Theor Appl Genet (2011) 122:875–884
DOI 10.1007/s00122-010-1493-4

ORIGINAL PAPER

Glycine max non-nodulation locus *rj1*: a recombinogenic region encompassing a SNP in a lysine motif receptor-like kinase (*GmNFR1α*)

Woo Kyu Lee · Namhee Jeong · Arief Indrasumunar · Peter M. Gresshoff · Soon-Chun Jeong

Received: 17 May 2010 / Accepted: 4 November 2010 / Published online: 23 November 2010
© Springer-Verlag 2010

Abstract The *rj1* mutation of soybean is a simple recessive allele in a single line that arose as a spontaneous mutation in a population; it exhibits non-nodulation with virtually all *Bradyrhizobium* and *Sinorhizobium* strains. Here, we described fine genetic and physical mapping of the *rj1* locus on soybean chromosome 2. The initial mapping of the *rj1* locus using public markers indicated that A343.p2, a sequence-based marker that contains sequence similar to a part of the *LjNFR1* gene regulating nodule formation as a member of lysin motif-type receptor-like kinase (LYK) family, maps very close to or cosegregates with the *rj1* locus. The sequence of A343.p2 is 100% identical to parts of two BAC clone sequences (GM_WBb0002O19 and GM_WBb098N11) that contain three members of the *LYK* family. We analyzed the sequence contig (262 kbp) of the two BAC clones by resequencing and subsequent fine genetic and physical mapping. The results indicated that *rj1* is located in a gene-rich region with a recombination rate of 120 kbp/cM: several fold higher than the genome average. Among the *LYK* genes, *NFR1α* is most likely the gene encoded at the *Rj1* locus. The non-nodulating *rj1* allele was created by a single base-pair deletion that results in a premature stop codon. Taken together, the fine genetic and physical mapping of the *Rj1*-residing chromosomal region, combined with the unexpected observation of a putative recombination hotspot, allowed us to demonstrate that the *Rj1* locus most likely encodes the *NFR1α* gene.

Communicated by H. T. Nguyen.

W. K. Lee · N. Jeong · S.-C. Jeong (✉)
Bio-Evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Cheongwon, Chungbuk 363-883, Republic of Korea
e-mail: scjeong@kribb.re.kr

A. Indrasumunar · P. M. Gresshoff
ARC Centre of Excellence for Integrative Legume Research, The University of Queensland, Brisbane, St. Lucia, QLD 4072, Australia

Introduction

Nodule formation in legumes is tightly controlled (for a review, see Ferguson et al. 2010). Legume plants initiate nodule formation by exuding phenolic flavonoids that attract bacteria to the root and activate *nod* (nodulation) gene expression, leading to the production and secretion of strain-specific lipo-chito-oligosaccharides known as nod factors (Caetano-Anollés and Gresshoff 1991; Dénarié et al. 1996; Spaink 2000). The presence of compatible rhizobia species and their corresponding nod factor is generally sufficient to trigger nodule development. Loss-of-function mutations in gene(s) involved in nod factor perception exhibit non-nodulation phenotype. Recently, phenotyping screens combined with molecular genetics approaches such as map-based cloning have led to the isolation of many plant genes causatively involved in nodule formation. Pioneering work with *Lotus japonicus* and *Medicago truncatula* have identified two nod factor receptor (NFR) kinases, namely *LjNFR1* and *LjNFR5* (Madsen et al. 2003; Radutoiu et al. 2003), *MtLYK3/MtLYK4* and *MtNFP* (Limpens et al. 2003; Arrighi et al. 2006) which are functional orthologs *GmNFR5* in soybean (*Glycine max*; Indrasumunar et al. 2010a) and *PsSYM2A* and *PsSYM10* in *Pisum sativum*. These two NFR kinases are presumed to dimerize and be involved in nod factor binding; no biochemical proof for binding exists as yet. The NFR kinases are located in root epidermal/root hair

cells and contain an intracellular kinase domain, a transmembrane domain, and an extracellular portion containing lysin modif (LysM) domains.

Three genetic loci, *rj1*, *rj5*, and *rj6* (Williams and Lynch 1954; Mathews et al. 1989a; Pracht et al. 1993; Palmer et al. 2004), were reported to control initial nodule formation in soybean. Economically *G. max* was the most important leguminous grain crop in 2010 but harbors a larger and more complex paleoalloploid genome relative to the other model legumes, *L. japonicus* and *M. truncatula* (Schmutz et al. 2010). Plants homozygous for recessive alleles of each of these genes do not form nodules under normal inoculation conditions.

The *rj1* allele was discovered as a simple recessive gene in a single line likely arisen as a spontaneous mutation in a population from the cross of the cultivars 'Lincoln' and 'Richland' (Williams and Lynch 1954). The single line, later named 'T181', and a progeny line from the cross between T181 and its sister line, later named 'T201', were deposited to US Plant Germplasm System (http://www.ars-grin.gov/npgs/) and have been used to study the function of the *rj1* gene and as non-nodulating controls in many nodulation studies (e.g., Devine and Weber 1977; Jeong et al. 2006). Homozygous presence of the *rj1* allele in soybean conditions the non-nodulating response with virtually all *Bradyrhizobium* and *Sinorhizobium* strains (Devine et al. 1980). In field tests, with over 200,000 soybean plants containing homozygous *rj1*, less than one nodule per thousand plants was recovered (Devine 1984). The field results are consistent with the controlled experiments that demonstrated high titers of some strains of *Bradyrhizobium*, including *B. elkanii,* have the ability to induce a low frequency of nodulation on *rj1* plants (Murphy and Elkan 1965; Devine and Weber 1977; Devine and Kuykendall 1996).

A non-nodulating mutant from the ethylmethane sulfonate-mutagenized soybean cultivar Bragg, *nod49*, also segregates as a Mendelian recessive allele, is allelic to the naturally occurring *rj1*, and occasionally formed one to five functional nodules when inoculated with high titers of *B. japonicum* (Carroll et al. 1986; Mathews et al. 1987; Mathews et al. 1989a). Both the *rj1* and *nod49* mutants showed lack of root hair curling and infection thread growth (Mathews et al. 1987; Mathews et al. 1989b) and normal root exudates for *B. japonicum nod* gene induction (Mathews et al. 1989c), indicating that the non-nodulation mutations affected an early stage of nodulation. The *rj1* locus was reported to be linked to the *f* (fascinated) locus (approximately 40 centiMorgans (cM) from *rj1*) and the *Idh1* locus (approximately 26 cM from *rj1*) in the classical linkage group 11 (Devine et al. 1983; Hedges et al. 1990). The *Idh1* locus was assigned to the soybean molecular linkage group (MLG) D1b (soybean chromosome 2) (Cregan et al. 1999). However, a linkage relationship between *rj1* and molecular markers had not yet been established.

A fine molecular genetic map may provide an opportunity to pin down the genetic relationship between known nodulation genes functionally understood in model legumes and the *rj1* locus. Here, we described the results of the high-resolution mapping on MLG D1b of the *rj1* gene. Molecular markers, including public microsatellite markers (Song et al. 2004) and sequence-based markers developed from the soybean genome sequence (Zhang et al. 2007; Schmutz et al. 2010), were used to map *rj1*. Although we analyzed a relatively small population consisting of 241 F_2 individuals, the high recombination rate in the chromosomal region surrounding the *rj1* locus allowed us to demonstrate that the *NFR1α* gene, which is involved in plant recognition of symbiotic bacteria (Radutoiu et al. 2003), cosegregates with the *rj1* locus. In addition, we demonstrated that the mutation in the *rj1* allele of T181 and T201 is a 1-bp deletion, causing a frameshift that creates a stop codon leading to a premature translational stop.

Despite the recent research focus for a central role of recombination, through its influences on mutation and selection, in the evolution of plant genome evolution (for a review, see Gaut et al. 2007), features of recombination rates and patterns in the soybean genome remain to be poorly understood. Only a few studies reported the estimated ratios of genetic and physical distances in soybean as byproducts of fine genetic and physical mapping efforts for map-based gene cloning: 110 kbp/cM at the *rhg1* chromosomal region (Ruben et al. 2006), 810 kbp/cM at the *Rps1* region (Bhattacharyya et al. 2005), and 1,680 kbp/cM at the *Rpg1-b* region (Ashfield et al. 2003). Evaluation of genetic and physical map near *rj1* demonstrated that *rj1* is located in a gene-rich region with a recombination rate of 120 kbp/cM: several fold higher than the genome average.

Materials and methods

Plant materials

A population of 241 F_2 individuals generated from a cross between 'V94-5152' (*Rj1*) and T181 (*rj1*) (referred to as the VT181 population) were used to map *rj1*. T201 (*rj1*) and 'T202' (*Rj1*) lines were used as additional check plants.

Nodulation evaluation

The nodulation phenotype of each F_2 plant from the VT181 population was determined by visual inspection of the whole root at the harvest time of F_3 seeds. The nodulation

Table 1 Attributes of sequence-based markers linked to the *rj1* locus developed in this study

Marker name	Accession number and/or description of template sequence	Primer specificity	Primer specificity and Sequence (5- > 3′)	Predicted product size (bp)
SM333	EF533702; soybean BAC GM_WBb098N11; microsatellite repeats	F	TAATGCATGTCATTTGATAA	240 (W)
		R	AAATCCTACTAATCCCCTTA	
NFR1a_d	EF533702; resequencing of soybean BAC GM_WBb098N11; *NFR1α*	FTS	CTGTTGGCAATTTAGTTCC	226 (T), 227 (V)
		FVS	CTGTTGGCAATTTAGTTGT	
		CR	TGGTTACATGAGGATTATTT	
LYK2_CT	EF533702; resequencing of soybean BAC GM_WBb098N11; *LYK2*	CF	GCAGTATGGCTGTACCTTTC	258 (T and V)
		RTS	CAACTTTCTGGGGAGTTG	
		RVS	CAACTTTCTGGGGAGTTA	
BAC2O19_T	HN280452 (T) and HN280450 (V); 5′ end region of soybean BAC GM_WBb0002O19 (AC235173)	F	GCGTCGATGAAGAATTAAACAAT	154 (T), 159 (V)
		R	GCGATAAAGGAGTAAAGCAACTG	
NIN_indel	HN280453 (T) and HN280451 (V); soybean genome sequences homologous to *LjNIN* (AJ238956)	F	TCCTCTTAATAATCAACCAA	119 (T), 116 (V)
		R	AAATGTGACGGTACAGATAG	

BAC bacterial artificial chromosome, *CF* common forward, *CR* common reverse, *F* forward, *FTS* forward T181 specific, *FVS* forward V94-5152 specific, *R* reverse, *RTS* reverse T181 specific, *RVS* reverse V94-5152 specific, *T* T181, *V* V94-5152, *W* Williams 82

phenotype of most of F_2 plants from the VT181 population was verified by growing 15–20 $F_{2:3}$ plants in the field in the summer of 2009 or in a greenhouse in winters of 2008–2009 and 2009–2010. Seeds of V94-5152 and T181 were included as controls in the experiment. For field cultivation, seeds were planted in sites where soybean was cultivated in the previous growing season. For greenhouse cultivation, seeds were planted in 15-cm plastic pots containing a 1:1 mixture of commercial potting soil and top soil from soybean-cultivating field. The roots of plants were visually classified as either non-nodulating or nodulating 1 week after R1 reproductive growth stage (Fehr and Caviness 1977).

DNA extraction

Genomic DNA was extracted from fresh leaves of field- and greenhouse-grown individuals of the VT181 population and soybean parents, in accordance with the methods described by Saghai Maroof et al. (1984). RNA degradation was performed by adding RNase A to the extracted DNA tubes. For quick preparation from the $F_{2:3}$ line plants, soybean genomic DNA was isolated using a FastDNA® Kit in accordance with the manufacturer's protocols (MP Biomedicals, Solon, OH, USA). The total DNA was quantified with a PicoGreen dsDNA quantification kit (Molecular Probes) and a BioQ™–mini fluorometer (Bioneer, Daejon, Korea).

Marker genotyping

DNA from the VT181 population was genotyped using publicly available markers (Cregan et al. 1999; Song et al. 2004; Yang et al. 2008), as well as primers designed for this study, which were derived from microsatellite regions observed in two bacterial artificial chromosome (BAC) clones, GM_WBb0002O19 and GM_WBb098N11 (Zhang et al. 2007). Allele-specific primers were designed from the nucleotide substitution and 1-bp insertion/deletion (indel) sites between the genomic DNA sequences of V94-5152 and T181. For the microsatellite markers designed in this study, microsatellite regions were identified via visual inspection of microsatellite repeats of the targeted sequence region, and primers were designed with the web-based Primer3 platform (Rozen and Skaletsky 2000).

Allele-specific PCR and simple-sequence-length polymorphism markers were analyzed as previously described (Jeong and Saghai Maroof 2004). Typically, a DNA polymerase chain reaction consisted of 3 min at 94°C, then 34 cycles of 30 s at 94°C, 30 s at a specific annealing temperature followed by 30 s at 72°C. The final extension was 5 min at 72°C. Primers for marker analysis were custom-made by Bioneer (Daejeon, Korea) (Table 1). Publicly available marker primers were synthesized in accordance with the previously reported sequences (Song et al. 2004; Yang et al. 2008).

Sequence analysis

Parts of the soybean genomic DNA sequence homologous to that of *LjNIN* (GenBank accession no. AJ238956, which was previously reported to be linked to *LjNFR1* (Hoffmann et al. 2007), were determined in V94-5152 and T181 to generate a 3-bp indel marker. Parts of the soybean genomic DNA sequence corresponding to the 5′ end of GM_WBb0002O19 were determined in V94-5152 and T181 to generate a 3-bp indel marker.

Genomic DNA extracted from V94-5152 and T181 was used as template to determine the sequences of the *NFR1α* and *LYK2* and their intergenic region. As the size of this genic region reached 12.3 kbp, several overlapping 1- to 1.5-kbp fragments were independently amplified, and their sequences were determined from DNA products cut from agarose gels. When necessary, a given PCR product was subcloned into a plasmid and then multiple clones were sequenced. The resulting sequences were aligned against the sequences of two BAC clones, GM_WBb0002O19 and GM_WBb098N11. Sequence comparison and marker development procedures were conducted as described in previous studies (Jeong and Saghai Maroof 2004; Yang et al. 2008).

Linkage analysis

MapMaker 3.0b (Lander et al. 1987) was utilized for linkage analysis. The marker loci were grouped at a LOD of 5.0 and a maximum genetic distance of 37.5 cM. Marker order within a group was determined using the Three point and Order commands. The marker order was rechecked via repetitive use of the Compare command. The Kosambi centimorgan function was used to calculate map distances with error detection off.

Results

Linkage mapping of *rj1*

The segregation for nodulation of F_2 individuals from the VT181 population displayed a 3:1 ratio (nodulating:non-nodulating, $\chi^2 = 1.49$, $P = 0.22$). Microsatellite markers, which distribute over 20 MLGs of the soybean genome with an average distance 30–40 cM, were genotyped in a subset of 27 F_2 individuals of the VT181 population (data not shown). At the same time, soybean MLG D1b, on which *Idh1* was found to be distantly linked to *rj1* (Devine et al. 1983; Hedges et al. 1990), was more closely examined in 5–10 cM intervals. The results indicated that the microsatellite marker Satt459 was possibly closely linked to the *rj1* locus. To locate the *rj1* locus in the context of the current public soybean molecular genetic map, additional public microsatellite markers in the vicinity of the Satt459 chromosomal region (Song et al. 2004) were tested for polymorphisms between the two parental lines, V94-5152 and T181, of the mapping population. The resultant polymorphic markers were genotyped in the mapping population consisting of 241 F_2 plants. Eight markers, Satg001, Sat_202, Sat_198, Satt274, Sat_183, Satt703, Satt172, and Sat_139, which were polymorphic between these two parental lines, were mapped to both sides of the *rj1* locus on MLG D1b (Fig. 1).

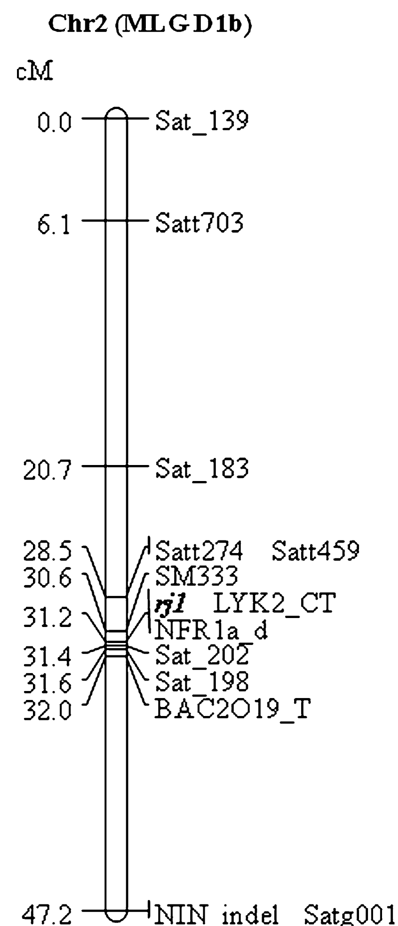

Fig. 1 Fine genetic map of the soybean chromosome 2 [molecular linkage group (MLG) D1b] in the vicinity of the soybean *Rj1* gene constructed in a segregating F_2 population derived from the cross of V92-5152 (*Rj1*) and T181 (*rj1*)

Simultaneously, to verify the nodulation phenotype as well as to determine homozygous or heterozygous nodulation genotypes of F_2 individuals that contained recombination events within 10 cM from the *rj1* locus, nodulation experiments were performed on 15–20 individuals of their $F_{2:3}$ families. Although we have not observed non-nodulating V94-5152 (*Rj1*) or nodulating T181 (*rj1*) plants in the present experimental conditions, individuals of $F_{2:3}$ families that formed 1–5 nodules, typically only at an isolated single lateral root out of the whole roots, were not counted in light of previous reports that *rj1* mutant showed occasional nodules when challenged with high doses of inoculum of specific strains (Devine and Weber 1977; Devine and Kuykendall 1996). Nodulation phenotypings of these $F_{2:3}$ families were repeated at least once. Sat_139, Satt703, Sat_183, Satt274, and Satt459 were mapped to the upper side of the *rj1* linkage map and Satg001 was mapped to the distal side (Fig. 1). Sat_198 cosegregated with *rj1*. However, Sat_202 was separated from *rj1* by one recombination event but could not be

mapped, because the recombinant individual was heterozygous only at the Sat_202 locus but contained the V94-5152 genotype at the surrounding markers. In other words, because the recombinant individual had recombination events on both sides of Sat_202, marker order among Satt459, *rj1*, Sat_202, and Sat_198 could not be determined without additional information including physical mapping or chromosome sequencing described below.

A fine genetic map of the *rj1* chromosomal region

Visual comparison of our microsatellite marker map to the genomewide map of Yang et al. (2008) suggested that A343.p2, a sequence-based marker that contains sequence homologous to a part of the *LjNFR1* gene, is expected to map very close to or cosegregate with the *rj1* locus. However, A343.p2 is not polymorphic in the VT181 population. A343.p1, a homeologous marker of A343.p2, was mapped between SL201 (11.9 cM away from A343.p1) and Satt126 (19.5 cM away from A343.p1) on MLG B2 (soybean chromosome 14) in our mapping population, as expected from the map of Yang et al. (2008). *LjNIN*, the gene encoding a putative transcription factor required for initial nodulation downstream of *LjNFR1* (Schauser et al. 1999), was reported to be linked to *LjNFR1* (Hoffmann et al. 2007). We sequenced a part of *LjNIN* from V94-5152 and T181. The V94-5152 sequence has one 3-base deletion compared with that of T181. This indel site was utilized to generate a codominant PCR marker, which was named NIN_indel. NIN_indel was mapped 16 cM away from the *rj1* locus, indicating that the soybean *NIN* gene is not responsible for the phenotype of *rj1*.

A BLAST search against the GenBank database indicated that sequence of A343.p2 from the soybean cultivar Hwangkeum is 100% identical to a part of two BAC clone sequences (GM_WBb098N11 and GM_WBb0002O19) from the soybean cultivar Williams 82 (Zhang et al. 2007). Sequences of the GM_WBb098N11 and GM_WBb0002O19 clones were aligned. The contig is 262,821 bp long and corresponds to positions 48,452,300-48,714,885 on chromosome 2 of the Williams 82 genome sequence (http://www.phytozome.net/soybean.php; Schmutz et al. 2010). The Williams 82 genome sequence was identical to the contig of the two BAC sequence except three ambiguous sequence components presented by N. This contig sequence contained polymorphic sites detected by markers Sat_198 and Sat_202. According to the soybean gene annotation database (accessible at Phytozome v5.0, http://www.phytozome.net; April 2010), the 262-kbp contig region contains 29 genes, suggesting that this region is highly gene-rich. Among these 30 genes, 3 genes (*Glyma02g43710.1*, *Glyma02g43750.1*, and *Glyma02g43760.1*) are members of the lysin motif-type receptor-like kinase (LYK) family including *LjNFR1* and *LjNFR5*, which function as receptors for bacterial nodulation factor (Madsen et al. 2003; Radutoiu et al. 2003; Zhang et al. 2007). *Glyma02g43710.1*, *Glyma02g43750.1*, and *Glyma02g43760.1* have been already named *GmLYK8*, *GmLYK2*, and *GmNFR1α*, respectively, by Zhang et al. (2007). According to the phylogeny of plant LYK proteins, *GmNFR1α* (*Glyma02g43760.1*) belongs to the *LjNFR1* ortholog subclade, *GmLYK2* belongs to the *LjNFR1* paralog subclade, and *GmLYK8* falls into an undefined subclade.

As the sequences of the two BAC clones were available, we attempted to delimit the *rj1* locus within the contig of the two BAC at sequence level. Four microsatellite regions near the ends of the contig were used to generate markers. One microsatellite marker, SM333, which is located between *NFR1α/LYK2* and *LYK8*, was mapped in our mapping population. SM333 is separated from *rj1* by three recombinations out of 241 F_2 individuals tested, thereby indicating that *LYK8* is not a candidate gene for *Rj1*. To generate a marker on the distal side of *rj1* where Sat_198 and Sat_202 are located on the contig, parts of the soybean genomic DNA sequence corresponding to 5′ end of GM_WBb0002O19 were determined in V94-5152 and T181 and a 4-bp indel marker BAC2O19_T was generated. BAC2O19_T is separated from *rj1* by two recombinations out of 241 F_2 individuals tested. Thus, the *rj1* locus mapped between BAC2O19_T and SM333. These results suggested that the gene responsible for the *rj1* locus was restricted to the physical region covered by the GM_WBb098N11-GM_WBb0002O19 contig (Fig. 2a).

Structural analysis of the *NFR1α-LYK2* genes

To identify polymorphic sites, and to investigate the structure of the soybean *NFR1α* and *LYK2* genes, we determined the 12.3 kbp sequence of the full *NFR1α-LYK2* genomic region from the V94-5152 and T181 lines, using primers designed from the GM_WBb0002O19-GM_WBb098N11 contig (Fig. 3). Sequence of the *LYK2* gene was 100% identical between Williams 82 (*Rj1*) and T181 (*rj1*). The *LYK2* sequence of T181 has only one single nucleotide polymorphism relative to that of V94-5152 (*Rj1*) in the first predicted intron (Fig. 3). Thus, this mutation appears to be neutral mutation to the nodulation phenotype. Using this sequence information, we designed specific forward and reverse primers in order to generate an allele-specific PCR marker, which was named LYK2_CT. LYK2_CT cosegregated with *rj1* in our maping population. Sequences of the intergenic region between *LYK2* and *NFR1α* were identical between V94-5152 and T181. Sequence of the *NFR1α* gene was 100% identical between V94-5152 and Williams 82. The *NFR1α* sequence of T181 (*rj1*) has only one 1-base deletion relative to that of V94-5152 in the fourth predicted exon (Fig. 3). Using this sequence

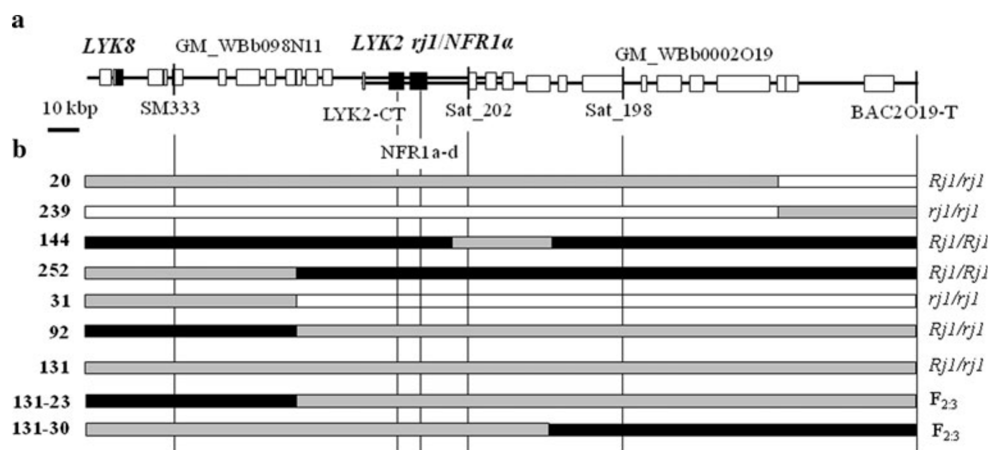

Fig. 2 Genetic and physical maps of *rj1*. **a** Locations of the lysin motif-type receptor-like kinase (LYK) genes (*NFR1α, LYK2*, and *LYK8*) and the mapped soybean markers on a BAC contig consisting of GM_WBb0002O19 (175,690 bp) and GM_WBb098N11 (132,396 bp). Sequence of the 3′ end of the BAC clone GM_WBb 098N11 is identical to sequence of 3′ end of GM_WBb0002O19. The predicted LYK genes are indicated by *filled rectangles* and the other genes from the soybean gene annotation database (accessible at Phytozome v5.0, http://www.phytozome.net; April 2010) by *open rectangles*. **b** Graphical genotypes of 6 F_2 recombinants and 2 F_3 recombinants derived from non-recombinant F_2 plant no. 131. *Black bars*, *white bars* and *gray bars* represent V94-5152 segments, T181 segments and heterozygous region, respectively. Nodulation genotypes determined by $F_{2:3}$ families of each recombinant are represented in the *right*

information, we designed allele-specific forward and reverse primers in order to generate an allele-specific PCR marker, named NFR1a_d. The *Rj1*/*rj1* genotypes in each recombinant completely coincided with the genotypes of NFR1a_d.

The *NFR1α* gene sequence between start and stop codons in V94-5152 was predicted to be 4,711 bp long with 11 introns. The predicted *NFR1α* cDNA encodes a putative protein of 623 amino acids. The 1-base deletion in T181 located at the fourth exon occurred at amino acid 257 of the NFR1α protein in V94-5152. The mutation results in a premature stop codon at residue 310 before the protein kinase domain. This would lead to the absence of the kinase domain of NRF1α. NFR1a_d was genotyped in T201 (*rj1*) and T202 (*Rj1*) (Fig. 4) and the NFR1a_d-containing sequence region was determined in T201 (data not shown). This shows that T201 contains the same mutation as T181. Sat_202, which could not be ordered on the genetic map, was located between NFR1a_d and Sat_198 based on the GM_WBb0002O19-GM_WBb098N11 contig. Taken together, the genetic, physical, and sequence relationship between the *NFR1α* mutation and the *rj1* alleles is strong evidence that the *Rj1* locus encodes the *NFR1α* gene.

Marker analysis of $F_{2:3}$ individuals derived from F_2 recombinants

The average ratio of genetic (cM) and physical (kbp) distances in soybean is estimated at approximately 480 kbp/cM based on the soybean genome size and total recombination distance (Arumuganathan and Earle 1991; Yang et al. 2008). Recombination rate between SM333 and BAC2O19_T and between SM333 and Sat_202 was estimated at 166 and 120 kbp/cM, 3- or 4-fold higher than genome average. This estimated recombination value include recombination events contributed by double crossovers in two F_2 recombinants, which would be rare in our mapping population consisting of 241 individuals (seeds of recombinant plants are available upon request). Recombination events occurred on both sides of Sat_202 in plant no. 144 (Fig. 2b) and on both sides of SM333 in plant no. 31 (data not shown). The genotypes of F_2 recombinants (no. 20, 31, 92, 239, and 252) having one recombination event between SM333 and BAC2O19_T were verified by the genotyping of markers SM333, LYK2_CT, NFR1a_d, Sat_202, Sat_198, and BAC2O19_T in 5–10 $F_{2:3}$ individuals. To verify the genotypes of the F_2 recombinant (no. 144) having double crossovers between SM333 and BAC2O19_T, additional $F_{2:3}$ individuals were genotyped (Fig. 2b). Collectively, marker genotypes of $F_{2:3}$ individuals derived from the F_2 recombinants spanning the *rj1* locus supported the genotypes of the F_2 recombinants. In addition, we genotyped 42 $F_{2:3}$ individuals of plant no. 131 heterozygous between SM333 and BAC2O19_T. Interestingly, we observed two novel $F_{2:3}$ recombinants arisen from the heterozygous region in the no. 131 (Fig. 2b). The results indicated that, although an exact spot could not be determined due to the small population size, a location between SM333 and BAC2O19_T is likely a recombination hotspot.

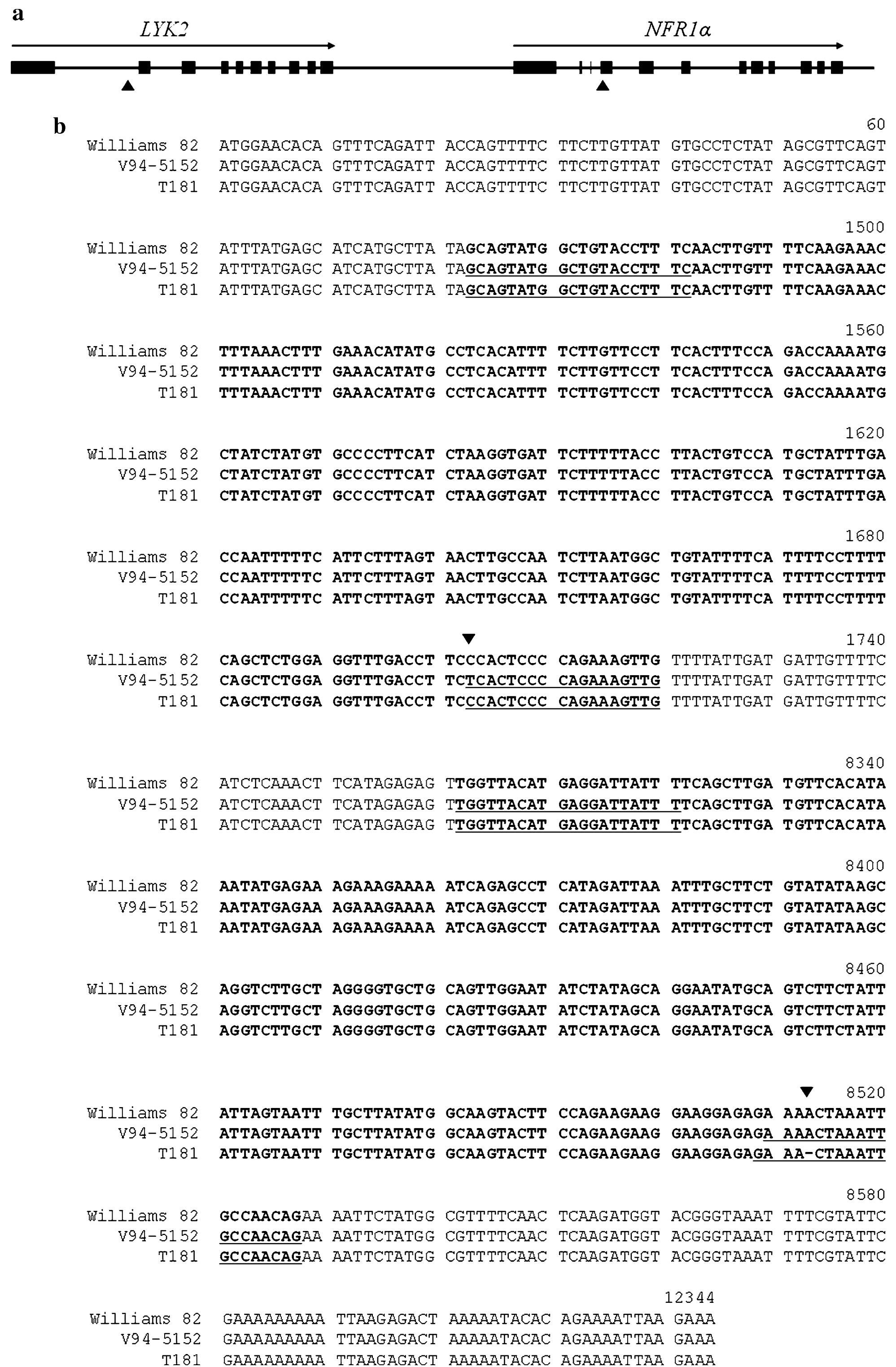

Fig. 3 Sequences of two tandem repeated lysin motif-type receptor-like kinase (LYK) genes located on soybean chromosome 2. **a** Exon–intron structure of the *LYK2* and *NFR1α* genes. Exons and introns were designated as described by Zhang et al. (2007). Exons are shown by *boxes*, introns and a 3′ UTR of NFR1α are shown by *lines*, and polymorphic sites between T181 and V94-5152 are designated with *filled triangles*. **b** Alignment of genomic *LYK2-NFR1α* sequences from Williams 82, V94-5152, and T181. Sequences from V94-5152 and T181 obtained in this study were aligned with the Williams sequence (GenBank accession no. EF533702). Parts of 5′ (60 b) and 3′ (44 b) ends and two parts encompassing polymorphic sites are shown in this alignment. Polymorphic sites detected are designated with *inverted filled triangles*. Amplified regions from mutation specific markers are shown in *bold*, with *horizontal bars* representing primer annealing location

Discussion

We have identified the genetic, physical, and sequence correlation between the *NFR1α* gene and the *Rj1* locus, one of the first gene conditioning nodulation in soybean to be defined genetically (Williams and Lynch 1954). Previous classical genetic mappings allowed us to quickly locate the *rj1* locus on a chromosomal region of soybean chromosome 2 (MLG D1b). Then, the sequence of marker A343.p2 on the genomewide molecular linkage map reported by Yang et al. (2008), combined with two completely sequenced BAC clones, permitted the identification

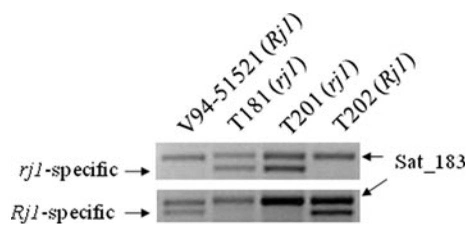

Fig. 4 Agarose gel assay of allele-specific polymerase chain reaction products that detect the one base deletion of *rj1* allele. *Upper bands* are Sat_183 bands and lower allele-specific bands

of a candidate gene for *rj1*. Resequencing and subsequent fine genetic mapping of the two BAC clones provided evidence that the *Rj1* locus encodes the *NFR1α* gene. The non-nodulating allele *rj1* appears to have been created by a base deletion that results in a translation product that terminates prematurely, rendering the *NFR1α* gene non-functional. However, this straightforward and seemingly simple finding was largely made possible because of the unusual recombination in the chromosomal region between SM333 and BAC2O19_T, part of which is likely a recombination hotspot. In this study, we have delimited the *rj1* locus within 100 kbp through the genotyping of 241 F_2 individuals several fold less than the numbers of individuals that have been used in recent map-based cloning studies of plant genes (e.g., Radutoiu et al. 2003; Xue et al. 2008; Watanabe et al. 2009).

Genome-wide patterns of recombination rates are available for several plant species including *Arabidopsis thaliana*, rice, maize, tomato, and wheat (Gaut et al. 2007). Although the extent of fine-scale variation along plant chromosomes is largely unexplored in plants, a few studies suggested that levels of recombination vary on kilobase scales as well as on chromosomal scales (e.g., Yao et al. 2002; Drouaud et al. 2006). We analyzed a small population that may make it difficult to conclude if the chromosomal region between SM333 and BAC2O19_T contain a recombination hotspot or not. However, in addition to the high recombination rate of 120 kbp/cM, several fold higher than the genome average, which is one of the highest recombination rates in the soybean genome reported thus far, the occurrence of double-crossover recombinants suggests that this region contains a recombination hotspot (Ng et al. 2010). It would be interesting to conduct a fine genetic and physical mapping research to delimit the predicted recombination hotspot within a few kilobases using a larger population in the future.

In the absence of transgenic complementation of the *rj1* allele with a functional copy of the *NFR1α* gene, the formal possibility exists that *Rj1* could encode another gene and that the frameshift mutation in the closely linked *NFR1α* gene in the two *rj1* genotypes is coincidental. Indrasumunar et al. (2010b) independently studied the function of *NFR1α* using a candidate-gene-cloning approach. In it, he demonstrated that *NFR1α*-transformed *rj1* and *nod49* mutants completely recovered the ability to form nodules. This result is complementary to our conclusion that, based on the fine genetic and physical maps, mutations in *NFR1α* are responsible for the phenotype of the *rj1* allele. Moreover, we have shown that the *LYK2* and *LYK8* genes, which are highly homologous to and located near *NFR1α*, are not candidate genes for the *Rj1* locus. Our molecular data on the structure of the *NFR1α* genes in the homozygous *Rj1* versus *rj1* genotypes agree with earlier positional cloning evidence (Radutoiu et al. 2003) that single nucleotide substitutions leading to premature stop codons in *NFR1* are responsible for the non-nodulating phenotype in *L. japonicus* mutants. Thus, the *Rj1* locus encodes a protein required for the regulation of nodulation in soybean. The association of the *rj1* locus with *NFR1α* also increases understanding of the genetic and molecular control of the nodulation pathway in soybean, an economically important crop.

The LysM is an ancient protein domain originally identified in bacterial autolycin (Joris et al. 1992; Bateman and Bycroft 2000). LysM proteins in plants species have attracted increasing attention since the identification of NFR1 and NFR5 in *L. japonicus* (Madsen et al. 2003; Radutoiu et al. 2003) and LysM-type receptor-like kinase3 (LYK3) and LYK4 from *M. truncatula* (Limpens et al. 2003). These four LysM proteins have an extracellular LysM domain, a single-pass transmembrane domain, and an intracellular Ser-Thr plant-specific protein kinase domain, reflecting the typical structure of plant kinase receptors. Genetic and molecular evidence suggests that they are receptors for the bacterial Nod factor, a GlcNAc lipo-chito-oligosaccharide with numerous modifications and structurally similar to peptidoglycan. Comprehensive characterization of plant LysM domains and molecular evolution and comparative genomics analysis of the plant *LYK* gene family revealed that plant LysM motifs are ancient, a minimum of six distinct types of LysM motifs exist in plant LysM kinase proteins, and the combination of LysM and receptor kinase domains occurs exclusively in plants (Zhang et al. 2007). Within plants, the gene family has evolved through both local and segmental duplications. Two of the most strongly conserved microsyntenies are observed between *GmNFR1α* and *GmNFR1β* and between *GmNFR5α* and *GmNFR5β*. *GmNFR1α* and *GmNFR1β* share a very high percentage sequence identity of 87%. Sequence analysis suggested that they are located at homeologous regions, originating during the soybean polyploidy that is estimated to have occurred approximately 15 million years ago (Schlueter et al. 2004; Pfeil et al. 2005), rather than remnants of older legume or dicot large-scale genomic duplications (Cannon et al. 2006). A homeologous

gene Gm*NFR1β* was mapped on MLG B2 (soybean chromosome 14). Thus, our mapping results suggested that *NFR1β* is not responsible for the phenotype of *rj1*. Indrasumunar et al. (2010b) showed that Gm*NFR1β* is non-functional with regards to nodulation.

Genetic and transformation studies of recessive loss-of-function non-nodulation mutants, the other study recently reported by Indrasumunar et al. (2010a), indicated that both *GmNFR5α* and *GmNFR5β* are functional and that *GmNFR5α and GmNFR5β* encode the *Rj5* and *Rj6* loci, respectively, which are the other two genetically identified genes conditioning initial nodule formation in soybean. The identification of all NFR kinases in soybean will facilitate the next phase of research, which includes elucidation of their downstream signaling pathway to further understand nodulation biology. Furthermore, detailed genetic and physical dissection of the predicted recombination hotspot residing at the *rj1* chromosomal region will facilitate the analysis of differential recombination rates along the plant chromosomes. The study may also provide a working hypothesis to address how this recessive gene arose as a spontaneous mutation in a breeding population.

Acknowledgments This work was supported by a grant from the BioGreen 21 Project (code no. 20080401034011 and 20050401034602), Rural Development Administration, the Republic of Korea, and, in part, by the KRIBB Research Initiative Program. PMG and AI thank the Australian Research Council for a Centre of Excellence grant.

References

Arrighi JF, Barre A, Ben Amor B, Bersoult A, Soriano LC, Mirabella R, de Carvalho-Niebel F, Journet EP, Ghérardi M, Huguet T, Geurts R, Dénarié J, Rougé P, Gough C (2006) The *Medicago truncatula* Lysine motif-receptor-like kinase gene family includes *NFP* and new nodule-expressed genes. Plant Physiol 142:265–279

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:211–215

Ashfield T, Bocian A, Held D, Henk AD, Marek LF, Danesh D, Peñuela S, Meksem K, Lightfoot DA, Young ND, Shoemaker RC, Innes RW (2003) Genetic and physical localization of the soybean *Rpg1-b* disease resistance gene reveals a complex locus containing several tightly linked families of NBS-LRR genes. Mol Plant Microbe Interact 16:817–826

Bateman A, Bycroft M (2000) The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). J Mol Biol 299:1113–1119

Bhattacharyya MK, Narayanan NN, Gao H, Santra DK, Salimath SS, Kasuga T, Liu Y, Espinosa B, Ellison L, Marek L, Shoemaker R, Gijzen M, Buzzell RI (2005) Identification of a large cluster of coiled coil-nucleotide binding site-leucine rich repeat-type genes from the *Rps1* region containing Phytophthora resistance genes in soybean. Theor Appl Genet 111:75–86

Caetano-Anollés G, Gresshoff PM (1991) Plant genetic control of nodulation. Annu Rev Microbiol 45:345–382

Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, Spannagl M, Monaghan E, Nicholson C, Humphray SJ, Schoof H, Mayer KF, Rogers J, Quétier F, Oldroyd GE, Debellé F, Cook DR, Retzel EF, Roe BA, Town CD, Tabata S, Van de Peer Y, Young ND (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc Natl Acad Sci USA 103:14959–14964

Carroll BJ, McNeil DL, Gresshoff PM (1986) Mutagenesis of soybean (*Glycine max* (L.) Merr.) and the isolation of non-nodulating mutants. Plant Sci 47:109–119

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean. Crop Sci 39:1464–1490

Dénarié J, Debelle F, Promé JC (1996) *Rhizobium* lipochitooligosaccharide nodulation factors: signalling molecules mediating recognition and morphogenesis. Annu Rev Biochem 65:503–535

Devine TE (1984) Genetics and breeding of nitrogen fixation. In: Alexander M (ed) Biological nitrogen fixation. Plenum Press, New York, pp 127–154

Devine TE, Kuykendall LD (1996) Host genetic control of symbiosis in soybean (*Glycine max* L.). Plant Soil 186:173–187

Devine TE, Weber DF (1977) Genetic specificity of nodulation. Euphytica 26:527–535

Devine TE, Kuykendall LD, Breithaupt BH (1980) Nodulation of soybeans carrying the nodulation-restrictive gene, *rj1*, by an incompatible *Rhizobium japonicum* strain upon mixed inoculation with a compatible strain. Can J Microbiol 26:179–182

Devine TE, Palmer JJ, Buzzell RI (1983) Analysis of genetic linkage in soybean. J Hered 74:457–460

Drouaud J, Camilleri C, Bourguignon PY, Canaguier A, Bérard A, Vezon D, Giancola S, Brunel D, Colot V, Prum B, Quesneville H, Mézard C (2006) Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination 'hot spots'. Genome Res 16:106–114

Fehr WR, Caviness CE (1977) Stages of soybean development, Iowa State University, Special Report 80. Iowa State University, Ames

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM (2010) Molecular analysis of legume nodule development and autoregulation. J Integr Plant Biol 52:61–76

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK (2007) Recombination: an underappreciated factor in the evolution of plant genomes. Nature Rev Genet 8:77–84

Hedges BR, Sellner JM, Devine TE, Palmer RG (1990) Assigning isocitrate dehydronase to linkage group 11 in soybean. Crop Sci 30:940–942

Hoffmann D, Jiang Q, Men A, Kinkema M, Gresshoff PM (2007) Nodulation deficiency caused by fast neutron mutagenesis of the model legume *Lotus japonicus*. J Plant Physiol 164:460–469

Indrasumunar A, Kereszt A, Searle I, Miyagi M, Li D, Nguyen CDT, Men A, Carroll BJ, Gresshoff PM (2010a) Inactivation of duplicated nod factor receptor 5 (NFR5) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean (*Glycine max* L. Merr.). Plant Cell Physiol 51:201–214

Indrasumunar A, Searle I, Lin MH, Kereszt A, Men A, Carroll BJ, Gresshoff PM (2010b) Nodulation factor receptor kinase 1α controls nodule organ number in soybean (*Glycine max* L. Merr.). Plant J. doi: 10.1111/j.1365-313X.2010.04398.x

Jeong SC, Saghai Maroof MA (2004) Detection and genotyping of SNPs tightly linked to two disease resistance loci, *Rsv1* and *Rsv3*, of soybean. Plant Breed 123:305–310

Jeong SC, Yang K, Park JY, Han KS, Yu S, Hwang T, Hur CG, Kim SH, Park PB, Kim HM, Park YI, Liu JR (2006) Structure, expression, and mapping of two nodule-specific genes identified by mining public soybean EST databases. Gene 383:71–80

Joris B, Englebert S, Chu CP, Kariyama R, Daneo-Moore L, Shockman GD, Ghuysen JM (1992) Modular design of the

Enterococcus hirae muramidase-2 and *Streptococcus faecalis* autolysin. FEMS Microbiol Lett 70:257–264

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA (1987) MapMaker: an interactive computer package for constructing genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Limpens E, Franken C, Smit P, Willemse J, Bisseling T, Geurts R (2003) LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. Science 302:630–633

Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J (2003) A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature 425:637–640

Mathews A, Carroll BJ, Gresshoff PM (1987) Characterisation of non-nodulation mutants of soybean (*Glycine max* (L.) Merr.): *Bradyrhizobium* effects and absence of root hair curling. J Plant Physiol 131:349–361

Mathews A, Carroll BJ, Gresshoff PM (1989a) A new nonnodulation gene in soybean. J Hered 80:357–360

Mathews A, Carroll BJ, Gresshoff PM (1989b) Development of *Bradyrhizobium* infections in supernodulating and non-nodulating mutants of soybean (*Glycine max* [L] Merrill). Protoplasma 150:40–47

Mathews A, Kosslak RM, Sengupta-Gopalan C, Appelbaum ER, Carroll BJ, Gresshoff PM (1989c) Biological characterization of root exduates and extracts from nonnodulating and supernodulating soybean mutants. Mol Plant Microbe Interact 2:283–290

Murphy SG, Elkan GH (1965) Nitrogen metabolism of some strains of *Rhizobium japonicum* having different nodulating capacities. Can J Microbiol 11:1039–1041

Ng S-K, Lo WS, Pun FW, Zhao C, Yu Z, Chen J, Tong KL, Xu Z, Tsang SY, Yang Q, Yu W, Nimgaonkar V, Stöber G, Harano M, Xue H (2010) A recombination hotspot in a schizophrenia-associated region of *GABRB2*. PLoS ONE 5:e9547

Palmer RG, Pfeiffer TW, Buss GR, Kilen TC (2004) Qualitative genetics. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses, 3rd edn. Argon. Monogr. 16. ASA, CSSA, and SSSA, Madison, pp 137–214

Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. Syst Biol 54:441–454

Pracht JE, Nickell CD, Harper JE (1993) Genes controlling nodulation in soybean: *Rj5* and *Rj6*. Crop Sci 33:711–713

Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Grønlund M, Sato S, Nakamura Y, Tabata S, Sandal N, Stougaard J (2003) Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature 425:585–592

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Ruben E, Jamai A, Afzal J, Njiti VN, Triwitayakorn K, Iqbal MJ, Yaegashi S, Bashir R, Kazi S, Arelli P, Town CD, Ishihara H, Meksem K, Lightfoot DA (2006) Genomic analysis of the *rhg1* locus: candidate genes that underlie soybean resistance to the cyst nematode. Mol Gen Genomics 276:503–516

Saghai Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Schauser L, Roussis A, Stiller J, Stougaard J (1999) A plant regulator controlling development of symbiotic root nodules. Nature 402:191–195

Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC (2004) Mining EST databases to resolve evolutionary events in major crop species. Genome 47:868–876

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Spaink HP (2000) Root nodulation and infection factors produced by rhizobial bacteria. Annu Rev Microbiol 54:257–288

Watanabe S, Hideshima R, Xia Z, Tsubokura Y, Sato S, Nakamoto Y, Yamanaka N, Takahashi R, Ishimoto M, Anai T, Tabata S, Harada K (2009) Map-based cloning of the gene associated with the soybean maturity locus E3. Genetics 182(4):1251–1262

Williams LF, Lynch DL (1954) Inheritance of a nonnodulating character in the soybean. Agron J 46:28–29

Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Xu S, Li X, Zhang Q (2008) Natural variation in Ghd7 is an important regulator of heading data and yield potential in rice. Nature Genet 40:761–767

Yang K, Moon JK, Jeong N, Back K, Kim HM, Jeong SC (2008) Genome structure in soybean revealed by a genomewide genetic map constructed from a single population. Genomics 92:52–59

Yao H, Zhou Q, Li J, Smith H, Yandeau M, Nikolau BJ, Schnable PS (2002) Molecular characterization of meiotic recombination across the 140-kbp multigenic *a1–sh2* interval of maize. Proc Natl Acad Sci USA 99:6157–6162

Zhang XC, Wu X, Findley S, Wan J, Libault M, Nguyen HT, Cannon SB, Stacey G (2007) Molecular evolution of lysine motif-type receptor-like kinases in plants. Plant Physiol 144:623–636