Molecular Evolution of Lysin Motif-Type Receptor-Like Kinases in Plants\[W\][OA]

Xue-Cheng Zhang*, Xiaolei Wu, Seth Findley, Jinrong Wan, Marc Libault, Henry T. Nguyen, Steven B. Cannon, and Gary Stacey

Division of Plant Sciences and National Center for Soybean Biotechnology (X.-C.Z., X.W., S.F., J.W., M.L., H.T.N., G.S.) and Division of Biochemistry, Department of Molecular Microbiology and Immunology (G.S.), University of Missouri, Columbia, Missouri 65211; and United States Department of Agriculture-Agricultural Research Service and Department of Agronomy, Iowa State University, Ames, Iowa 50011 (S.B.C.)

The lysin motif (LysM) domain is an ancient and ubiquitous protein module that binds peptidoglycan and structurally related molecules. A genomic survey in a large number of species spanning all kingdoms reveals that the combination of LysM and receptor kinase domains is present exclusively in plants. However, the particular biological functions and molecular evolution of this gene family remain largely unknown. We show that LysM domains in plant LysM proteins are highly diversified and that a minimum of six distinct types of LysM motifs exist in plant LysM kinase proteins and five additional types of LysM motifs exist in nonkinase plant LysM proteins. Further, motif similarities suggest that plant LysM motifs are ancient. Although phylogenetic signals are not sufficient to resolve the earliest relationships, plant LysM motifs may have arisen through common ancestry with LysM motifs in other kingdoms. Within plants, the gene family has evolved through local and segmental duplications. The family has undergone further duplication and diversification in legumes, where some LysM kinase genes function as receptors for bacterial nodulation factor. Two pairs of homeologous regions were identified in soybean (Glycine max) based on microsynteny and fluorescence in situ hybridization. Expression data show that most plant LysM kinase genes are expressed predominantly in the root and that orthologous LysM kinase genes share similar tissue expression patterns. We also examined synteny around plant LysM kinase genes to help reconstruct scenarios for the evolution of this important gene family.

The lysin motif (LysM) domain was originally identified in bacterial autolysin (Joris et al., 1992). Generally, it is found in bacterial enzymes involved in bacterial cell wall degradation (Jerse et al., 1990; Birkeland, 1994; Frankel et al., 1996; Pellegrini et al., 1999; Fonting et al., 1999). The LysM domain recognizes peptidoglycan, a linear form of N-acetylmuramic acid cross-linked with β(1-4)-linked GlcNAc (GlcNAc) by short peptides and a major component of the cell walls of both Gram-positive and Gram-negative bacteria. The LysM motif is usually about 40 amino acids in length (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF01476). Despite the fact that more than 1,500 LysM proteins are identified in Pfam (http://www.sanger.ac.uk/cgi-bin/Pfam/speciesdist.pl?id=PF01476&depth=all), the three-dimensional structure of the LysM motif has been determined for only two bacterial proteins, Escherichia coli membrane-bound lytic murein transglycosylase D (Bateman and Bycroft, 2000) and Bacillus subtilis YkuD (Bielnicki et al., 2006). These structures allowed sequence-based homology modeling of the LysM domain of the Nod factor perception (NFP) protein from the model legume Medicago truncatula (Mt; Mulder et al., 2006). In these three structures, the LysM motif is characterized as a βαβ secondary structure with the two α-helices stacking onto one side of a plate made up of a two-stranded antiparallel β-sheet.

More recently, increasing amounts of transcript and genomic sequence have allowed identification of LysM-encoding proteins in a broad range of organisms spanning all kingdoms except archaea (Bateman and Bycroft, 2000). This suggests that the LysM motif is ubiquitous, and that plant LysM domains remain elusive.

LysM proteins in plant species have attracted increasing attention since the identification of NFR1 and NFR5 in another model legume species, Lotus japonicus...
(Lj; 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 a typical structure of plant kinase receptors. Therefore, they are considered to be founding members of the plant LYK family. Genetic and molecular evidence suggested that they are receptors for the bacterial Nod factor, a GlcNAc lipochitooligosaccharide with various modifications and structurally similar to peptidoglycan. Consequently, homologous proteins were identified and characterized in other legume species. The LjNFR1 homologs include MtLYK3 and MtLYK4 (Limpens et al., 2003), pea (Pisum sativum) SYM2 (Limpens et al., 2003), and the NFR1a and NFR1b of soybean (Glycine max; P. Gresshoff, personal communication). The LjNFR5 homologs include MtNFP (Ben Amor et al., 2003; Arrighi et al., 2006), pea SYM10 (Madsen et al., 2003), and NFR5a and NFR5b of soybean (P. Gresshoff, personal communication). However, bioinformatic and genomic analyses suggest that Arabidopsis (Arabidopsis thaliana), rice (Oryza sativa), and M. truncatula genomes harbor a large number of LYK genes (Shiu et al., 2004; Arrighi et al., 2006). The characterized LYK genes listed above constitute only a small portion of the LYK gene families. Plant LYK genes likely have various functions other than establishing symbiotic relationships, especially in nonlegume plants.

Although the LysM domain is associated with a variety of protein domains across a large number of organisms spanning all kingdoms, it is intriguing that the linkage between the LysM domain and the protein kinase domain appears to occur exclusively in plants (Bateman and Bycroft, 2000). Two types of kinase domains, either predicted to be active or inactive, are found linked with the LysM domain (Madsen et al., 2003; Arrighi et al., 2006; Stacey et al., 2006). Compared to nonplant LysM proteins, plant LYKs possess additional unique features: (1) LysM domains embedded in plant LYK proteins are highly diversified; (2) there are no more than three LysM motifs in any individual LYK protein; and (3) plant LYKs seem to play a role in signal transduction rather than in enzymatic metabolism.

Based on the molecular phylogeny of plant kinase domains, the LYK proteins in Arabidopsis and rice were categorized into two clades, LysM-I and LysM-II (Shiu et al., 2004). Two subsequent studies assigned LjNFR1 and LjNFR5 to the LysM-I and LysM-II clades, respectively (Arrighi et al., 2006; Zhu et al., 2006). In a study tracing the nonlegume orthologs of legume LYK genes, the orthologs of LjNFR1 and LjNFR5 were identified in nonlegume plants, such as Arabidopsis and rice, based on sequence similarity and microsyntenic relationships (Zhu et al., 2006). Arrighi et al. (2006) also identified a large number of LYK genes in M. truncatula by in silico mining of EST and genome sequences and described the genomic distribution, gene structures, and tissue expression patterns of some of the MLTYK genes. However, our knowledge of the molecular evolution and comparative genomics of these genes remains limited due to poor availability of plant genome sequences. Detailed systematic studies using additional legume and nonlegume LYK sequences are needed to generalize the evolutionary and genomic characteristics of plant LYK genes across multiple species. These studies will provide valuable clues to investigate the biological functions of this newly formed, but very important, gene family.

Here we report comprehensive characterization of plant LysM domains and the molecular evolution and comparative genomics of the plant LYK gene family. Our data show that plant genomes harbor a minimum of 11 distinct types of LysM motifs. Plant LYK genes have duplicated locally, due to whole-genome duplications, and have subsequently functionally diversified. Plant LYK proteins fall into three major clades: two are represented by LjNFR1 and LjNFR5 and the third major clade has remained undescribed. We determined orthologous and paralogous relationships of plant LYK genes based on sequence similarities, molecular phylogenies, nucleotide substitution rates, genomic microsynteny (conserved gene content and order), and tissue expression patterns. We observed strong microsynteny in LjNFR1 and LjNFR5 orthologous regions across multiple species and dispersed microsynteny in ancestral LYK genes in more distantly related plant species.

RESULTS

Genome-Wide Exploration of Plant LYK Genes

Six plant species, Arabidopsis, rice, M. truncatula, L. japonicus, poplar (Populus trichocarpa), and soybean, are included in this study. The genomes of the first five species are either completed or close to being completed, whereas the random shotgun sequencing of the soybean genome is under way. We used the LysM domain sequences of LjNFR1 and LjNFR5 to search the public databases of Arabidopsis, rice, poplar, M. truncatula, and L. japonicus (see “Materials and Methods”). We identified soybean LYK genes by shotgun sequencing bacterial artificial chromosomes (BACs) with homologies to LysM-encoding ESTs (see “Materials and Methods”). The resulting putative LYK protein sequences from all species were then searched against the Pfam server to verify LysM and kinase domains. Collectively, a total of 48 LYK genes were identified in the six plant genomes (Supplemental Table S1).

Characterization of LysM Domains and LysM Motifs in Plants

In comparison to the LysM proteins in other kingdoms, plant LYK proteins possess unique features: (1) the combination of LysM and kinase domains exists
exclusively in the plant lineage; (2) plant LYK proteins have no more than three LysM motifs; (3) if more than two LysM motifs exist within a single plant LYK protein, they are always distinct from each other at the protein sequence level; and (4) the LysM domain sequences in plant LYK proteins are highly diversified due to different combinations of heterogenous LysM motifs. These facts led us to investigate the evolution of this fascinating plant LYK gene family and the phylogenies of these diversified LysM motifs. Based on the sequence phylogenies, LysM motifs (named LYKa, LYKb, and LYKc from the N to the C terminus) in plant LYK proteins largely fall into five clades (Fig. 1A; Supplemental Fig. S1). This distribution of LysM motifs was found in all six plant species studied (i.e. LysM motifs from dicots and rice are clustered together in each clade, suggesting that the diversification event of plant LysM motifs predated the divergence of monocot and dicot plants).

We further investigated the LysM motifs from non-kinase plant LysM proteins, retrieving these using BLAST searches against genomic sequence databases of Arabidopsis, rice, and poplar and EST sequences of soybean (see “Materials and Methods”). Based on their subcellular localization predictions and domain arrangements, nonkinase plant LysM proteins can be further categorized into three subgroups, including LysM-type receptor-like proteins (LYPs), extracellular LysM proteins (LysMe), and nonsecretory intracellular LysM proteins (LysMn; Fig. 1B). This grouping will be helpful in understanding the nature of each LysM protein and providing insightful clues to the biological functions. As predicted by Pfam, LYP proteins have exactly two LysM motifs and LysMe and LysMn proteins have only one LysM motif. Sequence alignments show that, among the 11 types of LysM motifs, motif sequences of LysMn (the motif within LysMn proteins, LysM motif type XI), one group of LysMe (the motif within LysMe proteins, LysM motif type X), and one group of LYPb (the second motif from the N terminus within LYP proteins, LysM motif VII) are extremely conserved (Supplemental Fig. S2). In these motifs, the amino acid identities averaged across the alignments are 91% for LysMe (type X), 86% for LysMn (type XI), and 75% for LYP (type VII; Supplemental Fig. S2). LysMn motif sequences always start with a His and end with a Pro. Similarly, LYPb motif sequences always end with a Pro. LYKa motifs are seven to 10 residues shorter.

To determine the phylogenetic relationships of plant LysM motifs, we calculated a majority-rule parsimony tree (Fig. 1A) using every plant LysM motif sequence (one to three LysM motif sequences from individual LysM proteins; Supplemental Table S2). Neighbor-joining and maximum-likelihood trees also showed generally similar topologies (data not shown). Plant LysM motifs can be separated into 11 early diverging clades (Fig. 1A; Supplemental Fig. 1). Most of these clades contain sequences from both monocots and dicots, indicating that the duplication events leading to these LysM proteins occurred prior to the monocot-dicot split. Domain and motif arrangements are shown in Figure 1B.

Figure 1. Eleven distinct types of LysM motifs in plants. A, The evolutionary relationships of plant LysM motifs. The majority-rule parsimony phylogeny was calculated using LysM motif sequences from six plant species. Maximum-likelihood branch lengths were calculated using Tree-Puzzle. CHLRE1 is a green algae LysM sequence used to root the tree. Numbers on the branches are the bootstrap values of 1,000 parsimony trees. A distance tree and a Bayesian likelihood tree were also calculated with similar topology. Each clade of plant LysM motifs was collapsed for simplifying the figure. LysM motifs in a given protein were named alphabetically from N to C termini. Eleven distinct types of LysM motifs were identified and denoted in the right by a symbol shown in B. B, Subcellular localization and LysM domain structures of LysM proteins in plants. Eleven types of LysM motifs, shown in roman numerals, are illustrated with specific symbols individually. The lengths of LysM proteins are roughly in scale.
The Origins of Plant LysM Motifs

To understand the origins of plant LysM motifs, we identified LysM motif sequences (Supplemental Table S2) from nonplant species, including bacteria, fungi, insects, and animals, and calculated phylogenetic trees using majority-rule parsimony (Fig. 2; Supplemental Fig. S3), neighbor joining (Supplemental Fig. S4), and maximum likelihood (Supplemental Fig. S5). Consistent with our notion that plant LysM motifs are highly diversified, they can be classified into several multi-kingdom clades characterized by distinct motifs (Fig. 2; Supplemental Figs. S3–S5). Although bootstrap support is generally low, most of the indicated clades are rooted with bacterial LysM motifs and at least two bacterial-rooted clades (XI and the clade above V) include sequences from fungi, worms, insects, plants, and animals in expected taxonomic order. This suggests that several LysM motifs may be very ancient, with common origin predating the divergence of fungal, insect, plant, and animal lineages.

Phylogenies of Plant LYK Proteins

We calculated plant LYK phylogenies using either LysM domain sequences (all LysM motifs sequences + spacer sequences) or the full protein sequence (LysM + kinase domain). Notably, the two sets of LYK phylogenies calculated using the parsimony method matched each other quite well. Therefore, only the full-length sequence trees are represented in this study. The plant LYK phylogenies calculated using parsimony (Fig. 3), distance, and maximum-likelihood methods showed similar topologies (Supplemental Figs. S6 and S7). A parsimony tree with maximum-likelihood branch lengths is shown in Figure 3, with supporting values calculated using the parsimony method shown for supporting branches. Generally, the plant LYK phylogeny reflects species phylogeny [i.e. in an evolutionary direction of (rice, Arabidopsis, (poplar, (legume)))) in most of the clades]. Five well-supported, distinct, multi-plant family clades are evident in Figure 3 (indicated by bold horizontal lines). Three of them contain more than 10 members, whereas the other two clades consist of only three genes each. LjNFR1 and LjNFR5 fall into separate clades, consistent with previous studies (Arrighi et al., 2006; Stacey et al., 2006; Zhu et al., 2006). A third large clade with no assigned biological functions was mentioned by Arrighi et al. (2006); we make additional evaluations of this clade in this study. The LjNFR5 clade and the undefined large clade are sisters, probably resulting from ancient duplication (predating the monocot-dicot split). Similar topologies are also evident between the two upper subclades embedded in the LjNFR5 and its sister clades. We term them sister subclades. Besides these large clades, the two small clades of LYK proteins are distantly related to and apparently more ancestral to the above major clades.

MtLYK13 (MtNFP) and GmNFR5α cluster with LjNFR5 with very high bootstrap values, consistent with their functional similarities (Arrighi et al., 2006), and are most likely orthologs. Similarly, MtLYK3 and GmNFR1α are considered to be LjNFR1 orthologs based on the phylogeny and functional similarities (Limpens et al., 2003). Interestingly, three nonlegume LYK proteins, PtLYK2, PtLYK11, and OsLYK5, are orthologous to LjNFR5, which functions in Nod factor
Figure 3. Phylogeny of plant LYK proteins. Parsimony tree with maximum-likelihood branch lengths was calculated using full-length LYK sequences. A distance tree (Supplemental Fig. S6) and a maximum-likelihood tree (Supplemental Fig. S7) were calculated with similar topologies. Bootstrap values larger than 70 on each branch were imposed from the majority-rule consensus of 1,000 maximum parsimony trees. The tree was rooted using two green algae LysM proteins, Vcchitinase (AAC13727) and Creye2 (AAF43040). Plant LYK proteins fall into three major clades and two minor clades separated by solid horizontal lines. Specific subclades listed on the right of vertical lines were characterized by incorporating data from synteny (Fig. 4). The LysM domain arrangements are represented by the LysM motifs (Fig. 2) listed from N to C termini. Numbers of exons...
recognition during nodulation. Accordingly, in its sister subclade, four nonlegume LYK proteins, PtLYK3, PtLYK6, PtLYK9, and AtLYK4, are orthologous to a group of legume LYK proteins, including GmLYK4, GmLYK7, LjLYK4, and MtLYK12. Unlike the above two clades, AtLYK1 is orthologous to the entire clade of legume members, but is distantly and weakly associated with the LjNFR1 ortholog subclade.

Synteny and Fluorescence in Situ Hybridization Results for Plant LYK Orthologous Regions

Phylogenetic relationships alone are usually not sufficient to infer orthology, but can be strengthened by information derived from genomic contexts. The most strongly conserved gene collinearity (also synteny or microsynteny) lies in LjNFR5 orthologous regions. These involve 11 blocks across the six species studied (Fig. 4A). Less-conserved synteny was observed in the LjNFR1 orthologous regions (Fig. 4D). More degraded synteny also exists in ancestral LYK regions involving mainly two species (Fig. 4, B and C). In all these cases, the conservation was not only observed in gene content and order, but also in the gene orientations, except that MtLYK3 and MtLYK4 genes are in reverse orientations. We took advantage of the LYK phylogenies, we claim the following four sets of orthology: LjNFR5 orthology, including MtLYK13 (MNF), GmNFR5a, GmNFR5b, PtLYK2, PtLYK11, and OsLYK5; LjNFR1 orthology, including MtLYK3, GmNFR1a, GmNFR1b, and AtLYK1; GmLYK4 orthology, including GmLYK7, LjLYK4, MtLYK12, PtLYK5, PtLYK6, PtLYK9, and AtLYK4; MtLYK4 orthology; including LjLY2, GmLYK2, GmLYK3, LjLY3, and MtLYK1. Interestingly, the two LjNFR5 sister subclades, LjNFR5 orthology and GmLYK4 orthology, are located head to head in the genome. We redefined GmLYK4 orthology subclade as the LjNFR5 paralog subclade I and the two subclades in LjNFR5 clades as LjNFR5 paralog subclades II and III, respectively (Fig. 3). Similarly, the MtLYK4 subclade, located head to tail with the LjNFR1 ortholog in the genome was redefined as the LjNFR1 paralog subclade (Fig. 3).

The most strongly conserved microsyntenies are observed between the GmNFR1a-GmNFR1b (Fig. 4D) and GmNFR5a-GmNFR5b regions (Fig. 4A). GmNFR1a and GmNFR1b share a very high percent sequence identity of 87%. Further analyses show that genes surrounding GmNFR1a are directly correlated with those surrounding GmNFR1b on a one-to-one basis with unusually high percent identity and conservation of predicted numbers of amino acids and gene orientations, but not gene structures (data not shown). Similarly, GmNFR5a is 94% identical to GmNFR5b and genes surrounding GmNFR5a and GmNFR5b are also highly conserved (data not shown). We suspect that they are recent 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). Seeking direct supporting evidence, we performed fluorescence in situ hybridization (FISH). A GmNFR5b-containing BAC (WBB095P01), labeled with Texas red hybridized to two pairs of two linked spots (Fig. 5A, middle). We speculate that the two brighter pairs of spots correspond to the generic GmNFR5b region, whereas the two dimmer pairs of spots correspond to a homeologous region. GmNFR5a-containing BAC (WBB035N07), labeled with AlexaFluor, hybridized to two pairs of spots at relatively lower brightness (Fig. 5A, right). Interestingly, the two pairs by GmNFR5a-containing BAC overlap with the two pairs of dimmer spots by GmNFR5b-containing BAC (Fig. 5A, red arrow). In the swapped color experiment (Fig. 5B), at least one overlapping spot pair (red arrow) is apparent. Similarly, GmNFR1a-containing BAC (WBB098N11) and GmNFR1b-containing BAC (WBB098N15) hybridized to at least one overlapping spot pair (Fig. 5, C and D). These facts strongly suggested that the two pairs of genes, GmNFR5a-GmNFR5b and GmNFR1a-GmNFR1b, indeed lie in homeologous regions.

We also noticed gene duplications, either large scale or local, followed by gene diversification or gene loss. Tandem duplications of LYK genes in the NFR5 syntenic regions are present in both legume and poplar genomes, but not in rice and Arabidopsis, suggesting that this duplication predates the split of poplar and legumes (both in Rosid I). It is also clear that LjNFR5 orthologous regions are remnants of segmental (probably whole-genome) duplications. At least two scenarios concerning the fates of duplicated LYK genes are possible. In the first scenario, LjNFR5 homologs in legumes evolved a new function involved in nodulation. In the second scenario, LjNFR5 homologs in MtLYK9 and PtLYK5 blocks were lost, most likely after the large-scale duplication. Gene duplications also occurred in the LjNFR1 syntenic regions, but seem to exist exclusively in legumes. Consistent with previous results (Limpens et al., 2003), our data suggest that genomic regions surrounding the MtLYK3 gene underwent significantly high levels of local gene rearrangement. It is also obvious that LjNFR1 orthologs evolved new
functions (related to nodulation) after duplication from an ancestral sequence and that the GmLYK2 counterpart is lost in the GmNFR1β region (Fig. 4).

**Nucleotide Substitution Rates**

We calculated rate changes in user-defined parsimony phylogenies based on synonymous (dS) sites, nonsynonymous (dN) sites, and dN/dS ratios. The topologies of the three trees agree for branches under purifying selection, but vary for branches that have undergone rapid changes. Figure 6 shows the dN/dS topology with average dN/dS ratios calculated for each clade and subclade only on terminal branches. As shown in the dN/dS tree (Fig. 6), the average terminal dN/dS ratios of the LjNFR5 clade and its sister clade are slightly greater than 1. However, the average dN/dS ratio of LjNFR5 orthologs is significantly less than 1. It is also notable that, in the codon alignments for calculating dN/dS ratios, all insertions and deletions

**Figure 4.** Collinearity of genomic blocks surrounding plant LYK genes. In each block, genes are shown in black boxes with orientations indicated by a triangle. The blocks are colored differently for each plant species. Plant LYK proteins are shown in inverted colors in each block and labeled from top to bottom. Orthologous genes are connected using solid black lines. Red lines indicate the loss of an orthologous gene in particular blocks. PtLYK6 and PtLYK2 block b is below block a, with a distance of about 30 genes.
were removed except for a gap of more than 30 nucleotides that was retained to demonstrate the lack of the p loop and the activation loop in the kinase domains of LjNFR5 orthologs (Limpens et al., 2003; Madsen et al., 2003; Arrighi et al., 2006). The retention of this gap may account for the higher dN/dS ratio. In short, the LjNFR5 orthologs are under purifying selection. LjNFR5 paralog subclade I has a ratio close to 1, whereas LjNFR5 paralog subclades II and III have a ratio greater than 1, suggesting that LjNFR5 paralogs are under neutral and diversifying selections. The LjNFR1 clade, in general, is under purifying selection. Moreover, LjNFR1 orthologs are under strong purifying selection. There is also strong correlation between the high degree of sequence identity, conserved synteny (Fig. 4), and low dN/dS ratios of LjNFR5 and LjNFR1 orthologs (Fig. 6). This indicates that our analyses mutually support each other and that LjNFR5 and LjNFR1 orthologs might have indispensable functions.

The LYK genes under rapid nucleotide changes can be classified into two categories: duplicated LYKs in syntenic blocks and singleton LYKs that show no or degraded synteny. In the first category, GmNFR5β, MtLYK9, LjLYK4, and PtLYK9 lie in syntenic blocks (Fig. 4). The higher rates of nucleotide changes of these genes are likely due to the relaxation of selection pressure after duplications. This is consistent with the observations that duplicated genes that are retained evolve slower than their singleton partners (after an
initial evolutionary rate increase), and that evolutionary rates for duplicated genes are inversely correlated with the copy of paralogs per gene (Jordan et al., 2004). Although they are not in a syntenic block, MtLYK10 and MtLYK11 are also tandem duplications and, therefore, also belong to the first category. The second category includes GmLYK3, GmLYK10, PtLYK7, AtLYK2, PtLYK4, AtLYK5, PtLYK3, GmLYK8, GmLYK3, AtLYK3, PtLYK8, and GmLYK11, which show little or no synteny. These suggest that they are located in genomic regions that are under rapid nucleotide change and may be critical in the plant’s adaptive evolution.

Comparative Tissue Expression Patterns of Plant LYK Genes

For the six plant species in this study, tissue expression levels of LYK genes were only reported for M. truncatula (Limpens et al., 2003; Arrighi et al., 2006) and Arabidopsis (https://www.genevestigator.ethz.ch/at/index.php?page=3). Therefore, we measured LYK gene expression using quantitative reverse transcription (RT)-PCR in different tissues of soybean, M. truncatula, and rice plants. Our data agree well with previous results on MtLYK expression levels (Limpens et al., 2003; Arrighi et al., 2006). Generally, we found that plant LYK expression is tissue regulated and that most plant LYK genes are expressed predominantly in the root in soybean, M. truncatula, rice (Fig. 7), L. japonicus (Madsen et al., 2003; Radutoiu et al., 2003), and Arabidopsis (https://www.genevestigator.ethz.ch/at/index.php?page=3), although a few genes are expressed in stems and leaves. As predicted from their orthologous relationships, GmNFR1a, GmNFR1b, MtLYK3, and LjNFR1 (Radutoiu et al., 2003) showed similar patterns of root-specific expression. Similarly, GmNFR5a, GmNFR5b, MtLYK13, OsLYK5, and LjNFR5 (Madsen et al., 2003) showed root-specific expression. These results are reasonable, from a biological perspective, because these receptors need to efficiently contact Nod factors secreted by soilborne symbiotic bacteria. Additionally, there are similar expression patterns for

![Figure 7](image-url)

**Figure 7.** Tissue expression patterns of LYK genes in soybean, M. truncatula, and rice. Expression levels of each LYK gene were displayed in artificial scales relative to particular housekeeping genes. Data were collected from three biological replicates. Error bars represent sds.
the following sets of orthologous genes: GmLYK4 and MtLYK12; GmLYK10 and AtLYK2 (https://www. genevestigator.ethz.ch/at/index.php?page=3); GmLYK8 (data not shown); and AtLYK5 (https://www. genevestigator.ethz.ch/at/index.php?page=3). Interestingly, several duplicated genes displayed different expression patterns. For example, GmLYK2 and MtLYKII expression is dramatically different from duplicated partners, GmNFR1a and MtLYK10, respectively. MtLYK9, paralogous to MtLYK13, is expressed differently from the latter. This clearly suggests the functional diversification of LYK genes after duplications.

Comparative Genomics of Plant LYK Genes

The *M. truncatula*, soybean, and poplar genomes harbor larger numbers of LYK genes than the Arabidopsis and rice genomes (Fig. 3; Supplemental Table S1). That poplar would harbor large numbers of LYK genes (11) is not surprising because it has changed more slowly (and lost duplicated genes more slowly) than Arabidopsis or *Medicago* since their common ancestors (Tuskan et al., 2006). The finding of larger numbers of LYK genes in *M. truncatula* than in Arabidopsis or rice is intriguing because each genome is suspected to have undergone the same numbers of independent genome duplications (Cannon et al., 2006; Tuskan et al., 2006). A plausible explanation is that more of the LYK duplicates have been retained in the legumes than in Arabidopsis or rice because the duplicated genes in the legumes have acquired important new functions. We would also predict that soybean, which has undergone another relatively recent round of whole-genome duplication, would also have the largest number of LYK genes.

Tandem duplications of plant LYK genes are common in legume and poplar plants at a percentage of more than 50% per LYK-containing region. In total, 10 pairs of tandemly duplicated LYK duos were identified in rice, soybean, *M. truncatula*, and poplar, and one LYK trio was identified in the *Lotus* genome (Figs. 3 and 4). A large cluster of LYK genes (MtLYK1–7) was identified (Limpens et al., 2003; Arrighi et al., 2006). However, MtLYK2 and MtLYK5, MtLYK6, and MtLYK7 were left out of this study because MtLYK2 does not have a kinase domain, MtLYK5 is considered to be a pseudogene (Limpens et al., 2003; Arrighi et al., 2006), and MtLYK6 and MtLYK7 do not have strongly predicted LysM domains at an E-value cutoff of 0.1. Arrighi et al. (2006) stated that the genomic distribution of LYK genes in *M. truncatula* is uneven. This seems to be true in Arabidopsis and rice, but not in the other three studied plant genomes, probably because the chromosome locations of the LYK genes are not completely determined.

There are two types of kinase domains, predicted to be either active or inactive, found in plant LYK proteins. It was predicted that LjNFR1 and its close homologs have an intact, functional kinase domain, whereas LjNFR5 and its orthologs each have a short, inactive kinase domain, due to the lack of the p loop and the activation loop (Limpens et al., 2003; Madsen et al., 2003; Arrighi et al., 2006). As predicted, MtLYK3 (LjNFR1 homolog) showed autophosphorylation activity, whereas MtNFP did not (Arrighi et al., 2006). Interestingly, the inactive kinase domain is the hallmark of LjNFR5 orthologs, not any of their paralogous partners.

LYK genes have various numbers of exons, ranging from one to 13 (Fig. 3). Comparisons of gene structures suggest that ancestral LYK genes, LjNFR1 orthologs, and paralogs have 10 to 13 exons, whereas the rest of the LYKs, including LjNFR5 orthologs and paralogs, have one to five exons. All LjNFR5 orthologs, except PtLYK11, which has two exons, have an intronless gene structure. PtLYK11 is unusually small compared to PtLYK2 and its orthologs. The presence of one intron and small size may be due to incorrect annotation, resulting from the low-quality genomic sequence generated via a random whole-genome shotgun approach.

DISCUSSION

Plant LysM and LYK genes are common and important in plants, yet their biological functions, molecular evolution, and comparative genomics have not been well understood. In this study, we performed comprehensive molecular evolutionary and comparative genomic studies of the LYK gene family. The goal was to reconstruct a plausible evolutionary scenario and to generalize genomic characteristics of the plant LYK gene family. Our focus is on kinase-type LysMs rather than nonkinase types for several reasons: (1) There are limited genomic sequences available in legumes so far; (2) biological functions have not yet been assigned to most nonkinase LysMs; and (3) the existing nonkinase LysMs in rice, Arabidopsis, and poplar are quite divergent from LYK genes, making it difficult to integrate them together with LYK genes for evolutionary and genomic analyses.

Recruitment of Plant Species and Recommended Nomenclatures

Compared to *M. truncatula*, *L. japonicus*, rice, and Arabidopsis, soybean and poplar have been the subject of relatively few evolutionary studies due to the lack of finished genomic sequence. Use of soybean and poplar in this study, by means of cloning soybean LYK genes in a near-saturated manner and utilizing the newly completed poplar genome sequence, allowed us to expand on previous studies of plant LYKs (Arrighi et al., 2006; Zhu et al., 2006). Until the completion of poplar genome sequencing (Tuskan et al., 2006), the closest nonlegume species in direct comparison to the Rosid I legumes was the Rosid II Arabidopsis. The use of poplar in this study allows us to better generalize key features of legume, dicot, and broader plant LYK genes. For example, Zhu et al. (2006) stated that the
legume-specific duplication of LYKs and maintenance of ancient gene duplication played important roles in legume-specific root nodule symbiosis. However, our data show that LjNFR5 orthologous genes also persist in the poplar genome and, therefore, must play other roles in plant biology besides nodulation.

In this study, we adapted the nomenclature of LYK from Limpens et al. (2003). We also introduced a nomenclature system, taking into account domain arrangements and phylogenetic relationships, for non-kinase-type LysM proteins (Fig. 1). We would like to recommend that the research community follow this nomenclature system for uncharacterized plant LysM proteins. The information in Supplemental Table S1 serves as a good index for plant LysM proteins.

Mode of Origins of Plant LYK Genes

There are several distinct types of LysM motifs in nature (Fig. 2). Although the alignments are difficult and bootstrap values are low for some branches, our data suggest that LysM motifs fall into several distinct eukaryotic clades, possibly with separate bacterial origins. However, given the great evolutionary distances and short and diverse sequences, it is not possible to rule out convergent evolution of motif patterns or misinformative groupings on the basis of alignment artifacts.

Plant LysM domains have evolved differently from nonplant LysM domains in several ways. If there are two LysM motifs in an individual plant LYK or LYP protein, they are heterogenous from each other. This is a phenomenon predominantly observed in plants, but not other lineages. The LysM motifs, if more than two, in an individual nonplant LysM protein are similar or almost identical. An extreme example is the nematode gene (NP_504862), which contains 12 nearly identical tandem LysM motifs.

The fact that LYK genes exist exclusively in plants suggests that plants gave birth to LYK genes via an ancient de novo event. The simplest explanation is that membrane-bound LYP proteins fused with plant-specific kinase domains to become LYK proteins. LYP recognize peptidoglycan-like molecules in the extracellular space of plant cells. It is likely that LYP transduces the signals into the inner part of plant cells via an intracellular kinase domain. We calculated a tree of the topology of plant LysM motifs (Fig. 1), but cannot ascertain whether LYP occurred prior to LYK proteins in plants. However, we are able to track the duplications and losses of LYK genes in plants (Figs. 3 and 4). As predicted from the phylogenetic tree (Fig. 3), there are probably multiple ancestral LYK genes in primitive plant species. These ancestral LYKs underwent multiple rounds of duplication and speciation in descendants and more recently evolved plant species (Fig. 4), producing a significant reservoir of genetic resources for neo- and subfunctionalization. One notable success resulting from the expansion of LYKs in plant genomes is the formation of Nod factor receptors (LjNFR1 and LjNFR5 and their legume orthologs) when legumes evolved nodulation functions approximately 60 to 70 million years ago.

The Diverse and Complex Nature of LysM Domains

Striking features about the LysM domains are their diversity and complexity in almost every organism. Besides the fact that we identify a minimum of 11 distinct types of LysM motifs in plants, the numbers and combinations of LysM motifs in individual plant LysM proteins vary from one to another. This probably reflects the complexity of the ligands, which are mostly peptidoglycan-like oligosaccharides. The presence of multiple, diverse LysM domains in a single protein may increase ligand affinity and allow for a wider range of ligand binding.

As stated above, LysM domains have undergone clearly different evolutionary paths in animals and plants. These differences likely reflect the differences in ligand chemistry encountered by the LysM motif. Mulder et al. (2006) predicted the docking sites of Nod factors on the LysM domain of MtNFP protein and proposed that approximately one LysM motif binds a tetrasaccharide moiety that is the backbone of the Nod factor. If true, one can imagine that plant LysM proteins with two or three motifs are capable of binding either multiple tetra- or pentasaccharide moieties or saccharides with a higher degree of polymerization, for example, an octosaccharide.

Duplication, Polyploidy, and Functional Diversification of the LYK Gene Family

Clearly, genomic duplications (such as polyploidy) have played an essential role in genomic expansion and functional diversification of plant LYK genes. At least two rounds of duplication, one local tandem duplication and one large scale (maybe genome wide), occurred in LjNFR5 orthologous and paralogous I regions in Rosid I plants (Fig. 4). The timing of these two duplications is therefore critical to understanding the dynamics of LYK duplications and genomic expansion. We believe that the local tandem duplication occurred before the large-scale duplication because it is improbable that LYK genes within large-scale duplicated fragments underwent tandem duplication with high similarity and stringency (head to head, low-nucleotide substitution rate, and similar expression pattern). Another piece of evidence comes from the fact that three blocks syntenic to the LjnFR5 region were identified in poplar, suggesting a scenario of one local duplication followed by two large-scale duplications. Cannon et al. (2006) proposed that a legume-specific whole-genome duplication occurred after the split between poplar and legumes. Our data are consistent with this. We also observe the results of two probable duplications in poplar (Tuskan et al., 2006) with trees, suggesting one of these large duplications may have occurred independently in poplar, following
the split with the legumes. Despite whole-genome duplication occurring in the common ancestor of *M. truncatula* and *L. japonicus*, we found only one syntenic block in *L. japonicus*. We predict that further genomic sequences will reveal the duplicated LjNFR5 microsyntenic block. We also observed two fates of LYK genes after duplications. The tandemly duplicated genes persist in the genome in a birth-and-death manner (Michelmore and Meyers, 1998) before speciation, whereas some duplicated genes were lost after large-scale duplications (Fig. 4).

Tandem duplications were also observed in LjNFR1 orthologous regions, but in a legume-specific manner. In contrast to the LjNFR5 orthologous regions, LjNFR1 regions underwent dynamic changes after duplication. The evidence for this includes: (1) GmNFR1β has a higher nucleotide substitution rate (Fig. 6); (2) the GmLYK2-like gene was lost in the GmNFR1β region (Fig. 4); (3) GmLYK2 has a distinct expression pattern from its partner GmNFR1α (Fig. 7); and (4) one or more rounds of tandem duplications occurred in *L. japonicus* and *M. truncatula*, respectively (Fig. 4). Presumably, these legume-specific duplications in LjNFR1 orthologous regions may play important roles in symbiosis. LjNFR1 and LjNFR5 both have orthologs in nonlegume species (Figs. 3 and 4) that are generally under purifying selection (Fig. 6), suggesting that LjNFR1, LjNFR5, and their legume orthologs might share additional functions with their nonlegume orthologs. For example, LjNFR1, LjNFR5, and their orthologs may play an important role in arbuscular mycorrhizal (AM) symbiosis. In our study, the LjNFR5 orthologous subclade is the only informative group that does not have a member from Arabidopsis, which is a nonmycorrhizal species (Figs. 3 and 4). Although AtLYK4 lies in the LjNFR5 syntenic regions, it actually is the ortholog to the duo partners that are tandemly duplicated from LjNFR5 orthologs as judged by sequence similarity (Fig. 3) and gene orientation (Fig. 4). These data suggest that the LjNFR5 orthologous group may be involved in AM symbiosis. However, this suggestion is contradicted by the fact that mutants of LjNFR1 and LjNFR5 are not impaired in the AM symbiosis (Madsen et al., 2003; Radutoiu et al., 2003). However, it remains to be seen whether mutations in other members of the NFR1 and NFR5 gene family in these plants would affect mycorrhizal symbiosis. In short, direct evidence is needed to clarify whether LjNFR5 orthologs are involved in mycorrhizal symbiosis.

### MATERIALS AND METHODS

**Retrieval of Plant LYK Sequences and LysM Motif Sequences from All Kingdoms**

Plant LysM protein sequences were first searched using the key word LysM and BLASTp (1e-20) using the LysM domains of LjNFR1 and LjNFR5 against the following databases: Arabidopsis (*Arabidopsis thaliana*; www.arabidopsis.org); rice (*Oryza sativa*; http://www.tigz.org/tlb/ezki/osa1); poplar (*Populus* spp.; http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html); *Medicago truncatula* (http://www.medicago.org/genome); and *Lotus japonicus* (http://www.kazusa.or.jp/lotus). Domain structures of the resulting potential LysM proteins were analyzed with Pfam (www.sanger.ac.uk/software/Pfam) and Inter-ProScan (http://www.ebi.ac.uk/Inter-ProScan) to identify LysM proteins. Soybean *Glycin max* LysM proteins were searched via tBLASTn (1e-5) using the same query sequences as above against the two EST databases: The Institute for Genomic Research (http://compbio.dii.c. harvard.edu/cgi/ cgi-bin/tgi/tgi/mipview.pl?gdb=soybean) and Monsanto (http://www.soyseseq.org). Primers were designed based on the resulting soybean EST sequences to probe a six-dimentional BAC pool for LysM-containing BACs via a PCR-based approach. The probed LysM-containing BACs were verified and shotgun sequenced to either finished phase (phase 3) at the Arizona Genome Sequencing Center or prefinished phase (phase 2) at the Washington University Genome Sequencing Center. BAC sequences were annotated using the dicot species model and Arabidopsis matrix of FGENESH (www.soyseseq.com). Annotated proteins were similarly analyzed to screen for LYK proteins. Signal peptides and transmembrane domains were predicted using SignalP (http://www.cbs.dtu.dk/services/SignalP) using both nearest-neighbor and hidden Markov model (HMM) algorithms and transmembrane HMM (http://www.cbs.dtu.dk/services/TMHMM-2.0), respectively. The GeneBank accession numbers of soybean BACs are EF533702 for GMWb069B11; EF533695 for GMWb068B15; EF533696 for GMWb026B19; EF533701 for GMWb026S01; EF533697 for GMWb035S07; EF533699 for GMWb060C12; EF33700 for GMWb080D08; and EF33698 for GMWb042D24. LysM protein sequences from species spanning all kingdoms were extracted from Pfam (http://www.sanger.ac.uk/cgi-bin/Pfam/speciesdist.pl?acc=PF01476&sid=LymM&depth=all) and searched for LysM motifs at an E-value cutoff of 0.1.

### Alignment, Phylogeny, Nucleotide Substitution Rates, and Microsynteny Analyses

Sequence alignments were performed using Clustal X 1.83 (Thompson et al., 1997) with PHYLIP output format and edited in Jalview (Clamp et al., 2004). The average identities across the alignments for LysM (type X), LysMn (type Xl), and LYPb (type VI) were calculated based on the exported annotations in Jalview. An HMM profile calculated using hmmer (Eddy, 1998) for each alignment was used to realign (hmmlign) sequences at matching states (-m) to identify and remove indel regions. Parsimony trees were generated using the program protpars of PHYLIP (Felsenstein, 2000), with maximum-likelihood branch lengths calculated using TREE-PUZZLE (Schmidt et al., 2002). Distance trees were calculated using the program Prodist and Fitch of the PHYLIP package. Maximum-likelihood trees were calculated using the program proml of the PHYLIP package. Bootstrap values were calculated using the program Fboot for the PHYLIP package. Trees were visualized and rooted using A Tree Viewer (Zmasek and Eddy, 2001). For calculation of nucleotide substitution rates, codon-aligned nucleic acid sequences were created using CodonAlign 2.0 (http://www.sinauer.com/hall). All insertions and deletions were removed except that a gap of more than 30 nucleotides was preferably retained to demonstrate the lack of the p loop and the activation loop in the kinase domains of LjNFR5 orthologs (Limpens et al., 2003; Madsen et al., 2003; Arrighi et al., 2006). Nucleotide substitution levels were calculated using the program codeml of the PAML package (Yang, 1997) with a user-defined parsimony tree (Fig. 3).

To build microsynteny maps, genomic sequences surrounding each LYK gene, about 0.5 to 0.9 Mb in length, were extracted from the above databases and from soybean BAC sequences, which are about 100 to 170 kb in length. The genomic sequences were annotated using dicot species model and Arabidopsis matrix of FGENESH for the five dicot plants, monocot species model and rice matrix for rice. The annotated protein sequences were compiled together into a peptide sequence database. Repetitive sequences were excluded from the databases. BLASTp was used to compare proteins against the database with an E-value cutoff of 1e-20 and a percent identity cutoff of 35% between species and 40% within same species and legumes. BLASTp results were then filtered once to remove retroelements. The microsynteny maps were finally drawn in Adobe Illustrator 10.0.

### FISH of Soybean Homeologous BACs

Soybean seeds were germinated at 30°C for 2 to 4 d. The terminal 2 to 4 cm of four to eight root tips were treated with pressurized nitrous oxide gas (Kato, 1999) for 1 h in moistened, perforated 1.6-mL microcentrifuge tubes. Root tips were fixed in ice-cold 90% acetic acid for 10 min and briefly washed twice with 190% methanol.
water, on ice. The terminal 3 to 4 mm of the root tips were then excised and individually digested in 20 μL of a solution of 1% pectolyase Y23 (ICN) and 2% Onozuka R-10 cellulase (Yakult Pharmaceutical Company) in citrate buffer (10 mM Na3 EDTA, 10 mM sodium citrate, pH 5.5) for 45 min at 37°C. Root tips were briefly rinsed twice with 70% ethanol and coarsely macerated with a blunt dissecting needle to dissociate cells. Samples were centrifuged at 2,000g for 3 s in a microcentrifuge to collect cells; cell pellets were resuspended in 40 μL of ice-cold 100% glacial acetic acid and vortexed for 1 s. Five microliters of the cell suspension were dropped onto a glass slide and allowed to dry for 5 min in a moistened paper towel-lined box. The samples were UV cross-linked (total energy, 120 mJ cm−2) in a Stratalinker UV cross-linker (Stratagene).

BACs were labeled for FISH using optimized nick-translation reactions as previously described (Kato et al., 2004). Cross-linked slides were cooled on ice and 5 μL of probe cocktail were dropped onto the center of each slide and a plastic coverslip applied. The slides were transferred to a moistened paper towel in a covered aluminum tray that was placed in a boiling water bath for 5 min to denature the DNA of the probe cocktail and chromosome spreads. Slides were transferred to a slide holder in a humidified plastic box and incubated for 18 to 24 h at 55°C. After hybridization, slides were dipped in 2× SSC buffer to remove cover slips and then washed in 2× SSC buffer at 55°C for 5 min. Slides were then mounted in 4% paraformaldehyde containing Vectashield mounting medium (Vector Laboratories). FISH images were acquired using a 100× oil immersion objective plan apo oil lens on an epifluorescence microscope equipped with a CCD camera. Raw image files were then imported into Adobe Photoshop 7.0 (Adobe Systems); background was subtracted and scope equipped with a CCD camera. Raw image files were then imported into Adobe Photoshop 7.0 (Adobe Systems); background was subtracted and highl

Plant Growth and Quantitative RT-PCR

Soybean, M. truncatula, and rice plants were grown in the greenhouse at 28°C to 30°C with a 16-h light/8-h dark cycle. Roots and vegetative tissues were sampled about 3 weeks after planting and flowers were sampled about 3 months after planting. Total RNAs were extracted using Trizol (Invitrogen) followed by Turbo DNase (Ambion) treatment to remove genomic DNA contamination. First-strand cDNAs were synthesized using Moloney murine leukemia virus reverse transcriptase (Promega). Quantitative RT-PCR was performed using a 7500 real-time PCR system (Applied Biosystems) following standard procedures. The primer sequences are listed in Supplemental Table S3.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers EF533695 to EF533702.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Parsimony phylogeny of all LysM motifs.

Supplemental Figure S2. Alignments of plant LysM motif types X, XI, and VII.

Supplemental Figure S3. Parsimony phylogeny of LysM motifs from all kingdoms.

Supplemental Figure S4. Distance phylogeny of LysM motifs from all kingdoms.

Supplemental Figure S5. Maximum-likelihood phylogeny of LysM motifs from all kingdoms.

Supplemental Figure S6. Distance phylogeny of plant LKY proteins from six plant species.

Supplemental Figure S7. Maximum-likelihood phylogeny of plant LKY proteins from six plant species.

Supplemental Table S1. List of LKY genes from six plant species.

Supplemental Table S2. LysM motif sequences from all kingdoms.

Supplemental Table S3. Plant LKY gene primers for qRT-PCR.

ACKNOWLEDGMENTS

We thank Dr. Peter Cresshoft, Attila Kereszt, and Arief Indrasumunar for providing GmNFR1a, GmNFR1b, GmNFR5a, and GmNFR5b genomic and cDNA sequences and Dr. John Walker for providing rice seeds. We are also grateful to Dr. Fangguo Han, Dr. Scott Jackson, and Christian Hans for helping with FISH experiments. We thank Dr. Chris Pires and Dr. Sherri Sachdev for critical reading of the manuscript.

Received January 31, 2007; accepted April 6, 2007; published April 20, 2007.

LITERATURE CITED

Arrighi J, Barre A, Ben Amor B, Bersoult A, Soriano L, Mirabella R, Carvalho-Niebel E, Joumet E, Gherardi M, Huguet T, et al (2006) The Medicago truncatula lysine motif–receptor-like kinase gene family includes NFP and new nodule-expressed genes. Plant Physiol 142: 265–279

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

Ben Amor B, Shaw SL, Oldroyd GED, Mailliet F, Penuzza RV, Cook D, Long SR, Denarie J, Gough C (2003) The NFP locus of Medicago truncatula controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. Plant J 34: 495–506

Bielnicki J, Devedjiev Y, Derewenda U, Dauter Z, Joachimiak A, Derewenda ZS (2006) R. subtulis yksD protein at 2.0 Å resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. Proteins 62: 144–151

Birkeland NK (1994) Cloning, molecular characterization, and expression of the genes encoding the lytic functions of lactococcal bacteriophage phi-Lc3—a dual lysis system of modular design. Can J Microbiol 40: 658–665

Cannon SR, Sterck L, Rombaats S, Sato S, Cheung E, Gouzy J, Wang X, Mudge J, Vasdevani J, Scheis T, et al (2006) Legume genome evolution viewed through the Medicago truncatula and Lotus japonicus genomes. Proc Natl Acad Sci USA 103: 14959–14964

Clamp M, Cuff J, Searle SM, Barton GJ (2004) The Jalview Java alignment editor. Bioinformatics 20: 426–427

Eddy SR (2001) Profile hidden Markov models. Bioinformatics 14: 755–763

Felsenstein J (2000) PHYLP (Phylogeny Inference Package), Ed 3.6. University of Washington, Seattle

Frankel G, Lider O, Hershkoviz R, Moul PD, Kachalsky SG, Candy DCA, Cahalon L, Humphries MJ, Dougan G (1996) The cell-binding domain of intimin from enteropathogenic Escherichia coli binds to beta integrins. J Biol Chem 271: 20339–20346

Jorde AE, Yu J, Tall BD, Kaper JB (1990) A genetic locus of enteropathogen- genic Escherichia coli necessary for the production of attaching and effecting lesions on tissue culture cells. Proc Natl Acad Sci USA 87: 7839–7843

Jordan K, Wolf Y, Koonin EV (2004) Duplicated genes evolve slower than singletons despite the initial rate increase. BMC Evol Biol 4: 22

Jorgensen L, Englebert S, Chu CP, Kariyama R, Daneemoore L, Shockman GD, Bateman A, Bycroft M, Bielnicki J, Devedjiev Y, Derewenda U, Dauter Z, Joachimiak A, Derewenda ZS (2006) R. subtulis yksD protein at 2.0 Å resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. Proteins 62: 144–151

Kato A (1999) Air drying method using nitrous oxide for chromosome counting in maize. Biotech Histochem 74: 160–166

Kato A, Lamb JC, Burchler JA (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci USA 101: 13554–13559

Limpens E, Franken C, Smit P, Willems J, Busseling 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, Szczylowski K, Sato S, Kaneko T, Tabata S, Sandal N, et al (2003) A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature 425: 637–640

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8: 1113–1130

Mulder L, Lefebvre B, Cullimore J, Imberty A (2006) LysM domains of Medicago truncatula NFP protein involved in Nod factor perception: glycosylation state, molecular modeling and docking of chitosan oligosaccharides and Nod factors. Glycobiochem 16: 801–809

Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci USA 96: 4285–4288
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

Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV (1999) Eukaryotic signalling domain homologues in archaea and bacteria: ancient ancestry and horizontal gene transfer. J Mol Biol 289: 729–745

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

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

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREEPUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504

Shiu SH, Karlowski WM, Pan RS, Tzeng YH, Mayer KFX, Li WH (2004) Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. Plant Cell 16: 1220–1234

Stacey G, Libault M, Brechenmacher L, Wan JR, May GD (2006) Genetics and functional genomics of legume nodulation. Curr Opin Plant Biol 9: 110–121

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal-X windows interface—flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313: 1596–1604

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13: 555–556

Zhu HY, Riely BK, Burns NJ, Ane JM (2006) Tracing nonlegume orthologs of legume genes required for nodulation and arbuscular mycorrhizal symbioses. Genetics 172: 2491–2499

Zmasek CM, Eddy SR (2001) ATV: display and manipulation of annotated phylogenetic trees. Bioinformatics 17: 383–384