Bioinformatics Analysis of the LysM-Type Receptor-Like Kinase Gene Family in Physcomitrella patens (Postprint)

Authors: Gao Mei, Xin Jiankang, Jiang Shan

Date: 2020-08-03T00:00:00+00:00

Abstract

Plant LysM-type receptor-like kinases (LYKs) represent an important class of receptor-like kinases (RLKs) in plants that play indispensable roles in plant growth, development, and stress resistance, constituting a research hotspot for gene function studies. To better understand LYK genes in Physcomitrella patens, this study employed bioinformatic methods to identify and analyze members of the LysM-type receptor-like kinase gene family. By examining basic physicochemical properties, gene structure, chromosomal localization, and phylogenetic relationships of P. patens LYK family members, we preliminarily explored the relationships among LYK gene structure, evolution, and function. The results revealed that P. patens contains 21 LYK genes, with amino acid sequence lengths ranging from 625 to 755 aa, molecular weights from 69.54 to 82.02 kDa, and isoelectric points between 5.98 and 7.78. Phylogenetic analysis of all P. patens LYK proteins together with those from three model plants (Oryza sativa, Arabidopsis thaliana, and Medicago truncatula) classified all LYK proteins into four subgroups (LYK-I, LYK-II, LYR-I, and LYR-II). Members within each subgroup of P. patens exhibited similar gene structures and conserved domain features, suggesting they may possess identical or similar functions. Chromosomal localization analysis showed that the 21 LYK genes were distributed across four chromosomes, forming small gene clusters that may also be functionally relevant. This study provides fundamental information on the LysM-type receptor-like kinase gene family in P. patens, laying the groundwork for future investigations into the physiological and biochemical functions of individual LYK family members.

Bioinformatics Analysis of the Lysin Motif Receptor-Like Kinase Gene Family in *Physcomitrella patens*

GAO Mei[1], **XIN Jiankang**[1], **JIANG Shan**[2]* [1]School of Life Science, Guizhou Normal University, Guiyang 550001, China
[2]School of International Education, Guizhou Normal University, Guiyang 550001, China

Abstract

Plant lysin motif receptor-like kinases (LYKs) represent an important class of receptor-like kinases (RLKs) that play essential roles in plant growth, development, and stress responses, making them a major focus of gene function research. To better understand LYK genes in the moss *Physcomitrella patens*, we identified and analyzed members of the LysM-type receptor-like kinase gene family using bioinformatics approaches. By examining basic physicochemical properties, gene structure, chromosomal localization, and phylogenetic relationships of LYK family members, we explored the connections between gene structure, evolution, and function. Our results identified 21 LYK genes in *P. patens* with amino acid sequences ranging from 625 to 755 residues, molecular weights from 69.54 to 82.02 kDa, and isoelectric points between 5.98 and 7.78. Phylogenetic analysis of all LYK proteins from *P. patens* together with those from three model plants (rice, *Arabidopsis*, and *Medicago truncatula*) revealed four subgroups (LYK-I, LYK-II, LYR-I, and LYR-II). Members within each *P. patens* subgroup exhibited similar gene structures and conserved domain features, suggesting they may possess identical or similar functions. Chromosomal mapping revealed that the 21 LYK genes are concentrated on four chromosomes and form small gene clusters, which may also correlate with functional relationships. This analysis provides fundamental information about the LysM-type receptor-like kinase gene family in *P. patens* and establishes a foundation for future investigations into the physiological and biochemical functions of its members.

Keywords: *Physcomitrella patens*, lysin motif receptor-like kinase, gene family, bioinformatics, gene structure, chromosome location, phylogenetic relationship

---

Introduction

Plant receptor-like protein kinases constitute an important group of protein kinases in plants. These kinases share structural similarity with animal receptor protein kinases but are termed "receptor-like" because their specific extracellular signaling molecules or ligands remain largely unidentified. A typical RLK comprises three domains: an extracellular domain, a transmembrane domain, and an intracellular kinase domain, with the extracellular region containing specific structural motifs that recognize cellular signals (Tian, 2016; Lehti-Shiu et al., 2009). Among these, LysM-type receptor-like kinases (LYKs) represent a crucial subclass characterized by the presence of 1-3 LysM motifs in their extracellular domains (Afzal et al., 2008).

The number of LYK family members varies across plant species. *Arabidopsis* and rice contain 5 and 10 LysM-RLKs, respectively (Zhang, 2007; Shiu et al., 2004), and the number of LysM motifs in the extracellular domain also differs among plant LYKs. In most organisms except archaea, LysM domains are approximately 40 amino acids long. Based on differences in their kinase domains, plant LYK family members can be divided into two main types: LYK (or LysM-I) and LYR (or LysM-II), with LYK further subdivided into LYK-I and LYK-II, and LYR into LYR-I and LYR-II (Limpens et al., 2003; Arrighi et al., 2006). Most LysM-RLKs are known to sense pathogen molecules or participate in plant-microbe interactions, including the establishment of defense responses or root endosymbiosis (Wan et al., 2008; Radutoiu et al., 2003).

While the importance of LysM-type receptor-like kinases in plant defense and microbial symbiosis has been well documented, few studies have investigated LysM-RLK family genes in *P. patens*. As a model non-vascular plant representing early land plant colonization, *P. patens* has a genome size of 511 Mb with 27 chromosomes and a fully sequenced genome (Rensing et al., 2002). In this study, we employed bioinformatics tools to identify LYK gene family members genome-wide in *P. patens* and performed protein sequence analysis, domain prediction, chromosomal localization, and evolutionary analysis to provide a reference for future functional studies of this gene family.

Methods

1.1 Research Subject and Sequence Data Acquisition

We used the model plant *P. patens* as our research subject. Genome annotation sequences for *P. patens* were downloaded from Ensembl Plants (http://plants.ensembl.org/index.html) (Kersey, 2018). *Arabidopsis* genome annotations were obtained from TAIR (https://www.arabidopsis.org/) (Poole, 2007). Rice genome sequences were downloaded from the National Rice Data Center (http://www.ricedata.cn/gene/) (Fan, 2009), and *M. truncatula* genome annotations were retrieved from Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html).

1.2 Identification of *P. patens* LYK Gene Family Members

Using the five *Arabidopsis* LYK protein sequences from TAIR as reference queries, we performed homology searches against the *P. patens* genome (http://asia.ensembl.org/index.html) to identify potential LYK family members. Candidate sequences were further screened using NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to ensure the presence of both N-terminal LysM and C-terminal Pkinase domains, yielding 21 candidate LYK genes. Conserved domains were visualized using TBtools (Chen et al., 2018).

1.3 Protein Characterization and Subcellular Localization Prediction

The 21 candidate receptor-like kinase protein sequences were analyzed using the ExPASy server (http://expasy.org/) to predict molecular weight and theoretical isoelectric point. Conserved domain positions were predicted using SMART (http://smart.embi-heidelberg.de/) (Chen, 2019). Subcellular localization was predicted with the CELLO online tool (http://cello.life.nctu.edu.tw/), where asterisks indicate highest confidence scores (Song, 2016).

1.4 Conserved Domain Structure Analysis of *P. patens* LYK Proteins

Protein sequence conservation was analyzed using MEME (http://meme-suite.org/) to identify high-similarity motifs, with visualization performed in TBtools.

1.5 Gene Structure and Genomic Localization Analysis

Using genomic information retrieved for *P. patens* LYK family members and chromosomal localization files (.gff3) from Ensembl Plants (http://plants.ensembl.org/index.html), we obtained exon-intron distribution patterns and chromosomal positions. Gene structure diagrams and chromosomal localization maps were generated using TBtools.

1.6 Phylogenetic Analysis of *P. patens* LYK Gene Family

For construction of the *P. patens* LYK phylogenetic tree, protein sequences were aligned using Muscle in MEGA7.0 (Edgar, 2004). Phylogenetic analysis was performed in MEGA7.0 using maximum likelihood inference based on the LG matrix model (LG+G+I) with partial deletion and bootstrap validation (1,000 replicates) (Sagulenko et al., 2018). Poisson correction and partial deletion parameters were applied for initial subgroup classification.

To further classify *P. patens* LYK family members and investigate their evolutionary relationships with other plants, we constructed a comprehensive phylogenetic tree including all identified LysM-RLK amino acid sequences from rice (10 sequences, monocot model), *Arabidopsis* (5 sequences, dicot model), and *M. truncatula* (23 sequences, legume model) using the same methodology.

Results

2.1 Basic Information of *P. patens* LYK Gene Family Members

Using the five *Arabidopsis* LYK (LysM-RLK) family sequences from TAIR as reference queries, we performed BLAST searches with TBtools Blast Wrapper against the *P. patens* genome, retrieving 160 candidate genes. After reciprocal

BLASTp analysis in NCBI and removal of duplicates, 50 candidate LYK genes remained. Final screening using NCBI Conserved Domain Search based on conserved domains yielded 21 candidate LYK genes, visualized in TBtools [Figure 1: see original paper]. Basic information including protein length, chromosomal location, molecular weight, and theoretical isoelectric point was analyzed and compiled .

Protein characterization revealed that *P. patens* LYK proteins have molecular weights ranging from 69.54 to 82.02 kDa with relatively stable variation, and theoretical isoelectric points from 5.98 to 7.78. The encoded proteins contain 625–755 amino acid residues. Analysis using SMART and Pfam databases confirmed that all 21 LysM-type receptor-like kinases possess both PKc_{like} and LysM characteristic domains, though the number of LysM domains varies among members—a feature similar to *Arabidopsis* LYK proteins (Shiu et al., 2004). Subcellular localization prediction using CELLO indicated that all *P. patens* LYK family members are localized to the plasma membrane .

2.2 Gene Structure and Protein Conserved Domain Analysis

MEME analysis of *P. patens* LYK gene sequences identified 10 highly conserved motifs [Figure 1: see original paper]B, with distinct arrangement patterns. The 3' end shows a consistent arrangement: motif 5-motif 4-motif 7-motif 1-motif 2-motif 3, while the 5' end displays two patterns: motif 9-motif 6-motif 10-motif 8 and motif 8-motif 6-motif 9-motif 10. All family members share the stable motif 5-4-7-1-2-3 structure at the 3' end, but exhibit variation in 5' motif order. MEME-predicted motif information is summarized .

Motif lengths range from 29 to 50 amino acids, with some motifs sharing identical lengths. Conserved domain analysis using SMART and Pfam revealed that all 21 LYK proteins contain both kinase and LysM domains, though the number of LysM domains varies [Figure 2: see original paper] and [Figure 1: see original paper]C. Phylogenetic tree construction using MEGA 7.0 divided the 21 LYK family members into three major branches [Figure 1: see original paper]A, indicating divergent evolutionary paths and potentially distinct functions among branches.

Analysis of exon-intron structures showed that approximately 62% of LYK genes contain 1–3 introns, while six genes (PpLYK9–PpLYK13 and PpLYK15) each contain 13 introns. PpLYK16 possesses 14 introns, the highest number among all members, suggesting potentially diverse biological functions. No intronless members were identified; PpLYK14 is the only gene lacking non-coding regions. Integration of conserved domains and gene structures revealed that both LysM and PKc_{like} domains reside within coding regions (CDS), indicating their functional importance. Notably, PpLYK9 through PpLYK13 show significant differences in exon-intron structure compared to other family members [Figure 1: see original paper]C.

2.3 Phylogenetic Analysis of *P. patens* LYK Proteins

Multiple sequence alignment and phylogenetic tree construction revealed three distinct branches in the *P. patens* LYK gene family [Figure 1: see original paper]A, suggesting different evolutionary trajectories and potentially divergent functions among branches.

To further classify *P. patens* LYK family members and examine their evolutionary relationships with other plant LYK genes, we constructed a comprehensive phylogenetic tree incorporating all LysM-RLK amino acid sequences from rice (10), *Arabidopsis* (5), and *M. truncatula* (23) [Figure 3: see original paper]. The analysis revealed four subgroups (LYK-I, LYK-II, LYR-I, LYR-II) across the four plant species. However, *P. patens* LYK proteins were restricted to only two subfamilies: LYK-II (eight members: PpLYK9–PpLYK16) and LYR-I (13 members). Within LYK-II, *P. patens* members formed two branches with >90% homology, and similarly, LYR-I members formed two branches with >90% homology, likely representing orthologous genes.

2.4 Chromosomal Localization of *P. patens* LYK Gene Family

Chromosomal localization data from Ensembl Plants positioned the 21 *P. patens* LYK genes across four chromosomes [Figure 4: see original paper]. Chromosome 7 harbors nine genes (the highest number), while chromosomes 1, 9, and 11 contain three, five, and four genes, respectively, demonstrating relatively concentrated distribution. Small gene clusters are evident on all four chromosomes, with closely positioned genes such as PpLYK6–PpLYK8 on chromosome 1 and PpLYK17–PpLYK21 on chromosome 9, likely resulting from tandem duplication events.

Discussion

This study identified and characterized 21 LysM-RLK genes in the *P. patens* genome. These genes exhibit relatively concentrated chromosomal distribution, residing on only four chromosomes and forming five small gene clusters, likely arising from tandem duplication. Each LYK contains 1–3 LysM conserved domains and one kinase domain [Figure 2: see original paper]. MEME analysis revealed that members within the same subfamily share identical or similar domains [Figure 1: see original paper]B, suggesting functional conservation throughout the moss life cycle.

In terms of family size, *P. patens* (21 members) contains more LYK genes than *Arabidopsis* (5) and rice (10) but similar numbers to *M. truncatula* (23). The prevalence of numerous family members in higher plants is well documented (e.g., 12 in apple, 17 in *Lotus japonicus*) (Zhou et al., 2014; Lohmann et al., 2010). As an early land plant representative, *P. patens* possesses more LysM-RLK members than *Arabidopsis* and rice, suggesting potential functional redundancy. The expansion of this family in some higher plants implies the acquisition of novel functions during evolution.

Phylogenetic analysis classified *P. patens* LysM-RLK genes into two subgroups (LYK-II and LYR-I) that cluster with subsets of homologs from rice, *Arabidopsis*, and *M. truncatula*. In contrast, higher plants contain four subgroups, suggesting that *P. patens* LysM-RLK family members may have relatively specialized functions. Consistent with evolutionary patterns, exon-intron arrangements within subgroups are similar and resemble those observed in rice, *Arabidopsis*, and *M. truncatula* (Shiu et al., 2004; Arrighi et al., 2006). The convergence of conserved domain features, homology relationships, and gene structure within subgroups supports the hypothesis of functional similarity among members.

LysM-type receptor-like kinase genes play crucial roles in plant development and stress responses, representing a major research focus in plant gene function. While this family exhibits numerical expansion and functional diversity in higher plants, functional characterization in *P. patens* remains limited, with most LYK genes uncharacterized. Based on database resources, we identified 21 LYK genes in *P. patens* and conducted comprehensive bioinformatics analyses of protein characteristics, chromosomal localization, gene structure, and evolutionary relationships. This preliminary classification of the *P. patens* LYK gene family reveals genomic features of this early land plant model and provides a theoretical framework for future functional studies of LysM-type receptor-like kinases.

References

Afzal AJ, Wood AJ, Lightfoot DA, 2008. Plant Receptor-Like Serine Threonine Kinases: Roles in Signaling and Plant Defense. *Mol Plant Microbe Interaction*, 21(5):507-517.

Arrighi JF, Barre A, Amor BB, et al., 2006. The *Medicago truncatula* Lysine Motif-Receptor-Like Kinase Gene Family Includes NFP and New Nodule-Expressed Genes. *Plant Physiol*, 142(1):265-279.

Chen CJ, Chen H, He YH, et al., 2018. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. *BioRxiv* 289660.

Chen ZH, 2019. Conservation and functional differentiation of E(z) homologous genes in plants [D].

Edgar RC, 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl Acid Res*, (5):5.

Fan YL, 2009. Simple repeated sequence analysis of the genome of *Xanthomonas oryzae* pv. Oryzae [J]. *Anhui Agricultural Sciences*, No.273 (20): 9387-9390.

Kersey PJ, 2018. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucl Acid Res*, (D1):D1.

Lehti-Shiu MD, Zou C, Hanada K, et al., 2009. Evolutionary History and Stress Regulation of Plant Receptor-Like Kinase/Pelle Genes. *Plant Physiol*, 150(1):12-26.

Limpens E, Franken C, Smit P, et al., 2003. LysM Domain Receptor Kinases Regulating Rhizobial Nod Factor-Induced Infection. *Science*, 302(5645):630-633.

Lohmann GV, Shimoda Y, Nielsen MW, et al., 2010. Evolution and Regulation of the *Lotus japonicus* LysM Receptor Gene Family. *Mol Plant Microbe Interaction*, 23(4):510-521.

Poole RL, 2007. The TAIR database. *Methods Mol Biol*, 406(406):179-212.

Radutoiu S, Madsen LH, Madsen EB, et al., 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature*, 425(6958):585-592.

Rensing SA, Rombauts S, Peer YVD, et al., 2002. Moss transcriptome and beyond. *Trends Plant Sci*, 7(12):535-538.

Sagulenko P, Puller V, Neher RA, 2018. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol*, 4(1):1.

Shiu S, Karlowski WM, Pan R, Tzeng Y, Mayer KFX, Li W, 2004. Comparative analysis of the Receptor-Like kinase family in *Arabidopsis* and rice. *Plant Cell*, 16: 1220-1234.

Shiu SH, Karlowski WM, Pan R, et al., 2004. Comparative Analysis of the Receptor-Like Kinase Family in *Arabidopsis* and Rice. *Plant Cell*, 16(5):1220-1234.

Song XQ, 2016. Cloning and functional analysis of the super-accumulating SaCAD gene of *Jingtiandongtian* [D].

Tian LM, 2016. Identification, expression pattern analysis and functional study of tomato IMPα/β and LYK family genes [D].

Wan J, Zhang XC, Neece D, et al., 2008. A LysM Receptor-Like Kinase Plays a Critical Role in Chitin Signaling and Fungal Resistance in *Arabidopsis*. *Plant Cell Online*, 20(2):471-481.

Zhang XC, 2007. Molecular evolution of Lysin motif-type receptor-like kinase in plants. *Plant Physiol*, 144.

Zhou Z, Zhang CX, Zhang LY, et al., 2014. Bioinformatics and Expression Analysis of the LysM Gene Family in Apple [J]. *Chinese Journal of Agricultural Sciences*, 000(013):2602-2612.

Note: Figure translations are in progress. See original paper for figures.