

# Co-Variation Among Major Classes of LRR-Encoding Genes in Two Pairs of Plant Species

Jiao Wang · Shengjun Tan · Li Zhang ·  
Ping Li · Dacheng Tian

Received: 10 February 2011 / Accepted: 10 May 2011 / Published online: 31 May 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** NBS-LRR (nucleotide-binding site-leucine-rich repeat), LRR-RLK (LRR-receptor-like kinase), and LRR-only are the three major LRR-encoding genes. Owing to the crucial role played by them in plant resistance, development, and growth, extensive studies have been performed on the NBS-LRR and LRR-RLK genes. However, few studies have focused on these genes collectively; they may co-vary as all of them contain LRR motifs. To investigate their common evolutionary patterns, all major classes of LRR-encoding genes were identified in 12 plant species, and particularly compared in two pairs of close relatives, *Arabidopsis thaliana*–*A. lyrata* (*At*–*Al*) and *Zea mays*–*Sorghum bicolor*. Our results showed that these genes co-vary significantly in terms of their numbers between species and that the genes with certain evolutionary parameters are most likely to have similar functions. The development-related genes have clear orthologous relationships between closely related species, as well as lower nucleotide divergence, and *Ka/Ks* ratio. In contrast, resistance-related genes have exactly opposite characteristics and favor 11–15 LRRs per gene. This association could be very useful in predicting the function of

LRR-encoding genes. The presence of co-variation suggests that LRRs, combined with other domains, can work better in some common functions. In order to cooperate efficiently, there should be balanced gene numbers among the different gene classes.

**Keywords** LRR-encoding gene · Co-variation · Resistance-related genes · Closely related species · Leucine-rich repeat

## Introduction

The leucine-rich repeat (LRR) domain is a common motif, which broadly exists in eukaryotes, prokaryotes, and viruses. Structural analysis of the LRR domain shows that this motif comprises 2–47 LRRs, with a length of between 20 and 30 amino acids in each repeat (McHale et al. 2006). Dozens of proteins with LRRs have resolved crystal structures. Although differences exist in the LRR structures, all share certain common features. Each LRR unit consists of one  $\beta$ -sheet and one  $\alpha$ -helix. The  $\beta$ -sheet and  $\alpha$ -helix are arranged parallel to the axis, which results in a horseshoe-shaped molecule with the  $\beta$ -sheet lining the concave surface of the horseshoe, and the  $\alpha$ -helix flanking its outer surface. Each  $\beta$ -sheet has a core consensus of  $L \times \times L \times L \times \times N/C$  (where L is leucine and  $\times$  is any amino acid). Leucine forms the conserved bone of each repeat, whereas the  $\times$  shows the variable feature of the putative solvent-exposed surface (Kajava 1998).

As an active motif, LRR usually combines with other motifs. The leucine-rich repeat receptor-like kinase (LRR-RLK) gene class contains an extracellular LRR domain and a cytoplasmic kinase domain. Another class contains an extracellular LRR domain exclusively (LRR-only),

---

Jiao Wang and Shengjun Tan contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-011-9448-1) contains supplementary material, which is available to authorized users.

---

J. Wang · S. Tan · L. Zhang · D. Tian (✉)  
State Key Laboratory of Pharmaceutical Biotechnology, School  
of Life Sciences, Nanjing University, Nanjing 210093, China  
e-mail: dtian@nju.edu.cn

P. Li (✉)  
Rice Research Institute, Sichuan Agricultural University,  
Wenjiang 611130, China  
e-mail: liping6575@163.com

including receptor-like proteins (RLPs), polygalacturonase-inhibiting protein (PGIP, the cell wall LRR), and others. Apart from the two extracellular LRR classes, nucleotide-binding site-leucine-rich repeat (NBS-LRR) exists within the cytoplasm. In addition, there are other LRR gene classes, e.g., 29 F-box LRR and eight LRX (LRR/EXTENSIN) identified in *Arabidopsis* (Baumberger et al. 2003; Kuroda et al. 2002). Of these LRR-encoding classes, the LRR-RLK, LRR-only, and NBS-LRR classes are the major ones in plants, because a large number of genes have been identified in each of the three classes in the genome. For example, 149 and 480 NBS-LRR genes were identified in *Arabidopsis thaliana* (At) and rice, respectively (Meyers et al. 2003; Zhou et al. 2004). Meanwhile, there were 217 LRR-RLKs in *A. thaliana* (Tang et al. 2010). In addition, 59 RLPs which belong to the LRR-only class were identified in *A. thaliana* (Fritz-Laylin et al. 2005).

The slender, arc-shaped structure of LRR provides a higher surface-to-volume ratio than the globular architecture, which makes it ideally suited to physical interaction (Padmanabhan et al. 2009). Furthermore, the repetitive structure of LRR makes it capable of the rapid generation of new variants by duplications and deletions of entire repeats (Ellis et al. 2000a). Indeed, LRR-encoding genes are widely involved in diverse pathways: *BRI1* and *RPK1* are LRR-RLK genes involved in hormone signaling (Li and Chory 1997; Osakabe et al. 2005); *TMM* and *CLAVATA2* are LRR-only genes involved in stomatal patterning and meristem development, respectively (Nadeau and Sack 2002; Yang and Sack 1995), *BAMI* (LRR-RLK) functions in the meristem, seed size, and nodule number differentiation; *ERECTA* (LRR-RLK) is involved in organ initiation and elongation (DeYoung et al. 2006; Torii et al. 1996); and the NBS-LRR protein, *CASI*, is involved in photomorphogenesis regulation (Faigon-Soverna et al. 2006).

Of the known functions of LRR-encoding genes, disease resistance may be uniquely important. LRR-encoding genes are widely involved in the two layers of the defense response in plants (Jones and Dangl 2006). In the first layer, plants use pattern-recognition receptors (PRRs) to perceive pathogen-associated molecular pattern (PAMP) which leads to PAMP-triggered immunity (PTI). Several LRR-RLK genes have been detected as PRRs; for example, *FLS2* and *EFR* recognize flagellin and the EF-Tu elongation factor, respectively, which are the conserved proteins from pathogens (Gomez-Gomez and Boller 2000; Kunze et al. 2004). Once pathogens escape from the first layer of defense, plants develop resistance genes (*R*-genes) that interact with pathogen effectors in a gene-for-gene manner and cause effector-triggered immunity (ETI). Most *R*-genes belong to the NBS-LRR gene class (Meyers et al. 2005) but some are also LRR-RLK genes, e.g., *Xa21* (Song et al. 1995), and the LRR-only class, such as *cf-9* (Jones et al. 1994).

Because of the irreplaceable role that LRR-encoding genes play in plants, there are many studies on these genes, particularly on the NBS-LRR and LRR-RLK classes (Meyers et al. 2003; Shiu et al. 2004; Tang et al. 2010; Yang et al. 2006). However, few studies have focused on these genes collectively. All the genes that contain the LRR motif may co-vary or may have some common evolutionary characteristics with respect to certain biological functions. In order to investigate the presence of co-variation, we identified all the three major classes of LRR-encoding genes from 12 plant genomes, including two pairs of close plant relatives, *At* and *Al*, *Zea mays L. ssp. mays* (maize), and *Sorghum bicolor* (sorghum). These two pairs of species only diverged at 5 and 11.9 million years ago, respectively (Koch et al. 2000; Swigonova et al. 2004). The comparisons between the closely related species could provide genetic measurements to detect the evolutionary patterns of LRR-encoding genes. Our comparative results showed that these genes indeed co-vary significantly and that genes with analogous functions have similar evolutionary patterns in different LRR gene classes. These findings could help to predict the functions of LRR-encoding genes.

## Materials and Methods

### Identification of LRR-Encoding Gene Classes

We used 12 complete plant genome sequences to identify LRR-encoding genes. Their sequences and corresponding annotations were downloaded from the online databases (Table S1). The annotation methods and quality could affect the gene identification. To estimate the bias, we used the ab initio gene finder FGENESH (<http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind>) to identify non-transposable-element-related genes in the repeat-masked sequences in *Sorghum bicolor*, *Medicago truncatula*, and *Brachypodium distachyon*, and those encoding proteins of less than 50 amino acids were omitted from the final set. A total of 35312 (35899), 51886 (53425), and 31807 (32255) non-transposable-element protein-coding sequences were predicted in *S. bicolor*, *M. truncatula*, and *B. distachyon*, respectively, with consistent gene numbers used in this article (numbers in the parenthesis).

We used the following steps to identify LRR-only genes in 12 genomes. First, both the amino acid consensus sequences of LRR\_1 domain (Pfam: PF00560) and C3 domain of cf-9 (Fritz-Laylin et al. 2005) were used as queries to conduct BLASTP searches in genome proteins with E value cutoff of 1.0. Second, the proteins which match LRR\_1 in PFAM library (<http://pfam.janelia.org/>) with an E value cutoff of 10 or LRR domain in SMART

(<http://smart.embl-heidelberg.de/>) were chosen as candidates. Third, these proteins were scored against PFAM HMMs, and the protein sequences with E values less than 0.5 to any non-LRR domains were removed. Then, the nucleotide sequences of candidate LRR-only sequences were used as queries to conduct BLASTN searches in genomes. All the new BLAST hits in the genomes, together with flank regions of 5,000–10,000 bp at both sides, were annotated using the gene-finding programs FGENESH to obtain information on complete open reading frames (ORFs). PFAM and SMART were used as before to confirm the new LRR-only candidates. To exclude potentially redundant candidate LRR-only genes, all the sequences were orientated by BLASTN, and sequences located in the same location were eliminated. This step was crucial to obtain the maximum number of candidate genes.

The NBS-LRR genes that have been identified previously in four genomes (*A. thaliana*, *A. lyrata*, maize, and Sorghum) (Chen et al. 2010; Li et al. 2010; Meyers et al. 2003) are used in this study. Meanwhile, LRR-RLK genes also have been revealed in *A. thaliana*, maize, and sorghum (Li et al. 2010). We identified the LRR-RLK and NBS-LRR genes for the other species by the same method as used in a previous study (Li et al. 2010; Tang et al. 2010). In addition, 51 housekeeping genes were randomly chosen as reference genes, based on the enzymes encoded by housekeeping genes in *At* (Scheideler et al. 2002).

To identify genes from whole-genome duplication (WGD) or segment duplication, whole genome annotation data of 10 genomes were used to do pairwise BLASTP ( $E < 1e-30$ ) within each genome to retrieve potential anchors. *Ricinus communis* and *Carica papaya* were not included for such analyses because their sequences have not been assembled into chromosomes. Then, we applied i-ADHoRe (Simillion et al. 2008) with suggested parameters to identify co-linearity blocks. Syntenic homologs (from WGDs or segment duplication) were calculated by finding segments on which there were at least five pairs of aligning genes. By these criteria, numbers of duplicates retention after WGD were identified consistent with previous studies (Schmutz et al. 2010; Schnable et al. 2009).

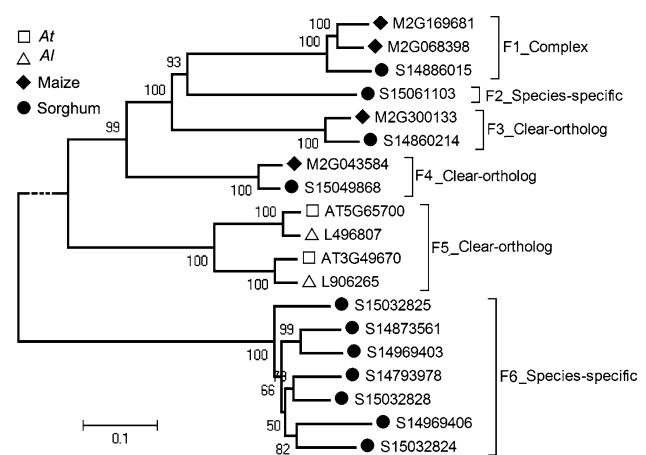
### Sequence Alignment and Data Analysis

Multiple alignments of amino acid sequences were performed by ClustalW with the default options (Thompson et al. 1994) and by MEGA Version 4.0 (Tamura et al. 2007) for manual corrections of alignment. The amino acid alignments were then used to guide the alignments of nucleotide coding sequences (CDSs). Phylogenetic trees were constructed based on the bootstrap neighbor-joining (NJ) method with a Kimura 2-parameter model by MEGA

v4.0. The stability of internal nodes was assessed by bootstrap analysis with 1,000 replicates.

A large-scale survey showed the nucleotide identity among *Rpl* homologs within or between species in Poaceae, an NBS-LRR locus, varied from 69 and 97% (Luo et al. 2011). Based on the range of nucleotide identity, the genes sharing at least 60% nucleotide identity and clustering in the same branch of phylogenetic tree with bootstrap value  $>70\%$  were defined as a family. Three types of families were identified based on the phylogenetic tree (Fig. 1 and Fig. S1). If all the genes in a family had one-to-one relationship between two close relatives, e.g., *At* versus *Al* or maize versus sorghum, the family was classified into clear-ortholog families. The second type was the species-specific family, in which the homolog gene(s) was(were) present only in one genome but absent in the other. The remaining families were categorized as complex families.

Nucleotide divergence was estimated using *Dxy* with the Jukes and Cantor correction using MEGA v4.0. The  $\times \times L \times L \times \times$  motif includes the solvent-exposed residues ( $L = \text{Leu or other aliphatic amino acid}$ ;  $\times = \text{any amino acid}$ ). To detect positive selection, we estimated the ratios of non-synonymous (*Ka*)-to-synonymous (*Ks*) substitutions of all CDSs and the  $\times \times L \times L \times \times$  motif between orthologs. *Ka* and *Ks* were calculated using MEGA v4.0 based on the Nei and Gojobori method (Nei and Gojobori 1986). All the estimates were calculated in a pairwise manner from orthologous comparisons, which included genes from two closely related species. In addition, each gene was paired only once with the least divergent counterpart in the other genome. The species-specific genes were excluded for calculation because there was no gene pair between the related species. Meanwhile, we also performed maximum likelihood (ML) analyses of *Ka*, *Ks*, and *Ka/Ks* in pairwise



**Fig. 1** Representative clades of the phylogenetic tree of LRR-encoding genes. The family names and types are denoted behind the square brackets, e.g., F1\_Complex represents family 1, a complex family

comparisons of gene family sequences. The branch models of Codeml from PAML were employed to analyze these parameters (Yang 2007).

### Identification of Numbers of Leucine-Rich Repeats

To define the precise start and end points of LRRs, we used all the predicted protein sequences encoded 5' to the kinase domain for LRR-RLK, 3' to the NBS domain for NBS-LRR, and all protein sequences for LRR-only as LRR regions to search for LRRs. Initially, MEME (Multiple Expectation Maximization for Motif Elicitation) (Bailey and Elkan 1995) was used for identifying conserved motif structure in the LRR regions (Meyers et al. 2003). This analysis was performed for the LRR regions of each LRR-encoding class. Prior analyses of LRR domain structure indicate that the regions typically have LRRs with a length of 20–30 amino acids. Multiple MEME analyses were performed with settings designed to identify three, five or ten motifs. Finally, the setting of three motifs with their sizes varying between 20 and 30 amino acids was chosen as the best parameters to recognize maximal non-overlapping motifs and filtered out most of non-LRR motif hits. However, the structure of LRRs varied greatly and MEME could not detect some individual LRRs. Therefore, manual proofreading was necessary. We manually used LRR features, such as the consensus sequences and repeated patterns, to identify the maximum number of LRR repeats (Meyers et al. 2003; Zhou et al. 2004). All the LRR repeats sequences were provided in supplement data online.

## Results

### Number Variation in Three Classes of LRR-Encoding Genes

To investigate the variation of LRR-encoding genes, three classes of them, LRR-only, LRR-RLK, and NBS-LRR, were identified in four plant species (Table S2). There were 122, 147, 131, and 195 LRR-only genes in *At*, *Al*, maize, and sorghum, respectively. The numbers of LRR-RLK genes were 228, 233, 235, and 253, and there were 147, 156, 95, and 184 NBS-LRR genes in these species, respectively (Chen et al. 2010; Li et al. 2010). With respect to the total numbers of protein-coding genes in these four genomes (30,925 in *At*, 32,670 in *Al*, 31,832 in maize, and 35,899 in sorghum), a nearly constant proportion (approximately 0.72%) was observed for LRR-RLK genes, whereas the proportion varied in LRR-only (0.39–0.54), and NBS-LRR genes (0.30–0.51; Table S2).

Given that the three LRR-encoding classes have related functions (Padmanabhan et al. 2009), we hypothesized that

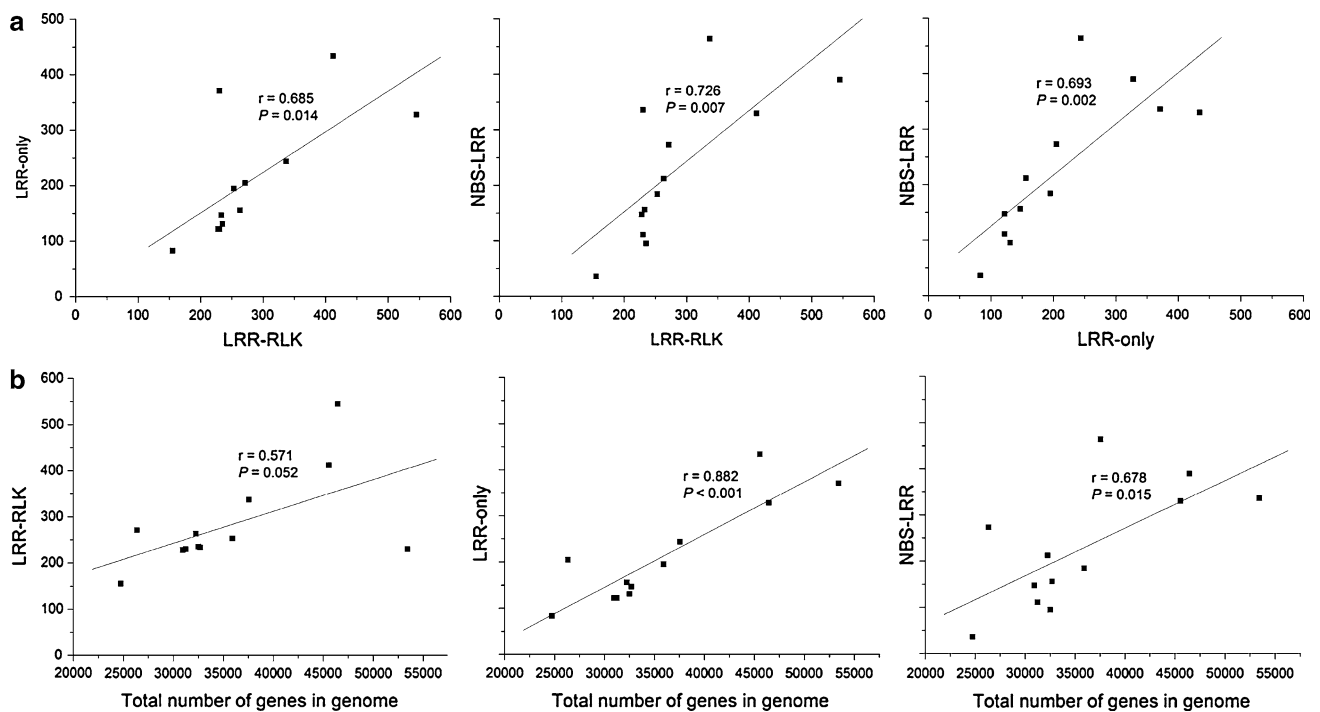
the sizes of these gene classes are supposed to co-vary among species because of similar selection pressures. To test this hypothesis, we identified the three classes of LRR-encoding genes in also eight plants, and a total of 12 plants were used in number correlation analysis (Table S1). A significant correlation ( $P < 0.05$ ) was observed between gene numbers from any of the two LRR-encoding classes in 12 species (Fig. 2a). The co-variation among species suggests that all classes of LRR-encoding genes either share a similar function or are shaped by the common forces of natural selection, most likely by pathogens, because the resistance function has been confirmed in selected genes from each of these three classes.

Interestingly, significant or nearly significant correlations are also present between each of LRR-encoding gene classes and the total number of genes in its genome ( $r = 0.571$ – $0.882$ ;  $P < 0.052$ ; Fig. 2b). The LRR-only genes had the highest correlation with total gene numbers ( $r = 0.882$ ;  $P < 0.001$ ). In contrast, only nearly significant correlations were observed between LRR-RLK and the total number of genes ( $r = 0.571$ ;  $P = 0.052$ ). This co-variation between the LRR-encoding and total number of genes in the genome indicates that these genes could expand neutrally with genome size and, therefore, contribute to the variation of the total gene numbers in a genome. In this scenario, the proportion of duplicates following WGD or segment duplication should be similar in each of LRR-encoding genes. We used i-ADHoRe to identify homologous blocks and counted such duplicates in three LRR-encoding gene classes and found that the proportions are significantly different ( $t$ -test,  $P < 0.01$ ) (Table S3). This result indicates that natural selection may also contribute to such co-variation.

### Identification of Gene Families Between Closely Related Species

The co-variation among LRR-encoding genes provides a common basis to distinguish their evolutionary similarities and differences, which could be described by evolutionary, genetic, and/or genomic parameters. To calculate such parameters, we identified clear-ortholog, complex, and species-specific families between two closely related species. There were 236 clear-ortholog families, 74 complex families, and 79 species-specific gene families in LRR-RLK gene class. The numbers of three types of families in LRR-only gene class were 79, 36, and 208, respectively. Meanwhile, NBS-LRR genes were classified into 58 clear-ortholog, 28 complex, and 205 species-specific gene families (Table S4).

Genes within three types of families were named as clear-orthologs, complex genes, and species-specific genes, respectively. The types of genes may be associated with the respective functions. For example, the clear-orthologs



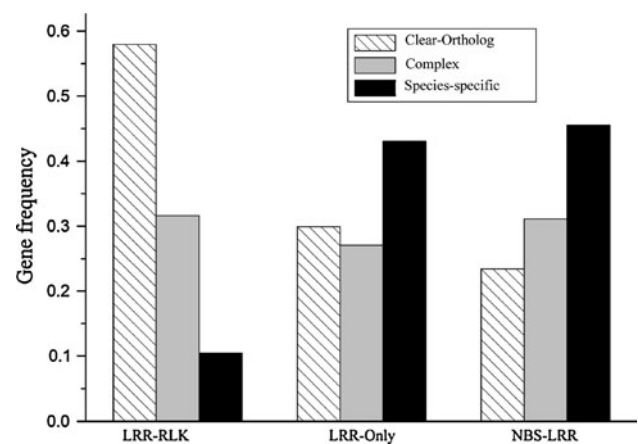
**Fig. 2** **a** Correlations of gene numbers between each two LRR-encoding classes in 12 sequenced plants, and **b** Correlations between gene numbers of each LRR-encoding class and the total number of

genes in 12 genomes. Both the correlation coefficients and *P* values are shown. The plant genomes used are listed in Table S1

might play some basic biotic functions maintained by both relative species. In contrast, the complex and species-specific genes may be more important in their response to variation of biotic factors, such as the variation of the spectrum of pathogens. In this scenario, NBS-LRR genes are expected to have a higher proportion of complex or species-specific genes because of their dominant role in disease resistance (DeYoung and Innes 2006). On the other hand, LRR-RLK genes are supposed to have a higher proportion of clear-orthologs, because many of the development-related genes were identified in this class (Morillo and Tax 2006). Indeed, the proportion of clear-orthologs was the highest (58.0%) and that of species-specific genes was the lowest (10.4%) in LRR-RLK class. On the contrary, in NBS-LRR class, the proportions of these two types families are less than half (23.4%) and more than four times (45.5%) those in the LRR-RLK class (Fig. 3; Table S4), respectively. In the LRR-only class, the proportions were intermediate.

#### Genetic Variation in Three LRR-Encoding Gene Classes

Given that the patterns of genetic variation in the genes with basic biotic function are different from those that respond to a variety of pathogens, there should be some



**Fig. 3** The proportions of three types of genes in LRR-encoding gene classes

differences in terms of genetic variation in the three classes of genes, as they contain different proportions of genes with basic biotic function. To distinguish their evolutionary patterns, we calculated the nucleotide divergence ( $D_{xy}$ ), synonymous ( $K_s$ ), and non-synonymous ( $K_a$ ) genes in each family of LRR-encoding genes. All the results were calculated in a pairwise manner from orthologous comparisons (Table 1), which included genes from two closely related species. In addition, only the least divergent

**Table 1** Genetic parameters of LRR-encoding genes within gene families

Classes	Family types <sup>a</sup>	Gene pairs <sup>b</sup>	<i>Dxy</i> <sup>c</sup>	<i>Ks</i> <sup>c</sup>	<i>Ka</i> <sup>c</sup>	<i>Ka/Ks</i> (CDS) <sup>d</sup>	<i>Ka/Ks</i> (LRR) <sup>d</sup>
TL							
LRR-RLK	Clear-ortholog	172	0.055	0.161	0.025	0.157	0.162
	Complex	37	0.073	0.171	0.045	0.266	0.321
	Total/Average	209	0.059	0.165	0.029	0.176	0.193
LRR-Only	Clear-ortholog	51	0.068	0.186	0.038	0.205	0.124
	Complex	29	0.102	0.216	0.070	0.343	0.469
	Total/Average	80	0.082	0.198	0.051	0.261	0.267
NBS-LRR	Clear-ortholog	44	0.106	0.210	0.079	0.351	0.536
	Complex	56	0.161	0.279	0.119	0.439	1.431
	Total/Average	100	0.137	0.249	0.101	0.400	1.029
MS							
LRR-RLK	Clear-ortholog	103	0.098	0.225	0.060	0.275	0.311
	Complex	64	0.094	0.222	0.057	0.256	0.332
	Total/Average	167	0.097	0.224	0.059	0.267	0.319
LRR-Only	Clear-ortholog	38	0.100	0.222	0.064	0.299	0.419
	Complex	22	0.130	0.273	0.088	0.304	0.272
	Total/Average	60	0.111	0.241	0.073	0.301	0.370
NBS-LRR	Clear-ortholog	19	0.100	0.269	0.061	0.284	0.257
	Complex	13	0.143	0.290	0.107	0.400	0.756
	Total/Average	32	0.117	0.277	0.079	0.331	0.460

TL, *A. thaliana* versus *A. lyrata*; MS, Maize versus sorghum

<sup>a</sup> Two types of families were included: clear-ortholog and complex families

<sup>b</sup> The number of orthologous gene pairs used in the calculation of genetic parameters

<sup>c</sup> Genetic parameters were calculated by pairwise comparison of genes from two close relatives for a gene family, e.g., *At* versus *Al* or maize versus sorghum. The species-specific gene families were excluded for calculation because there was no homolog from the other related species

<sup>d</sup> The ratios of non-synonymous (*Ka*) to synonymous (*Ks*) substitutions of all CDSs and the  $\times\times L\times L\times\times$  motif, respectively

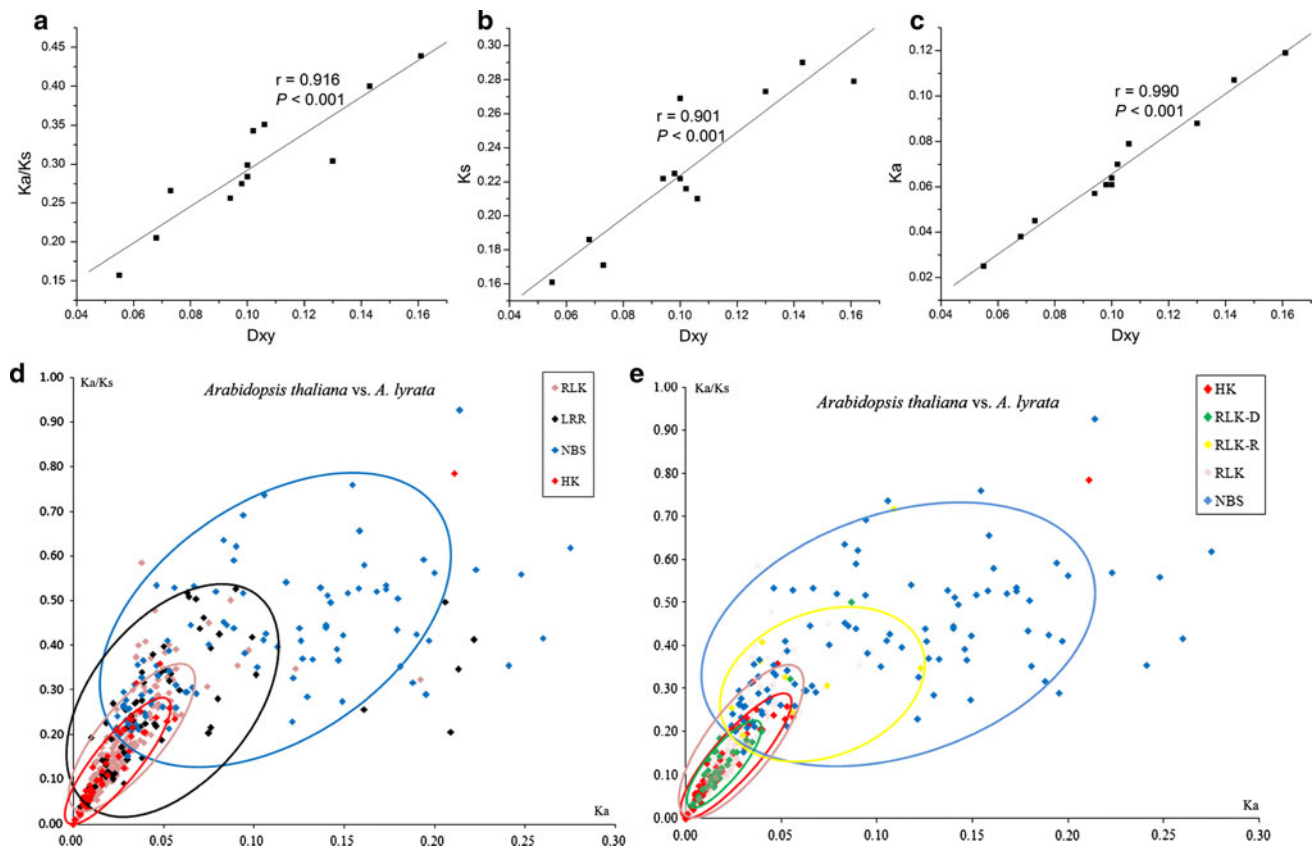
ortholog pairs were used for such calculation to give a conserved estimate for the complex families. The species-specific genes were excluded for calculation because there was no gene pair between the related species.

A significantly larger average nucleotide divergence was revealed in the NBS-LRR class than in the LRR-only (one-tailed *t*-test,  $P < 0.001$ ), and in the LRR-only class compared to the LRR-RLK gene class (one-tailed *t*-test,  $P < 0.01$ ) in the *At*–*Al* comparison. This pattern was also true in the maize–sorghum pair, although it was not significant. As the divergence time was constant between *At* and *Al* as well as between maize and sorghum, the *Dxy* values suggest that NBS-LRR class had the fastest rate of evolution, the LRR-only was intermediate and the LRR-RLK class was the slowest. When comparing this parameter between clear-ortholog and complex families, we found that complex families always have higher values than those of clear-ortholog families. This is consistent with the expectation that the function of clear-orthologs is more conserved than that of complex genes.

The *Ka/Ks* substitution rate was calculated to detect positive selection. Under the hypothesis of neutrality, *Ks*

should be equal to *Ka*.  $Ka > Ks$  indicates positive selection. In contrast, a *Ka/Ks* ratio of less than one provides evidence of negative selection. A significant linear relationship between the nucleotide divergence, *Dxy*, and the *Ka/Ks* ratio is shown in Fig. 4a ( $r = 0.916$ ;  $P < 0.001$ ). It indicated that the less divergent genes had suffering from more stringent selection pressure. Furthermore, as shown in Fig. 4b–c, the nucleotide divergence (*Dxy*) is correlated linearly with *Ka* and *Ks*. In addition, the correlation between *Dxy* and *Ka* ( $r = 0.990$ ;  $P < 0.001$ ) is much closer than that between *Dxy* and *Ks* ( $r = 0.901$ ;  $P < 0.001$ ). The stronger linear correlation between *Ka* and *Dxy* suggests that the rates of amino acid substitutions were more constant than the rates of synonymous substitutions following the divergence of genes in a short period of speciation time.

In summary, the *Ka/Ks* ratio in clear-orthologs of the three LRR-encoding gene classes was smaller than that of complex genes (Table 1). The NBS-LRR class exhibited a significantly greater *Ka/Ks* ratio than the other two classes (one-tailed *t*-test,  $P < 0.001$ ). As the  $\times\times L\times L\times\times$  motif was reported to be important for the specific recognition of pathogens (Ellis et al. 2000b); the *Ka/Ks* ratio in this region



**Fig. 4** Correlation and distribution of genetic parameters: **a** correlation between  $Dxy$  and  $Ka/Ks$ ; **b** between  $Dxy$  and  $Ks$ ; **c** between  $Dxy$  and  $Ka$ ; **d** distribution of  $Ka$  and  $Ka/Ks$  in the four gene classes in *At-Al*; **e**  $Ka$  and  $Ka/Ks$  in functionally different LRR-RLK genes with conserved HK or diversified NBS-LRR references. The values of  $Dxy$ ,  $Ks$ ,  $Ka$ , and  $Ka/Ks$  are adopted from Table 1 for **a–c**. The sizes of

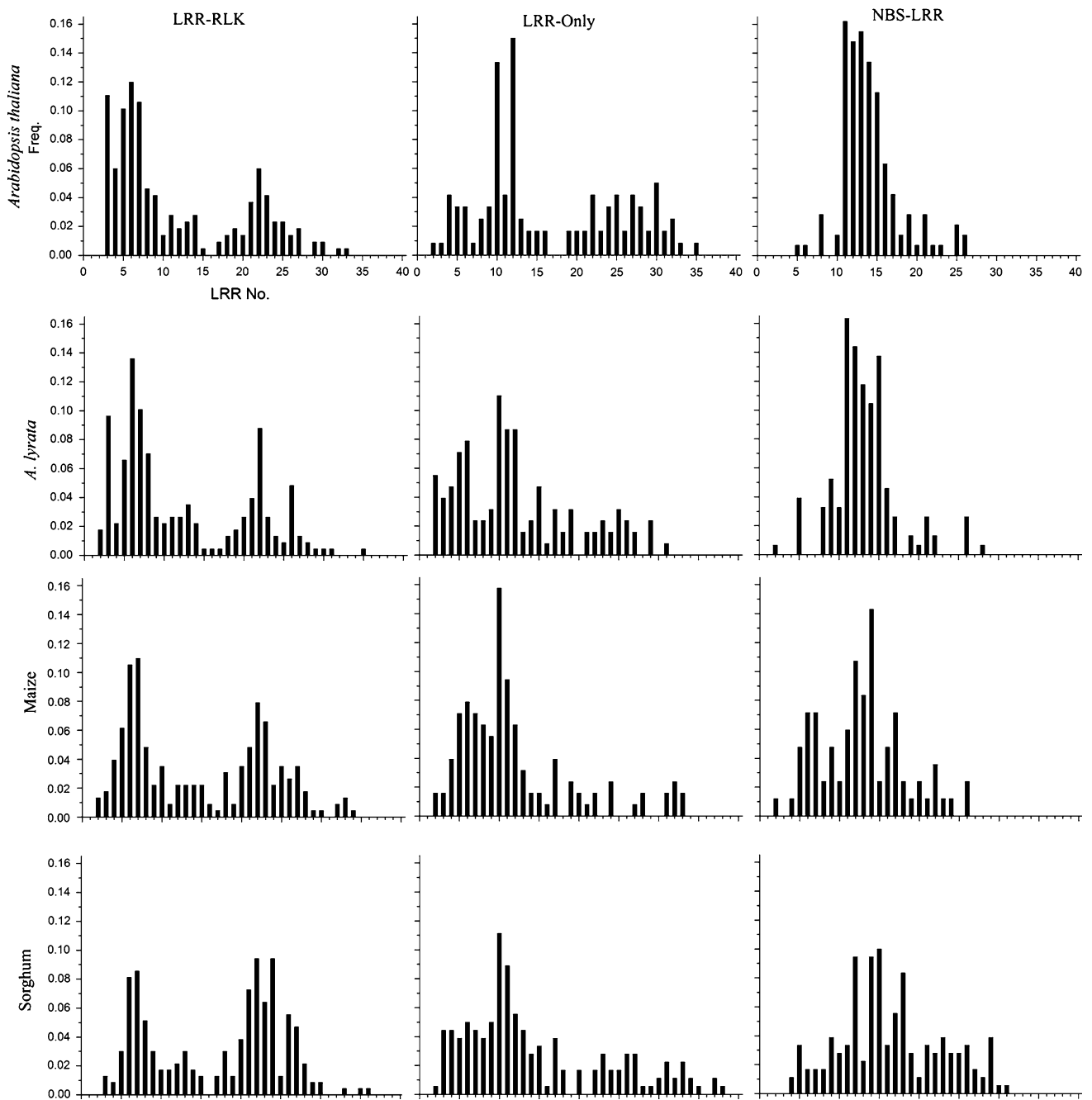
different colored circles were determined by including 90% of genes with the least areas, and the centers of circles were located at the points of average  $Ka$  and  $Ka/Ks$  values. HK, RLK-D, and RLK-R indicate housekeeping, development- and resistance-related LRR-RLK genes. The values of  $Ka$  and  $Ka/Ks$  were from a pairwise orthologous comparison of certain type of genes for **d–e**

(LRR core) was calculated separately. The  $Ka/Ks$  ratios in LRR core domain were greater than those in the coding sequence (CDS). The highest ratio ( $Ka/Ks = 1.431$ ) observed in complex families of NBS-LRR in *At-Al*, provided a sign of positive selection in NBS-LRR genes, although it was not significant ( $t$ -test,  $P = 0.128$ ). Meanwhile, the  $Ka/Ks$  substitution rate was also calculated based on pairwise ML analysis of PAML package (Table S5). The  $Ka/Ks$  ratios in complex families were generally larger than those in clear-ortholog families. Furthermore, NBS-LRR gene families showed larger  $Ka/Ks$  than the other two LRR-encoding gene classes. These results were consistent with the results calculated by Nei and Gojobori method. In addition, in order to avoid the bias from the least divergent ortholog pair estimates, we used the ML-based branch-specified model to estimate  $Ka$ ,  $Ks$ , and  $Ka/Ks$  in all the complex family members. The values of  $Ka$  and  $Ks$  were significantly larger from total complex family members than those from the least divergent ortholog pairs (two-tailed  $t$ -test,  $P < 0.01$ ). However, the  $Ka/Ks$  ratios were constant in the two sets of estimates.

Given that a lower  $Ka$  or  $Ka/Ks$  ratio in the constant coalescence time reflects the functional conservation of genes, the individual distribution of  $Ka$  or  $Ka/Ks$  could show the functional difference among the three classes of LRR-encoding genes. Indeed, when different color circles represent the distribution of 90% of the corresponding classes, the differences are clearly seen among those gene classes. From housekeeping genes to NBS-LRR genes, the sizes and ranges of circles increased gradually (Fig. 4d, Fig. S2a). These results indicate that the housekeeping gene class contained limited non-synonymous substitutions and had the strongest negative selection, which was consistent with the features of conserved families. In contrast, the other gene classes exhibited more non-synonymous substitutions.

#### Structural Difference of LRR Repeats Among Three Gene Classes

The LRR region characterized by leucine-rich repeats (LRRs) is shared by all three gene classes in this study. However, the number and distribution of LRRs has not



**Fig. 5** Frequency distribution of LRR repeat numbers in three gene classes. The x-axis indicates LRR numbers; and the y-axis indicates the frequency of genes with a number certain number of LRR motifs

been thoroughly investigated. Therefore, we analyzed the diversity of LRR numbers in all the predicted protein sequences from these three classes. Initially, MEME was used as described previously by two rounds of analysis. First, all the predicted protein sequences were analyzed together; then, sequences within each class were analyzed separately. Combined with the MEME analysis, we used LRR characteristics as defined previously (Kobe and Kajava 2001) to identify the individual repeats manually.

The pattern of copy number distribution is remarkably different among three classes of LRR-encoding genes (Fig. 5). In the LRR-RLK class, there are two peaks which concentrate on 5–7 and 21–24 LRRs, respectively. In general, the short LRRs genes are dominant in this class. In contrast, 11–15 LRRs genes are prevalent in NBS-LRR genes, which are the least common type in the other classes, and are particularly rare in LRR-RLK. The distribution in the LRR-only class seems to be the intermediate pattern



between the LRR-RLK and NBS-LRR classes. Overall, the short LRRs genes (10–12 repeats) are the major form in the LRR-only class. The average number of LRR repeats varies to a lesser degree among the three classes. For example, there were 11.61–17.06 LRRs in the LRR-RLK gene class from one species to another, 11.85–17.61 in the LRR-only class and 12.8–16.87 repeats in the NBS-LRR.

Further comparisons revealed the difference of LRR numbers between clear-ortholog and complex families of LRR-RLK in the *At* genome (Fig. S3). A large proportion of complex genes had rare LRR numbers, while clear-ortholog genes included both LRR-rich and LRR-rare classes. There were 28 complex genes that contained only three and four LRRs, which accounted for 52.8% (28/56) of the total number. Within these 28 complex genes, 25 genes were located within seven tandem repeats on chromosomes. This suggests that tandem duplications are the major cause for the expansion of three and four LRRs genes in *At*.

#### Resistance Candidate Genes in LRR-Encoding Genes

Our results demonstrate that the function of LRR-encoding genes is likely to be associated with their numeral, structural, genetic, and/or genomic features. In this scenario, this kind of association could be applicable to the gene families within a LRR-encoding gene class. LRR-RLK genes are known to be involved in both development and resistance (Shiu et al. 2004). Given that the genes in a family have a similar function, we classified the LRR-RLK gene class into two subclasses according to the functionally known genes in each family: the development-related subclass (RLK-D) and resistance-related subclass (RLK-R). Indeed, numerous differences were observed between these two subclasses. In RLK-D, the percentage of clear-orthologs in the total

number of genes of a subclass was 2.3–2.6 times higher than that of RLK-R (Table 2). The corresponding values of nucleotide divergence, synonymous, non-synonymous substitution rates, and *Ka/Ks* for RLK-R were significantly higher than those in RLK-D (two-tailed *t*-test,  $P < 0.05$ ), except for the synonymous substitution rate from *At–Al* (two-tailed *t*-test,  $P = 0.380$ ). When housekeeping and NBS-LRR genes were used as references (Table 2), RLK-D showed similar features to those of the housekeeping genes. In contrast, those parameters in RLK-R were more similar to the values in the NBS-LRR class.

These differences between RLK-D and RLK-R can be visually displayed by phylogenetic trees and scatter diagrams as the distribution profiles of *Ka* and *Ka/Ks*. Phylogenetic analysis of RLK-D shows its clear-ortholog relationship and stable copy numbers. By contrast, the phylogenetic tree of RLK-R reflects its complicated relationship between closely related species (Fig. S4). For example, *FLS2* and *PEPR*, involved in PTI, were located in clear-ortholog clades, as the RLK-D genes. However, the other RLK-R genes, such as *Xa21* and *Xa26* that are functionally important for specific recognition, were in lineage-specific clades. In general, RLK-R genes had much wider and higher *Ka* and *Ka/Ks* values (Fig. 4e, Fig. S2b). Therefore, a clear difference was displayed between these two subclasses.

## Discussion

#### Co-variation Among Major Classes of LRR-Encoding Genes

Although the gene size between the LRR-RLK and NBS-LRR classes was found to co-vary among a few selected

**Table 2** Characteristics of development- and resistance-related families in LRR-RLK gene class

Category	Gene no.		% <sup>c</sup>	<i>Dxy</i>	<i>Ks</i>	<i>Ka</i>	<i>Ka/Ks</i> (CDS)	<i>Ka/Ks</i> (LRR)
TL	At	Al						
House-keeping	52	54	98.2	0.052	0.157	0.023	0.134	
RLK-D <sup>a</sup>	36	37	95.9	0.050	0.155	0.020	0.125	0.152
RLK-R <sup>b</sup>	10	19	41.4	0.085	0.179	0.058	0.333	0.472
NBS-LRR	147	156	32.3	0.137	0.249	0.101	0.420	1.199
MS	Maize	Sorghum						
House-keeping	61	50	75.7	0.074	0.188	0.041	0.204	
RLK-D	41	42	43.4	0.081	0.203	0.046	0.221	0.281
RLK-R	34	63	16.5	0.157	0.358	0.104	0.306	0.511
NBS-LRR	95	184	13.6	0.117	0.277	0.079	0.331	0.460

TL, *A. thaliana* versus *A. lyrata*; MS, Maize versus sorghum

<sup>a</sup> Development-related genes in LRR-RLK gene class

<sup>b</sup> Resistance-related genes in LRR-RLK gene class

<sup>c</sup> Percentage of clear-ortholog genes in the total number of genes

species (Zhang et al. 2010), it is still surprising that such co-variation was found to be a general pattern not only among all sequenced plant genomes but also among all major LRR-gene classes. Such a phenomenon has been rarely reported in other gene classes (Demuth and Hahn 2009). Among the three major LRR classes, both their gene structures and cellular locations are quite different. Besides the LRR motif, these genes could contain a cytoplasmic kinase, a NBS motif, or the LRR motif exclusively. They could act as extracellular or cytoplasmic proteins. However, different gene classes co-vary significantly between each other in terms of their numbers. The co-variation could result from convergent effect following multiple rounds of whole-genome duplication in modern plants (Adams and Wendel 2005; Bowers et al. 2003; Tang et al. 2008). To evaluate such an effect, we identified duplicates following WGD or segment duplication in three LRR-encoding gene classes (Table S3). Indeed, the convergent effect was observed in the LRR-RLK gene class, in which the number of duplicates correlates significantly with the total number of LRR-RLK genes among the 10 analyzed species ( $r = 0.875$ ,  $P < 0.01$ ). However, when the duplicates were excluded, the number variations between any of the two LRR-encoding gene classes were still significantly correlated (Fig. S5). These results suggest that the co-variation revealed in this study was independent of individual WGD.

This co-variation suggests that all classes of LRR-encoding genes share some common functions. In this scenario, there are functional requirements for LRR-encoding genes to combine with different motifs or to locate at separate cellular sites, to achieve such functions. This may be true because all these classes contain disease resistance genes. To cope with various molecules from different pathogens, diverse gene structures and cellular sites are essential. This may explain why the pattern-recognition receptors (PRRs) are often found to be LRR-RLK genes, while the genes responsible for ETI commonly belong to the NBS-LRR class (Jones and Dangl 2006). This example demonstrates that they are doing different functions for the same purpose. To cooperate efficiently, there should have some balanced gene numbers among different gene classes. This kind of balance may be the underlying mechanism of co-variation among those gene classes. All of them evolve under the same pressure of pathogens, at least partly.

Such evolutionary pressure is believed to be different for certain species, and therefore, the numbers of LRR-encoding genes vary between species. Given that this mechanism is effective, the difference of gene numbers could be interpreted among species; however, it is still difficult to understand why there is a correlation between LRR-encoding genes and the total number of genes in a

species. This correlation might indicate that the strength of pathogenic pressure is positively correlated with the other stresses. Under strong stresses, a species should have more genes to cope with them. In this case, the increased total number of genes may be a consequence of these stresses. An expansion of LRR-encoding genes is likely to be a partial response to such an increase.

#### Association Between Evolutionary Pattern and Gene Function

The shared function among the three classes suggests that there may be an association of specific function with unique evolutionary characteristics. In other words, the genes with certain evolutionary parameters could have similar functions. Our analyses show that the genes, which are related to resistance or development, have a unique evolutionary pattern. Based on the functionally known genes, the most important parameters are numeral and genomic differences between resistance- or development-related genes. If the genes in a gene family have clear orthologous relationships and equal copy numbers between closely related species, they often have conserved development-related functions. In contrast, genes that vary in terms of copy numbers or are often distributed in species-specific manner are probably resistance-related genes.

The genetic differences are also important parameters to predict gene functions. Normally, resistance-related genes have larger nucleotide divergence, higher  $Ka/Ks$  ratio, and more frequent exchange of sequence information. This kind of evolutionary pattern was also detected commonly in NBS-LRR gene class (Chen et al. 2010). It suggests that the complex evolutionary patterns are general phenomena for the adaptively evolved gene family. The resistance genes face a variety of biotic factors, which make these genes co-evolve with pathogens. Consequently, the rapid divergence, high amino acid diversity, and frequent recombination are effective responses. On the contrary, the non-expanded genes with less divergence and lower  $Ka/Ks$  value, supposed to be under purifying selection, tend to have functions in plant growth and development (Fritz-Laylin et al. 2005; Krattinger et al. 2009; Tang et al. 2010; Yang et al. 2006). Development or growth genes often have critical functions. Therefore, they are more conservative in nucleotide sequences, and are also less prone to copy number variation in different species.

The structural differences may provide additional indication about gene function. It seems that the genes with resistant function prefer to have 11–15 LRRs per gene (e.g., the most NBS-LRR genes), although the relationship between LRR numbers and gene function is not clear yet. A high frequency of LRR-RLK centered on 21–24 LRRs. Interestingly, two *R*-genes identified in rice, *Xa21* and

*Xa26i* contain 22 and 26 LRRs, respectively. Compared with LRR numbers, however, it is clear that the combination of LRR with NBS tends to have resistance-related functions. To date, almost all of the NBS-LRR genes are found to be involved in disease resistance. When LRR combines with TM and kinase, there are a higher proportion of genes with growth and development functions. In this class, only a small proportion of LRR-RLK genes (39.2%) belong to the expanded group, which might be the candidates for *R* genes (Tang et al. 2010). The majority may be related to non-resistant functions. In addition, it was concluded that the majority of Cf-RLPs that belong to the LRR-only class are involved in resistance rather than development (Fritz-Laylin et al. 2005).

The association between evolutionary pattern and gene function could be useful in predicting gene function. For example, variable evolutionary patterns are found in the LRR-RLK gene class. Based on the evolutionary pattern in each family, function could then be predicted for genes within this family.

#### Origin of LRR-Encoding Genes

The LRR domain is an ancient and widespread domain. As a repetitive domain, the distribution of repeat numbers could be an important parameter to reflect the evolutionary history of these genes. Our analysis of the number distribution in Fig. 5 shows that each gene class favors having certain LRR numbers, e.g., 5–7 and 21–24 LRRs in LRR-RLK, 11–15 LRRs in NBS-LRR and 7–13 LRR-only genes. If all the classes of genes are analyzed together, it seems, then, that there is a plateau-like distribution from 5 to 25 LRRs, but with higher frequencies at 5 to 15 and 20 to 25 (shown by the black line in Fig. S6).

We hypothesized that all the LRR-encoding genes originated from the same ancestors, and in this case, we could expect a similar distribution in ancient species, such as bacteria, alga, fungi, or ancient land plants (e.g., *Physcomitrella patens* and *Selaginella moellendorffii*). After analyzing bacterial, algal, and fungal genes, and the genes from *P. patens* and *S. moellendorffii*, there were similar patterns of distribution but the patterns changed in a stepwise manner toward the pattern in land plants. The distribution of LRRs was still plateau-like with a higher frequency at 3–12 (the green line in Fig. S6). When species evolved to the stage of ancient land plants, the plateau-like line started to have higher frequency distribution at 5–14 and 20–23 LRRs, which was more consistent with that in angiosperms. All of these distribution patterns suggest the patterns in angiosperms were the result of stepwise evolution from ancient LRR genes and that the genes with long LRRs occurred later. It seems that certain numbers of LRRs prefer to combine with particular motifs, such as

11–15 LRRs combined with the NBS domain. Of course, a rapid recent expansion of some combination (e.g., NBS-LRR) could also explain the observed distributions. Perhaps both phenomena occurred during LRR-encoding gene evolution. Nevertheless, the LRR structures in different gene classes are most likely to evolve from the same ancestors, and the common LRR domain is likely to be the basis of co-variation among LRR-encoding gene classes and of the association between evolutionary pattern and gene function.

**Acknowledgments** This study was supported by the National Natural Science Foundation of China (31071062 & 30930049) and the National Key Project for Gene Transform in China (2009ZX08009-23B).

#### References

- Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. *Curr Opin Plant Biol* 8:135–141
- Bailey TL, Elkan C (1995) The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* 3:21–29
- Baumberger N, Doesseger B, Guyot R, Diet A, Parsons RL, Clark MA et al (2003) Whole-genome comparison of leucine-rich repeat extensins in Arabidopsis and rice. A conserved family of cell wall proteins form a vegetative and a reproductive clade. *Plant Physiol* 131:1313–1326
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438
- Chen Q, Han Z, Jiang H, Tian D, Yang S (2010) Strong positive selection drives rapid diversification of R-genes in Arabidopsis relatives. *J Mol Evol* 70:137–148
- Demuth JP, Hahn MW (2009) The life and death of gene families. *Bioessays* 31:29–39
- DeYoung BJ, Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat Immunol* 7:1243–1249
- DeYoung BJ, Bickle KL, Schrage KJ, Muskett P, Patel K, Clark SE (2006) The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in Arabidopsis. *Plant J* 45:1–16
- Ellis J, Dodds P, Pryor T (2000a) The generation of plant disease resistance gene specificities. *Trends Plant Sci* 5:373–379
- Ellis J, Dodds P, Pryor T (2000b) Structure, function and evolution of plant disease resistance genes. *Curr Opin Plant Biol* 3:278–284
- Faigon-Soverna A, Harmon FG, Storani L, Karayekov E, Staneloni RJ, Gassmann W et al (2006) A constitutive shade-avoidance mutant implicates TIR-NBS-LRR proteins in Arabidopsis photomorphogenic development. *Plant Cell* 18:2919–2928
- Fritz-Laylin LK, Krishnamurthy N, Tor M, Sjolander KV, Jones JD (2005) Phylogenomic analysis of the receptor-like proteins of rice and Arabidopsis. *Plant Physiol* 138:611–623
- Gomez-Gomez L, Boller T (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Mol Cell* 5:1003–1011
- Jones JD, Dangl JL (2006) The plant immune system. *Nature* 444:323–329
- Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JD (1994) Isolation of the tomato Cf-9 gene for resistance

- to *Cladosporium fulvum* by transposon tagging. *Science* 266: 789–793
- Kajava AV (1998) Structural diversity of leucine-rich repeat proteins. *J Mol Biol* 277:519–527
- Kobe B, Kajava AV (2001) The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* 11:725–732
- Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol* 17:1483–1498
- Krattinger SG, Lagudah ES, Spielmeier W, Singh RP, Huerta-Espino J, McFadden H et al (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science* 323:1360–1363
- Kunze G, Zipfel C, Robatzek S, Niehaus K, Boller T, Felix G (2004) The N terminus of bacterial elongation factor Tu elicits innate immunity in *Arabidopsis* plants. *Plant Cell* 16:3496–3507
- Kuroda H, Takahashi N, Shimada H, Seki M, Shinozaki K, Matsui M (2002) Classification and expression analysis of *Arabidopsis* F-box-containing protein genes. *Plant Cell Physiol* 43:1073–1085
- Li J, Chory J (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* 90:929–938
- Li J, Ding J, Zhang W, Zhang Y, Tang P, Chen JQ et al (2010) Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol Genet Genomics* 283:427–438
- Luo S, Peng J, Li K, Wang M, Kuang H (2011) Contrasting evolutionary patterns of the Rpl resistance gene family in different species of Poaceae. *Mol Biol Evol* 28:313–325
- McHale L, Tan X, Koehl P, Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. *Genome Biol* 7:212
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15:809–834
- Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. *Curr Opin Plant Biol* 8:129–134
- Morillo SA, Tax FE (2006) Functional analysis of receptor-like kinases in monocots and dicots. *Curr Opin Plant Biol* 9:460–469
- Nadeau JA, Sack FD (2002) Control of stomatal distribution on the *Arabidopsis* leaf surface. *Science* 296:1697–1700
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Osakabe Y, Maruyama K, Seki M, Satou M, Shinozaki K, Yamaguchi-Shinozaki K (2005) Leucine-rich repeat receptor-like kinase1 is a key membrane-bound regulator of abscisic acid early signaling in *Arabidopsis*. *Plant Cell* 17:1105–1119
- Padmanabhan M, Cournoyer P, Dinesh-Kumar SP (2009) The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. *Cell Microbiol* 11:191–198
- Scheideler M, Schlaich NL, Fellenberg K, Beissbarth T, Hauser NC, Vingron M et al (2002) Monitoring the switch from housekeeping to pathogen defense metabolism in *Arabidopsis thaliana* using cDNA arrays. *J Biol Chem* 277:10555–10561
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115
- Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH (2004) Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* 16:1220–1234
- Simillion C, Janssens K, Sterck L, Van de Peer Y (2008) i-ADHoRe 2.0: an improved tool to detect degenerated genomic homology using genomic profiles. *Bioinformatics* 24:127–128
- Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T et al (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270:1804–1806
- Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaça V, Bennetzen JL et al (2004) Close split of sorghum and maize genome progenitors. *Genome Res* 14:1916–1923
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. *Science* 320:486–488
- Tang P, Zhanga Y, Sun X, Tian D, Yang S, Ding J (2010) Disease resistance signature of the leucine-rich repeat receptor-like kinase genes in four plant species. *Plant Sci* 179:399–406
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Torii KU, Mitsukawa N, Oosumi T, Matsuura Y, Yokoyama R, Whittier RF et al (1996) The *Arabidopsis* ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell* 8:735–746
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Yang M, Sack FD (1995) The too many mouths and four lips mutations affect stomatal production in *Arabidopsis*. *Plant Cell* 7:2227–2239
- Yang S, Feng Z, Zhang X, Jiang K, Jin X, Hang Y et al (2006) Genome-wide investigation on the genetic variations of rice disease resistance genes. *Plant Mol Biol* 62:181–193
- Zhang M, Wu YH, Lee MK, Liu YH, Rong Y, Santos TS et al (2010) Numbers of genes in the NBS and RLK families vary by more than four-fold within a plant species and are regulated by multiple factors. *Nucleic Acids Res* 38:6513–6525
- Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K et al (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Genet Genomics* 271:402–415