Distinct evolution rate between receptor-like kinases and receptor-like proteins

Introduction

Plants are exposed constantly to the attacks from potential virulent microbes. As the sessile organisms, plants have developed innate immunity to detect and fight against the invading pathogens in order to survive.

Three components are involved in plant resistance and susceptibility to disease. Firstly, microbes have pathogen-associated molecular patterns (PAMPs) – the conserved components from pathogens, such as bacterial flagellin and fungal chitin, which plants have evolved the capacity to recognize and respond to, and this results in PAMP-triggered immunity (PTI) (Silke et al. 2002). Secondly, to achieve successful invasion, pathogens make and deliver effector proteins into plant cells to suppress PTI (Thomas et al. 2009). Thirdly, plants have employed a repertoire of resistance (R) genes that encode guarding proteins that sense and recognize specific pathogen effectors, either directly or indirectly via effector action on host cellular components. R gene-dependent recognition results in effector-triggered immunity (ETI), resulting in restored resistance (Thomas et al. 2009). To understand resistance, we need to better understand PTI, not only because of its intrinsic interest and important role in crop protection, but also because host components involved in PTI are targets of pathogen effectors.

To trigger PTI, PAMPs from microbial pathogens need to be perceived by cell surface receptors, known as pattern recognition receptors (PRRs) (Silke et al. 2002). PRRs include receptor-like kinases (RLKs) and receptor like proteins (RLPs). RLKs and RLPs both contain a single transmembrane domain, an extracellular ligand binding domain. Detection or binding of PAMPs will trigger intracellular responses transduced by the RLKs cytoplasmic Ser/Thr kinase domain which is absent in RLPs (Christiaan et al. 2012). Instead, RLPs are likely to transduce signals by recruiting RLKs partners upon ligand binding (Christiaan et al. 2012). Being the most frontiers to detect the existence of potential invaders by binding PAMPs, RLKs and RLPs are likely to be under the similar selective pressure. On the other hand, the intracellular domain of RLKs is involved in signal transduction and interactions with different partners upon different stimuli, as well as becoming the targets of effectors of pathogens, all of could render RLKs to a higher selective pressure (Mariana et al. 2005).

In this study, all the available validated RLKs and RLPs protein sequences in different plant species are collected from Plant Resistance Gene database (PRGdb). The sequences are then analyzed to determine the evolution rate of RLKs and RLPs with regards to the similarities and dissimilarities in their roles in recognizing PAMPs and downstream signaling initiation.

Methods

Sequence Alignment

All the protein sequences downloaded from PRGdb were aligned using Mafft installed in HPC-class (Iowa State University) with "auto" strategy and default setting of BLOSUM62 model and gap penalty -1.53, +0.00, +0.00.

Construction of Phylogenetic Tree

Neighbor-joining Tree Construction: Neighbor-joining tree was constructed using online FastME program available at the ATGC bioinformatics platform website. RLKs and RLPs amino acid sequence alignment file was uploaded for FastME with substitution model LG and MtREV, respectively, Gamma distribution rate across the sites and tree refinement with SPR were chosen to run the program. Random sampling using bootstrap was used to evaluate the support of clade in inferred neighbor-joining tree.

Maximum Likelihood-based Phylogenetic Tree Construction: RAxML program available on HPC-class is used to infer phylogenetic tree. Instead of choosing a particular amino acid substitution model, "Auto" is used to allow the program to automatically determine the best model based on likelihood score. Random sampling using bootstrap was used to evaluate the support of clade in inferred neighbor-joining tree.

Result

Phylogenetic analyses reveal different evolution rates between RLKs and RLPs

A total of 23 protein sequences of the receptor-like kinases (RLKs) and receptor-like proteins (RLPs) are downloaded from PRGda. Genes encoding the 23 RLKs and RLPs have been investigated with published data supporting their roles in recognizing PAMPs and therefore the participation in PTI response. Detailed information about the corresponding interactive pathogens and the related diseases is provided from PRGda website. The inferred neighbor-joining (NJ) tree by FastMe suggests distinct evolution rates between RLKs and RLPs (Figure 1). With LG model, the overall amino acid substitution rate indicated by branch length shows that RLKs have a significantly higher substitution rate than that of the RLPs, ranging from 0.004 of Serk3A in *Nicotiana benthamiana* to 2.39 of RFO in *Arabidopsis thaliana*. The branch lengths of most RLK members range from 0.94 to 2.39. In contrast, the branch lengths of most RLPs members are from 0.00 to 0.15. There are exceptions from both groups: RLKs BAK1 from *Arabidopsis thaliana* and the related SERKs from *Nicotiana benthamiana* show very low level of substitution rate, similar to that of the RLPs members; meanwhile, two proteins of RLPs - PGIP and Hp-1 have much longer branch lengths compared to the other RLPs. NJ tree built using MtREV model shows the same topology as compared to the tree inferred by LG model. Although the overall branch lengths using MtREV model are slightly longer than those using LG model: 0.002 - 3.7 for RLKs and 0.00 – 0.22 for RLPs, the distinct evolution rates for the 2 groups remain similar to that inferred using LG model (Figure 1). Phylogenetic tree using maximum likelihood method further validate this observation (Figure 3).

RLKs intracellular domain is not responsible for the higher substitution rate

The extracellular domain of RLKs and RLPs involved in PTI response contain multiple leucine rich repeat (LRRs) units for ligand perception (Shin-Han et al. 2001). Besides extracellular domain, both RLKs and RLPs contain a similar single transmembrane domain. However, the cytoplasmic kinase domain for initiating downstream signal transduction in RLKs is absent in RLPs. RLPs contain a short cytoplasmic tail without any identified functional domain. To investigate whether the intracellular kinase domain of RLKs, which might be under different selection, may contribute to the higher substitution rate in RLKs, the kinase domain and the adjacent N terminal amino acids are removed from all RLKs protein sequences. The trimmed RLKs amino acid sequences containing extracellular domain, transmembrane domain and a short intracellular tail are re-aligned with RLPs sequences to infer the NJ trees using exactly the same parameters and substitution models used in previous analysis. Surprisingly, although the topology looks different between the "new" and the "old" NJ trees, the branch length of most RLKs remain significant higher than RLPs (Figure 2), indicating that kinase domain is not responsible for the higher substitution rate among RLKs receptors. Maximum likelihood analysis supports this result (Figure 4).

Discussion

The neighbor-joining trees using protein sequences of RLKs and RLPs involved in plant PTI response reveal distinct evolution rate between RLKs and RLPs. Most RLKs protein sequences tested in this analysis have longer branch lengths. In contrast, the branch lengths of all the RLP protein sequences are significantly shorter, indicating a much lower substitution rate among RLP members. Three sequences in RLKs show extremely short branch length are BAK1 from *Arabidopsis thaliana* and Serk3A and Serk3B from *Nicotiana benthamiana*. The branch lengths of BAK1, SERK3A and SERK3B are similar to those of the RLPs. Compared to other RLKs, BAK1 has less LRR repeats and may not be involved in direct ligand perception. Growing evidence shows that BAK1 can be instantaneously recruited by other RLKs, FLS2 for example. It acts as a co-receptor of FLS2 and is required for the full activation of FLS2 triggered immunity (Antje et al. 2007, Yadong et al. 2013). Notably, the closely related SERK proteins have gain increasing support to be associated with RLKs and regulate their function, suggesting their role as co-receptors upon ligand binding. RLPs are capable of binding PAMPs directly, however, they also act as co-receptors in a way that additional RLKs are required to transduce the signals upon ligand perception. The role of RLPs, RLK BAK1 and SERKs acting as co-receptors may somehow distinguish them from other RLKs with different substitution rate. NJ trees, constructed using RLKs partial amino acid sequences by removing the intracellular kinase domain sequences, indicate the kinase domain alone doesn't contribute to the relatively fast evolution rate of RLKs. The distinct substitution rate may be a result of dominant and ancillary roles in PTI initiation and signal transduction between different receptor-like kinases and receptor-like proteins. There are many factors capable of exerting effects on protein evolution including protein functions, family size and function redundancy among members of the protein families. Further investigations need to be done, and meanwhile, more RLKs and RLPs to be identified and included for further analysis may help address this question.

Reference

1 Antje Heese, Dagmar R. Hann, Selena Gimenez-lbanez, Alexandra M.E. Jones, Kai He, Jia Li, Julian I. Schroeder, Scott C. Peck and John P. rathjen. 2007. The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proc.Natl.Acad.Sci.* 104:12217-12222

2 Christiaan Greeff, Milena Roux. John Mundy and Morten Petersen. Receptor-like kinase complexes in plant innate immunity. 2012. *Frontiers in Plant Science* 3: 209-216

3 Mariana Mondragon-Palomino, Brandon S. Gaut. 2005. Gene conversion and the evolution of three leucine-rich repeat gene families in Arabidopsis thaliana. *Molecular Biology and Evolution* 22: 2444-2456

4 Shin-Han Shiu and Anthony B. Bleecker. 2001. Plant receptor-like kinase gene family: diversity, function and signaling. *Science's STKE* 113: 22-28

5 Silke Stracke, Catherine Kistner, Satoko Yoshida, Lonneke Mulder, Shusei Sato, Takakazu Kaneko, Satoshi Tabata, Niels Sandal, Jens Stougaard, Krzysztof Szczyglowski and Martin Parniske. 2002. A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* 417: 959-962

6 Thomas Boller and Sheng Yang He. 2009. Innate immunity in plants: an arm race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science* 324: 742-744

7 Yadong Sun, Lei Li, Alberto P. Macho, Zhifu Han, Zehan Hu, Cyril Zipfel, Jian-Min Zhou, Jijie Chai. 2013. Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex.

Figures

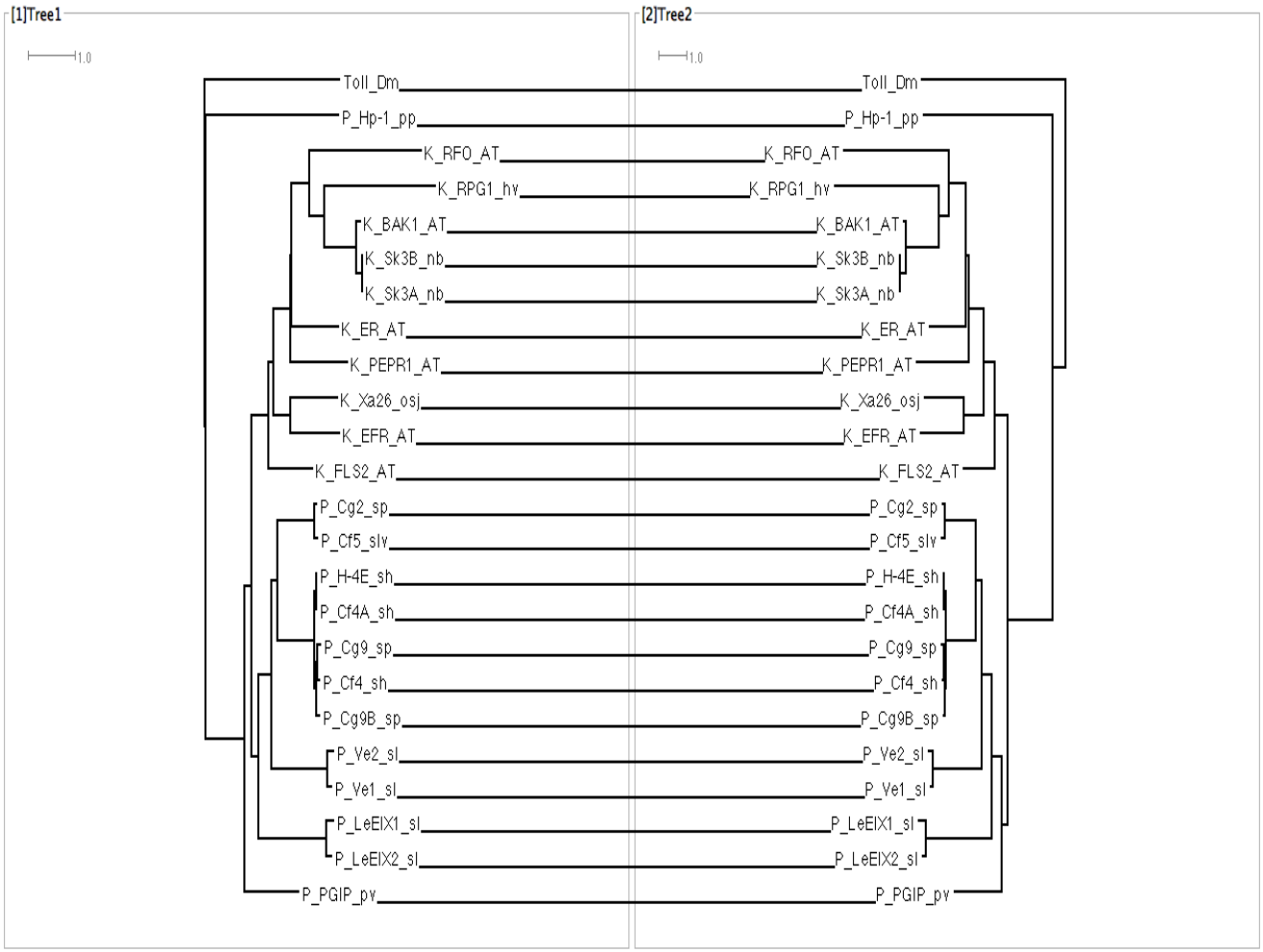

Figure 1. Neighbor-joining trees using full length protein sequences of RLKs and RLPs with model LG and MtREV, respectively. Tree viewer software Dendroscope was used to display NJ trees under 2 substitution models side by side. Tree 1 on the left side was built using LG model and Tree 2 on the right side was built using MtREV model.

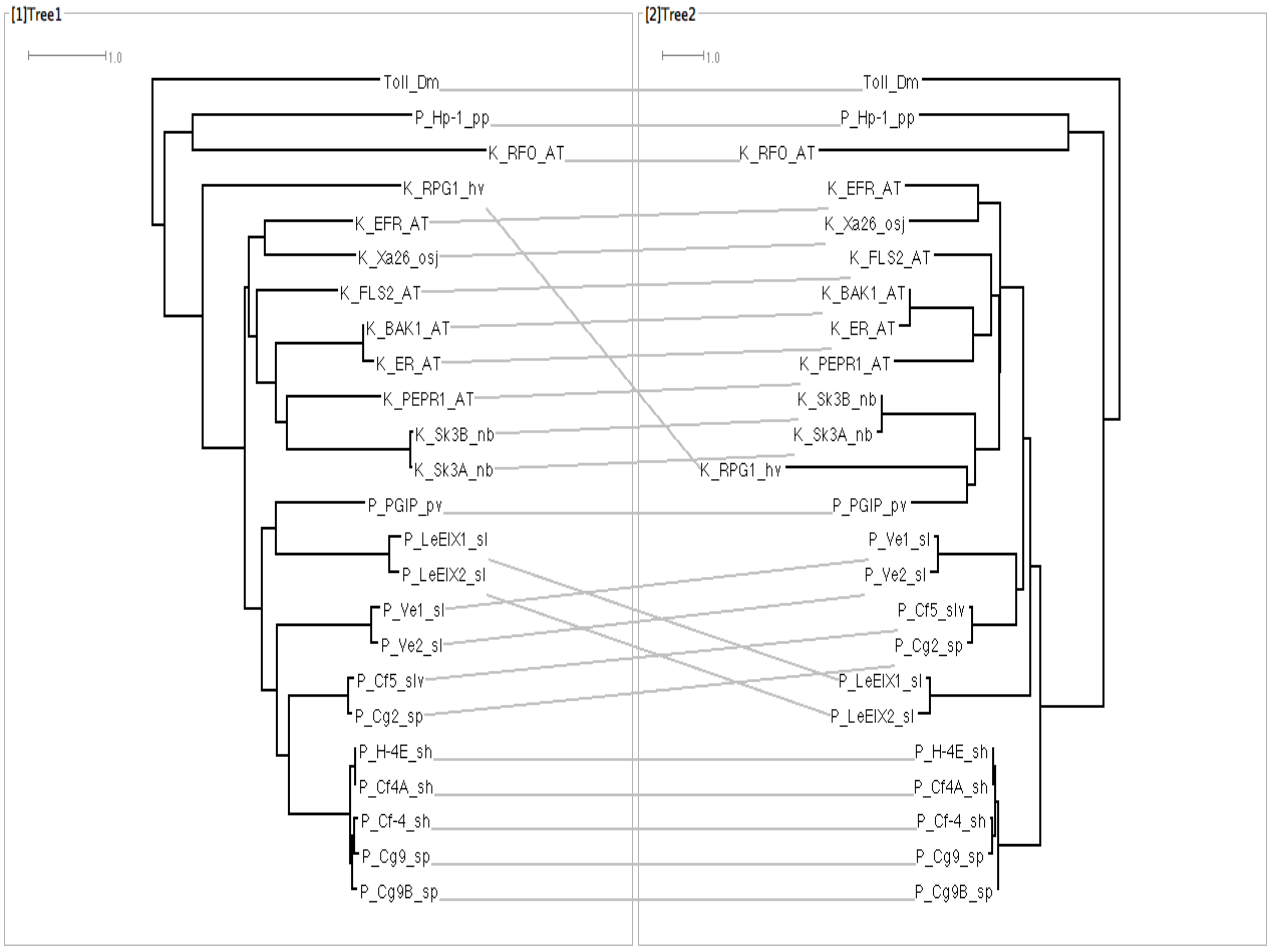

Figure 2. Neighbor-joining trees using truncated RLK protein sequences and full length RLP protein sequences with substitution model LG and MtREV. RLK protein sequences with extracellular and transmembrane domains, but without intracellular kinase domain were used to construct phylogenetic tree with RLP members with full length protein sequences. Tree viewer software Dendroscope was used to display NJ trees under 2 substitution models side by side. Tree 1 on the left side was built using LG model and Tree 2 on the right side was built using MtREV model.

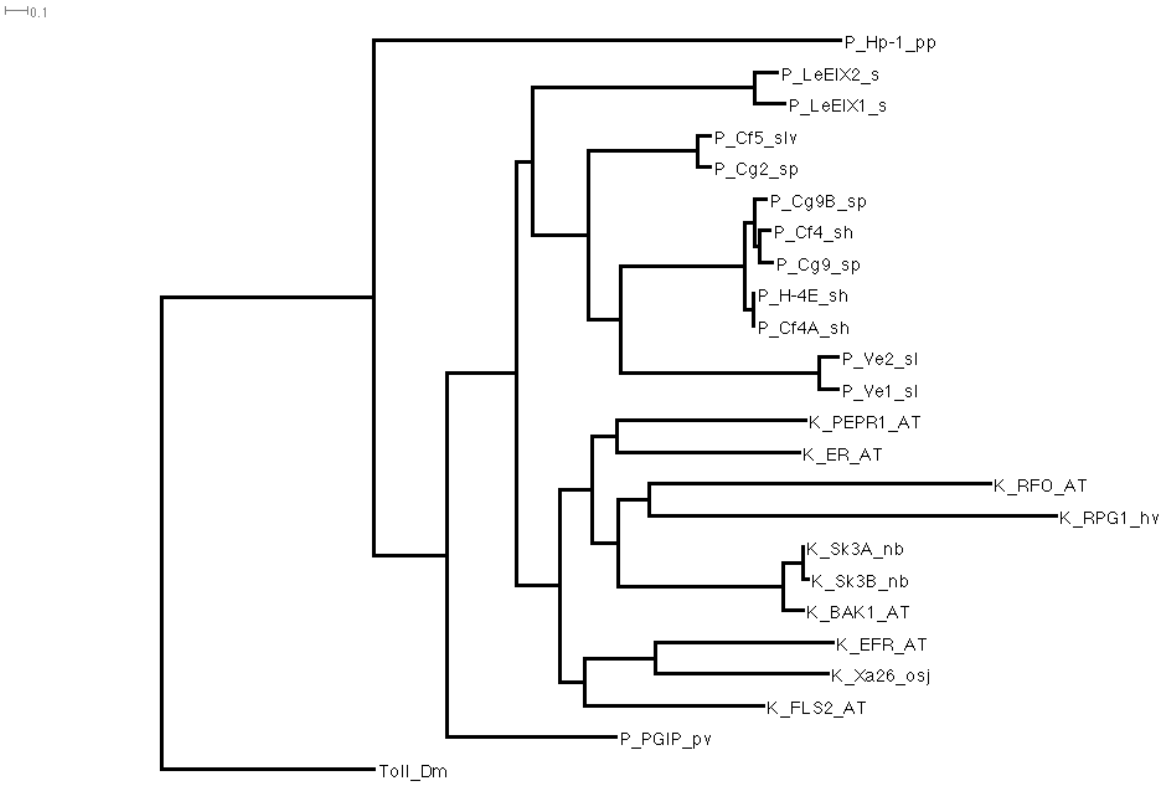

Figure 3. Phylogenetic tree for RLKs and RLPs based on maximum likelihood method. The tree was constructed using RAxML program. Full length of protein sequences of RLKs and RLPs were used for tree construction.

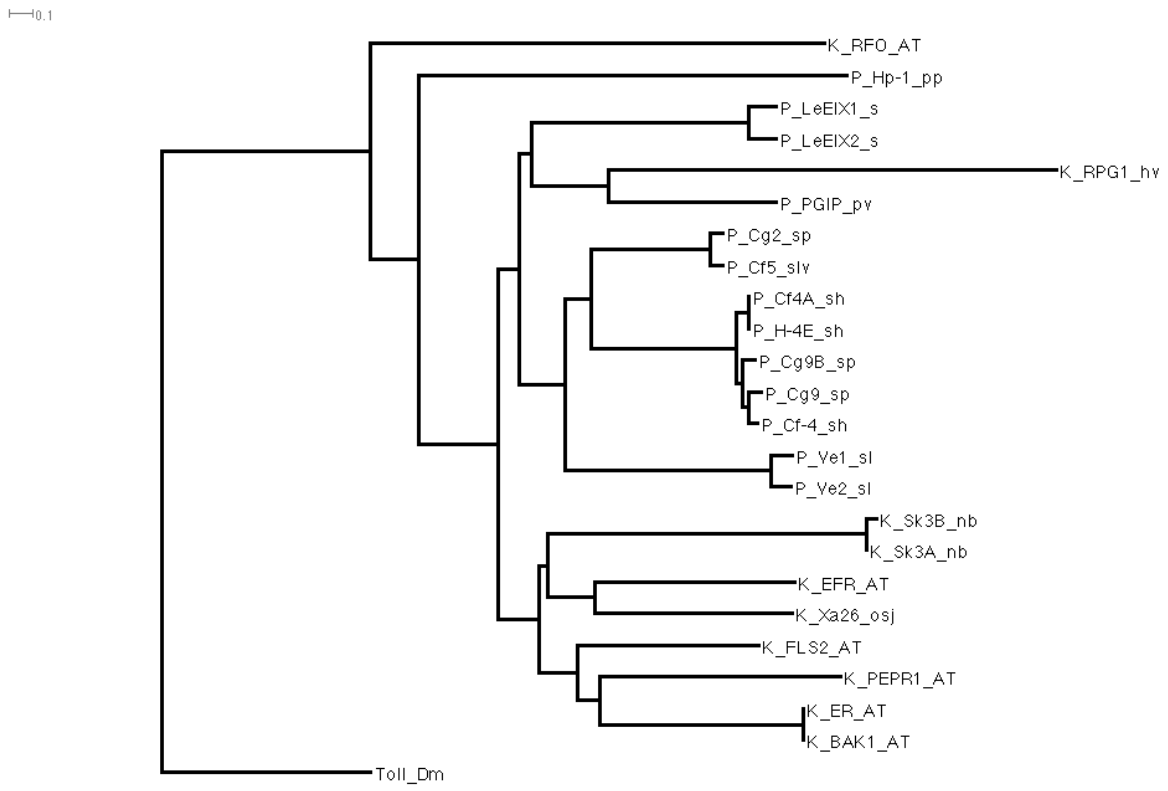

Figure 4. Phylogenetic tree for RLKs and RLPs based on maximum likelihood method. The tree was constructed using RAxML program. The truncated RLK protein sequences with only extracellular and transmembrane domains and full length RLP protein sequences were used to construct the tree.