Theor Appl Genet (2004) 108:442–449
DOI 10.1007/s00122-003-1438-2

ORIGINAL PAPER

S. Stracke · S. Sato · N. Sandal · M. Koyama · T. Kaneko · S. Tabata · M. Parniske

Exploitation of colinear relationships between the genomes of *Lotus japonicus*, *Pisum sativum* and *Arabidopsis thaliana*, for positional cloning of a legume symbiosis gene

Received: 6 May 2003 / Accepted: 11 August 2003 / Published online: 14 October 2003

Abstract The *Lotus japonicus LjSYM2* gene, and the *Pisum sativum* orthologue *PsSYM19*, are required for the formation of nitrogen-fixing root nodules and arbuscular mycorrhiza. Here we describe the map-based cloning procedure leading to the isolation of both genes. Marker information from a classical AFLP marker-screen in *Lotus* was integrated with a comparative genomics approach, utilizing *Arabidopsis* genome sequence information and the pea genetic map. A network of gene-based markers linked in all three species was identified, suggesting local colinearity in the region around *LjSYM2/PsSYM19*. The closest AFLP marker was located just over 200 kb from the *LjSYM2* gene, the marker SHMT, which was converted from a marker on the pea map, was only 7.9 kb away. The *LjSYM2/PsSYM19* region corresponds to two duplicated segments of the *Arabidopsis* chromosomes AtII and AtIV. *Lotus* homologues of *Arabidopsis* genes within these segments were mapped to three clusters on LjI, LjII and LjVI, suggesting that during evolution the genomic segment surrounding *LjSYM2* has been subjected to duplication events. However, one marker, AUX-1, was identified based on colinearity between *Lotus* and *Arabidopsis* that mapped in physical proximity of the *LjSym2* gene.

Communicated by J.S. Heslop-Harrison

S. Stracke · M. Koyama · M. Parniske (✉)
The Sainsbury Laboratory,
John Innes Centre,
Colney Lane, Norwich NR4 7UH, UK
e-mail: martin.parniske@sainsbury-laboratory.ac.uk
Tel.: +44-1603-450249
Fax: +44-1603-450011

S. Sato · T. Kaneko · S. Tabata
Kazusa DNA Research Institute,
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

N. Sandal
Laboratory of Gene Expression,
Department of Molecular Biology,
University of Aarhus,
Gustav Wieds Vej 10, 8000 Aarhus C, Denmark

Present address:
M. Koyama, John Innes Centre,
Department of Crop Genetics,
Colney Lane, Norwich NR4 7UH, UK

Introduction

Nitrogen-fixing root nodules and arbuscular mycorrhiza are of prime agricultural importance. Mutants affected in these root symbioses have been isolated from important crop species such as soybean, pea and faba bean (Carroll et al. 1985; Duc and Messayer 1989; Duc et al. 1989). However, the isolation of the genes affected in these mutants is hindered by the large and complex genomes of these plants. Soybean, for instance, is an ancient polyploid and the genome comprises 1,115 Mb (Arumuganathan and Earle 1991; Shoemaker et al. 1996). Pea is diploid, with a genome size of 4,000 Mb (Michaelson et al. 1991). Model legumes such as *Lotus japonicus* and *Medicago truncatula* are believed to harbour the core set of legume genes, and their genomes are predicted to be to a large extent colinear to those of other legumes, but their genome content is estimated to be just under 500 Mb. Therefore, the development of comparative maps between model legumes and crop legumes should facilitate the cloning of agriculturally important genes from the crops, and the genetic maps of the model legumes are developing rapidly (Thoquet et al. 2002; Sandal et al. 2003).

Such considerations have motivated subsequent efforts to look for symbiotic mutants in diploid legumes with smaller genome size, such as *M. truncatula*, *L. japonicus* and *Phaseolus vulgaris* in order to accelerate the map-based cloning of the affected genes, with the ultimate aim to subsequently clone the orthologous genes from the crops by homology. Although this concept is promising, there are only few published examples in which a comparative approach has been used to clone a gene from a crop legume using the shortcut model legume.

Like *PsSYM19*, the orthologous gene from pea, the *Lotus LjSYM2* is required for the formation of nitrogen-fixing root nodules in symbiosis with rhizobia, as well as

the formation of arbuscular mycorrhiza (AM) in the interaction with AM-fungi (Duc et al. 1989; Schauser et al. 1998; Wegel et al. 1998; Schneider et al. 1999). The sequences of the *LjSYM2* and the *PsSYM19* genes have been published previously. They both encode receptor-like kinases which play a role in early steps of symbiotic signal transduction (Stracke et al. 2002). The *LjSYM2* gene has therefore been renamed to *Sym*biosis *r*eceptor-like *k*inase (SYMRK, Stracke et al. 2002). Another likely orthologue called NORK (nodule receptor kinase) has been cloned independently from *Medicago sativa* by Endre et al. (2002). These authors have also established that the *DMI2* gene of *M. truncatula* and the *PsSYM19* genes are likely to represent orthologues of *NORK*.

Here we describe the map-based cloning procedure that led to the isolation of the *Lotus LjSYM2* gene, and the subsequent cloning of the pea orthologue *PsSYM19* by homology. This study serves as an excellent example for the feasibility of the model legume strategy to clone genes from the more complex crop genomes. Map-based cloning relies on the generation of markers closely linked to the target gene. In this project we initially performed a classical screen for AFLP-markers (Vos et al. 1995) linked to *LjSYM2* in combination with bulked segregant analysis (Giovannoni et al. 1991; Michelmore et al. 1991), which resulted in linked markers. In addition, we developed *Lotus* markers using two comparative mapping strategies. To exploit marker resources developed for pea, we used gene-based (and hence potentially transferable) markers from the pea genetic map to test for possible colinearity between the *LjSYM2* and *PsSYM19* genes.

In addition, we demonstrate that it is possible to use the total genome sequence from *Arabidopsis* (The *Arabidopsis* Genome Initiative 2000) to develop closely linked markers in a legume. We first identified genomic segments in *Arabidopsis* related to the *LjSYM2/PsSYM19* region, and subsequently used individual genes within these *Arabidopsis* segments to design markers for *Lotus*. This strategy resulted in one marker in physical vicinity to *LjSYM2*. However, the vast majority of *Arabidopsis* markers mapped in *Lotus* revealed the presence of possibly duplicated segments within the *Lotus* genome. Despite the restrictions that such duplication impose on a comparative mapping strategy, we were able to exploit local colinearity between the genomes of *Arabidopsis* and the two legumes *Lotus* and pea to generate markers physically linked to the *LjSYM2* gene of *L. japonicus*.

Materials and methods

Plant material

Two populations from inter- and intra-specific crosses were generated for mapping of *LjSYM2* based on mutant alleles generated in a transposon/T-DNA-tagging attempts in the *L. japonicus* ecotype 'Gifu' (Schauser et al. 1998; J. Stougaard, unpublished). The F_1 plants of the interspecific cross between mutant cac41.5 (*sym2-3*) and *Lotus filicaulis* set very little seed, but an individual heterozygous F_2 plant (genotype 14F) set abundant seed and mutants segregated in the expected 3:1 ratio in the F_3 progeny of this plant (*L. japonicus sym2–3* × *L.filicaulis* F_3). The high level of polymorphism between the two *Lotus* species was useful for marker development. In total, 327 *L. japonicus sym2-3* × *L.filicaulis* F_3 mutants and 1,279 F_2 mutants of an intraspecific cross between mutant 282–287 (*sym2-1*) and ecotype Miyakojima MG20 (Kawaguchi et al. 2001) were screened for chromosomes carrying recombination events flanking the *LjSYM2* gene.

For phenotypic scoring of nodulation and arbuscular mycorrhiza *Lotus* seedlings were transferred into pots containing *Glomus intraradices*-colonized chive nurse plants in a substrate composed of baked expanded red clay particles (Biosorb, PDI Agrochemical Ltd., Essex, UK) and the seedlings were inoculated with *Mesorhizobium loti* strain R7A (Sullivan and Ronson 1998). After 3–5 weeks plants were scored visually for nodulation and microscopically for their arbuscular mycorrhiza phenotype.

DNA extraction

DNA for the AFLP procedure was extracted from the leaf tissue of *Lotus*. Three young emerging leaves were ground using liquid nitrogen and subsequently resuspended in 600 μl of nuclear lysis buffer (120 mM Tris, 30 mM EDTA, 1.2 M NaCl, 1.2% CTAB, 0.01% Na-bisulphite, pH 7.5) and 160 μl of 5% N-laurylsarcosine, and incubated at 65°C for at least 20 min. Samples were extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) followed by centrifugation at 16,000 rcf for 20 min. The nucleic acid was precipitated using 0.6 vol. of isopropanol and centrifugation at 16,000 rcf for 20 min. The resulting pellet was washed in 70% ethanol over-night and subsequently resuspended in TE.

DNA samples used in the recombinant screen were ground directly in 96-well plates in lysis buffer by employing a mixer mill (MM 300, Retsch, Germany). The extraction procedure applied is described by Steiner et al. (1995).

AFLP analysis

We employed bulked segregant analysis (BSA, Michelmore et al. 1991) to identify AFLP fragments linked to *LjSYM2*. Bulks were established from *L. japonicus sym-2–3* × *L.filicaulis* F_3 individuals and composed of eight plants that were homozygous for either mutant or wild-type *LjSYM2*. Plants homozygous for the mutation were identified by their mutant phenotype. Absence of mutant plants in the progeny identified homozygous wild-type plants.

In total, 110 *Pst*I+2/*Mse*I+3 enzyme-primer-combinations (including P11-P13, P15-P17, P19-P22 and M31-M51, M59, M61 according to Keygene, Wageningen, http://www.keygene.com/html/index_research.htm) were screened, using the radioactive AFLP procedure described by Vos et al. (1995) and the fluorescence-based AFLP protocol described by Dehmer et al. (2001). For fluorescence-based AFLP and all subsequent sequencing reactions, ABI sequencers (ABI PRISM 377 and 3700, PE Biosystems) were employed. AFLP marker fragments were gel-excised, re-amplified, cloned and sequenced. Markers were developed from AFLP fragments by extending fragment sequence using the genome walker strategy (Siebert et al. 1995; CLONTECH), and subsequent identification of sequence polymorphism between mapping parents.

Genome comparison

BLASTN searches of all predicted *Arabidopsis* cDNAs within the identified orthologous regions were performed against a database of *Lotus* EST sequences using the local blast function of the GCG package (GCG, Wisconsin Package, Version 10, Genetics Computer Group, 575 Science Drive, Madison, Wis., 53711 USA). Resulting hits were further evaluated and primer sets were designed based on alignment of *Lotus* ESTs with *Arabidopsis* genomic and cDNA (GCG; Staden 1996), such that the PCR product encom-

passed intron sequences. In the case of a polymorphism between the crossing parents, the gene was mapped in the general mapping population *L. filicaulis* × *L. japonicus*, employing the software JoinMap version 2.0 (Biometris, Wageningen) as described in Sandal et al. (2003).

Physical mapping

A TAC (transformation-competent artificial chromosome) library of genomic DNA of ecotype MG20 (Sato et al. 2001) and a BAC library carrying genomic DNA of ecotype Gifu (provided by N. Sandal) were screened with markers closely linked to *LjSYM2*. The selected TAC clones were sequenced according to the bridging shotgun method (Sato et al. 1997, 2001).

For survey sequencing, Gifu BAC clone 8M6 was sonicated and, after blunting the ends, fragments were ligated into *Sma*I-digested pUC18. In total, 96 subclones were end-sequenced. SNPs were detected after aligning the obtained Gifu sequences with the MG20 sequence represented in TAC clones LjT01N11 and LjT11E23.

Sequences are deposited in GenBank with the following accession numbers. Lotus SYMRK, AF492655; pea SYM19, AF491997. Lotus TAC clones: LjT11E23 (containing the Lotus SYMRK gene), AP004579; LjT01N11, AP004576; LjT10N22, AP004577; LjT17H01, AP004578.

Results

Lotus AFLP markers linked to *LjSYM2*

AFLPs in combination with bulked segregant analysis resulted in an initial set of linked markers. Based on AFLP data, up to 30% DNA sequence polymorphism has been observed between the species *L. japonicus* and *L. filicaulis*. To accelerate the identification of linked markers, homozygous plants for either mutant or wild-type *LjSYM2* have been selected from a *L. japonicus sym2-3* × *L.filicaulis* F_3 population and DNA pools generated. A screen of 110 enzyme-primer-combinations (EPC) identified two marker fragments, M37 (EPC P17M37) and P12 (EPC P12M51), which are distally located 6 cM and 0.3 cM, respectively, to *LjSYM2* (Fig. 1A). Marker fragment M37 showed homology to a Cyclin gene, for which a gene-specific primer set was generated (CYC-1). Sequencing of one end of a BAC clone (61A22) derived from the *L. japonicus* ecotype Gifu harbouring CYC-1 exhibited homology to an ubiquitin-like gene, from which a gene-specific marker was developed (UBL). Markers CYC-1, UBL and P12 were scored in the *L. filicaulis* × *L. japonicus* ecotype 'Gifu' general mapping population (Sandal et al. 2003), which together with fluorescence in situ hybridization of BAC clone 61A22 (Pedrosa et al. 2002) resulted in the positioning of the *LjSYM2* region on the long arm of linkage group LjII (Fig. 1A). The general *Lotus* map was inspected for additional markers in this region, and the SSR marker TM0008 was found to be linked at a distance of 1 cM (Fig. 1A).

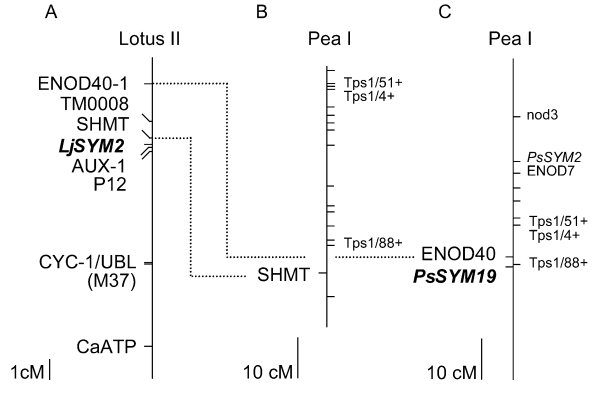

Fig. 1A–C Colinear relationship between the *LjSYM2* region of *L. japonicus* (**A**, Lotus II) and *PsSYM19* region of pea (Pea I), according to (**B**) Laucou et al. (1998) and (**C**) Schneider et al. (1999). *Dotted lines* join markers common between the maps

Marker development based on colinearity between *Lotus* and pea

Pea mutants affected in the *PsSYM19* gene (Schneider et al. 1999) share a very similar phenotype with *Lotus Ljsym2* mutants, in that the early stages of both root symbioses are impaired. Therefore, the *PsSYM19* gene was a candidate for the pea orthologue of *Lotus LjSYM2*. Markers SHMT and ENOD40 are closely linked to *PsSYM19* (Turner et al. 1993; Laucou et al. 1998; Schneider et al. 1999). Consequently, we developed *Lotus*-specific markers for both genes in order to establish possible colinearity between *Lotus* and pea in the *PsSYM19* region. For this purpose, *Lotus* ESTs with strong homology to pea *SHMT* were assembled into a contig and a gene-specific PCR (CAPS) marker was developed, which was mapped proximal to *LjSYM2*, at a distance of 0.19 cM (Fig. 1). Specific primer pairs corresponding to the two reported *ENOD40* genes of *Lotus* (Flemetakis et al. 2000) were designed. *ENOD40-1* was mapped on LjII, 3-cM proximal to *LjSYM2*, while *ENOD40-2* was positioned on LjIV (see Fig. 3). The close linkage of markers *PsSYM19*, SHMT and ENOD40 in pea, and *LjSYM2*, SHMT and ENOD40-1 in *Lotus*, strongly suggested that *PsSYM19* and *LjSYM2* reside in colinear regions of the genomes (Fig. 1). This, together with the similar phenotype of *Ljsym2* and *Pssym19* mutants, further supported the idea that these two genes might be an orthologous gene pair.

Marker development based on colinearity between *Lotus* and *Arabidopsis*

The gene-based markers CYC-1, UBL and SHMT are linked in *Lotus*. Inspection of the genome sequence of *Arabidopsis* indicated that homologues of these genes were linked in a duplicated region on chromosomes AtII and AtIV (Fig. 2). This finding suggested the presence of two segments in the genome of *Arabidopsis* that are colinear with the *LjSYM2* region. We therefore tried to develop additional markers in the target interval based on

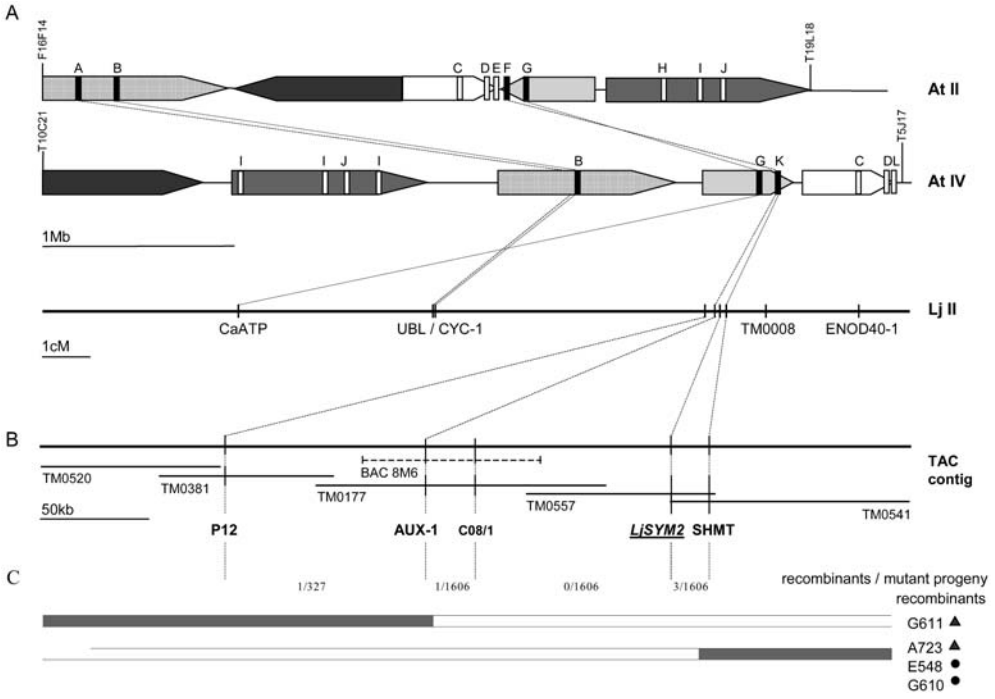

Fig. 2A–C Genetic and physical map of the *LjSYM2* region of *L. japonicus* and related segments of the *A. thaliana* genome. **A** Alignment of the *Lotus LjSYM2* region on LjII and the duplicated segments on *Arabidopsis* AtII and AtIV (according to Blanc et al. 2000). *Identically coloured arrows* represent the position and orientation of corresponding duplicated blocks on both At chromosomes. The relative positions of *Arabidopsis* genes for which *Lotus* markers could be developed are indicated by *rectangles* (A–L, gene annotations in Table 1) and those that mapped in the *LjSYM2* region are represented by black rectangles. **B** Physical map of a TAC contig based on the *LjSYM2*-flanking markers P12 (0.3 cM), AUX-1 (0.06 cM) and SHMT (0.19 cM) which comprise the *L. japonicus* ecotype MG-20 TAC clones LjT01N11, LjT11E23, LjT47I01, LjT10N22 and LjT17H01, spanning in total 405 kb. Marker C08/1 was derived from sequence alignment of TAC LjT01N11 and *L. japonicus* ecotype Gifu BAC 8M6 and was found to co-segregate with *LjSYM2*. **C** Numbers of recombinants identified among mutant progeny and identifiers of individual recombinants, originating from a *L. japonicus Ljsym2-3* × *L. filicaulis* F_3 (▲) or *L. japonicus Ljsym2-1* × MG-20 F_2 (●) population, are given below the markers. The physical target interval, flanked by AUX-1 and SHMT, has been delimited to 118 kb. The physical position of the *LjSYM2* gene is shown

both *Arabidopsis* genomic regions. This was complicated by the finding that the potentially orthologous *Arabidopsis* region (approximately 4 Mb) has not only been duplicated, but has subsequently undergone several rearrangements and differential gene loss, resulting in a complex patchwork of imperfectly duplicated blocks (Fig. 2A; Blanc et al. 2000). It was difficult to predict a priori, whether one of the two *Arabidopsis* genomic regions was more closely related to the *LjSYM2* region, and whether a similar duplication was present in *Lotus*. Therefore genes from both *Arabidopsis* chromosomal locations were used for analysis.

In order to develop corresponding *Lotus* markers, all predicted *Arabidopsis* cDNA sequences within the potentially orthologous regions on AtII and AtIV were extracted from the database, and a *Lotus* EST database was searched using BLASTN. The idea being that strong sequence similarity at the nucleotide level would indicate closer evolutionary relatedness than merely homology at the amino acid level. Highly homologous *Lotus* ESTs were identified for approximately 30% of the *Arabidopsis* cDNAs. In total, 26 of the resulting *Lotus* genes, corresponding to 20 different gene families, were selected for marker development. This resulted in 11 markers that were mapped (Table 1). The remaining primer combinations either gave no uniformly sized amplicon population or no detectable polymorphism between the mapping parents.

Two of the 11 mapped markers are located in the vicinity of *LjSYM2*. Marker CaATP is 10 cM distal to *LjSYM2*. The second marker, AUX-1, is separated from *LjSYM2* by only 0.06 cM (Fig. 1A).

Physical delimitation and cloning of *LjSYM2*

A total of 1,606 homozygous *Ljsym2* mutant plants were screened for the occurrence of recombinant chromosomes, using the markers UBL and TM0008 located on either side of the *LjSYM2* gene (Fig. 2). Recombinants identified were further genotyped using markers P12, SHMT and AUX-1. One homozygous *Ljsym2* mutant plant carrying a recombinant chromosome (genotype G611; Fig. 2C) was identified that was heterozygous at locus P12 (1 out of 327 mutants derived from the interspecific cross) and at locus AUX-1 (1 out of 1606

Table 1 Gene-based *Lotus* markers derived from *Arabidopsis* genomic segments which are related to the *LjSYM2/PsSYM19* region. The list of *Arabidopsis* genes located on a pair of duplicated segments on AtII and AtIV. Homologous *Lotus* ESTs were used as templates for marker generation and subsequently mapped in the mapping population *L. filicaulis* × *L. japonicus* ecotype 'Gifu' (Sandal et al. 2003)

1	2	3	4	5	6
A	At2g17200	none	Ubiquitin-like protein	UBL	II
B	At2g17620	At4g35620	Cyclin 2	CYC-1+2	I,II
C	At2g21600	At4g39220	AtRer1 A, 1B	Rer	I
D	At2g22240	At4g39800	Myo-inositol-1-phosphate	Myo-1+2	IV,VI
E	At2g22450	none	GTP cyclohydrolase	GTPcase	VI
F	At2g22670	none	Auxin-regulated protein	AUX-1+2	II,IV*)
G	At2g22950	At4g37640	CaIcium-ATPase	CaATP	II
H	At2g24580	none	Sarcosine oxidase	SO	I*)
I	At2g25060	At4g30590 At4g31840 At4g32490	Early nodulins, ENOD16/Putative protein	ENP	III*)
J	At2g25340	At4g32150	Synaptobrevin	SYN	I
K	none	At4g37930	Serine hydroxymethyltransferase	SHMT	II
L	none	At4g39980	2-dehydro-3-deoxyphosphoheptonate aldolase	2,3aldo2	VI

[1] Gene reference corresponding to Fig. 1
[2] Gene family member on AtII
[3] Gene family member on AtIV
[4] Annotation according to TAIR (<exref type="URL">http://www.Arabidopsis.org/</exref>)
[5] Name of mapped gene-specific marker in *Lotus*
[6] Chromosomal location of mapped markers in *Lotus* (Sandal et al. 2003 and *) this work)

mutants derived from the inter- and intra-specific crosses). For SHMT three recombinants out of the 1,606 mutants were detected (genotypes A723, E548, G610; Fig. 2C).

The screening of a TAC library, generated from ecotype MG20, with the flanking markers P12, AUX-1 and SHMT, resulted in a contig (LjT01N11, LjT11E23, LjT47I01, LjT10N22, LjT17H01) spanning 405 kb (Fig. 2B). The mutually exclusive recombination events described above demonstrated that the SHMT and AUX-1 markers flanked either side of the *LjSYM2* gene. Consequently, the target gene was physically delimited on the contig. The complete contig has been sequenced.

In an attempt to further reduce the physical target interval, we developed additional SNP markers between the closest flanking markers SHMT and AUX-1 (Fig. 2). To this end, both markers were used to screen a BAC library generated from ecotype Gifu. The screening with SHMT was not successful; however, one clone (BAC clone 8M6; Fig. 2B) corresponding to marker AUX-1 was identified and survey sequenced. The resulting Gifu sequences were aligned with the MG20 sequence, originating from TAC clone LjT01N11, to reveal SNPs. On average, one SNP per 1.7 kb was detected. A PCR marker (C08/1; Fig. 2B) which detected polymorphism in both mapping populations was used for genotyping recombinants. However, the marker C08/1 was co-segregating with *LjSYM2*, indicating that the recombination event in G611 had occurred between AUX-1 and C08/1 (Fig. 2C).

The genomic sequence of the target interval delimited by AUX-1 and SHMT spanned 118 kb, and was subjected to gene prediction and annotation. The subsequent sequencing of candidate genes from the *Ljsym2* mutants resulted in the identification of a gene encoding a receptor-like kinase, which was subsequently called SYMRK, for *sym*biosis *r*eceptor *k*inase (Stracke et al. 2002).

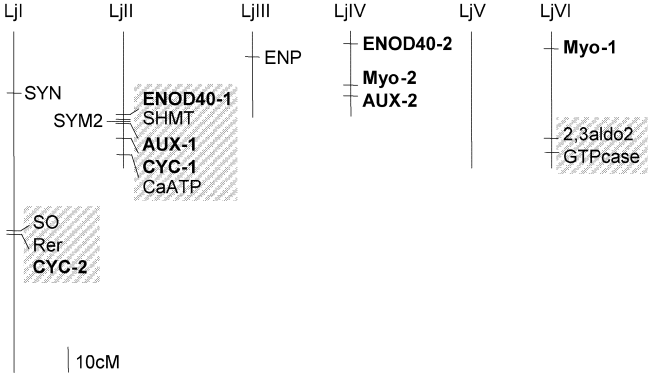

Fig. 3 Relative map position of *L. japonicus* genes with homology to the approximately 4 Mb duplicated segment in the *Arabidopsis* genome related to the *LjSYM2* region (see Fig. 2 and Table 1). Gene families with two mapped members are shown in *boldface*. Linked marker sets that indicate possible duplicated segments within the *Lotus* genome *are shaded*. Homologues of *ENOD40* and *LjSYM2* were not found in the corresponding *Arabidopsis* regions

Duplication of genomic segments in *Lotus*?

The 11 gene-specific *Lotus* markers, that were constructed based on the duplicated 4 Mb *Arabidopsis* AtII/AtIV region, revealed evidence for possible duplication events during the evolution of the *Lotus* genome (Figs. 2 and 3; Table 1). Nine of these markers did not map randomly in *Lotus*, but appeared to cluster in three distinct segments on LjI, LjII and LjVI (Fig. 3). Marker GTPcase and marker 2,3aldo2 are linked at a distance of 4 cM on the bottom of LjVI. The two corresponding *Arabidopsis* homologues are in similar positions on duplicated blocks on AtII and AtIV, respectively (genes 'E' and 'L'; Fig. 2A). The markers of the second cluster, CYC-2, Rer and SO are linked in *Lotus* on LjI. CYC-2 and Rer are co-segregating, while SO was mapped at a distance of 1 cM (Fig. 3). All three genes have homologues on AtII that are

separated by 1.8 Mb and 1.2 Mb, respectively (genes 'B', 'C' and 'H'; Fig. 2A). Counterparts of CYC-2 and Rer are also represented in the duplicated region on AtIV separated by 1.4 Mb (genes 'B' and 'C'; Fig. 2A).

A possible third cluster was revealed by the mapping of gene family members. Genome duplications could result in gene families with two or more members. In such cases the mapping of the corresponding gene family members could reveal evidence for segmental or total genome duplication.

We developed gene-specific markers for two members each of the AUX, ENOD40, CYC and Myo families in *Lotus*, resulting in the marker pairs AUX-1 and AUX-2, ENOD40-1 and ENOD40-2, CYC-1 and CYC-2, and Myo-1 and Myo-2 (Boldface in Fig. 3). Only 3-cM separate, AUX-1 and ENOD40-1 on LjII, while the corresponding gene family members AUX-2 and ENOD40-2 are linked on LjIV, but at a distance of 20 cM. Also the family members CYC-1 and CYC-2 map to linked clusters of markers on LjII and LjI respectively (Fig. 3). Myo-1 and Myo-2 are located on LjVI and LjIV, respectively, but neither shows close linkage to another gene-specific marker. It is possible that genes mapped outside the observed clusters are not the only members of this gene family in *Lotus* and that other members are represented in the potentially duplicated segments in *Lotus*.

In summary, the observed linkage between these markers suggests the possibility that LjII might have duplicated segments on LjI and LjVI. Traces of ancestral and recent genome scale duplications have been observed in *Arabidopsis* (Blanc et al. 2003) and other eukaryotes such as yeast and vertebrates (review in Wolfe and Li 2003); so it is not unexpected to obtain evidence for their occurrence in *Lotus*.

Discussion

Two strategies, AFLP marker-screening and comparative mapping, led to the identification of markers which were physically linked to the *Lotus LjSYM2* gene. Comparative mapping of *Lotus* was performed with two representatives of sister clades within the eurosids: pea a closely related legume (eurosid I), and *Arabidopsis*, a more distally related member of the eurosid II clade. While the closest AFLP marker (P12, Fig. 2B), obtained after screening a total of 110 EPC was at a distance of over 200 kb, the closest markers obtained through comparative genetics, were as close as 7.9 kb (SHMT, derived from the pea map) and 100 kb (AUX-1, derived from the *Arabidopsis* genome sequence) to *LjSYM2*. This result shows that the latter strategy can be at least as powerful in generating linked markers as an AFLP marker-screen. SHMT and ENOD40, two pea markers linked to *PsSYM19*, were converted into *Lotus* markers. Their position revealed a colinear arrangement of the ENOD40 *LjSYM2*/*PsSYM19* and SHMT genes in *Lotus* and pea. Therefore, the well-developed pea genetic map led to the development of two of the key markers that helped the isolation of the *Lotus* gene. Subsequently, the sequence corresponding to the *LjSYM2* orthologue from pea was isolated by homology. A probe derived from the *LjSYM2* 5′ region was used to screen a pea genomic DNA phage library. A lambda clone was isolated and entirely sequenced. This sequence information was used to design primers for RT-PCR amplification of a cDNA corresponding to a pea *LjSYM2* homologue. The predicted protein encoded by the pea cDNA was 82.8% overall identical to the predicted *Lotus* LjSYM2/LjSYMRK protein. The gene was sequenced in two pea lines carrying two different *PsSYM19* mutant alleles, and point mutations affecting the amino acid sequence could be identified (Stracke et al. 2002). Other recent examples for the successful application of the model legume genomes for the identification of crop genes include the NIN orthologue from pea (Borisov et al. 2003), the cloning of the HAR1 orthologue from soybean and pea (Krusell et al. 2002; Nishimura et al. 2002) as well as the ongoing comparative map-based cloning approach of the pea *PsSym2* gene (Gualtieri et al. 2002, Limpons et al. Science published online 28 August 03 [doi: 10.1126/science.1090074]).

Synteny between *Arabidopsis* and the legumes has been described (Grant et al. 2000). This was further substantiated by the observed micro-colinearity with *Arabidopsis* around the *LjHar1* locus in *Lotus* (Krusell et al. 2002; Nishimura et al 2002), in the *NORK* region of *Medicago* (Endre et al. 2002) and in the vicinity of *GmNARK* in soybean (Searle et al. 2003).

The cloning of the *LjSYM2* gene from *L. japonicus* demonstrates the successful exploitation of local synteny between the *Arabidopsis* and the legume genomes for positional cloning. We used *Arabidopsis* genes in a duplicated region of the genome to develop *Lotus* markers. After BLASTN analysis of all predicted genes in the two *Arabidopsis* segments of interest against a *Lotus* EST database, 26 *Lotus* ESTs were chosen for mapping. A total of 42% (11/26) of the tested *Lotus* ESTs could be converted into gene-specific markers and mapped in the *Lotus* genome. Of these, the proportion of markers linked to the target gene was 18% (2 out of 11 markers tested). The remaining gene-specific markers revealed the potential presence of duplicated segments in the *Lotus* genome (Fig. 3). Such duplications influence the success rate of developing markers linked to the target gene. If the events that gave rise to duplicated regions occurred after the legume-*Arabidopsis* lineages were separated, each of the duplicates should be equidistantly related to the *Arabidopsis* homologue; hence any marker developed from the *Arabidopsis* sequence should map with a similar likelihood to either of the duplicates. In our case, with at least three duplicated segments present, the theoretical hit rate per segment was only 33%. With the progress of the *L. japonicus* genome sequencing project (Cyranoski 2001; Sato et al. 2001) it will be possible to analyze the extent of duplication events in *Lotus*, and to answer the question whether such duplications in *Lotus* and *Arabidopsis* are the results of the same events

predating the divergence of the species or are more recent and hence independent events.

Detailed analysis of the *Arabidopsis* genome suggests the occurrence of at least two independent large-scale genome duplication events (Blanc et al. 2003; Ziolkowski et al. 2003). The most recent event was dated 25–40 MYA after the *Brassica* lineage split from the Legume lineage about 95 MYA (Gandolfo et al. 1998). The duplications of the AtII and AtIV segments displayed in Fig. 2 are estimated to originate from this recent event (Blanc et al. 2003). Of the genes, mapped in the vicinity of *LjSYM2*, only Cyclin2 and Calcium-ATPase are located on both duplicated *Arabidopsis* intervals. Useful web-resources for the identification of duplicated regions in Arabidopsis are the MIPS (http://mips.gsf.de/proj/thal/db/gv/rv/rv_frame.html) and Wolfe laboratory web-pages (http://wolfe.gen.tcd.ie/athal/dup), the latter featuring complementary material to Blanc et al. (2003).

In practice, gene duplications of any origin impose a serious limitation to a comparative mapping project. *Arabidopsis* has 25,498 predicted protein-encoding genes from 11,601 distinct gene families with one or more members. A third (35%) of these are unique genes and another third (37.4%) comprise large gene families with five or more members (The *Arabidopsis* Genome Initiative 2000). The major limitation of the colinearity approach is the presence of highly homologous but unlinked gene family members in the target plant genome. Ideally a sequence alignment of all members of a gene family in the target plant would be performed, and gene-specific primer pairs could be designed to map all members of the gene family, one of which is predicted to map in the target region. However, the availability of all sequences of a gene family requires either the genome to be completely sequenced, which would abolish the need for a colinearity-based cloning strategy, or an infinite depth of EST sequences from all organs and possible environmental conditions, which is not available for any plant species. Therefore, sequences available for this approach are genes which are well represented in the EST libraries, although they might not represent the gene copy linked to the target gene.

Another major limitation is unpredictable discontinuities in colinearity between the genomes studied. The frequency of such rearrangements remains to be determined, but the likelihood of encountering a breach of colinearity increases with the distance from the target gene and depends on the phylogenetic distance between the compared organisms (Vision et al. 2000).

In our study, the flanking gene-specific markers SHMT and UBL/CYC-1 spanning approximately 6 cM, allowed the *LjSYM2* region to be linked to potentially orthologous regions in *Arabidopsis* (Fig. 2). However, within this interval at least two breaches of colinearity have been observed, one of them within 1 cM from *LjSYM2* (data not shown). The closer the spacing between the initial set of gene-based markers, the more reliably orthologous regions in *Arabidopsis* can be defined. Despite the highlighted limitations, the degree of colinearity between *Lotus* and *Arabidopsis* in the *LjSYM2* region was sufficient for the successful development of linked markers in the physical vicinity of the target gene.

Acknowledgements We thank Thomas Lahaye for useful discussions and Satoko Yoshida, Kate Vickers and Julia Mattner for critical reading of the manuscript. We thank Jillian Perry, Jodie Pike and Julia Schmitz for help with recombinant screens, David Baker, Patrick Bovil and Nigel Hartley for shotgun sequencing of BAC 8M6 and Sara Perkins and Fiona Alderson for plant care. Research at the Sainsbury Laboratory is supported by the Gatsby Charitable Foundation.

References

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Blanc G, Barakat A, Guyot R, Cooke R, Delseny I (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. Plant Cell 12:1093–1101

Blanc G, Hiokamp K, Wolfe KH (2003) A recent polyploid superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Res 13:137–144

Borisov AY, Madsen LH, Tsyganov VE, Umehara Y, Voroshilova VA, Batagov AO, Sandal N, Mortensen A, Schauser L, Ellis N, Tikhonovich IA, Stougaard J (2003) The *Sym35* gene required for root nodule development in Pea is an ortholog of *Nin* from *Lotus japonicus*. Plant Physiol 131:1009–10017

Carroll BJ, McNeil DL, Gresshoff PM (1985). Isolation and properties of soybean glycine-max (L) Merr mutants that nodulate in the presence of high nitrate concentrations. Proc Natl Acad Sci USA 82:4162–4166

Cyranoski D (2001) Japanese legume project may help to fix the nitogen problem. Nature 409:272

Dehmer KJ (2001) Conclusions on the taxonomy of the *Solanum nigrum* complex by molecular analyses of IPK germplasm accessions. In: van den Berg RG, Barendse GWM, van der Weerden GM, Mariani C (eds) *Solanaceae* V: Advances in taxonomy and utilization. Proc Vth Int Solanaceae Conference, July 23–29th (2000), Nijmegen. Nijmegen University Press, pp 85–96

Duc G, Messager A (1989) A mutagenesis of pea (*Pisum sativum* L.) and the isolation of mutants for nodulation and nitrogen fixation. Plant Sci 60:207–213

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

Endre G, Kereszt A, Kevei Z, Mihacea S, Kaló P, Kiss GB (2002) A receptor kinase gene regulating symbiotic nodule development. Nature 417:962–966

Flemetakis E, Kavroulakis NEM, Quaedvlieg HP, Spaink HP, Dimou M, Roussis A, Katinakis P (2000) *L. japonicus* contains two distinct ENOD40 genes that are expressed in symbiotic, nonsymbiotic, and embryonic tissues. Mol Plant-Microbe Interact 13:987–994

Gandolfo MA, Nixon KC, Crepet WL (1998) A new fossil flower from the Turonian of New Jersey: *Dressiantha bicarpellata* gen. et sp. nov. (Capparales). Am J Bot 85:964–974

Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Res 19:6553–6558

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

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

Kawaguchi M, Motomura T, Imaizumi-Anraku H, Akao S, Kawasaki S (2001) Providing the basis for genomics in *L. japonicus*: the accessions Miyakojima and Gifu are appropriate crossing partners for genetic analyses. Mol Genet Genomics 266:157–166

Krusell L, Madsen LH, Sato S, Aubert G, Genua A, Szczyglowski K, Duc G, Kaneko T, Tabata S, de Bruijn F, Pajuelo E, Sandal N, Stougaard J (2002) Shoot control of root development and nodulation is mediated by a receptor-like kinase. Nature 420:422–426

Laucou V, Haurogne K, Ellis N, Rameau C (1998) Genetic mapping in pea. 1. RAPD-based genetic linkage map of *Pisum sativum*. Theor Appl Genet 97:905–915

Michaelson MJ, Price HJ, Ellison JR, Johnston JS (1991) Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow-cytometry. Am J Bot 78:183–188

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis—a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Nat Acad Sci USA 88:9828–9832

Nishimura R, Hayashi M, Wu GJ, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M, Harada K, Kawaguchi M (2002) HAR1 mediates systemic regulation of symbiotic organ development. Nature 420:426–429

Pedrosa A, Sandal N, Stougaard J, Schweizer D, Bachmair A (2002) Chromosomal map of the model legume *Lotus japonicus*. Genetics 161:1661–72

Sandal N, Krusell L, Radutoiu S, Olbryt M, Pedrosa A, Stracke S, Parniske M, Bachmair A, Ketelsen T, Stougaard J (2003) An AFLP-based genetic linkage map of the model legume *L. japonicus* developed from an interspecific F2 mapping population. Genetics (in press)

Sato S, Kotani H, Nakamura Y, Kaneko T, Asamizu E, Fukami M, Miyajima N, Tabata S (1997) Structural analysis of *Arabidopsis thaliana* chromosome 5. I. Sequence features of the 1.6-Mb regions covered by 20 physically assigned P1 clones. DNA Res 4:215–230

Sato S, Kaneko T, Nakamura Y, Asamizu E, Kato T, Tabata S (2001) Structural analysis of a *L. japonicus* genome. I. Sequence features and mapping of 56 TAC clones which cover the 5.4 Mb regions of the genome. DNA Res 8:311–318

Schauser L, Handberg N, Sandal N, Stiller J, Thykjaer T, Pajuelo E, Nielsen A, Stougaard J (1998) Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *L. japonicus*. Mol Gen Genet 259:414–423

Schneider A, Walker SA, Poyser S, Sagan M, Ellis THN, 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

Searle IR, Men AE, Laniya TS, Buzas DM, Iturbe-Ormaetxe I, Carroll BJ, Gresshoff PM (2003) Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. Science 299:109–112

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). Genet 144:329–338

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995) An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res 23:1087–1088.

Staden R (1996) The Staden sequence analysis package. Mol Biotechnol 5:233–241

Steiner JJ, Poklemba CJ, Fjellstrom RG, Elliott LF (1995) A rapid one-tube genomic DNA extraction process for PCR and RAPD analyses. Nucleic Acids Res 23:2569–2570

Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szcyglowski K, Parniske M (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature 417:959–962

Sullivan JT, Ronson CW (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc Natl Acad Sci USA 95:9059–9059

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

Thoquet P, Gherardi M, Journet EP, Kereszt A, Ane JM, Prosperi JM, Huguet T (2002) The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. BMC Plant Biol 2:1

Turner SR, Hellens R, Ireland R, Ellis N, Rawsthorne S (1993) The organization and expression of the genes encoding the mitochondrial glycine decarboxylase complex and serine hydroxymethyltransferase in Pea (*Pisum sativum*). Mol Gen Genetics 236:402–408

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. Science 290:2114–2117

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T et al. (1995) AFLP—a new technique for DNA-fingerprinting. Nucleic Acids Res 23:4407–4414

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

Wolfe KH, Li W-H (2003) Molecular evolution meets the genomics revolution. Nature Genet 33:255–265

Ziolkowski PA, Blanc G, Sadowski J (2003) Structural divergence of chromosomal segments that arose from successive duplication events in the *Arabidopsis* genome. Nucleic Acids Res 31:1339–1350