with *M. truncatula* ESTs[d] Probable Transport ATP-binding protein MSBA, *Escherichia coli*
16	MtgHC2.8 (2.8 kbp)	C2	C2	ND
17	MtgLRR52 (0.43 kbp)[e]	C2	C2	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	ND
19	*PscLRR52*	Tightly linked to pea sym2	C2	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	Receptor-like protein kinases
21	MtgG5	C2	C2	No homology to ESTs Polyproteins from retrotransposons, reverse transcriptase, polyproteins (*Sorgum bicolor*) (*Oryza sativa*)
22	MtgG6	C2	C2	No homology to Ests Putative sulfite reductase (*E. coli*)

[a] The weak hybridization of *Mtc424* with a single fragment in C1 and C2 suggests that this cDNA is probably not encoded by any of these two contigs

[b] No cDNA hybridizing with this sequence was found in the *M. truncatula* root-hair library

[c] A single transcript hybridized with this clone on Northern blots containing 4-days old non-inoculated root total RNA

[d] These sequences had homology with cDNAs from several *Medicago* EST libraries

[e] These C1/C2 sequences correspond to RFLP markers tightly linked to *SYM2*

[f] These C1/C2 sequences correspond to RFLP markers that delimitated the *SYM2*-orthologous region to 350 kbp

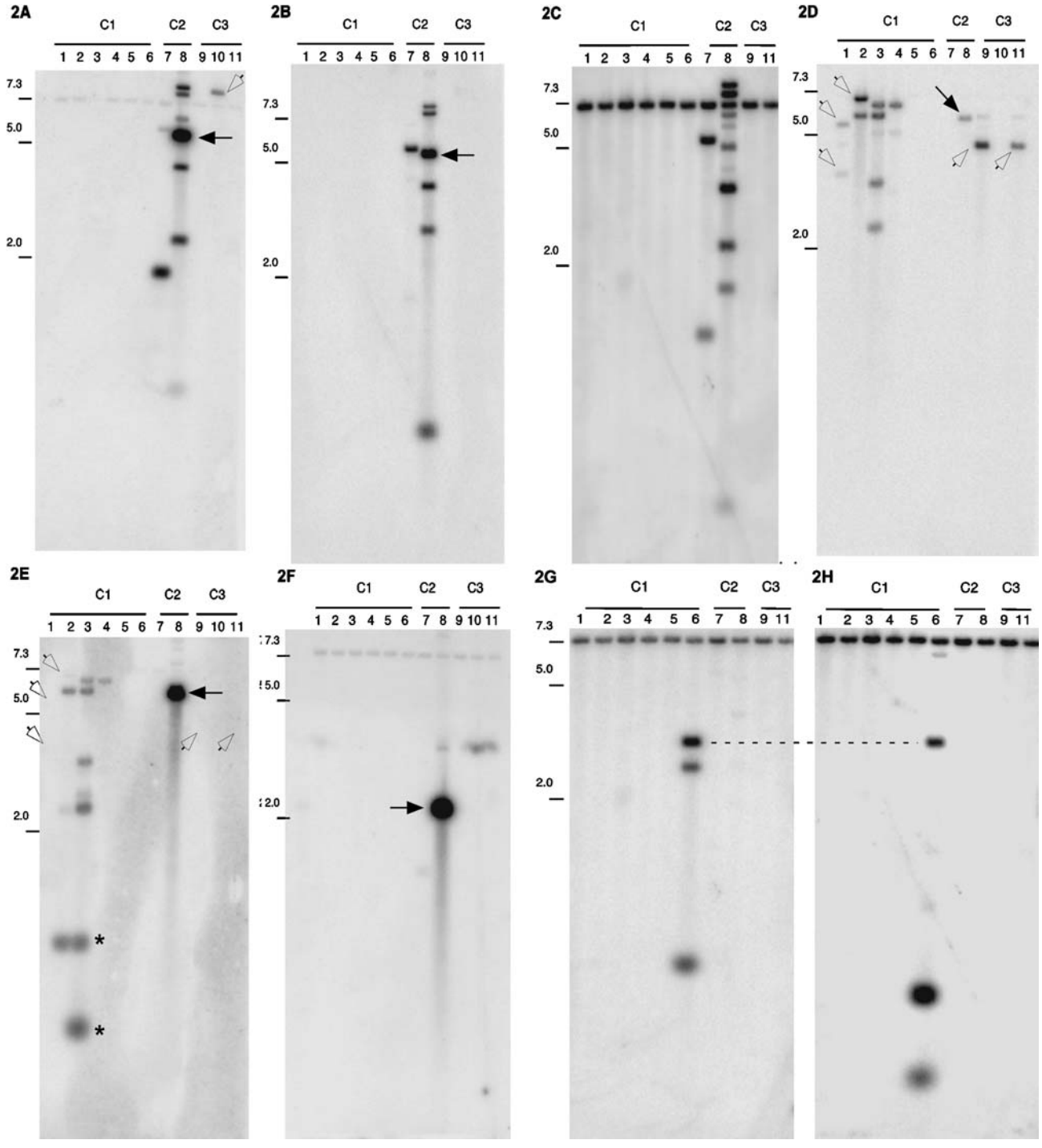


Fig. 2A–H Southern-blot hybridization of kinase-like and LRR-containing sequences with the contigs. BAC clones with minimal overlap and representing the complete sequence from the three contigs were digested with *Hin*dIII. *Lanes 1 to 6* corresponds to C1 DNA: *lane 1*, BAC 63010; *lane 2*, BAC 11I01; *lane 3*, BAC 20K04; *lane 4*, BAC 52010; *lane 5*, BAC 56F17; *lane 6*, BAC 59K07. *Lanes 7 and 8* correspond to C2 DNA: *lane 7*, BAC 21J05; *lane 8*, BAC 15B03. The 10-kbp overhanging sequence from the C2 end BAC 63C24 (Gualtieri et al. 2002) does not hybridize with any of the sequences described in this manuscript (data not shown). *Lanes 9 to 11* correspond to C3 DNA: *lane 9*, BAC 27H19; *lane 10*, BAC 30F10; *lane 11*, BAC 46I13. A size marker in kbp units is included at the left side of the blot pictures. The 7.3-kbp band observed in all the lanes of blots **A, C, F, G** and **H** corresponds to linear BAC vector DNA. Similar BAC DNA amounts were loaded in all lanes, but since *lanes 1 and 11* have a lower concentration, conclusions about the hybridization intensity of these lanes are based on several additional experiments not shown in this figure. The probes used in each blot were Mtg3552 in **A**, MtgLRR52 in **B**, *PscLRR52* in **C**, *PscW62-1* in **D**, Mtg3556 in **E**, Mtg1957 in **F**, *Mtc831* in **G**, *Mtc156* in **H**. *Black arrows* in **A, B, E**, and **F** indicate the fragment from which the probe originates. In **D** the *white arrowheads* show the C1 and C3 sequences most homologous to *PscW62-1* (weaker signals in *lanes 1 and 11* are due to a lower DNA concentration). The *black arrow* in **D** points to the Mtg3556 *Hin*dIII fragment that shows a weak hybridization with *PscW62-1*. In **E** the *white arrowheads* show that Mtg3556 does not hybridize with sequences most homologous to *PscW62-1* (indicated by a *white arrowhead* in **D**) and the *asterisks* indicate the fragments that hybridize with Mtg3556 but not with *PscW62-1*. **G** and **H** show that *Mtc831* and *Mtc156*, respectively, represent C1 specific kinase-like genes that do not hybridize with the *PscW62-1* homologues in C1 and C3, or with Mtg3556. In addition, these two cDNAs co-hybridize with a *Hin*dIII fragment of BAC 59K07 (indicated by a *dashed line*). *Mtc831* and *Mtc156* do not hybridize with BAC 30F10 (data not shown in **G** and **H**)

The majority of the studied sequences only hybridize with the contig *Hin*dIII fragments from which they originate (Fig. 1). The C1 *Mtc831* (Fig. 2G), *Mtc923*, *Mtc156* (Fig. 2H), Mtg2511, and the C2 Mtg63EB4, Mtg2114, Mtg2128, Mtg1957 (Fig. 2F), MtgHC2.8, MtgX3.2, are located on a single contig region (Fig. 1). The C2 *Mtc7x*-1, MtgG5 and MtgG6, are located on two adjacent regions of this contig (Fig. 1). MtgG28, is a subclone of C1 (region 6) that hybridized with two additional fragments within the same contig and these two fragments also hybridizes with *PscW62-1* (Fig. 1). The 5.2-kbp C2 genomic subclone Mtg3552 contains an end sequence, MtgLRR52 (0.43 kbp), which has homology with the LRR domain of the Cf4 and Cf9 tomato (*Lycopersicon*) disease resistance proteins (Parniske et al. 1997). RFLP analysis using *Eco*RI to digest pea DNA, showed that MtgLRR52 is tightly linked to *SYM2*. MtgLRR52 was used to isolate the pea cDNA *PscLRR52* from the pea root-hair cDNA library. RFLP mapping showed that *PscLRR52* is tightly linked to *SYM2* (Gualtieri 2001). *PscLRR52* also contains a sequence homologous to the LRR domains of Cf4 and Cf9. Mtg3552 (Fig. 2A), MtgLRR52 (Fig. 2B) and *PscLRR52* (Fig. 2C) hybridize with several fragments in C2 (Fig. 1, Fig. 2) (see next section).

Some sequences hybridize with other contigs apart from the one from which they originate (Fig. 1). The C2 5.6-kbp subclone Mtg3556 hybridizes with several fragments of C1 (Fig. 1, Fig. 2E). In addition to MtgLRR52, Mtg3552 contains a sequence homologous to *Mtc424*, a cDNA with high homology to beta glucosidases that was isolated by screening the *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. 1). 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.

Out of the 22 studied sequences only four hybridized with C3 (Fig. 1). *Mtc411* and Mtg3552 (Fig. 2A), and Mtg2.4 (C2-end genomic subclone), hybridize with C3 with low and high intensity, respectively (Fig. 1). In addition, the pea cDNA clone *PscW62-1* hybridizes with a similar high specificity to restriction fragments of C1 and C3, and with low specificity to a fragment in C2 (Fig. 1, Fig. 2D). These four sequences reveal homology between C1/C2 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

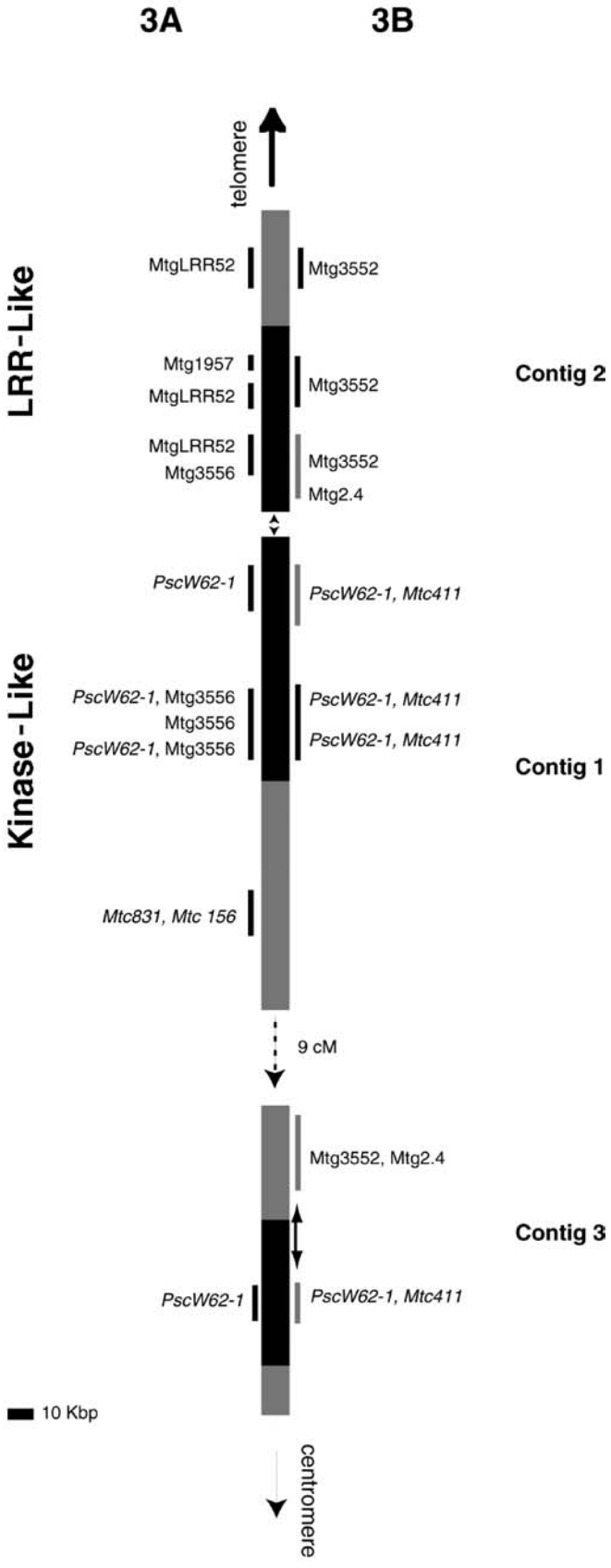
The LRR-containing sequences MtgLRR52 and *PscLRR52*, and the kinase-like sequences *PscW62-1* and Mtg3556, hybridized to many fragments that occur in several regions within the contigs (Fig. 1, Fig. 2B, C, D, E). On the other hand, the LRR-containing subclone Mtg1957 is a single-copy sequence in the *M. truncatula* genome and hybridizes to a single C2 contig region (Fig. 1, Fig. 2F).

Most kinase-like sequences are localized on C1 (Fig. 3A) and include the *PscW62-1* weakly and strongly hybridizing homologues (Fig. 2D), the Mtg3556 weakly hybridizing homologues (Fig. 2E), and the fragments containing the genes of the cDNAs *Mtc156* (Fig. 2H) and *Mtc831* (Fig. 2G). In C1, the *PscW62-1* homologues are located on two contig regions separated by about 50 kbp (Fig. 1 region 2 and region 6; and Fig. 3A). The C1 fragments that weakly hybridize with *PscW62-1* also hybridize with Mtg3556 (the C2 subclone containing kinase-like sequences), but this *M. truncatula* subclone hybridizes with two additional small C1 *Hin*dIII fragments located in region 6 (Fig. 1, indicated by asterisks in Fig. 2E) that are not detected by *PscW62-1*. This reveals that Mtg3556 contains unique kinase motives that are not present in *PscW62-1*. 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 E, white arrows). Thus, Mtg3556 is related to the C1 *PscW62-1* distant homologues but not 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. 1, region 12; Fig. 3A) form a separate group of kinase sequences that does not cross-hybridize with *PscW62-1* and Mtg3556. One *Hin*dIII fragment of BAC 59K07 hybridizes with both *Mtc831* (Fig. 2G) and *Mtc156* (Fig. 2H), indicating the presence of common sequence motives in these two cDNAs (shown by a dashed line connecting these *Hin*dIII fragments between Fig. 2G and Fig. 2H).

LRR-like sequences exclusively occur in C2 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 (Fig. 1, Fig. 3A). These three contig regions contain restriction fragments that hybridize with the *M. truncatula* genomic sequence MtgLRR52 that is homologous to the LRR domain of the Cf4/Cf9 proteins. Figure 2B shows that seven *Hin*dIII 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 Figs. 2A, B and C).

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 motives of the TMV N-like and RPS4 disease resistance proteins of *Arabiposis thaliana* (Table 1), and is located on a single *Hin*dIII 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 in C2: the



Cf4/Cf9 LRR motif-like sequences that hybridize with MtgLRR52, and the TMV N-like and RPS4 LRR motif-like sequence of Mtg1957. Neither C1 nor C3 contain homologues of these two types of LRR-like sequences.

Similar to *Arabidopsis* (The *Arabidopsis* Genome Initiative, 2000), kinase-like, LRR-containing, and other sequences from C1/C2 belong to families with several duplicated members. However, when some of these sequences are present in C3 it does so as a single-copy sequence. The latter is comparable to what is found in *Arabidopsis* where members of a given sequence family are present as a single-copy on a certain chromosomal segment whereas several copies are located within another chromosomal region. These copy number differences have been postulated to result from tandem duplication and/or gene loss after segmental duplication (The *Arabidopsis* Genome Initiative, 2000). Furthermore, it seems likely that sequence insertions in C1/C2 and/or deletions in C3 have occurred (see below).

Organization of homologous sequences within C1/C2 and C3

Figures 1 and 3B show that members of the four sequence families that hybridize with both C1/C2 and C3 are distributed in these contigs in a similar way. The C3 region 1 (about 50 kbp) contains the homologues of Mtg2.4 and Mtg3552, and this is separated by about 50 kbp from region 8 (about 20 kbp) containing the homologues of *PscW62-1* and *Mtc411* (Fig. 1, Fig. 3B). In C1, *Mtc411* as well as *PscW62-1* homologues are localized in region 2 (about 20 kbp) that is at a 10-kbp distance from the proximal contig-end; while in C2, Mtg3552 homologues are localized in region 10 (about 20 kbp) that is only 3-kbp apart from region 13 (about 7 kbp) containing Mtg2.4 (Fig. 1, Fig. 3B). Considering

Fig. 3A, B Distribution of kinase-like and LRR-containing sequences and sequences occurring in a duplicated segment of C1/C2 and C3. *Black/gray thick vertical lines* represent the contigs as in Fig. 1. C1 and C2 are drawn in the orientation as they occur in *M. truncatula* chromosome 5 (Gualtieri et al. 2002). Since the orientation of C3 in *M. truncatula* chromosome 5 is unknown, its orientation in this figure is arbitrary. The *short double arrowhead line* in between C1 and C2 indicates the sequence of C1/C2 not studied in this manuscript and corresponding to a region of about 10 kbp. Contig regions containing the cDNAs and subclones (and their homologues) presented in this figure are drawn as *lines aside the contigs*, and when these regions are contiguous or close to each other they are drawn as *a single line*. **A** (left side) Distribution of kinase and LRR-containing sequences in the three contigs. Contig regions containing these sequences are indicated as *black lines* at the left side of the contigs. **B** (right side) Distribution in C1/C2 and C3 of sequence family members corresponding to Mtg2.4, Mtg3552, *Mtc411* and their homologues, and homologues of *PscW62*-1. *Gray lines* aside the contigs represent contig regions containing a similar grouping of these sequences in both the 70-kbp C1/C2 and the 120-kbp C3 duplicated contig segments, while *black lines* represent other contig regions in C1/C2 containing additional homologues of these four sequences. The *double arrowhead line* aside C3 indicates the 50-kbp distance separating contig regions 1 and 8 as described in the text

that C1 and C2 were separated on chromosome 5 by about 10 kbp (Gualtieri et al. 2002), then the C2 region 10 and region 13 are together separated by about 20 kbp from the C1 region 2 (Fig. 1, Fig. 3B). Therefore, a *M. truncatula* chromosome-5 segment of about 70 kbp, which spans from C1 region 10 to C2 region 2, has a distribution/grouping of these four sequences that is similar to that of a 120-kbp segment of C3, which spans from region 1 to region 8. This indicates that these two chromosome-5 regions arose through a chromosomal segmental duplication, since the probability that these two regions originate independently is very low. However, since the orientation of C3 in chromosome 5 has not been determined it is not known whether these duplicated genomic regions have a direct or inverse chromosomal orientation. The duplication of these C1/C2 and C3 segments, separated by 9 cm within the same chromosome arm, is comparable to the duplications of large chromosomal segments within *Arabidopsis* chromosomes 1,4 and 5 (The *Arabidopsis* Genome Initiative, 2000) and within rice chromosome 9 (Foote et al. 1997).

It is worth noting that, as already mentioned, additional members of three (*PscW62-1*, *Mtc411*, and Mtg3552) of these four sequences families are present in C1/C2 beyond the duplicated segment of this contig (Fig. 1, Fig. 3B). These additional family members probably arose by duplications within C1/C2. The absence in C3 of these additional family members, particularly of *PscW62-1* and *Mtc411*, could be due to deletions in this region. Moreover, post-duplication deletions within C3 and/or insertions within the C1/C2 region could account for the absence and presence of MtgG5-, MtgG6-, MtgG28-, Mtg2511-, MtgLRR52- and Mtg3556-like sequences in these two duplicated chromosomal regions, respectively.

Our studies confirm that C1/C2 represents the *M. truncatula SYM2*-orthologous genomic region, since no homologues of the *SYM2*-linked RFLP markers MtgG28, Mtg3556, MtgLRR52 and *PscLRR52* are present in the 120-kbp C3 duplicated segment, but these markers or their homologues are present within the 70-kbp C1/C2 duplicated segment. In addition, the hybridization of Mtg3552 and *Mtc411* within the C3 duplicated segment is very weak as compared to their hybridization in C1/C2. However, the cloned sequences described in this manuscript represent only about 2% of the total sequence of the three contigs, and the complete sequencing of the contigs could reveal additional sequence homologues present both in C1/C2 and C3, which are linked to the pea *SYM2* gene. The latter becomes relevant if *M. truncatula* is used as a functional test-system to identify the *M. truncatula SYM2*-orthologue, because *SYM2* itself could belong to a family of homologous sequences present in both C1/C2 and C3.

Acknowledgements We are grateful to Sharon Long for generously providing the root hair-enriched *M. truncatula* cDNA library.

References

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12:1021–30

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

Boutin S, Young ND, Olson T, Yu ZH, Schoemaker RC, Vallejos CE (1995) Genome conservation among three legume genera detected with DNA markers. Genome 38:928–937

Cook DR (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 KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12: 637–646

Doyle JJ (1995) DNA data and legume phylogeny: a progress report. In: Crisp MD, Doyle JJ (eds) Advances in legume systematics. The Royal Botanic Gardens, Kew, Britain, pp 11–30

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, Downie J (1993) Resistance to nodulation of cv Afghanistan 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, Moore G (1997) Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. Genetics 147:801–807

Geurts R, Heidstra R, Hadri AE, 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

Gualtieri G (2001) *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. PhD thesis, Wageningen University (ISBN 90-5808-439-6), The Netherlands

Gualtieri G, Bisseling T (2000) The evolution of nodulation. In: Doyle JJ, Gaut BS (eds) Plant molecular evolution. Kluwer Academic Publishers, Dordrecht, The Netherlands. Plant Mol Biol 42, pp 181–194

Gualtieri G, Kulikova O, Limpens E, Kim D-J, Cook DR, Bisseling T, Geurts R (2002) Microsynteny between pea and *Medicago truncatula* in the SYM2 region. Plant Mol Biol 50(2):225–235

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

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

Josey DP, Beylon JL, Johnston AWB, Beringer JE (1979) Strain identification in *Rhizobium* using intrinsic antibiotic resistance. J Appl Bacteriol 46:343–350

Kozik A, Matvienko M, Scheres B, Paruvangada VG, Bisseling T, Van Kammen A, Ellis TH, LaRue TA, Weeden NF (1996) The pea early nodulin gene *PsENOD7* maps in the region of linkage group I containing *sym2* and leghaemoglobin. Plant Mol Biol 31:149–156

Kulikova O, Gualtieri G, Geurts R, Kim DJ, Cook DR, Huguet T, De Jong H, Franz PF, Bisseling T (2001) Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. Plant J 27:49–58

Menancio-Hautea D, Fatokun CA, Kumar L, Danesh D, Young ND (1993) Comparative genome analysis of mungbean (*Vigna radiata* L. Wilczek) and cowpea (*Vigna unguiculata* L.

Walpers) using RFLP mapping data. Theor Appl Genet 86: 797–810

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

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

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming R, Wright RJ (2000) Comparative genomics of plant chromosomes. Plant Cell 12:1523–1540

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

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

Temnykh SV, Kneen BE, Weeden NF, LaRue TA (1995a) Localization of *nod-3*, a gene conditioning hypernodulation, and identification of a novel translocation in *Pisum sativum* L. cv Rondo. J. Hered 86:303–305

Temnykh SV, Weeden NF, LaRue TA (1995b) *Sym*-2 and *nod*-3 are independent but closely linked genes influencing nodule development in pea. Pisum Genet 27:26–28

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

Weeden NF, Kneen BE, LaRue TA (1990) Genetic analysis of sym genes and other nodule-related genes in *Pisum sativum.* In: Gresshoff, Roth, Stacey, Newton (eds) Nitrogen fixation: achievements and objectives. Chapman and Hall, New York, pp 323–330

Weeden NF, Muehlbauer FJ, Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83:123–129