MPMI Vol. 23, No. 4, 2010, pp. 510–521. doi:10.1094/MPMI-23-4-0510.  *e*-Xtra*

# Evolution and Regulation of the *Lotus japonicus LysM Receptor* Gene Family

**Gitte Vestergaard Lohmann,[1] Yoshikazu Shimoda,[2,3] Mette Wibroe Nielsen,[1] Frank Grønlund Jørgensen,[4] Christina Grossmann,[1] Niels Sandal,[1] Kirsten Sørensen,[1] Søren Thirup,[1] Lene Heegaard Madsen,[1] Satoshi Tabata,[2] Shusei Sato,[2] Jens Stougaard,[1] and Simona Radutoiu[1]**

[1]Centre for Carbohydrate Recognition and Signalling, MBI, University of Aarhus, Gustav Wieds Vej 10, Aarhus C, Denmark; [2] Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan; [3]Plant-Microbe Interaction Research Unit, National Institute of Agrobiological Sciences (NIAS), 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan; [4]Bioinformatics Research Center (BiRC), University of Aarhus, C.F. Møllers Alle, Bldg. 1110, DK-8000, Aarhus C, Denmark

Submitted 26 October 2009. Accepted 3 December 2009.

**LysM receptor kinases were identified as receptors of acylated chitin (Nod factors) or chitin produced by plant-interacting microbes. Here, we present the identification and characterization of the LysM receptor kinase gene (*Lys*) family (17 members) in *Lotus japonicus*. Comprehensive phylogenetic analysis revealed a correlation between *Lys* gene structure and phylogeny. Further mapping coupled with sequence-based anchoring on the genome showed that the family has probably expanded by a combination of tandem and segmental duplication events. Using a sliding-window approach, we identified distinct regions in the LysM and kinase domains of recently diverged *Lys* genes where positive selection may have shaped ligand interaction. Interestingly, in the case of NFR5 and its closest paralog, LYS11, one of these regions coincides with the predicted Nod-factor binding groove and the suggested specificity determining area of the second LysM domain. One hypothesis for the evolutionary diversification of this receptor family in legumes is their unique capacity to decipher various structures of chitin-derived molecules produced by an extended spectrum of interacting organisms: symbiotic, associative, endophytic, and parasitic. In a detailed expression analysis, we found several *Lotus Lys* genes regulated not only during the symbiotic association with *Mesorhizobium loti* but also in response to chitin treatment.**

Plants are highly dependent on the abiotic and biotic environment. Their growth and development is affected by the availability of water, light, and nutrients and their interaction with bacteria, fungi, viruses, nematodes, and insects. Interactions causing disease limit the plant's chances for optimized growth; however, symbiotic interactions between plants and mycorrhizal fungi or rhizobial bacteria can promote plant growth when soil nutrients are scarce. The symbiotic interaction between legume plants and rhizobial bacteria is a highly controlled process involving Lysin motif-containing receptor kinases (LysM-RLK) that perceive secreted bacterial lipochitooligosaccharides (Nod factor) signals (Limpens et al. 2003; Madsen et al. 2003; Radutoiu et al. 2003). In *Lotus japonicus*, two LysM-RLK, NFR1 and NFR5, initiate a signal transduction cascade after perception of Nod factors (Madsen et al. 2003; Radutoiu et al. 2003, 2007). Genetic studies of receptor mutants suggest that NFR1 and NFR5 may form a receptor complex or that two independent signal-transduction pathways converge shortly after signal perception. Expression of *LjNfr1* and *LjNfr5* in *Medicago truncatula* showed that both genes are required for extending the *Medicago truncatula* host range to include *Mesorhizobium loti* and a modified *Rhizobium leguminosarum* DZL, which was engineered to infect *L. japonicus* (Radutoiu et al. 2007). Similar observations were obtained in transgenic *L. filicaulis*, where the DZL strain nodulates plants transformed with both *LjNfr1* and *LjNfr5*. These experiments, together with domain swaps between *L. japonicus* and *L. filicaulis* proteins, show that Nod factor receptors (NFR) are responsible for host specificity and that the second LysM motif of NFR5 is a major determinant in Nod factor recognition (Radutoiu et al. 2007).

In addition to the Nod factor receptors in legumes, LysM-RLK have been identified in nonlegume plants, including *Arabidopsis thaliana* (five genes), *Populus trichocarpa* (11 genes) and rice (*Oryza sativa*) (eight genes) (Shiu et al. 2004, Zhang et al. 2007). In *Arabidopsis*, one of these, CERK1, was shown to be essential for chitin signaling and resistance to fungal infection (Miya et al. 2007, Wan et al. 2008). Protein localization and kinase activity experiments indicate that CERK1 perceives fungal chitin at the cell surface and, subsequently, initiates a signal cascade through the intracellular kinase domain leading to a pathogen-associated molecular pattern (PAMP)-triggered immunity response (Miya et al. 2007). Recently, CERK1 was shown to be a target of the bacterial AvrPtoB effector protein, leading to bacterial virulence in *Arabidopsis* (Gimenez-Ibanez et al. 2009). In rice, a plasma membrane glycoprotein named CEBiP, which contains two extracellular LysM motifs but lacks an intracellular kinase, plays a role in chitin perception and signal transduction (Kaku et al. 2006). Therefore, LysM domains appear to be involved in recognition of chitin and chitin derivatives such as the Nod factors and, possibly, also related ligands.

The presence of *LysM-RLK* genes in nonlegume species suggests that some of the legume *LysM-RLK* genes are involved in processes other than the symbiotic association between legumes and rhizobia. Interestingly, the *Medicago truncatula Lyr1* gene, which is the closest paralog of *Medicago truncatula Nfp*, was recently shown to be upregulated upon arbuscular mycorrhiza colonization of *Medicago truncatula* roots (Gomez et al. 2009).

G.-V. Lohmann and Y. Shimoda contributed equally to this work.

Corresponding author: Simona Radutoiu; E-mail: sir@mb.au.dk

*The *e*-Xtra logo stands for "electronic extra" and indicates three supplemental figures and three supplemental tables are published online.

In this study, we characterized the *LysM-RLK* (*Lys*) gene family in *L. japonicus* based on full-length cDNA and genomic sequences, gene structure, phylogeny, chromosome localization, and expression pattern. An in silico search identified a total of 17 *LysM-RLK* genes in the *L. japonicus* genome, including *Nfr1* and *Nfr5*. Finally, a detailed expression analysis allowed us to divide the 17 genes into four classes with distinct patterns of expression in different organs and to identify a number of *Lys* genes regulated either during the symbiotic association between *L. japonicus* and its rhizobial symbiont, *Mezorhizobium loti*, or in the chitin-induced signaling, or both.

## RESULTS

### Identification of *LysM-RLK* genes in *L. japonicus*.

Two *Lotus LysM-RLK* genes, *Nfr1* and *Nfr5*, were characterized and shown to be involved in Nod factor signaling (Madsen et al. 2003; Radutoiu et al. 2003, 2007). In addition, two tandem duplicated paralogs of *Nfr1* and one of *Nfr5* was identified on the same genomic clones (Madsen et al. 2003; Radutoiu et al. 2003; Zhang et al. 2007) and one unlinked paralog was described by Zhang and associates (2007). This led to the hypothesis that *L. japonicus* contains a family of LysM-RLK. Therefore, in silico searches for LysM-RLK in *L. japonicus* genomic and expressed sequence tag (EST) databases were initiated. Fifteen other *LysM-RLK* (*Lys*) genes were identified, including the three known paralogs of *Nfr1* and *Nfr5* (Table 1). In all, eight of the 17 *Lys* genes are represented in EST public collections.

The *Lys* genes were grouped and named according to their sequence similarity to *Nfr1* and *Nfr5*. The first group (LYS-I) contains seven genes, *Lys1* to *Lys7*, which are closely related to *Nfr1*. The second group (LYS-II) contains of six genes, *Lys11* to *Lys16*, that are similar to *Nfr5* and the last group (LYS-III) contains two genes, *Lys20* and *Lys21*, that are markedly different from both *Nfr1* and *Nfr5*. *Lys21* has a stop codon in the sequence encoding subdomain VIb of the kinase and possibly encodes a truncated protein. All newly identified *Lys* genes are predicted to encode proteins with an N-terminal signal peptide, three extracellular LysM motifs, a transmembrane spanning region, and a Ser-Thr kinase domain in the cytoplasmic part of the protein. LYS proteins share sequence identities ranging from 23 to 61% for group I and 21 to 61% among the proteins of group II, whereas LYS20 (the third group) has 19% identity to NFR1 and 20% to NFR5 (Table 1). The lowest level of conservation was generally found in the regions corresponding to the LysM domains.

### Analysis of *Lys* genes exon-intron structure, phylogeny, and kinase domain conservation.

To determine the structure of *L. japonicus Lys* genes, we amplified, cloned, and sequenced full-length cDNAs of the 15 newly identified *Lys* genes. The protein coding regions of *Lys* genes range from 1,776 to 2,040 bp, whereas corresponding genomic regions vary from 1,776 to 7,201 bp. Comparison of cDNA with the sequence of the corresponding genomic clones allowed separation of *L. japonicus Lys* genes into three groups based on their exon-intron structure (Fig. 1).

Using the amino acid sequence alignment of the kinase domains, we constructed a maximum likelihood phylogeny (Guindon et al. 2005) of the *Lotus Lys* genes. This phylogeny also separates the *Lys* genes into the three main clades: LYS-I containing *Nfr1* paralogs; LYS-II, including genes similar to *Nfr5*; and LYS-III, with two genes that differ significantly from both *Nfr1* and *Nfr5* (Fig. 2). A strong correlation was found between the exon-intron structure of the *Lys* genes and the main groupings in the phylogeny. Genes belonging to the LYS-I group have 10 to 13 exons. This group is further divided into two subgroups (*Nfr1*, *Lys1-2*, and *Lys6-7*) and (*Lys3* to *Lys5*) based on the presence or absence of a 5- to 8-bp microexon, respectively (Figs. 1 and 2). The second (LYS-II) and third (LYS-III) groups are represented by genes having one and two exons, respectively. The degree of conservation of the Ser-Thr kinase domain in the predicted LYS proteins was also consistent with the gene structure. All *Lys* genes of the LYS-I and LYS-III groups contain all the conserved protein kinase subdomains. However, *Lys* genes belonging to the LYS-II group encode proteins that lack or have substituted conserved residues in various kinase subdomains: the P-loop, the DFG motif, or the activation loop (Table 2).

### Investigations of past selective forces that have shaped the *Lys* gene family.

To infer the average selective pressure exerted on a set of protein-coding genes, the rate of nonsynonymous substitutions

**Table 1.** Amino acid identities among *Lotus japonicus* LYS proteins based on full-length protein, LysM motifs, and the kinase domain[a]

| | Amino acid identity (full length/LysM/kinase) (%) | | |
|---|---|---|---|
| **Phylogeny group** | **To NFR1** | **To NFR5** | **No. of amino acids** |
| LYS-I | | | |
| NFR1 | – | 22/17/34 | 621/623 |
| LYS1[b] | 55/51/67 | 20/16/33 | 601 |
| LYS2[b] | 53/49/63 | 20/18/34 | 631 |
| LYS3[b] | 25/11/44 | 18/15/32 | 621 |
| LYS4 | 23/12/42 | 19/16/33 | 636 |
| LYS5 | 25/12/46 | 19/15/32 | 665 |
| LYS6 | 61/55/79 | 21/19/34 | 623 |
| LYS7 | 54/40/76 | 19/18/34 | 622 |
| LYS-II | | | |
| NFR5 | 22/17/34 | – | 596 |
| LYS11 | 20/18/32 | 61/66/66 | 592 |
| LYS12[b] | 18/17/33 | 26/29/35 | 634 |
| LYS13 | 21/21/34 | 24/24/32 | 668 |
| LYS14 | 20/21/33 | 24/24/33 | 667 |
| LYS15 | 17/16/28 | 21/21/29 | 603 |
| LYS16 | 18/15/31 | 21/21/27 | 672 |
| LYS-III | | | |
| LYS20 | 19/17/35 | 20/20/28 | 669 |
| LYS21 | –/18/– | –/19/– | 484 (truncated) |

[a] Percentages of identity were defined by ClustalW.
[b] These genes are referred by Zhang and associates (2009) under the following names: *Lys1-LjLyk2*, *Lys2-LjLyk3*, *Lys3-LjLyk6*, and *Lys12-LjLyk4*.

($d_N$) to the rate of synonymous substitutions ($d_S$) ($\omega = d_N/d_S$) can be used. In a purely neutrally evolving set of sequences such as an alignment of pseudogenes, this ratio is expected to be equal to 1. Any significant deviation from 1 is a clear indication of non-neutrality. Purifying selection will lower the expected number of nonsynonymous changes that become fixed in a population and, therefore, result in a $d_N/d_S$ ratio that is smaller than 1 whereas positive selection will do the opposite and increase the number of nonsynonymous changes that eventually becomes fixed, resulting in a $d_N/d_S$ ratio larger than 1.

The entire set of *Lotus Lys* kinase domains has an estimated average $d_N/d_S$ ratio of 0.12. The sequences of *LysM-RLK* genes are quite divergent; therefore, we performed pairwise comparisons in selected branches of the phylogeny, containing more recently diverged genes. The subset of pairwise comparisons found suitable for $d_N/d_S$ ratio analyses are shown in Table 3. The *Lys13-Lys14* and *Lys4-Lys5* gene pairs show fairly low $d_N/d_S$ ratios, indicating that, on average, purifying selection has been the major determinant in the evolution of these genes. When studying the kinase domains, the estimates between the *Lys11-Nfr5* and the *Lys1-Lys2* pairs show the highest rates of 0.38 and 0.37, respectively. Moreover, analysis of the regions containing the LysM domains in the case of *Lys11-Nfr5* and *Lys20-Lys21* gene-pairs showed even higher levels of the $d_N/d_S$ ratio (Table 3). This increase shows that one or both of the genes in the pair have undergone more nonsynonymous changes in the kinase- or LysM-containing domains compared with the average of this gene family, indicating that one or both of the genes could have experienced a period of positive selection or, alternatively, lost its functional importance.

**Positive selection has likely shaped defined regions of *Lotus* LysM receptors.**

Mutant and transformation experiment analyses revealed that the NFR are host determinants in legumes, and that the second LysM motif of NFR5 plays a major role in Nod factor recognition (Radutoiu et al. 2007). For the newly identified *Lys* genes, no knowledge about their function or possible ligands is available. Moreover, our initial analyses of the $d_N/d_S$ ratio generally indicate that the genes are evolving under strong functional constraints ($d_N/d_S < 1$). In order to investigate whether fixation of amino acid replacements by positive selection has played a role in the evolution of the extracellular domain or other functionally important parts of the proteins, we have

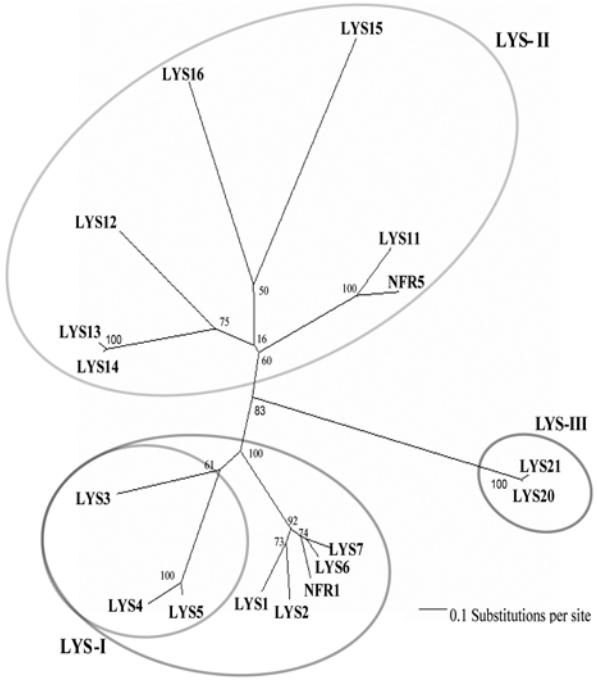


**Fig. 2.** Unrooted maximum likelihood phylogenetic tree based on the amino acid sequence alignment of *Lotus japonicus* LYS kinase domains. Numbers on branches denote bootstrap support based on 100 repetitions. The three main groups are designated LYS-I, LYS-II, and LYS-III.

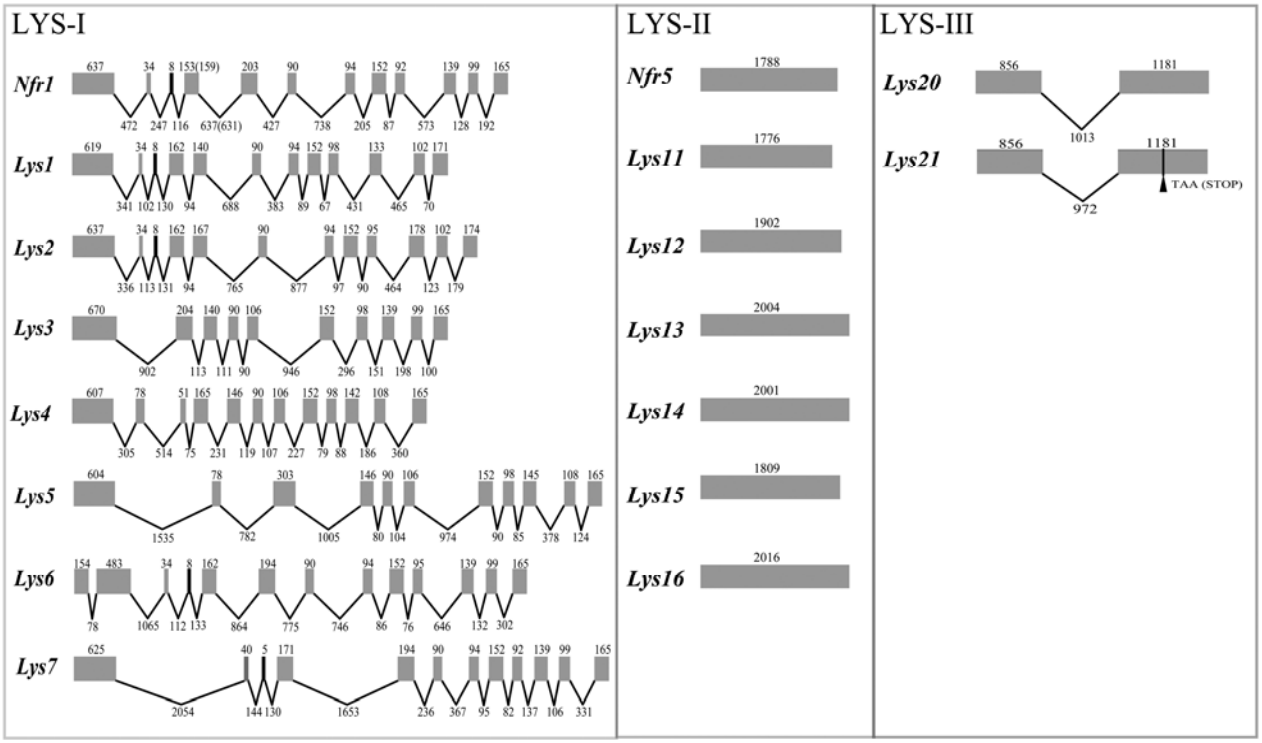


**Fig. 1.** *Lotus japonicus Lys* gene structures. Exons are illustrated by gray boxes and introns by lines. The microexon in the LYS I group is marked in black. The nucleotide numbers of each intron and exon and the location of the premature stop codon in *Lys21* are shown.

conducted a sliding-window analysis of the $d_N/d_S$ ratio in recently diverged gene pairs. A sliding window of 30 amino acids with an overlap of 15 was used to detect regions wherein the $d_N/d_S$ ratio was larger than 1 within the two major LYS groups, LYS-I and LYS-II. Separate analyses on the LysM-containing extracellular region and the intracellular kinase were performed.

Similar to the average $d_N/d_S$ ratio based on the entire gene family, the sliding-window analysis on individual groups and regions generally show fairly low $d_N/d_S$ values along the analyzed regions. However, a closer look at more recently diverged gene pairs, NFR5-LYS11 and LYS1-LYS2, revealed the presence of small specific regions where the $d_N/d_S$ ratio was above one (Figs. 3 and 4).

In the extracellular regions of the *Nfr5-Lys11* genes, two windows corresponding to the start of the second LysM and the middle end of the third LysM domain have a $d_N/d_S$ ratio above 1 (Fig. 3A and B). For the *Lys1-Lys2* gene pair, one region with a $d_N/d_S$ ratio larger than 1 was detected in the middle of the first LysM domain, while a second region with a distinctively higher $d_N/d_S$ ratio, though below 1, corresponded to the second LysM domain (Fig. 3D and E). None of these regions have a $d_N/d_S$ ratio significantly higher than 1 according to Fishers exact test but this does not exclude the possibility that positive selection may have played a part in the recent divergence of these gene regions. Interestingly, for the NFR5-LYS11 pair, the first region includes the NFR5-L118 that defines specificity toward decorations present on the nonreducing end of the Nod factor molecule (Radutoiu et al. 2007) (Fig. 3B). In order to obtain a more detailed insight into the possible effect of positive selection in the second LysM domain between NFR5 and LYS11, we took advantage of the NFR5-LysM2 homology model and mapped the location of amino acids that differ between the two proteins (Fig. 3C). Supporting the positive selection indicated by the $d_N/d_S$ ratio, most of the differences were found to be present in the suggested Nod factor-binding groove and in the area determining specificity (Bek et al. 2010; Radutoiu et al. 2007). Nevertheless, three of the four residues which have been suggested to define the Nod-factor binding cleft were maintained aromatic (Y114 to H, F124 to Y, and Y130), suggesting that possible sugar-aromatic stacking interaction might have been maintained (Fig. 3B and C).

**Table 2.** Kinase domain structure of *Lotus japonicus* Lysin motif-containing receptor kinases (LysM-RLK) based on protein domain predictions and amino acid sequence alignments[a]

| Group | LYS protein | P-loop (sd. I) | DFG motif (sd. VII) | Activation loop (sd. VIII) |
|---|---|---|---|---|
| LYS-II | NFR5, LYS11 | – | Substitutions | – |
| LYS-II | LYS12 to LYS16 | – | Substitutions | + |
| LYS-III | LYS20, LYS21 | + | + | + |
| LYS-I | NFR1, LYS1 to LYS4, LYS6 to LYS7 | + | + | + |

[a] sd. = kinase subdomain; + or – = presence or absence, respectively, of the subdomain.

**Table 3.** Ratios of nonsynonymous substitutions ($d_N$) to synonymous substitutions ($d_S$) ($\omega = d_N/d_S$) for recently diverged *Lys* genes

| | $d_N/d_S$ for each region | |
|---|---|---|
| **Genes** | **LysMs** | **Kinase** |
| *Nfr1-Lys1* | 0.27 | 0.26 |
| *Nfr1-Lys2* | 0.33 | 0.22 |
| *Lys1-Lys2* | 0.31 | 0.37 |
| *Lys1-Lys6* | 0.26 | 0.28 |
| *Lys4-Lys5* | 0.12 | 0.19 |
| *Nfr5-Lys11* | 0.49 | 0.38 |
| *Lys13-Lys14* | Identical | 0.17 |
| *Lys20-Lys21* | 0.72 | 0.16 |

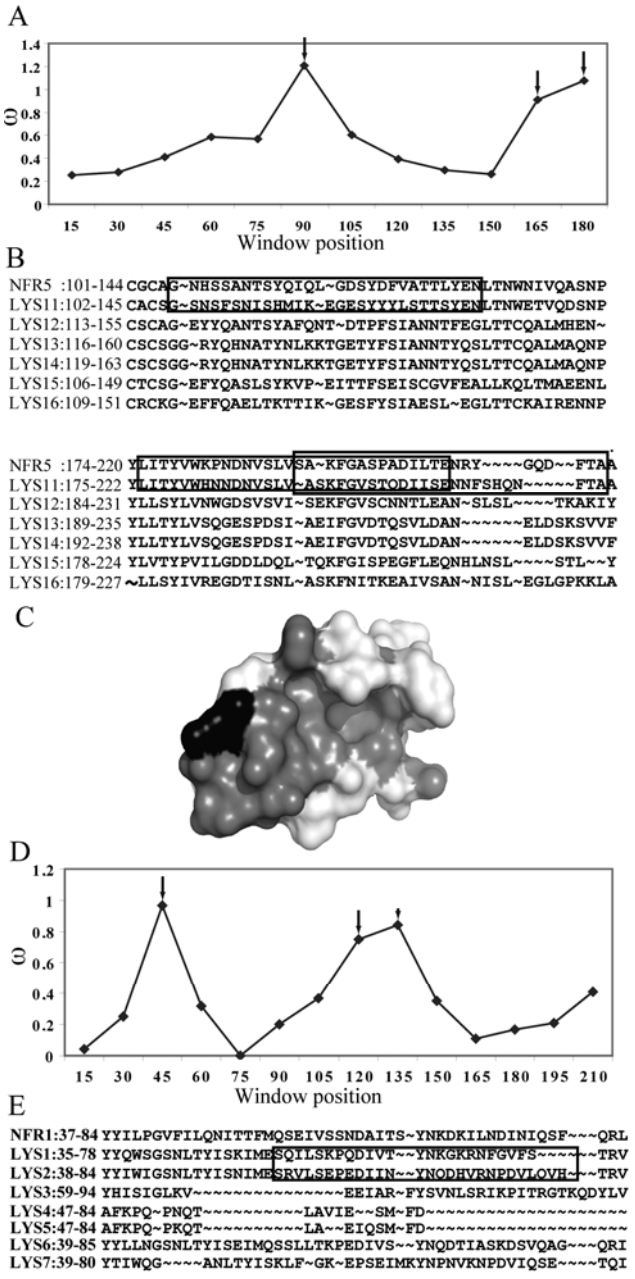


**Fig. 3.** Sliding-window plots and alignments for the LysM regions. **A,** Sliding-window plot of $d_N/d_S$ ratio calculated for NFR5 versus LYS11. **B,** Amino acid alignment for the LYS-II group used for the sliding-window model. **C,** Homology model of NFR5-LysM2 domain. Amino acids identical to LYS11 are in white, different amino acids are in gray, and the NFR5-Leucine 118 is shown in black. **D,** Sliding-window plot of $d_N/d_S$ ratio calculated for LYS1 versus LYS2. **E,** Amino acid alignment for the LYS-I group used for the sliding-window model. Arrows on panels A and D show the $d_N/d_S$ ratios, which correspond to the regions boxed in panels B and E, respectively. Numbers in B and E indicate the amino acid positions in individual proteins.

The intracellular kinase regions of the NFR5 and LYS11 share approximately 70% identity at the amino acid level and both proteins lack the same conserved subdomains (Table 2). However, using the sliding-window analysis of the $d_N/d_S$ ratio, we identified two regions where positive selection may have been an important evolutionary factor (Fig. 4A). The first one is located between the subdomains V and VI and the second one corresponds to the elusive subdomain X, where NFR5 has seven additional amino acids (Fig. 4B). A similar situation was identified when the kinases of LYS1 and LYS2 were analyzed (Fig. 4C). However, in this case, the two regions are located only in subdomain X, and the first window corresponds to 14 of the 17 additional amino acids of LYS2 (Fig. 4D).

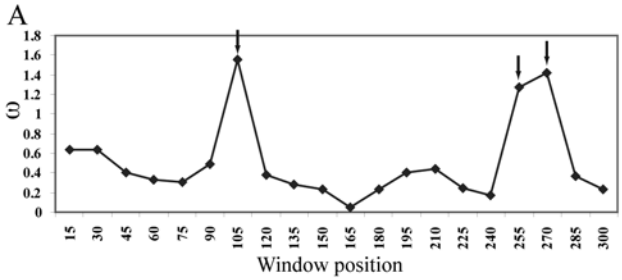


**Fig. 4.** Sliding-window plots and alignments for the kinase regions. **A,** Sliding-window plot of $d_N/d_S$ ratio calculated for NFR5 versus LYS11. **B,** Amino acid alignment for the LYS-II group used for the sliding-window model. **C,** Sliding-window plot of $d_N/d_S$ ratio calculated for LYS1 versus LYS2. **D,** Amino acid alignment for the LYS-I group used for the sliding-window model. Arrows on panels A and C show the $d_N/d_S$ ratios, which correspond to the regions, boxed in panel B and D, respectively. Numbers in B and D indicate the amino acid positions in individual proteins.

**Chromosomal localization of *LysM-RLK* genes in *L. japonicus*.**

To investigate the genomic organization of the *Lys* genes in *L. japonicus*, we determined their chromosomal localization. The genes are positioned on five of the six chromosomes (excluding chromosome number 5) (Sato et al. 2008) (Fig. 5). Seven of them localize on chromosome 2 in three different gene clusters (*Nfr1-Lys1-Lys2*, *Nfr5-Lys12*, and *Lys13-Lys14*) and one as a single gene (*Lys3*). Evidence of segmental genome duplications was identified for the *Nfr1* and *Nfr5* chromosomal regions, which show synteny with the genomic regions containing *Lys6* (chromosome 6) and *Lys11* (chromosome 4), respectively (Supplementary Fig. S1). However, despite sequence similarity between the genomic regions surrounding *Nfr1* and *Lys6* or *Nfr5* and *Lys11*, no tandem-duplicated paralogous genes were found in the region surrounding *Lys6* or *Lys11*. Considering the higher sequence similarity of *Nfr1* to *Lys6* compared with *Lys1-Lys2*, the tandem duplication events in the *Nfr1* cluster appear to predate the genome-wide segmental duplication. Consequently, the paralogous genes next to *Lys6* may have been lost after the segmental genome duplication. A similar scenario might apply to *Nfr5* and its paralogous genes (*Lys11* and *Lys12*). Therefore, we conclude that tandem and segmental duplication events appear to be the major source for expansion of the *Lys* gene family in *L. japonicus*.

***Lys* genes have specific expression patterns in different *Lotus* organs.**

In *L. japonicus*, the only LysM receptors with a known function, *Nfr1* and *Nfr5*, are preferentially expressed in root tissues (Madsen et al. 2003; Radutoiu et al. 2003). We analyzed the expression patterns of the newly identified genes by quantita-

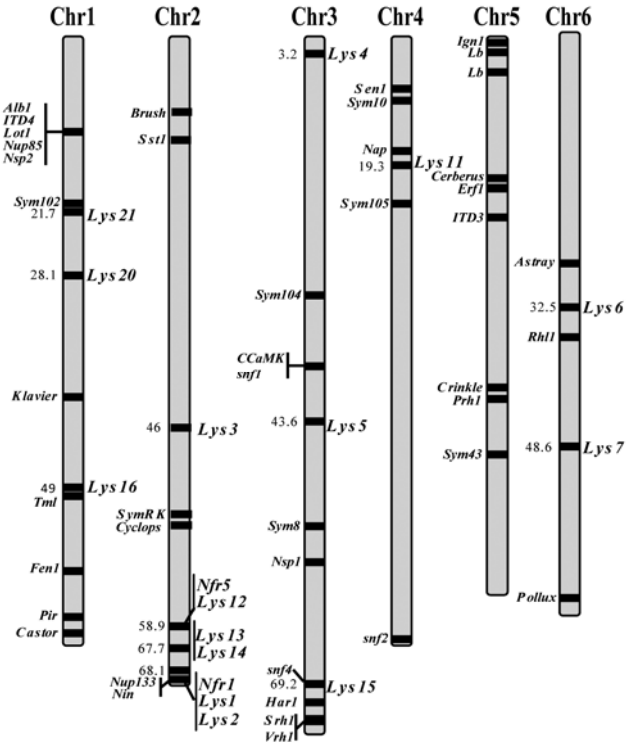


**Fig. 5.** Genetic map positions of *Lotus japonicus Lys* genes. Positions of the three groups of duplicated paralogous (*Nfr5-Lys12*, *Lys13-Lys14*, and *Nfr1-Lys1-Lys2*) are denoted by a single line for the particular group. Positions of known symbiotic loci are shown to the left of each chromosome. Distances are in centimorgans (cM). Genetic positions correspond to *L. japonicus* Miyakojima MG-20.

tive real-time polymerase chain reaction (PCR) and separated the genes into four groups according to their pattern of expression in leaf, stem, root, and nodule (Fig. 6). Two genes, *Lys4* and *Lys13,* were grouped together, with *Nfr1* and *Nfr5* being preferentially expressed in roots (Fig. 6A). Another group of six genes (*Lys2, Lys3, Lys7, Lys12, Lys15,* and *Lys20*) were mainly expressed in roots or nodules or both (Fig. 6B). Three genes (*Lys1, Lys11,* and *Lys16*) were preferentially expressed in the aerial tissues. *Lys1* and *Lys11* were upregulated in leaf and stem while *Lys16* was primarily expressed in the stem tissue (Fig. 6C). Finally, four genes (*Lys5*, *Lys6*, *Lys14*, and *Lys21*) were found ubiquitously expressed to similar levels in all or most of the tested organs (Fig. 6D).

### *Lys* genes are regulated by *M. loti* or chitin treatment or both.

Expression pattern analyses in different *Lotus* organs revealed that eight *Lys* genes (*Lys2-4*, *Lys7*, *Lys12-13*, *Lys15*, and *Lys20*) are expressed mainly in nodules, root tissues, or both. In order to obtain a better insight into their regulation during rhizobium-legume symbiosis, we studied their expression pattern in the whole root 3 days postinoculation (dpi) (Fig. 7A, C, and E) and in the susceptible zone at different time points after rhizobial treatment (Fig. 7B, D, and F). The *Nin* gene, included as reference for nodulin upregulation after *M. loti* inoculation, was found together with *Lys13* to be highly induced in the whole root, while lower levels of induction were detected for *Lys3* and *Lys14*. In the susceptible zone, *Nin* and *Lys3* transcript levels increase from 8 h postinoculation (hpi) to 2 dpi (Fig. 7A and B). A second group of genes, which contains *Nfr1*, *Nfr5*, *Lys2*, and *Lys16,* was found to be slightly but statistically significant upregulated after *M. loti* inoculation in the susceptible zone (Fig. 7C and D). Interestingly, *Lys16* that is preferentially expressed in the stem was found regulated up to twofold during the early time points of *L. japonicus–M. loti* interaction. Another group of four genes (*Lys5*, *Lys6*, *Lys12*, and *Lys15*) had significantly lower transcript levels in the whole roots upon inoculation (Fig. 7E), and three of them (*Lys5*, *Lys12*, and *Lys15*) were identified to be slightly regulated at the early time points in the susceptible zone (Fig. 7F). A similar

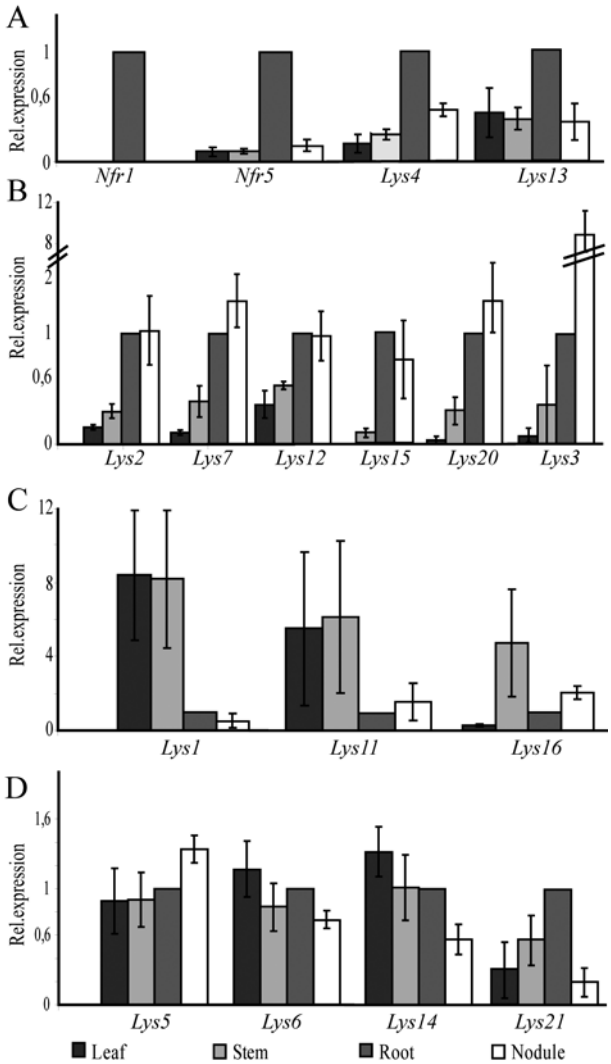


**Fig. 6.** Expression of *Lotus japonicus Lys* genes in leaf, stem, root, and nodule tissues. **A,** *Lys* genes preferentially expressed in roots. **B,** *Lys* genes expressed preferentially in roots and nodules. **C,** *Lys1*/*Lys11* and *Lys16* expressed preferentially in aerial and stem tissue, respectively. **D,** *Lys* genes ubiquitously expressed in the tested organs. All values are relative to the expression value in roots set to 1. Panels have different scales. The error bars represent the corresponding upper and lower 95% confidence interval.

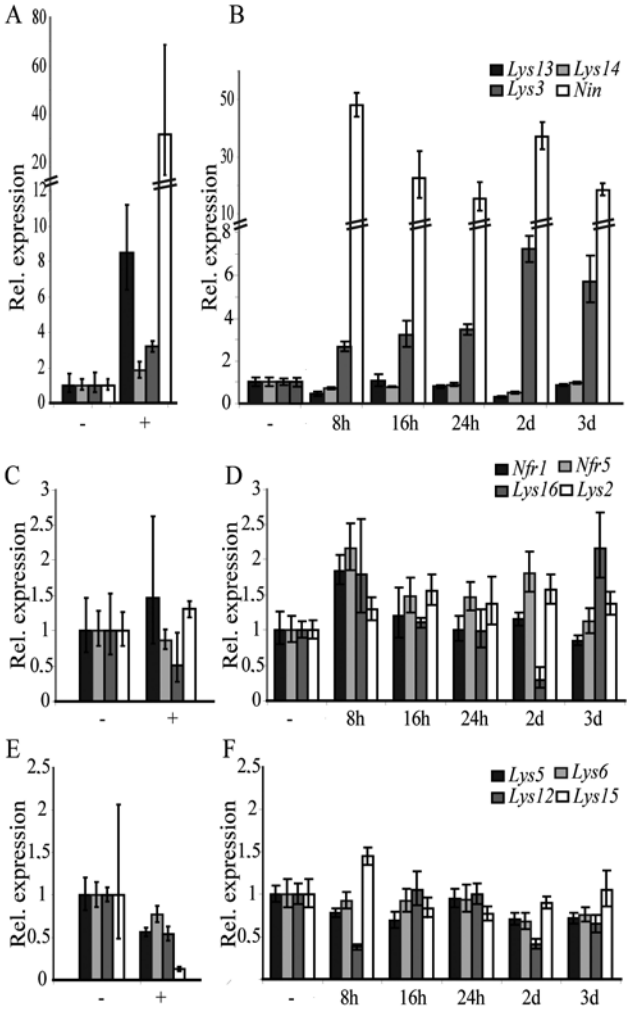


**Fig. 7.** Expression of *Lotus japonicus Lys* genes during symbiosis with *Mesorhizobium loti*. **A, C,** and **E,** Whole root uninoculated (–) and *M. loti* inoculated (+). **B, D,** and **F,** Susceptible zone uninoculated (–) and at 8, 16, and 24 h 2 and 3 days upon inoculation with *M. loti*. **A** and **B,** *Lys* genes that are upregulated in the susceptible zone or at the whole-root level. **C** and **D,** Genes are all induced in the susceptible zone but not at the whole-root level; **E** and **F,** genes that are downregulated at the whole-root level. Values are calculated as ratios relative to **A, C,** and **E,** uninoculated whole root or **B, D,** and **F,** uninoculated susceptible zone set to 1. Note the change in scale in panels A and B. Error bars represent the corresponding upper and lower 95% confidence interval.

pattern was observed when the *Lotus* 52K Affy chip (Hogslund et al. 2009) was used to evaluate *Lys* gene expression in various organs or in roots after *M. loti* inoculation (data not shown).

In *Arabidopsis*, one of the five LysM-RLK was shown to be involved in chitin signaling and its transcript levels were upregulated shortly after chitin treatment (Miya et al. 2007). However, in *L. japonicus*, mutations in *Nfr1* and *Nfr5* did not affect chitin signaling (Wan et al. 2008). In order to identify *Lotus LysM-RLK* genes whose transcripts are regulated by chitin, we analyzed the expression of all *Lys* genes in shoot and root following chitin treatment. Only three genes (*Lys13*, *Lys14*, and *Lys20*) were identified to be regulated by chitin in both shoot and root tissues (Fig. 8A and B). The levels of induction were similar to *Wrky33*, *Wrky53*, and *Mpk3,* which are known as chitin-regulated genes in *Arabidopsis* and *L. japonicus* (Wan et al. 2004, 2008). Interestingly, *Nfr5* was found to be slightly downregulated by chitin treatment in the root tissues (Fig. 8B).

Our detailed analysis of *Lys* expression upon *M. loti* inoculation identified nine genes (*Lys2*, *Lys3*, *Lys5*, *Lys6*, and *Lys12-16*) significantly regulated after inoculation and initiation of the symbiotic interaction in either the whole root, the susceptible zone, or both. Furthermore, three genes (*Lys13, Lys14*, and *Lys20*) were upregulated by chitin. These data indicate that, in addition to *Nfr1* and *Nfr5*, other *Lotus LysM* receptor genes might be required for the nitrogen-fixing or chitin signaling pathways, or both.

## DISCUSSION

### The *L. japonicus* genome contains a large family of *LysM-RLK* genes.

Searches for genes encoding LysM proteins in the *L. japonicus* genome identified a large gene family consisting of at least 17 LysM-RLK genes. This number is larger than in nonlegume plants such as *Arabidopsis* (five genes) and rice (six genes). A similar number of *LysM-RLK* genes in the *Medicago truncatula* (17 genes) and soybean (12 genes) genomes were also identified by in silico searches in EST and whole-genome databases (Arrighi et al. 2006; Zhang et al. 2007). This indicates that LysM-RLK-encoding genes have expanded in legume genomes, possibly acquiring important new functions.

All newly identified *LysM-RLK* genes are predicted to encode a protein with an N-terminal signal peptide, three LysM motifs, a transmembrane spanning region, and a Ser-Thr kinase domain in the predicted cytoplasmic part of the protein. The extracellular regions of *Lotus* LYS proteins were predicted to contain three LysM motifs. Furthermore, the cysteine pairs (CXC), which are present between the individual LysM domains of *Lotus*, *Medicago*, soybean, poplar, *Arabidopsis*, and rice LysM-RLK, suggest that this pattern is a characteristic trait for plant proteins containing LysM domains fused to a kinase. Cysteine pairs were also found flanking the leucine-rich repeats (LRR) of plant LRR-receptor kinases (Torii 2004). These cysteines are spaced by six amino acids and seem to be important for receptor function and for receptor dimerization (Trotochaud et al. 1999). Disulfide bridges are as well known to stabilize protein folding (Baldwin 2007); therefore, the cysteine pairs bordering individual LysM motifs could participate in correct folding of the extracellular region. In general, the individual LysM motifs of LYS proteins have a higher similarity to the equivalent motifs in the closely related proteins than to its own neighboring LysM motif. Exceptions from this distinctive organization of the extracellular region are the LYS3, LYS4, and LYS5, which have an atypical assembly of the individual LysM domains. Further analyses remain to determine whether the organization of the extracellular region of LYS3, LYS4, and LYS5 proteins represent a more ancient trait.

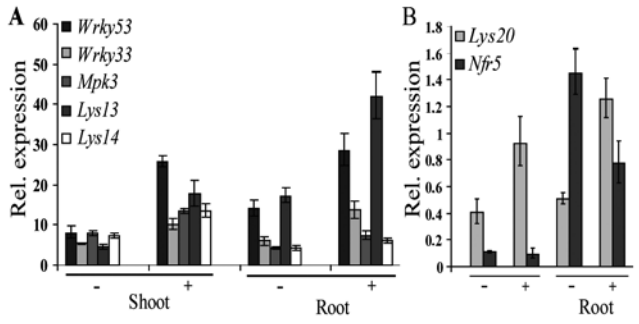


**Fig. 8.** *Lys* genes regulated upon chitin treatment in *Lotus* shoots and roots. **A,** Relative expression of *Lys13*, *Lys14*, *Wrky53*, *Wrky33*, and *Mpk3*. **B,** Relative expression of *Lys20* and *Nfr5*. Mock-treated (–) and chitin-treated (+) plants have been analyzed. Panels have different scales. Error bars represent the corresponding upper and lower 95% confidence interval.

The extracellular regions of LYS proteins are highly variable in sequence, a situation similar to plant disease resistance *RLK* genes containing extracellular LRR. The LRR domains have been shown to determine recognition specificity (Rairdan and Moffett 2007) and, therefore, are subjected to positive selection to ensure a large potential for recognition diversity (Mondragon-Palomino et al. 2002; Rairdan and Moffett 2007). Similarly, the highly divergent LysM domains among different LysM-RLK could contribute to an enlarged spectrum of ligand recognition. Our detailed analysis of recently diverged *Lys* gene pairs using the sliding-window analysis allowed identification of specific LysM regions that have a $d_N/d_S$ ratio above 1. In the NFR5-LYS11 case, one of these regions contains most of the residues (except F159), which defined the putative ligand-binding cleft and the specificity-determining area of the homology model (Fig. 3 B and C) (Radutoiu et al. 2007). Together, this suggests that selection may have shaped the carbohydrate-binding and recognition properties of this domain (Fig. 3C). However, some of the *L. japonicus LysM-RLK* genes show high similarity in their LysM domains. The most obvious example is the LysM domains of LYS13 and LYS14, in which the amino acids are identical in all positions. In this case, the identity extends from the region between the signal peptide and the first LysM motif to a region between kinase subdomains V and VIa. This high degree of identity could be explained by a gene conversion event where unequal crossing-over transferred the LysM and part of the kinase domain from one of the genes to the other. In a study of three gene families in *Arabidopsis*, gene conversion events were frequent among highly similar and directly repeated genes located close to each other (Mondragon-Palomino and Gaut 2005). The sequences of *Lys13* and *Lys14* are very similar (92% amino acid identity) and are located 5 kb apart from each other on chromosome 2. Therefore, it is plausible that the LysM domains and part of the kinase domain of the two genes were homogenized by a gene conversion event.

Analysis of the LYS kinase domains revealed a high prevalence of atypical kinases in the family. The absence of the P-loop and DFG motif in all LYS-II proteins are indications of their monophyletic origin followed by a second separation of the NFR5-LYS11 pair, which has lost the activation loop as well. In a standard phosphorylation assay, the NFP protein, which is the ortholog of NFR5 in *Medicago truncatula*, lacks autophosphorylation activity (Arrighi et al. 2006), suggesting that LYS-II proteins might signal through phosphorylation-independent mechanisms. However, kinases lacking conserved catalytic residues are not necessarily inactive. Mammalian pseudokinases have amino acids in other kinase domains which substitute for the missing residues (Xu et al. 2000). However, if the kinase is inactive, the pseudokinase might play important roles by binding

to the kinase domain of functional protein kinases, thereby controlling their activity (Boudeau et al. 2006).

**The separation of *Lys* genes into three groups is mutually supported by the exon-intron structure, phylogeny, and kinase domain conservation.**

The *LysM-RLK* genes of *L. japonicus* were separated into three major groups, LYS-I to LYS-III. Low bootstrap values in the phylogeny can complicate the definition of subfamilies when based solely on the sequence similarities. However, gene structure provides additional criteria to support or refute the groupings based on phylogeny alone (Shiu and Bleecker 2003). A strong correlation between the phylogeny and gene structure was observed among the *Nfr* and *Lys* genes, providing additional support to the definition of *LysM-RLK* gene subfamilies in *L. japonicus*. Furthermore, the LYS-I group was divided in two subgroups based on the presence or absence of a microexon.

A phylogeny based on full-length LysM-RLK amino acid sequences from *L. japonicus*, *Medicago truncatula*, *Glycine max*, *P. trichocarpa*, *A. thaliana*, and *O. sativa* reveals that all proteins belonging to the LYS-II group deviate from the consensus or lack conserved residues of typical protein kinases, supporting the grouping of these proteins (Supplementary Fig. S2). Furthermore, the phylogeny shows that all the plant species tested contain LysM-RLK that group in either LYS-I or LYS-II; however, the LYS-III group contains only *L. japonicus* and *Medicago truncatula* LysM-RLK. Additional in silico analyses identified two LysM-RLK from *Ricinus communis* that share 53% identical residues with LYS20. This indicates that diversification of LYS-I and LYS-II types of LysM-RLK predated the divergence of dicot and monocot plants, whereas the events leading to the LYS-III group might be of a more recent origin because no monocot genes cluster in this group. The kinase domain of LYS20 shows 70% identity to plant wall-associated kinases, which are involved in pathogen resistance (He et al. 1998), heavy-metal tolerance (Sivaguru et al. 2003), and plant development (Lally et al. 2001). The fusion of the LysM triplet to a special type of kinase by gene conversion could contribute to molecular complexity of these types of receptors.

The fact that the *Lys* gene phylogeny correlates with exon-intron structure provides some information regarding which group of genes might be of a more ancient origin. By comparing intron positions in orthologous genes from various eukaryotic lineages, the common ancestor of animals and plants was shown to posses an intron-rich genome (Rogozin et al. 2005). Therefore, the LYS-I group of the *L. japonicus* might be closer to ancestral genes whereas intron-loss could have occurred in the potentially more recent genes of the LYS-II and LYS-III groups. This has also been hypothesized by Zhang and associates (2007) in a phylogenetic study of LysM-RLK from six plant species. In this analysis, two small clades containing intron-rich LysM-RLK were judged to be ancestral.

**Tandem and segmental duplications led to *Lys* gene expansion in *L. japonicus*.**

Localization of the *LysM-RLK* genes on the *L. japonicus* genome reveals the contribution of both tandem and segmental duplications in the expansion of the gene family. The contribution of tandem duplication is supported by the localization of seven *LysM-RLK* genes in tandem repeats (*Nfr1-Lys1-Lys2*, *Nfr5-Lys12*, and *Lys13-Lys14*). Furthermore, the occurrence of segmental duplication events are supported by the observation that *Lys11*, the closest paralog of *Nfr5*, is located on a duplicated segment on chromosome 4, showing synteny to the *Nfr5-Lys12* region, while a duplicated segment of the *Nfr1* region is present on chromosome 6, where *Lys6* is located. The importance of tandem and segmental duplications has also been reported for the expansion of *LysM-RLK* genes and *RLK* gene families in *A. thaliana* and *Medicago truncatula,* respectively (Shiu and Bleecker 2003; Arrighi et al. 2006).

The conservation of genes between large duplicated blocks within the *L. japonicus* genome is only 13%, whereas 60% of the genes in syntenic regions between *Medicago truncatula* and *L. japonicus* are conserved, indicating that these large segmental duplications predated the common ancestor of *Lotus* and *Medicago* spp. (Sato et al. 2008). The syntenic relationships between *L. japonicus Nfr1-Lys1-Lys2* and the *Medicago truncatula Lyk1-7* region as well as the syntenic relationship between *L. japonicus Nfr5-Lys12* and *Medicago truncatula Nfp-Lyr3* have previously been documented (Zhu et al. 2006; Zhang et al. 2007). Furthermore, these reports show that a close paralog to *Nfp* (*Lyr1*) is located on a duplicated segment of the *Nfp*, *Lyr3* region on chromosome 8 of *Medicago truncatula*. Our data extends this syntenic relationship between the two legume species to include close homologs such as *Lys3* and *Lyk10*, *Lys13-Lys14* and *Lyr4*, and *Lys20-Lys21* and *Lyr5-Lyr6* (Supplementary Fig. S3, yellow, red, and blue boxes, respectively). The three groups are all positioned within syntenic regions between *L. japonicus* and *Medicago truncatula* and the genes within each group cluster closely together in a phylogeny.

Therefore, the syntenic relationship between *L. japonicus* and *Medicago truncatula LysM-RLK* genes is extensive. This is supported by sequence similarities among the genes; a phylogeny of *Lotus* and *Medicago* LysM-RLK protein sequences reveals that nearly all *L. japonicus* LysM-RLK cluster with a potential *Medicago* ortholog. Together, these results strongly indicate that most tandem and segmental duplication events of the *LysM-RLK* genes occurred before the separation of the two legume species. However, there is also evidence of species-specific duplication events. One example is the presence of seven *Lyk* genes of *Medicago truncatula* in the syntenic region of only three *LysM-RLK* genes (*Nfr1*, *Lys1*, and *Lys2*) in *L. japonicus*. Conversely, an extra duplication event seem to have occurred in the *L. japonicus Lys13-14* tandem repeat as opposed to only one close homolog, *Lyr4*, in *Medicago truncatula*. Furthermore, based on the phylogeny, four genes of *L. japonicus*—*Lys4, Lys5, Lys15,* and *Lys16*—show no close *Medicago truncatula* homologs and—for *Lys4, Lys5*, and *Lys15*—no syntenic relationship between the two legume species has been identified, indicating that these genes could be missing in *Medicago truncatula*. However, because soybean, *Arabidopsis*, and poplar all have one gene that clusters together with *L. japonicus Lys4* and *Lys5* and the fact that the genome of *Medicago truncatula* is not yet fully sequenced, close homologs of *L. japonicus Lys4, Lys5, Lys15,* and *Lys16* could exist in yet unsequenced regions of the genome.

In *Arabidopsis* and rice, *RLK* genes involved in defense or resistance mechanisms have been subject to preferential expansion whereas genes involved in development have rarely been duplicated following the monocot-dicot separation (Shiu et al. 2004). The large number of duplications within the *L. japonicus LysM-RLK* gene family could be related to the possible involvement of some LysM-RLK in plant innate immunity, as indicated by the regulation of some *Lotus* LysM-RLK family members upon chitin treatment. Furthermore, the LysM-RLK of *Medicago truncatula*, *L. japonicus*, and most of the *G. max* LysM-RLK which were shown to be involved in symbiosis are all localized in tandem repeats (Limpens et al. 2003; Zhu et al. 2006; Zhang et al. 2007), indicating that genes involved in the perception of the rhizobial Nod factor have been duplicated or retained at high rates, similar to genes involved in defense or resistance. The syntenic region between *Medicago truncatula Lyk1-7* and *L. japonicus Nfr1-Lys1-Lys2* contains seven and three *LysM-*

*RLK* genes, respectively, and this expansion may have influenced the Nod factor perception suggested by the difference in *nfr1* and *lyk3* mutant phenotypes (Radutoiu et al. 2003; Smit et al. 2007). An attractive hypothesis is that extension of the LysM-RLK family in legumes allows perception and recognition of various interacting organisms that could be symbiotic or pathogenic by nature.

**Past selective forces shaping the *Lys* genes.**

In order to elucidate whether the *Lys* genes have experienced positive or negative selection, we calculated the rate of nonsynonymous substitutions ($d_N$) to the rate of synonymous substitutions ($d_S$) ($\omega = d_N/d_S$). The calculated ratios on different gene pairs of the *L. japonicus Lys* family revealed a couple of gene pairs having a higher rate than the average of the gene family. One explanation for these results is that one or both of the genes in the pair could have experienced a period of positive or diversifying selection. Another alternative is that one or both of the genes have lost their functional importance and, therefore, have experienced relaxed constraints on the amino acid sequence conservation. Because the $d_N/d_S$ ratios are both below 1, it is not possible to discriminate between the two explanations solely from these analyses. Interestingly, for the two sets of genes (*Nfr5-Lys11* and *Lys1-Lys2*) that show a high $d_N/d_S$ ratio in the kinase domains, the two genes of a pair show a divergent pattern of expression. *Nfr5*, the Nod factor receptor, is expressed specifically in roots while *Lys11* seems to be preferentially expressed in the aerial tissues (Fig. 6). *Lys1* is expressed mainly in the aerial tissues while *Lys2* is expressed in root and nodule tissues (Fig. 6). Together, this indicates that the two genes of the *Nfr5-Lys11* and *Lys1-Lys2* pairs have diverged in their pattern of expression and, perhaps, in their respective functions as well. Functional constraints in the kinase domains might reflect the stronger purifying selection in this domain. In several cases, positive selection seems to be restricted to certain residues of the domains (Mondragon-Palomino et al. 2002; Shiu et al. 2004). Our analysis of the $d_N/d_S$ ratio on the kinase domains using the sliding-window model identified defined regions that might have been subject to diversifying selection (Fig. 4). Interestingly, in both the NFR5-LYS11 and the LYS1-LYS2 pair, one of these regions includes amino acid insertions or deletions in one of the compared proteins. These insertions contain serine/threonine residues that could be phosphorylated and allow interaction with specific downstream signaling components.

**Expression pattern of several *Lys* genes in plant organs and at time points during early symbiosis suggests an involvement in symbiosis.**

We present here a detailed expression analysis of the entire *LysM-RLK* gene family in *L. japonicus*. Eight genes (*Lys2*, *Lys 3, Lys4*, *Lys7*, *Lys12, Lys13*, *Lys15,* and *Lys20*) were expressed mainly in nodules, root tissues, or both (Fig. 6A and B), fewer than previously reported for *Medicago truncatula* (Arrighi et al. 2006). This might reflect the different developmental processes controlling nodulation in the two model legume species. The determinate and indeterminate nodule formation in *L. japonicus* and *Medicago truncatula*, respectively, was also suggested to explain the different expression patterns between *Lotus Nfr1* and *Nfr5* and *Medicago Nfp* and *Lyk3* (Arrighi et al. 2006).

We analyzed the expression of the *LysM-RLK* gene family in roots upon *M. loti* inoculation and at different time points during infection in the zone susceptible to bacterial infection. At 8 hpi, Nod factor–induced balloon-shaped root hairs are visible and *Nin* is highly expressed (Fig. 7B), indicating that the perception of Nod factors and initiation of the symbiotic response has occurred. At 16 and 24 hpi, branching and root-hair curling, respectively, are observed while microcolony formation and infection thread initiation are observed at 2 and 3 dpi, respectively. The steady-state transcript levels of nine *Lys* genes (*Lys2*, *Lys3*, *Lys5*, *Lys6*, and *Lys12* to *Lys16*) showed signs of regulation after inoculation at different time points in the susceptible zone. For a number of genes (*Nin*, *Nfr1*, *Nfr5*, *Lys15*, and *Lys16*), an increased transcript level was detected at 8 hpi. This observation correlates well with the function of NFR1 and NFR5 as Nod factor receptors acting at the initial Nod factor perception stage. Interestingly, another group of four genes (*Lys5*, *Lys6*, *Lys12*, and *Lys15*) was found to be downregulated in the whole root during the rhizobial infection process (Fig. 7E).

Our results on *Lys* gene expression indicate that, in addition to *Nfr1* and *Nfr5*, other *Lys* genes might play a role in the symbiotic interaction between *M. loti* and *L. japonicus*. In comparison, analysis of *Lys* genes expression at 4 and 28 dpi with the AM fungus *Gigaspora margarita*, using the *Lotus* 52K Affy chip (Guether et al. 2009), did not reveal transcriptional changes (data not shown) when compared with untreated roots.

**Expression patterns upon chitin treatment indicate the involvement of some LysM-RLK in plant innate immunity.**

Chitin treatment has been shown to induce a number of PAMP receptor genes, such as *Fls2* and *Efr*, and the putative chitin receptor-gene in *Arabidopsis*, *Cerk1* (Miya et al. 2007). We identified three *Lys* genes (*Lys13*, *Lys14*, and *Lys20*) to be upregulated upon chitin treatment in *L. japonicus*. *Lys13* and *Lys14* were upregulated by chitin in both root and shoot tissues to a level similar to the well-known chitin-induced genes, *Wrky33*, *Wrky53*, and *Mapk3* (Fig. 8). The LysM domains of LYS13 and LYS14 are 100% identical, suggesting recognition of identical or very similar ligands and, therefore, a redundant effect. However, their kinase domains differ, which might suggest, as well, that specific signaling cascades might be activated by LYS13 or LYS14.

In legumes, which evolved symbiosis with rhizobia and fungi to become independent of soil nitrogen and for enhanced phosphor uptake, the family of LysM receptor kinase has expanded. The evolutionary diversification of *Lys* genes in *L. japonicus* is reflected in specific expression patterns and a large diversity in the exposed extracellular regions combined with variations in the kinase domains. This diversity may possibly account for the capacity of legumes to decipher various structures of chitin-derived molecules produced by an extended spectrum of interacting bacteria, fungi, nematodes, and insects. As in the case of other gene families (Cano-Delgado et al. 2004; DeYoung et al. 2006), functional analyses of the *Lys* genes most likely necessitate complex genetic resources to reveal their role in signaling events induced by symbiotic, associative, endophytic, and parasitic organisms.

## MATERIALS AND METHODS

**Plant material and growth conditions.**

*L. japonicus* ecotypes MG-20 Miyakojima and Gifu B-129 were used for cloning of LysM-RLK genes. Plants were grown in greenhouse conditions: temperature cycles between 18 to 21 and 15°C in day and night, respectively, and relative humidity of 70%.

**Retrieval of *L. japonicus Lys* genes.**

A reiterated search in the *L. japonicus* whole-genome sequence and EST database (Gene Index and Kazusa DNA Research Institute websites) using the keyword "LysM" and the

blastp algorithm with the sequence of *A. thaliana* and *L. japonicus* LysM-RLK was performed. Domain structure of retrieved genes was analyzed using web-based programs and, in the case of LysM domains, partly by visual inspection. LysM and Ser-Thr protein kinase domains were analyzed by Pfam and Inter-ProScan. Signal peptide and transmembrane domains were predicted using SignalP and TMHMM server v. 2.0, respectively.

**Cloning of *LysM-RLK* genes and determination of exon-intron structure.**

Primers for amplification of *Lys* genes were designed based on the gene prediction that was assigned by Kazusa Annotation PipelineE (Sato et al. 2008). Full-length coding regions from the signal peptide to the translational stop codon was amplified using a cDNA template that was synthesized from total RNA of nodulated (*M. loti* MAFF303099) *L. japonicus* MG-20 Miyakojima plants. Additionally, full-length cDNA clones of *Lys2*, *Lys3*, *Lys11*, *Lys12*, and *Lys20* were isolated by 5′ rapid amplification of cDNA ends (RACE) PCR on 5′-RACE-ready cDNA (Clontech, Palo Alto, CA, U.S.A.) prepared from messenger (mRNA) extracted (Dynabeads mRNA direct kit) from 17-day-old *M. loti* (NZP2235)-inoculated or uninoculated *L. japonicus* Gifu roots, and the amplified cDNA products were cloned (pCR 2.1-TOPO or pENTR D-TOPO; Invitrogen, Carlsbad, CA, U.S.A.). The full-length cDNA clones of *Lys1* and *Lys13* were obtained from the Kazusa DNA Research Institute. Exon-intron structure was determined by comparing the sequence of the cDNA clones and the sequence of the corresponding genomic region.

*Lys* gene sequences have been deposited in the DNA Data Bank of Japan (DDBJ) under accession numbers AB503681 to AB503696 and AB506699 to AB506704.

**Expression analyses.**

For expression analysis of *Lys* genes on plant organs, *L. japonicus* ecotype Gifu B-129 was grown in soil (vermiculite) supplied with nitrogen-free B&D medium (Broughton and Dilworth 1971) and inoculated with *M. loti* MAFF303099. Four weeks after rhizobial inoculation, the various organs (leaf, stem, root, and nodules) were collected. Total RNA was extracted using the RNeasy plant mini kit (Qiagen, Hilden, Germany) followed by treatment with RNase-Free DNase (Qiagen) for genomic DNA removal. First-strand cDNA was synthesized using Superscript III reverse transcriptase (Invitrogen) from 0.25 µg of total RNA. Real-time reverse-transcriptase (RT)-PCR was performed using DyNAmo HS SYBR Green qPCR Kit (Finnzymes, Espoo, Finland) and 1 µl of a fourfold-diluted cDNA template was used for each RT-PCR reaction. Results were quantified using the DNA Engine Opticon2 system (Bio-Rad). Transcript levels were normalized to *LjeIF-4A* (Uchiumi et al. 2002).

The *Lys* gene transcript levels at different time points during the symbiotic interaction with *M. loti* MAFF303099 were analyzed on *L. japonicus* Gifu plants grown in greenhouse conditions. The plants were inoculated with rhizobia at a specified time point to ensure tissue harvest at the same time and plant age. The untreated and treated whole root or the susceptible zone was harvested as mentioned. The RNA was extracted using a borate buffer (20 mM Na-Borate, 30 mM EGTA, 5 mM EDTA, and 1% [wt/vol] Na-Deoxycholate) preheated to 95°C and DNAse (Promega, Madison, WI, U.S.A.) treated, and the cDNA was synthesized using oligodT primer and reverse transcriptase (Fermentas, St. Leon-Rot, Germany). The cDNA samples were tested for contaminating DNA using PCR primers specific for the *Nin* promoter. Real-time PCR using target and housekeeping gene-specific primers was performed using the LightCycler (Roche Molecular Biochemicals, Mumbai, India) using the FastStart DNA Master SYBR green I kit (Roche Molecular Biochemicals). *ATP-synthase*, *ubiquitin-conjugating enzyme*, *protein phosphatase 2A*, and *TIP41* were used as reference genes (Czechowski et al. 2005). Melting curve analysis and sequencing of the amplified products were used to determine their identity. The relative quantification software (Roche Molecular Biochemicals) was used to determine the efficiency-corrected relative transcript, normalized to a calibrator sample. The geometric mean (Vandesompele et al. 2002) of the relative expression ratios for the three biological and technical replicates and the corresponding 95% intervals of confidence have been calculated. Primers used for RT-PCR are listed in Supplementary Tables S1, S2, and S3.

**Chitin treatments.**

For the chitin treatment, seeds of *L. japonicus* Gifu were germinated on a metal grid positioned on top of plant containers containing clay granules, one-quarter B&D (Broughton and Dilworth 1971) and 5 mM $KNO_3$. Plants were grown for 3 weeks at 21°C under a regime of 16 h of light and 8 h of darkness. For the treatment, the metal grids containing *Lotus* seedlings were submerged for 30 min in liquid medium containing one-quarter B&D (Broughton and Dilworth 1971) supplemented with 5 mM $KNO_3$. A concentration of 1 µM chitooctaose (Sigma) was used for the chitin treatment.

**Mapping of *Lys* genes on the *L. japonicus* genome.**

Microsatellite and derived cleaved amplified polymorphic sequence markers linked to *Lys* genes were identified based on sequences of the *L. japonicus* genome. A test for polymorphism was performed with the parents and the F1 plants of various mapping partners (Sandal et al. 2006), *L. japonicus* Gifu B-129, *L. japonicus* Miyakojima MG-20, *L. filicaulis*, and *L. burttii*. If a polymorphism was identified, the marker was scored in the corresponding F2 or recombinant inbred line populations and a map position was obtained.

**Phylogenetic studies.**

An unrooted maximum likelihood phylogenetic tree was estimated from the amino acid sequence alignment under the WAG + I (invariant sites) + $\gamma$ (four discrete categories) model (Whelan and Goldman 2001) with 100 bootstrap replications. The phylogenetic tree was estimated using the online version of PhyML (Guindon et al. 2005).

The average selective forces that have shaped the evolution of the *Lys* gene family were estimated from the nucleotide alignment using the one-ratio codon model (Goldman and Yang 1994) implemented in the PAML software package (Yang 2007). Less divergent gene pairs, where the synonymous divergence does not show high saturation, were investigated in greater detail using pairwise comparisons. The pairwise comparisons were conducted with the YN00 program of the PAML software package (Yang 2007) using both the likelihood method (Yang and Nielsen 2000) and the simpler pathway method (Nei and Gojobori 1986). The full-sequence pairwise alignment as well as a sliding-window approach with a window size of 30 codons and an overlap of 15 codons was used to estimate the $d_N/d_S$ ratios.

## ACKNOWLEDGMENTS

This work was supported by the Danish National Research Foundation; the Danish Research Agency grant 2113-04-0018; KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas "Comparative Genomics" from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; and the Kazusa DNA Research Institute Foundation.

## LITERATURE CITED


Arrighi, J. F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L. C., Mirabella, R., de Carvalho-Niebel, F., Journet, E. P., Gherardi, M., Huguet, T., Geurts, R., Denarie, J., Rouge, P., and Gough, C. 2006. The *Medicago truncatula* LysM-receptor kinase gene family includes NFP and new nodule-expressed genes. Plant Physiol. 142:265-279.

Baldwin, R. L. 2007. Energetics of protein folding. J. Mol. Biol. 371:283-301.

Bek, A. S., Sauer, J., Thygesen, M. B., Duus, J. Ø., Petersen, B. O., Thirup, S., James, E. , Jensen, K. J., Stougaard, J., and Radutoiu, S. Improved characterization of Nod factors and genetically based variation in LysM receptor domains identify amino acids expendable for Nod factor recognition in *Lotus* spp. Mol. Plant-Microbe Interact. 23:58-66.

Boudeau, J., Miranda-Saavedra, D., Barton, G. J., and Alessi, D. R. 2006. Emerging roles of pseudokinases. Trends Cell Biol. 16:443-452.

Broughton, W. J., and Dilworth, M. J. 1971. Control of leghaemoglobin synthesis in snake beans. Biochem. J. 125:1075-1080.

Cano-Delgado, A., Yin, Y., Yu, C., Vafeados, D., Mora-Garcia, S., Cheng, J. C., Nam, K. H., Li, J., and Chory, J. 2004. BRL1 and BRL3 are novel brassinosteroid receptors that function in vascular differentiation in *Arabidopsis*. Development 131:5341-5351.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W. R. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiol. 139:5-17.

DeYoung, B. J., Bickle, K. L., Schrage, K. J., Muskett, P., Patel, K., and Clark, S. E. 2006. The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in *Arabidopsis*. Plant J. 45:1-16.

Gerard, C., Andrejka, L. M, and Macina, R. A. 2000. Mitochondrial ATP synthase 6 as an endogenous control in the quantitative RT-PCR analysis of clinical cancer samples. Mol. Diagn. 5:39-46.

Gimenez-Ibanez, S., Hann, D. R., Ntoukakis, V., Petutschnig, E., Lipka, V., and Rathjen, J. P. 2009. AvrPtoB targets the LysM receptor kinase CERK1 to promote bacterial virulence on plants. Curr. Biol. 19:423-429.

Goldman, N., and Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11:725-736.

Gomez, S. K., Javot, H., Deewatthanawong, P., Torres-Jerez, I., Tang, Y., Blancaflor, E. B., Udvardi, M. K., and Harrison, M. J. 2009. *Medicago truncatula* and *Glomus intraradices* gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. BMC Plant Biol. 9:10.

Guether, M., Balestrini, R., Hannah, M., He, J., Udvardi, M. K., and Bonfante, P. 2009. Genome-wide reprogramming of regulatory networks, transport, cell wall and membrane biogenesis during arbuscular mycorrhizal symbiosis in *Lotus japonicus*. New Phytol. 182:200-212.

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. 2005. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33:W557-559.

He, Z. H., He, D., and Kohorn, B. D. 1998. Requirement for the induced expression of a cell wall associated receptor kinase for survival during the pathogen response. Plant J. 14:55-63.

Hogslund, N., Radutoiu, S., Krusell, L., Voroshilova, V., Hannah, M. A., Goffard, N., Sanchez, D. H., Lippold, F., Ott, T., Sato, S., Tabata, S., Liboriussen, P., Lohmann, G. V., Schauser, L., Weiller, G. F., Udvardi, M. K., and Stougaard, J. 2009. Dissection of symbiosis and organ development by integrated transcriptome analysis of *Lotus japonicus* mutant and wild-type plants. PLoS One 4:e6556.

Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., Minami, E., and Shibuya, N. 2006. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. Proc. Natl. Acad. Sci. U.S.A. 103:11086-11091.

Lally, D., Ingmire, P., Tong, H. Y., and He, Z. H. 2001. Antisense expression of a cell wall-associated protein kinase, WAK4, inhibits cell elongation and alters morphology. Plant Cell 13:1317-1331.

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

Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N., and Stougaard, J. 2003. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature 425:637-640.

Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., Narusaka, Y., Kawakami, N., Kaku, H., and Shibuya, N. 2007. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. Proc. Natl. Acad. Sci. U.S.A. 104:19613-19618.

Mondragon-Palomino, M., and Gaut, B. S. 2005. Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. Mol. Biol. Evol. 22:2444-2456.

Mondragon-Palomino, M., Meyers, B. C., Michelmore, R. W., and Gaut, B. S. 2002. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. Genome Res. 12:1305-1315.

Nei, M., and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418-426.

Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Gronlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., and Stougaard, J. 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature 425:585-592.

Radutoiu, S., Madsen, L. H., Madsen, E. B., Jurkiewicz, A., Fukai, E., Quistgaard, E. M., Albrektsen, A. S., James, E. K., Thirup, S., and Stougaard, J. 2007. LysM domains mediate lipochitin-oligosaccharide recognition and Nfr genes extend the symbiotic host range. EMBO (Eur. Mol. Biol. Organ.) J. 26:3923-3935.

Rairdan, G., and Moffett, P. 2007. Brothers in arms? Common and contrasting themes in pathogen perception by plant NB-LRR and animal NACHT-LRR proteins. Microbes Infect. 9:677-686.

Rogozin, I. B., Sverdlov, A. V., Babenko, V. N., and Koonin, E. V. 2005. Analysis of evolution of exon-intron structure of eukaryotic genes. Brief. Bioinf. 6:118-134.

Sandal, N., Petersen, T. R., Murray, J., Umehara, Y., Karas, B., Yano, K., Kumagai, H., Yoshikawa, M., Saito, K., Hayashi, M., Murakami, Y., Wang, X., Hakoyama, T., Imaizumi-Anraku, H., Sato, S., Kato, T., Chen, W., Hossain, M. S., Shibata, S., Wang, T. L., Yokota, K., Larsen, K., Kanamori, N., Madsen, E., Radutoiu, S., Madsen, L. H., Radu, T. G., Krusell, L., Ooki, Y., Banba, M., Betti, M., Rispail, N., Skot, L., Tuck, E., Perry, J., Yoshida, S., Vickers, K., Pike, J., Mulder, L., Charpentier, M., Muller, J., Ohtomo, R., Kojima, T., Ando, S., Marquez, A. J., Gresshoff, P. M., Harada, K., Webb, J., Hata, S., Suganuma, N., Kouchi, H., Kawasaki, S., Tabata, S., Hayashi, M., Parniske, M., Szczyglowski, K., Kawaguchi, M., and Stougaard, J. 2006. Genetics of symbiosis in *Lotus japonicus*: recombinant inbred lines, comparative genetic maps, and map position of 35 symbiotic loci. Mol. Plant-Microbe Interact. 19:80-91.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., Sasamoto, S., Watanabe, A., Ono, A., Kawashima, K., Fujishiro, T., Katoh, M., Kohara, M., Kishida, Y., Minami, C., Nakayama, S., Nakazaki, N., Shimizu, Y., Shinpo, S., Takahashi, C., Wada, T., Yamada, M., Ohmido, N., Hayashi, M., Fukui, K., Baba, T., Nakamichi, T., Mori, H., and Tabata, S. 2008. Genome structure of the legume, *Lotus japonicus*. DNA Res. 15:227-239.

Shiu, S. H., and Bleecker, A. B. 2003. Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. Plant Physiol. 132:530-543.

Shiu, S. H., Karlowski, W. M., Pan, R., Tzeng, Y. H., Mayer, K. F., and Li, W. H. 2004. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. Plant Cell 16:1220-1234.

Sivaguru, M., Ezaki, B., He, Z. H., Tong, H., Osawa, H., Baluska, F., Volkmann, D., and Matsumoto, H. 2003. Aluminum-induced gene expression and protein localization of a cell wall-associated receptor kinase in *Arabidopsis*. Plant Physiol. 132:2256-2266.

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., and Bisseling, T. 2007. *Medicago* LYK3, an entry receptor in rhizobial nodulation factor signaling. Plant Physiol. 145:183-191.

Torii, K. U. 2004. Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. Int. Rev. Cytol. 234:1-46.

Trotochaud, A. E., Hao, T., Wu, G., Yang, Z., and Clark, S. E. 1999. The CLAVATA1 receptor-like kinase requires CLAVATA3 for its assembly into a signaling complex that includes KAPP and a Rho-related protein. Plant Cell 11:393-406.

Uchiumi, T., Shimoda, Y., Tsuruta, T., Mukoyoshi, Y., Suzuki, A., Senoo, K., Sato, S., Kato, T., Tabata, S., Higashi, S., and Abe, M. 2002. Expression of symbiotic and nonsymbiotic globin genes responding to microsymbionts on *Lotus japonicus*. Plant Cell Physiol. 43:1351-1358.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3:RESEARCH0034.

Wan, J., Zhang, S., and Stacey, G. 2004. Activation of a mitogen-activated protein kinase pathway in *Arabidopsis* by chitin. Mol. Plant Pathol. 5:125-135.

Wan, J., Zhang, X. C., Neece, D., Ramonell, K. M., Clough, S., Kim, S. Y., Stacey, M. G., and Stacey, G. 2008. A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in *Arabidopsis*. Plant Cell 20:471-481.

Xu, B., English, J. M., Wilsbacher, J. L., Stippec, S., Goldsmith, E. J., and Cobb, M. H. 2000. WNK1, a novel mammalian serine/threonine protein kinase lacking the catalytic lysine in subdomain II. J. Biol. Chem. 275:16795-16801.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586-1591.
Yang, Z., and Nielsen, R., 2000. Estimating synonymous and nonsyonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17:32-43.
Zhang, X. C., Wu, X., Findley, S., Wan, J., Libault, M., Nguyen, H. T., Cannon, S. B., and Stacey, G. 2007. Molecular evolution of lysin motif-type receptor-like kinases in plants. Plant Physiol. 144:623-636.
Zhang, X. C., Cannon, S. B., and Stacey, G. 2009. Evolutionary genomics of LysM genes in land plants BMC Evol. Biol. 9:183.
Zhu, H., Riely, B. K., Burns, N. J., and Ane, J. M. 2006. Tracing nonlegume orthologs of legume genes required for nodulation and arbuscular mycorrhizal symbioses. Genetics 172:2491-2499.

## AUTHOR-RECOMMENDED INTERNET RESOURCES

Center for Biological Sequence (CBS) SignalP server: www.cbs.dtu.dk/services/SignalP
CBS TMHMM prediction of transmembrane helices server: www.cbs.dtu.dk/services/TMHMM
EMBL Inter-ProScan sequence search site: www.ebi.ac.uk/Tools/InterProScan
Gene Index Project DFCI *L. japonicus* gene index database: compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=l_japonicus
Kazusa DNA Research Institute *L. japonicus* sequence webpage: www.kazusa.or.jp/lotus/index.html
The Sanger Institute Pfam database: pfam.sanger.ac.uk